# Array-Ready Oligo Sets™ for the Pig Genome, Version 1.0, and the Pig Genome Oligo Extension Set, Version 1.0

We are pleased to announce Version 1.0 of our Pig Genome Oligo Set and our Pig Genome Oligo Extension Set. The Pig Genome Oligo Set contains 10,665 70mer probes representing 10,665 *Sus scrofa* gene sequences with a hit to a known human, mouse, or pig gene transcript and some sequences containing a 3' expressed sequence tag (EST). The Pig Genome Oligo Extension Set contains 2632 70mer probes representing 2632 *Sus scrofa* gene sequences with at least one 3' EST. 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 sequence source and selection

All probes are designed from The Institute of Genome Research (TIGR*)* Gene Index Database SsGI Release 5.0, released on October 1, 2002. TIGR develops and maintains this database at http://www.tigr.org/tdb/tgi/ssgi/.

SsGI Release 5.0 contains a total of 49,201 unique sequences including 17,354 tentative consensus sequences (TCs), singleton 494 expressed transcripts (ETs), and 31,353 singleton ESTs. ETs represent mature transcripts and start with an "*NP*."

| TIGR Pig SsGI Release 5.0 Database | Number of sequences in SsGI Release 5.0 | Pig Genome Oligo Set Version 1.0 | Pig Genome Oligo Extension Set Version 1.0 |
|---|---|---|---|
| TCs | 17,354 | 10,313 | 2632 |
| Singleton ETs | 494 | 352 | 0 |
| Singleton ESTS | 31,353 | 0 | 0 |
| Total | 49,201 | 10,665 | 2632 |

## Pig Genome Oligo Set sequence selection

All TCs and singleton ETs were aligned using BLAST versus known gene transcripts for human, mouse, and pig. More specifically, these TCs and singleton ETs were aligned to 27,628 known human gene transcripts from ENSEMBL, 28,097 known mouse gene transcripts, and 1897 known pig gene transcripts from the Pig UniGene Build #10

(http://www.ncbi.nlm.nih.gov). Both the human and mouse ENSEMBL databases are from January 2003 and were obtained from http://www.ensembl.org .
All TCs and singleton ETs with a > 75% identity over at least 100 bases to a known human, mouse, or pig gene transcript and yielding a designed probe <=70 cross-hybridization identity is included in this set. For more details on cross-hybridization identity, see the probe design and selection rules below.

## Pig Genome Oligo Extension Set sequence selection

All component ESTs used to make the TCs were obtained from GenBank at http://www.ncbi.nlm.nih.gov. All ESTs with the keyword 3', denoting a 3 prime EST, were marked. A total of 7739 TCs were found to contain at least one 3' EST. The TCs that contain at least one 3' EST and are not present in the Pig Genome Oligo Set are included in the Pig Genome Oligo Extension Set.

| | Pig Genome Oligo Set Version 1.0 | Pig Genome Oligo Extension Set Version 1.0 |
|---|---|---|
| Number of oligos designed from a TC with at least one 3' EST | 5005 | 2632 |
| Number of oligos designed from a sequence with a hit to a known human, mouse, or pig gene transcript* | 10,665 | 172 |
| Number of oligos that have a <=70% cross-hybridization identity to another sequence[†] | 10,665 | 2538 |
| Total | 10,665 | 2632 |

*Aligning the original sequence versus a known human, mouse, or pig gene transcript at >75% over at least 100 bases.

[†] For more information, see the cross-hybridization identity step in the probe design and selection rules.

## Sequence orientation

TIGR obtains and predicts orientation for all the tentative consensus sequences and singletons based on various techniques including alignments to known proteins and poly A trimming. After SsGI 5.0 was released, TIGR later gave us updated orientation of 7218 of the sequences in this database. A total of 628 of these are TCs and the rest are singleton ESTs. Probes for these 628 TCs that appear in the sets are therefore designed in the updated orientation. In the gene list, a column indicates the orientation of the probe to the original TC sequence. All probes are designed in the sense strand as given by TIGR.

## 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. Sufficient numbers of 70mer candidate probes for each gene are selected using the following criteria for our Pig Genome Oligo Set and our Pig Genome Oligo Extension Set.

