

CCQM Microbial Identity 16S rRNA Interlaboratory Study

Supplemental Results

Nathan Olson

March 30, 2014

Contents

1	Biologically Conserved Positions	5
2	Biologically Variable Positions	7
3	Likely sets of variant combinations	10
4	Appendix	11

List of Figures

1	<i>L. monocytogenes</i> base ratios	8
2	<i>E. coli</i> base ratios	9

List of Tables

1	Biological conserved base quality statistics	5
2	<i>E. coli</i> variant combinations	10
3	<i>L. monocytogenes</i> variant combinations	10
4	<i>E. coli</i> positions pipeline comparison	11
5	<i>L. monocytogenes</i> positions pipeline pomparison	11

1 Biologically Conserved Positions

None of the variants for the biologically conserved positions were called using both variant callers, indicating the variants were potential false positives (Manuscript Table 2, tables 4 and 5). Consensus base quality statistics for biologically conserved positions (table 1).

Table 1: **Biologically Conserved Position Base Qualities** Characteristics of consensus based calls for conserved bases. Normalized quality values were obtained by dividing the quality values by the raw quality score assigned by GATK for each biologically conserved base position

Org	Plat	Lab	Rep	Raw Qual	Normalized	Min	Max
Ecoli	454	LGC	1	32792.22	0.65	0.45	1.46
Ecoli	454	LGC	2	32792.22	1.34	0.59	3.01
Ecoli	454	LGC	3	32792.22	0.74	0.47	2.67
Ecoli	454	NMIA	1	12602.23	2.69	0.38	3.02
Ecoli	ION	NIMC	1	1294.23	2.96	0.81	3.15
Ecoli	ION	NIST	1	1302.23	2.71	0.50	3.12
Ecoli	Sanger	ATCC	1	34.23	17.11	9.31	31.24
Ecoli	Sanger	ISP	1	31.24	31.23	-10.00	31.24
Ecoli	Sanger	LGC	1	169.23	3.60	0.51	3.97
Ecoli	Sanger	NIST	1	115.23	3.97	-1.43	10.06
Lmono	454	LGC	1	19276.23	2.55	1.01	8.65
Lmono	454	LGC	2	32792.22	0.65	0.41	2.04
Lmono	454	LGC	3	32792.22	0.74	0.43	2.29
Lmono	454	NMIA	1	13451.23	2.61	1.00	8.65
Lmono	ION	NIMC	1	1285.23	2.98	1.39	3.16
Lmono	ION	NIST	1	1404.23	2.78	1.51	3.04
Lmono	Sanger	ATCC	1	34.23	17.11	-10.00	31.24
Lmono	Sanger	ISP	1	34.23	17.11	-10.00	31.24
Lmono	Sanger	LGC	1	169.23	3.45	0.77	3.71

A number of false positive variant calls were due to low sequencing coverage because the targeted sequencing strategy was responsible for false positive variant calls in six of the eight “454” datasets. For those six datasets, a variant was called at the last position in the gap between the two sequencing regions, bases 940 and 963 relative to reference sequences for *E. coli* and *L. monocytogenes*, respectively. A 40 bp region that was not part of the targeted sequencing region had significantly lower median coverage than the targeted region (2 X vs. 30,110 X, respectively) for all “454” datasets combined (Main Paper Figures 1 and 2).

A number of false positive variants were called due to contaminants. A low level of contaminating reads (150) present in the LGC *L. monocytogenes* rep 1 dataset. A BLAST analysis of a representative of these reads indicated that they were from *E. coli* (E value of 0.0), a well known contaminant of molecular biology reagents (section 4). A number of *Escherichia coli* strains E value of 0.0. in LGC Lmono rep 1 454 dataset. False positive variant calls were also attributable to the sequencing strategy and the variant calling algorithm. Resulting in a number of variants called due to strand bias. Strand bias was identified as the cause of the false positive variant call because greater than 99% of the reads were covering the variant bases were in the same direction. The strand bias was a product of the amplicon-based sequencing. For the *E. coli* dataset the variant was at the 3 end of the region 1 amplicon (Main Paper Figure 1) and the variant in the *L. monocytogenes* was at the 5 end of the region 2 amplicon (Main Paper Figure 2). As a result a majority of the reads covering the variants were in a single direction as the reads in the other direction were not long enough to cover the variant. For whole genome sequencing data, read direction biases can indicate a systematic error. The UnifiedGenotyper variant caller takes into consideration strand bias resulting in the false positive variant calls and reports strand bias using the Fisher exact test statistic (see Main Paper Table 2). A filtering step is commonly performed when calling SNPs that would have identified these as false positive variants due to the low number of reads with the variant base.

