

### *Bordetella pertussis* Array-Ready Oligo Set™ (Version 1.0)

The Operon *Bordetella pertussis* AROS Version 1.0 contains 3582 70-mer oligonucleotide probes representing 3800 open reading frames (ORFs) of *B. pertussis* strain Tohama I genome. Each probe contains an amino linker at its 5' end.

#### Sequence source and ORF selection

3436 ORF sequences for *B. pertussis* were obtained from the Refseq entry NC\_002929.1 and additional ORF sequences were obtained from Sanger Institute GeneDB database (<http://www.genedb.org/>). The Refseq NC\_002929.1 can be accessed through the National Center for Biotechnology website (<http://www.ncbi.nlm.nih.gov/genomes/frames.cgi?db=Genome&gi=315>). The combined dataset has 3806 ORF sequences. There are six ORFs of this combined dataset that have short or highly GC rich sequence, and can not yield suitable array oligos.

#### Probe design and selection rules

The 70-mer probes are selected with an optimal set of parameters as described below:

1. The melting temperatures ( $T_m$ ) of the probes are restricted within the range of  $80 \pm 5$  °C.  $T_m$  is calculated using the following formula:  
$$T_m = 81.5 + 16.6 \times \log[\text{Na}^+] + 41 \times (\#G + \#C) / \text{length} - 500 / \text{length}$$
 where  $[\text{Na}^+] = 0.1$  M and  $\text{length} = \#A + \#C + \#G + \#T$
2. The contiguous single nucleotide repeat or poly (N) tract within a probe is limited to 7 bases or shorter.
3. The hairpin stem length of a probe is controlled at 8 bases or shorter.
4. The cross hybridization score for a probe against other non-representing (non-self) ORFs or genes in the genome is set 70% or less of BLAST percent identity score.
5. The contiguous base match to other non-self ORFs or genes is constrained at 20 bases or less.
6. The selection distance of the candidate probes is 40 bases away from the 3' end of the ORF or gene. This design restriction is intended for the 3'-end priming of ORFs or genes.

The probes with the highest specificity (or the least cross-hybridization scores) are selected from a pool of candidates satisfying all the rules as described above.

The exceptions (relaxation of one or more selection rules) are made for the probe candidates of 213 ORFs (5.9%), which don't meet the rules above. 18 probes of them are matched to multiple ORFs or genes at the identity score of 100% due to the highly conserved sequences of ORFs or genes in *B. pertussis* genome.

SUMMARY

| Selection rules                                | Threshold | Probe Number |
|--|-----------|--------------|
| Probe length (bases)                           | 70        | 3369         |
| Melting temperature (°C)                       | 80 ± 5    |              |
| Poly (N) tract length (bases)                  | < 8       |              |
| Hairpin stem length (bases)                    | > 110     |              |
| Distal distance from 3' end                    | < 9       |              |
| Cross-hybridization score (identity %)         | ≤ 70      |              |
| Contiguous base match to non-self ORFs (bases) | ≤ 20      |              |
| Exceptions                                     |           | 213          |
| Total  |           | 3582         |

The following illustrations show the distribution of all 3582 probes for melting temperature, GC content, distance from 3' end, hairpin stem length, and cross-hybridization identity.

