AUN/SEED-Net JICA

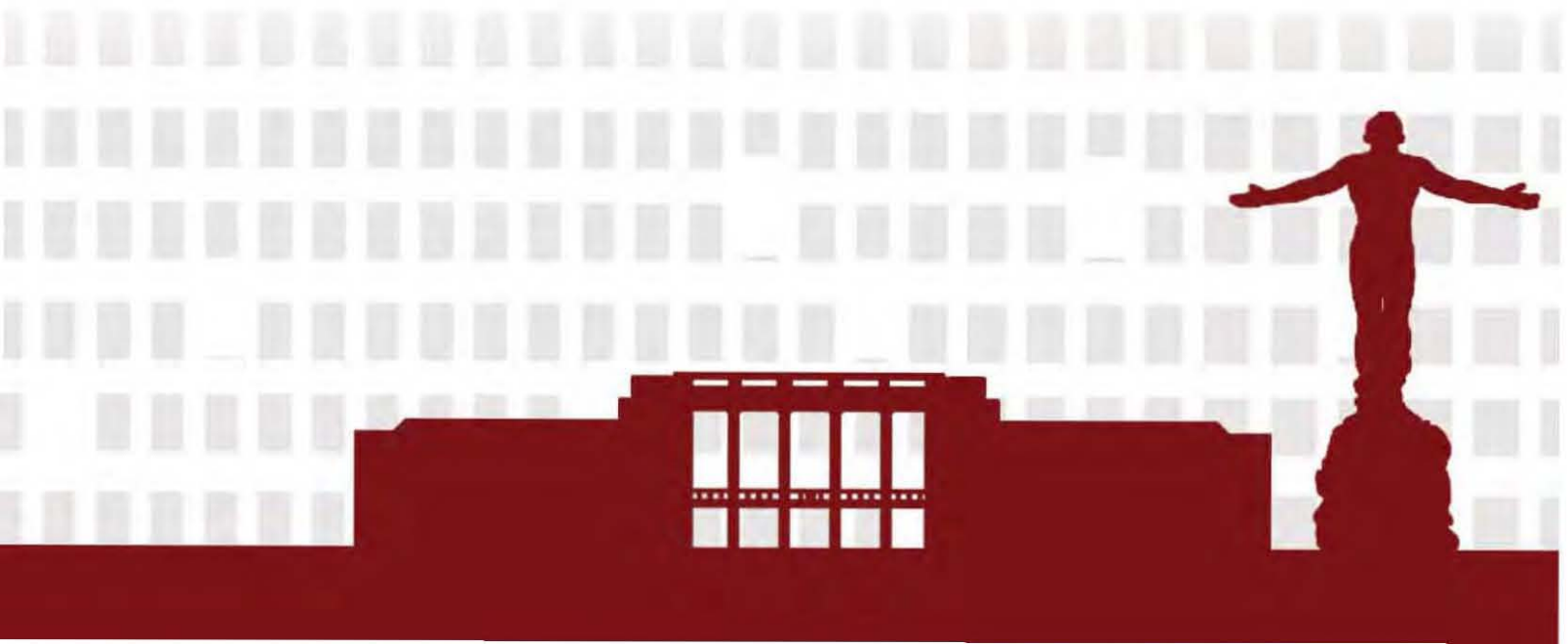
8th AUN/SEED-Net
REGIONAL CONFERENCE ON ELECTRICAL AND ELECTRONICS ENGINEERING

Envision, Enable, and Empower
Smarter and Resilient Societies

co-located with

11th ERDT Conference
on Semiconductor and Electronics, Information and Communications Technology and Energy

16-17 November 2015
Metro Manila, Philippines



Proceedings of the 8th AUN/SEED-Net RCEEE 2015 and 11th ERDT Conference
on Semiconductor and Electronics, Information and Communications Technology, and Energy

Editors:
Dr. Joel Joseph S. Marciano Jr.
Dr. Jhoanna Rhodette I. Pedrasa
Dr. Rhandley D. Cajote

ISBN: 978-616-406-075-3

Published by: ASEAN University Network / Southeast Asia Engineering Education Development Network (AUN/SEED-Net) JICA Project
Faculty of Engineering, Bldg. 2
Chulalongkorn University, Bangkok
Thailand 10330

Printed in the Philippines by: ERZALAN PRINTING PRESS
45 Cotabato Street, Luzviminda Village, Batasan Hills, Quezon City, Philippines

8th AUN/SEED-Net Regional Conference
on Electrical and Electronics Engineering 2015

co-located with

11th ERDT Conference
on Semiconductor and Electronics, Information and Communications Technology, and Energy

Envision, Enable and Empower Smarter and Resilient Societies

Published by: ASEAN University Network / Southeast Asia Engineering Education Development Network (AUN/SEED-Net) in partnership with Engineering Research and Development for Technology (ERDT) and University of the Philippines Diliman.

ISBN: 978-616-406-075-3

AN EFFICIENT EXACT SOLUTION TO THE (*L, D*)-PLANTED MOTIF PROBLEM

Maria Clara Isabel D. Sia*, Julieta Q. Nabos, and Proceso L. Fernandez

Department of Information Systems and Computer Science,
Ateneo de Manila University, Quezon City, PHILIPPINES.
*E-mail: aia.sia1995@gmail.com

ABSTRACT

DNA motif finding is widely recognized as a difficult problem in computational biology and computer science. Because of the usual large search space involved, exact solutions typically require a significant amount of execution time before discovering a motif of length *l* that occurs in each sequence *Si* from an input set {*S1,…,St*} of sequences, allowing for at most *d* substitutions.

In this paper, we propose a novel algorithm that operates on a compact bit-based representation of the search space and takes advantage of distance-related patterns in this representation in order to compute the exact solution for any arbitrary problem instance up to *l*=17, *d*=6.

A Java implementation—run on synthetic datasets for various challenge instances of the (*l, d*) motif finding problem—shows the proposed algorithm to be highly competitive against PMS8 and qPMS9, two current state-of-the-art exact motif search algorithms. The proposed algorithm works extremely well for problems involving short motifs, outperforming the current best algorithm for the challenge problem instances (13,4) and (15,5) with a run-time reduction of at least 50% and 20% respectively for these instances, while ranking second to qPMS9 for test instances involving *l*=16 and *l*=17.

Keywords: DNA motif finding, exact motif search, Java