NanoOK report for UCSC_MARC_Phase_Ia_Run_2

Pass and fail counts

Type	Pass	Fail
Template	0	49464
Complement	0	20940
2D	0	7538

Read lengths

Type	NumReads	TotalBases	Mean	Longest	Shortest	N50	N50Count	N90	N90Count
Template	49464	306859868	6203.70	210796	5	8525	13014	4055	32278
Complement	20940	126484843	6040.35	136833	10	8285	5516	3810	13809
2D	7538	48411005	6422.26	37842	115	8448	2132	3818	5251
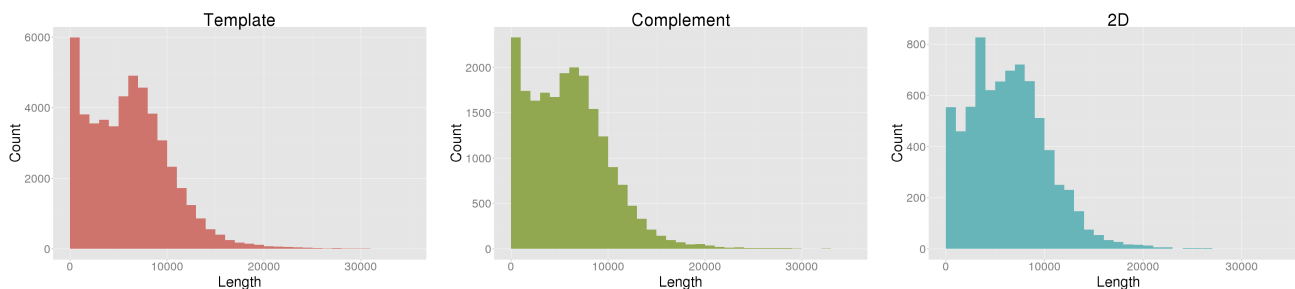
Template alignments

Number of reads	49464	
Number of reads with alignments	27081	(54.75%)
Number of reads without alignments	22383	(45.25%)

ID	Size	Number of Reads	% of Reads	Mean read length	Aligned bases	Mean coverage	Longest Perf Kmer
Control sequence	3560	943	1.91	2851.42	2412728	677.73	52
Escherichia coli	4641652	26138	52.84	7795.08	175625228	37.84	78

Complement alignments

Number of reads	20940	
Number of reads with alignments	11486	(54.85%)
Number of reads without alignments	9454	(45.15%)

ID	Size	Number of Reads	% of Reads	Mean read length	Aligned bases	Mean coverage	Longest Perf Kmer
Control sequence	3560	370	1.77	2682.67	862387	242.24	40
Escherichia coli	4641652	11116	53.09	7357.51	70456923	15.18	63

2D alignments

Number of reads	7538	
Number of reads with alignments	5884	(78.06%)
Number of reads without alignments	1654	(21.94%)

ID	Size	Number of Reads	% of Reads	Mean read length	Aligned bases	Mean coverage	Longest Perf Kmer
Control sequence	3560	315	4.18	2874.39	917282	257.66	160
Escherichia coli	4641652	5569	73.88	7309.84	40874051	8.81	237

