

# The Condor Hatches in Africa

Throughput Computing 2023 (HTC23)

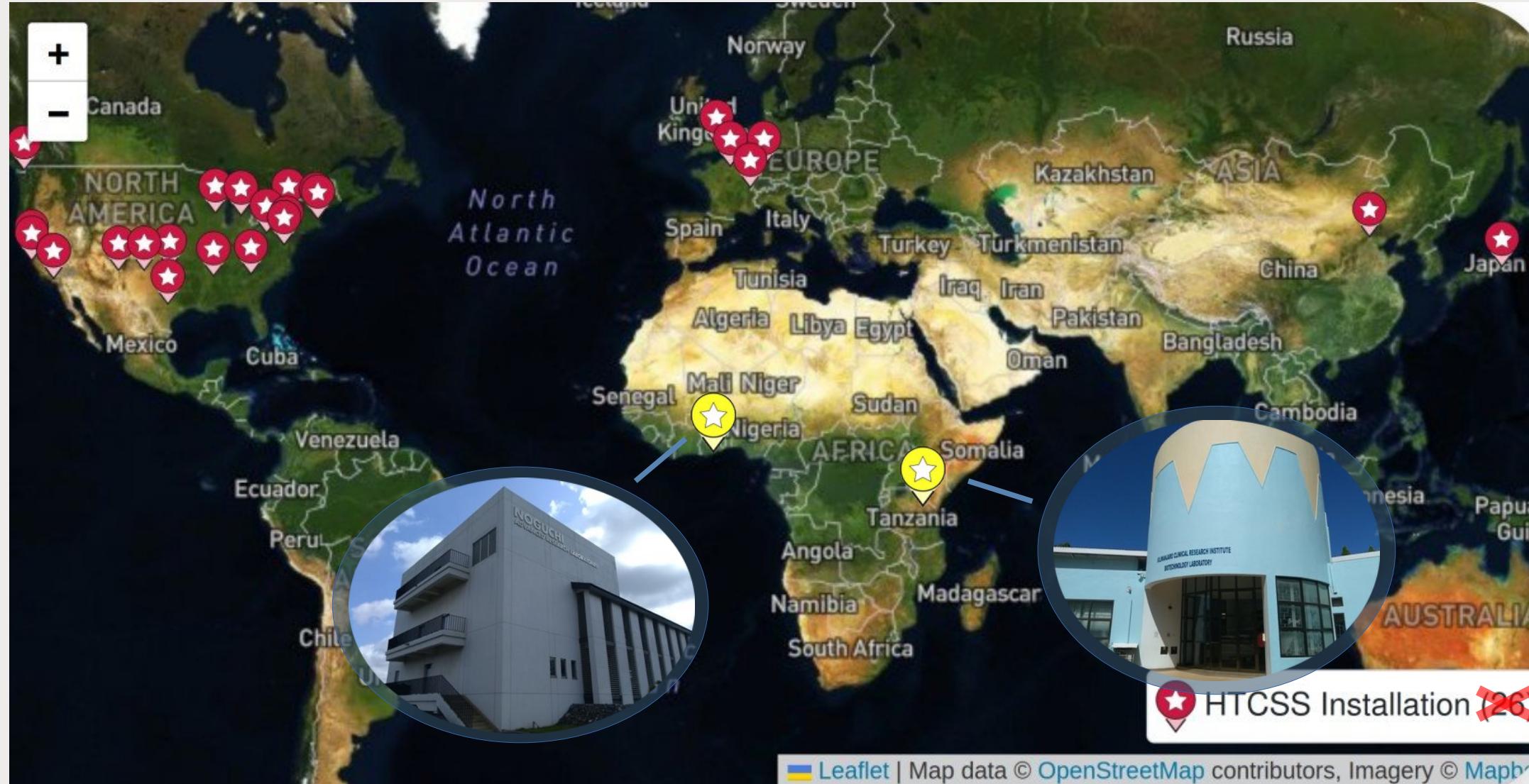


14 July 2023

Marco van Zwetselaar

[io@zwets.it](mailto:io@zwets.it) / The SeqAfrica Project





