

Array-Ready Oligo Set™ for the *Arabidopsis thaliana* Genome  
Version 3.0

The *Arabidopsis thaliana* Genome Oligo Set Version 3.0 contains 29,110 longmer probes and represents 26,173 protein-coding genes, 28,964 protein-coding gene transcripts, and 87 microRNA genes from *A. thaliana*. Pseudogenes are excluded from oligo design. The design is based on the ATH1 release 5.0 of the TIGR *A. thaliana* genome annotation database and release 4.0 of the miRNA Registry at the Sanger Institute and allows detection of alternative splicing variants using common, partial common, or individual transcript oligos. For 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 the file ATH1.cdna from the ATH1 release 5.0 of the TIGR *A. thaliana* genome annotation database (<http://www.tigr.org/tdb/e2k1/ath1/>) and release 4.0 of the miRNA Registry at the Sanger Institute (<http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml>).

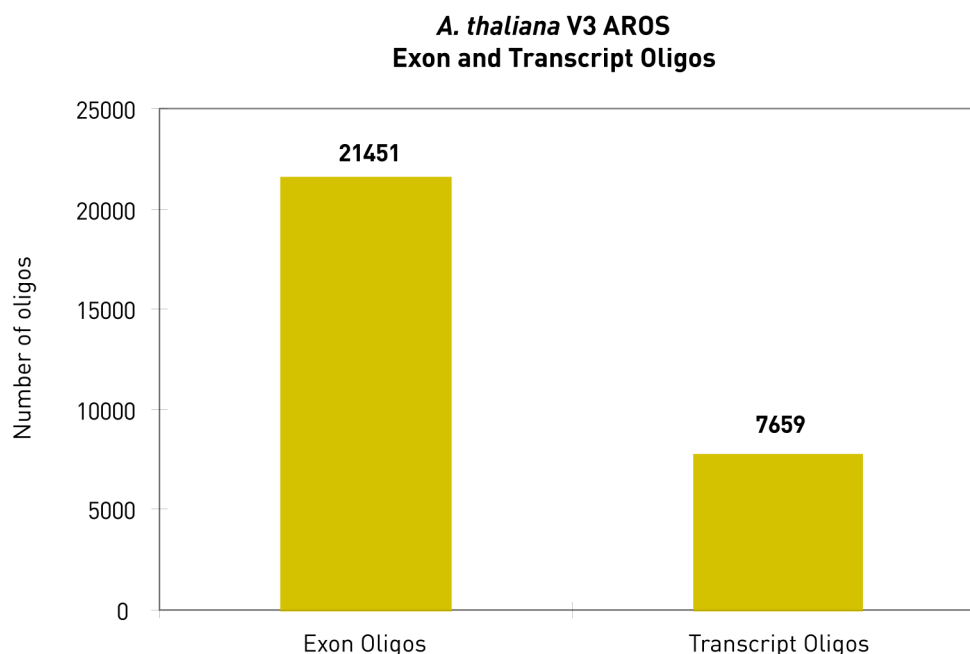
Probe Design and Selection Rules

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	2257
Partial Common Oligo	P	The oligo represents a subset of transcripts of one gene	26
Individual Transcript Oligo	I	The oligo represents only one transcript of one gene	26,742

The concept of an “exon oligo,” which is an oligo fully contained in an exon, was previously used to experimentally identify alternative splice variants in the human genome in a study by Shoemaker et al. (1). A “transcript oligo” is an oligo contained in multiple exons. In the *A. thaliana* V3 AROS, there are 21,451 exon oligos and 7659 transcript oligos.

Figure 1: Number of *Exon* and *Transcript* Oligos



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 set:

- 1) All oligos are within  $78^{\circ}\text{C} \pm 5^{\circ}\text{C}$  using the following formula:  
$$T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C) / \text{length} - 500 / \text{length}$$
 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 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 non-self 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 a number of probes in the set that did not yield oligos satisfying all the above criteria, certain rules were relaxed. 507 probes are less than 70mers.

SUMMARY

Oligo Selection Criteria	Criteria Values	Percent of oligos in the <i>A. thaliana</i> V3 satisfying these criteria
Length	70mer	98.26%
Contiguous base match to any other gene <sup>§</sup>	≤20	92.00%
Cross-hybridization to all other genes <sup>§</sup>	≤70%	93.67%
Poly(N)tract length	≤8	99.97%

<sup>§</sup>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 oligos for melting temperature, GC content, location from 3' end, longest stem length, and cross-hybridization identity.

