Building a Reference Panel and Assessing Imputation Methods for Low-Coverage Sequencing in Black Soldier Fly *(Hermetia illucens)*

Peter Muchina

PhD student, JKUAT

pmuchina@qgg.au.dk

AARHUS UNIVERSITY

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

The Need for Genomic Resources

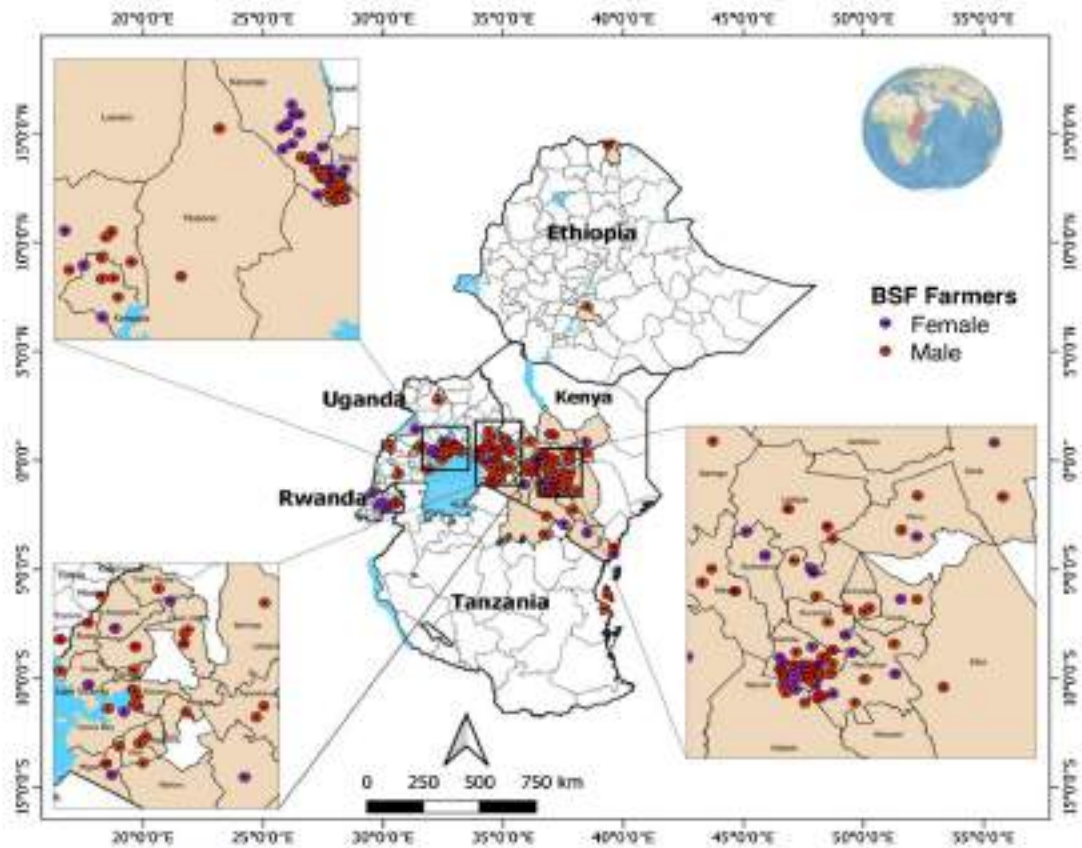
Importance of BSF:

- An alternative source of protein for animal feed.
- Organic waste bioconversion and environmental sustainability.

How Genomic Resources Can Help:

- Enhance understanding of BSF biology and population structure.

- Optimize desirable traits, e.g., growth rate, reproduction, and waste conversion efficiency through breeding programs.



BSF Farming in East Africa: *Tanga et al., 2021*

Limitation

- Limited existing BSF genomic data
  - High sequencing cost
- Lack of standardized resources for BSF
  - Computational challenges