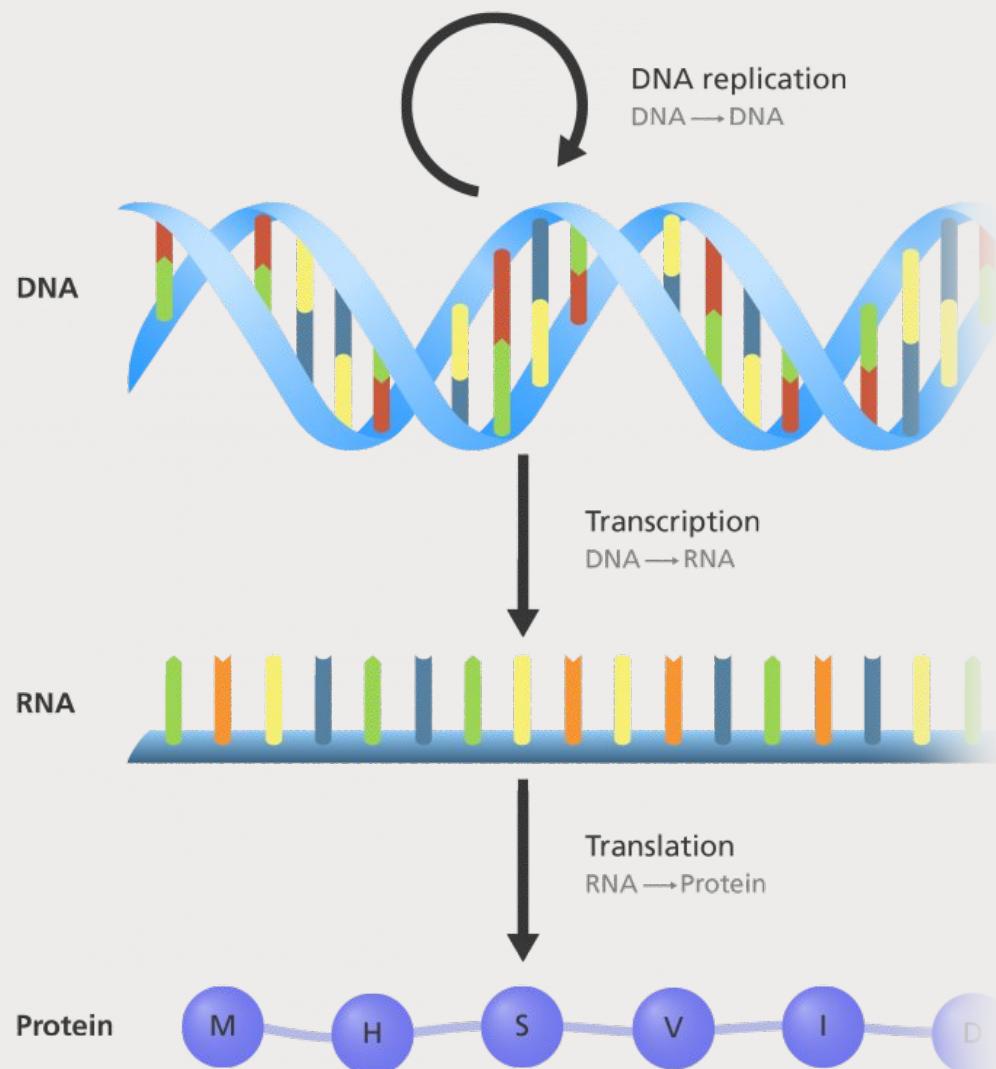


NOGUCHI MEMORIAL INSTITUTE  
FOR MEDICAL RESEARCH  
UNIVERSITY OF GHANA, LEGON

# SEQAFRICA