1) All oligos are within 78°C ±5°C using the following formula:
   $T_m = 81.5 + 16.6 \times \log[Na+] + 41 \times (\#G + \#C)/\text{length} - 500/\text{length}$
   where $[Na+] = 0.1$ 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 genes. For all oligos in the Pig Genome Oligo Set, using BLAST, each oligo is aligned against all 49,201 sequences (both TCs and singletons) in the Pig SsGI 5.0 database. Using the alignment with the candidate oligo versus the highest scoring non-self gene, a BLAST percent identity score is computed. The highest scoring non-self gene is defined as the sequence that yields the most matched bases in an alignment. This is referred to as the cross-hybridization identity of the oligo.

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

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

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.

Note that for 58 of the sequences in the Pig Genome Oligo Set and 214 of the sequences in the Pig Genome Oligo Extension Set that did not yield oligos satisfying all the above criteria, certain rules were relaxed.

SUMMARY

| Oligo selection criteria | Criteria values | Number of oligos in the Pig Genome Oligo Set Version 1.0 satisfying these criteria | Number of oligos in the Pig Genome Extension Set Version 1.0 satisfying these criteria |
|---|---|---|---|
| Length | 70mer | 10,607 | 2418 |
| Melting temperature | 78°C ±5°C | | |
| Location from 3' end | <=1000 | | |
| Poly(N)tract length | <8 | | |
| Stem length in potential hairpin | <9 | | |
| Cross-hybridization identity to all other sequences | <=70% | | |
| Contiguous base match to any other sequence | <=20 | | |
| Total number of oligos not satisfying one or more of the above criteria | | 58 | 214 |
| Length | 50mer | 27* | 10[†] |
| Location from 3' end | >1000 | 31* | 6[†] |
| Contiguous base match to any other sequence | >20 | 0* | 176[†] |
| Cross-hybridization to all other sequences | >70% | 0* | 94[†] |
| Total | | 10,665 | 2632 |

*Out of 58 probes.

[†] Out of 214 probes.

The following illustrations show the distribution of all 10,665 oligos representing the Pig Genome Oligo Set Version 1.0 followed by the 2632 oligos from the Pig Genome Oligo Extension Set for melting temperature, GC content, location from 3' end, longest stem length, and cross-hybridization identity.

