

Announcements

- 1) Dinner together here at 6 PM tonight
- 2) See Dan's email about names on certificates

Todos Santos workshop - day 3

A metagenomics workflow



Computational Biology
Workshop

Todos Santos Center
May 9-12, 2022

Mark Stenglein, PhD
Associate Professor
Center for Vector-Borne Infectious Diseases
Department of Microbiology, Immunology, and Pathology
Mark.Stenglein@colostate.edu
StengleinLab.org

Math
undergrad →

7 years as a
computer
programmer →

PhD in mol.
biology /
virology →

Postdoc using
microarrays, NGS,
and bioinformatics →

Assoc.
Professor at
Colorado State

1999, Bangkok, Thai Airways test facility



Department of
**MICROBIOLOGY,
IMMUNOLOGY & PATHOLOGY**
COLLEGE OF VETERINARY MEDICINE & BIOMEDICAL SCIENCES

**CENTER FOR VECTOR-BORNE
INFECTIOUS DISEASES**

UCSF
University of California
San Francisco
advancing health worldwide™

Today, we will perform a metagenomic workflow

Get sequence data

Clean up data

Read mapping

Assembly

Metagenomic classification

Annotation

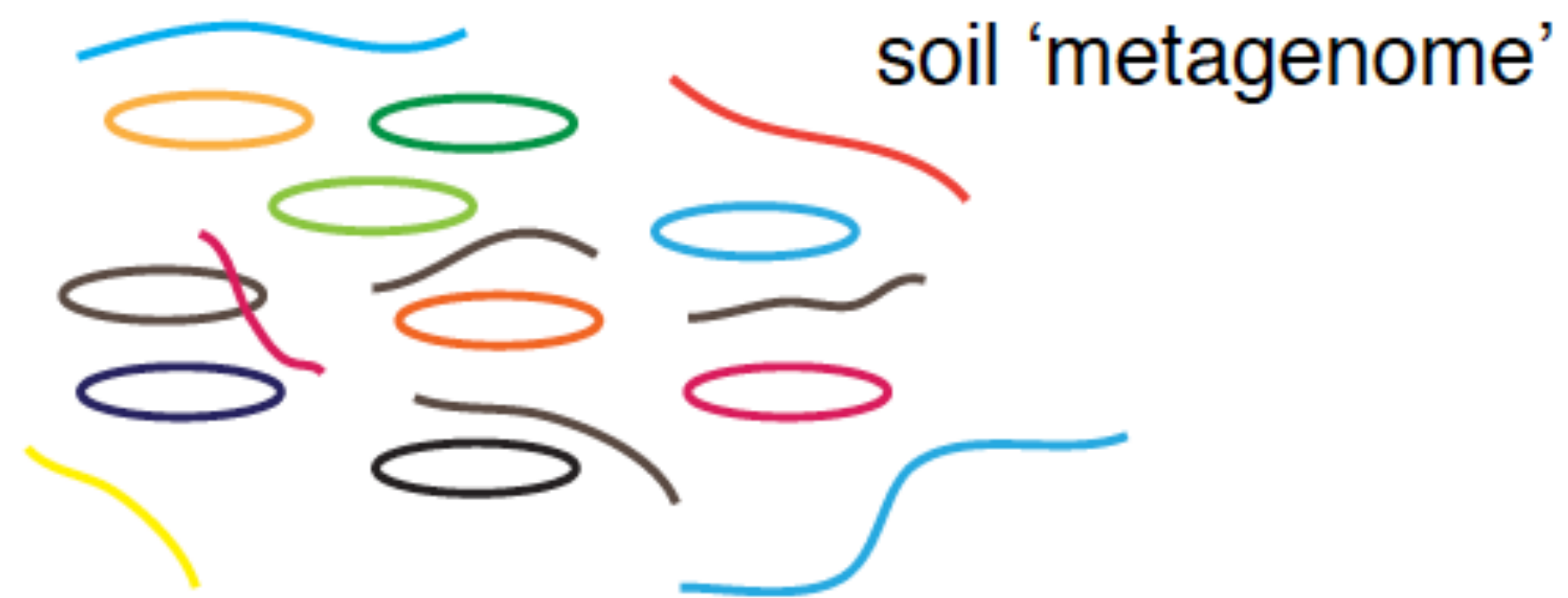
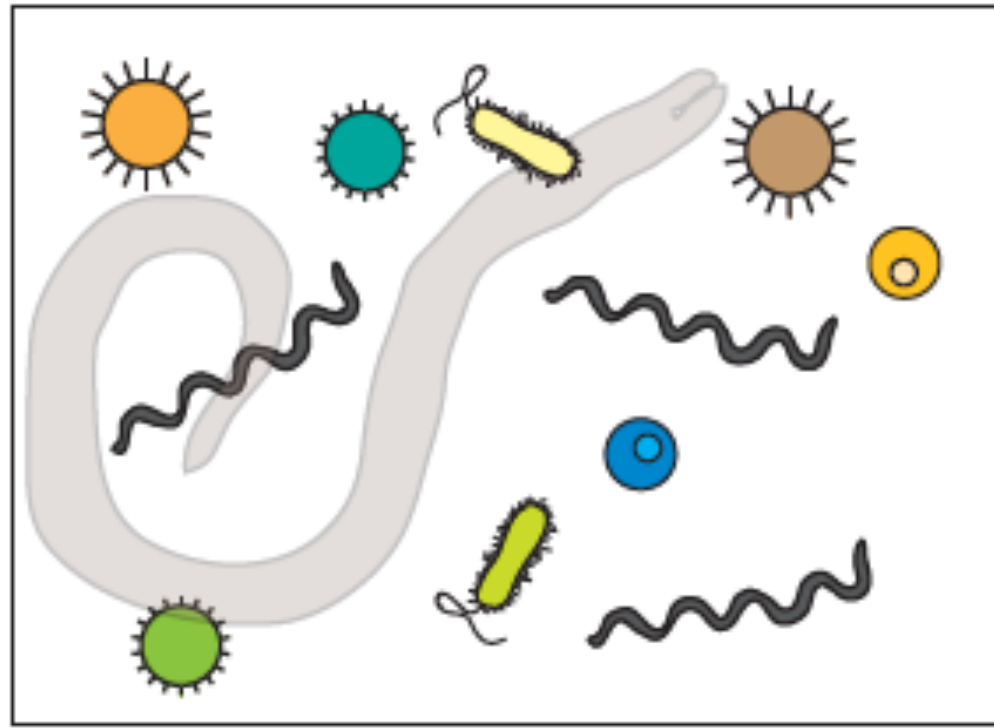