2 Biologically Variable Positions

To determine the ratio of bases at the biologically variant positions, a novel Bayesian analysis based on binomial sampling theory was developed. According to the binomial distribution, the observed base ratios, while precise (due to high coverage), differed significantly from all potential copy ratios. Subsequently given the observed base ratios a Bayesian approach was used to identify the most probable copy ratio out of the possible abundant base ratios assuming *E. coli* and *L. monocytogenes* have seven and six 16S gene copies respectively (figs. 1 and 2).

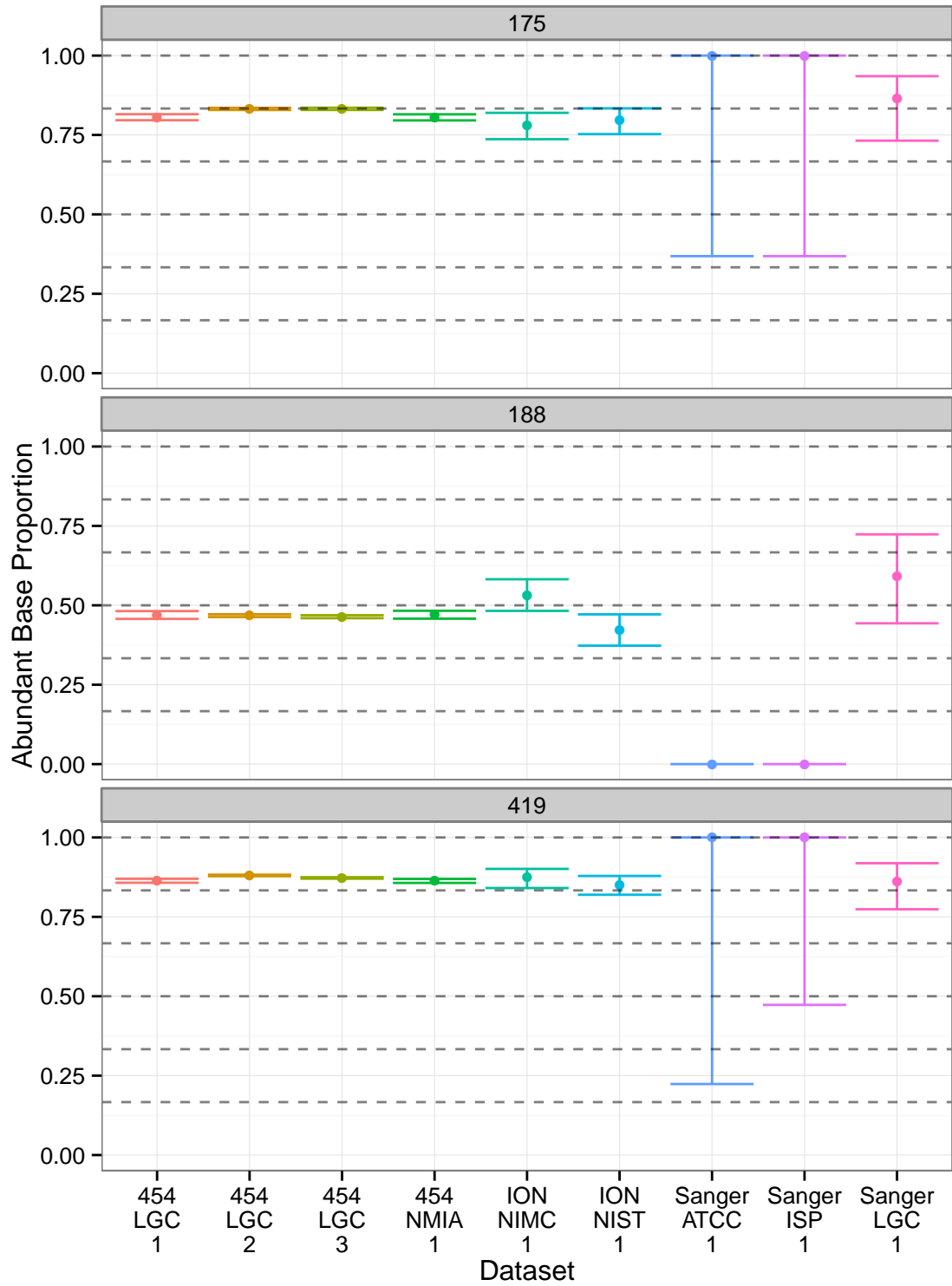


Figure 1: *L. monocytogenes* biologically variable position base ratios. Error bars represent the 95 % posterior credibility interval estimated from a beta binomial distribution. Grey dashed lines indicate the potential base ratios assuming seven gene copies.

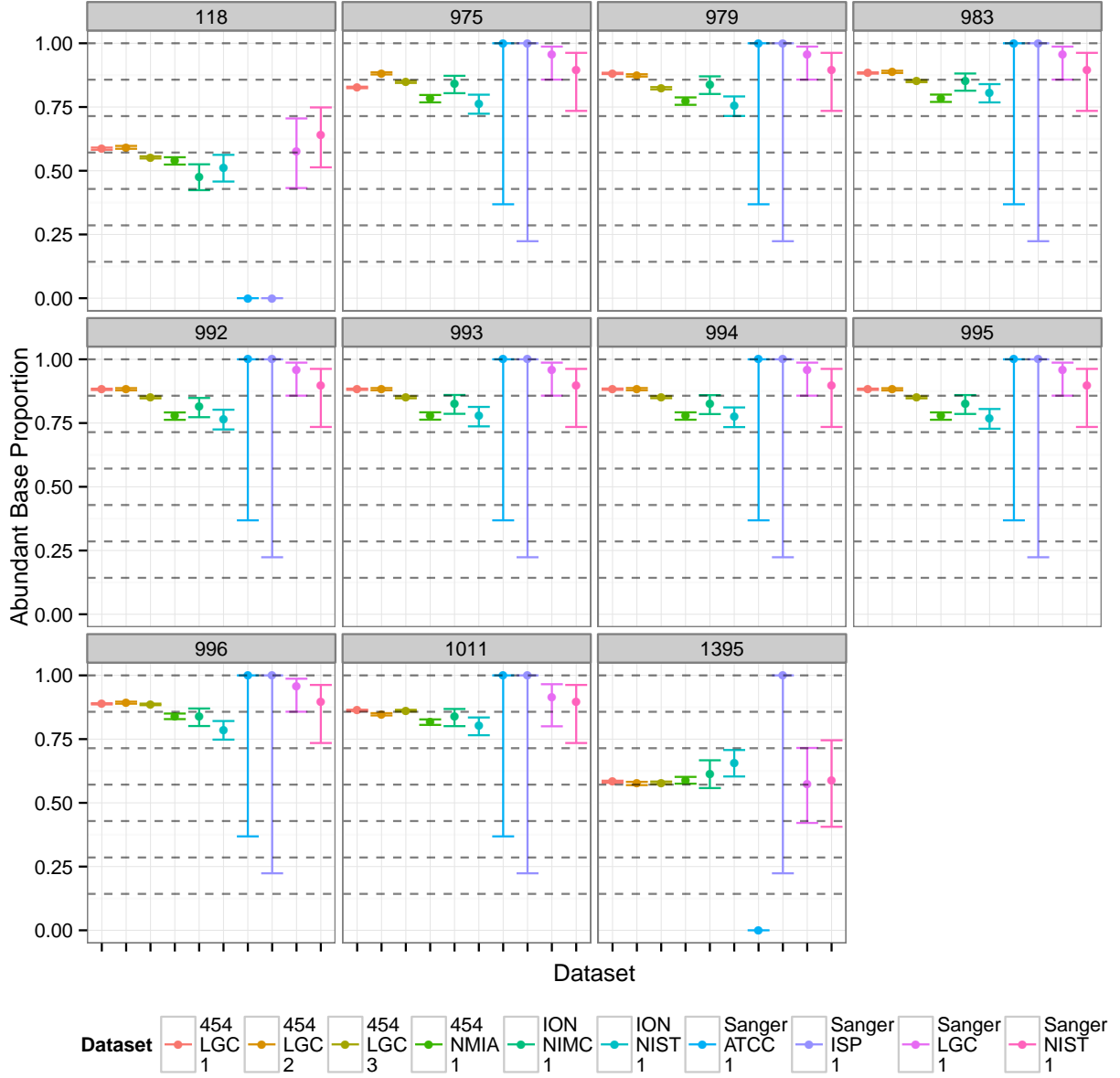


Figure 2: *E. coli* biologically variable position base ratios. Error bars represent the 95 % posterior credibility interval estimated from a beta binomial distribution. Grey dashed lines indicate the potential base ratios assuming seven gene copies.