Control sequence error analysis

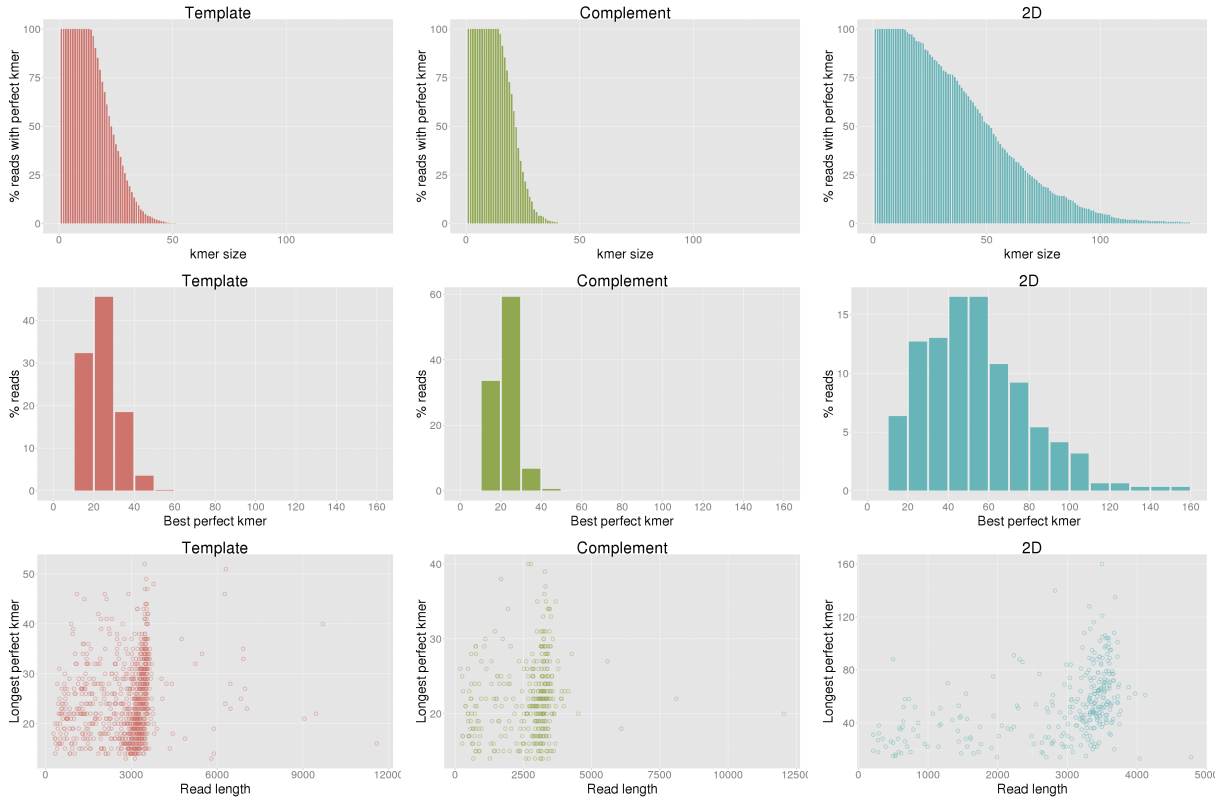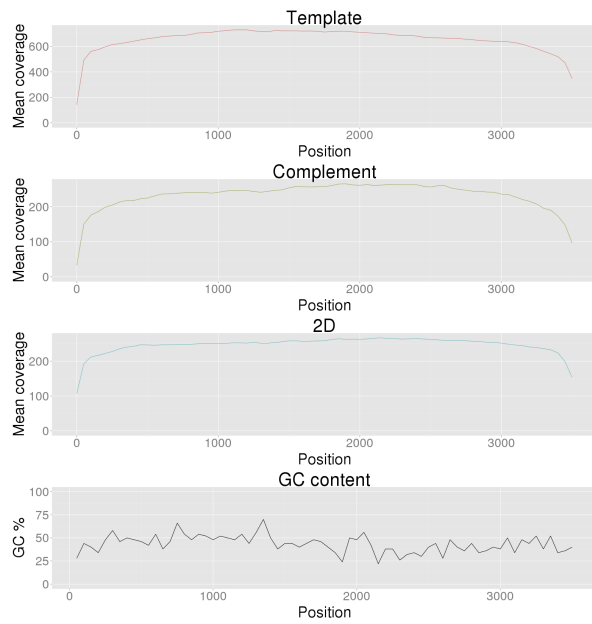|	Template	Complement	2D
Overall base identity (excluding indels)	59.39%	56.47%	76.99%
Aligned base identity (excluding indels)	79.45%	79.21%	87.07%
Identical bases per 100 aligned bases (including indels)	66.19%	65.00%	75.99%
Inserted bases per 100 aligned bases (including indels)	4.90%	4.53%	4.57%
Deleted bases per 100 aligned bases (including indels)	11.80%	13.42%	8.15%
Substitutions per 100 aligned bases (including indels)	17.12%	17.06%	11.28%
Mean insertion size	1.57	1.53	1.52
Mean deletion size	1.73	1.82	1.63
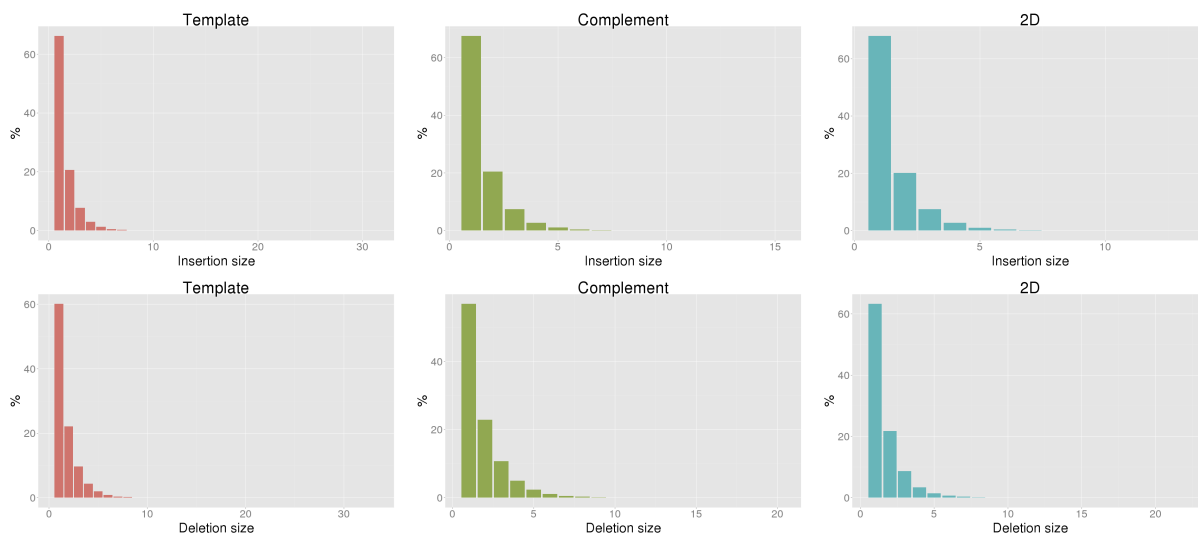
Control sequence read identity
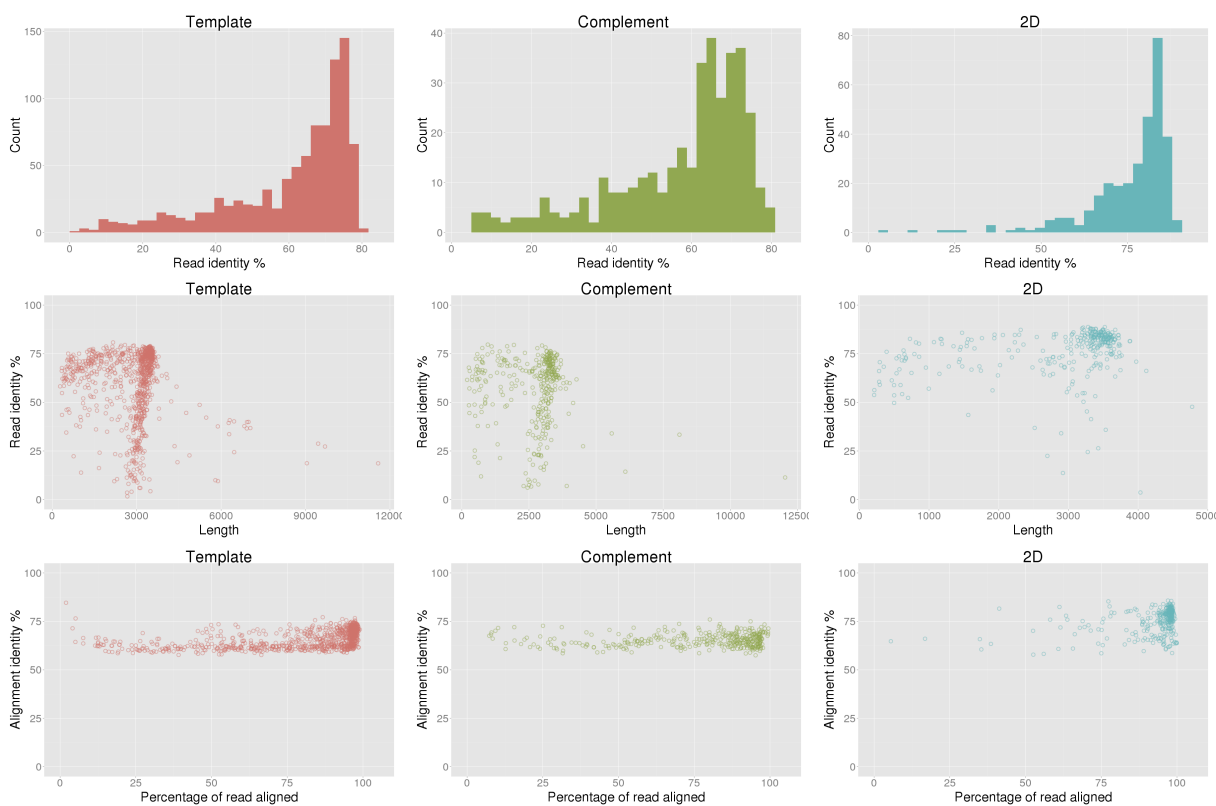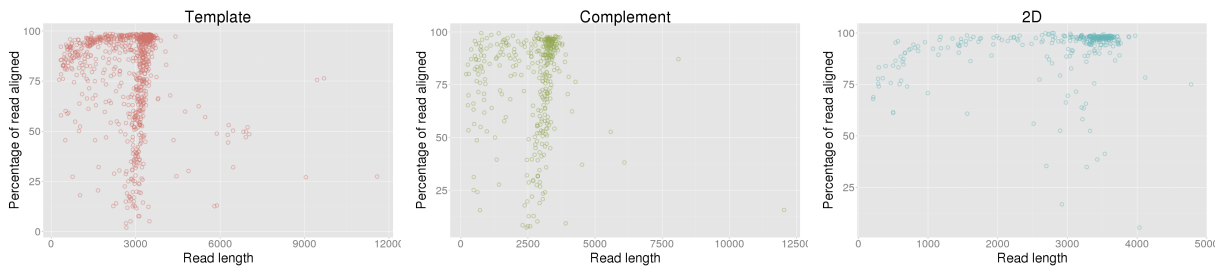
## Control sequence perfect kmers

