

Array-Ready Oligo Set™ for the *Sinorhizobium* Genome Version 1.0

Our *Sinorhizobium* Genome Oligo Set contains 6205 70mer probes representing 6205 ORFs from *Sinorhizobium meliloti* strain 1021. For our probe design we use state-of-the-art methods and proprietary software.

ORF Sequence Source and Selection

The 6205 probes for *S. meliloti* strain 1021 are designed from the three National Center of Biotechnology Information (NCBI) Reference Sequences (RefSeqs): *NC_003047* (chromosome) using the database file *NC_003047.ffn*, *NC_003037* (pSymA) using the database file *NC_003037.ffn*, and *NC_003078* (pSymB) using the database file *NC_003078.ffn* (updated in March 2002; source: NCBI at ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Sinorhizobium_meliloti/). More information is available at the following website: <http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/eframk?db=Genome&gi=188>.

Probe Design and Selection Rules

A probe is selected with an optimal set of parameters. Large numbers of 70mer candidate probes for each ORF are selected using the following criteria for the *Sinorhizobium* Genome Oligo Set:

1) An oligo is within $80^{\circ}\text{C} \pm 5^{\circ}\text{C}$. 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) To accommodate both random and gene-specific priming, the oligo is located more than 50 bases away from both the 5' and 3' end of the ORF. In other words, the 5' end of the oligo is more than 50 bases from the 5' end of the ORF and the 3' end of the oligo is at least 50 bases from the 3' end of the ORF.

3) An oligo cannot have a contiguous single nucleotide base repeat or poly (N) tract longer than 7 bases.

4) An oligo cannot have a potential hairpin structure with a stem length longer than 9 bases.

5) Each oligo has less than or equal to 70% identity to all other ORFs. Each oligo is aligned against all 6205 ORF sequences in *NC_003047*, *NC_003037*, and *NC_003078*. Using the alignment with the candidate oligo versus the highest scoring non-self sequence, a BLAST percent identity score is computed. The highest scoring non-self sequence is defined as the sequence that yields the most matched bases in an alignment. This BLAST percent identity is also referred to as cross-hybridization identity of the oligo.

6) Each oligo of any length cannot have greater than 20 contiguous bases common to any other gene.

Once oligo candidates have been selected satisfying all the selection rules mentioned above, each oligo is ranked based on BLAST percent identity as computed in Step 5. One final oligo for each gene is selected with the minimum cross-hybridization identity.

Note that for 292 (4.7%) ORFs that did not yield oligos satisfying all the above criteria, certain rules were relaxed. For those ORFs, the oligo is located less than 50 bases to either the 3' or 5' end or the cross-hybridization identity is greater than 70%.

SUMMARY

Oligo selection criteria	Value	Number of oligos in genome set satisfying these criteria
Length Melting temperature Location from 5' and 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other ORFs Contiguous base match to all other ORFs	70mer 80°C ± 5°C > 50 ≤ 7 ≤ 9 ≤ 70% ≤ 20	5913
Total number of oligos not satisfying one or more of the above criteria		292
Location from 5' and 3' end	≤ 50	99
Length	60mer	1
Cross-hybridization identity to all other ORFs	> 70%	194
Contiguous base match to all other ORFs	> 20	129
Total		6205

The following illustrations show the distribution of all 6205 oligos for melting temperature, GC content, location from 3' end of gene sequence, length of maximum stem length, and cross-hybridization identity.