3 Likely sets of variant combinations

Most likely combination of variant strings for “454” and Sanger Clone library datasets table 2 and table 3).

Table 2: **Estimated most likely set of variant combinations for *E. coli***

dataset	likelihood	chimera	ACCGATTGTA	ACCGATTGTG	GGTAGAATCA
Ecoli-454-LGC-1	0.04	353.03	3	3	1
Ecoli-454-LGC-2	0.03	258.60	3	3	1
Ecoli-454-LGC-3	0.08	33.30	3	3	1
Ecoli-454-NMIA-1	0.10	13.50	3	3	1
Ecoli-LGC-Sanger-Clones.csv	0.05	3.55	3	4	0
Ecoli-NIST-Sanger-Clones.csv	0.12	4.54	3	4	0
Consensus	0.04	342.58	3	3	1

Table 3: **Estimated most likely set of variant combinations for *L. monocytogenes***

dataset	likelihood	chimera	GTA	GTG	TCA	TCG
Lmono-454-LGC-1	0.00	63.60	2	2	1	1
Lmono-454-LGC-2	0.01	658.52	2	2	1	1
Lmono-454-LGC-3	0.01	410.99	2	2	1	1
Lmono-454-NMIA-1	0.00	62.80	2	2	1	1
Lmono-LGC-Sanger-Clones.csv	0.00	5.13	2	2	1	1
Lmono-NIST_Sanger-Clones.csv	0.01	8.38	2	2	1	1
Consensus	0.01	1009.81	2	2	1	1

4 Appendix

Full List of False Positive Variants

All variants called by the 8 pipelines used during the pipeline validation along with the suspected cause of the variant.

Table 4: ***E. coli* Pipeline Comparison** Characteristics of variant calls for different bioinformatic pipelines.

Org	Plat	Lab	Rep	Map	Var	POS	DP	QUAL	MQ	FS	Cause
Ecoli	454	LGC	1	TMAP	gatk	325	250	82.77	89.18	61.10	Strand bias
Ecoli	454	LGC	1	bwa	sam	396	2545	70.00	60.00		End of read
Ecoli	454	LGC	1	TMAP	sam	396	3013	72.00	56.00		End of read
Ecoli	454	LGC	1	bwa	gatk	940	19	215.77	60.00	28.54	Non-target region
Ecoli	454	LGC	1	TMAP	gatk	940	19	215.77	85.03	28.54	Non-target region
Ecoli	454	LGC	1	bwa	gatk	959	250	1274.77	60.00	11.77	End of read
Ecoli	454	LGC	2	bwa	gatk	106	250	92.77	60.00	0.00	End of read
Ecoli	454	LGC	2	bwa	gatk	959	250	602.77	59.98	17.81	End of read
Ecoli	454	LGC	3	bwa	gatk	325	250	38.77	59.83	57.10	End of read
Ecoli	454	LGC	3	bwa	gatk	348	250	634.77	59.86	14.01	End of read
Ecoli	454	LGC	3	bwa	sam	417	1032	38.00	60.00		Homopolymer
Ecoli	454	LGC	3	TMAP	sam	417	1020	39.00	58.00		Homopolymer
Ecoli	454	LGC	3	bwa	gatk	940	9	91.05	60.00	6.99	Non-target region
Ecoli	454	LGC	3	TMAP	gatk	940	10	91.05	84.55	6.99	Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	313	250	5383.77	78.77	459.55	Strand bias
Ecoli	454	NMIA	1	TMAP	gatk	508	250	834.77	83.88	0.00	End of read
Ecoli	454	NMIA	1	TMAP	gatk	509	250	875.77	83.88	0.00	End of read
Ecoli	454	NMIA	1	TMAP	gatk	510	250	929.77	83.88	0.00	End of read
Ecoli	454	NMIA	1	TMAP	gatk	514	250	784.77	83.88	0.00	End of read
Ecoli	454	NMIA	1	TMAP	sam	514	6337	86.00	59.00		End of read
Ecoli	454	NMIA	1	TMAP	gatk	901	208	8198.77	84.53	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	904	208	8152.77	84.53	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	934	250	9424.77	75.21	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	935	250	9425.77	75.21	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	938	250	9335.77	75.02	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	sam	938	2746	9.54	56.00		Non-target region
Ecoli	454	NMIA	1	TMAP	gatk	939	250	9128.77	75.02	0.00	Non-target region
Ecoli	454	NMIA	1	TMAP	sam	939	2746	15.20	60.00		Non-target region
Ecoli	454	NMIA	1	TMAP	sam	941	2746	9.52	55.00		Non-target region
Ecoli	ION	NIMC	1	TMAP	sam	1463	168	22.50	60.00		End of reference
Ecoli	Sanger	NIST	1	TMAP	sam	1463	29	139.00	60.00		End of reference
Ecoli	Sanger	NIST	1	TMAP	sam	1464	29	214.00	60.00		End of reference