Control sequence coverage

Control sequence 5-mer analysis

Under-represented 5-mers

		Template				Complement				2D		
Rank	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %
1	TTTTT	0.759	0.138	-0.621	TTTTT	0.759	0.077	-0.682	TTTTT	0.759	0.044	-0.714
2	AAAAA	0.478	0.088	-0.390	AAAAA	0.478	0.047	-0.430	AAAAA	0.478	0.041	-0.437
3	TGATG	0.393	0.131	-0.262	TGATG	0.393	0.160	-0.233	TGATG	0.393	0.164	-0.229
4	GATGT	0.309	0.091	-0.218	AAAAC	0.337	0.106	-0.231	AAAAC	0.337	0.141	-0.196
5	AAAAC	0.337	0.120	-0.217	GATGT	0.309	0.097	-0.212	GATGT	0.309	0.131	-0.178
6	CTGAT	0.309	0.110	-0.199	GCAAT	0.309	0.127	-0.182	CTTTT	0.253	0.101	-0.152
7	GCTGA	0.281	0.109	-0.172	TTATC	0.309	0.130	-0.179	CTGAT	0.309	0.158	-0.151
8	GCAAT	0.309	0.139	-0.170	AACAA	0.281	0.104	-0.177	TTATC	0.309	0.165	-0.144
9	AATAT	0.309	0.139	-0.170	CTGAT	0.309	0.133	-0.176	GCAAT	0.309	0.169	-0.140
10	TAATA	0.281	0.114	-0.166	AATAT	0.309	0.141	-0.168	AATAT	0.309	0.174	-0.135

Over-represented 5-mers

		Template				Complement				2D		
Rank	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %
1	CTTTG	0.028	0.143	0.115	TACTT	0.000	0.124	0.124	CTTTG	0.028	0.126	0.097
2	TACTT	0.000	0.114	0.114	ACCCC	0.000	0.123	0.123	GCATC	0.084	0.179	0.095
3	ATCTA	0.000	0.113	0.113	CTTTG	0.028	0.150	0.122	TCAGC	0.028	0.121	0.093
4	GCTCC	0.000	0.111	0.111	TCAGC	0.028	0.149	0.121	ATCTA	0.000	0.092	0.092
5	TCAGC	0.028	0.138	0.110	CCCCA	0.000	0.114	0.114	CAGCA	0.056	0.147	0.090
6	CATCT	0.000	0.109	0.109	CTTAC	0.000	0.112	0.112	AACCA	0.028	0.117	0.089
7	ACTCT	0.000	0.104	0.104	TATAC	0.000	0.112	0.112	CATCT	0.000	0.089	0.089
8	CCCCA	0.000	0.102	0.102	GAGGA	0.000	0.110	0.110	CCCGC	0.000	0.088	0.088
9	GAGGA	0.000	0.100	0.100	GGAAT	0.112	0.222	0.109	GTCGA	0.000	0.087	0.087
10	GTCGA	0.000	0.099	0.099	TACAT	0.028	0.136	0.107	TCTAC	0.000	0.086	0.086
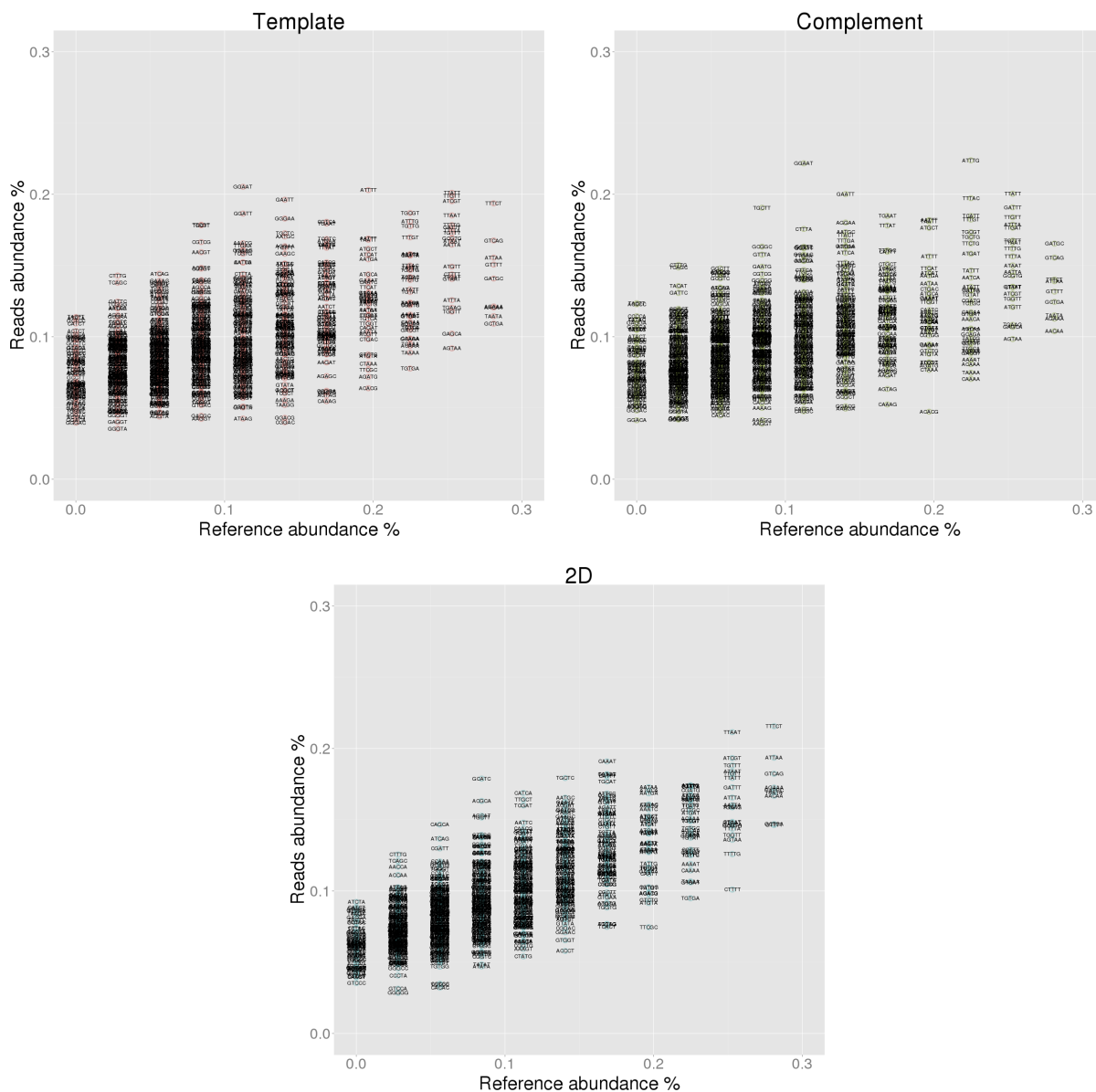
Control sequence GC content
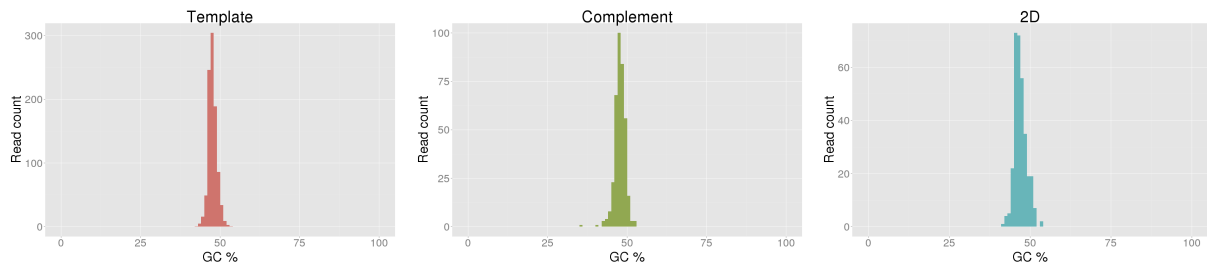
Escherichia coli error analysis

