

## THE TRANSFORMATION DISTANCE PROBLEM REVISITED<sup>1</sup>

BEHSHAD BEHZADI and JEAN-MARC STEYAERT

*LIX, Ecole Polytechnique,  
Palaiseau cedex 91128 France*

*e-mail: {behzadi, steyaert}@lix.polytechnique.fr*

### ABSTRACT

Evolution acts in several ways on biological sequences: either by mutating an element, or by inserting, deleting or copying a segment of the sequence. Varré et al. [13] defined a transformation distance for the sequences, in which the evolutionary operations are copy, reverse copy and insertion of a segment. They also proposed an algorithm to calculate the transformation distance. This algorithm is  $O((n + m)^4)$  in time and  $O((n + m)^4)$  in space, where  $n$  is the size target sequence and  $m$  is the size of the source string. of the sequences. In this paper, we propose an improved algorithm which costs  $O(n(n + m))$  in time and  $O(n^2 + m)$  in space. Furthermore, we extend the operation set by adding point deletions. We present an algorithm which is  $O(n^2m)$  in time and  $O(n^2 + m)$  in space for this more general model.

*Keywords:* Dynamic Programming, Pattern Matching

### 1. Introduction

Building models and tools to quantify evolution is an important domain of biology. Evolutionary trees or diagrams are based on statistical methods which exploit comparison methods between genomic sequences. Many comparison models have been proposed according to the type of physico-chemical phenomena that underly the evolutionary process [6]. Different evolutionary operation sets are studied. Mutation, deletion and insertion were the first operations dealt with [8]. Duplication and contraction were then added to the operation set [1, 2]. All these operations were acting on single letters, representing bases, amino-acids or more complex sequences: they are called point transformations. Segment operations are also very important to study. In a number of papers [12, 13, 14], Varré et al. have studied an evolutionary distance based on the amount of segment moves that Nature needed (or is supposed to have needed) to transfer a sequence from one species to the equivalent sequence in another one. Their model is concerned with segments copy with or without reversal and on segment insertion: it is thus a very simple and robust model which can easily be

---

<sup>1</sup>Full version of a submission presented at the *Prague Stringology Conference* (Czech Technical University in Prague, Czech Republic, September 22–24, 2003).