



**HAL**  
open science

## Volume reduction and automation of DNAPrep libraries for low pass sequencing

Marcuzzo Camille, Cervin Guyomar, Amandine Suin, Adrien Castinel, Carole Iampietro, Claire Kuchly, Julie Demars, Nathalie Iannuccelli, Christophe Klopp, Christine Gaspin, et al.

### ► To cite this version:

Marcuzzo Camille, Cervin Guyomar, Amandine Suin, Adrien Castinel, Carole Iampietro, et al.. Volume reduction and automation of DNAPrep libraries for low pass sequencing. Assemblée générale France Génomique, Jun 2023, Paris, France. hal-04590691

**HAL Id: hal-04590691**

**<https://hal.science/hal-04590691v1>**

Submitted on 28 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

Marcuzzo Camille<sup>1</sup>; Guyomar Cervin<sup>2,3</sup>; Suin Amandine<sup>1</sup>; Castinel Adrien<sup>1</sup>; Iampietro Carole<sup>1</sup>; Kuchly Claire<sup>1</sup>; Demars Julie<sup>2</sup>; Iannuceli Nathalie<sup>2</sup>; Klopp Christophe<sup>3</sup>; Gaspin Christine<sup>3</sup>; Milan Denis<sup>1,2</sup>; Donnadiou Cécile<sup>1</sup>

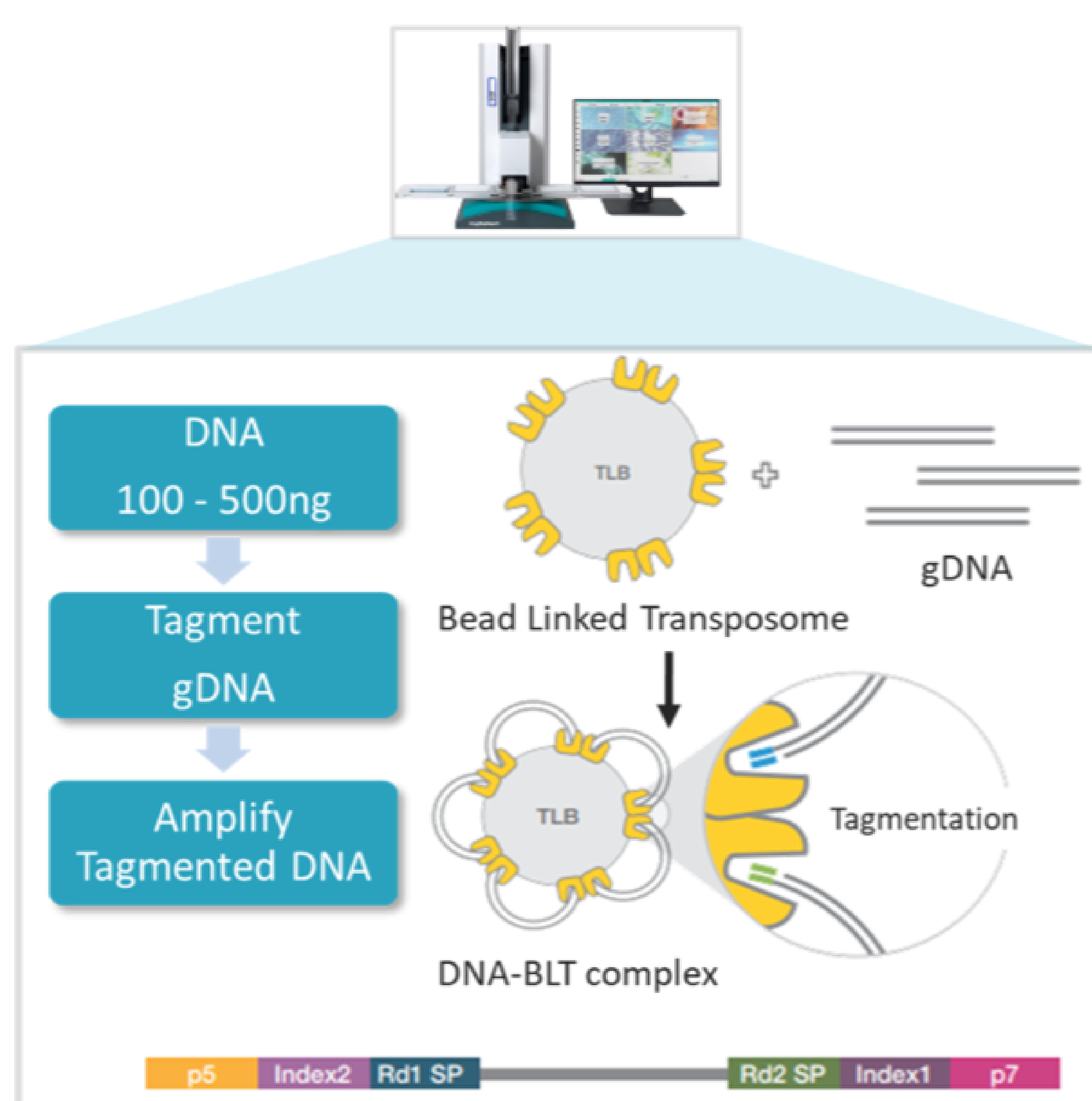
<sup>1</sup> INRAE, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan ; <sup>2</sup> GenPhySE, Université de Toulouse, INRAE, INPT, ENVT, Castanet-Tolosan Cedex, F-31326 ; <sup>3</sup> Plateforme Bio-informatique Genotoul, Mathématiques et Informatique Appliquées de Toulouse, INRAE, Castanet-Tolosan

