High quality **MinION** and **Flongle** long-read nanopore genome assemblies of ***Mycoplasma bovis*** using **taxon-specific training** of the Bonito basecaller


@methenickname

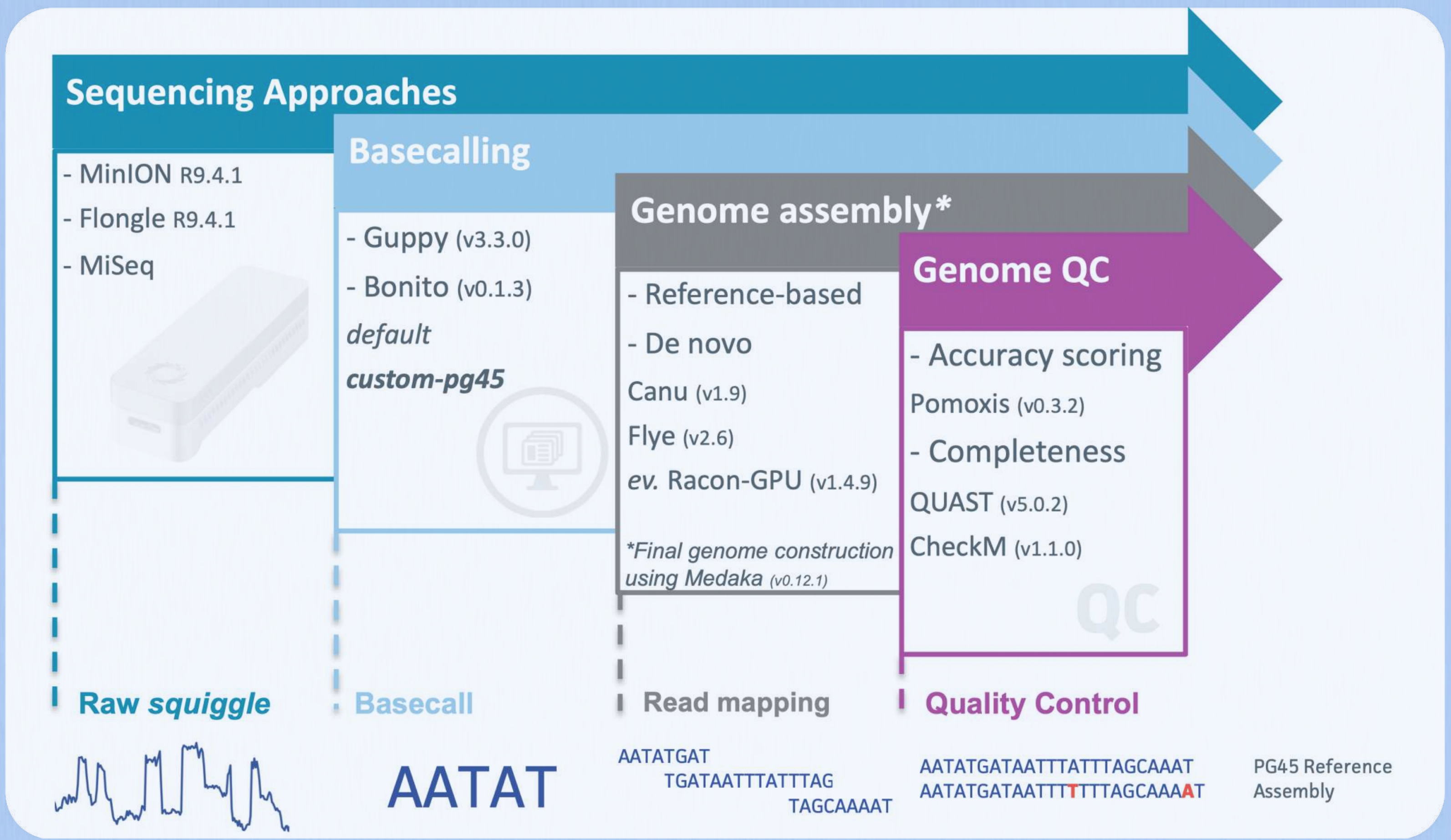
Nick Vereecke[a,d], **Jade Bokma**[b], **Freddy Haesebrouck**[c], **Hans Nauwynck**[a,d], **Filip Boyen**[c], **Bart Pardon**[b], **Sebastiaan Theuns**[a,d]

[a] Ghent University, Faculty of Veterinary Medicine, Department of Virology, Parasitology and Immunology, [b] Department of Large Animal Internal Medicine, [c] Department of Pathology, Bacteriology and Avian diseases, Salisburylaan 133, 9820 Merelbeke, Belgium; [d] PathoSense, Merelbeke, Belgium