Alternative

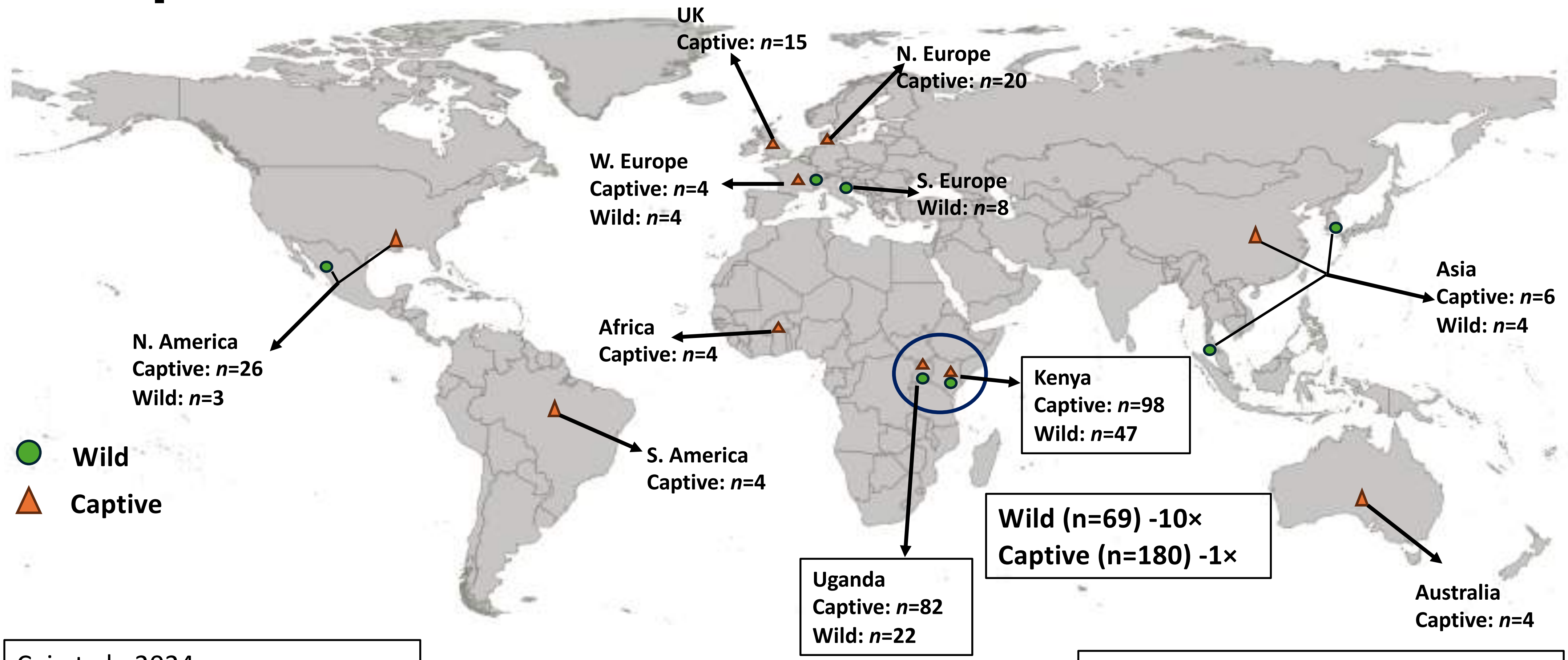
- Low-coverage sequencing followed by imputation using a reference panel

lcWGS: < 5×, few variants | **Reference panel** | **Imputed genomic data**

Objectives

- Build a reliable reference panel
- Test and optimize imputation tools that can handle low-coverage whole genome sequence (lcWGS) data
- Provide guidelines for implementing lcWGS imputation pipelines in BSF

