Operon *Lactobacillus sakei* Genome Array Ready Oligo Set™ (Version 1.0)

Lactobacillus sakei Array Ready Oligo Set™ (AROS) Version 1.0 was designed based on the genome annotations of *L. sakei* strain 23K from Flore Lactique et Environnement Carne (FLEC) Laboratory (Institut National de la Recherche Agronomique, JOUY-EN-JOSAS, France). The design was performed using Operon proprietary oligo design platform under the guidance of the genome experts of FLEC Laboratory. *Lactobacillus sakei* AROS Version 1.0 contains 2,000 oligonucleotide probes. It consists of eight subsets of probes: 1) main subset - the single sense-strand probes originated from *L. sakei* strain 23K genome; 2) companion subset – the probes derived from the sources other than the source used for the main subset (the details described in the section I); 3) dual sense-strand controls – the probes placed in the 5'-half and 3'-half regions of the selected genes; 4) dual sense-strand positive controls – the probes located within the 5'-half and 3'-half regions of the same gene; 5) paired sense and antisense controls – the probes positioned on both sense and antisense strands of the same genes; 6) Stratagene alien spike controls – the probes complement to the ten exogenous alien mRNA spikes; 7) Hybridization stringency controls – the probes with varied sequence homology (antisense, 50 - 90% identity) to Stratagene alien spike controls and; 8) production tracking controls – the probe placed randomly within the AROS plates for quality assurance.

I. The sources of gene sequences and the naming conventions

- The genome annotations (with LSAxxxx names) of *L. sakei* str. 23K are from FLEC Laboratory, and available in GenBank and EMBL (accession: CR936503).
- The non-coding sequences with GHOxxxx names of *L. sakei* str. 23K are unpublished data from the accession CR936503.
- The gene sequences with GSUxxxx names are originated from GenBank accession AY697434 of *L. sakei* strain Kg15.
- The gene sequences with LCSxxxx names are originated from GenBank accession Z54312 of *L. sakei*.
- The gene sequences with PRVxxxx names are originated from GenBank accession AF438419 of *L. sakei* plasmid pRV500.
- The gene sequences with SKAxxxx are originated from Genbank accession Z46867 of *L. sakei* strain Lb706.
- The gene sequences with SKGxxxx names are originated from GenBank accession AF395533 of *L. sakei*.
- The gene sequences with SKPxxxx names are originated from GenBank accession AF002276 of *L. sakei* strain LTH673.
- The gene sequences with SKQxxxx names are originated from GenBank accession AJ844595 of *L. sakei* strain LTH673.
- The gene sequences with SKXxxxx names are originated from GenBank accession AY206863 of *L. sakei*.
- The gene sequences with TETxxxx names are originated from GenBank accession AY149596 of *L. sakei* subsp. sakei isolate DG525.
- The gene sequences with TNLxxxx names are unpublished work of Monique Zagorec and Stéphane Chaillou (Institut National de la Recherche Agronomique, FLEC Laboratory, JOUY-EN-JOSAS, France). The sequences are originated from *L. sakei* strain T332.

II. The design criteria and selection rules:

The dataset for oligo design was from a combination of sequence sources described above. The oligo candidate generation and the optimal oligo selection were conducted using Operon proprietary platform with the following criteria and selection rules.

1) The lengths of oligo sequences are set at 70 bases for the generation of oligo candidates.

2) The melting temperatures (T_m) are determined within 73 ± 5°C for oligo screening. The T_m is calculated using the formula: T_m = 81.5 + 16.6 * log[Na+] + 41 * (#G + #C)/length – 500/length, where [Na^+] = 0.1 M and length = #A + #C + #G + #T.

3) The distances of oligos to the 3'-ends of the transcripts are limited within 2,000 bases or less with the exception on the oligos for the 5'-half regions of the selected genes.

4) The oligo candidates are screened against the low complexity with two filtering criteria. (a) The contiguous single nucleotide base repeat or poly (N) tract is restricted to 8 bases or less. (b) The oligo candidates are selected against a pre-set normalized simple repeat score.

5) The oligos candidates are checked against the potential secondary structures. (a) Hairpin structures --- the hairpin stem length is restricted to 9 bases or less. (b) Self-dimers --- the self annealing scores of oligo candidates are kept at 120 or less. The self-annealing score is calculated as the alignment score of the optimum local alignment between the oligo sequence and its reverse compliment using the Smith-Waterman algorithm.

