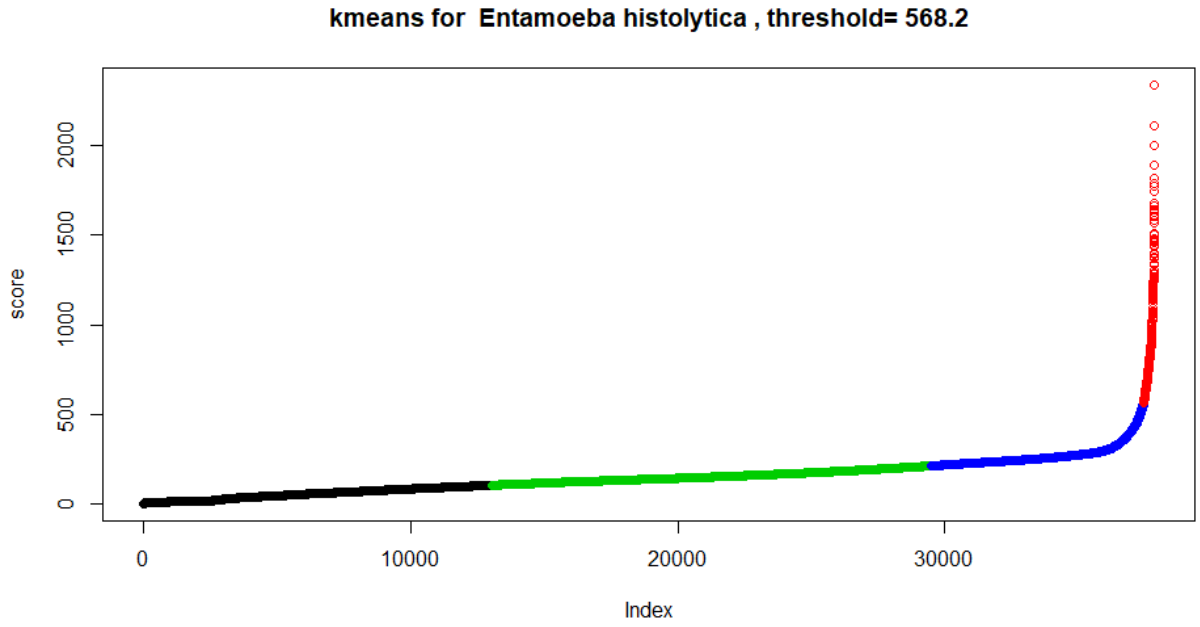
Description

Using a paralog search pipeline, ORFs of *Entamoeba histolitica* were used to search the **chembl_25** database for paralog targets and drugs[1],[2],[3, 4]. Similarity scores were uploaded to the **hmmer_statistics table** created in that database as described [1].

Kmeans clustering was used to determine a filtering threshold for selecting report results [1, 5].



341 distinct drugs were found.

Spreadsheet showing drug and target information was saved as **E_histolytica_hmmer_drugs.xlsx**.

Spreadsheet with annotations from **chembl_25** database for each drug/target was saved as **E_histolytica_hmmer_drugs_annotated.xlsx**.

Spreadsheets are included in supplements in git repository **https://github.com/jeremy-b-singer/paralog_targets.git.**

References

1. Singer JB: **Finding Paralog Targets for Neglected Diseases**. In. Edited by paralog_targets.pdf; 2020: Methods for finding targets and drugs for a diisease caused by a pathogen searching the ChEMBL database for existing targets with protein sequences similar to ones in the pathogen genome.
2. **AmoebaDB Download Files** [https://amoebadb.org/common/downloads/Current_Release/EhistolyticaHM1IMSS/fasta/data/AmoebaDB-46_EhistolyticaHM1IMSS_ORFs_AA.fasta]
3. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B *et al*: **ChEMBL: a large-scale bioactivity database for drug discovery**. *Nucleic Acids Res* 2012, **40**(Database issue):D1100-1107.
4. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, Mutowo P, Atkinson F, Bellis LJ, Cibrián-Uhalte E *et al*: **The ChEMBL database in 2017**. *Nucleic Acids Res* 2017, **45**(Database issue):D945-954.
5. **forgy: Initialization of cluster prototypes using Forgy's algorithm in inaparc: Initialization Algorithms for Partitioning Cluster Analysis**. 2020.