Table 5: ***L. monocytogenes* Positions Pipeline Comparison** Characteristics of variant calls for different bioinformatic pipelines.

Org	Plat	Lab	Rep	Map	Var	POS	DP	QUAL	MQ	FS	Cause
Lmono	454	LGC	1	bwa	gatk	315	250	4949.77	46.30	85.97	Strand bias
Lmono	454	LGC	1	bwa	gatk	328	250	5119.77	46.34	80.09	Strand bias
Lmono	454	LGC	1	TMAP	gatk	334	250	4082.77	70.33	294.21	Strand bias
Lmono	454	LGC	1	bwa	gatk	508	250	1480.77	51.44	15.48	Contaminants
Lmono	454	LGC	1	bwa	gatk	533	166	1419.77	46.35	0.00	Contaminants

Lmono	454	LGC	1	bwa	sam	533	1762	176.00	38.00		Contaminants
Lmono	454	LGC	1	TMAP	gatk	533	250	2270.77	51.58	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	536	250	1859.77	51.58	4.25	Contaminants
Lmono	454	LGC	1	bwa	gatk	537	166	1362.77	46.35	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	537	1762	100.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	538	166	1705.77	46.35	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	538	1622	163.00	38.00		Contaminants
Lmono	454	LGC	1	TMAP	gatk	539	250	1942.77	51.58	4.39	Contaminants
Lmono	454	LGC	1	bwa	gatk	548	166	1991.77	46.35	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	548	1762	221.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	549	166	1979.77	46.35	0.83	Contaminants
Lmono	454	LGC	1	bwa	sam	549	1762	189.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	550	166	1949.77	46.35	0.80	Contaminants
Lmono	454	LGC	1	bwa	sam	550	1762	218.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	555	167	1933.77	46.45	0.81	Contaminants
Lmono	454	LGC	1	bwa	sam	555	1763	222.00	37.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	559	167	1959.77	46.45	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	559	1762	222.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	574	168	2070.77	46.34	0.79	Contaminants
Lmono	454	LGC	1	bwa	sam	574	1764	189.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	585	166	1526.77	46.54	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	585	1736	222.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	587	166	1644.77	46.54	0.82	Contaminants
Lmono	454	LGC	1	bwa	sam	587	1740	209.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	595	166	2013.77	46.54	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	595	1740	222.00	38.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	677	250	4981.77	58.51	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	677	4535	222.00	58.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	700	249	5253.77	56.78	13.26	Contaminants
Lmono	454	LGC	1	bwa	sam	700	4614	222.00	58.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	703	249	4381.77	56.78	4.39	Contaminants
Lmono	454	LGC	1	bwa	sam	703	4614	222.00	58.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	712	249	4734.77	56.78	9.10	Contaminants
Lmono	454	LGC	1	bwa	sam	712	4612	222.00	58.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	716	249	3824.77	56.78	1.19	Contaminants
Lmono	454	LGC	1	bwa	sam	716	4612	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	729	223	3226.07	59.12	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	729	4400	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	731	223	2379.77	59.12	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	731	4377	112.00	60.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	733	223	2457.77	59.12	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	733	4401	222.00	60.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	738	223	3785.77	59.12	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	738	4384	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	740	223	3911.77	59.12	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	740	4390	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	741	224	3901.77	59.10	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	741	4394	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	742	224	3916.77	59.10	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	742	4400	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	743	224	4026.77	59.10	0.00	Contaminants
Lmono	454	LGC	1	bwa	sam	743	4400	222.00	59.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	753	250	1583.77	55.75	44.19	Contaminants
Lmono	454	LGC	1	bwa	sam	753	4613	222.00	59.00		Contaminants