A typical metagenomics workflow

Metagenomics is simply the study of >1 genome

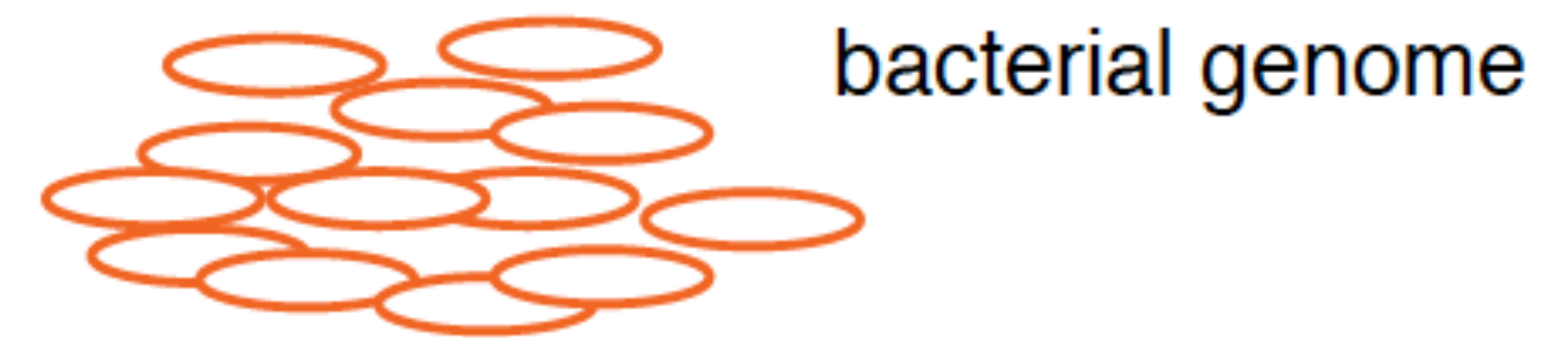
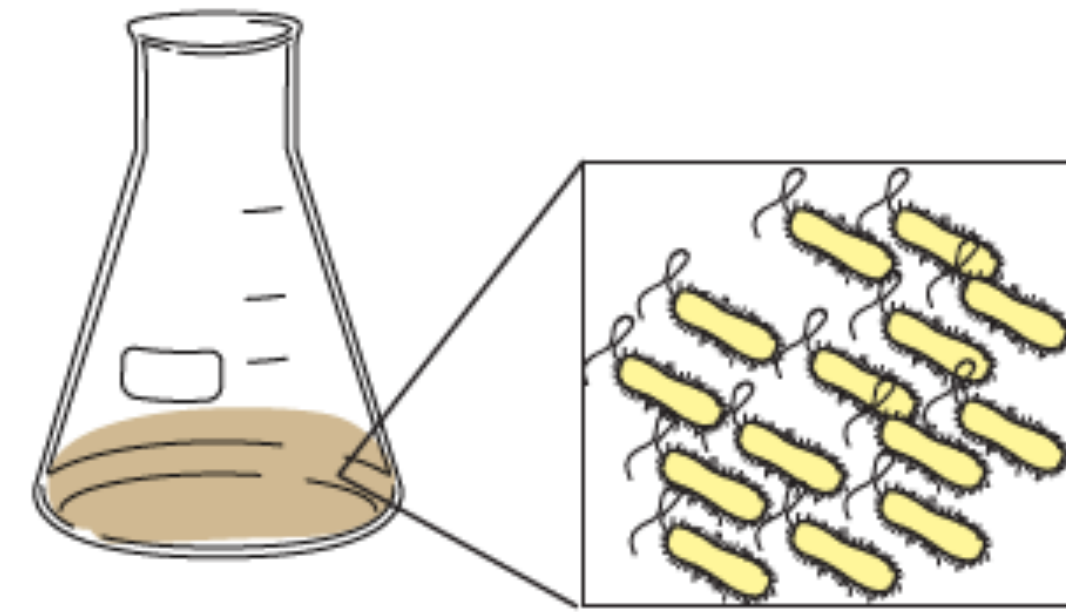
Many genomes

