

Array-Ready Oligo Set™ for the *Mycobacterium tuberculosis* Genome, Version 1.1

Our Tuberculosis Genome Oligo Set Version 1.1 contains 4505 70mer probes representing 3926 Open Reading Frames (ORFs) from *Mycobacterium tuberculosis H37RV* strain and also representing 3811 ORFs from the CDC-1551 strain.

ORF Sequence Source and Selection

Primary source

A total of 3932 probes are designed from the set of ORFs from *Mycobacterium tuberculosis H37RV* strain obtained from the Sanger Institute (<http://www.sanger.ac.uk>).

Secondary source

Another 573 probes are designed directly from the set of ORFs from *Mycobacterium tuberculosis CDC-1551* strain obtained from The Institute for Genomic Research (<http://www.tigr.org>).

An oligo is said to represent a certain CDC-1551 or H37RV ORF if the oligo has greater than 97% identity over the length of the oligo. Using this criterion, 3811 ORFs from the CDC-1551 strain are represented.

	<i>H37RV</i>	<i>CDC-1551</i>
Number of ORFs represented	3926	3511

Probe design and selection rules

For each ORF sequence, a probe is selected with an optimal set of parameters. Sufficient numbers of 70mer candidate probes for each ORF are selected using the following criteria for the Tuberculosis Genome Oligo Set.

1) Each oligo is within 79°C ±5°C using the following formula:

$$T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C) / \text{length} - 500 / \text{length} \text{ where } [\text{Na}^+] = 0.1 \text{ M and length} = \#A + \#C + \#G + \#T$$

2) Each oligo is located anywhere in the ORF to allow for random or gene-specific priming.

3) Each oligo is scanned for potential hairpins stems.

4) Using BLAST, each oligo designed from a *H37RV* ORF is aligned against all the *H37RV* ORF sequences. Each oligo designed from a *CDC-1551* ORF is aligned against all the *CDC-1551* ORFs. Using the alignment with the candidate oligo versus the highest scoring non-self ORF, a percent identity score is computed. The highest scoring non-self ORF is defined as the sequence that yields the most matched bases in an alignment. This percent identity is referred to as the cross-hybridization identity of the oligo.

One final oligo for each ORF is selected with the minimum cross-hybridization identity.

## SUMMARY

Oligo selection criteria	Value	Number of oligos in genome set satisfying these criteria
Length (bases) Melting temperature	70mer 79°C ±5°C	4409
Length (bases)	40–69mer	92
Length (bases)	71mer	4
<b>Total</b>		<b>4505</b>

All 4295 oligos were aligned against one another to obtain a percent identity of each probe versus all other non-self probes. The following table summarizes oligo versus oligo percent identity. The percent identity is the number of matched bases divided by the length of the oligo.

Percent identity of oligo versus another oligo	Cumulative number of oligos	Cumulative percent of oligos
≤ 40%	4205	93%
≤ 70%	4376	97%
≤ 90%	4424	98%
100%	63	1.4%

## Design Criteria Distributions

The following illustrations show the distribution of all 4505 oligos for melting temperature, GC content, location from 3' end of ORF sequence, longest hairpin stem length, and cross-hybridization identity. The cross-hybridization identity data is based on oligos originally designed from *H37RV* ORF sequences aligned versus *H37RV* ORF sequences, and oligos originally designed from *CDC-1551* ORFs aligned versus *CDC-1551* ORFs.