Figure 1. Melting Temperature — Pig Genome Oligo Set

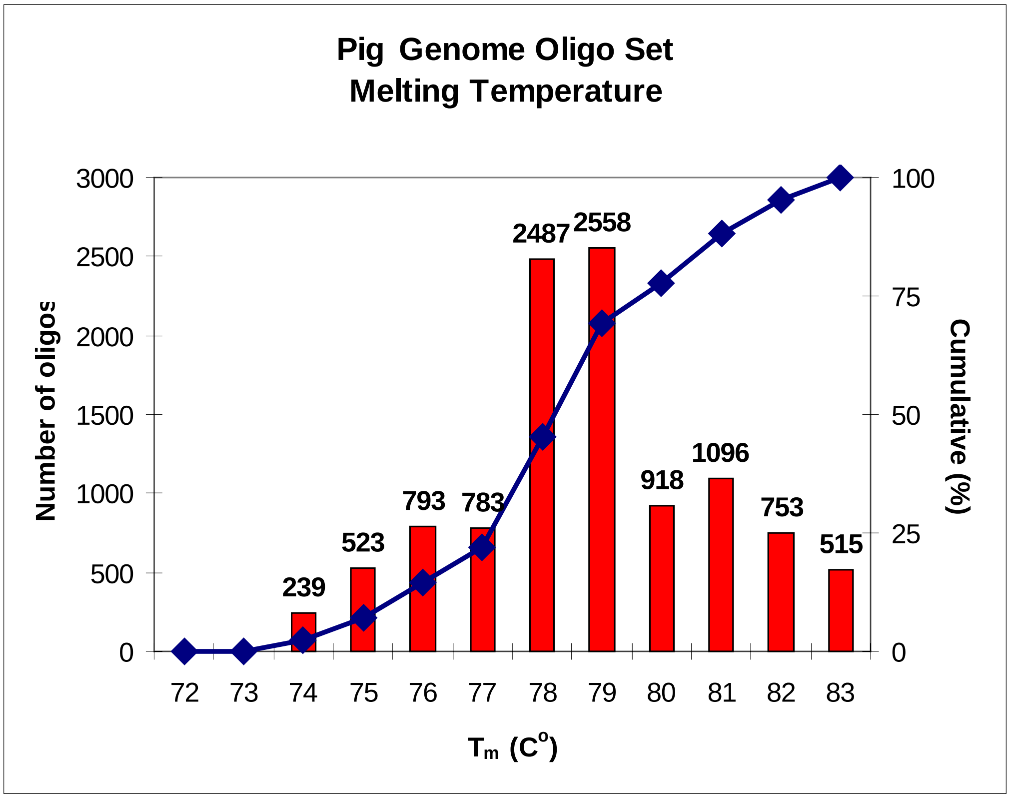


Figure 2. GC Content — Pig Genome Oligo Set

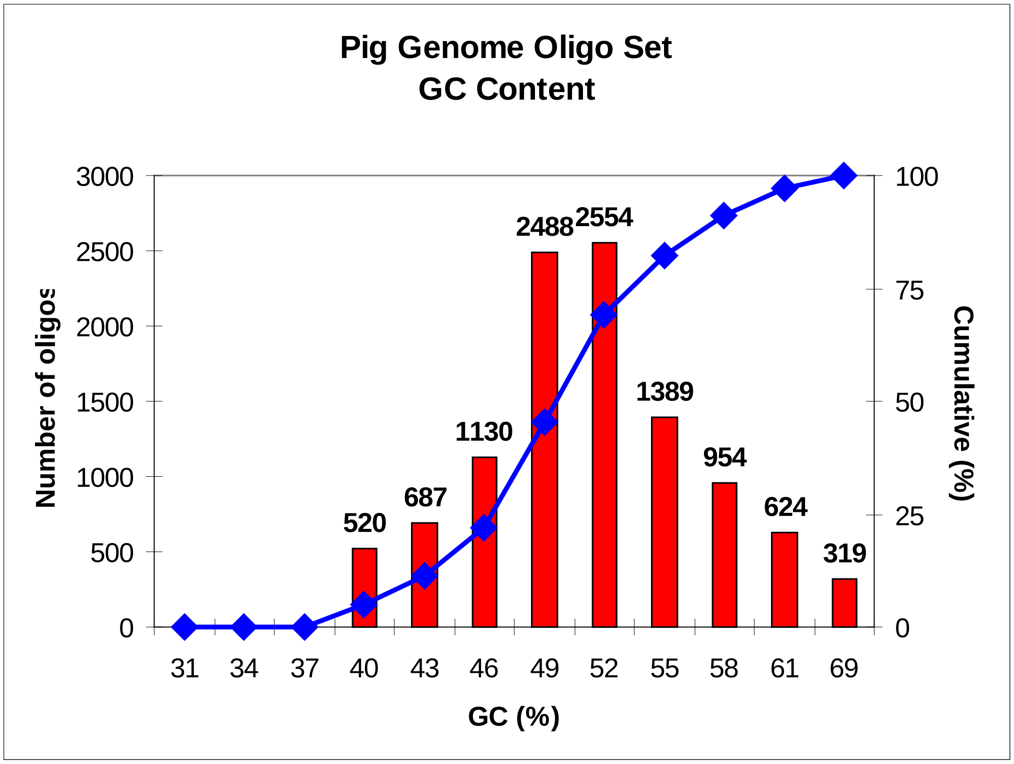