soil community

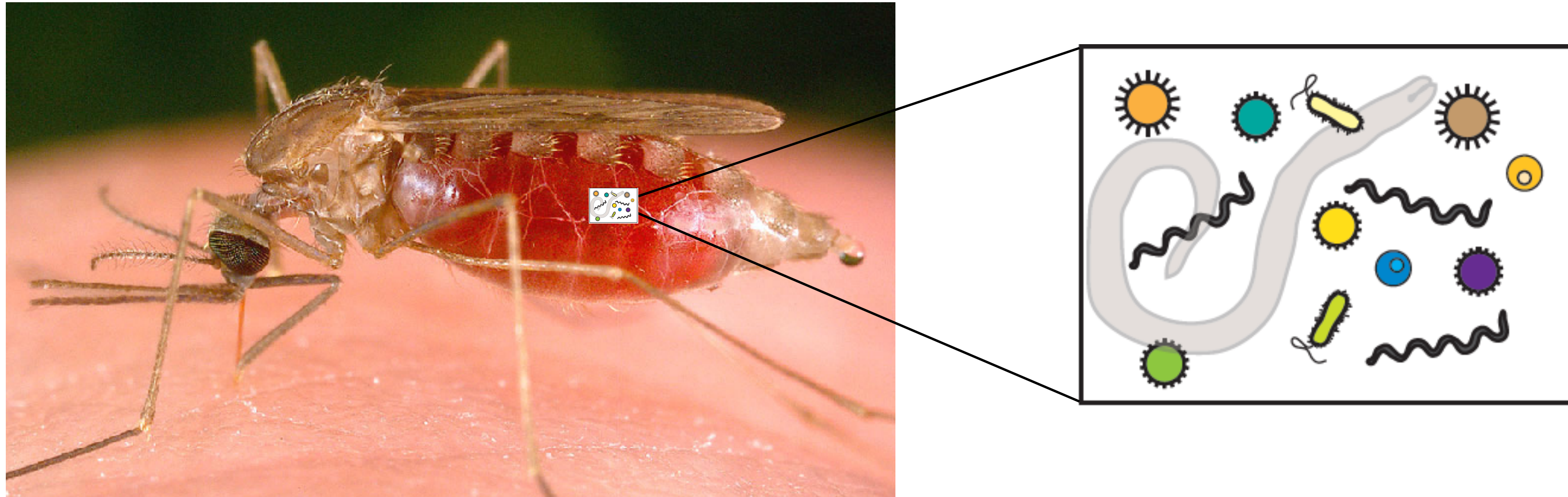


1 genome

bacterial isolate



If you sequence total nucleic acid from a multicellular organism,
you are doing metagenomics



Today, we will perform a metagenomic workflow using existing sequence data from a sick snake

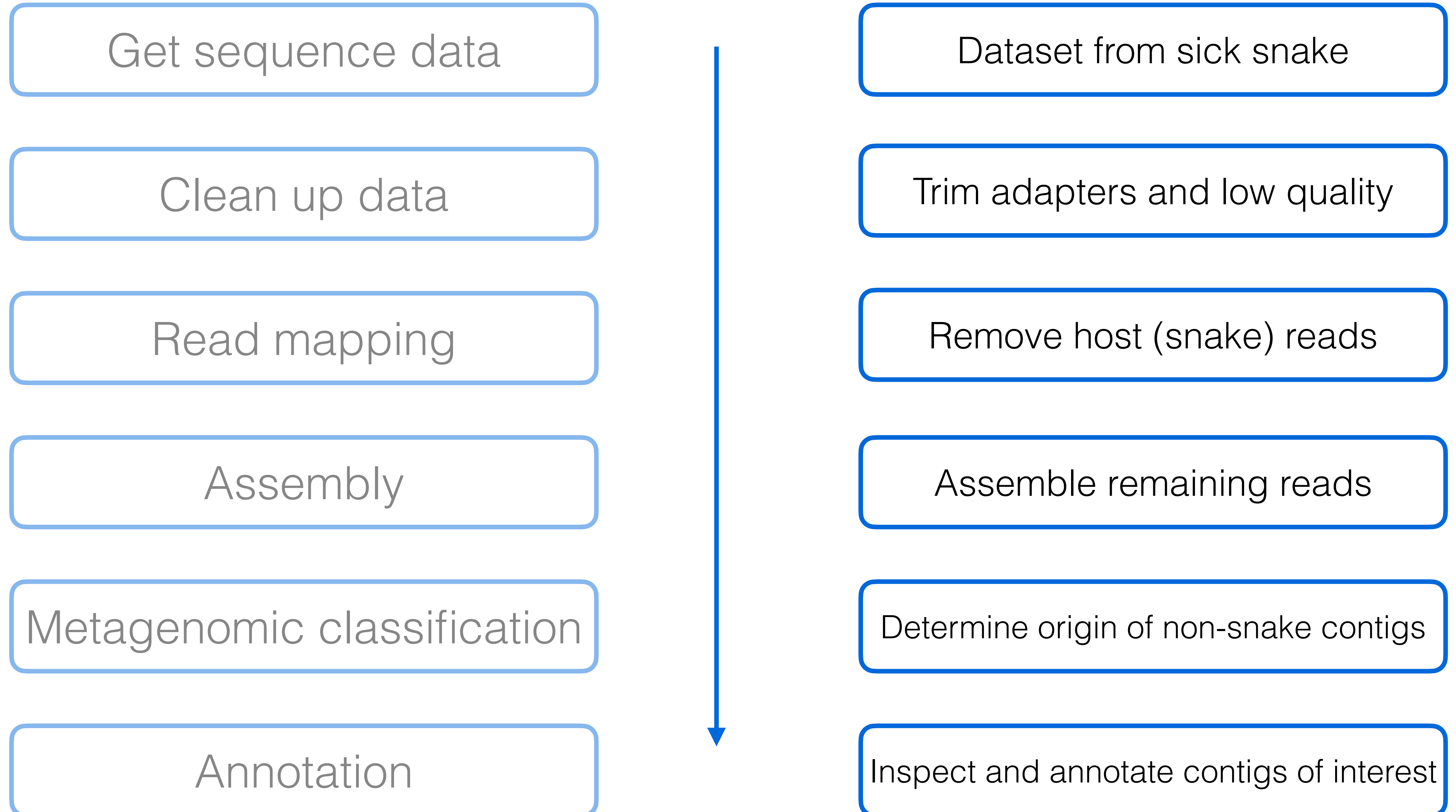
A python with inclusion body disease (IBD)



