Study of the association between eating habits, the composition of the gut microbiota and the incidence of colon cancer in the Moroccan population using a metagenomic approach

Imane Allali
PhD Student
Faculty of Sciences, Rabat
Mohammed First University, Oujda
Morocco
allali.imane@gmail.com
Interest of the Project

Colon cancer is the 3rd cause of death in Morocco

<table>
<thead>
<tr>
<th>Location</th>
<th>Incidence standardisée</th>
<th>Homme</th>
<th>Femme</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA 2004 - 2008 [9]</td>
<td>38,3</td>
<td>30,6</td>
<td></td>
</tr>
<tr>
<td>France (Francim) 2005 [10]</td>
<td>37,7</td>
<td>24,5</td>
<td></td>
</tr>
<tr>
<td>Chine (Hong Kong) 1998 - 2002 [14]</td>
<td>23,8</td>
<td>18,9</td>
<td></td>
</tr>
<tr>
<td>Algérie (Oran) 1996 - 2004 [13]</td>
<td>4</td>
<td>3,4</td>
<td></td>
</tr>
<tr>
<td>Maroc RCRC 2005 - 2007</td>
<td>4,7</td>
<td>3,2</td>
<td></td>
</tr>
</tbody>
</table>

Attributed to the westernization of the diet in many African populations
Objectives

• make a comprehensive analysis of the composition of the microbiota in the Moroccan population

• define the complexity of communities in the intestinal tract

• demonstrate the associations between the microbiota and colon cancer according to dietary and genetic characteristics
Metagenomic approach

Filtering

Short Reads
Remove adapters
Remove of contaminant DNA
Assembly reads into contigs:
SOAPdenovo
Velvet

Reads removal

Assembly

MG-RAST
IMG

Annotation & Function
Metagenomic analysis: Galaxy

Sample: Saliva Microbiome

HMIWGS/HMASM - Illumina WGS Reads and Assemblies
In the first phase of WGS sequencing, 764 samples were sequenced, comprising 16 body sites. Of these, 749 samples underwent assembly. Reads for all 764 samples, and 749 assemblies are provided here.

Reads and assemblies were subjected to QC assessment, including identification of outliers by mean contig & ORF density, human hits, tRNA hits and size. 690 samples passed this QC and were included in downstream wgs analyses.

This dataset includes over 35 billion human contaminant-screened reads in FASTQ format, which are 2.3 TB in size, compressed. Reads from each individual sample were assembled using SOAP, generating 46.3 million scaffolds with a total compressed size of 13 GB.

- Data Table
- Protocols and Tools
- Related Pages

http://www.hmpdacc.org/HMASM/
Metagenomic analysis: Galaxy

The Workflow for metagenomic analysis of the sample of Saliva

https://main.g2.bx.psu.edu/
Metagenomic analysis: Galaxy

Phylogeny tree

Result of the different species found after the metagenomic analysis
Questions

• How can we choose the appropriate software for each step of the metagenomic analysis?

• How can we be sure that the assembly is done correctly?

• What is the minimum score to consider for annotation?
Thank you