Lmono	454	LGC	1	bwa	gatk	757	250	2654.77	55.75	154.47	Contaminants
Lmono	454	LGC	1	bwa	sam	757	4613	222.00	59.00		Contaminants
Lmono	454	LGC	1	TMAP	gatk	924	187	1838.77	44.01	0.00	Contaminants
Lmono	454	LGC	1	TMAP	sam	924	616	6.19	11.00		Contaminants
Lmono	454	LGC	1	TMAP	gatk	926	187	1954.77	44.01	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	928	187	1913.77	44.01	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	930	187	1830.77	44.01	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	952	250	499.77	53.84	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	953	250	33179.29	53.84	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	954	250	468.77	53.84	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	955	250	3643.77	53.84	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	957	250	33182.29	53.86	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	958	250	4897.77	53.86	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	959	250	3563.77	53.86	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	961	250	3566.77	53.86	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	962	250	509.77	53.86	0.00	Contaminants
Lmono	454	LGC	1	TMAP	gatk	963	250	33182.29	53.86	0.00	Contaminants
Lmono	454	LGC	1	bwa	gatk	982	250	1846.77	60.00	677.62	Contaminants
Lmono	454	LGC	1	TMAP	gatk	982	250	448.77	56.52	2.54	Contaminants
Lmono	454	LGC	1	TMAP	sam	997	6560	19.10	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1018	6693	4.77	41.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1035	6753	4.77	41.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1036	6754	4.13	42.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1047	250	3244.77	59.39	22.08	Contaminants
Lmono	454	LGC	1	bwa	sam	1047	8006	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1047	8006	23.00	43.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1055	250	3148.77	59.41	4.04	Contaminants
Lmono	454	LGC	1	bwa	sam	1055	8011	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1055	8008	30.00	43.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1072	250	3166.77	59.06	2.38	Contaminants
Lmono	454	LGC	1	bwa	sam	1072	8022	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1072	8013	11.30	44.00		Contaminants
Lmono	454	LGC	1	bwa	sam	1077	7975	215.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1077	7932	19.10	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1113	7985	10.40	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1114	8010	34.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1115	7987	28.00	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1116	8011	28.00	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1145	8009	17.10	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1148	8011	7.80	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1150	8011	4.77	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1158	8011	9.52	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1192	250	5548.77	58.66	43.31	Contaminants
Lmono	454	LGC	1	bwa	sam	1192	8008	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1192	8011	24.00	43.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1201	250	4652.77	58.66	46.30	Contaminants
Lmono	454	LGC	1	bwa	sam	1201	8006	206.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1201	8008	4.13	43.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1208	250	3745.77	58.66	14.75	Contaminants
Lmono	454	LGC	1	bwa	sam	1208	8009	134.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1208	8012	15.10	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1213	250	4297.77	58.66	48.89	Contaminants
Lmono	454	LGC	1	bwa	sam	1213	8010	223.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1213	8012	25.00	44.00		Contaminants