Figure 3. Location from 3' End — Pig Genome Oligo Set

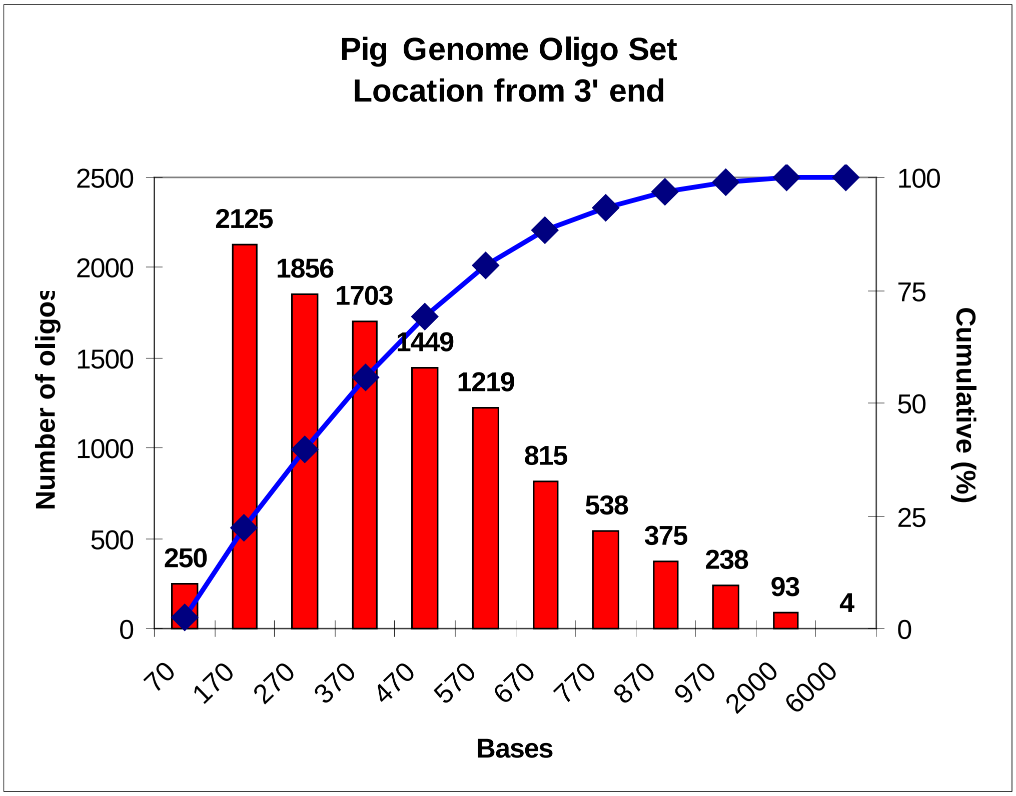


Figure 4. Longest Hairpin Stem Length — Pig Genome Oligo Set

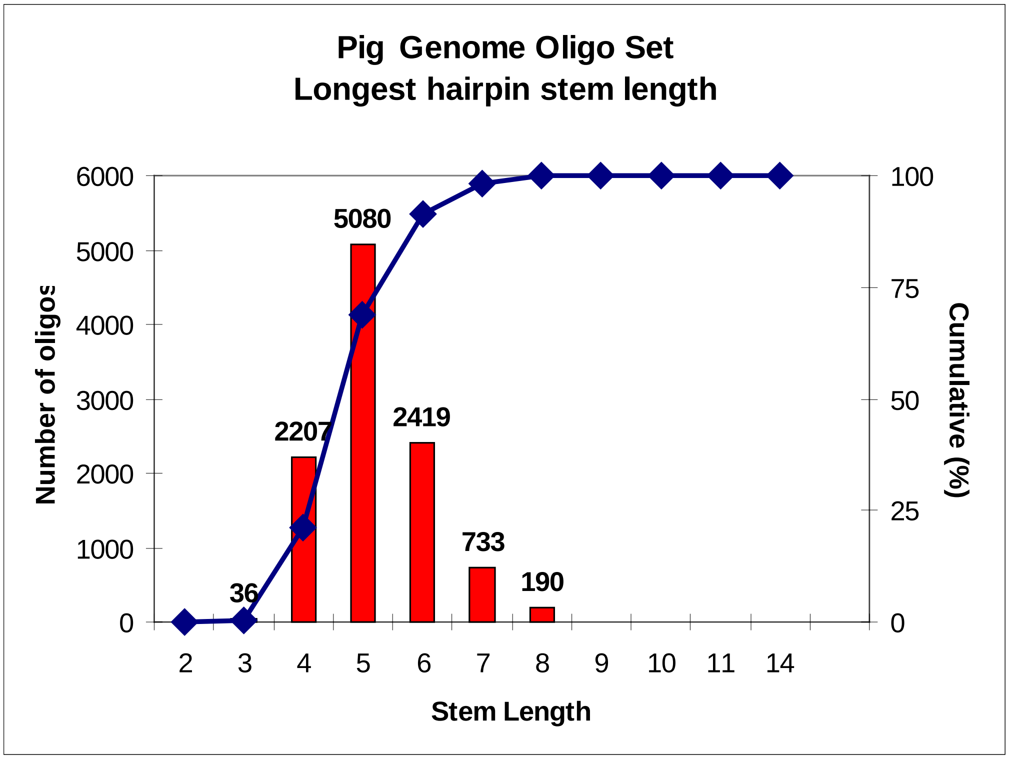