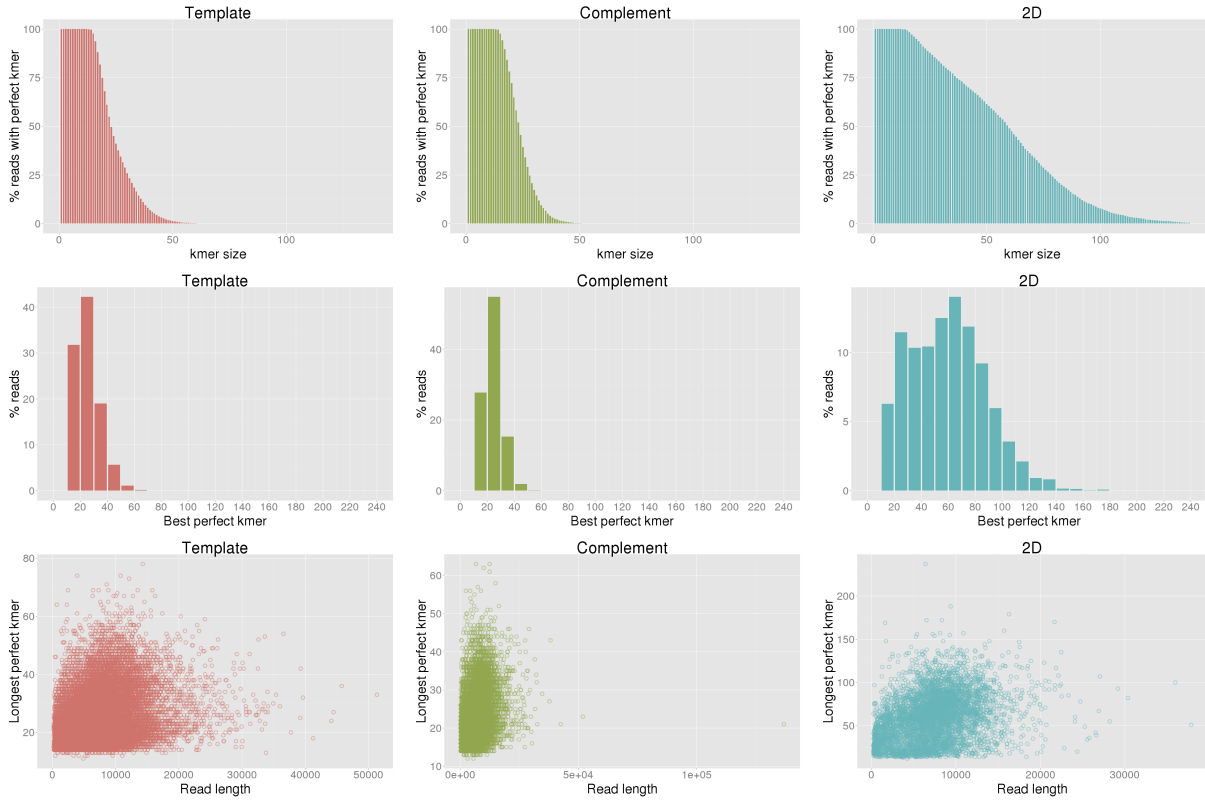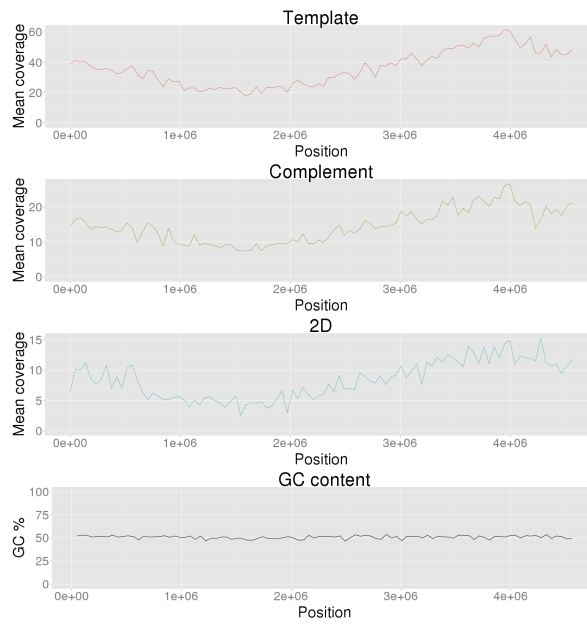|	Template	Complement	2D
Overall base identity (excluding indels)	55.68%	55.74%	75.55%
Aligned base identity (excluding indels)	78.39%	79.59%	86.64%
Identical bases per 100 aligned bases (including indels)	64.59%	64.71%	75.25%
Inserted bases per 100 aligned bases (including indels)	5.28%	4.83%	5.18%
Deleted bases per 100 aligned bases (including indels)	12.32%	13.87%	7.97%
Substitutions per 100 aligned bases (including indels)	17.81%	16.60%	11.60%
Mean insertion size	1.61	1.58	1.59
Mean deletion size	1.76	1.85	1.66
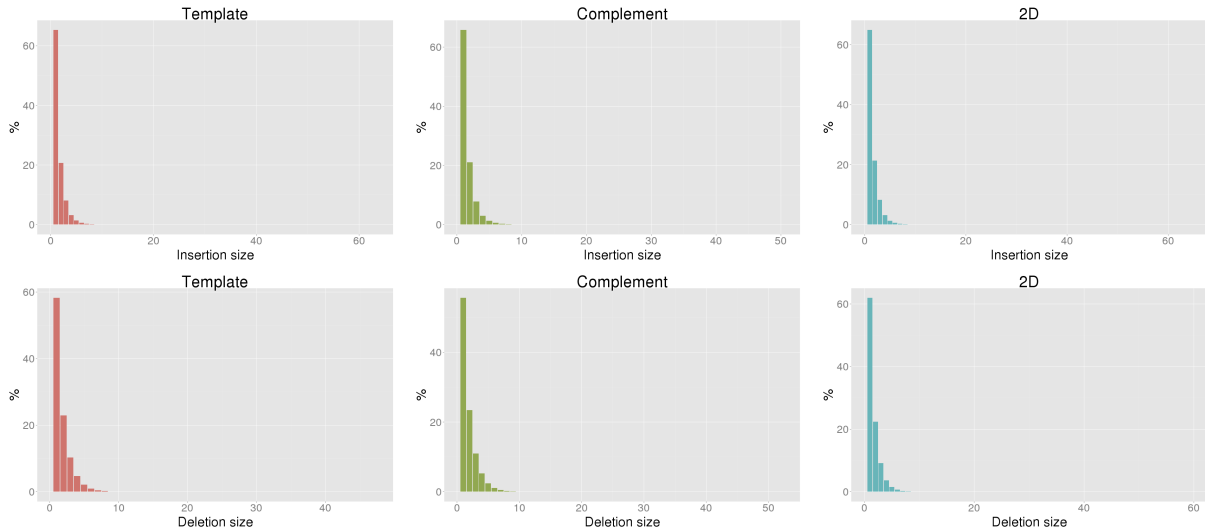
Escherichia coli read identity
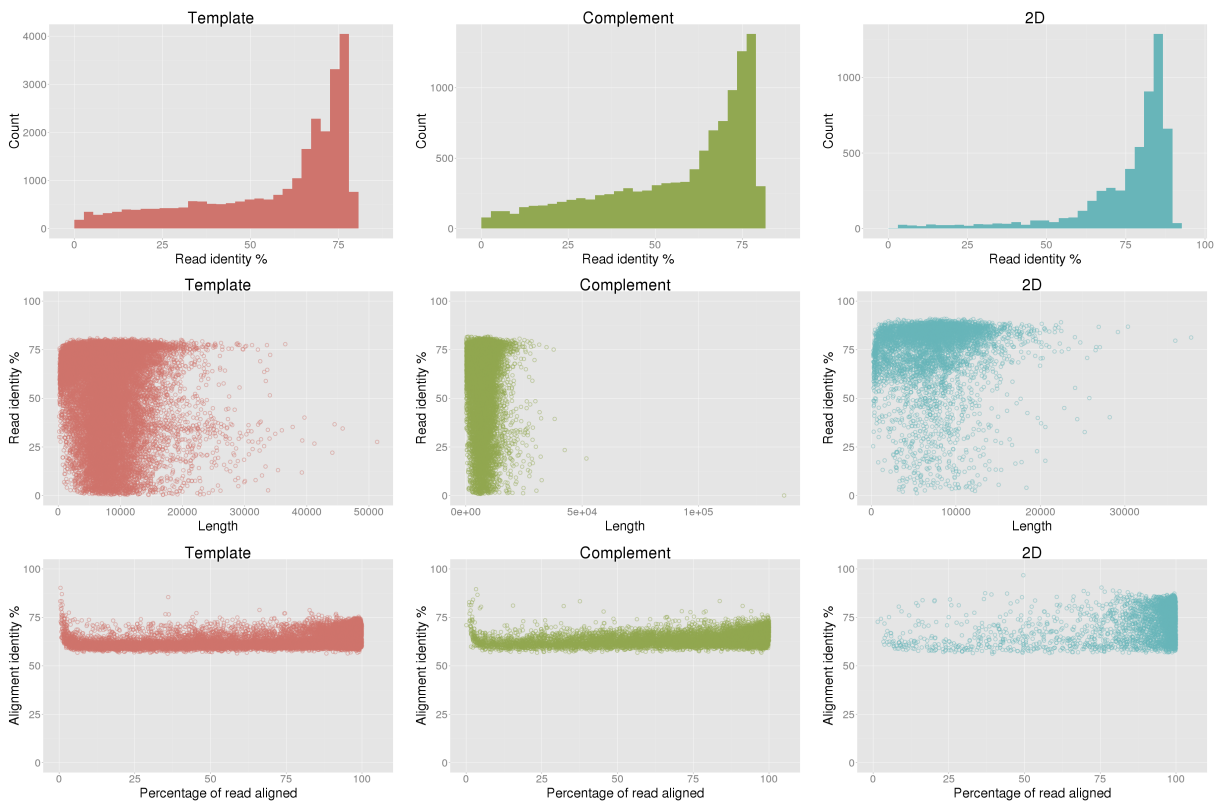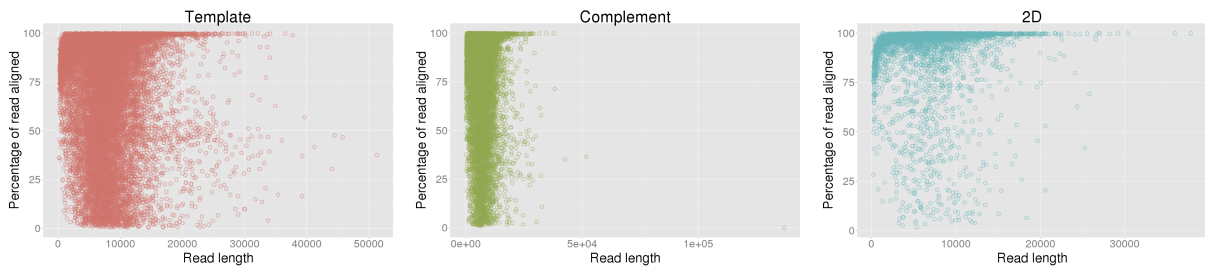
## Escherichia coli perfect kmers

