UNDERSTANDING PANGENOMICS AND COMPARATIVE STUDY OF EXIGUOBACTERIUM STRAINS

BACKGROUND

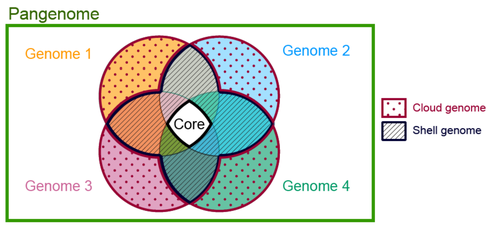
The abundance of genome data being produced by the new sequencing techniques is providing the opportunity to investigate gene diversity at a new level. A pan-genome analysis can provide the framework for estimating the genomic diversity of the dataset at hand and give insights towards the understanding of its observed characteristics. Here we provide a systematic approach for constructing the groups inherently associated with a pan-genome analysis, using the complete proteome data of Exiguobacterium genomes as the driving case study. As opposed to similar studies, the presented method requires a complete information system (i.e. complete genomes) in order to produce meaningful results.

PANGENOMICS

Pangenomics is a genomic concept that takes into account the entire set of genes, including both core and accessory genes, within a group of related organisms. Unlike traditional genomics, which often focuses on a single reference genome, pangenomics considers the genomic diversity across multiple individuals or strains of a species. The pangenome is the complete set of genes found in the entire population, and it is often divided into different parts to better understand the genetic makeup of a species.

Parts of the pangenome:

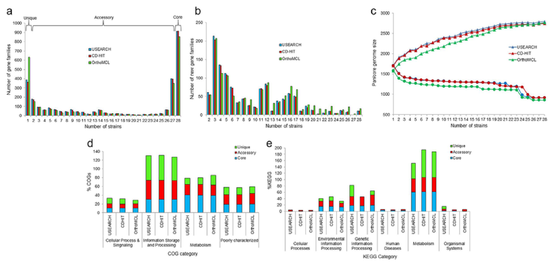
1. Core Genome:
   - The core genome represents the set of genes shared by all individuals within a species.
   - These genes are often essential for basic cellular functions and are considered the genetic backbone of the species.
2. Shell Genome:
   - The shell genome includes genes that are shared by a significant portion of the population but not by all individuals.
   - These genes contribute to common traits and characteristics within subgroups of the species.
3. Cloud Genome:
   - The cloud genome consists of genes that are unique to specific individuals or subgroups.
   - These genes contribute to the diversity of the species and may be associated with specific adaptations or environmental niches.
4. Singletons:
   - Singleton genes are unique to individual organisms and are not shared with any other members of the population.
   - These genes may result from recent mutations or horizontal gene transfer.



BACTERIAL PAN GENOMIC ANALYSIS PIPELINE (BPGA)

BPGA is an ultra-fast software package that provides comprehensive pan genome analysis of microorganisms. In addition to all types of routine pan genomic analyses (Pan genome Profiles, Pan/Core Phylogeny etc.), BPGA includes a number of novel downstream analysis features like Exclusive Gene Family Analysis, Atypical GC Content Analysis, Subset Analysis, MLST based on housekeeping genes and KEGG Distribution etc.

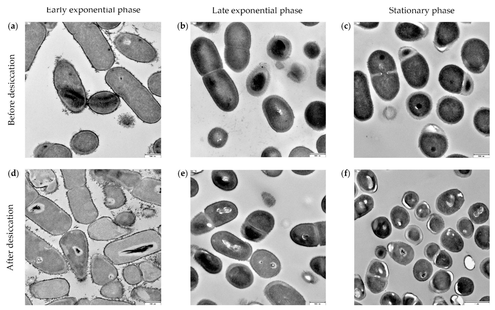
Another unique feature of BPGA is that it allows the user to select from three different tools for ortholog clustering – USEARCH, CD-HIT & OrthoMCL, the first one being the default clustering tool



EXIGUOBACTERIUM

Exiguobacterium are globally diverse organisms that are found in a variety of environments including microbialites (Thrombolite from Pavilion Lake, BC and Stromatolites from Laguna Socompa, Argentina), ocean, freshwater lakes, Himalayan ice, Himalayan soil, hydrothermal vents, brine shrimp, gastrointestinal tract of marine fish and in microbial biofilms

Twenty genomes from the genus have been completed as either complete (one circular chromosome, with plasmids) or in a draft format (containing multiple unassembled contigs)
Their ability to degrade plastic and to survive in varying temperature extremes makes them an important area of study.

