GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

CiBreed Center for Integrated Breeding Research

Assessing Nanopore Sequencing for Detection of Large Structural Variation in the Goettingen Minipig

J. Geibel[1*], S. Hansen[2*], A. Abdelwahed[2], S. Böhlken-Fascher[2], A. R. Sharifi[1], C.-P. Czerny[2], H. Simianer[1] & C. Reimer[1]

Background and Aim

- Structural variation is expected to underlie severe phenotypic differences
- Large structural variants are not directly observable from short sequencing reads [1] → validating results is challenging
- Latest developments in long read sequencing technologies promise increased accuracies and read lengths up to mega bases, possibly spanning structural variants [1]

⇒ **Assessment of nanopore sequencing performance for a Goettingen Minipig trio on a MinION device**

⇒ **Evaluation of usability for detection of large structural variants**

Conclusions

- Preliminary results, especially the achieved read lengths, are promising for a larger scaled use of nanopore sequencing for detection of structural variants in farm animals
- Strong increase in coverage is necessary for reliable results
- 1D² calling and barcoding lead to avoidable yield losses for mammalian genomes

⇒ **Future strategy: Combination of unsheared 1D library (ultra long reads) and sheared 1D library (increased yield and quality) without any barcoding, potentially in combination with short Illumina reads**

General Approach

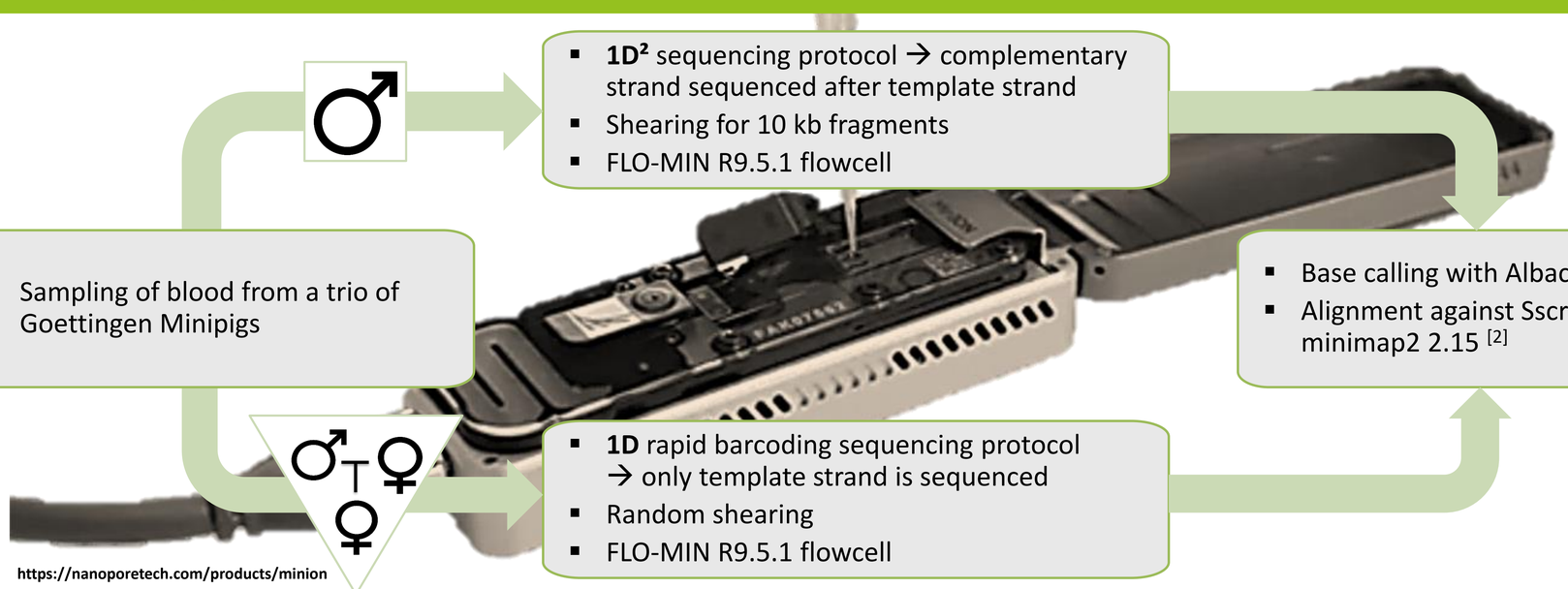
- Sampling of blood from a trio of Goettingen Minipigs