Escherichia coli coverage

Escherichia coli 5-mer analysis

Under-represented 5-mers

		Template				Complement				2D		
Rank	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %
1	CGCTG	0.259	0.074	-0.185	AAAAA	0.247	0.040	-0.208	TTTTT	0.251	0.034	-0.217
2	AAAAA	0.247	0.070	-0.177	CGCCA	0.288	0.085	-0.203	AAAAA	0.247	0.034	-0.213
3	CGCCA	0.288	0.123	-0.165	TTTTT	0.251	0.062	-0.189	CGCCA	0.288	0.175	-0.113
4	GCTGG	0.279	0.115	-0.164	CGCTG	0.259	0.109	-0.150	GCCAG	0.280	0.172	-0.108
5	TTTTT	0.251	0.106	-0.145	CTGGC	0.278	0.130	-0.148	CGCTG	0.259	0.151	-0.108
6	CTGGC	0.278	0.135	-0.143	CCAGC	0.289	0.149	-0.140	GCTGG	0.279	0.172	-0.107
7	GCCAG	0.280	0.139	-0.141	TGGCG	0.275	0.137	-0.138	CCAGC	0.289	0.192	-0.097
8	CCAGC	0.289	0.150	-0.139	GCCAG	0.280	0.146	-0.134	AAAAT	0.195	0.101	-0.094
9	TGGCG	0.275	0.140	-0.135	CGCGC	0.201	0.070	-0.131	TGGCG	0.275	0.183	-0.093
10	CGCCG	0.219	0.105	-0.114	CAGCA	0.261	0.131	-0.130	CTGGC	0.278	0.190	-0.088