Figure 2. Melting Temperature

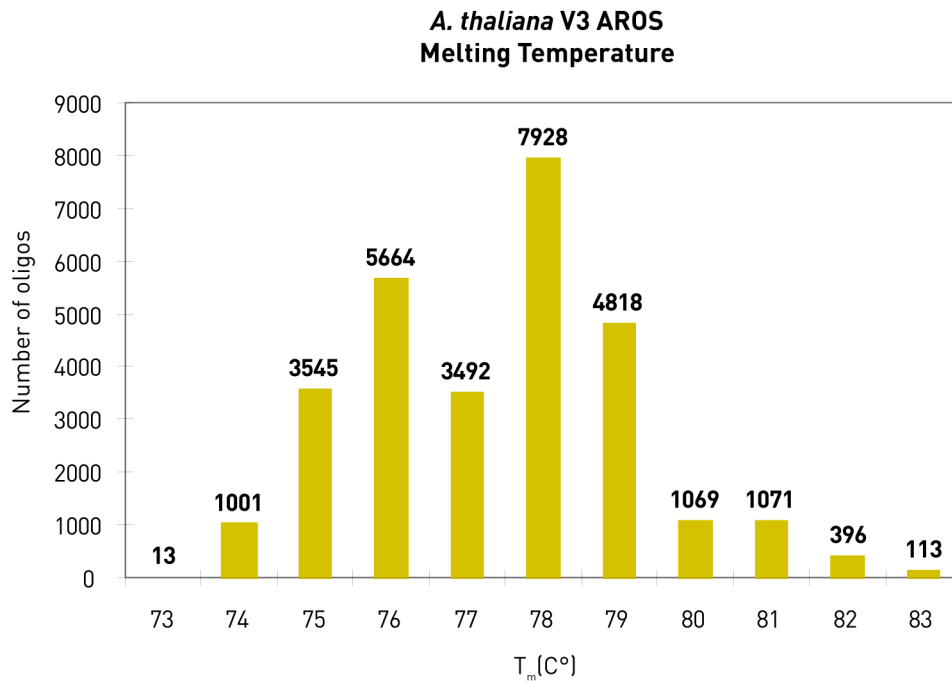


Figure 3. GC Content

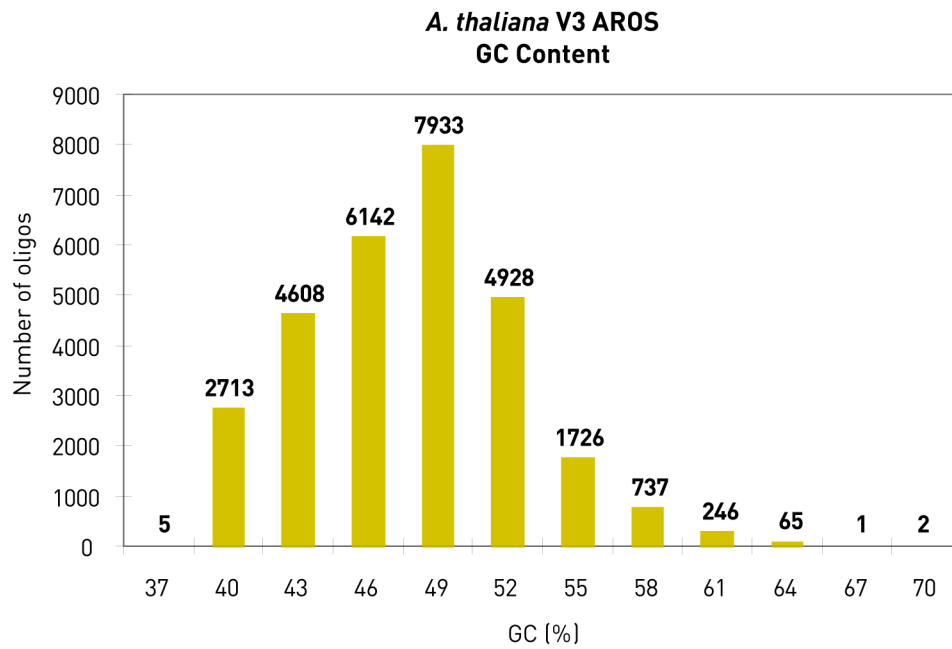


Figure 4. Locations from 3' End

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.

