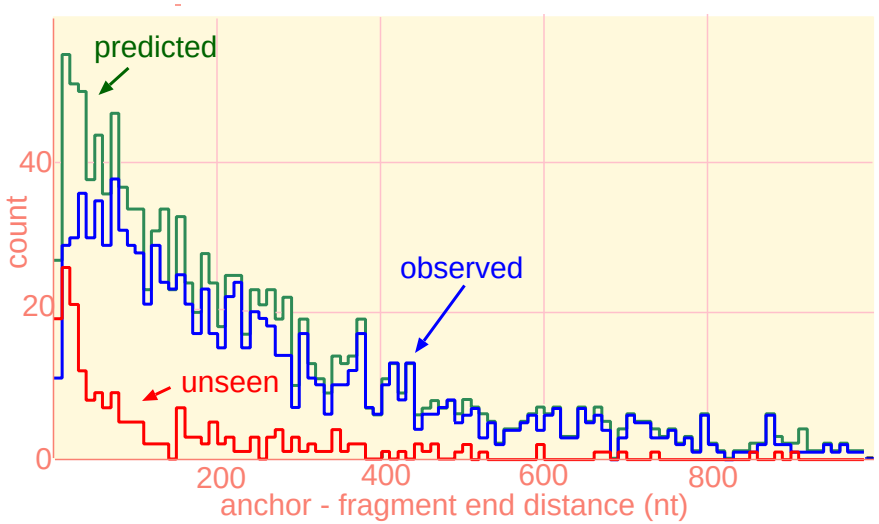
Tag: CATG + ...	Frequency (or count)
ATCTGGAGGTTCGGTTC	**65**
CGTCATCTCGCTGAACG	45
CCCTGCGGTACGGGAGC	34
GATGTATTTACGGCGTC	34
GCTGCATTGGCACCGTT	23
CCAGCATCAGCCAGCGC	22
TAGGCTCGAGCCGCGCC	20
ATGACCAGCGGCAAGCC	16
CTGCTGTTCGTGCTGCT	16
GTGATGGGGGGCGCTTC	16
GTGCGCTTAGGGGCAAC	16
AGCTTGGGGGTACCGGA	15
GACGGGAAGTCGAAAGA	15
GCACCAATGCGCTCACA	15
ACATCTTGTGGGCTGAT	14
AGGAGTTCCCTTCCCCA	14
ATGGCCGACTCCGGTGC	14
CCATCGCGATCAGCGAA	14
CTAATGCTTCATCGGGG	14
GCGGTGATGTGATTGAT	14
GTTCGGATGGCTGCAAA	14
AGCCCCATCAATTGAAC	13
ATTAACATTTGTAGCAG	13
CAGCGGGCCAAGATCAC	13
CAGTCGCTGGTATTCAC	13
CGCTCATCGAAATTATC	13
CTGATAAACCGGGATCG	13
GAAAAGTGTATTCCGAT	13
GCATCGCTTAACGCCCA	13
GCCAATGCGCGGTGAGA	13
GTGGCTTGCAACAACCC	13
GTTAGCCATAGTGAAGC	13
GTTCGCCCGAAGCCTAA	13
TCGTTCAAATCAAAGGA	13
ATCCGGCTAACTGGCCG	12
CAACTTTGGCTTAACGA	12
CAGCATCGCCGCCCATT	12
CAGGAGAACGGCGAAGG	12
CGCCGATAGCACCGATG	12
GCCAGCGCGCGCCAGGA	12
GCTCCGTGACCAATGAA	12
GCTTGCAAGATGGACAT	12
GGTGTTCCCCCGCAGCC	12
GTGCGGCTCGATGGAAA	12
TCAAAACGATTACCACC	12
ACCGTAATGCCATCCAC	11
ACTCGTGGGGCAGAAGC	11
AGCCTCGGAAACCAGCG	11
ATGCAGCCCATTCCAGG	11
ATGCTATCCGGCCCACC	11
ATTATGCGTTATGCGCC	11
CCCCAGCCCTGGCCCGC	11
.	.
.	.
.	.

5432 GSTs were collected in total, from concatemers. These were then compared with a small database of 1326 GSTs that was generated from the *Y. pestis* genome sequence (CO92). 97% exactly matched 1133 GST sites in the genome. The remainder were almost all matched imperfectly (sequencing errors) or matched other NlaIII sites.

This tag has a high frequency because it occurs in a repeat element in the genome.

209 tags present in the database were not observedin concatemers. While this is partly due to the non-oversampled statistics, a histogram of anchor-fragment end distances for observed and unseen tags indicates a distinct bias towards the short end. A longer biotinylated cassette should rectify this problem.



Nearly all tags matched the Genome or could be otherwise accounted for.

tag (CATG+...)	dir	Bam frag	offsets 3'	offsets 5'	number collecte
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
ATCGACAGCGATACCGG	f	521	65	5145	2
ATCGCCGCCATCGGTGC	f	625	266	538	3
ATCGCCTGATATTCCCG	r	110	334	9958	4
ATCGCGATGGGTATCTT	f	50	21	7547	1
ATCGCGGCAATCTGCCA	f	403	383	5967	7
ATCGCGGTGTGTCAGGG	f	560	81	857	3
ATCGGATCTTGTTGGTC	f	483	49	9728	1
ATCGGCATAAAAGTTGC	f	121	72	1063	1
ATCGGCCGTAAACGCGC	r	57	442	5595	9
ATCGTCGGTGCGGTATG	r	625	254	550	1
ATCGTCGTCTCGCGCAG	f	368	592	4195	3
ATCGTTAATGTTCTGTT	r	398	902	9667	1
ATCGTTCCTTACGGTTT	f	434	226	5852	2
ATCTATTACCCGCTGTC	r	654	38	5627	1
ATCTCGACCGCACACAT	r	550	137	2347	9
ATCTGGAGGTTCGGTTC	f	685	137	6707	65
	f	679	137	7419	
	f	658	137	43143	
	r	603	137	12357	
	r	587	137	10959	
	f	172	137	13165	
	f	122	137	4734	
	r	13	137	1621	
ATCTGGTTGCTGCGTTC	f	54	265	3413	5
ATCTTCCAGAATTTCCA	f	221	258	4777	4
ATGAAAACCACGATCTT	f	385	108	11698	3
ATGAAATGCGATCCCGT	r	105	228	17144	3
ATGACACGCACCAACGG	r	4	299	9167	3
ATGACCAGCGGCAAGCC	r	548	777	100	16
ATGACCGTTGATGGAAA	f	666	39	5326	1
ATGACCTTGGGTGTAAT	r	645	59	14077	1
ATGACGATAGCGATGAA	f	171	72	820	3
ATGAGGTCAAGCGGGAG	f	34	38	17789	3
ATGAGTTCTATGGTGAA	r	50	799	6769	5
ATGATCCGCTTGAGTCG	f	600	69	32659	3
ATGCACTCGGTATTACC	f	570	197	10304	3
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.