<https://antimicrobialresistance.dk/seqafrica.aspx>



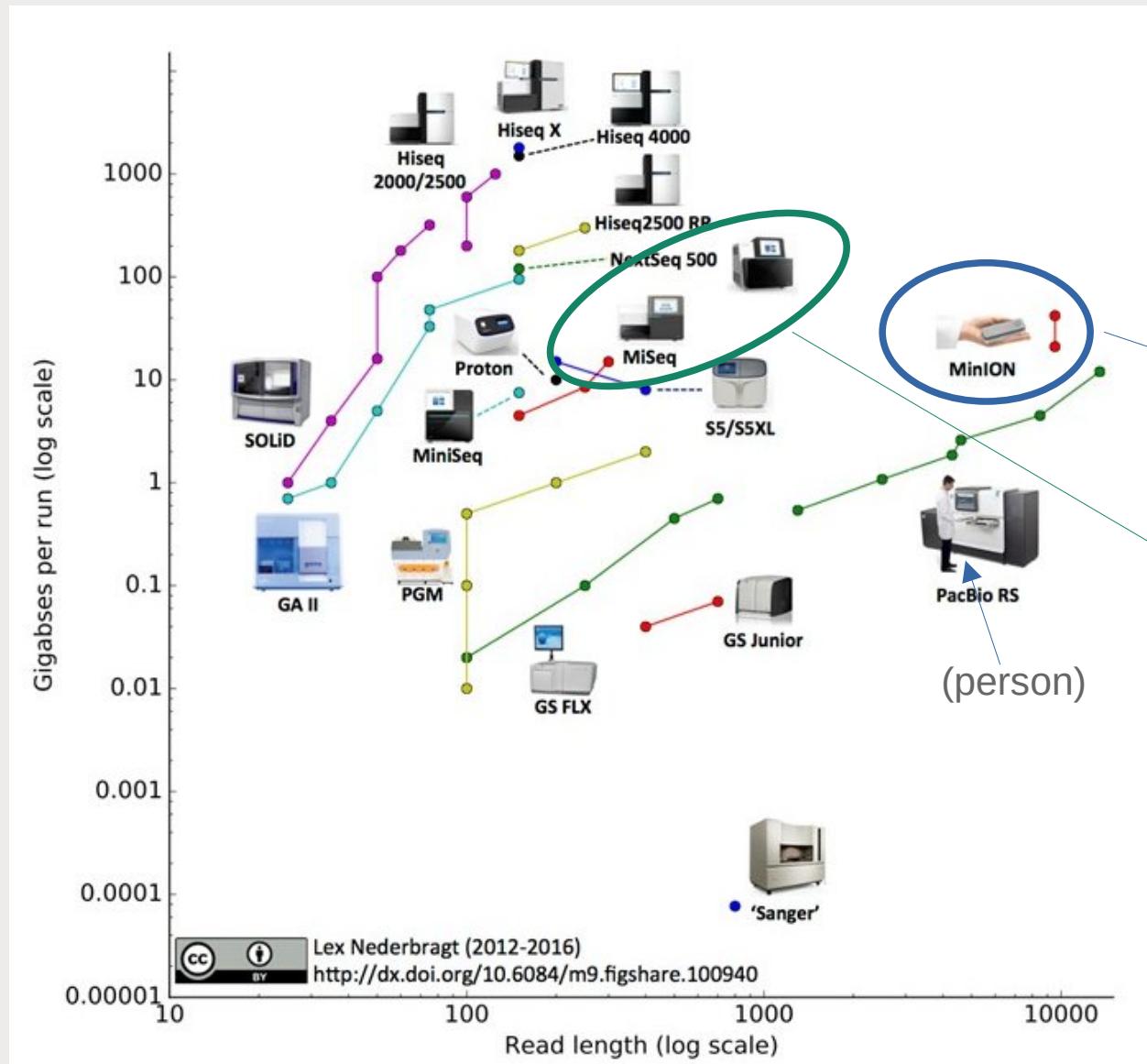


## Genomics in a Nutshell

- Adenine (A)
- Thymine (T)
- Cytosine (C)
- Guanine (G)
- Uracil (U)
- Amino acid