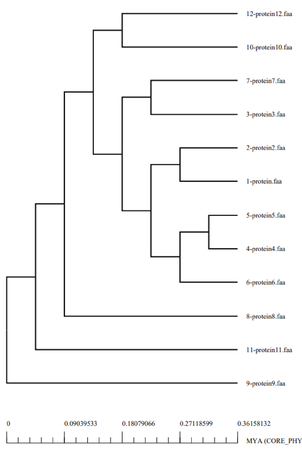


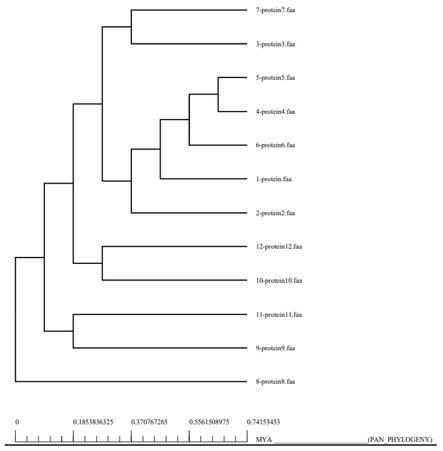
PRINCIPLE

The main objective is to isolate enzymes capable of breaking down polystyrene (PS) by adopting a methodology similar to that used for previously identified plastic-degrading enzymes like PETase, MHETase, Demetra, Ceres, Cora, and Cibeles. This involves isolating and studying the genome of Exiguobacterium to identify potential PS-degrading enzymes.
Work flow involved accessing proteomes of Exiguobacterium sp. from NCBI database, clustering and pangenomic analysis using BPGA Pipeline, data analysis using visual aids (histogram, bar graphs, etc.) and data interpretation.

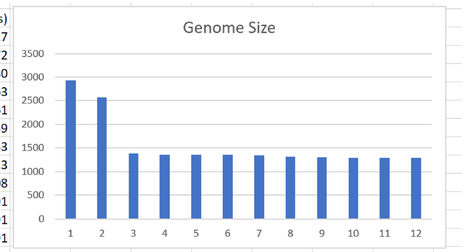
ANALYSIS

Org code.	Scientific name	Modifier
1	Exiguobacterium sibiricum 255-15	255-15 (strain)
2	Exiguobacterium acetylicum	SI17 (strain)
3	Exiguobacterium profundum	TSS-3 (strain)
4	Exiguobacterium alkaliphilum	B-3531D (strain)
5	Exiguobacterium acetylicum	EYLC1 (strain)
6	Exiguobacterium acetylicum	UTDF19-27C (strain)
7	Exiguobacterium acetylicum	AMCC 101217 (strain)
8	Exiguobacterium mexicanum	A-EM (strain)
9	Exiguobacterium aurantiacum	CX2 (strain)
10	Exiguobacterium antarcticum B7	Eab7 (strain)
11	Exiguobacterium marinum	a-1 (strain)
12	Exiguobacterium sp. AT1b	AT1b (strain)

