Journal of Automata, Languages and Combinatorics **10** (2005) 5/6, 671–685

A LINEAR ALGORITHM FOR THE DETECTION OF EVOLUTIVE TANDEM REPEATS[1]

RICHARD GROULT
LaRIA, Université de Picardie Jules Verne
33 Rue Saint Leu, 80039 Amiens cedex 01, France
e-mail: `richard.groult@u-picardie.fr`

MARTINE LÉONARD
LITIS Research Laboratory, Avenue de l'Université - BP 8
76801 Saint-Étienne-du-Rouvray Cedex, France
e-mail: `martine.leonard@univ-rouen.fr`

and

LAURENT MOUCHARD
LITIS Research Laboratory, Avenue de l'Université - BP 8
76801 Saint-Étienne-du-Rouvray Cedex, France, and
Dept. Computer Science, King's College London, London WC2R 2LS, England
e-mail: `laurent.mouchard@univ-rouen.fr`

ABSTRACT

We present here a linear algorithm for the detection of evolutive tandem repeats. An evolutive tandem repeat consists in a series of almost contiguous copies, every copy being similar (using Hamming distance in this article) to its predecessor and successor. From a global view point, evolutive tandem repeats extend the traditional approximate tandem repeat where each copy has to be in a neighbourhood of a given model. Due to the lack of algorithms, these repeats have been discovered in genomic sequences only recently. In this article, we present a two-stage algorithm, where we first compute an array containing all the Hamming distances between candidates, then we visit this array to build a complete evolutive tandem repeat from insulated pairs of copies. Moreover, we explain how it is still consistent with the usual technique devoted to dynamic programming which consists in filling a comparison matrix and backtracking through it to find an optimal alignment.

Keywords: Linear algorithm, evolutive tandem repeats, Hamming distance

1. Introduction

The notion of approximate tandem repeat is generally well-defined, from the formal view point [2, 12], it uses a consensus model, every copy participating to this repeat

---

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