Lmono	454	LGC	1	TMAP	sam	1237	8002	24.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1238	8004	11.30	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1248	8004	21.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1251	8004	13.20	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1257	8004	27.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1258	8004	4.13	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1265	8004	4.77	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1266	8004	10.40	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1268	8004	16.10	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1272	8002	7.80	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1273	8002	6.20	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1275	7999	23.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1277	8004	5.46	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1284	8004	27.00	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1285	8004	3.54	43.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1286	8004	23.00	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1287	8004	17.10	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1288	8000	10.40	44.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1291	8004	17.10	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1304	249	5021.77	59.83	67.31	Contaminants
Lmono	454	LGC	1	bwa	sam	1304	7998	187.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1304	8005	26.00	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1307	250	5365.77	59.83	74.40	Contaminants
Lmono	454	LGC	1	bwa	sam	1307	7999	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1307	8005	7.80	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1318	250	5089.77	59.78	83.81	Contaminants
Lmono	454	LGC	1	bwa	sam	1318	8002	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1318	8006	17.10	44.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1321	250	4776.77	59.78	80.08	Contaminants
Lmono	454	LGC	1	bwa	sam	1321	8002	219.00	60.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1329	250	4978.77	59.72	80.67	Contaminants
Lmono	454	LGC	1	bwa	sam	1329	8005	225.00	60.00		Contaminants
Lmono	454	LGC	1	bwa	gatk	1356	249	3948.77	59.63	31.03	Contaminants
Lmono	454	LGC	1	bwa	sam	1356	8010	225.00	60.00		Contaminants
Lmono	454	LGC	1	TMAP	sam	1356	8005	15.10	44.00		Contaminants
Lmono	454	LGC	2	bwa	gatk	315	250	3407.77	51.57	0.00	Contaminants
Lmono	454	LGC	2	bwa	gatk	328	250	3430.77	51.57	0.00	Contaminants
Lmono	454	LGC	2	TMAP	gatk	347	250	144.77	91.85	62.73	Strand bias
Lmono	454	LGC	2	bwa	gatk	963	122	628.78	60.00	0.00	Non-target region
Lmono	454	LGC	2	TMAP	gatk	963	21	54.05	74.82	11.76	Non-target region
Lmono	454	LGC	2	bwa	gatk	1047	250	3642.77	59.92	2.94	Contaminants
Lmono	454	LGC	2	bwa	gatk	1055	250	3403.77	59.92	5.74	Contaminants
Lmono	454	LGC	2	bwa	gatk	1072	250	2432.77	59.92	9.74	Contaminants
Lmono	454	LGC	2	bwa	gatk	1077	249	2055.77	59.92	9.74	Contaminants
Lmono	454	LGC	2	bwa	gatk	1192	250	4319.77	59.63	11.73	Contaminants
Lmono	454	LGC	2	bwa	gatk	1201	250	4239.77	59.63	6.44	Contaminants
Lmono	454	LGC	2	bwa	gatk	1208	250	4215.77	59.63	9.57	Contaminants
Lmono	454	LGC	2	bwa	gatk	1213	250	4229.77	59.63	9.53	Contaminants
Lmono	454	LGC	2	bwa	gatk	1304	250	4197.77	60.00	9.64	Contaminants
Lmono	454	LGC	2	bwa	gatk	1307	250	4324.77	60.00	11.73	Contaminants
Lmono	454	LGC	2	bwa	gatk	1318	250	4258.77	60.00	11.93	Contaminants
Lmono	454	LGC	2	bwa	gatk	1321	250	4187.77	60.00	11.93	Contaminants
Lmono	454	LGC	2	bwa	gatk	1329	250	4346.77	60.00	12.14	Contaminants
Lmono	454	LGC	2	bwa	gatk	1356	250	4254.77	60.00	9.87	Contaminants

Lmono	454	LGC	3	bwa	gatk	963	111	302.48	60.00	0.00	Non-target region
Lmono	454	LGC	3	TMAP	gatk	963	10	78.77	67.25	6.02	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	330	250	4404.77	81.62	633.82	Strand bias
Lmono	454	NMIA	1	TMAP	gatk	334	250	4362.77	81.62	659.56	Strand bias
Lmono	454	NMIA	1	TMAP	gatk	335	250	3790.77	81.62	659.56	Strand bias
Lmono	454	NMIA	1	bwa	gatk	381	250	91.77	60.00	16.13	End of read
Lmono	454	NMIA	1	TMAP	gatk	533	249	1263.77	69.42	0.00	End of read
Lmono	454	NMIA	1	TMAP	gatk	932	95	3636.77	75.01	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	936	93	3633.77	75.42	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	954	250	8799.77	79.13	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	957	250	8734.77	79.13	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	961	250	8757.77	79.13	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	962	250	8716.77	79.13	0.00	Non-target region
Lmono	454	NMIA	1	TMAP	gatk	963	250	8669.77	79.13	0.00	Non-target region
Lmono	ION	NIST	1	TMAP	gatk	792	273	189.77	85.59	66.65	Strand bias
Lmono	Sanger	LGC	1	bwa	sam	390	81	25.50	60.00		End of read
Lmono	Sanger	LGC	1	bwa	sam	1409	44	13.70	60.00		End of read
Lmono	Sanger	LGC	1	TMAP	sam	1505	41	71.20	60.00		End of reference
Lmono	Sanger	LGC	1	TMAP	sam	1506	41	71.20	60.00		End of reference

Contaminants - BLAST results

BLAST reports for representative sequences of reads responsible for false positive variant calls in the LGC *L. monocytogenes* "454" rep 1 dataset.