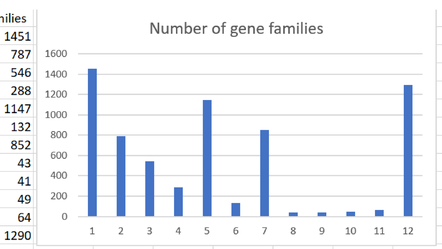


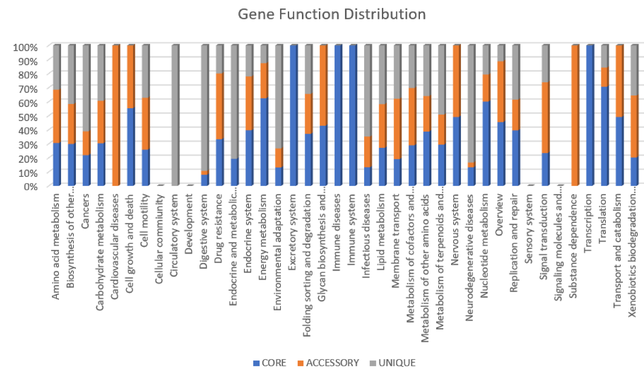


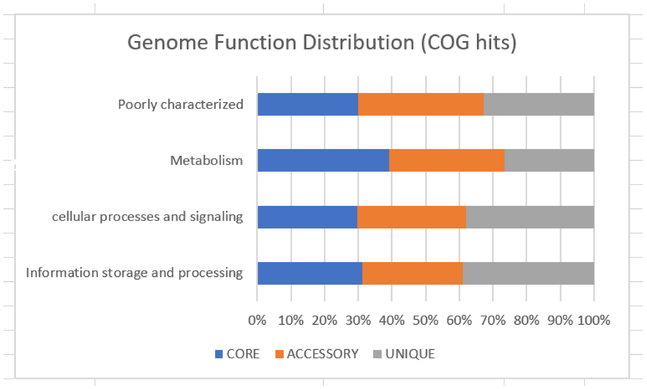
Org	Genome size (residues)
1	2927
2	2572
3	1380
4	1363
5	1361
6	1359
7	1343
8	1313
9	1308
10	1291
11	1291
12	1291



Organism	Number of gene families
1	1451
2	787
3	546
4	288
5	1147
6	132
7	852
8	43
9	41
10	49
11	64
12	1290







RESULTS & CONCLUSION

Core gene involvement high(>50%) in transcription, translation, immune system, immune diseases, energy metabolism and cell growth and death.

Accessory gene involvement high(>50%) in cardiovascular diseases, glycan biosynthesis and metabolism, nervous system, signal transduction, substance dependence, transport and catabolism and xenobiotics biodegradation.

Genes of interest (plastic degradation) for xenobiotics biodegradation lie in accessory gene indicating moderate conservation of these genes or later stage evolution.

The genes in accessory pangenome region might have evolved later as xenobiotic abundance increased in their niche, i.e. plastic and other pollution increased. Since Exiguobacterium are extremophiles, factors such as high scarcity of food and high competition must have driven the evolution to find alternate sources of nutrition.

FUTURE ASPECT

After pinpointing promising candidates, enzymatic assays should be conducted to confirm their plastic-degrading activity. Subsequently, spectroscopic studies must be employed to accurately characterize the structure and function of these enzymes. This comprehensive approach aims to isolate and understand PS-degrading enzymes from Exiguobacterium for potential applications in addressing plastic pollution.

REFERENCES

- Marschall,T. et al. (2016) Computational Pan-Genomics: Status, Promises and Challenges
- S.K. Aggarawal et al. (2022) Pangenomics in Microbial and Crop Research: Progress, Applications, and Perspectives
- María Castillo López et al. (2021) Xerotolerance: A New Property in Exiguobacterium Genus
- Yang, Li et al. (1 January 2021). "Biodegradation of expanded polystyrene and low-density polyethylene foams in larvae of Tenebrio molitor Linnaeus (Coleoptera: Tenebrionidae): Broad versus limited extent depolymerization and microbe-dependence versus independence"
- B.Y. Peng et al. Biodegradation of polystyrene by dark (Tenebrio obscurus) and yellow (Tenebrio molitor) mealworms (Coleoptera: Tenebrionidae) Environ. Sci. Technol., 53 (2019), pp. 5256-5265
- Deepika Chauhan et al. (2018) Biofilm formation by Exiguobacterium sp. DR11 and DR14 alter polystyrene surface properties and initiate biodegradation
- Lalina Maroof et al. (2022) Microbial degradation of low density polyethylene by Exiguobacterium sp. strain LM-IK2 isolated from plastic dumped soil
- Wikipedia:
  https://en.wikipedia.org/wiki/Exiguobacterium
  https://en.wikipedia.org/wiki/Category:Organisms_breaking_down_plastic
  https://en.wikipedia.org/wiki/File:Parts_of_the_pangenome.png
- Genome - NCBI - NLM (nih.gov)