

Array-Ready Oligo Set™ for the *Arabidopsis thaliana* Genome
Version 1.0 Upgrade

We are pleased to announce the *Arabidopsis thaliana* Genome Oligo Set Version 1.0 Upgrade. This upgrade contains 4104 arrayable 70mers representing 4039 genes and 4223 transcripts. All oligos in the upgrade represent either new genes or new gene transcripts not represented in the *A. thaliana* Genome Oligo Set Version 1.0. Probes in the upgrade are designed from gene transcripts from Release 4.0 of the TIGR (The Institute for Genomic Research) *A. thaliana* Genome Annotation Database. For our probe design we use state-of-the-art methodology and proprietary software. An amino linker is attached to the 5' end of each oligo.

Sequence source and selection

All probes are designed based on ATH1.cdna Release 4.0 of the TIGR *A. thaliana* Genome Annotation Database (<http://www.tigr.org/tdb/e2k1/ath1/ath1.shtml>).

All gene transcript sequences in the ATH1.cdna Release 4.0 were aligned to oligos from our *A. thaliana* Genome Oligo Set Version 1.0. Those sequences that did not yield an alignment greater than 95% over the length of the oligo were considered for probe design in the upgrade.

Probe design and selection rules

Oligos are classified as three oligo types depending on the number of transcripts represented: “common oligo,” “partial common oligo,” and “individual transcript oligo.” These three oligo classifications are essential for differentiating alternative splice variants and maximizing the number of represented transcripts. The common oligo type is used for representing all transcripts of one gene. The design platform makes use of these oligo type classifications. Most oligos in the upgrade are individual transcript oligos.

Oligo Type	Oligo Type Symbol	Definition	Number of Oligos
Common Oligo	C	The oligo represents all transcripts of one gene	183
Partial Common Oligo	P	The oligo represents a subset of transcripts of one gene	4
Individual Transcript Oligo	I	The oligo represents only one transcript of one gene	3917

Non-self transcripts for common oligos are all transcripts of other genes. Non-self transcripts for partial common oligos are all transcripts not represented by the oligo. This classification of self and non-self transcripts is used below for computing certain design criteria.

Once a sequence 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 *A. thaliana* Genome Oligo Set Version 1.0 Upgrade.

1) All oligos are within $78^{\circ}\text{C} \pm 5^{\circ}\text{C}$ 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) 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 8 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 transcripts. Using BLAST (Basic Local Alignment Search Tool), each oligo is aligned against all transcripts in ATH1.cdna. A cross-hybridization identity score is computed versus the top non-self transcript. A non-self transcript is defined above.

6) Each oligo cannot have greater than 20 contiguous bases common to any nonself transcripts.

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.

For 587 sequences in the upgrade that did not yield oligos satisfying all the above criteria, note that certain rules were relaxed.

SUMMARY

Oligo Selection Criteria	Value	Number of Oligos in Genome Set Satisfying These Criteria
Length	70mer	3517
Melting temperature	78°C ± 5°C	
Poly(N)tract length	≤ 8	
Stem length in potential hairpin	≤ 9	
Cross-hybridization to all other genes§	≤ 70%	
Contiguous base match to any other gene§	≤ 20	
Total number of oligos not satisfying one or more of the above criteria		587
Contiguous base match to any other gene§	> 20	470*
Cross-hybridization to all other genes§	> 70%	359*
Total		4104

*Out of 587 probes.

§For common oligos, the top non-self transcript is always a transcript of another gene. For a partial common oligo, the top non-self transcript can be any transcript other than the transcripts represented by the partial common oligo.

The following illustrations show the distribution of all 4014 oligos representing the *A. thaliana* Genome Oligo Set Version 1.0 Upgrade for melting temperature, GC content, location from 3' end, longest stem length, and cross-hybridization identity.