Introduction

Implementation of Third-Generation Sequencing for WGS all-in-one **diagnostics in human and veterinary medicine** requires rapid and accurate generation of consensus genomes. Nevertheless, obtaining high quality genomes from non-standard organisms is still not guaranteed. This is the case for *Mycoplasma bovis*, an important respiratory pathogen in cattle. Current complete diagnostics require molecular and time-consuming culture-based approaches. Implementation of cheap and quick long read *all-in-one* WGS approaches are only feasible if increased genome accuracies and completeness can be obtained. Here, a **taxon-specific custom-trained Bonito v.0.1.3 model (*custom-pg45*)** was implemented in various WGS assembly bioinformatics pipelines.

Workflow


Sequencing Approaches
- MinION R9.4.1
- Flongle R9.4.1
- MiSeq

Basecalling
- Guppy (v3.3.0)
- Bonito (v0.1.3)
default
custom-pg45

Genome assembly*
- Reference-based
- De novo
Canu (v1.9)
Flye (v2.6)
ev. Racon-GPU (v1.4.9)

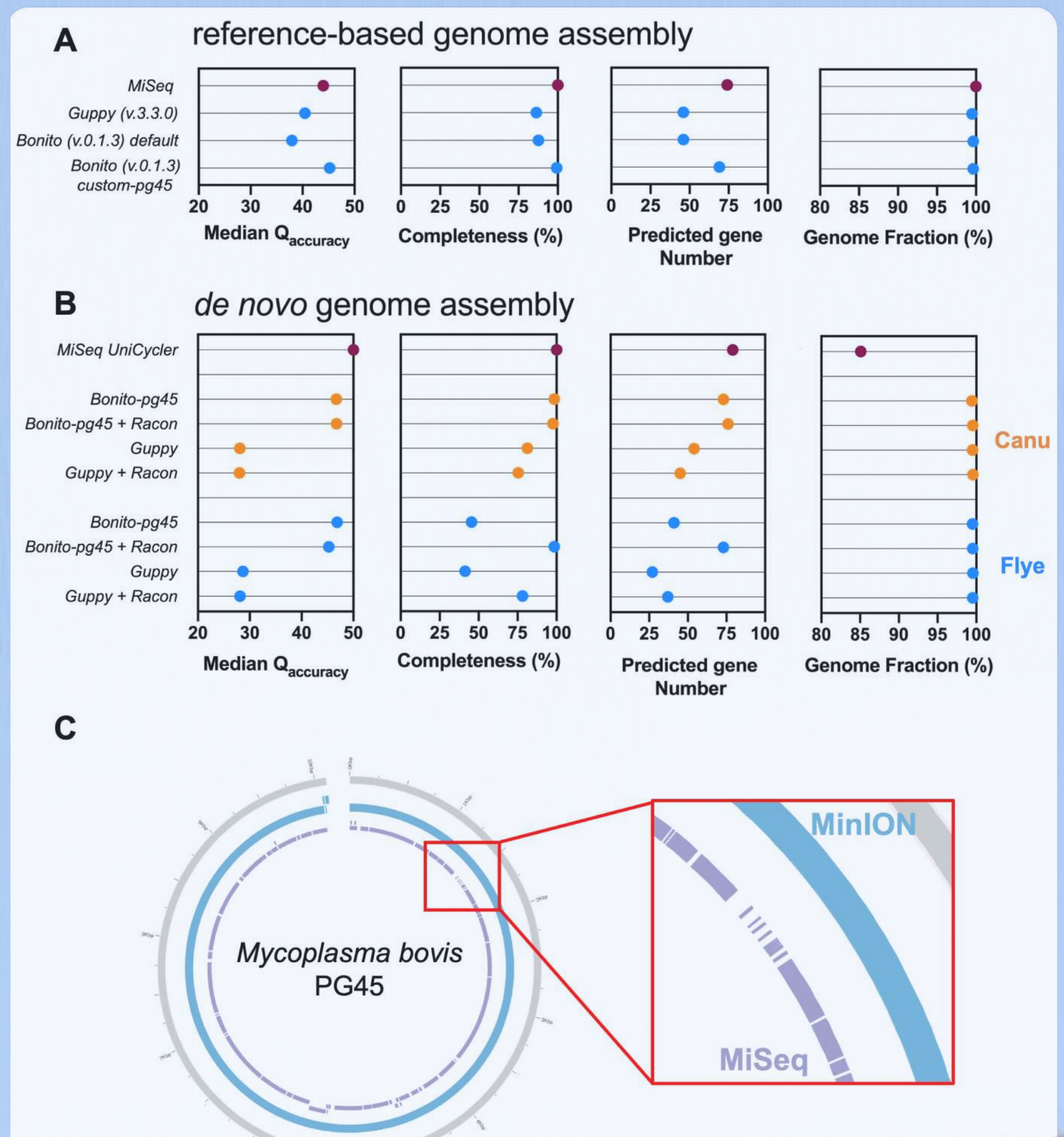
**Final genome construction using Medaka (v0.12.1)*

