Whole-genome comparison among bacteria

A pan-genome clustering pipeline

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

FRIEDRICH-LOEFFLER-INSTITUT
F L I
Bundesforschungsinstitut für Tiergesundheit
Federal Research Institute for Animal Health

RNA
BIOINFORMATICS & HIGH-THROUGHPUT ANALYSIS

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Lisa-Marie Barf, Molecular Life Sciences
lisa-marie.barf@uni-jena.de
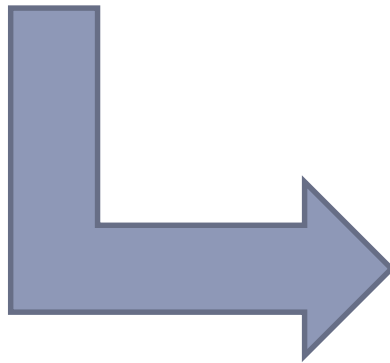
NGS facilitates analysis of bacteria genetics and evolution

Illumina

Roche 454

PacBio

MinION

- NGS leads to flood of sequence data
- Increasing number of genomes
- Analyze relation among organisms

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Motivation

Identify genes responsible for specific bacterial traits
e.g. host specificity, pathogenicity, virulence

Input

Genomes (Assemblies) in FASTA format

Output

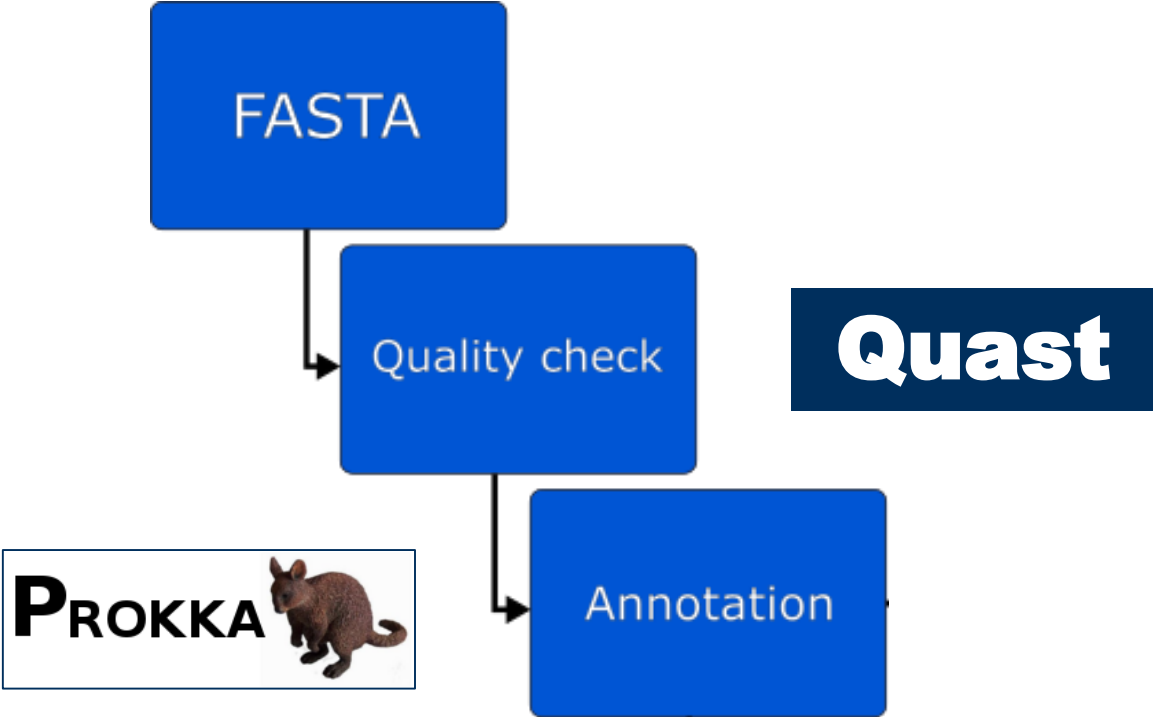Homology-based gene presence/absence table
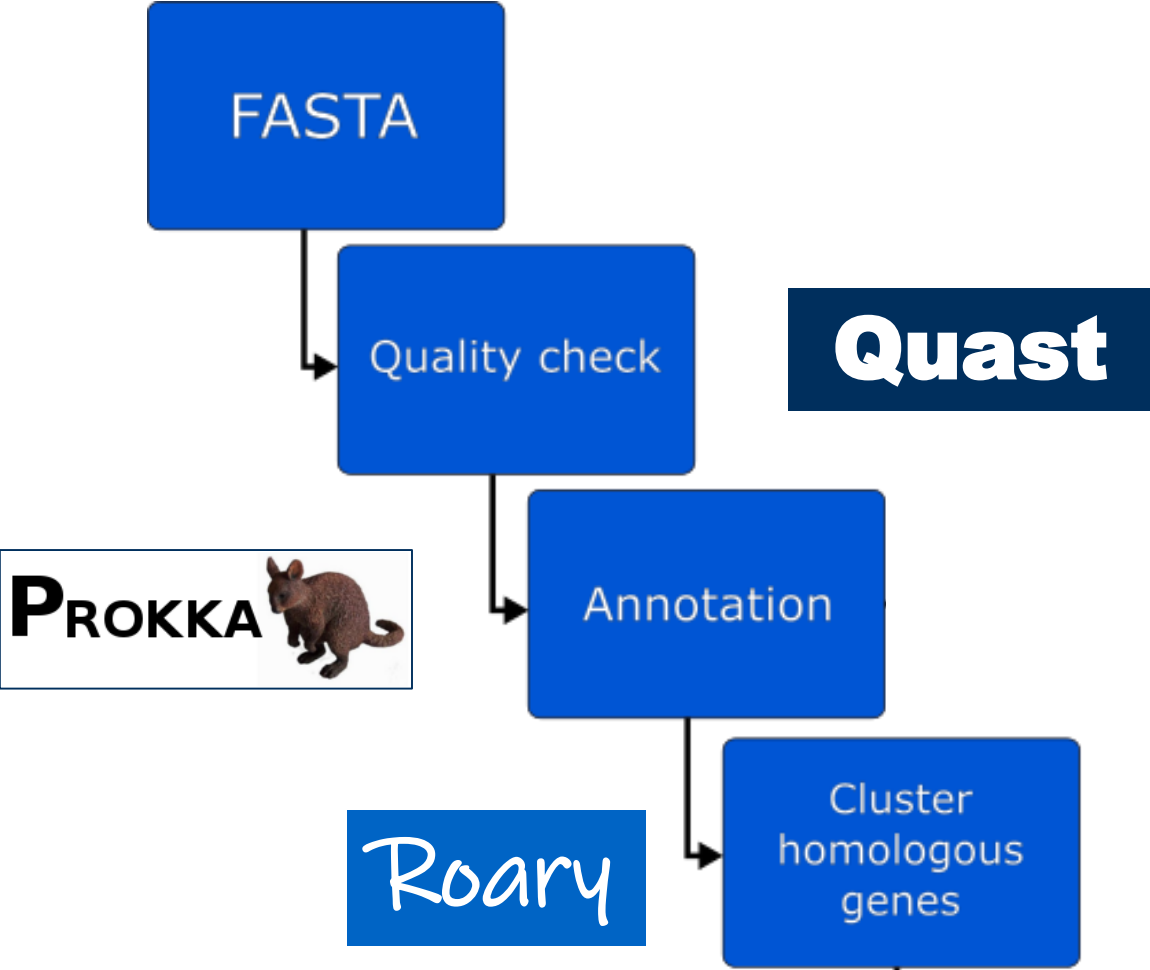
→ Extract subsets of genes

