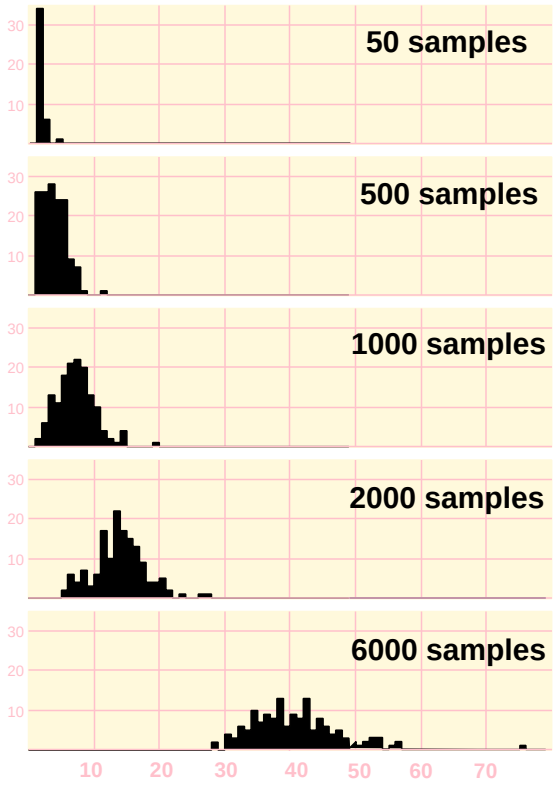
GST sampling is democractic. Tag frequencies follow a binomial distribution. For low sample numbers, the distribution is approximately Poisson. As the number of samples grows, the distribution becomes normal (Gaussian).

Shown here is a series of Monte Carlo samplings of *Y. pestis Bam*HI fragmement defined *Nla*III GSTs, taken from the genome.



undersampled

oversampled

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
.	.
.	.
.	.