

Operon *Candida albicans* Genome AROS™ Upgrade Set
Version 1.1

Operon is pleased to announce the availability of an upgrade set to supplement the *Candida albicans* Genome AROS™ (version 1.1) product line for microarray manufacture. The upgrade set was designed based on a consolidated data-set of both manually-curated and automatically-predicted gene sequences from *C. albicans* Genome Assembly 19. The upgrade set is intended to cover the newly-defined gene space, which hasn't been mapped by the probes of the *C. albicans* Genome AROS (version 1.1). The same design principles and selection criteria were implemented for the upgrade set as done for the *C. albicans* AROS™ (version 1.1). The upgrade set consists of 1,659 probes.

Release date: February, 2006

I. The sequence sources The human-curated gene sequences of *C. albicans* genome were published by Braun and coworkers [Braun et. al., 2005], and are accessible at the website of the *C. albicans* Annotation Work Group (<http://candida.bri.nrc.ca/candida/index.cfm>). The *C. albicans* genome assemblies, sequence data and annotations were obtained from the *C. albicans* Genome Database (<http://www.candidagenome.org>). The gene sequences and annotations of *C. albicans* mitochondrial genome were acquired from the Genbank accession NC_002653 (GI:12585565), which is accessible at the website of National Center of Biotechnology Information (<http://www.ncbi.nih.gov>).

II. The design and selection criteria: Once an ORF or gene of the *C. albicans* genome or its mitochondrial genome was mapped out for the probe design, a pool of 70-mer candidates were generated for each sequence, and screened with an optimal set of parameters as illustrated in Table 1.

Table 1: Probe selection criteria

Criteria	Value
Probe length (bases)	70
Melting temperature (T_m °C)	73 ± 5
Maximal distance from the 3' end (bases)	≤ 1000
Poly(N) tract length (bases)	≤ 8
Hairpin stem length (bases)	≤ 9
Simple repeats	Pre-set
Cross-hybridization to all other sequences (identity %)	$\leq 70\%$
Contiguous base match to other sequence (bases)	≤ 20

Notes and explanations:

- The melting temperature was 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 $[\text{Na}^+] = 0.1 \text{ M}$ and $\text{length} = \#A + \#C + \#G + \#T$
- Poly (N) tract is defined as the contiguous single nucleotide repeat within the sequence.
- A normalized score for simple repeat was assigned to each oligo based on the number of repeats. Oligos with more repeats having a normalized score greater than a certain threshold were filtered out.
- The cross hybridization score was calculated by the NCBI BLAST program in the following way. A specific oligo

was blasted against the sequence database for all detectable alignments. The highest scoring non-self sequence, defined as the sequence that yields the most matched bases with the oligo except itself, was computed for a BLAST score (identity %). This BLAST score (identity %) was referred to as the cross-hybridization score. The calculated BLAST score (identity %) is dependent on the size of the sequence database, oligo sequence, and the use of gapped or no-gap alignment option.

III. The characteristics of *C. albicans* AROS upgrade set (version 1.1)

The following figures illustrate characteristics of T_m , GC content, distances from 3' end, hairpin stem length, and cross-hybridization score of 1,659 probes in the upgrade set.

Figure 1. Melting Temperature (T_m)

