Attrition Analysis

Name	Sequences	Trimmed	Aligned	% of Sequences	Less than Three	% of Aligned	Unique Hits	% of Less than 3	Set Size	% of Unique	Passing Set Size	% of Set Size	Homopolymeric correction	% of Total
Wang-VSVGgfp-Jurkat-454-Avr	92321	71259	39518	55.46%	37461	94.79%	35035	93.52%	25682	73.30%	20067	78.14%	19962	28.01
Wang-VSVGgfp-Jurkat-454-Mse	73351	73237	37454	51.14%	35051	93.58%	32943	93.99%	26985	81.91%	20694	76.69%	20607	28.13

Sequences - number of sequences with obtained from pyrosequencing. Reads are in one direction only.

Trimmed - number of sequences with perfect primers (Avr) or those trimmed by cross_match (Mse).

Aligned - number of sequences that returned any alignment in BLAT using default client-server settings.

Less Than Three - number of sequences whose alignment (qStart) began at the first,second, or third base pair position.

Unique Hits - number of sequences that have a unique alignment with the best number of matches. If a sequences aligns to two genomic positions with equally good match scores, it is omitted from the set.

Set size - size of the set after converging hits are compressed. Converging hits refer to sequences that align to the same position. Only the sequence with the best identity is kept.

Passing set size - size of set after a 98% identity criteria is enforced. Sequences must align with 98% identity in the first 50bp, first 100bp, or over the entire alignment.

Homopolymeric correction - distinct sequences that converged within 3 base pairs of each other and showed evidence of homopolymeric sequencing errors were converged.