♂

- **1D²** sequencing protocol → complementary strand sequenced after template strand
- Shearing for 10 kb fragments
- FLO-MIN R9.5.1 flowcell

♂ ♀ ♀

- **1D** rapid barcoding sequencing protocol → only template strand is sequenced
- Random shearing
- FLO-MIN R9.5.1 flowcell



https://nanoporetech.com/products/minion

- Base calling with Albacore 2.3.4
- Alignment against Sscrofa11.1 with minimap2 2.15 [2]

- Quality Control with NanoPlot 1.20.1 [3] and poretools 0.6.0 [4]
- Individual based structural variant detection (>50bp) with SNIFFLES 1.0.11 [5]
- Population-wide structural variant calling with SNIFFLES 1.0.11 [5] and SURVIVOR 1.0.7 [6]

Results

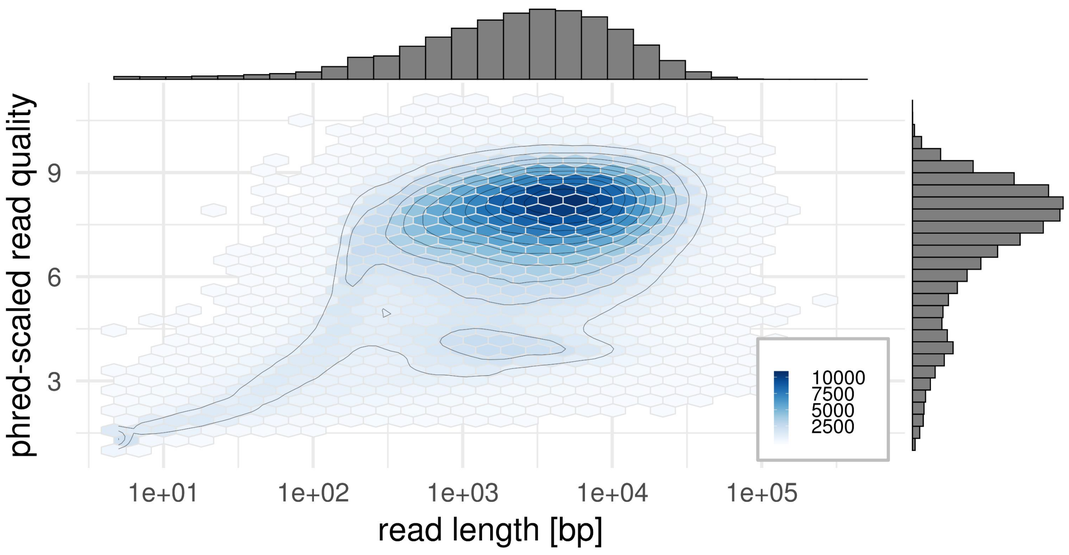


Fig. 1: Read length vs. read quality from the 1D barcoded sequencing.

- 1D² workflow with only 71 / 512 active pores → not satisfying yield of 121 Mb / 21 h
- 1D workflow with 508 / 512 active pores → 5.6 Gb / 25 h, could be increased with longer runtimes
- Only 1 Mb 1D² consensus reads could be restored from 121 Mb 1D reads
- De-barcoding failed for 50 % of the reads (2 Mb out of 5.6 Mb)
- Average phred-scaled read quality of 7.9 (83 % read accuracy; Fig. 1)
- 1D² increased read quality to 9.5 (88 % read accuracy)
- Read length N50 of 9.1 kb (1D² run) and 11.8 kb (1D run)
- Longest read: 505,912 bp (> 1000 short reads)

