

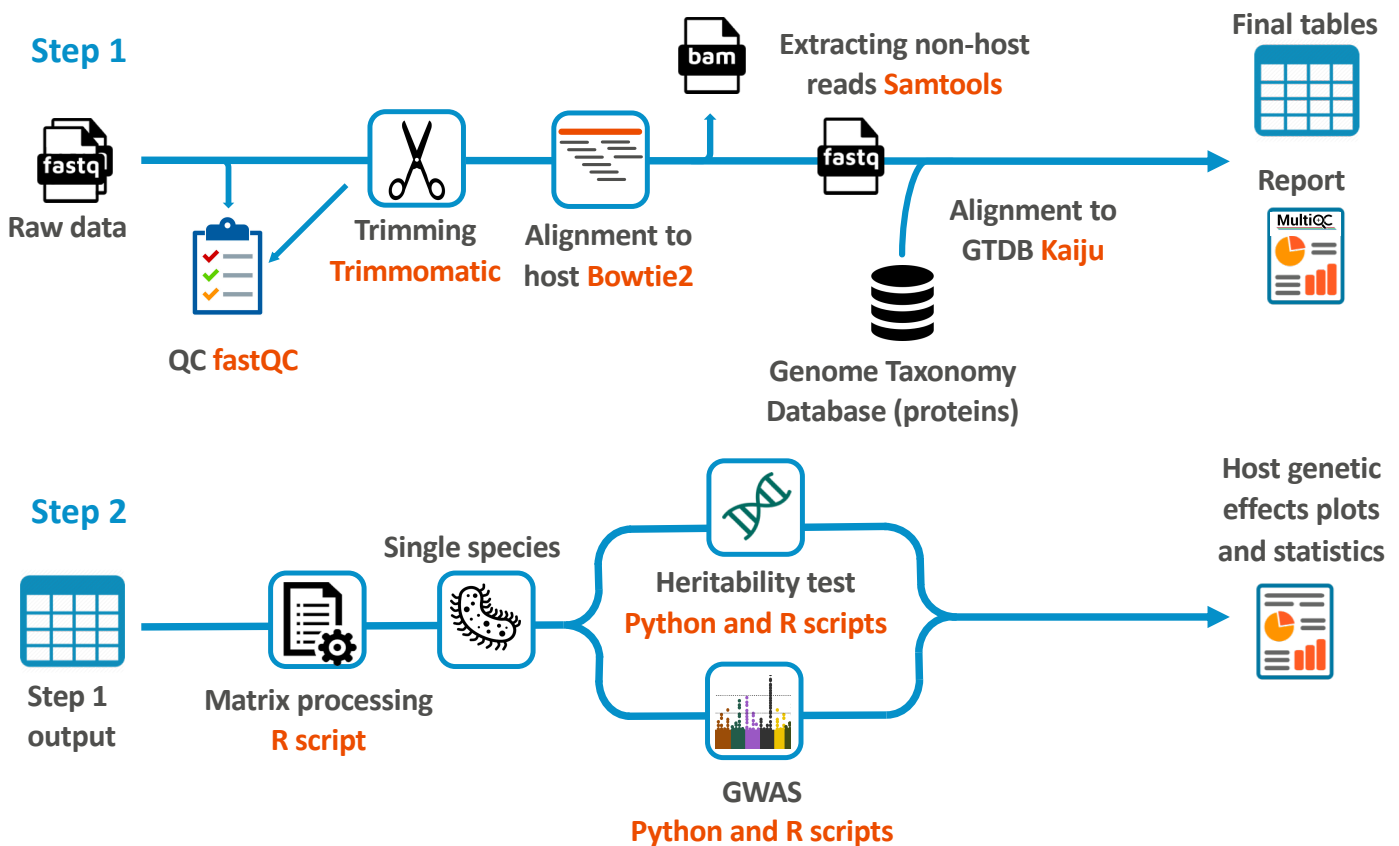
A suite of Nextflow pipelines for profiling microbiomes from shallow shotgun data and then test for host genetic effects.

Luca Cozzuto^{1,2}, Felipe Morillo^{1,2}, Helene Tonnele^{1,2}, Fernando Duarte^{1,2}, Julia Ponomarenko^{1,2}, Toni Hermoso^{1,2}, and Amelie Baud^{1,2}

1. Centre for Genomic Regulation (CRG), Barcelona, Spain
2. Universitat Pompeu Fabra (UPF), Barcelona, Spain

Biological background

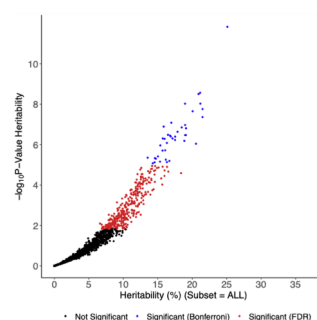
One of the main questions in microbiome studies is to what extent the host genetics influence their microbiome. To answer this we analysed the gut microbiome of 796 outbred rats using shallow NGS sequencing (~1mi pairs of reads per sample). We then used two Nextflow pipelines for the whole analysis. The first step infers the bacterial composition by looking at the translated gene content from a comprehensive catalogue (GTDB), with reads mapping to multiple genomes being weighted based on the univocally assigned ones. The second one removes low quality samples with low abundant and low prevalent bacteria from the dataset, transforms the data based on centred-log ratios and regresses out known covariates. Then it uses the genotypes of the animals (~3mi SNPs) to detect aggregate genetic effects (Heritability test) and find correlations between many host's genetic loci and bacterial composition (GWAS).



Preliminary results

1. Heritability tests

Taxonomic rank	Total number of mapped taxa	Significant taxa (FDR<10%)	Significant taxa (Bonferroni<5%)
Phylum	15	0	0
Class	18	0	0
Order	35	2	0
Family	58	11	0
Genus	224	45	4
Species	1,175	331	36
Total	1,525	389	40



2. GWAS

