

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

Our *Medicago* Genome Oligo Set Version 1.0 contains 16,086 70mer probes representing 16,086 *Medicago truncatula* genes. For our probe design we use state-of-the-art methodology and proprietary software.

Gene Sequence Source and Selection

All probes are designed from The Institute of Genome Research (TIGR) Gene Index Database MtGI Release 5.0, released on May 3, 2002. TIGR develops and maintains this database at <http://www.tigr.org/tdb/tgi/mtgi/>.

MtGI Release 5.0 contains a total of 33,765 unique sequences including 16,086 tentative consensus sequences (TCs) and 17,679 singletons. All 16,086 TCs are used to design oligos.

TIGR <i>Medicago</i> MtGI Release 5.0 Database	Number of sequences	Number of probes designed in set
Tentative consensus sequences (TCs)	16,086	16,086
Singletons	17,679	0
Total	33,765	16,086

TIGR obtains and predicts orientation for all the tentative consensus sequences and singletons based on various techniques including alignments to known proteins and poly A trimming. After MtGI 5.0 was released, TIGR later updated the orientation of 242 of the TCs. Probes for these 242 sequences are therefore designed in the updated orientation.

Probe Design and Selection Rules

Once a gene has been selected to be included in the set, a probe is selected with an optimal set of parameters. Sufficient numbers of 70mer candidate probes for each gene are selected using the following criteria for the *Medicago* Genome Oligo Set.

- 1) All oligos are within 78°C ± 5°C using the following formula:
 $T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C)/\text{length} - 500/\text{length}$ where $[\text{Na}^+] = 0.1 \text{ M}$ and $\text{length} = \#A + \#C + \#G + \#T$
- 2) Each oligo is within 1000 bases from the 3' end of the available gene sequence.
- 3) An oligo cannot have a contiguous single nucleotide base repeat or poly (N) tract longer than 9 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 genes. For all oligos in the *Medicago* Genome Oligo Set, using BLAST, each oligo is aligned against all 33,765 sequences (both TCs and singletons) in the *Medicago* MtGI 5.0 database. Using the alignment with the candidate oligo versus the highest scoring non-self gene, a BLAST percent identity score is computed. The highest scoring non-self gene 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.

This calculated cross-hybridization identity is dependent on the size of the sequence database used to BLAST against, oligo sequence, and use of either gapped or no-gap alignment method.

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 a small number of genes (3073 genes or 19%) that did not yield oligos satisfying all the above criteria, certain rules were relaxed.

SUMMARY

Oligo selection criteria	Criteria values	Number of oligos in genome set satisfying these criteria
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other genes Contiguous base match to any other gene	70mer 78°C ± 5°C ≤ 1000 ≤ 9 ≤ 9 ≤ 70% ≤ 20	13,013
Total number of oligos not satisfying one or more of the above criteria		3073
Length	45 ≤ x ≤ 50	3*
Melting temperature	66°C < x < 73 or 83°C < x < 86	83*
Location from 3' end	> 1000	1851*
Stem length in potential hairpin	9 < x ≤ 14	2*
Contiguous base match to any other gene	>20	935*
Cross-hybridization to all other genes	>70%	760*
Total		16,086

* Out of 3073 probes.

The following illustrations show the distribution of all 16,086 oligos representing the *Medicago* Genome Oligo Set 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