Figure 5. Cross-Hybridization Identity — Pig Genome Oligo Set

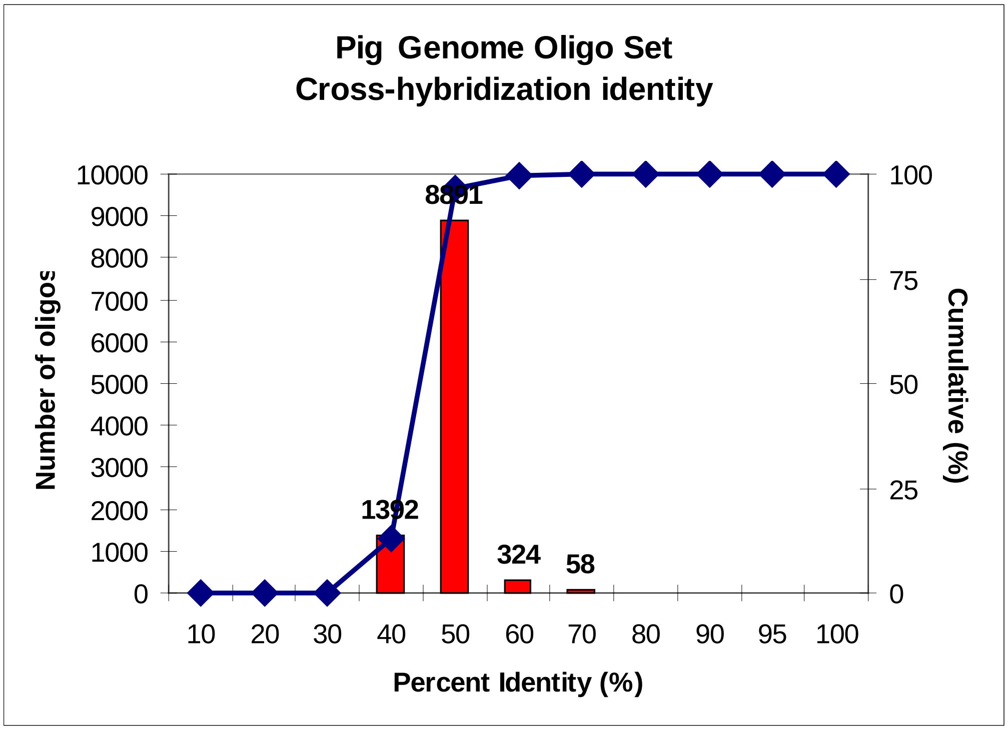


Figure 6. Melting Temperature — Pig Genome Oligo Extension Set