Figure 1. Melting Temperature

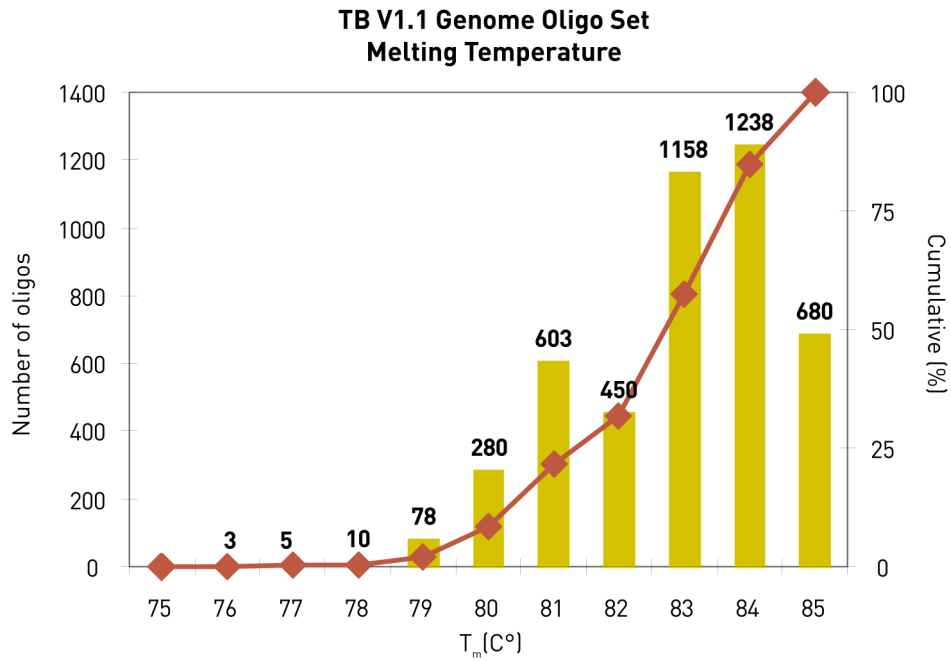


Figure 2. GC Content

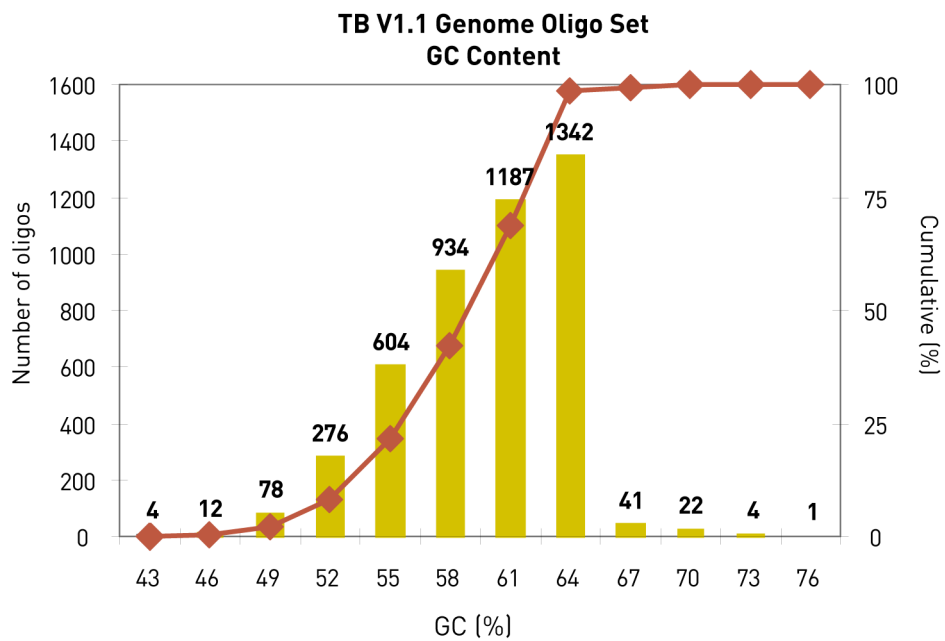


Figure 3. Location from 3' End