Genome QC
- Accuracy scoring
Pomoxis (v0.3.2)
- Completeness
QUAST (v5.0.2)
CheckM (v1.1.0)

Results

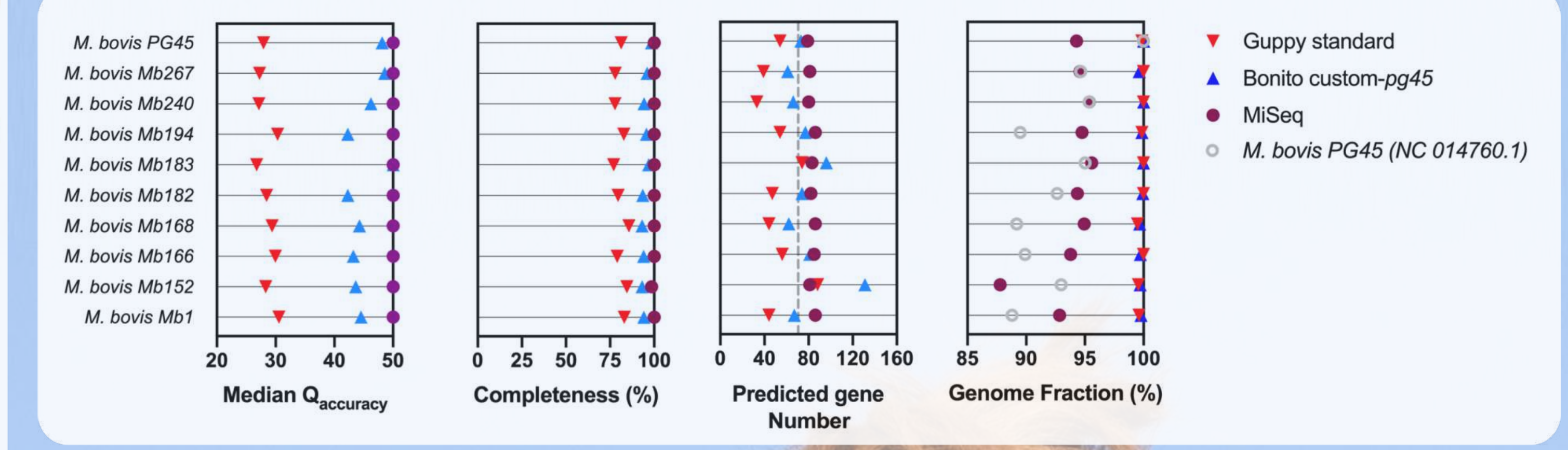
1 - Improved reference-based and *de novo* assembled *M. bovis* Genomes

- *Reference-based assembly*
  - Improved genome quality with a **Q45.2** accuracy score
    +Q1.3 as compared to MiSeq
  - More complete genomes (99.3% and 69 out of 75 predicted genes)
- *De novo assembly*
  - **Canu** without Racon polishing preferred assembler
  - Improved *de novo* genome quality with a **Q46.7** accuracy score
  - Higher genome fraction covered
    +14.3% genome fraction covered as compared to MiSeq