BLASTN 2.2.29+

Reference: Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", J Comput Biol 2000; 7(1-2):203-14.

RID: KH9SY3U8014

Database: Representative Chromosomes

2,857 sequences; 5,609,140,793 total letters

Query=

Length=558

Sequences producing significant alignments:		Score (Bits)	E Value
ref NC_000913.3	Escherichia coli str. K-12 substr. MG1655, c...	979	0.0
ref NC_018658.1	Escherichia coli O104:H4 str. 2011C-3493 chr...	979	0.0
ref NC_017634.1	Escherichia coli O83:H1 str. NRG 857C chromo...	979	0.0
ref NC_011751.1	Escherichia coli UMN026 chromosome, complete...	979	0.0
ref NC_011750.1	Escherichia coli IAI39 chromosome, complete ...	979	0.0
ref NC_011740.1	Escherichia fergusonii ATCC 35469 chromosome...	979	0.0
ref NC_007384.1	Shigella sonnei Ss046 chromosome, complete g...	979	0.0
ref NC_002695.1	Escherichia coli O157:H7 str. Sakai chromoso...	979	0.0
ref NC_004337.2	Shigella flexneri 2a str. 301 chromosome, co...	974	0.0
ref NC_007613.1	Shigella boydii Sb227 chromosome, complete g...	974	0.0

ALIGNMENTS

>ref|NC_000913.3| Escherichia coli str. K-12 substr. MG1655, complete genome
Length=4641652

Features in this part of subject sequence:

rRNA-16S ribosomal RNA of rrnH operon

Score = 979 bits (530), Expect = 0.0

Identities = 539/543 (99%), Gaps = 2/543 (0%)

Strand=Plus/Plus

Query	3	CCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCTTACGGGTTGT-AAGTACGTTTCAGC	61
Sbjct	224155	CCTGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTAC-TTTCAGC	224213
Query	62	GGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCAC	121
Sbjct	224214	GGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAGCAC	224273
Query	122	CGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTA	181
Sbjct	224274	CGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAATCGGAATTA	224333
Query	182	CTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAAC	241
Sbjct	224334	CTGGGCGTAAAGCGCACGCAGGCGGTTTGTTAAGTCAGATGTGAAATCCCCGGGCTCAAC	224393
Query	242	CTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGG	301
Sbjct	224394	CTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGGGGGGTAGAATTCCAGG	224453
Query	302	TGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGA	361
Sbjct	224454	TGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCGGCCCCCTGGA	224513
Query	362	CGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAG	421
Sbjct	224514	CGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAG	224573
Query	422	TCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCT	481
Sbjct	224574	TCCACGCCGTAAACGATGTCGACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCT	224633
Query	482	AACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAACTCAAATGAATTGA	541
Sbjct	224634	AACGCGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAACTCAAATGAATTGA	224693
Query	542	CGG	544
Sbjct	224694	CGG	224696

Database: Representative Chromosomes

Posted date: Mar 21, 2014 12:17 AM
Number of letters in database: 5,609,140,793
Number of sequences in database: 2,857

Lambda	K	H
1.33	0.621	1.12

Gapped

Lambda	K	H
1.28	0.460	0.850

Matrix: blastn matrix:1 -2

Gap Penalties: Existence: 0, Extension: 0

Number of Sequences: 2857

Number of Hits to DB: 6177

Number of extensions: 6

Number of successful extensions: 6

Number of sequences better than 10: 1

Number of HSP's better than 10 without gapping: 0

Number of HSP's gapped: 3

Number of HSP's successfully gapped: 3

Length of query: 558

Length of database: 5609140793

Length adjustment: 30

Effective length of query: 528

Effective length of database: 5609055083

Effective search space: 2961581083824

Effective search space used: 2961581083824

A: 0

X1: 13 (25.0 bits)

X2: 32 (59.1 bits)

X3: 54 (99.7 bits)

S1: 13 (25.1 bits)

S2: 21 (39.9 bits)