6) The oligo specificity is screened through the computation of the cross-hybridization identity score on the top non-self transcripts by blasting against a combined multiple-genome database (see rule #7 below). The non-self transcripts are all other transcripts not represented by a specific oligo. (a) The oligos candidates are selected for the cross-hybridization identity score of 70% or less. (b) The selected oligos must not have more than 20 contiguous bases matched to any non-self transcripts.

7) In the specificity screening, the oligo candidates are checked not only against the database of *L. sakei* str. 23K genome annotations, but also against the genome annotations of four *Escherichia. coli* strains (CFT073, K12, O157H7, O157H7_EDL933).

8) Certain exceptions are made for a few oligos during the selection (see figures in Section III). For a few transcripts that don't yield the oligos satisfying the cutoff scores described above, the selection rules are gradually relaxed in the next round of selection process.

III. The specifications and characteristics of *Lactobacillus sakei* AROS Version 1.0.

The descriptions on the eight subsets of *L. sakei* AROS Version 1.0 are detailed as the below.

- The main subset is comprised of 1,850 single sense-strand probes generated from the genes of *L. sakei* str. 23K.

- The companion subset contains 82 probes from a variety of sources (see the details in the section I). They're intended for the extra gene coverage of *L. sakei*.

- The dual sense-strand controls (44) were designed from the 22 genes listed below. 22 probes are at the 5'-half regions of the genes, and 22 at the 3'-half regions. Their layout in 384-well plates is arranged as such manner. Each 384-well plate has four 5'-half region probes and four 3'-half region probes from the different genes. The controls are arranged in the wells immediately after the first two empty wells (A1, A2).

  LSA0172, LSA0212, LSA332, LSA362, LSA534, LSA624, LSA708, LSA0912, LSA0955, LSA1034, LSA1141, LSA1261, LSA1335, LSA1406, LSA1568, LSA1704, LSA1730, LSA1731, LSA1800, GSU0001, 16S and 23S ribosomal RNA genes

- The dual sense-strand positive controls (2) were selected from the gene (LSA1606) with one at the 5'-half region and another at the 3'-half region of the gene.

- The paired sense and antisense controls (22) were genenrated from the eleven gray-hole sequences (ids listed below), which are the intergenic sequences of *L.sakei* str. 23K genomes. 11 probes are on the sense strand, 11 on the antisense strand.

  GHO0001, GHO0002, GHO0003, GHO0004, GHO0005, GHO0006, GHO0007, GHO0008, GHO0009, GHO0010 and GHO0011

- Stratagene alien spike controls (10) are licensed directly from Stratagene (www. Stratagene.com). In coupling with the alien mRNA spikes in Stratagene SpotReport® Alien® Oligo Array Validation System, they are intended as the internal controls for the normalization and standardization of dye incorporation, microarray hybridization and data analyses.

- Hybridization stringency controls (60) were produced based on the Stratagene alien spike controls with varied sequence homology of 50%, 60%, 70%, 80%, 90% and anti-sense. They're intended as the markers to assess the microarray hybridization stringency.

- Production tracking oligo (1) is a randomly-generated oligo sequence with a length of 30 bases. It has been selected against the cross-hybridizations with the gene sequences used in this AROS design. The tracking oligo is randomly positioned in the 384-well and 96-well plates so that each 384-well plates has 4, and each 96-well plate has one for production quality assurance in the process of oligo synthesis and microarray printing.

The following figures illustrate the characteristics of *Lactobacillus sakei* AROS Version 1.0.

Figure 1.