Over-represented 5-mers

		Template				Complement				2D		
Rank	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %	kmer	Ref %	Read %	Diff %
1	CGGGC	0.116	0.199	0.084	ACCCC	0.040	0.129	0.089	CTAGA	0.003	0.063	0.060
2	TTCGT	0.090	0.173	0.083	TAGGA	0.012	0.094	0.082	TCTAG	0.003	0.063	0.059
3	TCCGT	0.066	0.147	0.082	TGCTT	0.099	0.174	0.074	TCTAA	0.025	0.082	0.057
4	TAGGA	0.012	0.093	0.081	CCTAG	0.003	0.074	0.071	CTCGT	0.042	0.098	0.056
5	CCTAG	0.003	0.084	0.081	TCCTA	0.013	0.084	0.071	GGGTC	0.040	0.096	0.055
6	TCGTA	0.053	0.133	0.081	GGAAT	0.089	0.158	0.068	TTAGA	0.026	0.080	0.053
7	TCGTC	0.094	0.174	0.080	CCTAT	0.028	0.094	0.066	CCCAA	0.047	0.099	0.052
8	GGATT	0.098	0.177	0.079	CCCCG	0.055	0.120	0.065	ATCTA	0.033	0.084	0.052
9	TCGGG	0.060	0.136	0.076	ACCTA	0.027	0.091	0.064	GGGGT	0.039	0.091	0.051
10	ACCCC	0.040	0.115	0.074	CTAGG	0.003	0.066	0.064	TAGAT	0.035	0.085	0.051
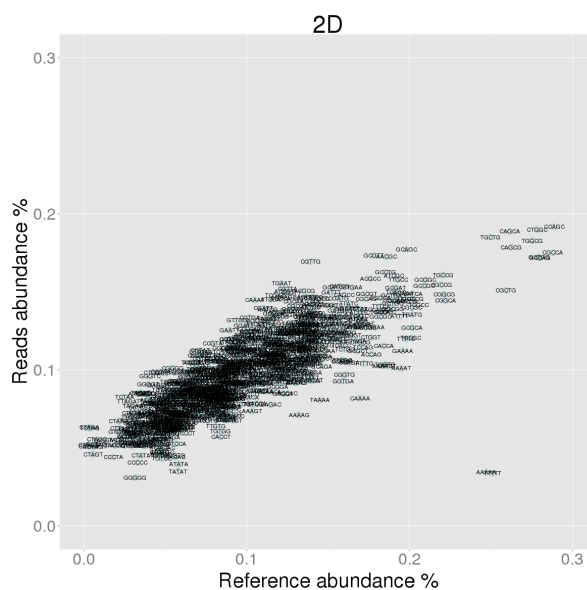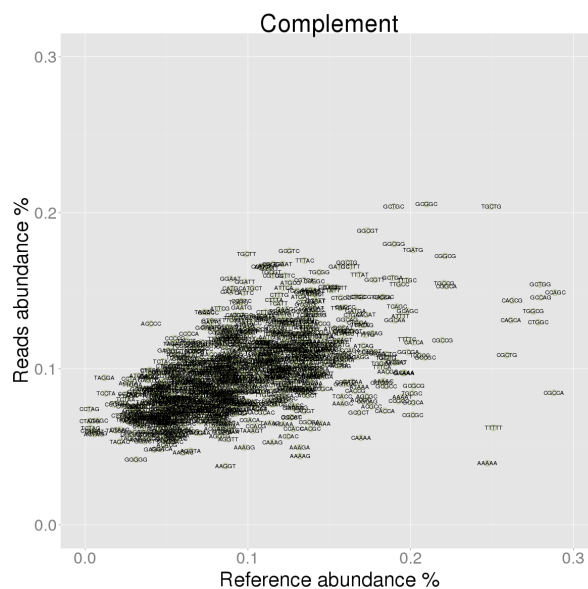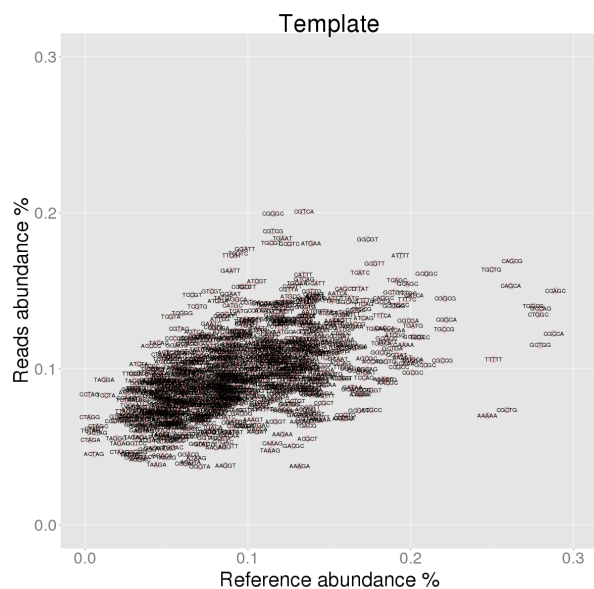
Escherichia coli GC content
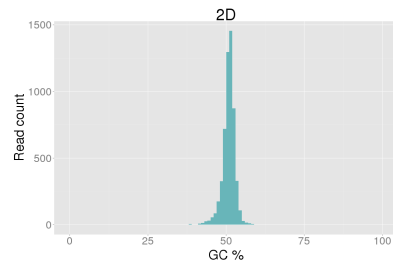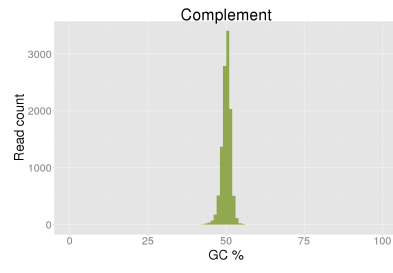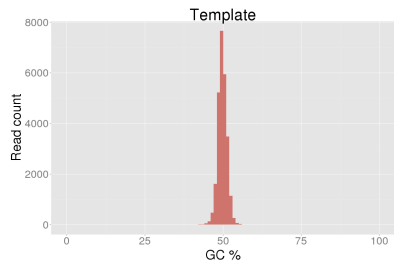
All reference 21mer analysis

Template · Complement · 2D (Number of perfect 21mers vs Read length)
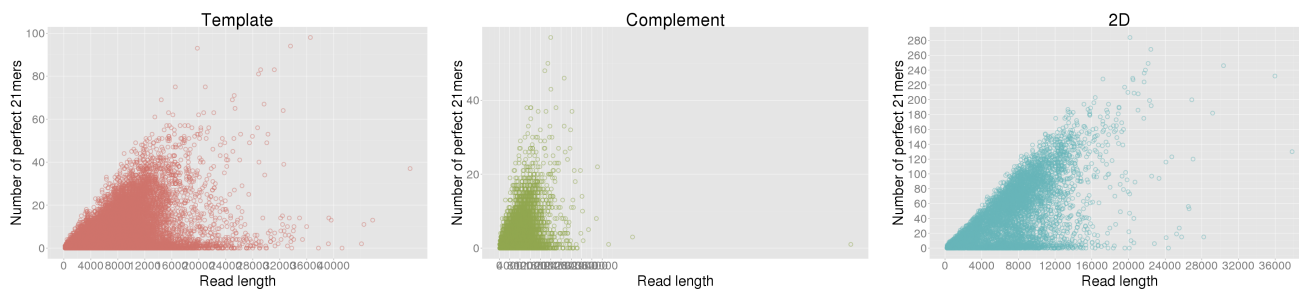
All reference substitutions

		Template substituted %				Complement substituted %				2D substituted %			
		a	c	g	t	a	c	g	t	a	c	g	t
Reference	A	0.00	8.54	8.70	5.29	0.00	8.65	8.58	5.34	0.00	8.77	8.80	4.76
	C	8.85	0.00	8.87	10.07	9.36	0.00	8.63	9.73	8.99	0.00	9.95	9.03
	G	9.60	8.92	0.00	8.58	9.14	8.72	0.00	9.00	9.02	10.06	0.00	8.58
	T	5.51	8.71	8.35	0.00	5.61	8.66	8.58	0.00	4.81	8.67	8.56	0.00