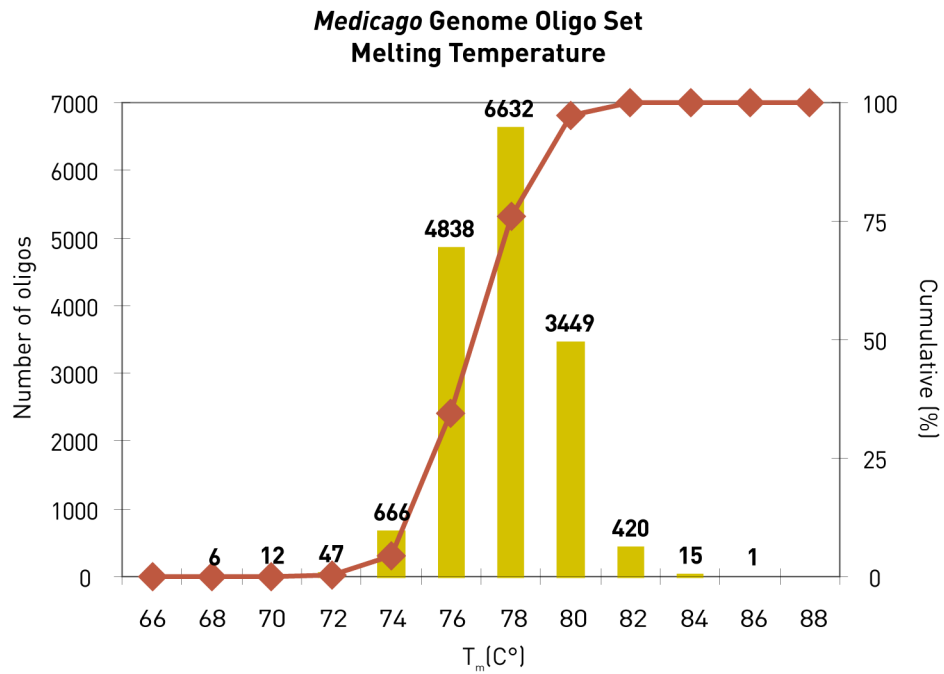


Figure 2. GC Content

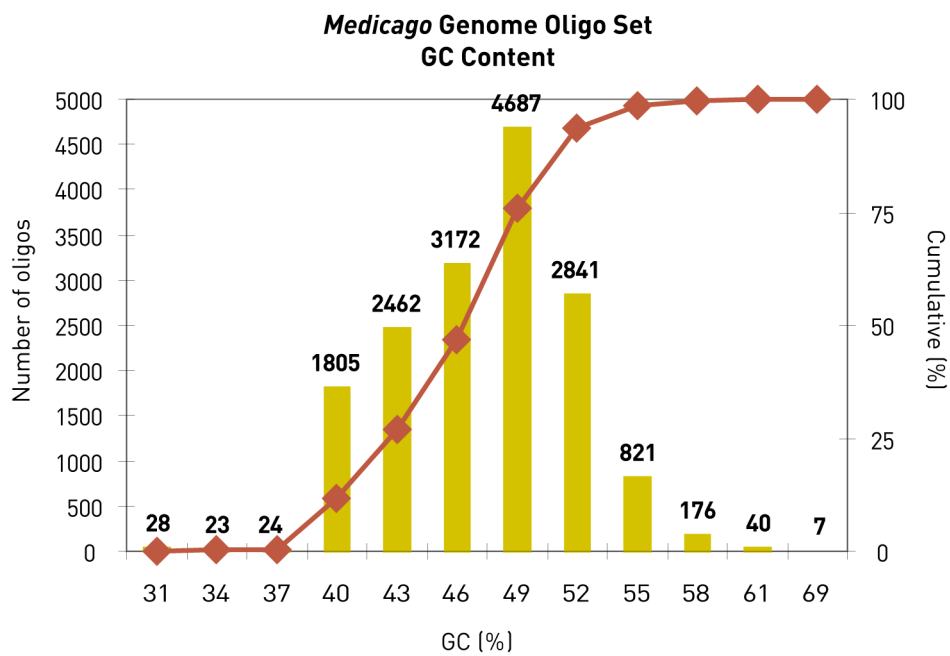


Figure 3. Location from 3' End

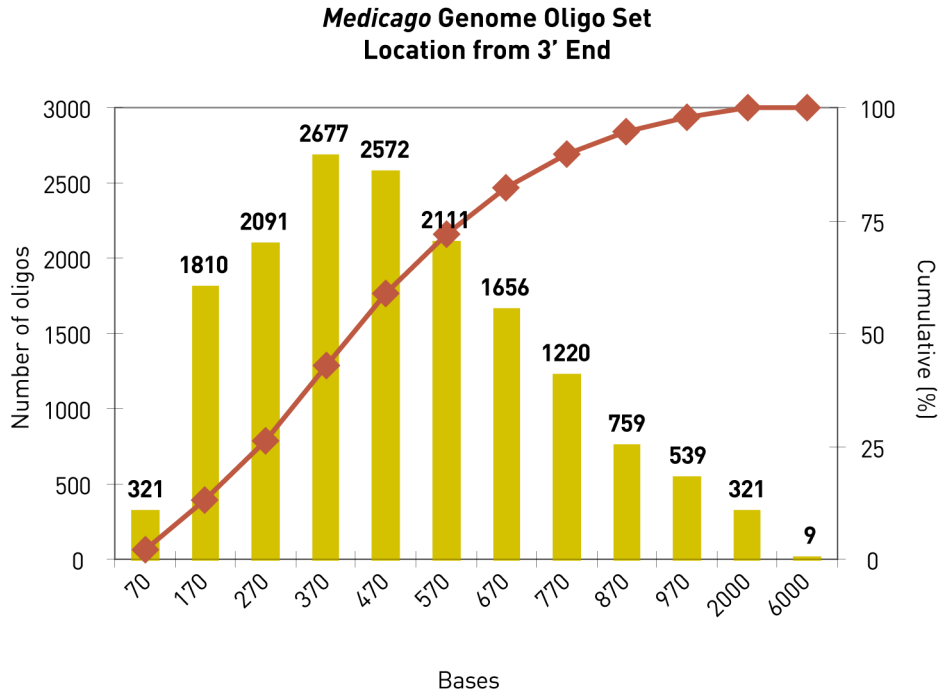


Figure 4. Longest Hairpin Stem Length

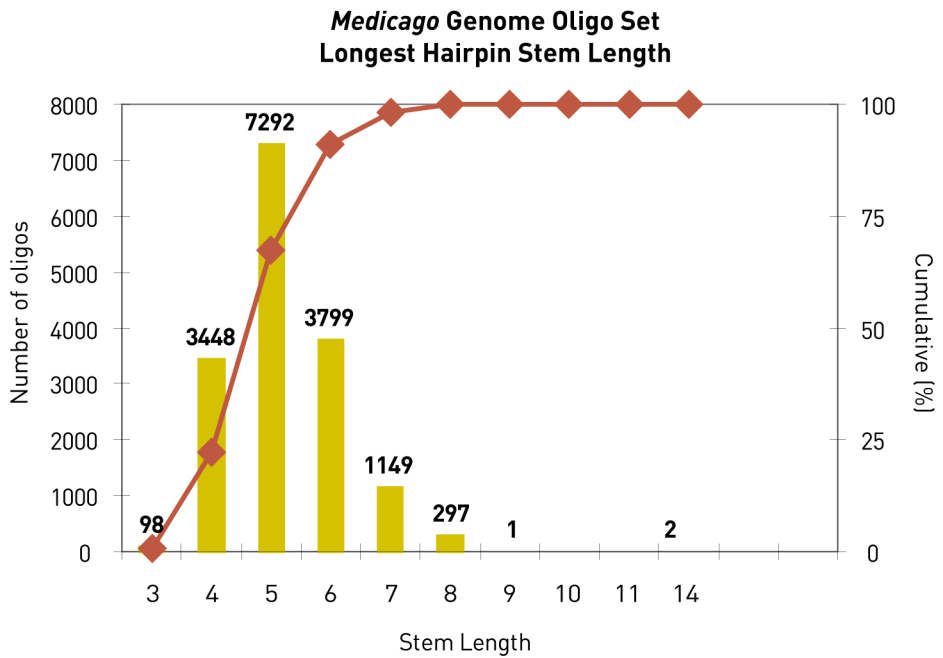
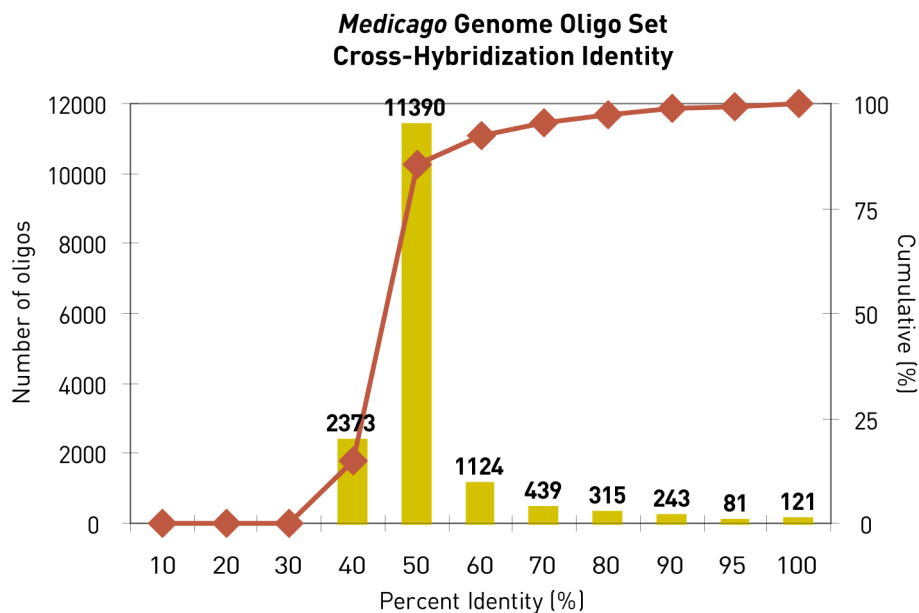


Figure 5. Cross-Hybridization Identity



Quality Check of Probe Design Specifications

Once the final oligo set 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 16,086 probes.

Probe design specification	Expected value	Verified range	Number of oligos
Melting temperature [C°]	78°C ± 5°C	73.6–82.9	16,003
Melting temperature [C°]	66°C–73°C	66.5°C–72.9°C	82
Melting temperature [C°]	83°C–86°C	85.9°C	1
Cross-hybridization identity [%]	≤ 70	31–70	15,326
Cross-hybridization identity [%]	> 70	71–100	760