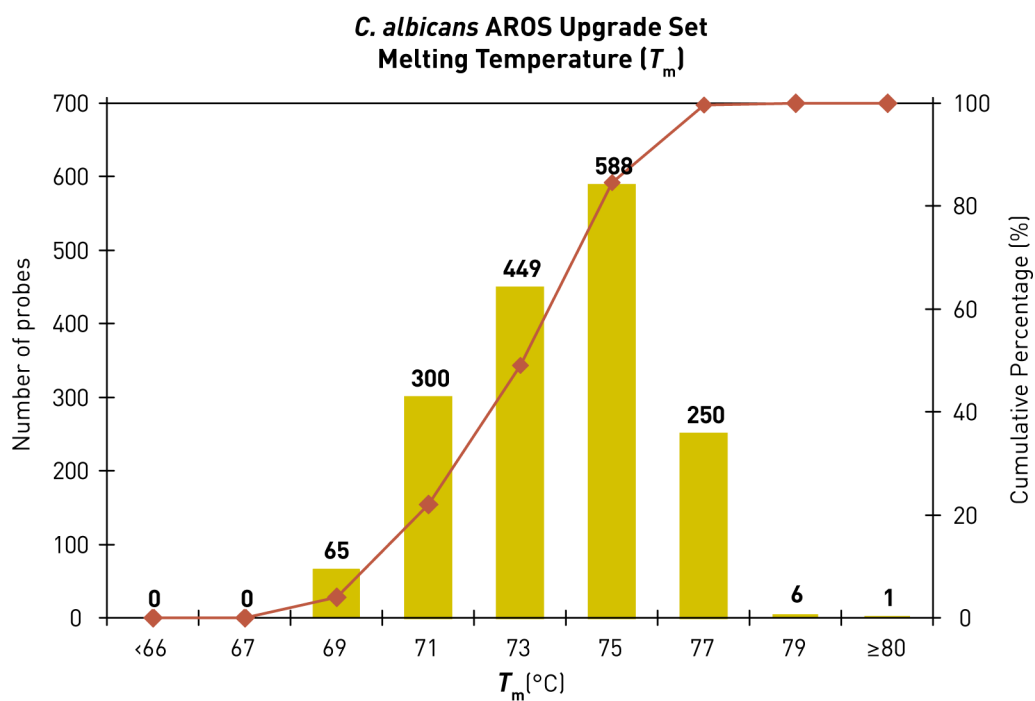


Figure 2. GC Content

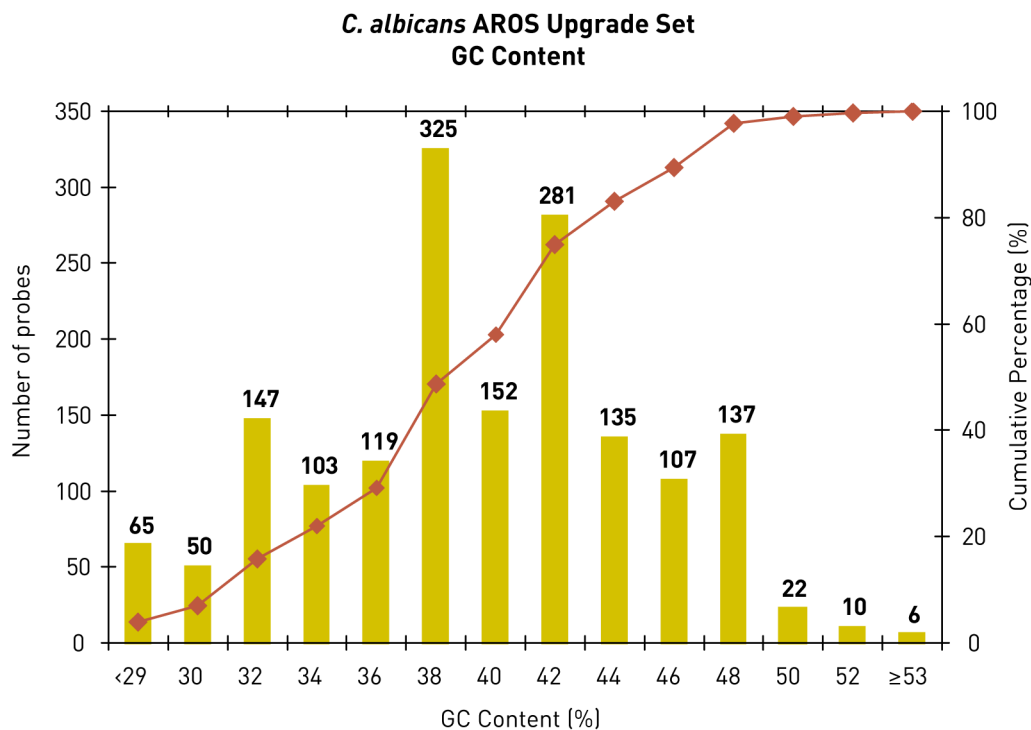


Figure 3. Distance to the 3' End of Transcripts

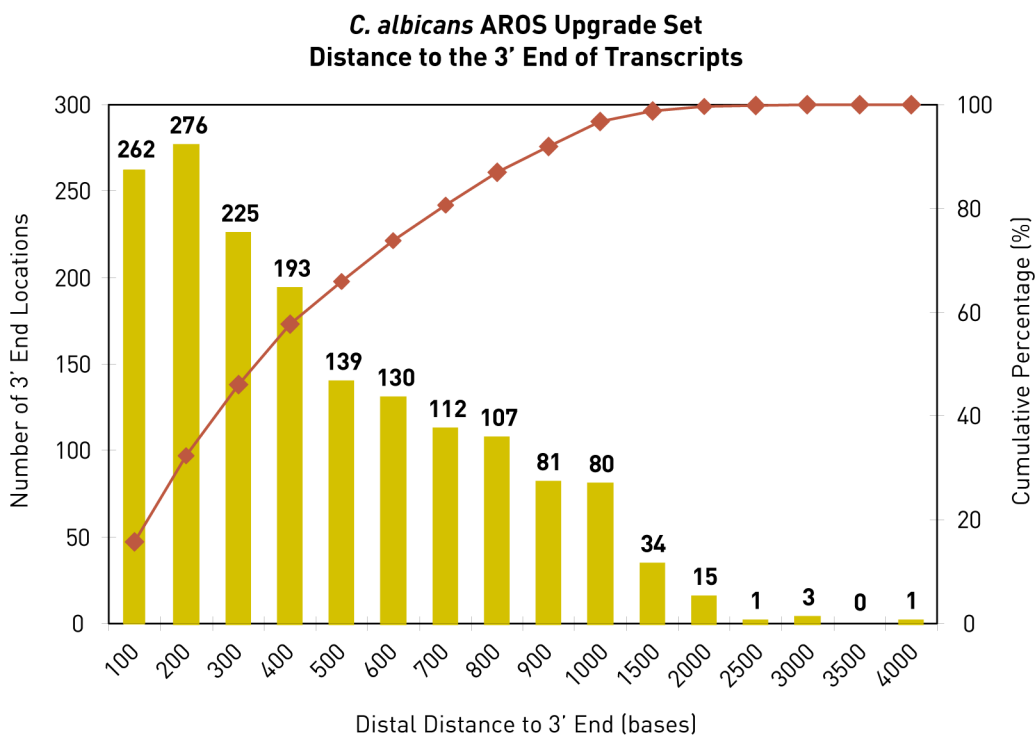


Figure 4. Hairpin Stem Length

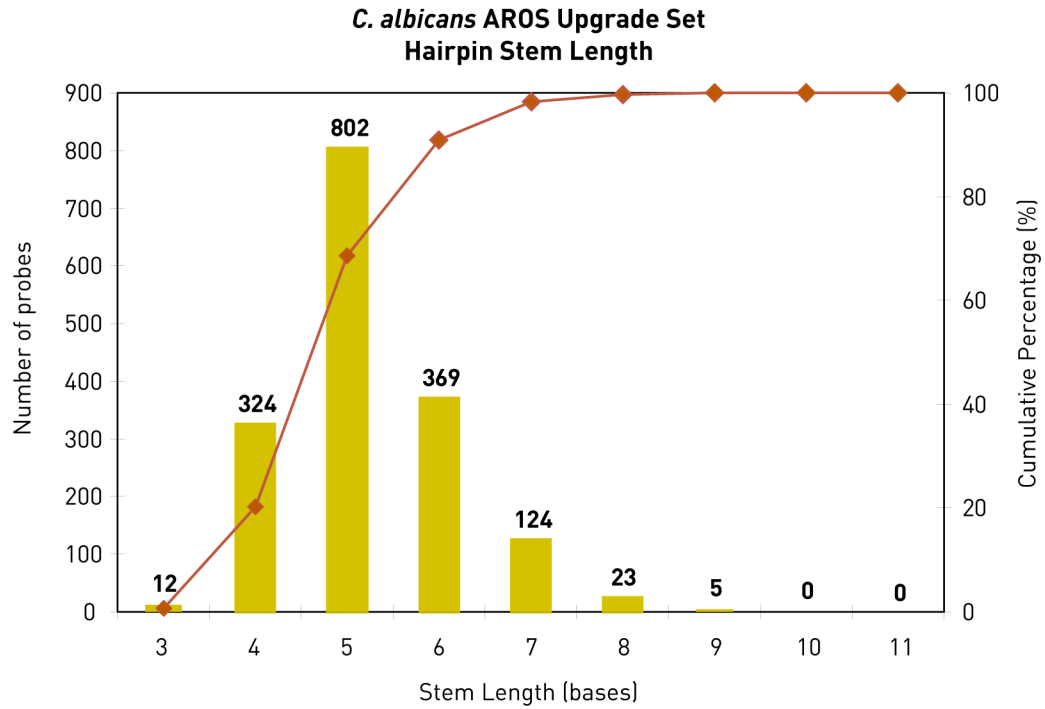
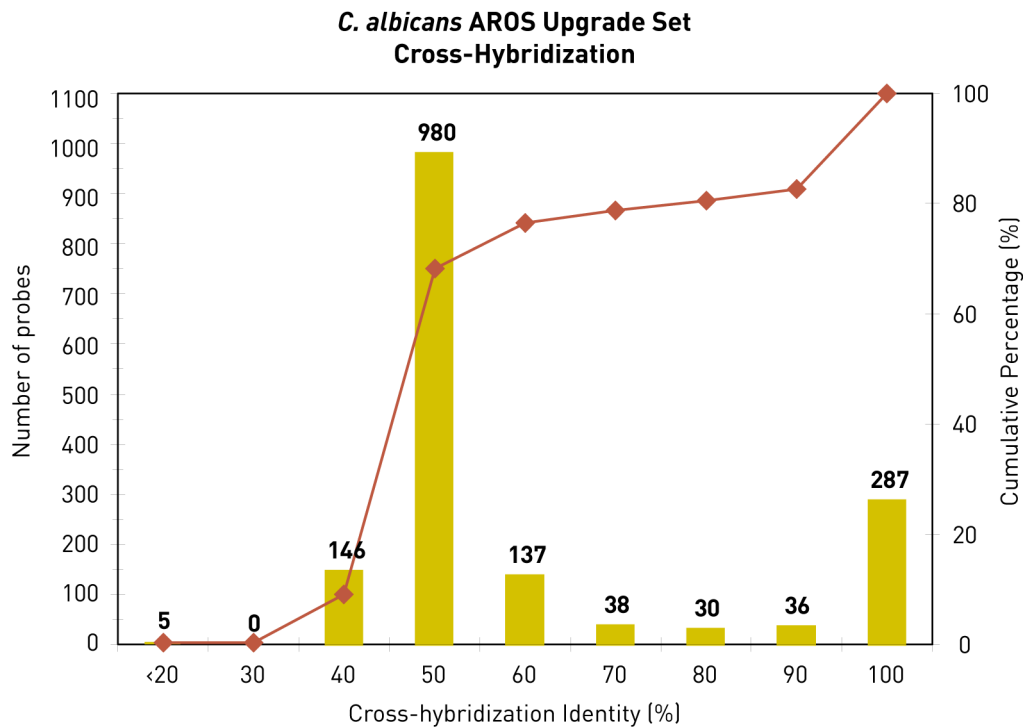


Figure 5. Cross-Hybridization



IV. References: Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, Inglis DO, Uhl MA, Hogues H, Berriman M, Lorenz M, Levitin A, Oberholzer U, Bachewich C, Harcus D, Marcil A, Dignard D, Iouk T, Zito R, Frangeul L, Tekaiia F, Rutherford K, Wang E, Munro CA, Bates S, Gow NA, Hoyer LL, Kohler G, Morschhauser J, Newport G, Znaidi S, Raymond M, Turcotte B, Sherlock G, Costanzo M, Ihmels J, Berman J, Sanglard D, Agabian N, Mitchell AP, Johnson AD, Whiteway M, and Nantel A (2005). A human-curated annotation of the *Candida albicans* genome. *PLoS Genet.* 1:36-57.