RIBAP – Roary-ILP Bacterial Annotation Pipeline

RIBAP – Roary-ILP Bacterial Annotation Pipeline
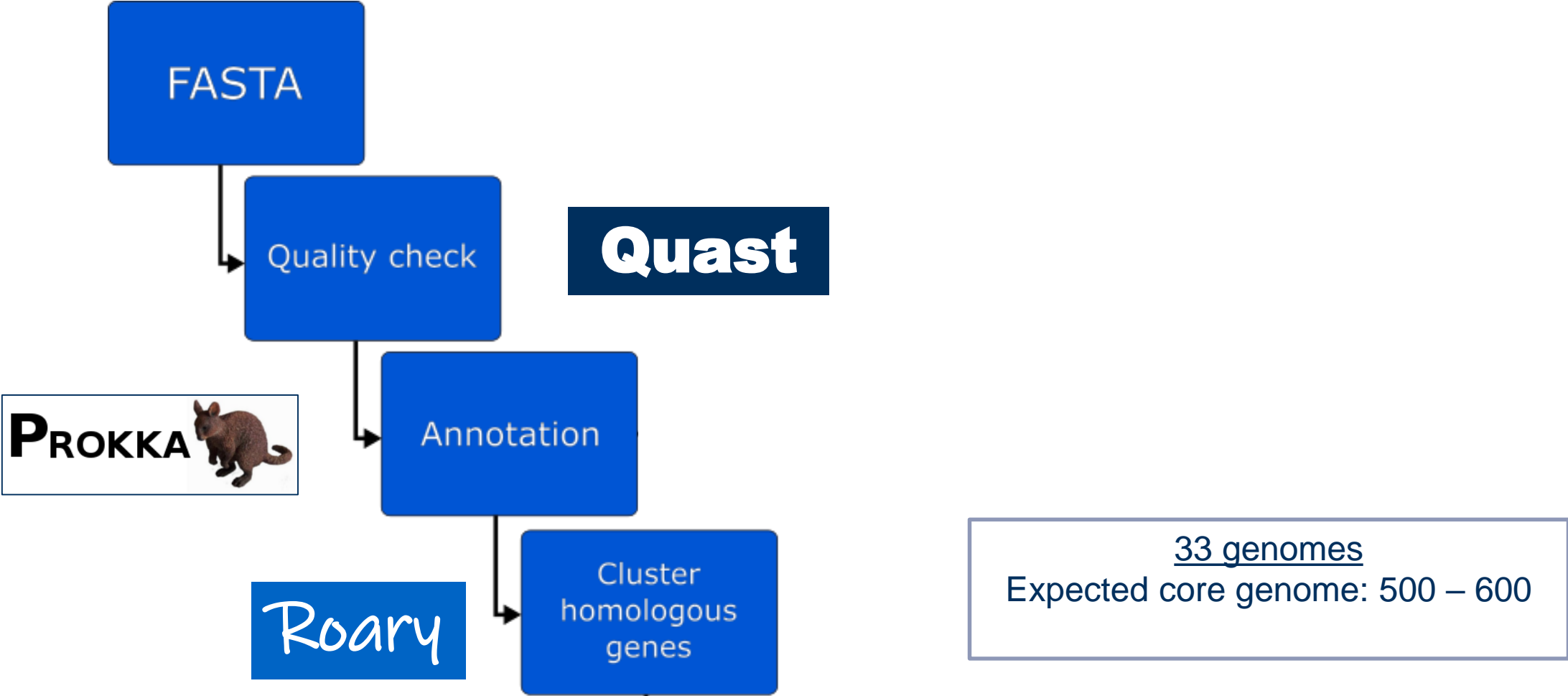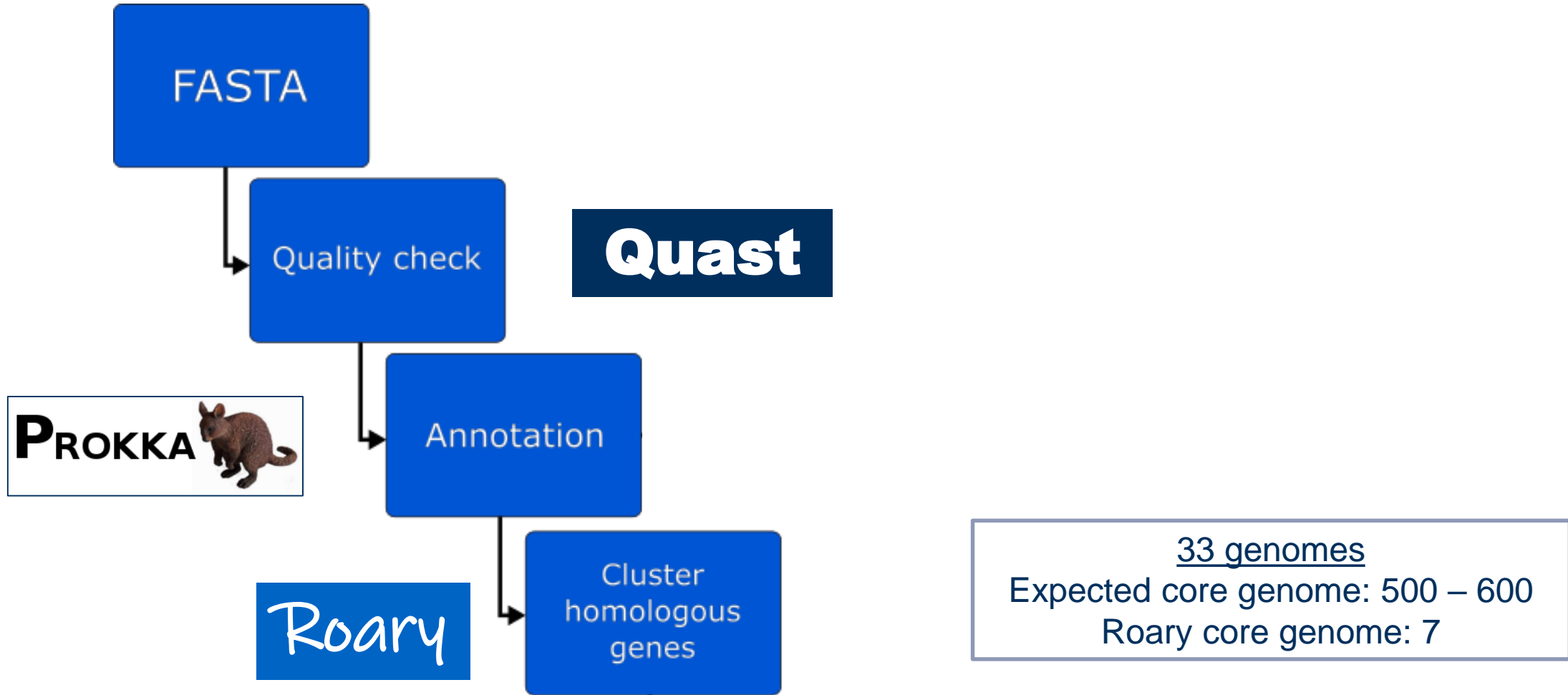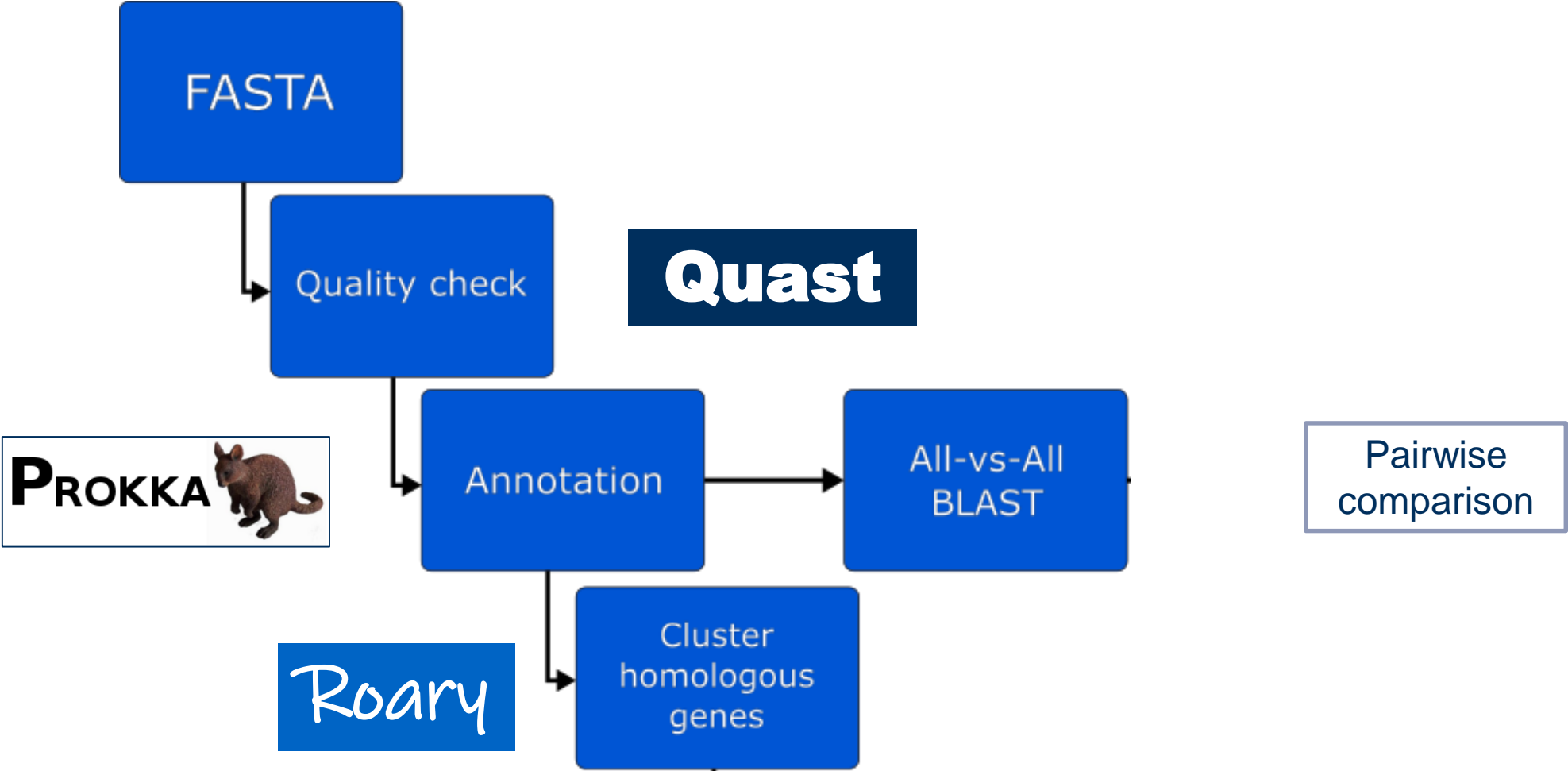