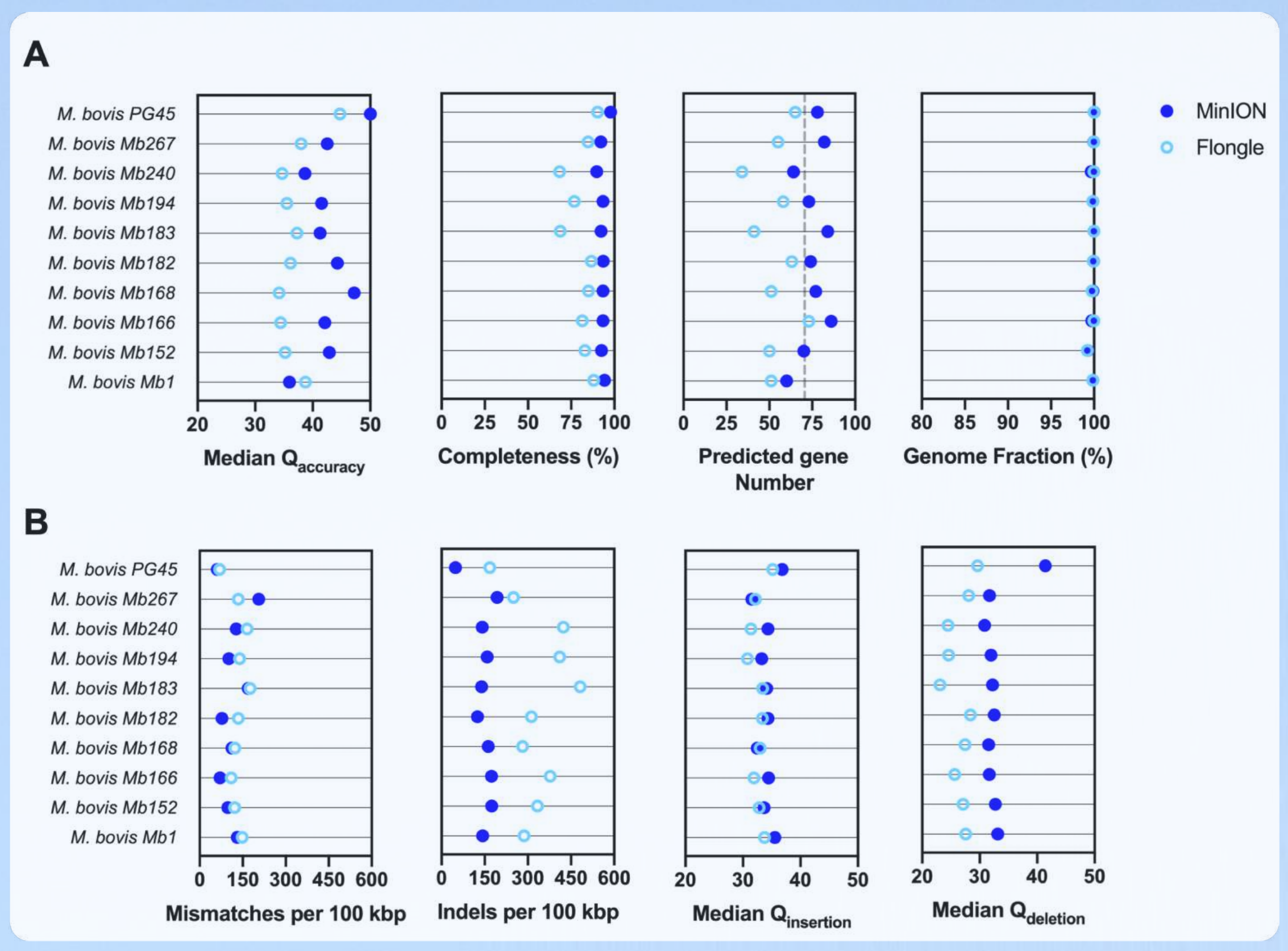


2 - Validation of the *custom-pg45* basecalling model in nine *M. bovis* field strains

- Improved *de novo* genome qualities for all field strains reaching to **Q42 - Q50** accuracies
- Higher genome fraction covered (**99.7% - 100%**) as compared to MiSeq (87.8% - 95.6%)



3 - Application of single-use Flongle flow cells in an optimized Bioinformatics workflow for veterinary diagnostics



- Taxon-specific training results in similar improvements for **Flongle reads**
  - Lower accuracy genomes (*av.* Q36)
    = 267 errors per 1 Mbp
  - Lower genome completeness (*av.* 80%)
  - Full Genome fraction covered (*av.* 100%)
- Increased error suggested due to **Indels**

Conclusions

Advantages of Nanopore sequencing
- Sequencing native DNA – *incl. epigenetics (Distinct 6mA methylation patterns for M. bovis)*
- Increased genome coverage for AT-rich and highly repetitive genomes

Implementation of **taxon-specific Bonito basecalling**
- Increased genome quality and completeness for both reference-based (**Q45.2**) and *de novo* (**Q46.7**) assemblies
- ***Canu*** (without Racon polishing) preferred *de novo* assembler
- Validation shown for nine additional *M. bovis* field strains

Taxon-specific basecalling of **MinION** and **Flongle** reads are of great value in rapid *all-in-one* WGS tools

Acknowledgments

Industrial Research Fund of Ghent University