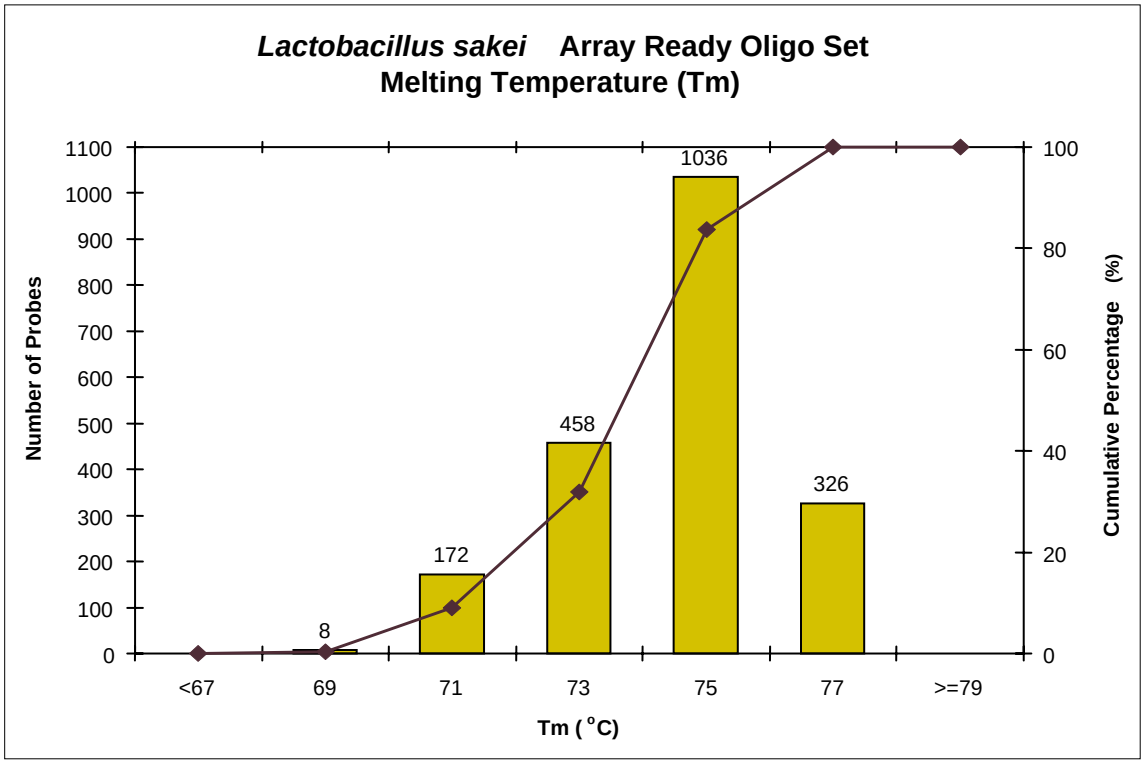


Figure 2.

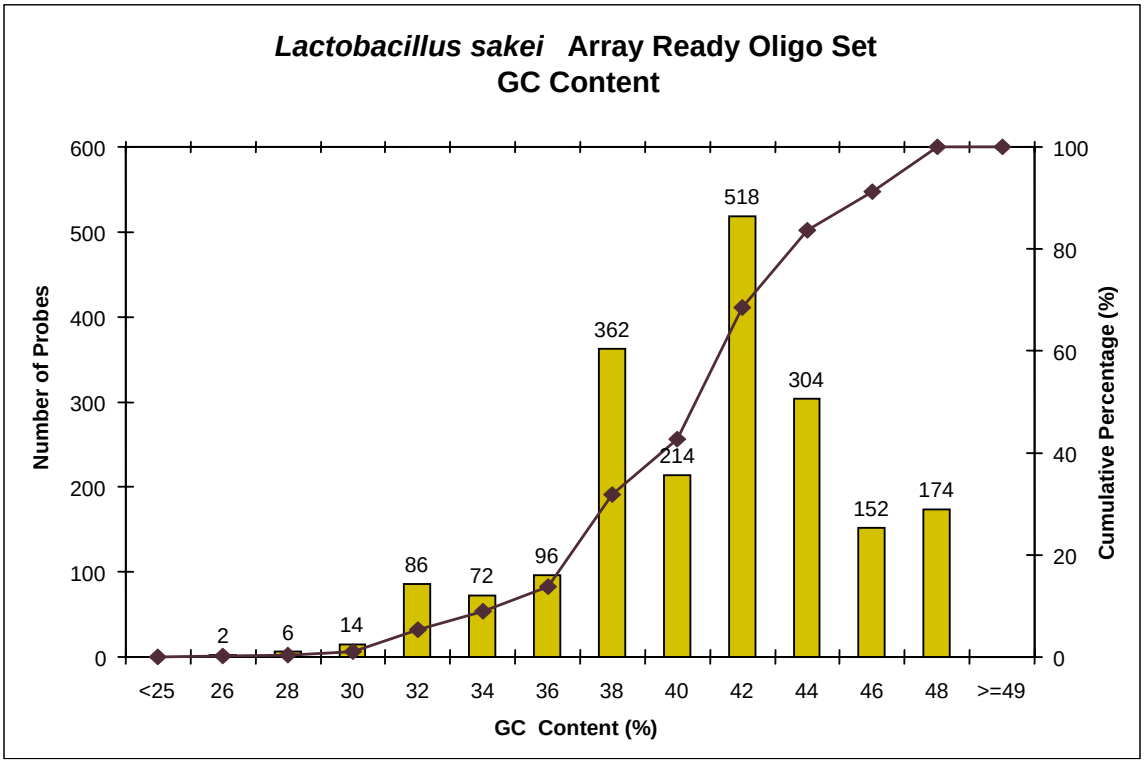


Figure 3.

