NITheCS
National Institute for
Theoretical and Computational Sciences

NITheCS COLLOQUIUM:

Bioinformatics – a personal journey

Prof Hugh Patterton (University of Stellenbosch)

Monday, 3 October 2022 | 16h00 – 17h00 SAST

ABSTRACT

A broad consensus view is that life originated in the enclosed spaces of alkaline hydrothermal vents. Although it seems feasible that complex chemistries can originate in these enclosures, it does not explain how the information encoding catalysts originated. Although it is likely that RNA ribozymes were the original catalysts, it is unclear how the information was derived to encode a catalyst for the complex chemistries that occurred in the vent enclosures. This represents the two big groups in the argument: Information First or Chemistry First.

The relationship between a complex chemistry and the information required to drive the chemistry and a von Neumann automaton is striking, and it is useful to think of living organisms as von Neumann automata. The information encoding the establishment and maintenance of the chemistry is encoded in DNA. When looking at the total length of DNA used to encode genes, it is striking that there is a significant increase in genome size with the appearance of eukaryotes. What happened here in the evolutionary timeline? The appearance of histones to allow lengthwise compaction of the DNA is almost certainly the reason. The degree of information compaction in chromatin is higher than found in any encoding scheme. Histones provide a regulatory interface to genome function.

To get to the information in DNA for analysis, it must be sequenced. One of the most widely used sequencing technologies is the Illumina method. To analyse DNA, sub-sequences must be 'found'. There are many different text searching algorithms. We examine why the Bower-Wheeler algorithm is so useful in sequence read mapping. We apply the BW algorithm to protein associated DNA sequences and map the distribution of proteins in the genome. I show how barcoding can be used to select specific populations from an experimental set, and how mapping protein distributions on the genome can provide insight into biological effects such as aging.

I further demonstrate how the positions of all nucleosomes in a genome can be derived, and how this information can be used to understand gene expression.

Finally, I look at the analysis of proteomic data, profile matching algorithms, and the mapping of the genome-wide distribution of specific chemical modifications in the genome. I discuss how the specific location of these modifications can impact on local gene regulation.

BIOGRAPHY

Prof Patterton is the Director of the Centre for Bioinformatics and Computational Biology (CBCB) at SU. He received his PhD in Biochemistry from UCT in 1991.

In 1992 he was awarded a Fogarty Visiting Fellowship and spent the next 4 years at the National Institutes of Health in Bethesda, USA. In 1996 he moved to Pennsylvania State University as a Research Associate. In this time, he completed writing DNAtools, the Stider equivalent for Windows in C++.

In 1998 he returned to South Africa and was awarded a Wellcome Trust International Senior Research Fellowship in Biomedical Science.

Prof Patterton was part of group that set up the National Bioinformatics Network to establish Bioinformatics in South Africa. He then launched the Bioinformatics node at UCT, the Facility for Genomics and Proteomics at UFS, and the CBCB at SU. He has also set up an undergraduate programme and BScHons, MSc and PhD degrees in Bioinformatics and Computational Biology at SU.

REGISTER

Visit https://bit.ly/3y6h0vv
or scan/click:

MEET THE SPEAKER AFTER THE EVENT

visit:
kumospace.com/nithecs_social