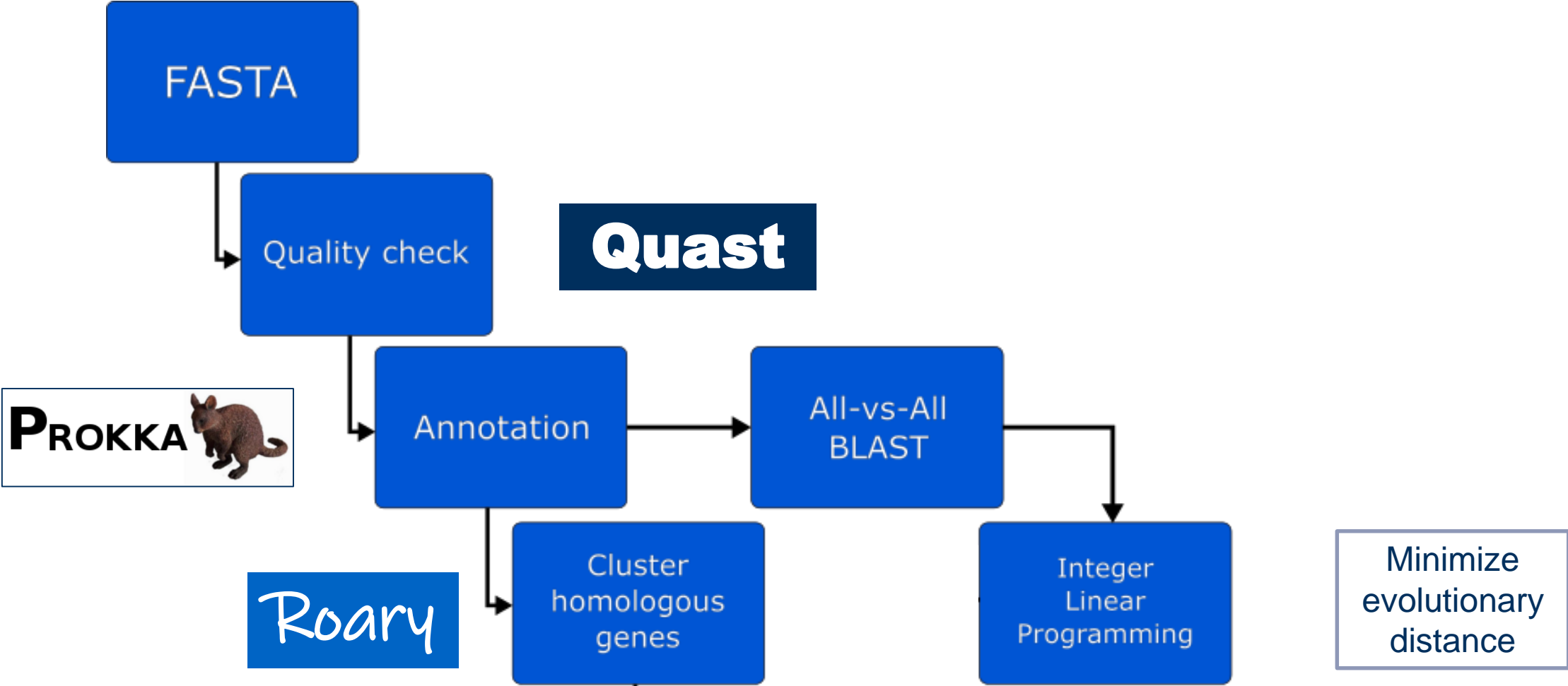
RIBAP – Roary-ILP Bacterial Annotation Pipeline

