Non-redundant Clustered Gene Index
Institute for Genome Sciences, UMD School of Medicine

Author: Brandi Cantarel
Version:
Effective Date:

1 Abstract

2 Introduction

This SOP describes creation of a non-redundant catalog of bacterial genes by body site. This was done by clustering the gene predictions coming out of the Illumina metagenomic wgs sequence, using the same parameters used by the Metahit project[1] to cluster their human intestinal tract data.

3 Requirements

3.1 Software requirements
USEARCH (http://www.drive5.com/usearch/)

4 Procedure

4.1 Clustering
Clustering was performed as described in Qin et al[1], with the exception that clustering was done here using USEARCH, rather than BLAT.

MetaGeneMark predicted ORFs (available for download at http://hmpdacc.org/HMGI/) were aligned to one another using Usearch[2] (http://www.drive5.com/usearch/), version 5.0.144. The following options were used:

    -id 0.95
    -targetfract 0.90

5 Implementation

6 Discussion

7 Related Documents & References

[1] Qin, J. et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature;464(7285):59-65.

[2] Edgar, RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics; 26(19):2460-1.

Author: Brandi Cantarel
Version:
Effective Date:

8 Revision History

Version	Author/Reviewer	Date	Change Made
1.01		11/20/2011	Establish SOP