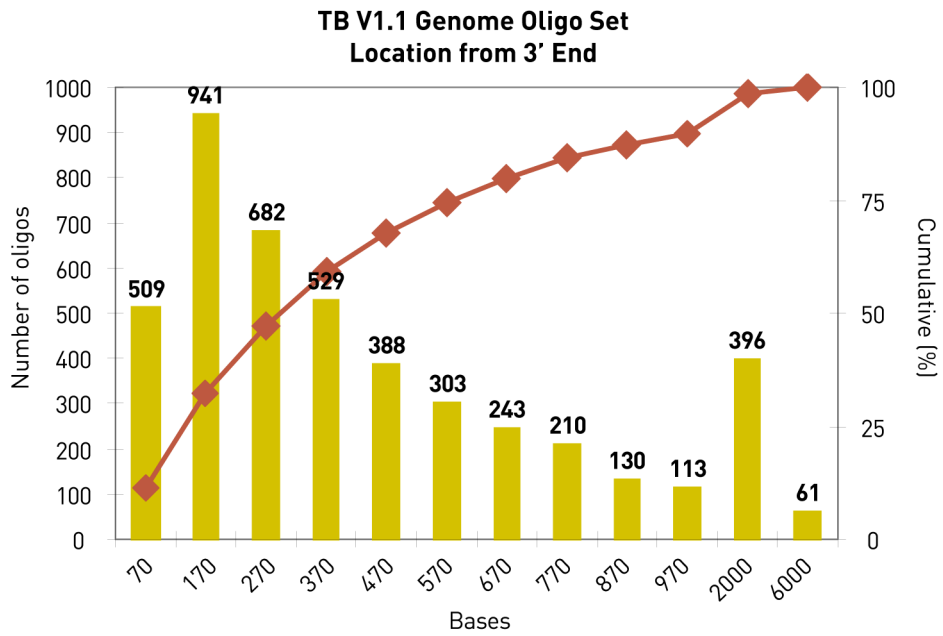


Figure 4. Longest Hairpin Stem Length

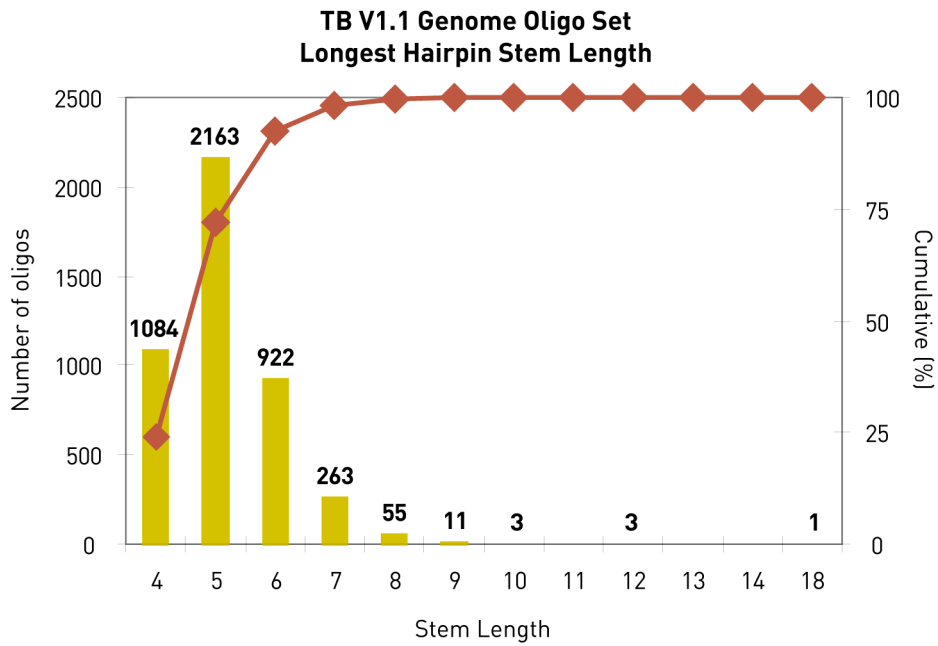


Figure 5. Cross-Hybridization Identity