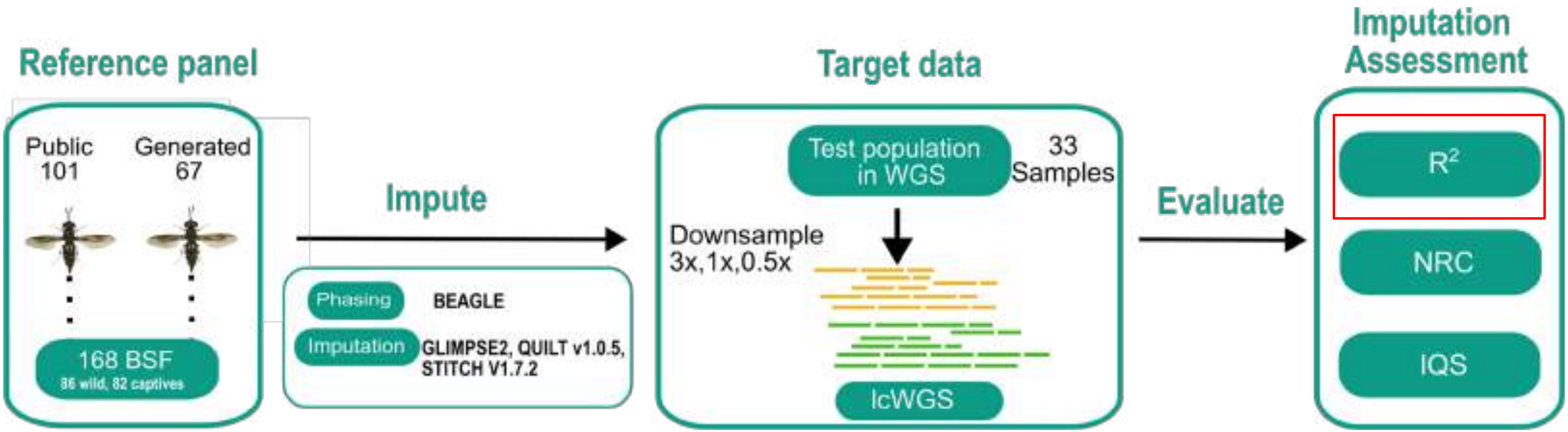
Sample Collection



Cai et al., 2024
Generalovic et al., 2021, 2023
Kaya et al., 2021
Vicoso & Bachtrog, 2013