Figure 1. Melting Temperature

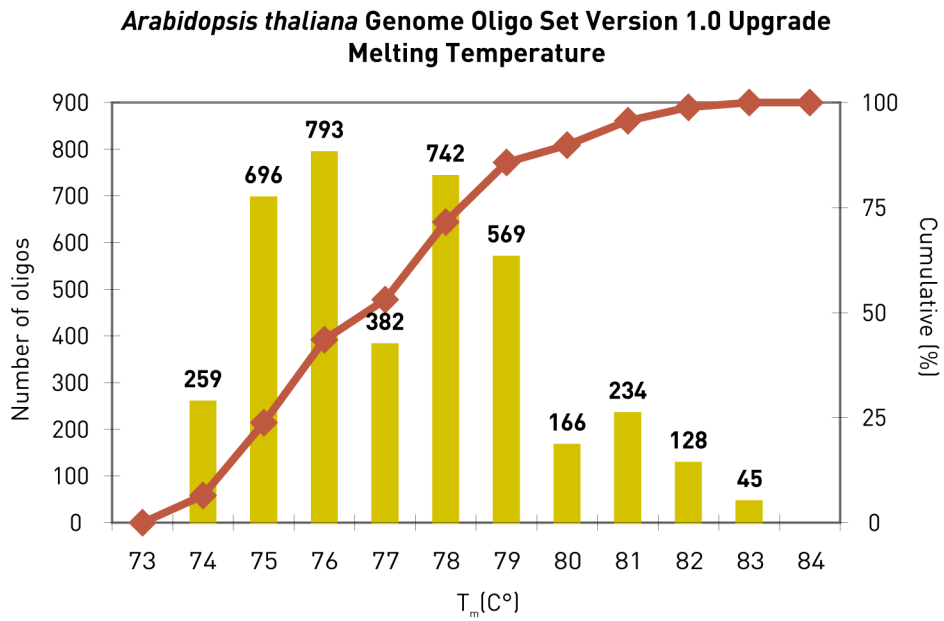


Figure 2. GC Content

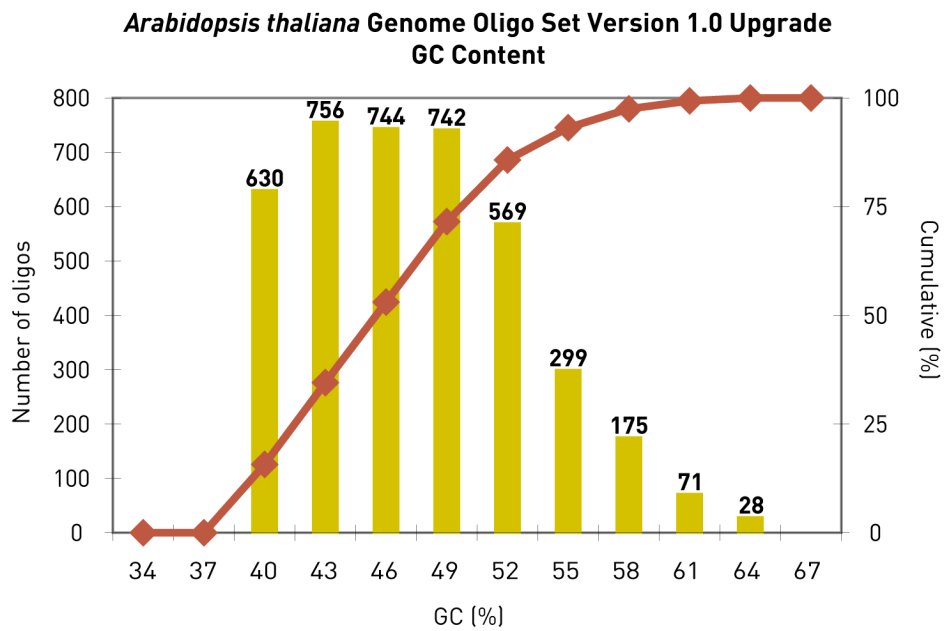


Figure 3. Locations from 3' End

Common and partial common oligos have multiple locations from 3' end shown as they represent multiple transcripts. Individual transcript oligos have only one location from 3' end shown. Total number of 3' end locations shown is 4322.