Kmer motifs before errors

3-mer error motif analysis

Rank	Template Insertion	Template Deletion	Template Substitution	Complement Insertion	Complement Deletion	Complement Substitution	2D Insertion	2D Deletion	2D Substitution	
1	TTC (3.34%)	TTC (3.34%)	AAA (4.07%)	TTC (2.93%)	GGC (2.82%)	AAA (3.96%)	GCA (3.00%)	AAA (2.63%)	AAA (3.76%)	Most common
2	AAA (2.80%)	AAA (2.91%)	TTC (3.59%)	TGC (2.80%)	AAA (2.79%)	GCA (3.50%)	TTC (2.88%)	GCA (2.63%)	GCA (3.51%)	
3	GCA (2.76%)	TGC (2.90%)	GCA (3.28%)	GCA (2.73%)	TGC (2.77%)	GAA (3.24%)	AAA (2.83%)	GGC (2.62%)	TTC (3.02%)	
4	TGC (2.66%)	GCA (2.79%)	GAA (3.01%)	AAA (2.72%)	GCA (2.74%)	TTC (3.12%)	TCA (2.44%)	TGC (2.58%)	GAA (3.01%)	
5	ATC (2.52%)	TCA (2.47%)	TGC (2.64%)	GAA (2.45%)	TTC (2.66%)	TTT (2.56%)	TGC (2.42%)	TTC (2.54%)	TTT (2.68%)	
6	TCA (2.40%)	GCC (2.45%)	TTT (2.57%)	CAG (2.40%)	GAA (2.48%)	TGC (2.54%)	ATC (2.39%)	GCG (2.51%)	GTT (2.47%)	
7	GCC (2.26%)	ATC (2.32%)	AAT (2.53%)	TCA (2.33%)	GCC (2.35%)	TCA (2.33%)	GAA (2.34%)	TCA (2.46%)	GCC (2.39%)	
8	GAA (2.24%)	GAA (2.32%)	ATC (2.28%)	ATC (2.30%)	TCA (2.30%)	ATC (2.26%)	GCG (2.29%)	CGC (2.25%)	AAT (2.39%)	
9	TTT (2.21%)	GGC (2.31%)	TCA (2.27%)	GGC (2.27%)	CAG (2.23%)	GCC (2.24%)	CGC (2.22%)	GAA (2.24%)	TGC (2.28%)	
10	AAT (2.12%)	AAT (2.29%)	GCC (2.22%)	GCC (2.15%)	AAT (2.21%)	AAT (2.20%)	AAT (2.20%)	GCC (2.22%)	GCG (2.25%)	
	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	
-10	AGT (0.99%)	AGG (0.92%)	GGT (0.90%)	AGT (0.98%)	AGT (0.95%)	CCC (0.92%)	TGT (1.02%)	CTC (1.03%)	ACT (0.92%)	Least common
-9	CTC (0.91%)	CTT (0.90%)	AGA (0.87%)	AGA (0.93%)	GGA (0.93%)	CTC (0.91%)	CTC (1.01%)	CGA (0.96%)	CCC (0.92%)	
-8	CCC (0.83%)	CCT (0.90%)	AGG (0.86%)	GGA (0.86%)	CCC (0.92%)	AGT (0.88%)	GAG (0.83%)	ACT (0.94%)	AGG (0.90%)	
-7	GGA (0.80%)	GGG (0.89%)	GGG (0.85%)	CTC (0.85%)	AGA (0.92%)	CCT (0.88%)	CCC (0.82%)	CCC (0.91%)	CGA (0.85%)	
-6	AGA (0.80%)	CGA (0.87%)	TGT (0.85%)	CCC (0.81%)	CCT (0.90%)	AGG (0.84%)	GGA (0.81%)	CTT (0.90%)	CTT (0.78%)	
-5	GAG (0.77%)	GAG (0.84%)	AGT (0.83%)	AGG (0.74%)	CTC (0.86%)	ACT (0.79%)	AGA (0.80%)	CCT (0.85%)	GAG (0.78%)	
-4	AGG (0.75%)	AGA (0.73%)	CTT (0.83%)	GAG (0.74%)	GAG (0.84%)	GGG (0.77%)	AGG (0.77%)	AGA (0.79%)	GGA (0.70%)	
-3	GGG (0.68%)	GGA (0.72%)	GAG (0.66%)	GGG (0.63%)	GGG (0.78%)	GAG (0.61%)	GGG (0.74%)	GGA (0.75%)	AGA (0.65%)	
-2	CTA (0.50%)	TAG (0.49%)	TAG (0.39%)	CTA (0.53%)	CTA (0.53%)	CTA (0.43%)	CTA (0.58%)	CTA (0.64%)	TAG (0.47%)	
-1	TAG (0.44%)	CTA (0.47%)	CTA (0.36%)	TAG (0.46%)	TAG (0.52%)	TAG (0.39%)	TAG (0.47%)	TAG (0.61%)	CTA (0.41%)	
	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	[logo]	

Kmer space for 3-mers: 64   Random chance for any given 3-mer: 1.56%

4-mer error motif analysis

Rank	Template Insertion	Template Deletion	Template Substitution	Complement Insertion	Complement Deletion	Complement Substitution	2D Insertion	2D Deletion	2D Substitution	
1	TTTC (1.04%)	TTTC (1.05%)	AAAA (1.34%)	ATCA (0.88%)	CGGC (1.03%)	CAAA (1.08%)	GAAA (0.84%)	TGGC (0.94%)	AAAA (1.08%)	Most common
2	GAAA (0.88%)	TTCA (0.97%)	TTTC (1.21%)	CAGC (0.85%)	CAGC (0.99%)	AAAA (1.07%)	TTTC (0.83%)	CAGC (0.89%)	GGCA (1.07%)	
3	AAAA (0.87%)	TGCC (0.95%)	GAAA (1.13%)	TTGC (0.84%)	TGGC (0.96%)	TGAA (0.95%)	AACA (0.82%)	TTCA (0.87%)	GAAA (1.03%)	
4	TTCA (0.87%)	TTCC (0.91%)	CAAA (0.90%)	CAAA (0.82%)	TTGC (0.86%)	GAAA (0.94%)	ATCA (0.82%)	CGGC (0.79%)	TTTC (0.93%)	
5	ATCA (0.83%)	AAAA (0.90%)	TTTT (0.90%)	CTGC (0.82%)	CTGC (0.85%)	GGCA (0.91%)	GGCA (0.81%)	CTGC (0.75%)	GGAA (0.88%)	
6	TTGC (0.81%)	TTGC (0.89%)	GGCA (0.89%)	CGGC (0.81%)	CAAA (0.84%)	TTTC (0.91%)	GCCA (0.80%)	TGCC (0.75%)	CGCC (0.87%)	
7	GTTC (0.80%)	CAGC (0.86%)	AAAT (0.88%)	TTTC (0.79%)	TGCC (0.83%)	AGCA (0.88%)	CGCC (0.79%)	ATCA (0.75%)	CGTT (0.87%)	
8	AACG (0.80%)	GAAA (0.83%)	TTCA (0.88%)	AAAA (0.76%)	TTCC (0.82%)	ATCA (0.88%)	CAAA (0.78%)	CGCC (0.74%)	TGCA (0.85%)	
9	TGCC (0.80%)	CTGC (0.82%)	GTTC (0.87%)	CCAG (0.75%)	ATCA (0.79%)	TAAA (0.86%)	TTCA (0.78%)	CAAA (0.74%)	CAAA (0.85%)	
10	CTTC (0.76%)	AACG (0.82%)	GCAA (0.86%)	TTCA (0.75%)	TTCA (0.79%)	GGAA (0.84%)	CAGC (0.76%)	GGCG (0.74%)	TGCC (0.84%)	
-10	TAGT (0.12%)	CTAT (0.13%)	CGAG (0.10%)	GGGG (0.12%)	GGGG (0.12%)	ACTA (0.11%)	GAGG (0.14%)	CGGA (0.15%)	CTAA (0.12%)	Least common
-9	AGGG (0.12%)	CGGA (0.12%)	CCCT (0.10%)	GTGT (0.12%)	CCCC (0.12%)	TTAG (0.11%)	TAGT (0.14%)	ACTT (0.15%)	CGGA (0.11%)	
-8	TTAG (0.12%)	TAGT (0.12%)	ACTA (0.10%)	CTAA (0.11%)	ACCT (0.12%)	TAGA (0.11%)	CCCT (0.13%)	CTAT (0.15%)	TATA (0.11%)	
-7	GAGG (0.11%)	GGAC (0.11%)	TAGT (0.09%)	GAGG (0.10%)	GTGT (0.12%)	CTAT (0.10%)	TCTA (0.13%)	CCCC (0.14%)	CTAT (0.11%)	
-6	CTAA (0.11%)	CCCT (0.11%)	TAGA (0.08%)	GGAC (0.10%)	CTAA (0.11%)	CGAG (0.09%)	CTAT (0.13%)	CTAA (0.13%)	ACTA (0.11%)	
-5	GGAC (0.09%)	CTAA (0.10%)	TCTA (0.08%)	CCCT (0.09%)	TAGA (0.09%)	GGAC (0.09%)	CTAA (0.11%)	CCCT (0.11%)	TCTA (0.10%)	
-4	TAGA (0.07%)	TAGG (0.07%)	GGAC (0.08%)	TAGA (0.09%)	CCCT (0.08%)	CCCT (0.07%)	TAGA (0.08%)	TAGG (0.09%)	TAGG (0.07%)	
-3	TAGG (0.06%)	TAGA (0.06%)	TAGG (0.07%)	CCTA (0.07%)	TAGG (0.07%)	TAGG (0.07%)	CCTA (0.08%)	TAGA (0.08%)	TAGA (0.06%)	
-2	CCTA (0.06%)	CCTA (0.05%)	CCTA (0.05%)	TAGG (0.06%)	CCTA (0.06%)	CCTA (0.05%)	TAGG (0.06%)	CCTA (0.07%)	CCTA (0.05%)	
-1	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.01%)	CTAG (0.02%)	CTAG (0.01%)	