Total: 351 WGS samples
171 ≥ 10× - Reference
180 = 1× - Target

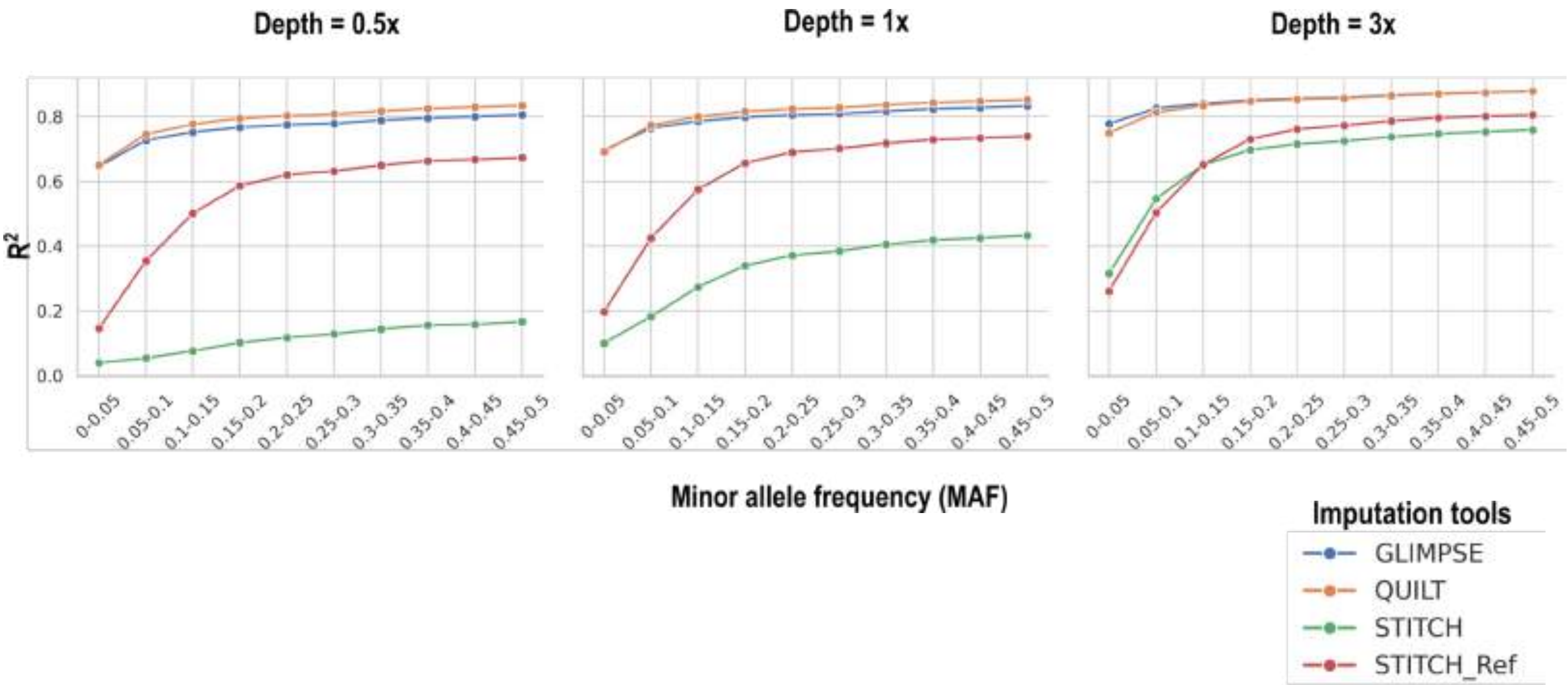
Study Design



~ 29 Million SNPs

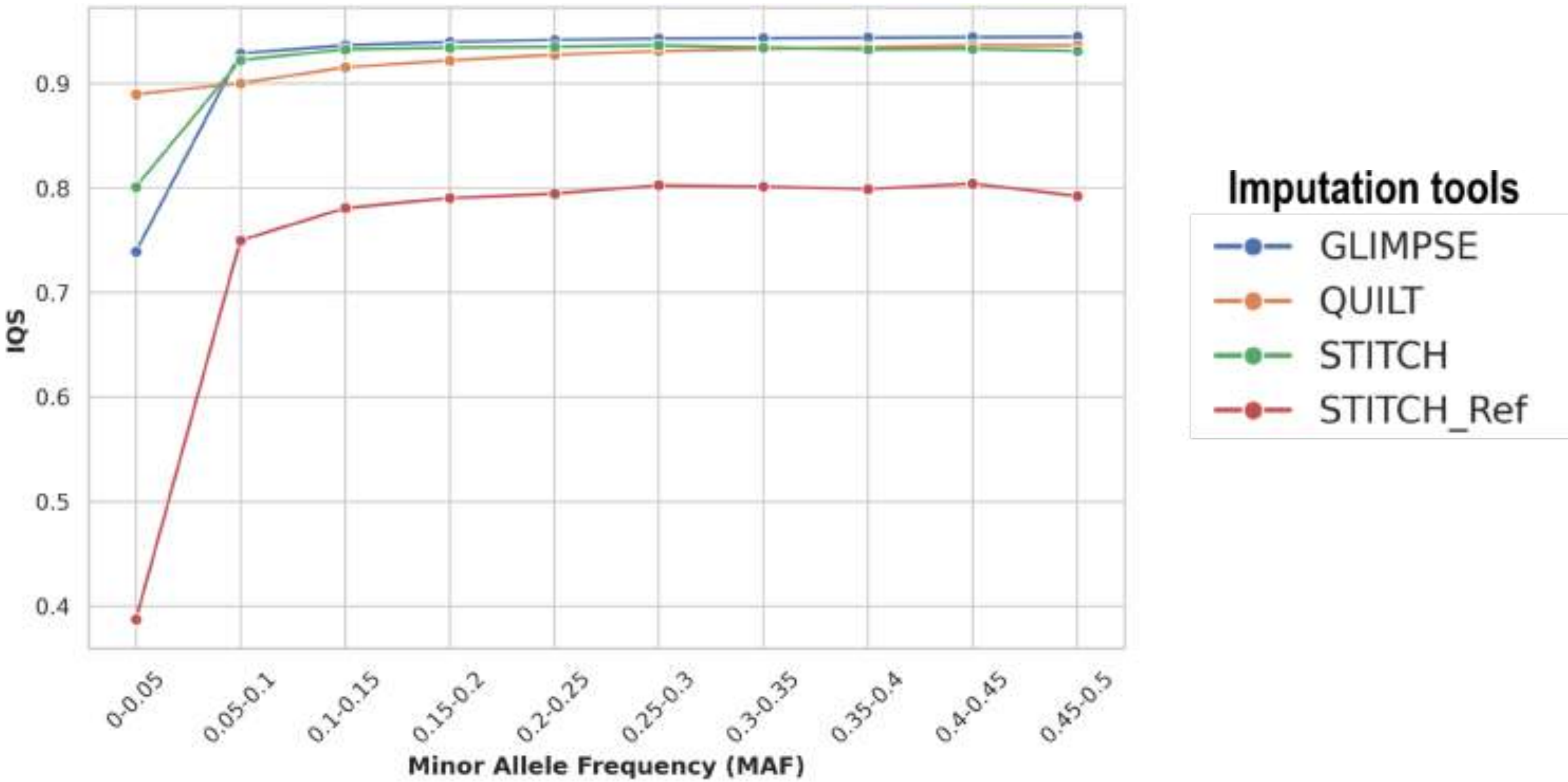
- **R^2** indicates the correlation between the true genotype dosages (actual values) and the imputed genotype dosages.
- A higher R^2 (closer to 1) means better accuracy.

