Vrije Universiteit Brussel

Faculty of Sciences and Bioengineering Sciences

The Research Group of

Microbiology

has the honour to invite you to the public defence of the PhD thesis of

Ahmed MOHAMED MYSARA ABDELWAHAB

to obtain the degree of Doctor of Bio-engineering Sciences

Title of the PhD thesis:

From Sequencing Reads to Microbial Diversity: Bioinformatic Algorithms for Processing Amplicon Sequencing Data

Joint PhD with KUL

Promotor:

Prof. dr. D. CHARLIER

The defence will take place on

September 27th 2016 at 15.00 h

in Auditorium D.2.01 at the Campus Etterbeek of the Vrije Universiteit Brussel, Pleinlaan 2 in 1050 Elsene, and will be followed by a reception

Members of the jury

Prof. Dr. D. SPRINGAEL (KUL, chairman)
Prof. Dr. W. VRANKEN (VUB, secretary)
Prof. Dr. J. RAES (KUL, co- promotor)
Dr. P. MONSIEURS (SCK-CEN, co-promotor)
Prof. Dr. S. WECKX (VUB)
Prof. Dr. V. VAN NOORT (KUL)
Prof. Dr. P. VANDAMME (UG)
Dr. C. QUINCE (Univ. of Warwick, UK)

Curriculum vitae

Mohamed Mysara Ahmed (born on February 17th 1986 in Nigran, Saudi Arabia) obtained his BSc in Pharmaceutical Sciences at Cairo University (Egypt) and MSc in Information Technology at the University of Nottingham (UK), winning the best dissertation on the ITI Programme. Mohamed authored seven publications in peer-reviewed journals and a book chapter. He presented six talks and six research posters in national and international conferences. He supervised different Master thesis projects. His PhD was financially supported by an SCK●CEN grant of the Belgian Nuclear Research Centre

Abstract of the PhD research

The development of high-throughput sequencing technologies has revolutionized the field of microbial ecology by offering a cost-efficient method to assess microbial diversity at an unseen depth using 16S rRNA amplicon sequencing approaches. Different preprocessing algorithms need to be performed to obtain a collection of highly reliable sequencing reads, ending with a clustering step to group them into Operational Taxonomic Units (OTUs) However, this approach is posing various challenges: the removal of PCR artefacts (called chimera), correction of sequencing errors resulting from the sequencing technologies and clustering those sequences into OTUs. In this work various bioinformatics tools were developed to tackle those challenges.

First, an ensemble classifier for chimera detection was developed named CATCh, which obtained a higher performance on different types of sequencing data compared to existing tools. Secondly, two artificial intelligence-based algorithms, NoDe and IPED, able to treat sequencing errors within 454 Pyrosequencing and Illumina MiSeq data respectively, were introduced. A benchmarking study was performed comparing NoDe and IPED, showing a more pronounced decrease of the error rate compared to other state-of-the art tools. Thirdly, a new method was developed introducing an adaptive cut-off score in the OTU clustering step, as such making the results of the OTU clustering less sensitive to variations in evolutionary rates between taxonomic lineages and to the region of the 16S rRNA gene targeted for amplification. Implementing such a dynamic cut-off value resulted in closer correspondence between the number of OTUs and the actual diversity of the samples. Finally, a benchmark analysis comparing existing pipelines for 16S rRNA metagenomics data processing was performed, showing that an integration of our in-house developed algorithms achieved the highest accuracy. Conclusively, the newly developed pipeline within this PhD translates amplicon sequencing data into high-quality OTUs tendering robust diversity estimates.