- Coverage < 0.3 X for parents and 0.73 X for offspring
- > 50 % of coverage due to reads > 10kb
- 30 % of reference genome uncovered (Fig. 2 A)
- 110,950 structural variants (> 50 bp) called when requesting one supporting read
- 436 structural variants (> 50 bp) called when requesting five supporting reads
- 237 structural variants (≥ 5 supporting reads) segregating in all animals
- Mostly insertions and deletions between 100 and 1000 bp (Fig 2 B)

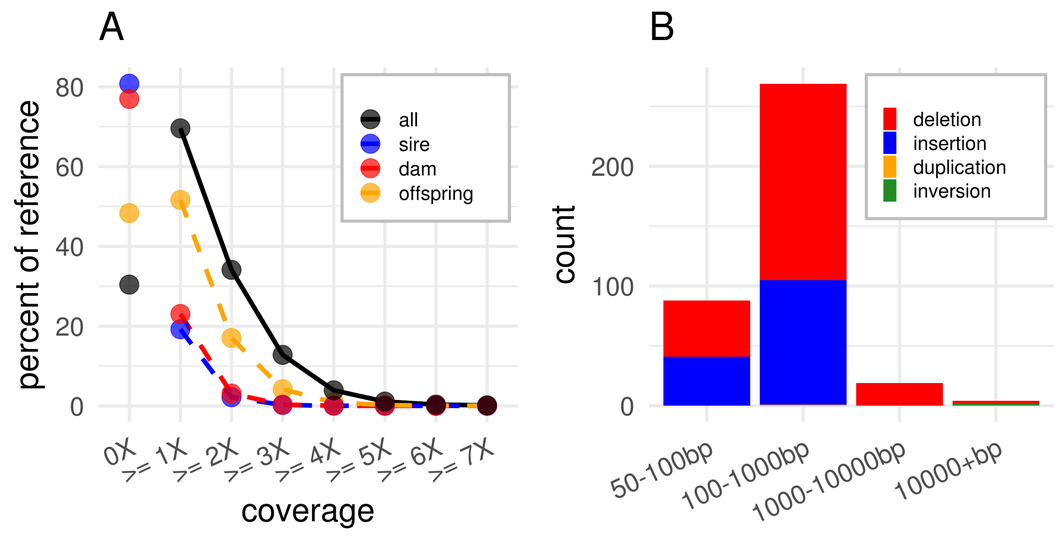


Fig. 2: Achieved (cumulative) coverage (A) and called structural variants when requesting at least 5 supporting reads in the population (B).

References

[1] Sedlazeck FJ et al. **(2018)** *Piercing the dark matter: Bioinformatics of long-range sequencing and mapping.* Nature Reviews Genetics 19:329–346

[2] Li H **(2018)** *Minimap2: Pairwise alignment for nucleotide sequences*. Bioinformatics 34:3094–3100

[3] Coster W de et al. **(2018)** *NanoPack: Visualizing and processing long-read sequencing data.* Bioinformatics 34:2666–2669

[4] Loman NJ, Quinlan AR **(2014)** *Poretools: A toolkit for analyzing nanopore sequence data*. Bioinformatics 30:3399–3401

[5] Sedlazeck FJ et al**. (2018)** *Accurate detection of complex structural variations using single-molecule sequencing*. Nat Methods 15:461–468

[6] Jeffares DC et al. **(2017)** *Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast*. Nature Communications 8:14061

Affiliations

[1] *University of Goettingen, Animal Breeding and Genetics Group, Center for Integrated Breeding Research, Albrecht-Thaer-Weg 3, 37075 Goettingen, Germany*

[2] *University of Goettingen, Microbiology and Animal Hygiene, Burckhardtweg 2, 37077 Göttingen, Germany*

[*] these authors contributed equally to the work

Corresponding author: Johannes Geibel, johannes.geibel@uni-goettingen.de