

Appendix for Konovalov *et al.* (2005)

We converted the partition-distance algorithm of Almudevar & Field (1999) to this paper notation

$$C_1(a_i, b_j) = a_i \cap b_j, \quad (1)$$

$$C_2(a_i, b_j) = a_i \oplus b_j, \quad (2)$$

$$D(A, B) = \min(D_1, D_2), \quad (3)$$

$$D_k = D(A \cap \bar{C}_k, B \cap \bar{C}_k) + |C_k|, \quad (4)$$

$$\text{where } C_k = C_k(a_i, b_j), \quad k = 1, 2 \quad (5)$$

and achieved the following more efficient version:

- (1) Remove all clusters common in A and B .
- (2) Select the first remaining non-empty cluster, a_i , in A and a first overlapping with a_i cluster, b_j , in B .
- (3) Perform AND and XOR operations on the clusters using Equations (1) and (2).
- (4) The distance can now be recursively calculated via Equations (3) and (4).

The above algorithm is more efficient then the original because there may be a number of common elements that will be processed at once when $|C_1| > 1$. In the case when $|C_1| = 1$ our formulation is identical to the original.

REFERENCES

- Almudevar, A. and Field, C. (1999) Estimation of single-generation sibling relationships based on DNA markers. *Journal of Agricultural Biological and Environmental Statistics*, **4**, 136-165.
- Konovalov, D.A. *et al.* (2005) Partition Distance via the Assignment problem. *Bioinformatics*, **21**, 2463-2468.