

Array-Ready Oligo Set™ for the Tomato Genome Version 1.0

The Tomato Genome Oligo Set contains 12,160 oligos including 300 controls and represents 11,862 genes from tomato. Oligos in this set are mostly designed from gene transcript sequences from the *Lycopersicon Combined Build #3* unigene database from Cornell University and a remaining number of oligos are designed from sequences from GenBank. 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

11,769 probes are designed based on gene transcript sequences from the *Lycopersicon Combined Build #3* unigene database (<http://www.sgn.cornell.edu/>) and 93 probes are designed from GenBank gene transcript sequences (<http://www.ncbi.nlm.nih.gov>).

Probe Design and Selection Rules

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) An oligo is 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 the *Lycopersicon Combined unigene* database. A cross-hybridization identity score is computed versus the top non-self transcript.
- 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 cross-hybridization 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.

SUMMARY

Oligo selection criteria	Criteria Values	Percent of oligos in the Tomato V1 AROS satisfying these criteria
Length	70mer	95.6%
Contiguous base match to any other sequence	≤ 20	95.5%
Cross-hybridization to all other sequences	≤ 70%	92.3%
Poly(N)tract length	≤ 8	99.9%

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 1. Melting Temperature

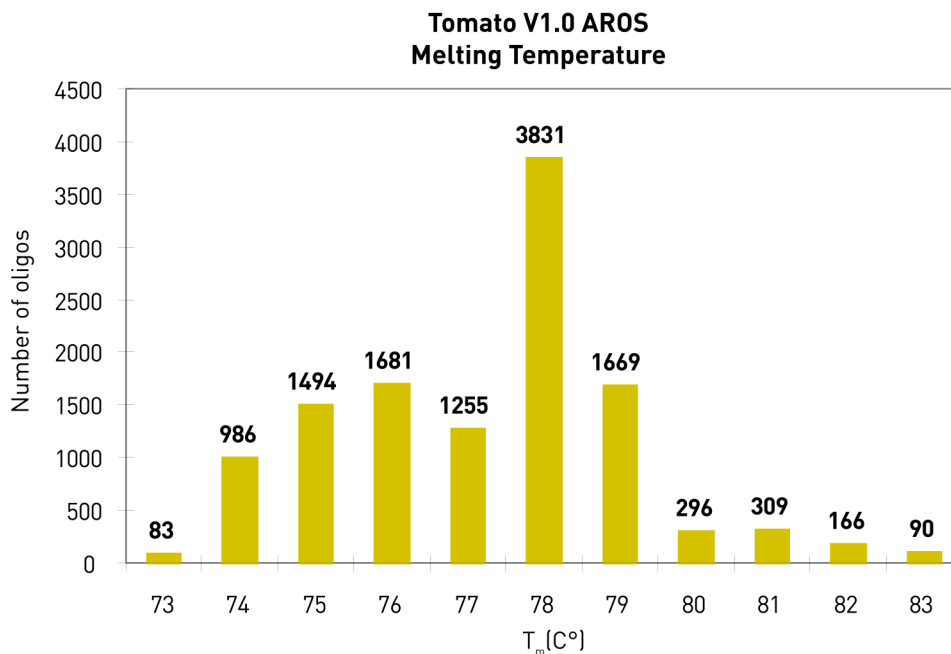


Figure 2. GC Content

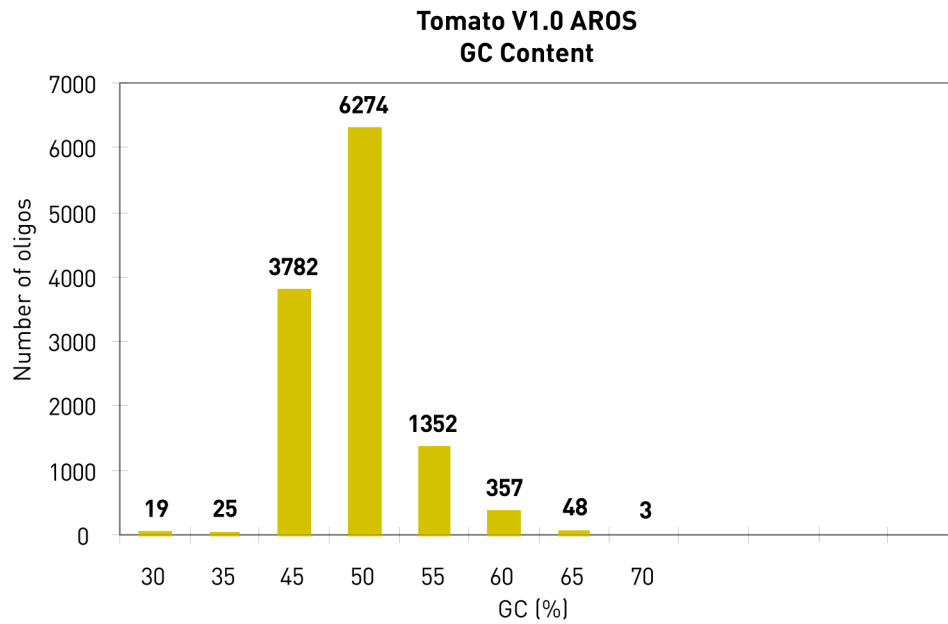


Figure 3. Location from 3' End

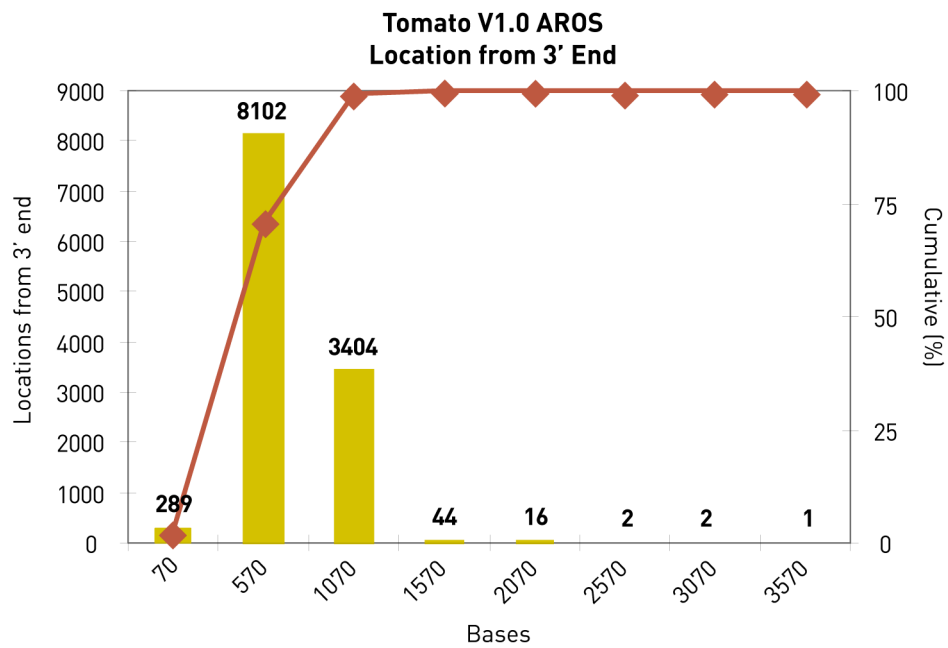


Figure 4. Longest Hairpin Stem length

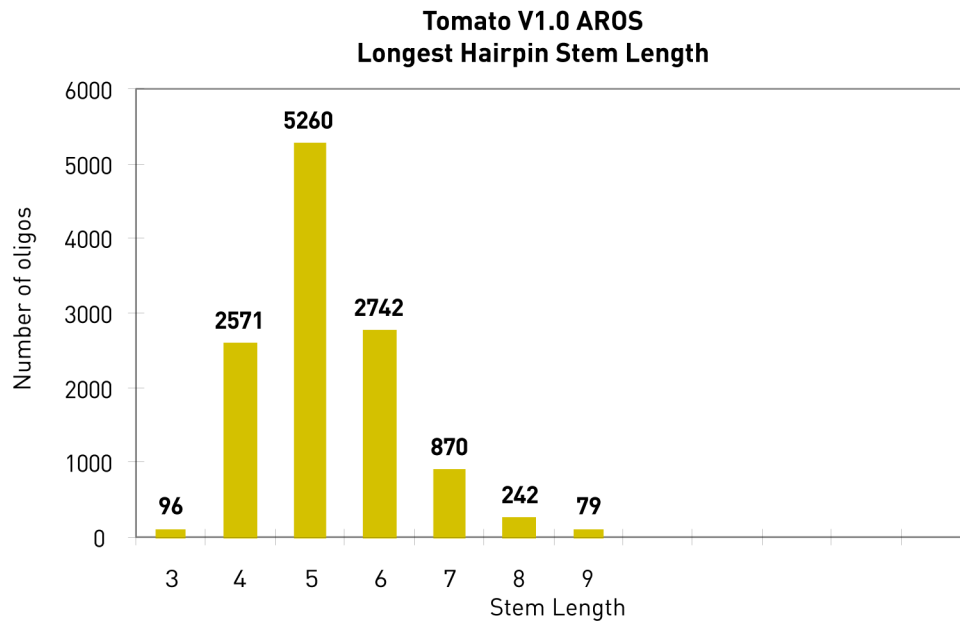


Figure 5. Cross-hybridization identity

