

Array-Ready Oligo Set™ for the Drosophila Genome
Version 1.1

We are pleased to announce the release of our complete Drosophila melanogaster Genome Oligo Set containing 14,593 70mer probes representing 13,664 genes and 17,899 transcripts. The design directly deals with alternative splicing variants using common, partial common, or individual transcript oligos. This Array-Ready Oligo Set™ (AROS) comes with Gene Ontology (GO) annotation. For our probe design we use state-of-the-art methods and proprietary software. An amino linker is attached to the 5' end of each oligo.

Gene Transcript Sequence Source

All 14,593 probes are designed from gene transcript sequences from the Gadfly release 3.1 database whole_genome_transcript_dmel_RELEASE3-1.FASTA [Berkeley Drosophila Genome Project [<http://www.bdgp.org>]].

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.

Oligo type	Oligo type symbol	Definition	Number of oligos
Common oligo	C	The oligo represents all transcripts of one gene.	1912
Partial common oligo	P	The oligo represents a subset of transcripts of one gene.	308
Individual transcript oligo	I	The oligo represents only one transcript of one gene.	12,373

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.

Sufficient numbers of 70mer candidate probes for each gene transcript are selected using the following criteria:

- 1) All oligos are within $77^{\circ}\text{C} \pm 5^{\circ}\text{C}$ 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 transcript sequence.
- 3) An oligo cannot have a contiguous single nucleotide 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, each transcript oligo is aligned against all 18,860 transcripts in whole_genome_transcript_dmel_RELEASE3-1.FASTA. A cross-hybridization identity score is computed versus the top non-self transcript. A non-self transcript is defined above.

6) Each transcript oligo cannot have greater than 20 contiguous bases common to any non-self transcripts.

For a number of transcripts that did not yield oligos satisfying all the above criteria, certain rules were relaxed. Certain selection rules such as oligo length, location, crosshybridization identity, and contiguous bases were relaxed.

Once oligo candidates have been selected satisfying the selection rules mentioned above, each oligo is ranked based on cross-hybridization identity.

By selecting a combination of common oligos, partial common oligos, and individual transcript oligos from both exon and transcript oligos based on minimized crosshybridization identity scores, this set was designed to differentiate alternative splicing variants.

A summary of the selection criteria is shown in the table below. Complete data for the 3' end location criteria is shown in Figure 3.

SUMMARY

Oligo selection criteria	Value	Number of oligos in genome set satisfying these criteria
Length Melting temperature Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other ORFs [†] Contiguous base match to all other ORFs [†]	70mer 77°C ± 5°C ≤8 ≤9 ≤70% ≤20	13,797
Length	<70	151*
Stem length in potential hairpin	9 < x ≤13	5*
Contiguous base match to all other transcripts [†]	>20	658*
Cross-hybridization identity to all other transcripts [†]	>70%	347*
Number of oligos not satisfying one or more of the above criteria		796
Total		14,593

*Out of 796 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 14,593 oligos for melting temperature, GC content, location from 3' end of gene sequence, hairpin stem length, and cross-hybridization identity.