Biological questions:

- What is making this snake sick?
- What causes inclusion body disease?

A metagenomic workflow



Most of bioinformatics is based on sequence alignments

Sequence 1:

A C G A C A T C

| | | | | | |

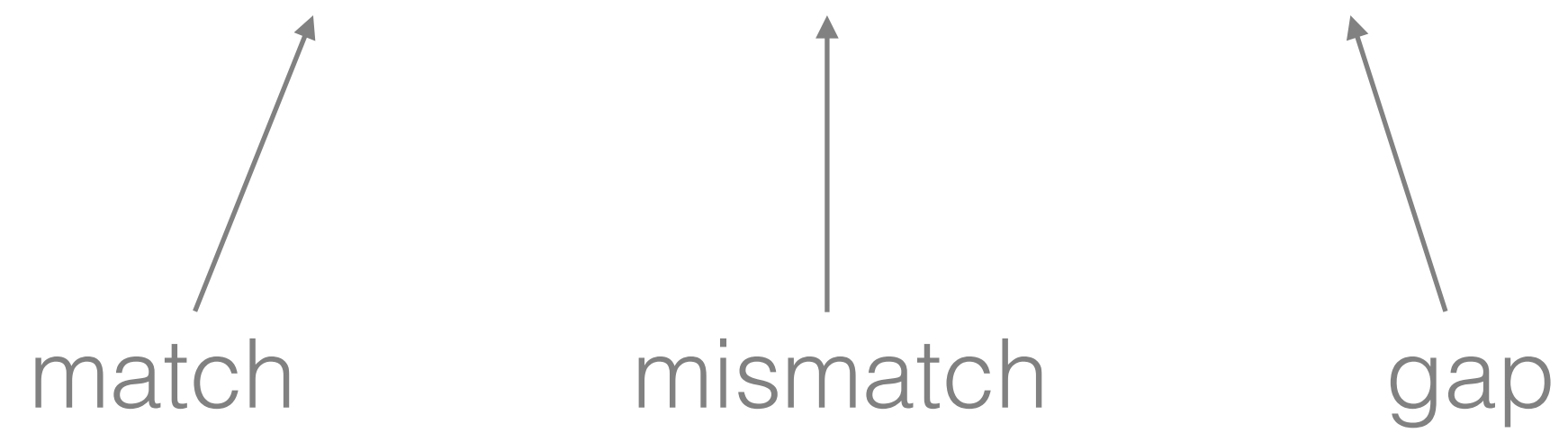
Sequence 2:

A C G C C A - C

match

mismatch

gap



Types of sequence alignment

Alignment type	Purpose	Example uses
Pairwise alignment	Align two sequences	How are two sequences different?
Multiple sequence alignment	Align > 2 sequences	Phylogenetic tree building
Mapping	Align reads to a reference sequence	RNA-seq Variant calling Read filtering
Assembly	Create a new reference sequence	Working with non-model organisms
BLAST search	Find the most closely related sequence in a database of sequences	Taxonomic assignment of contigs

We are switching servers: RNA -> Thoth

