

Array-Ready Oligo Set™ for the *Caenorhabditis elegans* Genome
Version 1.1

We are pleased to announce the release of our *Caenorhabditis elegans* Genome Oligo Set Version 1.1 containing 19,873 70mer probes representing 20,445 *C. elegans* gene transcripts. Version 1.1 contains all 7125 probes from Version 1.0, plus 12,748 probes (upgrade). 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.

Gene Transcript Sequence Source and Selection

The 19,873 probes are designed from the WormPep DNA Release 75 (March 2002) database, developed and maintained by the Sanger Institute (<http://www.sanger.ac.uk>). The *C. elegans* sequencing project was completed by the Sanger Institute and Washington University School of Medicine. All probes are designed within the open reading frame (ORF).

Assigning the probes to WormPep DNA Release 93

To find out which transcripts are represented in WormPep DNA Release 93 released in December 2002 (also developed and maintained by the Sanger Institute), the 19,873 probes were aligned using BLAST to the Release 93 transcripts. Furthermore, all the Release 75 transcripts were aligned using BLAST versus the Release 93 transcripts. A probe is marked to represent a certain Release 93 transcript if the probe has > 97% identity over the length of the oligo. Additionally, the Release 93 transcript and the Release 75 transcript, used to design the oligo, have an E value score < 1e-20 in the BLAST alignment (the E value score is a measure of statistical significance using BLAST).

	WormPep DNA Release 75	WormPep DNA Release 93
Total number of genes	19,430	19,517
Total number of gene transcripts	20,391	21,197
Number of genes represented by a probe in V. 1.1	19,430	19,060
Number of gene transcripts represented by a probe in V. 1.1	20,391	20,445

Splice variants in WormPep DNA

A gene can undergo alternative splicing and produce multiple transcripts or splice variants. A probe designed for a splice variant can be designed from a unique region such that the probe can distinguish that transcript from all others. Such a probe is said to be differential. A probe can also be designed from a region that is common with other splice variants or transcripts. That probe is said to be common to a set of transcripts.

In WormPep DNA Release 75, 1747 splice variants are present. A fraction of the splice variants yield probes such that the probe is unique to that one splice variant or differential. Of these splice variants, 757 are represented by at least one probe that is unique. This unique probe occurs in only one transcript. Additionally, 1397 of these splice variants have at least one probe that is common to more than one transcript.

In WormPep DNA Release 93, 2927 splice variants are present. Using the probe assignment mentioned in the above section to the transcript sequences from WormPep DNA Release 93 using a greater than 97% identity over the length of the oligo, the number of splice variants represented is as follows. A total of 2522 splice variants are represented by at least one oligo. Furthermore, 791 of these splice variants are represented by at least one oligo

where that oligo represents only one transcript. Additionally, 2120 splice variants are represented by at least one oligo that is also common to another transcript.

Splice Variant	Number of splice variants in WormPep DNA Release 75	Number of splice variants in WormPep DNA Release 93
Total	1747	2927
Represented by at least one oligo	1747	2522
Represented by at least one oligo that is unique (an oligo that represents only one transcript)	757	791
Represented by at least one oligo that is common to more than one transcript	1397	2120

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. A sufficient number of 70mer candidate probes for each transcript are selected using the following criteria for the *C. elegans* Genome Oligo Set Version 1.1.

1) All oligos are within $76^{\circ}\text{C} \pm 5^{\circ}\text{C}$ 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 ORF 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) A normalized score is assigned to each oligo based on the number of repeats. Oligos with more repeats having a normalized score greater than a certain threshold are filtered out.

6) Each oligo has less than or equal to 70% identity to all other transcripts. Each oligo is aligned against all 20,391 transcripts in WormPep DNA Release 75 (March 2002). Using the alignment with the candidate oligo versus the highest scoring non-self sequence, a BLAST percent identity score is computed. The highest scoring non-self sequence is defined as the sequence that yields the most matched bases in an alignment. The cross-hybridization identity of the oligo is obtained using the number of matched bases and the oligo length.

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

7) Each oligo of any length cannot have greater than 20 contiguous bases common to any other transcript.