Figure 1. Melting Temperature

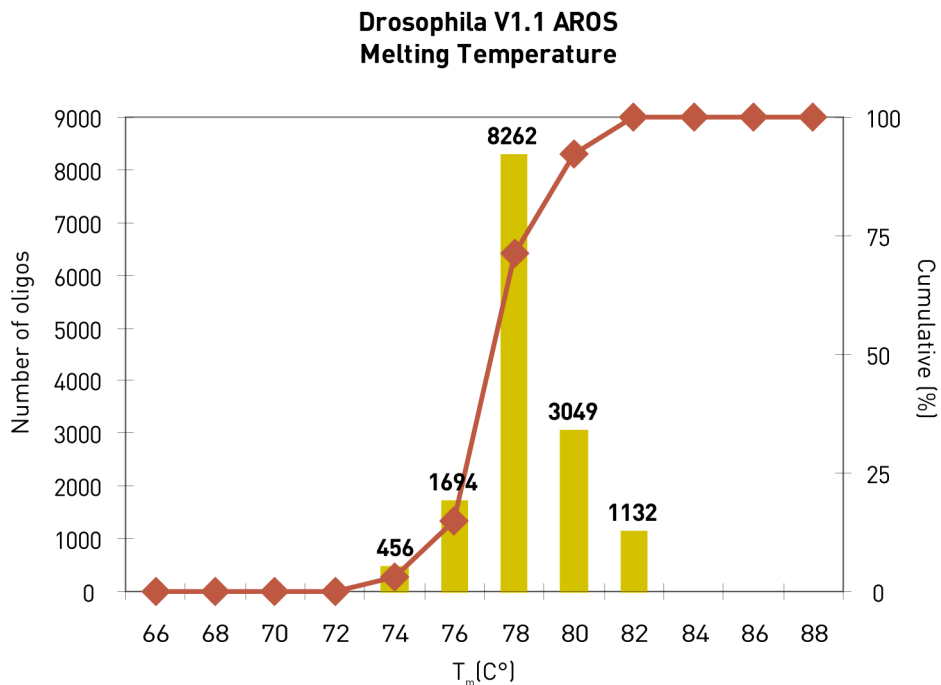


Figure 2. GC Content

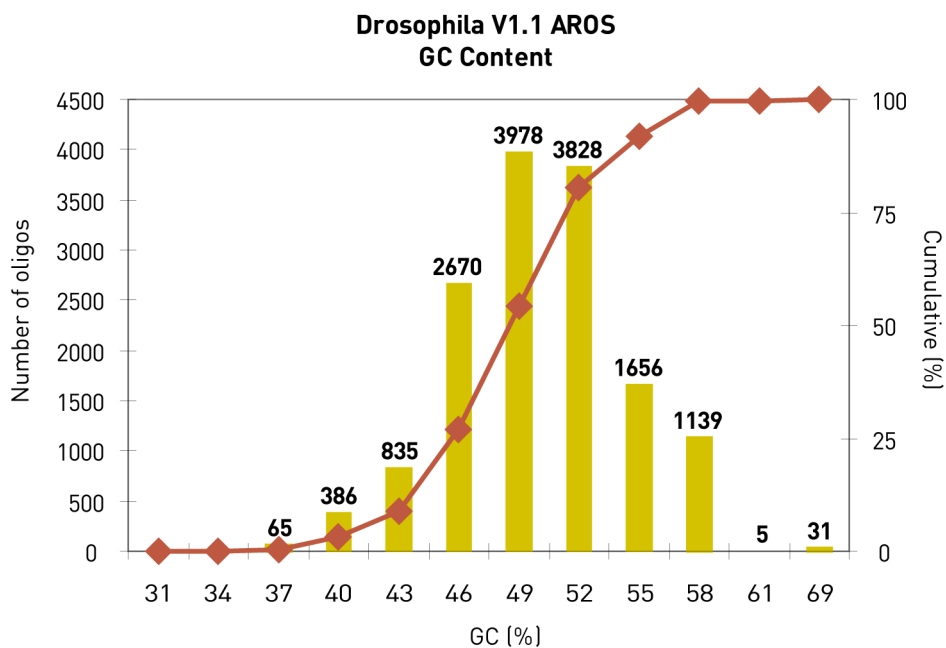


Figure 3. Locations from 3' End

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 18,176.