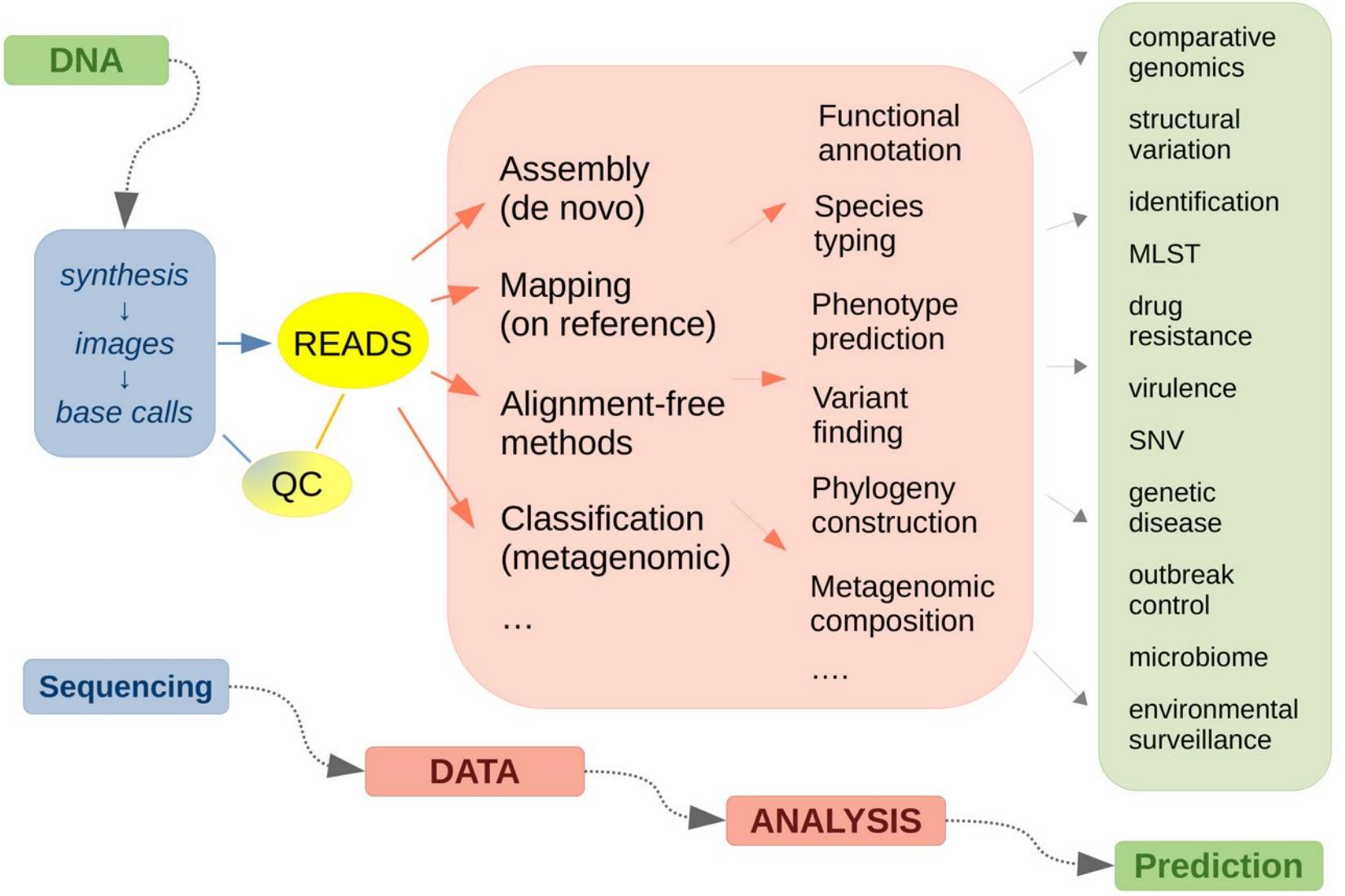
Image credit: Genome Research Limited (CC-BY)  
<https://www.yourgenome.org/facts/what-is-the-central-dogma/>

## DNA Sequencer Evolution

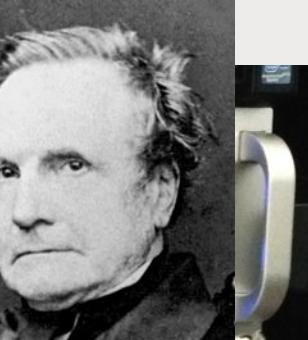
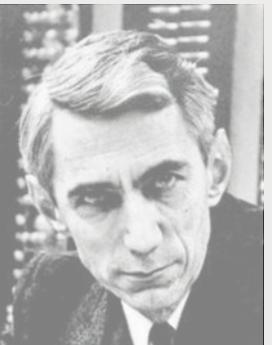
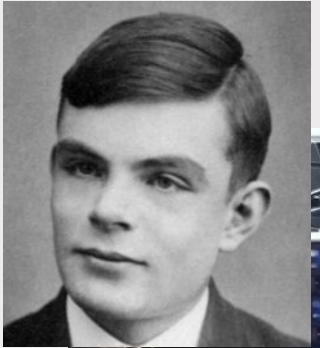