Test Dataset (n=33), ~ 29 Million SNPs

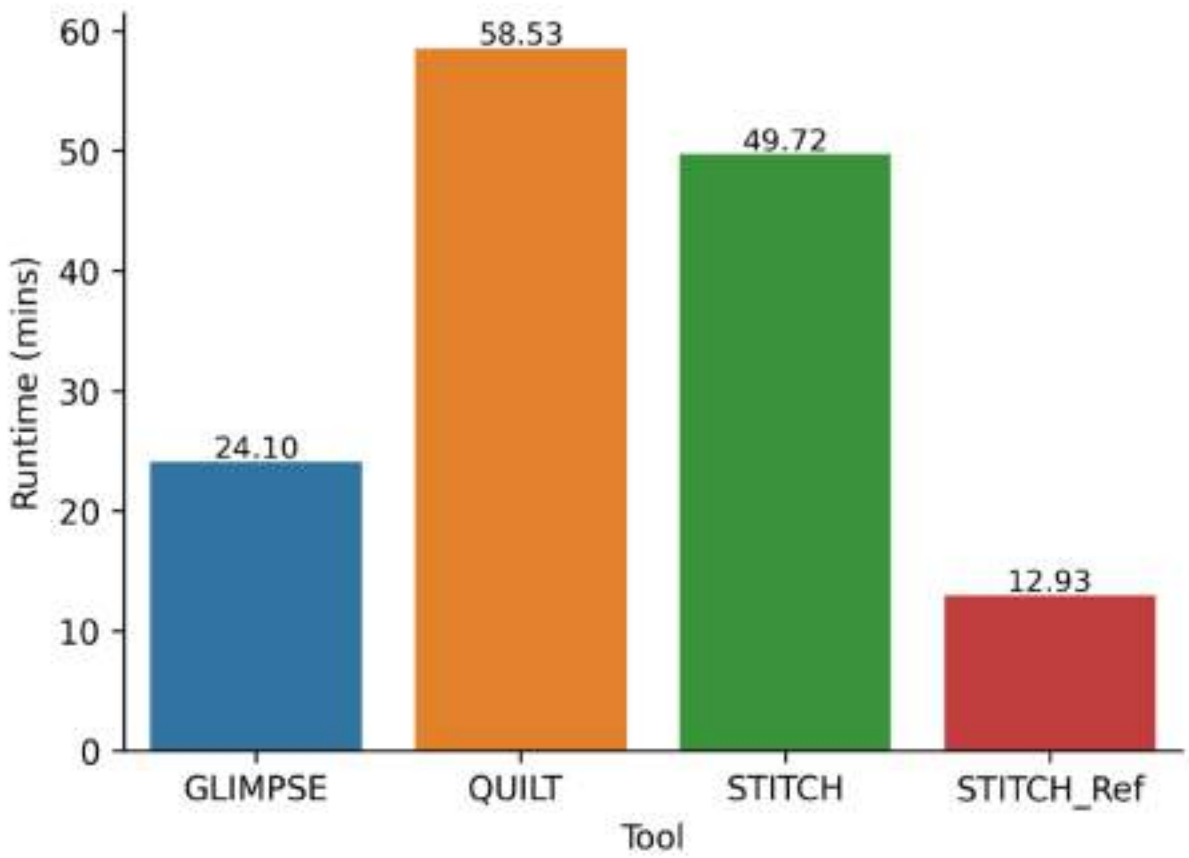


Imputation Quality Score (IQS)

Target Dataset - (n=180,1×),~ 29 Million SNPs



Runtime – Chr1, ~ 6.9 Million SNPs



Linux cluster (Intel Xeon & AMD EPYC 9374F CPUs) with 35 computation nodes

Conclusion

- Built a reliable and robust reference panel for imputation.
- Tested and optimized the performance of low-coverage imputation tools.
- **QUILT** outperformed other tools in accuracy and reliability.
- Expanded the genomic dataset for future research.
- Code and data will be made publicly available to support reproducibility.

Acknowledgment

INSECT-IMP

Supervisors:

Prof. Johnson Kinyua
Dr. Fathiya Khamis
Dr. Chrysantus Tanga
Dr. Grum Gebreyesus
Dr. Zexi Cai
Dr. Goutam Sahana

APU Lab *(ICIPE):*

Levi Ombura
Emmanuel Kiprono

Field Survey

Prof. Dorothy Nakimbugwe (UG)
Dr. Geoffrey Ssepuuya (UG)
Mark Ojungu (UG)
Kentosse Gutu (KE)
Eric Rachami (KE)
Dr. Rawlynce Bett (KE)
Phillip Tare (KE)
Abigael Nuna (KE)

FLYgene

INSECT PRO
MARULA PROTEEN LIMITED
AARHUS UNIVERSITY
UNIVERSITY OF NAIROBI
icipe
QGG