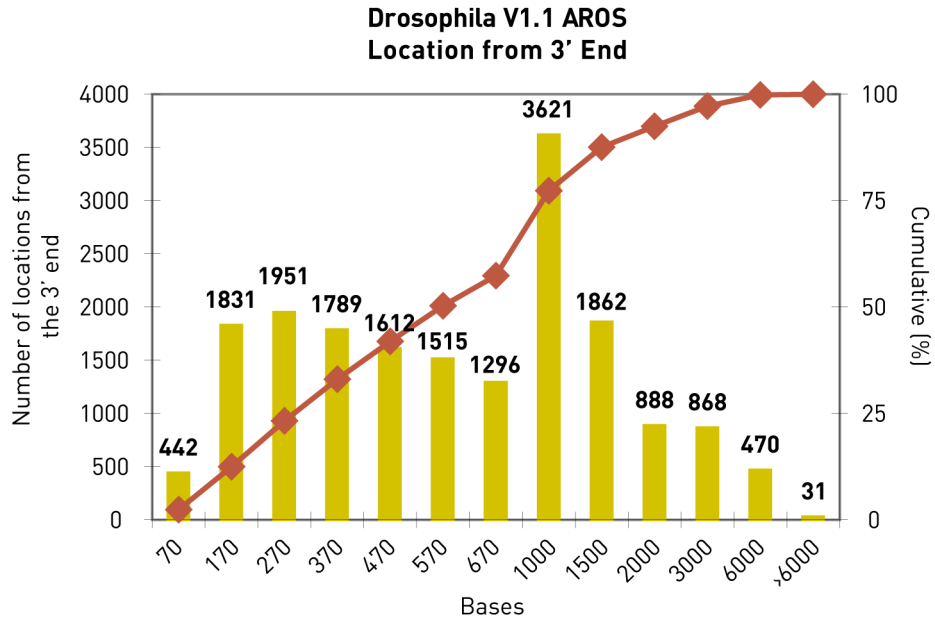


Figure 4. Longest Hairpin Stem Length

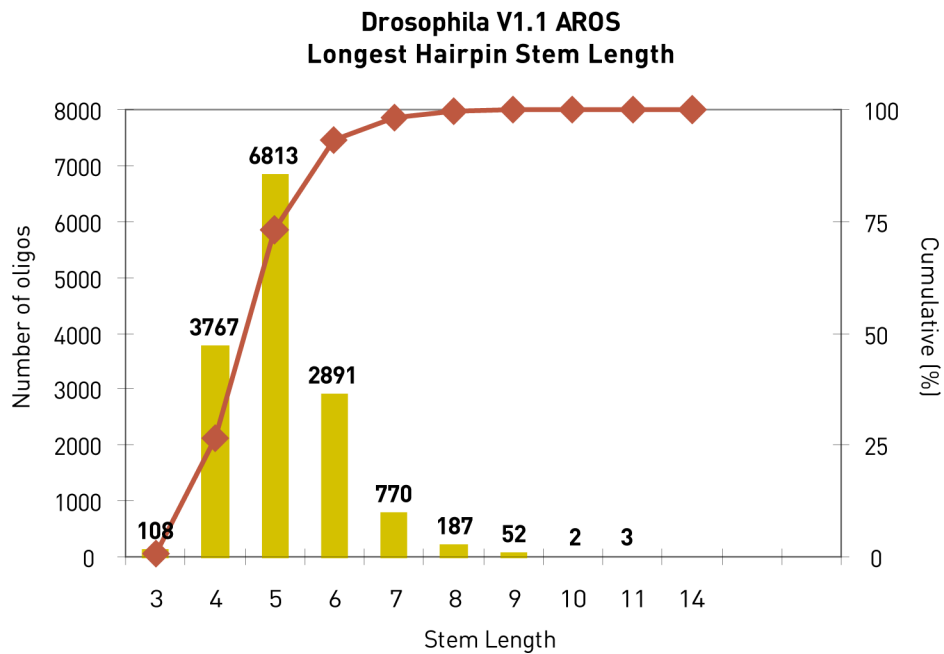
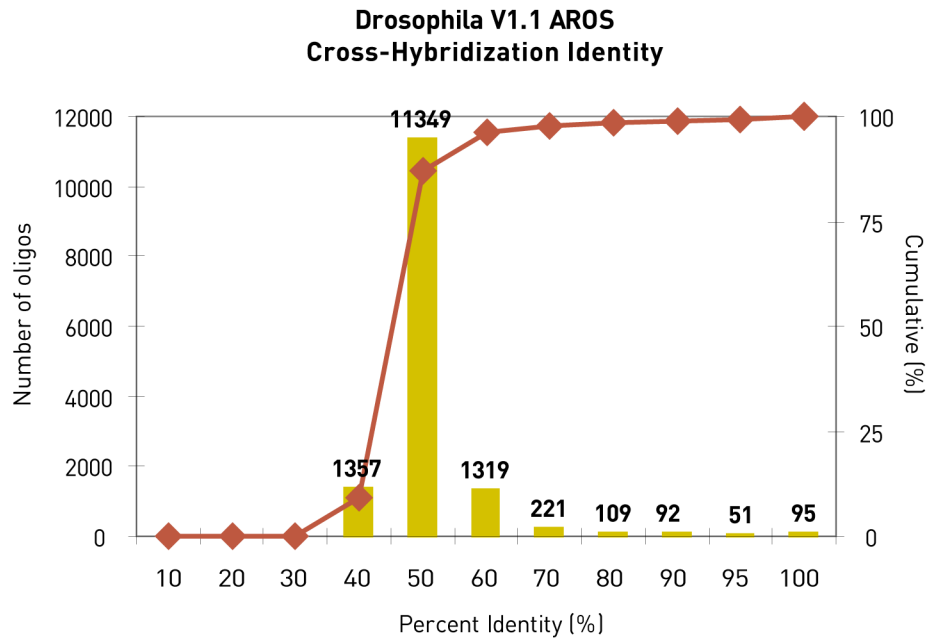


Figure 5. Cross-Hybridization Identity



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 14,593 oligos in the set.

Probe design specification	Expected value	Verified range	Number of oligos
Melting temperature (C°)	77°C ± 5°C	72.1–82.0	14,593
Hairpin stem length	≤9	3–9	14,588
Hairpin stem length	>9	10–11	5
Cross-hybridization identity (%)	≤70	34–70	14,246
Cross-hybridization identity (%)	71–100	71–100	347