Once oligo candidates have been selected satisfying all the selection rules mentioned above, each oligo is ranked based on the cross-hybridization identity as computed in Step 6. One final oligo for each sequence is selected with the minimum cross-hybridization identity.

Note that for 1978 (10%) transcripts that did not yield oligos satisfying all the above criteria, certain rules were relaxed. For those transcripts, the oligo is located beyond 1000 bases from the 3' end, has greater than 70% cross-hybridization identity, or has a slightly higher or lower T_m .

SUMMARY

Oligo Selection Criteria	Value	Number of oligos in genome set satisfying these criteria (Version 1.0)	Number of oligos in genome set satisfying these criteria (Version 1.1- Upgrade)
Length Melting temperature Location from 3' end Poly(N)tract length Stem length in potential hairpin Cross-hybridization identity to all other transcripts Contiguous base match to all other transcripts	70mer 76°C ± 5°C ≤ 1000 ≤ 8 ≤ 9 ≤ 70% ≤ 20	7109	10,786
Number of oligos not satisfying one or more of the above criteria		16	1962
Location from 3' end	> 1000	16*	684*
Length	< 70	0	275*
Melting temperature	62°C < X < 71°C or 81°C < X < 82°C	0	8*
Cross-hybridization identity to all other transcripts	> 70	0	1099*
Contiguous base match to all other transcripts	> 20	0	788*
Total		7125	12,748

* Out of 1962.

The following illustrations show the distribution of all 19,873 oligos for melting temperature, GC content, location from 3' end of gene sequence, length of longest stem length, and cross-hybridization identity.