Kmer space for 4-mers: 256   Random chance for any given 4-mer: 0.39%

5-mer error motif analysis

Rank	Template Insertion	Template Deletion	Template Substitution	Complement Insertion	Complement Deletion	Complement Substitution	2D Insertion	2D Deletion	2D Substitution	
1	CAGCA (0.34%)	CAGCA (0.35%)	CAGCA (0.42%)	CAGCA (0.41%)	CAGCA (0.43%)	CAGCA (0.52%)	CAGCA (0.37%)	CTGGC (0.37%)	CAGCA (0.44%)	Most common
2	TTATC (0.30%)	TTGCC (0.34%)	GAAAA (0.40%)	CATCA (0.33%)	CTGGC (0.36%)	GCAAA (0.37%)	CGCCA (0.34%)	CAGCA (0.34%)	TGGCA (0.39%)	
3	ATTTC (0.33%)	CTGGC (0.33%)	CAAAA (0.38%)	GCTGC (0.31%)	GCGGC (0.35%)	ATAAA (0.35%)	CGGCA (0.29%)	CGCCA (0.30%)	GAAAA (0.38%)	
4	CTGGC (0.29%)	TTTGC (0.31%)	GCAAA (0.35%)	CGGCA (0.29%)	CATCA (0.32%)	GAAAA (0.34%)	CTGGC (0.29%)	GCGGC (0.30%)	CGGCA (0.37%)	
5	CGTTC (0.29%)	TTTCA (0.30%)	CGTTC (0.33%)	CTGGC (0.29%)	CCAGC (0.31%)	CATCA (0.34%)	GCAAA (0.28%)	CCAGC (0.29%)	GCAAA (0.32%)	
6	CATCA (0.28%)	TTTTC (0.29%)	TTGCC (0.32%)	CCAGC (0.28%)	GCTGC (0.31%)	CATCA (0.34%)	TGGCA (0.28%)	TTTCA (0.29%)	GCGTT (0.32%)	
7	TTGCC (0.28%)	CATCA (0.29%)	TTTGC (0.31%)	GCAAA (0.28%)	TTGCC (0.31%)	AAGAA (0.32%)	TGGCG (0.27%)	TGGCG (0.28%)	CAAAA (0.31%)	
8	TTTGC (0.28%)	GAAAA (0.29%)	TGTTC (0.30%)	GCGGC (0.28%)	ATAAA (0.30%)	ACAAA (0.31%)	CAACA (0.27%)	GCCAG (0.28%)	TTGCC (0.30%)	
9	GAAAA (0.28%)	CGTTC (0.29%)	TGAAA (0.30%)	AATCA (0.28%)	TCAGC (0.30%)	ACGCA (0.31%)	CCAGC (0.27%)	GCAGC (0.28%)	TGAAA (0.30%)	
10	CAAAA (0.28%)	GCTGC (0.29%)	TTATC (0.30%)	TCTTC (0.27%)	AATCA (0.29%)	ATGAA (0.31%)	CATCA (0.26%)	CATCA (0.27%)	CGCCA (0.29%)	
-10	CCCTA (0.01%)	GGACC (0.01%)	CTTAG (0.01%)	CCCTA (0.01%)	ACCTA (0.01%)	ACCTA (0.01%)	CCCTA (0.01%)	CCCTA (0.01%)	TCCTA (0.01%)	Least common
-9	GGACC (0.00%)	CCCTA (0.01%)	CCCTA (0.01%)	ACCTA (0.01%)	CTAGC (0.01%)	CCCTA (0.01%)	TAGGA (0.01%)	TAGGA (0.01%)	TAGGA (0.01%)	
-8	CTAGC (0.00%)	GCTAG (0.00%)	ACTAG (0.00%)	CTAGC (0.01%)	CCCTA (0.01%)	CTAGC (0.00%)	CTAGC (0.01%)	GCTAG (0.01%)	CTAGT (0.01%)	
-7	GCTAG (0.00%)	ACTAG (0.00%)	GCTAG (0.00%)	GCTAG (0.00%)	GCTAG (0.00%)	GCTAG (0.00%)	GCTAG (0.00%)	CTAGC (0.01%)	CTAGC (0.00%)	
-6	ACTAG (0.00%)	CTAGC (0.00%)	CTAGC (0.00%)	ACTAG (0.00%)	CTAGT (0.00%)	CTAGT (0.00%)	CCTAG (0.00%)	ACTAG (0.01%)	GCTAG (0.00%)	
-5	CTAGT (0.00%)	CTAGT (0.00%)	CTAGT (0.00%)	CTAGT (0.00%)	ACTAG (0.00%)	CTAGA (0.00%)	CTAGT (0.00%)	CTAGT (0.00%)	ACTAG (0.00%)	
-4	CCTAG (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	CCTAG (0.00%)	CTAGA (0.00%)	ACTAG (0.00%)	ACTAG (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	
-3	CTAGG (0.00%)	CTAGG (0.00%)	CCTAG (0.00%)	CTAGA (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	TCTAG (0.00%)	
-2	CTAGA (0.00%)	CCTAG (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	CCTAG (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	CTAGA (0.00%)	CTAGA (0.00%)	
-1	TCTAG (0.00%)	CTAGA (0.00%)	CTAGA (0.00%)	TCTAG (0.00%)	CTAGG (0.00%)	CCTAG (0.00%)	CTAGA (0.00%)	CCTAG (0.00%)	CCTAG (0.00%)	

Kmer space for 5-mers: 1024   Random chance for any given 5-mer: 0.10%