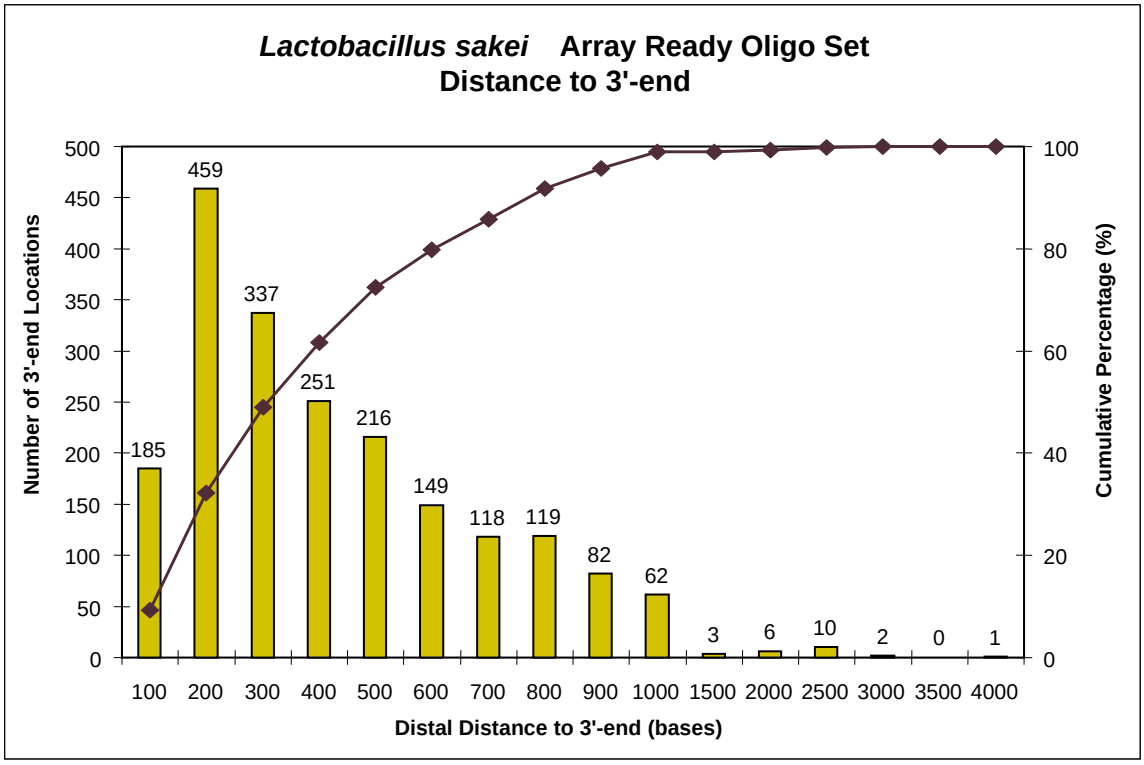


Figure 4.

