COMPUTING THE REPETITIONS IN A BIOLOGICAL WEIGHTED SEQUENCE [1]

COSTAS ILIOPOULOS

Department of Computer Science, King's College London Strand
London WC2R 2LS, England
e-mail: `csi@dcs.kcl.ac.uk`

LAURENT MOUCHARD

ABISS, Atelier Biology, Informatics, Statistics and Sociolinguistics, Université de Rouen
76821 Mont Saint Aignan Cedex, France
e-mail: `laurent.mouchard@univ-rouen.fr`

KATERINA PERDIKURI

Department of Computer Engineering and Informatics, University of Patras
26500 Patras, Greece
e-mail: `perdikur@ceid.upatras.gr`

and

ATHANASIOS TSAKALIDIS

Research Academic Computer Technology Institute
61 Riga Feraiou Str., 26221 Patras, Greece
e-mail: `tsak@cti.gr`

ABSTRACT

One of the most important goals in computational molecular biology is allocating repeated patterns in nucleic or protein sequences, and identifying structural or functional motifs that are common to a set of such sequences. Although the problem of computing the repetitions in biological sequences has been extensively studied, in the relevant literature, the problem of computing the repetitions in biological weighted sequences has not been efficiently solved. In this work we present an $O(n^2)$ algorithm for computing the set of repetitions in a biological weighted sequence with probability of appearance larger than $1/k$, where k is a given constant. Our algorithm can be applied in the detection of the repeated patterns in biological weighted sequences such as assembled DNA sequences.

Keywords: Repetitions, weighted sequences, DNA assembly

---

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