Development and Beta-Testing of a Bioinformatics Pipeline for Metagenomic Analysis

Students: Anna Noyvert, Malika Kopbayeva
Mentors: Cynthia Ronkowski,
Anton Katsuba, Serghei Mangul PhD

MetaSUB

MetaSub's goal is to map the urban microbiome to build a molecular profile of cities around the globe. http://metasub.org/

Global City Sampling Day (gCSD) is an event that takes place annually on June 21st. During gCSD, samples from surfaces in subways, buses, airports and other well-traveled public spaces are collected across the world's cities.

The collected samples are then sequenced, using Illumina HiSeq, and analyzed to develop a genetic and epigenetic map detailing the microbiome in each participating city.

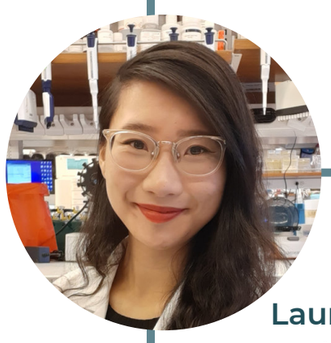
The MetaSUB bioinformatics working group is developing a core set of tools that can be run on metagenomes collected during gCSD and is highly reproducible and extendable.

MetaSUB

CAMP

Core Analysis Modular Pipeline

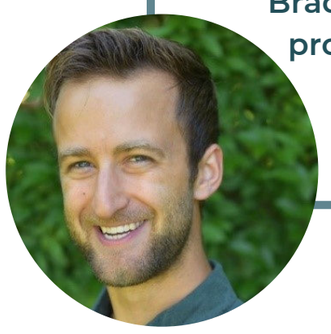
- Standardized quality control and abundance outputs
- Alternative to traditional pipeline
- Easily updated by any software developer
- Distributed, long-term development approach
- Integrates cutting-edge bioinformatics tools



Lauren Mak - project leader

Modules under development:

- Short read quality control
- Short read assembly
- **Gene catalog**
- Consensus binning
- Short read taxonomy
- Assembled virus characterization



Braden Tierney - project leader

QR code for CAMP

Gene catalog module

- ORF prediction/annotation using Prokka
- Non-redundant ORF clustering using MMseqs2
- Relative ORF abundance quantification with our own script

snakemake

Beta-testing

- Include rulegraph in the main README.
- Document rules in Snakefile with comments.
- References to {{Cookiecutter.module_slug}} are unclear, especially if the meaning of slug is unknown.
- Usage of configuration .yaml files is unclear.
- Link for fixing cases of dependency conflicts is in question.

Acknowledgements

SMTB
TARTU ÜLIKOOL · UNIVERSITAS TARTUENSIS · 1632
ZIMIN FOUNDATION
hhmi | Howard Hughes Medical Institute

Разработка и бета-тестирование биоинформатического пайплайна для метагеномного анализа

Студенты: Анна Нойверт, Малика Копбаева
Наставники: Синтия Ронковски, Сергей Мангул, Антон Кацубо

MetaSUB

Цель MetaSub - составить карту городского биома для создания молекулярного профиля городов по всему миру, чтобы улучшить их дизайн, функциональность и влияние на здоровье.

Глобальный день сборки проб в городах - это мероприятие, которое проводится ежегодно 21 июня. Во время гСП в разных городах мира собираются образцы с поверхностей в метро, автобусах, аэропортах и других общественных местах с интенсивным трафиком.

Затем собранные образцы секвенируются с использованием Illumina HiSeq и анализируются для составления генетической и эпигенетической карты микробиома каждого города-участника.

Рабочая группа MetaSUB разрабатывает основной набор инструментов, который может быть использован на каждом метагеноме, собранным во время gCSD, и обладает высокой воспроизводимостью и расширяемостью.

MetaSUB

CAMP

Core Analysis Modular Pipeline

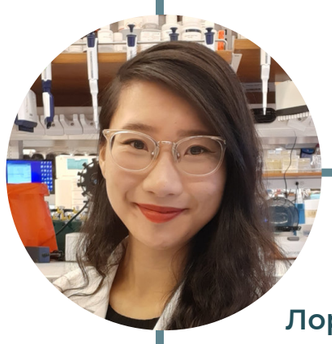
Стандартизированный контроль качества и выходные данные по изобилию

Альтернатива традиционному пайплайну

Легкое для обновления любым разработчиком программного обеспечения

Распределенный, долгосрочный подход к разработке

Интеграция передовых инструментов биоинформатики



Лорен Мак, ведущий проекта

Модули в стадии разработки:

- Контроль качества коротких ридов
- Сборка коротких ридов
- **Каталог генов**
- Консенсусный биннинг
- Таксономия коротких ридов
- Характеризация собранных вирусов

Брэйден Тирни, ведущий проекта

QR код для CAMP

Модуль генного каталога

- **Прогнозирование & аннотация ОРС (открытых рамок считывания) с помощью Prokka**
- **Кластеризация неизбыточных ОРС с помощью MMseqs2**
- **Количественная оценка относительного изобилия ОРС с помощью своего кода**

snakemake

Beta-testing

- Включение графика правил в основной README.
- Документация правил в Snakefile с комментариями.
- Дополнение документации {{Cookiecutter.module_slug}} с объяснением значения slug.
- Использование конфигурационных файлов **.yaml** неясно.
- Исправление вспомогательной ссылки для случаев конфликтов зависимостей.

Благодарность

SMTB
TARTU ÜLIKOOL · UNIVERSITAS TARTUENSIS · 1632
ZIMIN FOUNDATION
hhmi | Howard Hughes Medical Institute