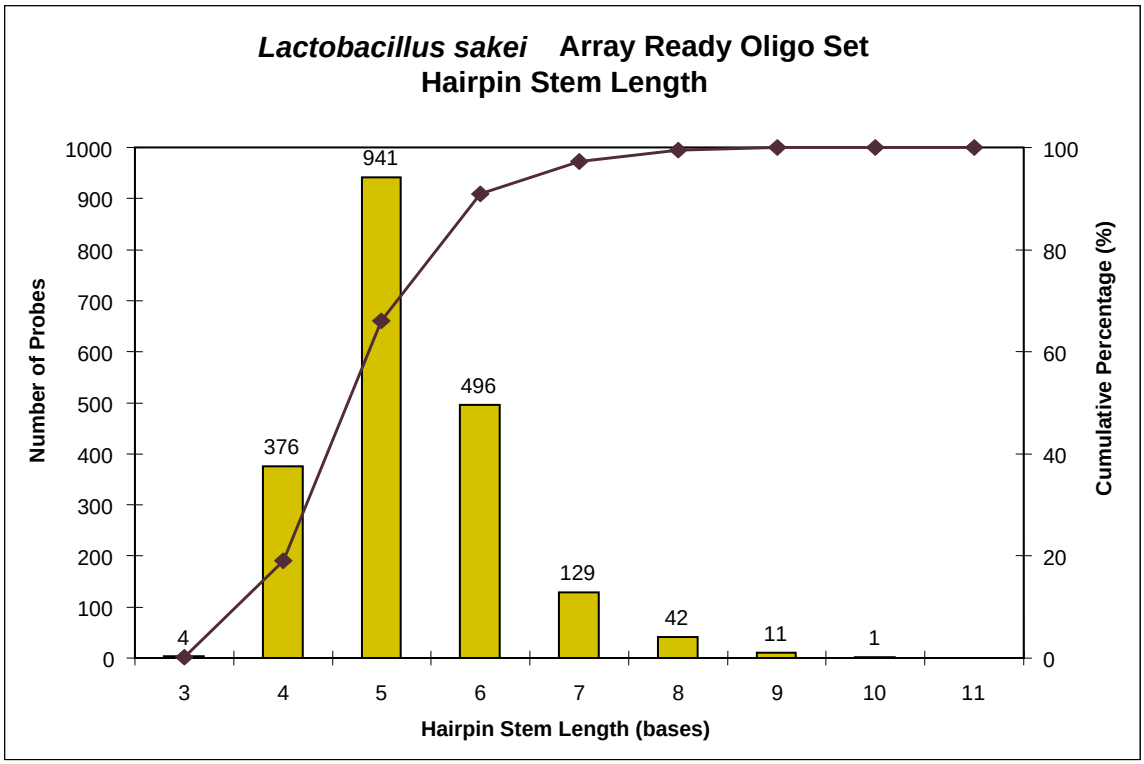
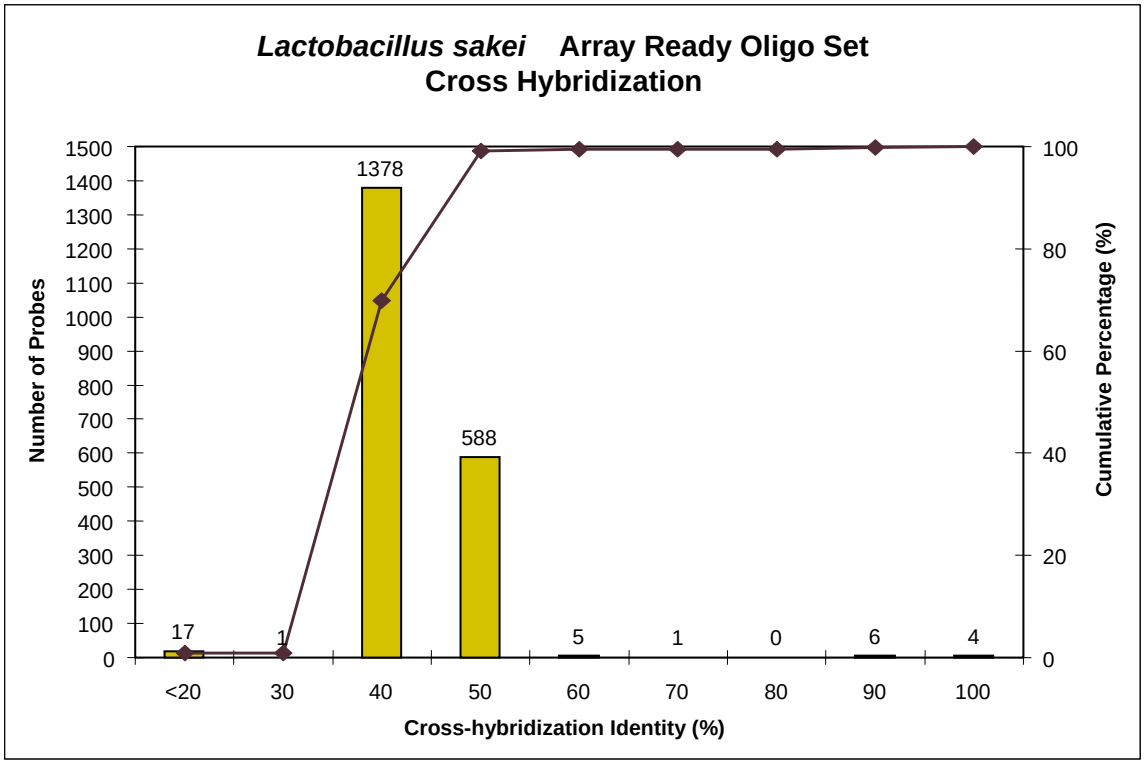


Figure 5.


Lactobacillus sakei Array Ready Oligo Set
Cross Hybridization
Number of Probes
Cumulative Percentage (%)
Cross-hybridization Identity (%)
<20
30
40
50
60
70
80
90
100
17
1
1378
588
5
1
0
6
4