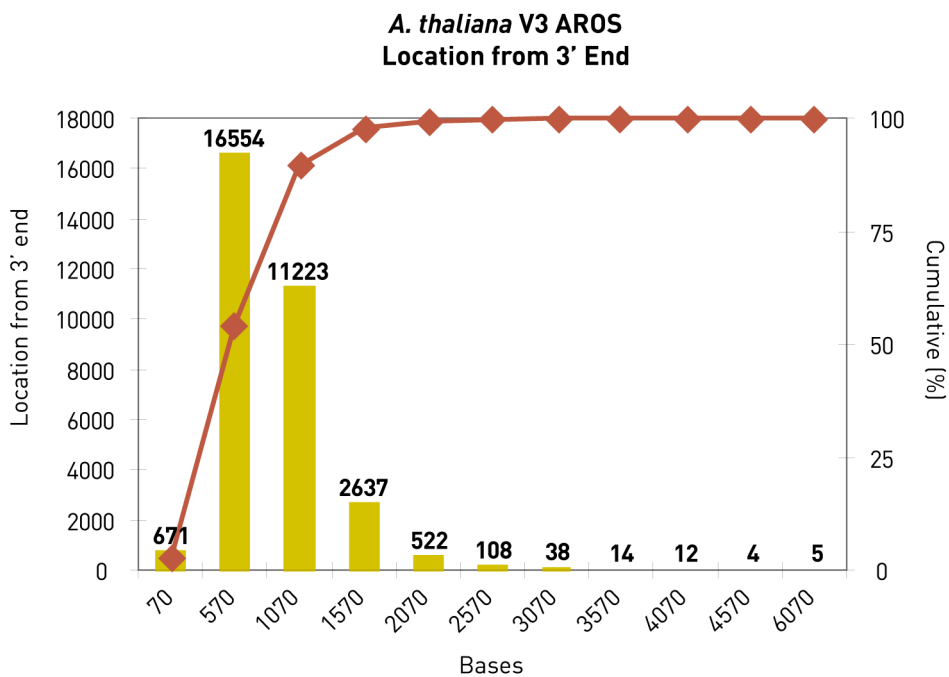


Figure 5. Longest Hairpin Stem Length

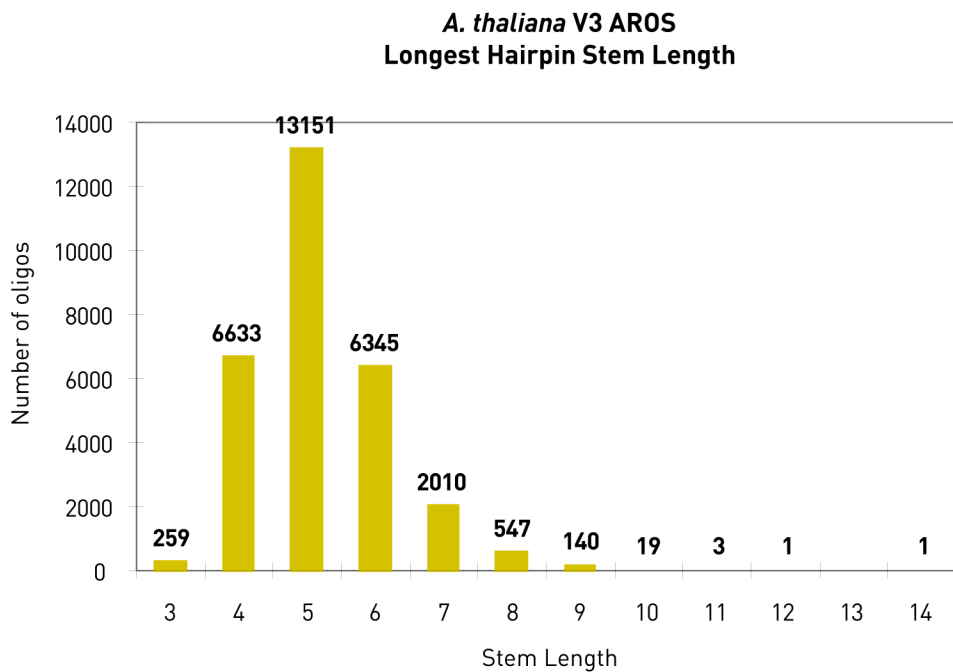
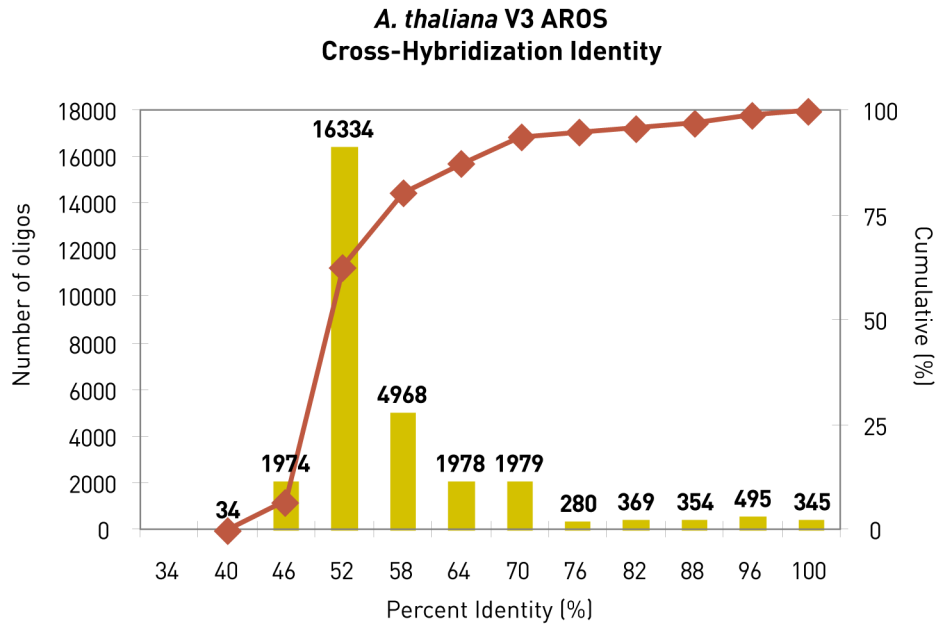


Figure 6. Cross-Hybridization Identity



References

1. Shoemaker, D.D. et al. (2001) *Experimental annotation of the human genome using microarray technology. Nature.* 409, 922-7.