

# A Significant Difference in output among Microarray Experiment Front-end Tools\*

Ezekiel F. Adebiyi, Segun A. Fatumo and Victor C. Osamor\*\*

## 1 Introduction and Motivation

It has been noted in [3] that 25-mer oligo arrays (Affymetrix gene Chip) could be hybridization problematic and longer oligo (50-70 mers) arrays might be needed to generate accurate expression profiles. Existing oligos selector algorithms include that of Kaderali and Schliep, Li and Stormo, Rouillard, Herbert and Zuker and Rahamann[4]. Other algorithms include Lockhart et al.[2] and Bozdech et al.[1]. The last two algorithms had been used in microarray experiments[3, 1]. The algorithms of Rahamann and Lockhart et al. are designed to find short oligos, while the algorithms of Bozdech and Rouillard et al. are designed to produce long oligos. Li and Stormo algorithm can select both short and long oligos. The algorithms for finding long oligos are found to be slow. Bozdech et al. program takes 21 hours to design gene specific 70mer oligos for 15.5MB *Neurospora crassa* genes on a 3.0 GHz Intel 4, linux computer, while Rouillard et al. takes 36 hours for the same task on the same platform. In our attempt to develop another algorithm for selection of long oligos, we started with the computational analysis of existing methods. Note that a computational analysis of the methods use to select oligos is important since the correctness of the front-end, of course has great influence on the results we will obtain at the middle and back ends

## 2 Computational analysis of existing methods: Discussion and Conclusion

Since each oligo should be specific for exactly one gene and the cDNA or mRNA from the sample need not contain the perfect Watson-Crick complement of an oligo to hybridize, the oligos produced by any of the existing algorithms for finding oligos (short or long) should be the same or at least closely the same.

In this analysis, we compare the output of the algorithm of Bozdech et al. and Rouillard et al., namely the ArrayOligoSelector (Pick70) and OligoArray2.1. We apply them to design oligos for three organism sequences, namely the 6343 yeast (*Sacharomyces cervisiae*) ORFs (a total of 9.5 MB), 10082 *Neurospora crassa* (NC) genes (of about 15.5 MB) and the 12 MB *Plasmodium falciparum* coding sequences. The program parameters used

for the two algorithms targeted 1 oligo of length 70 per gene that contains 28 percent GC content. Note that in Pick70 output, oligos containing meaningless sequences like “GGGGG”, “CCCCC”, “TTTTT”, and “AAAAA” are been implicitly removed. It is expedient here to state that the algorithm of Rouillard et al. hung on gene PFA0135w of chromosome 1, while designing oligos for the publicly available genome of *Plasmodium falciparum*. We discovered that their algorithm does not behave this way when we used it on yeast and *Neurospora crassa* genes. We do not know the reason for this behavior. On both yeast and *Neurospora crassa*, the two algorithms performed as depicted in the table and figure attached.

The results below actually speak for themselves and lead to the discovery that there is a significant variant among the microarray experiment front-end tools in term of the oligos they produced. That is, for example, the results of two microarray experiments on the same organism genes, whose oligos were designed by two different algorithms will be at variant. At this junction, a nice question to pose is: *what is the threshold (that is, edit distance) that oligos from different programs can have between themselves and still be considered to be the same, in the sense of how close/equal are the expression of the genes at the gene expression profiling experiment.* Finally, looking at the patterns of the distributions in fig 1 attached, it is straight forward to see that the patterns are the same. This implies that the difference in the algorithms is due to their design and not because the algorithms are been applied to different organisms.

## References

1. Bozdech, Z. Zhu, J., Joachimiak, M. P., Cohen, F. E., Pulliam, B., and DeRisi, J. L. Expression profiling of the schizont and trophozoite stages of *Plasmodium falciparum* with a long-oligonucleotide microarray. *Genome Biology*, 4(2), R9, 2003.
2. Lockhart, D. J. et al. *Expression monitoring by hybridization to high-density oligonucleotide arrays.* *Nat. Biotechnol.*, 14, 1675-1680, 1996.
3. Le Roch, K. G., et al. *Discovery of gene function by expression profiling of the malaria parasite life cycle.* *Science*, 301, 1503-1508, 2003.
4. Rahmann, S. *Fast large scale oligonucleotide selection using the longest common factor approach.* *JBCB*, 2002.

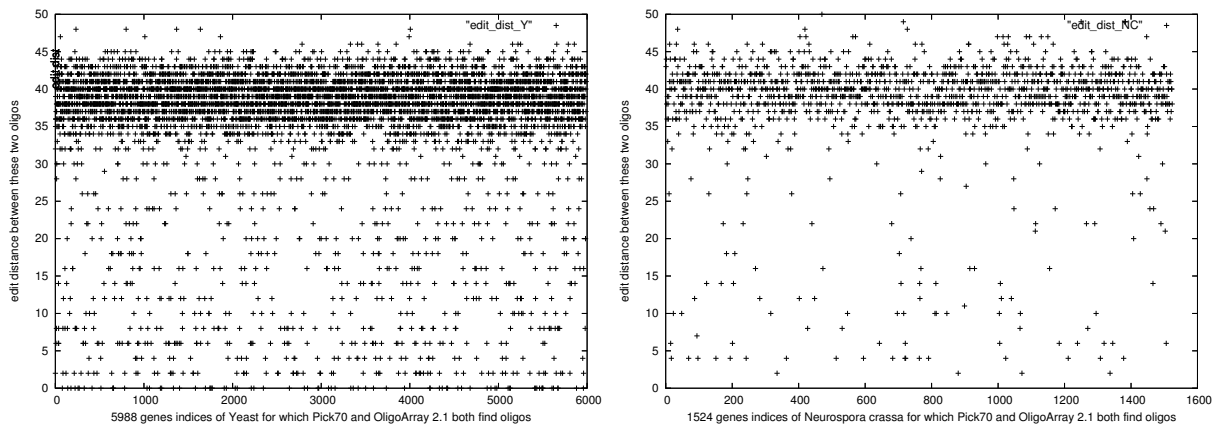
\* We thank Rouillard, J.-M. for sending us the OligoArray program.

\*\* Department of Computer and Information Sciences, Covenant University, PMB 1023, Ota, Nigeria. E-mail: eadebiyi@sds.edu

organism	Sacharomyces cerevisiae	
Program	Pick70	OligoArray2.1
Found	6302	6018
Failed	30	314
	Both Failed	
	11	
	Both Found	
	5988	

organism	Neurospora crassa	
Program	Pick70	OligoArray2.1
Found	9950	1525
Failed	1	8426
	Both Failed	
	131	
	Both Found	
	1524	

**Table 1.** The number of genes for which oligos was found or not found by the two programs under consideration.



**Fig. 1.** The edit distances distributions between the two oligos as emitted by the two algorithms from a gene.