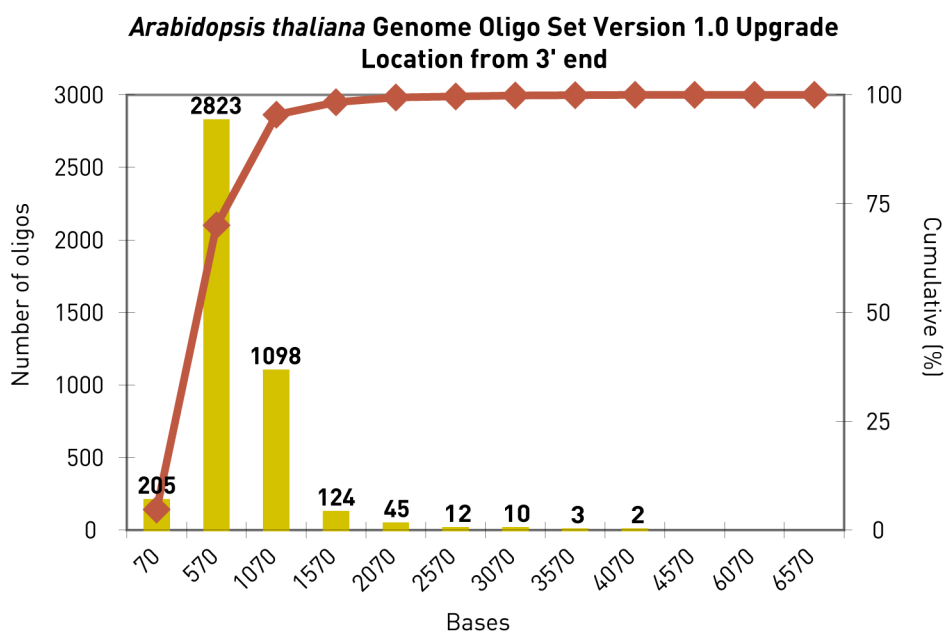


Figure 4. Longest Hairpin Stem Length

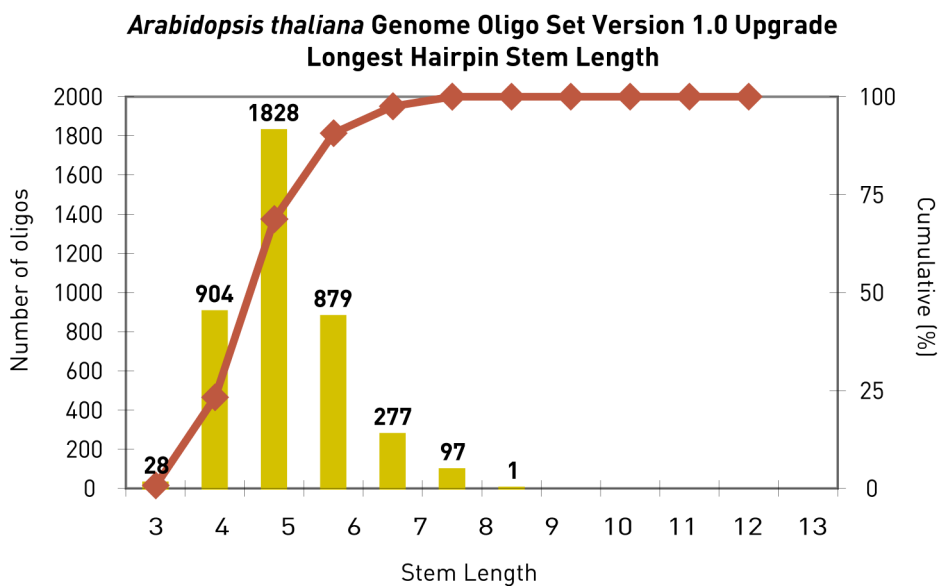
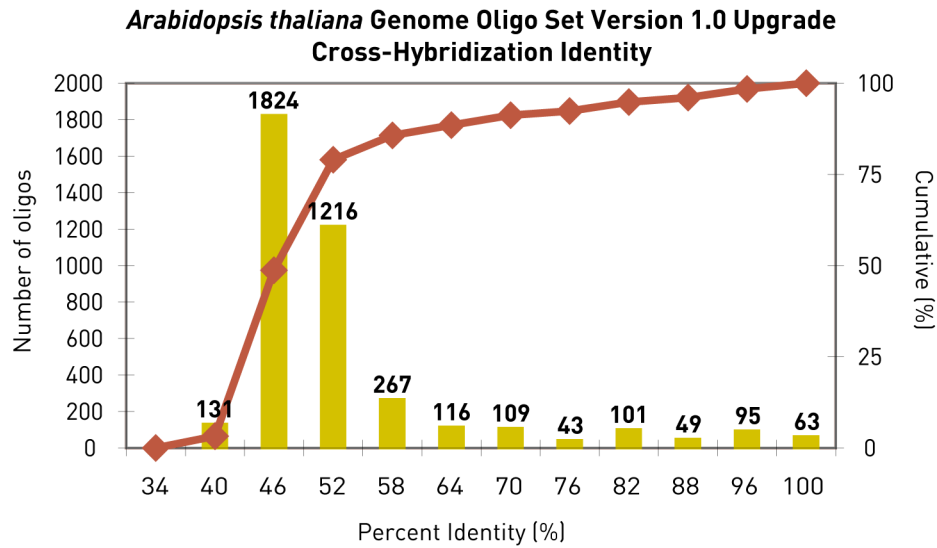


Figure 5. Cross-Hybridization Identity



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 probes.

Probe Design Specification	Expected Value	Verified Range	Number of Oligos <i>A. thaliana</i> Genome Oligo Set Version 1.0 Upgrade
Melting Temperature (°C)	78°C ± 5°C	73.6 - 82.9	4014
Cross-Hybridization Similarity (%)	≤ 70	37 - 70	3745
Cross-Hybridization Similarity (%)	> 70	71 - 100	359