Figure 1. Melting temperature

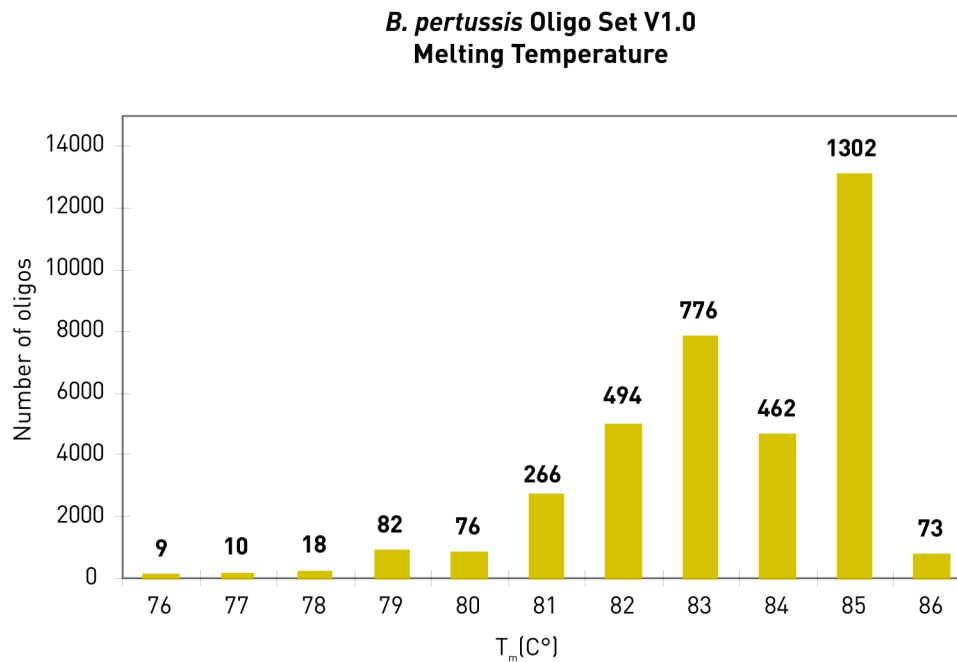


Figure 2. GC Content

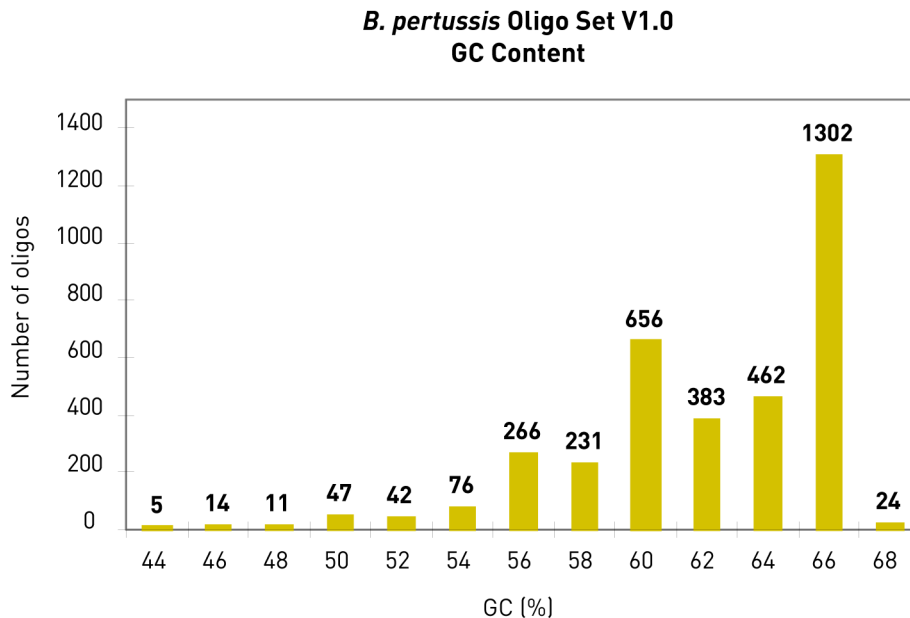


Figure 3. Distance from 3' End