RIBAP – Roary-ILP Bacterial Annotation Pipeline

RIBAP – Roary-ILP Bacterial Annotation Pipeline

33 genomes
Expected core genome: 500 – 600

RIBAP – Roary-ILP Bacterial Annotation Pipeline

33 genomes
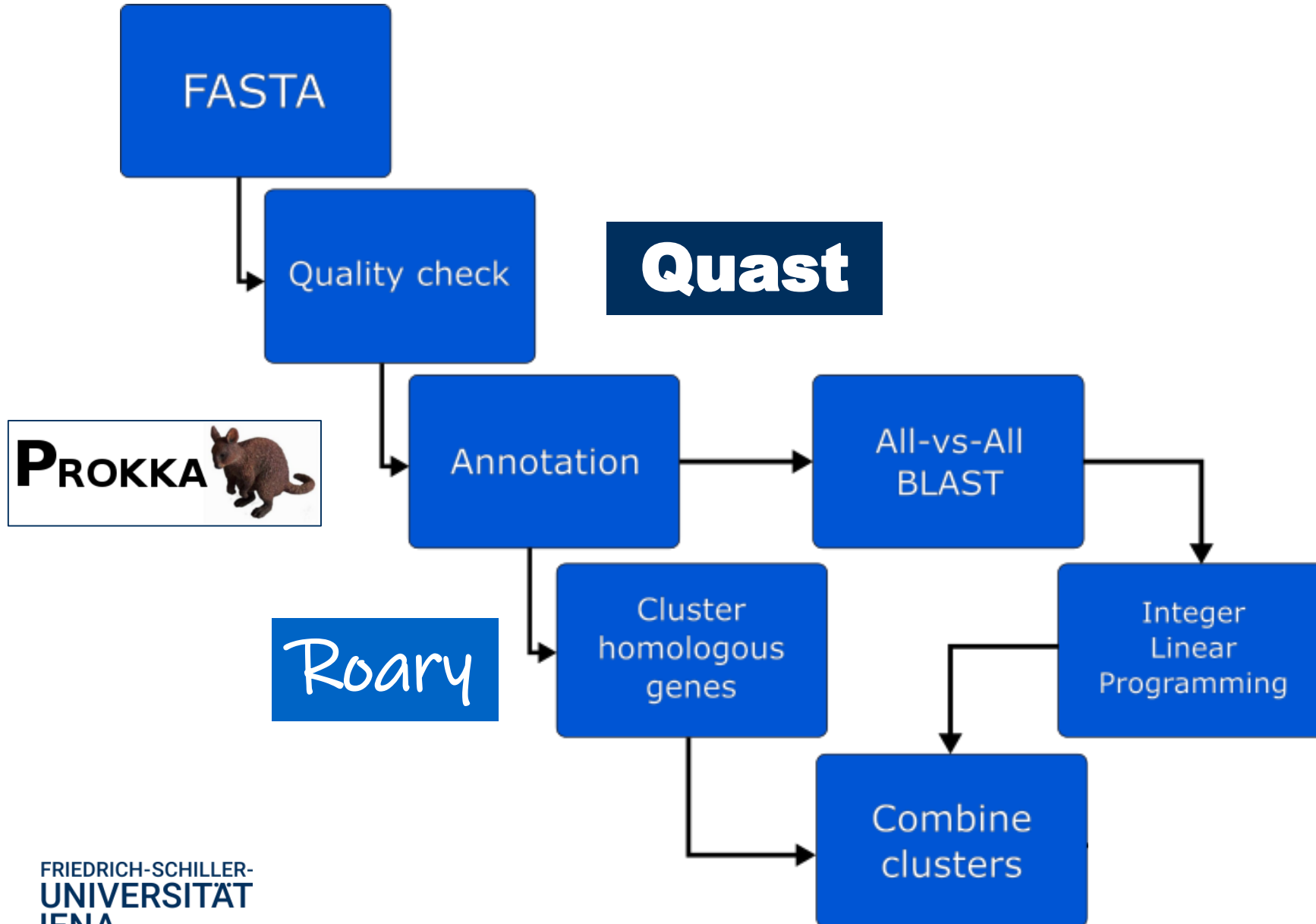Expected core genome: 500 – 600
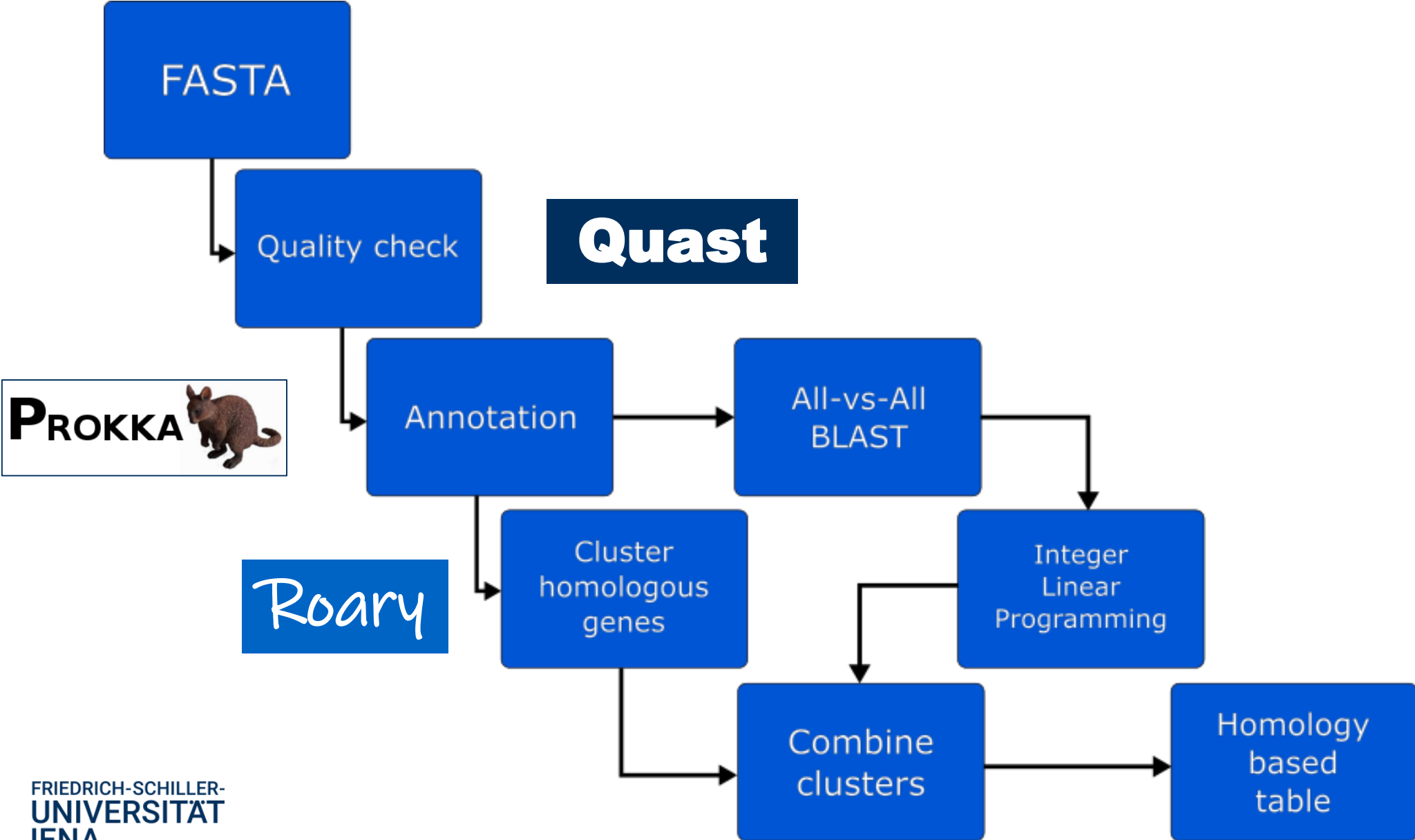Roary core genome: 7