Figure 1. Melting Temperature

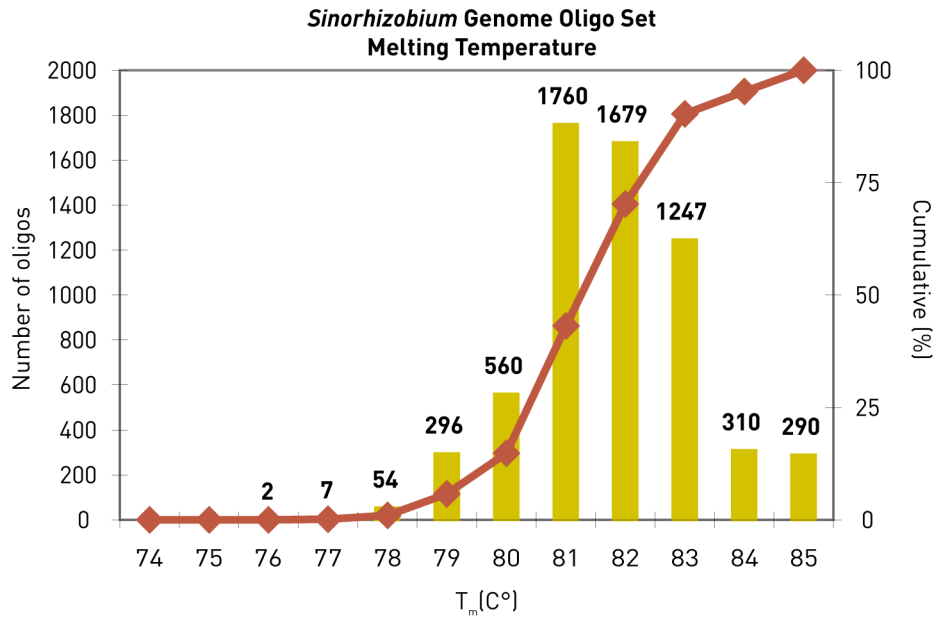


Figure 2. GC Content

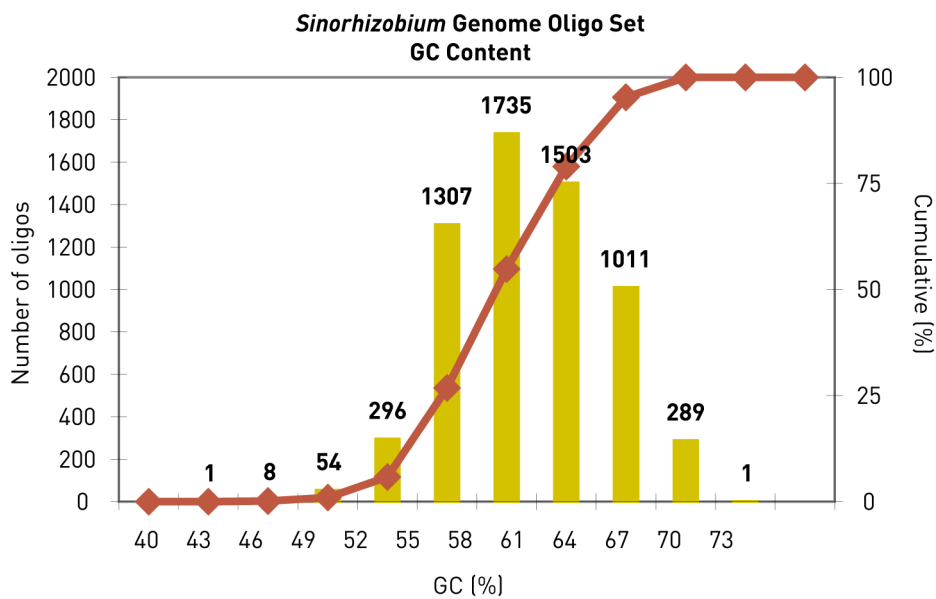


Figure 3. Location from 3' End

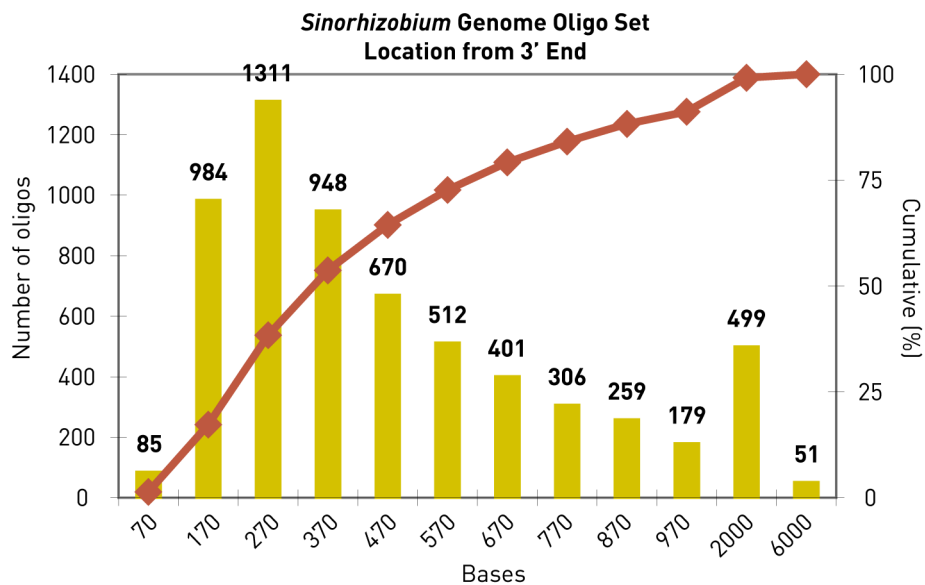


Figure 4. Length of the Longest Hairpin Stem

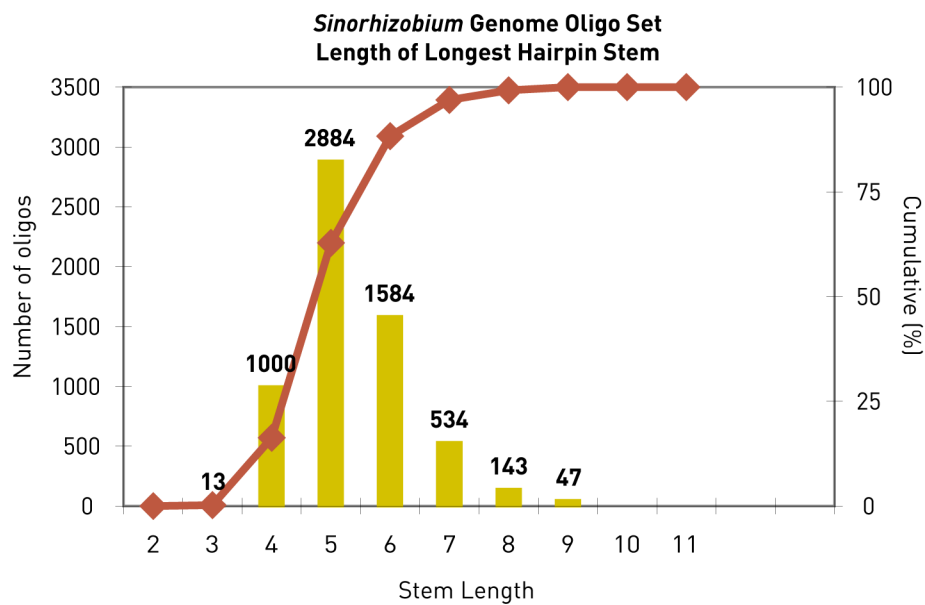
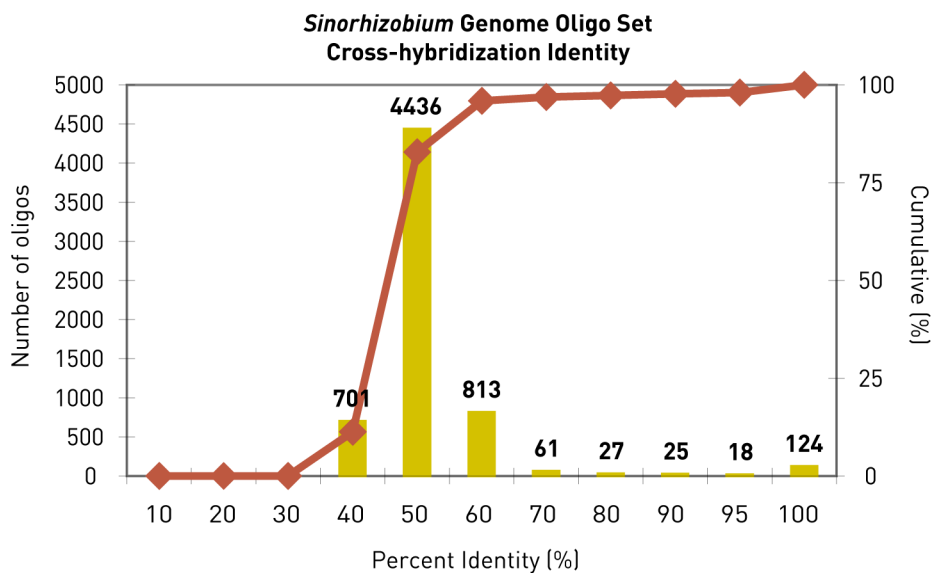


Figure 5. Cross-Hybridization Identity



Quality Check of Probe Design Specifications

Once the final oligo has been selected to represent a gene, each oligo undergoes design specifications quality control where we use an independent method to confirm that all oligos have met the specified design specifications. The table below summarizes data from our quality check for probe design specifications for all 6205 oligos in the set.

Probe design specification	Expected value	Verified Range	Number of oligos
Melting temperature (C°)	80°C ± 5°C	75.33–84.70	6205
Cross-hybridization similarity (%)	≤ 70	27–67	6011
Cross-hybridization similarity (%)	71–100	71–100	194