## Bioinformatics for Genomics

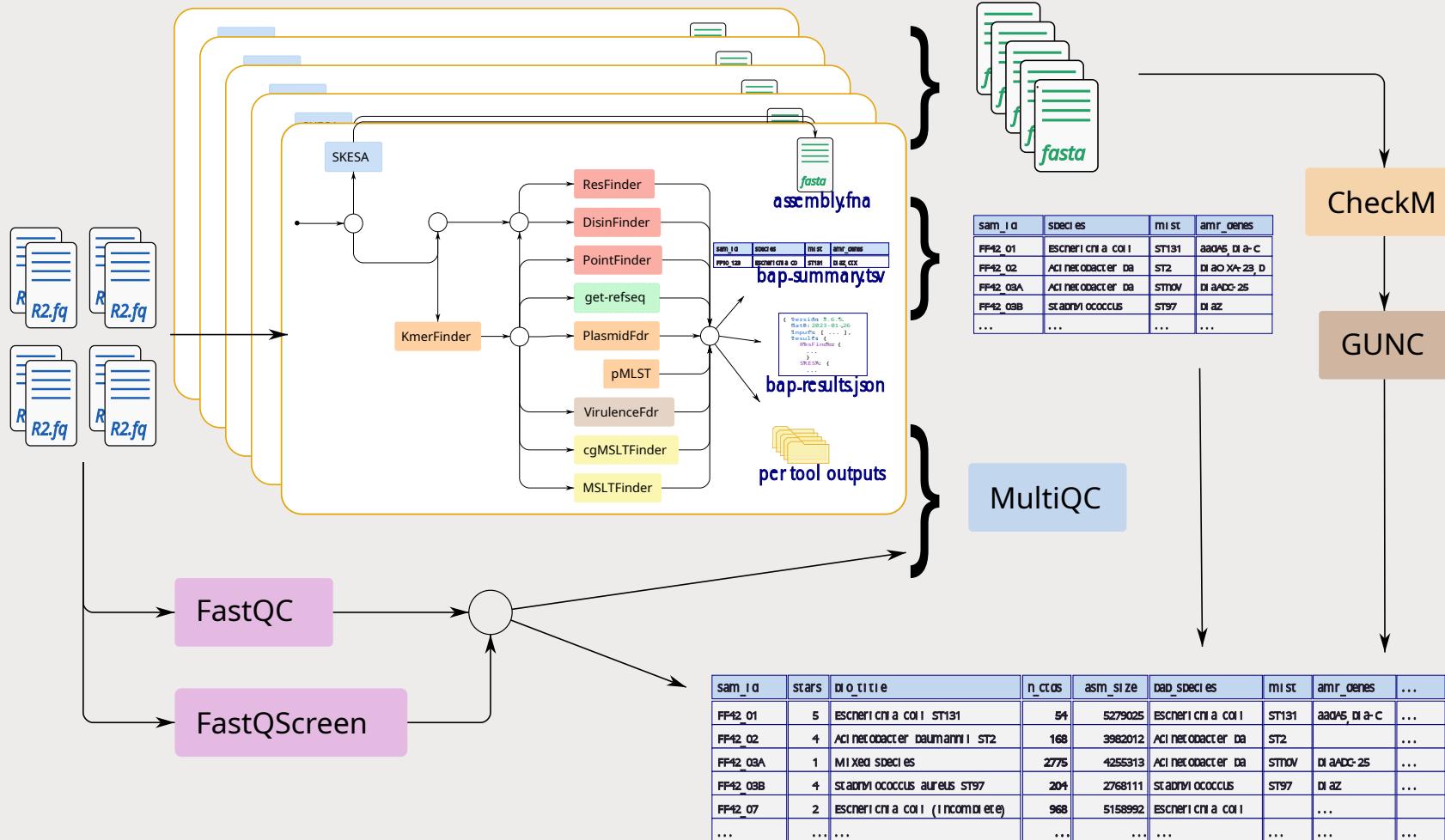


# How the Condor Landed on Kilimanjaro



Let's not upset Disney's lawyers

# Bacterial Analysis Pipeline (the “BAP”)



## Project BAP

- “Batching the BAP”
  - Process all isolates of a project
  - Collate & aggregate
  - Bundle for customer
- Add QC, flag issues
  - Genome completeness
  - Contamination & Mixes
  - Quality rating: 0-5 star
  - “Label” assembly title
- Utilise HPC facilities
  - HTCondor job batching
  - HTCondor workflow control
  - Include “heavyweight” tools

# Thank you



NOGUCHI MEMORIAL INSTITUTE  
FOR MEDICAL RESEARCH

UNIVERSITY OF GHANA , LEGON



M MOTT  
MACDONALD



This programme is being funded by the UK Department of Health and Social Care.  
The views expressed do not necessarily reflect the UK Government's official policies.