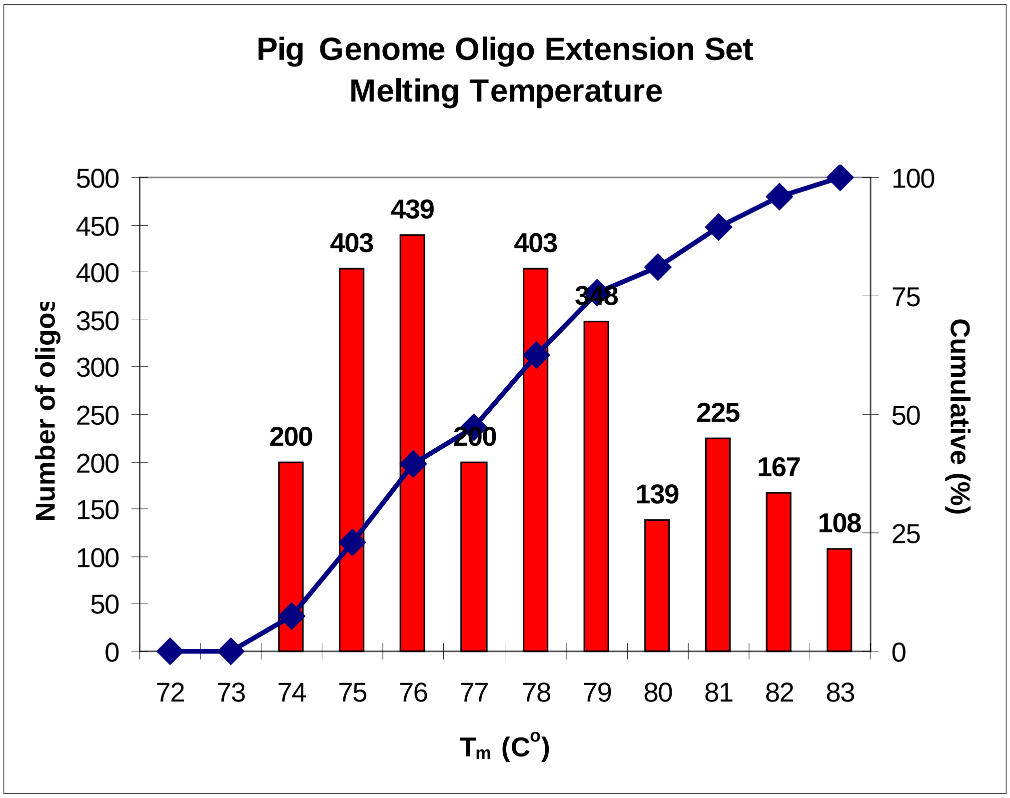


Figure 7. GC Content — Pig Genome Oligo Extension Set

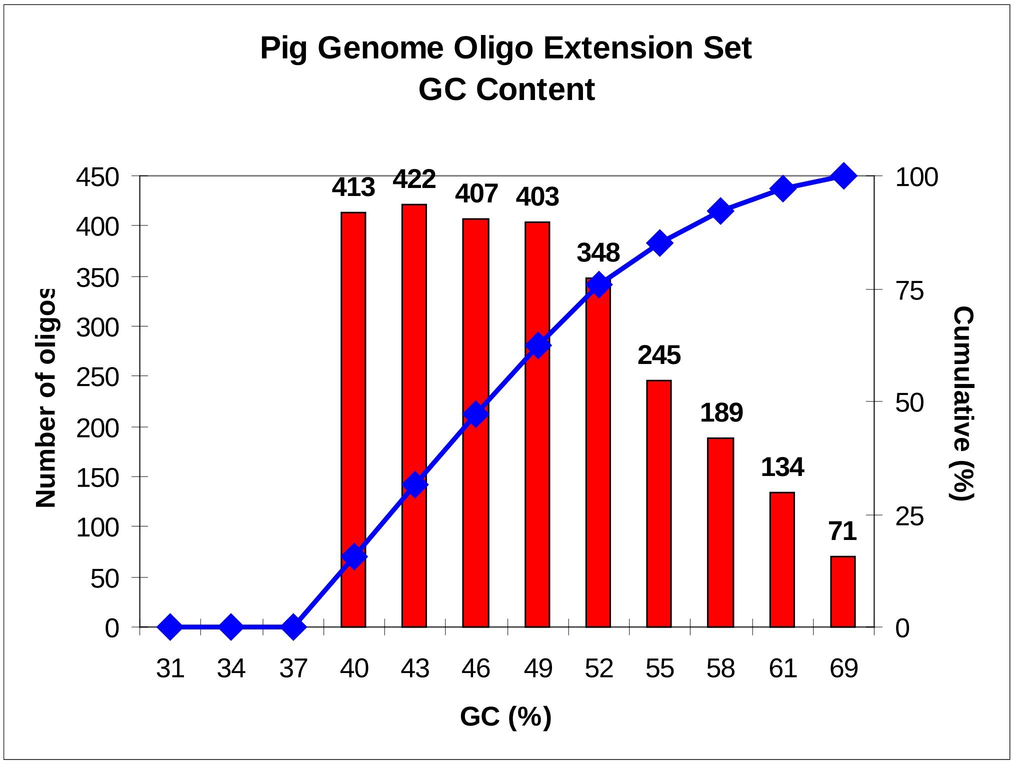


Figure 8. Location from 3' End — Pig Genome Oligo Extension Set