RIBAP – Roary-ILP Bacterial Annotation Pipeline

RIBAP – Roary-ILP Bacterial Annotation Pipeline
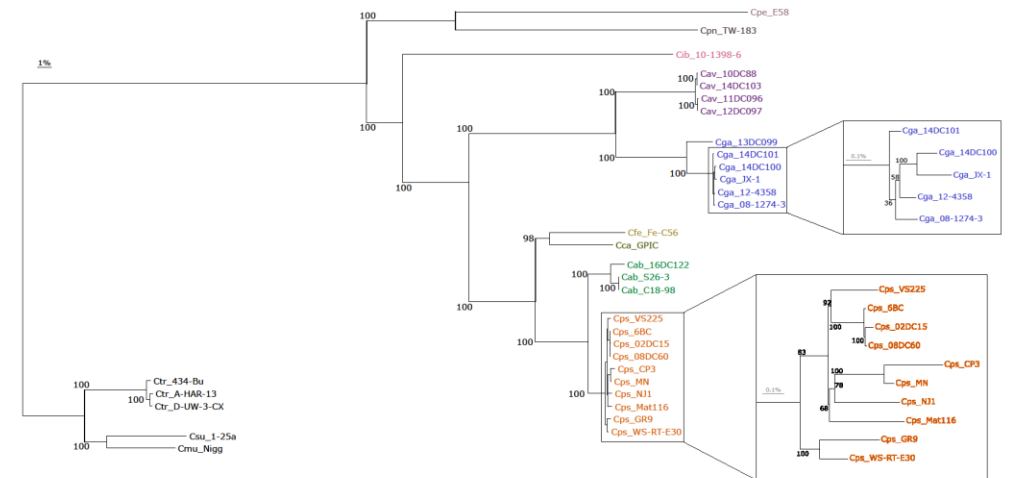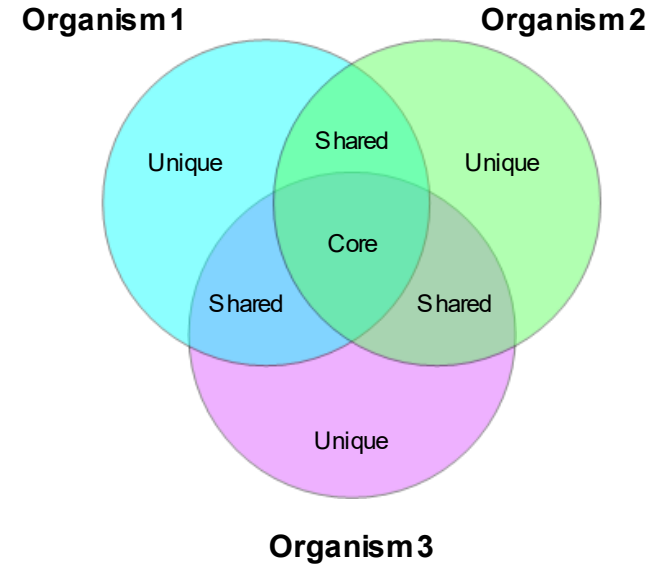
RIBAP – Roary-ILP Bacterial Annotation Pipeline

RIBAP – Roary-ILP Bacterial Annotation Pipeline

Next steps

- Expand annotation database

Next steps

- Expand annotation database
- Benchmarking

Next steps

- Expand annotation database
- Benchmarking
- Visualization webpage

Special thanks
Manja Marz
Konrad Sachse
Martin Hölzer
Kevin Lamkiewicz

Roary

FRIEDRICH-SCHILLER-
**UNIVERSITÄT
JENA**