Figure 1. Melting Temperature

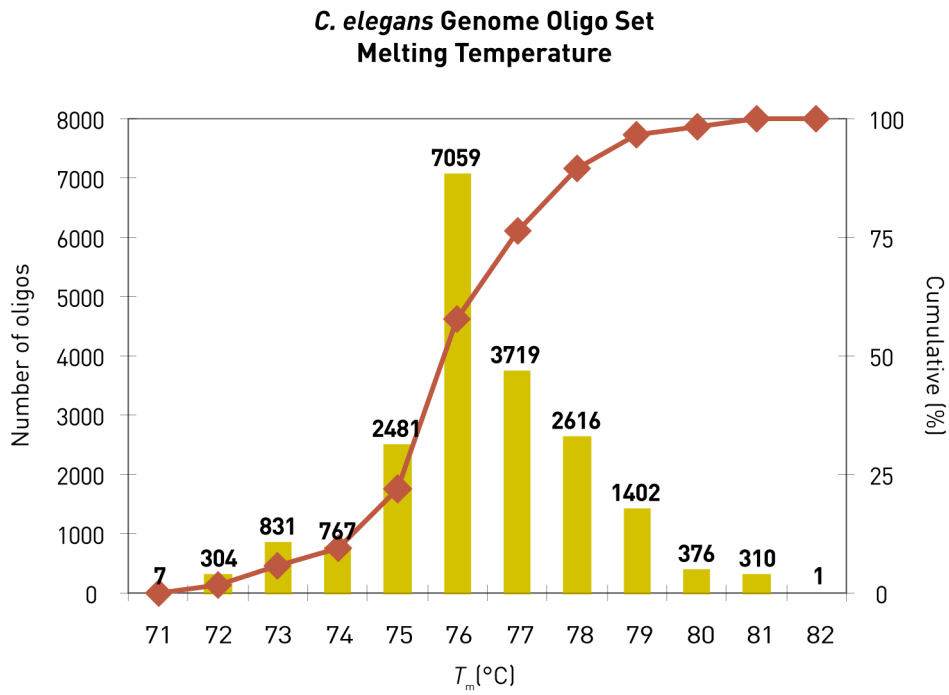


Figure 2. GC Content

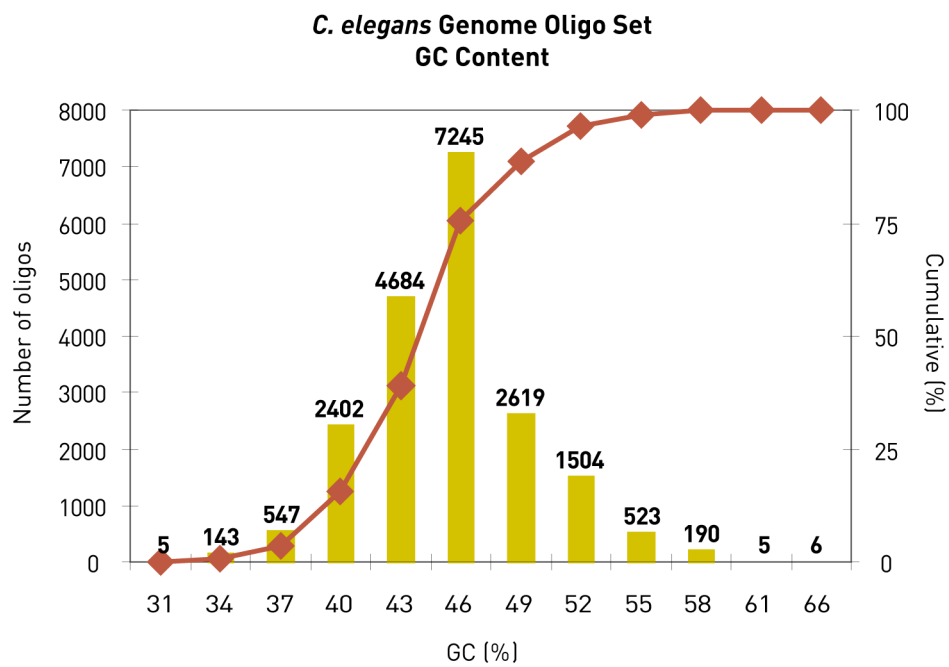


Figure 3. Location from 3' End

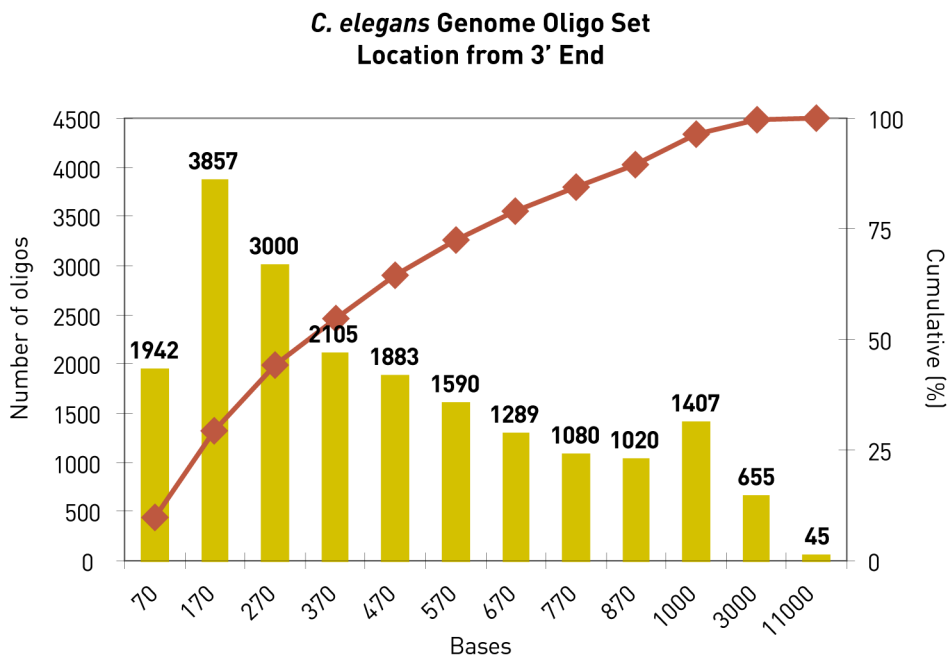


Figure 4. Length of the Longest Hairpin Stem

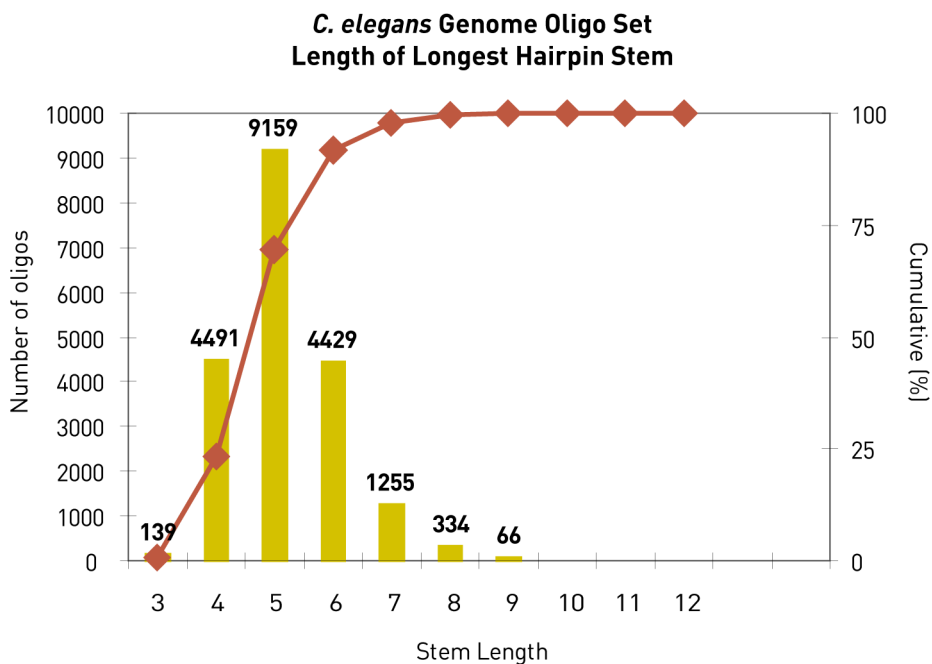
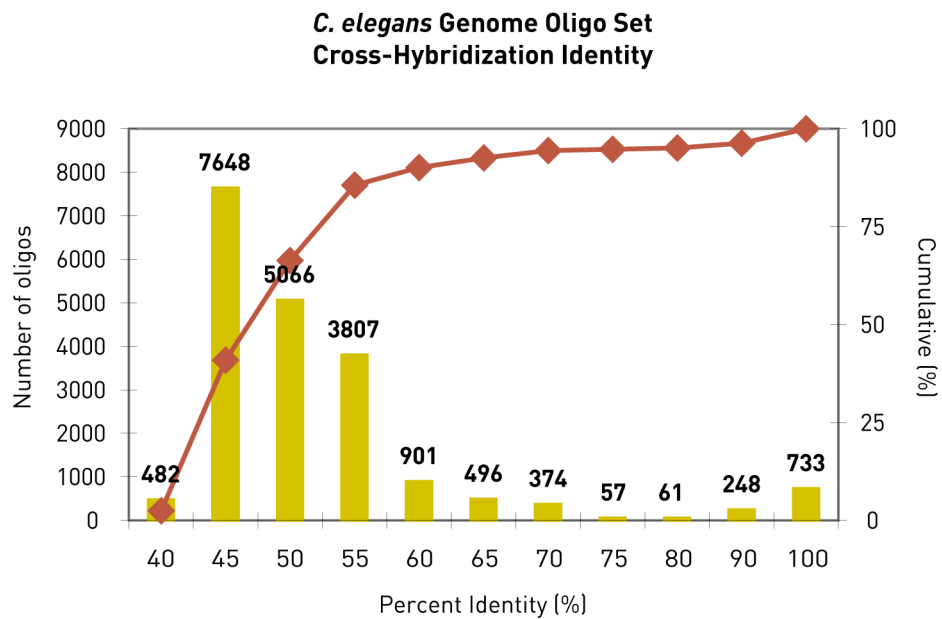


Figure 5. Cross-Hybridization Identity



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 19,873 oligos in the set.

Probe Design Specification	Expected Value	Verified Range	Number of Oligos
Melting temperature [C°]	76°C ± 5°C	71.2-80.6	19,865
Melting temperature [C°]	62°C < X < 71°C	62.4-70.5	7
Melting temperature [C°]	81°C < X < 82°C	81.6	1
Hairpin stem length	≤ 9	3-9	19,873
Cross-hybridization identity (%)	≤ 70	30-70	18,774
Cross-hybridization identity (%)	> 70	71-100	1099