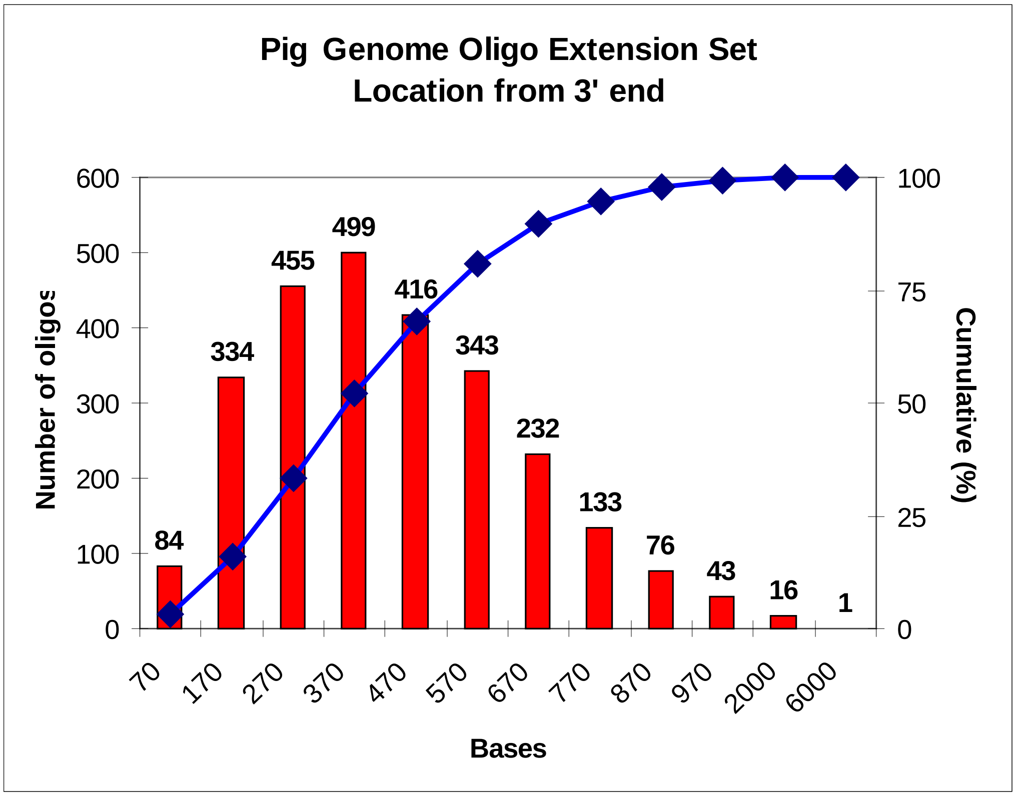


Figure 9. Longest Hairpin Stem Length — Pig Genome Oligo Extension Set

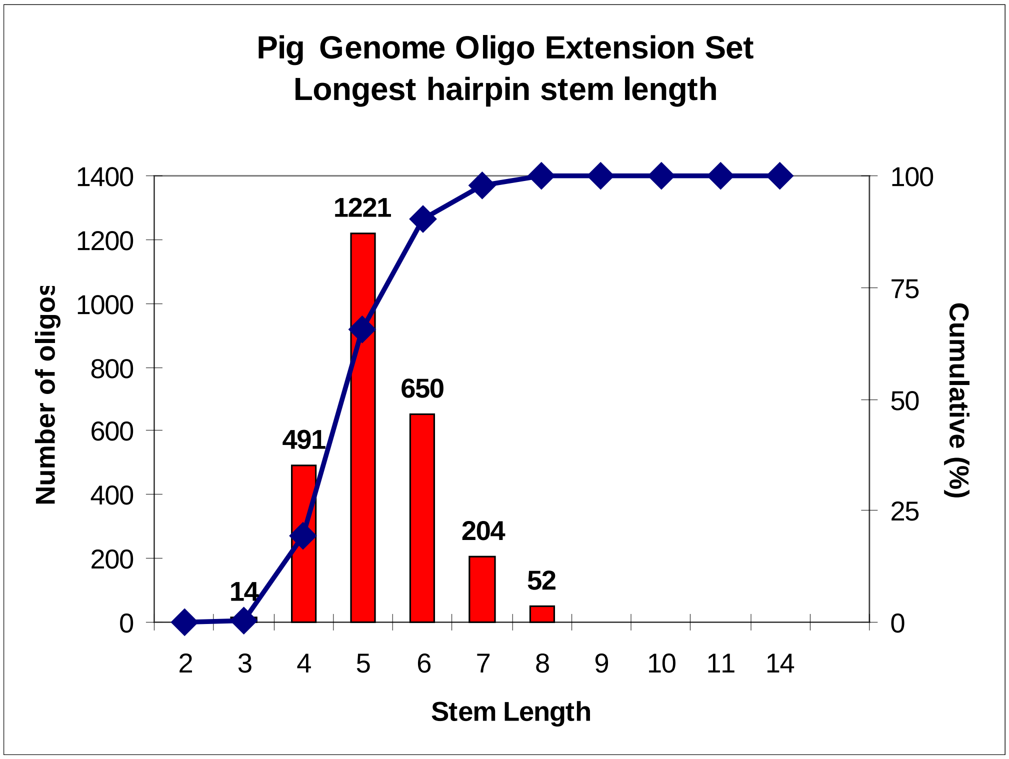
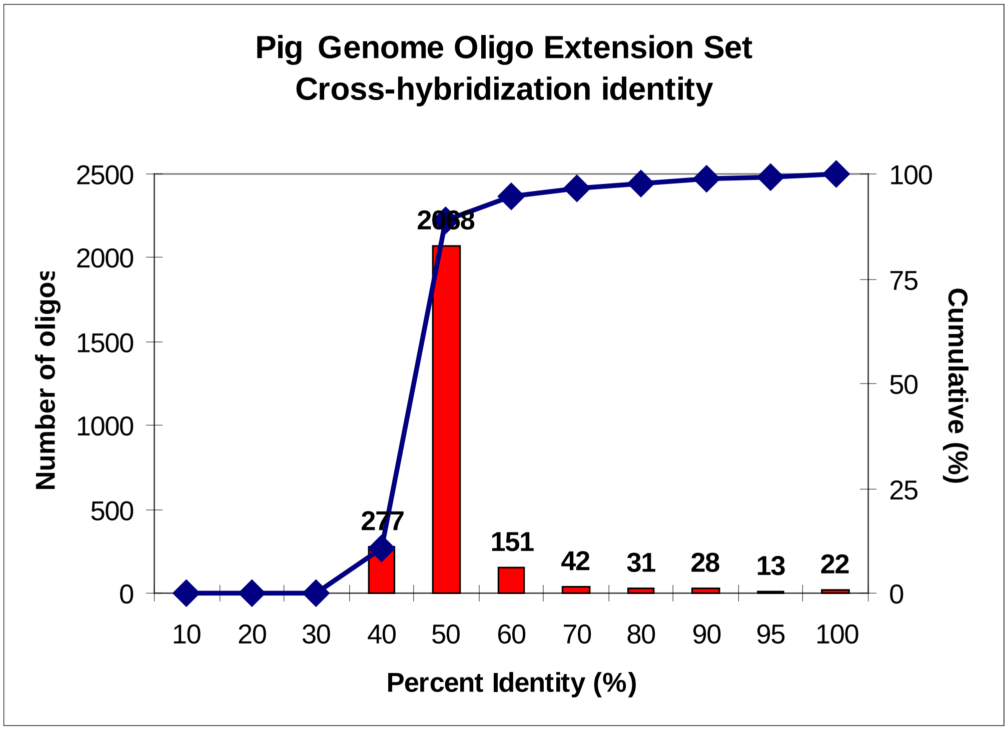