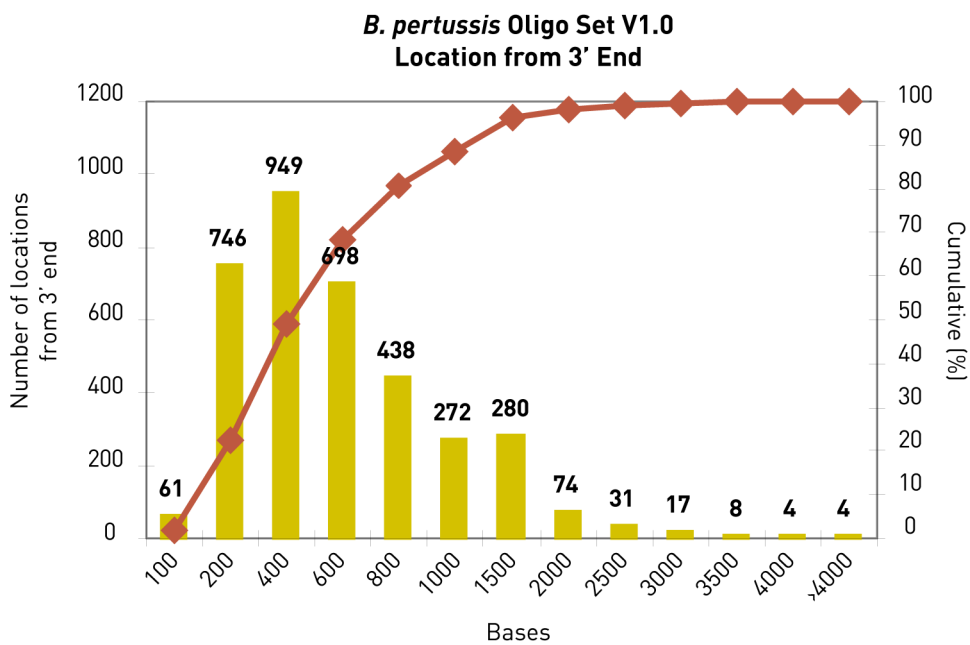


Figure 4. Hairpin Stem Length

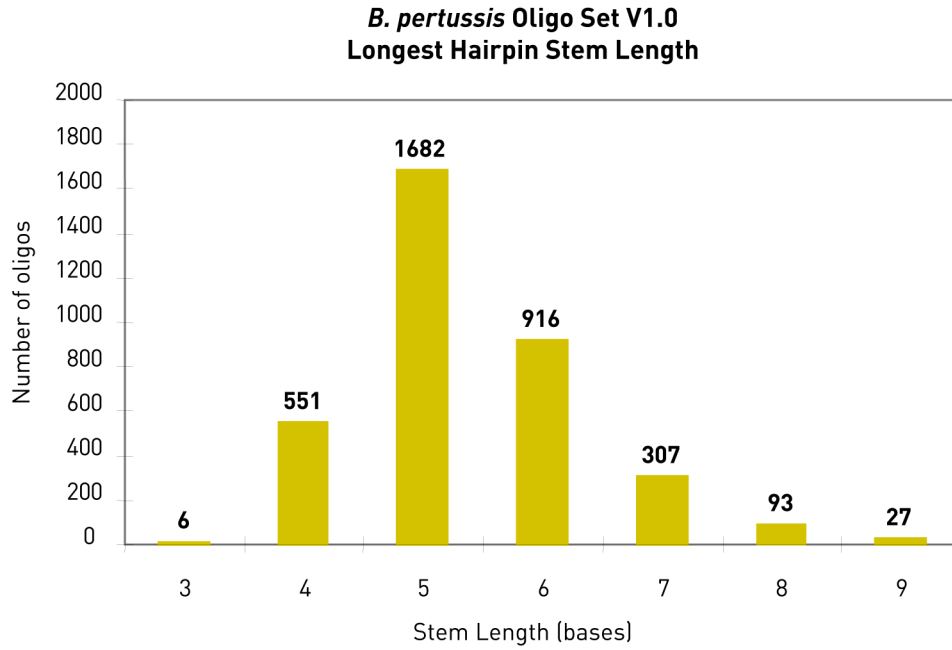


Figure 5. Cross-Hybridization Score