Figure 10. Cross-Hybridization Identity — Pig Genome Oligo Extension Set



## Controls

The last plate in the Pig Genome Oligo Set Version 1.0 represents the control plate. It contains twelve unique positive control oligos selected from other plates and represents twelve unique housekeeping genes. Twelve unique negative controls are also included; seven of these are randomly generated negative controls that have a minimal cross-hybridization identity to all the pig gene sequences in SsGI 5.0. Five of these negative controls are designed from Arabidopsis genes and the oligos sequences have been obtained from the study by Wang and et. al [1]. In this study, these five oligos are used as printing control 70mers where Arabidopsis cRNA is spiked in the labeling reaction at different relative copy numbers. These five controls can be used both as negative controls as they share a minimal cross-hybridization identity with the pig gene sequences in SsGI 5.0 or can be used as printing controls in exactly the same way as in the referenced study.

## 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 Pig Genome Oligo Set Version 1.0 | Number of oligos Pig Genome Oligo Extension Set Version 1.0 |
|---|---|---|---|---|
| Melting temperature (C°) | 78°C±5°C | 73.6–82.9 | 10,665 | 2632 |
| Cross-hybridization identity (%) | < =70 | 31–70 | 10,665 | 2538 |
| Cross-hybridization identity (%) | >70 | 71–100 | 0 | 94 |

## Acknowledgements

We wish to thank the following for their valuable feedback and input: Dr. Max Rothschild, USDA Pig Genome Coordinator, Dr. Chris Tuggle, and Dr. Zhiliang Hu, all from Iowa State University, and Dr. Jon Beever, University of Illinois.

## Reference

1. Wang, H., Malek, R., Kwitek, A.E., Greene, A.S., Truong, L.V., Babak, B., Frank, B., Quackenbush, J., Lee, N. (2003) *Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays*. *Genome Biology* ***4*:R5**, *1–13*.