## Background and objectives

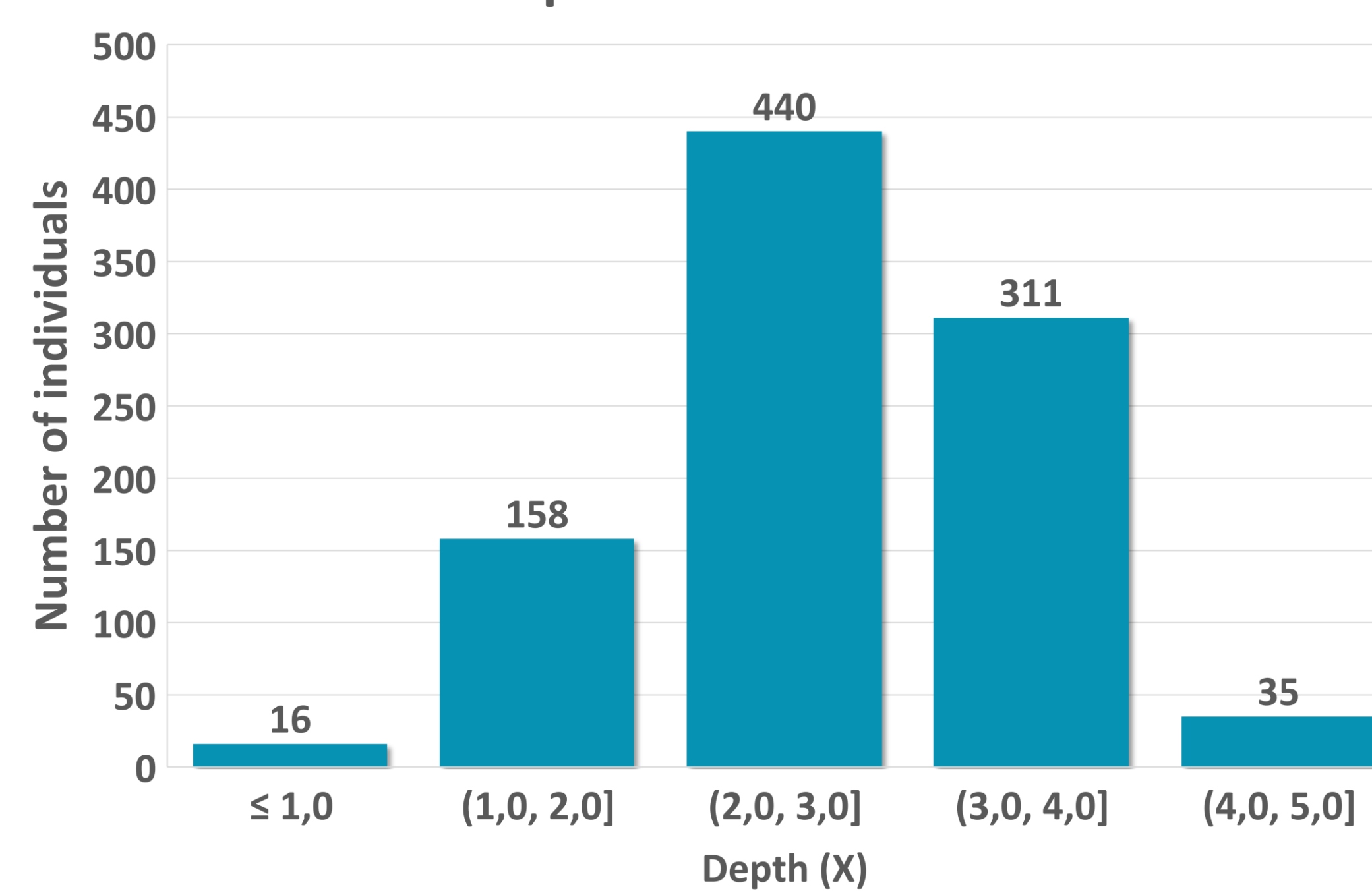
To characterize a population's haplotypic variability, **low pass sequencing** represents an interesting alternative to genotyping DNA-microarrays. Although this new technology is also costly, it offers greater precision of genetic information. We have therefore reduced reagent and DNA volumes by automating our Illumina **DNA Prep library** preparation protocol with the **Mosquito HV** (SPT LabTech, UK). For this purpose, we intend to evaluate this strategy for the **porcine species**. We used a family F2 protocol based on crossbreeds between Large White (LW) and Meishan (MS) individuals. It comprises around **1000 pigs**, including 12 F0 (6 LW and 6 MS), 30 F1 (6 males and 24 females) and 1000 F2. In addition to the evaluation of bioinformatic phasing methods for the **imputation** and reconstruction of **haplotypes**. The exploitation of these data will enable the precise characterization of **recombination sites**. This will provide an understanding of the reasons why a large region of the X chromosome recombines or not. The SeqOccIn **porcine pilot project** gave us the opportunity to (i) **automate the DNA prep protocol** to significantly reduce volumes with very good results and (ii) generate **good quality low pass sequencing data** for genotype imputation (validated by Mendelian compatibility).

## Protocol optimisation: automation with volume reduction

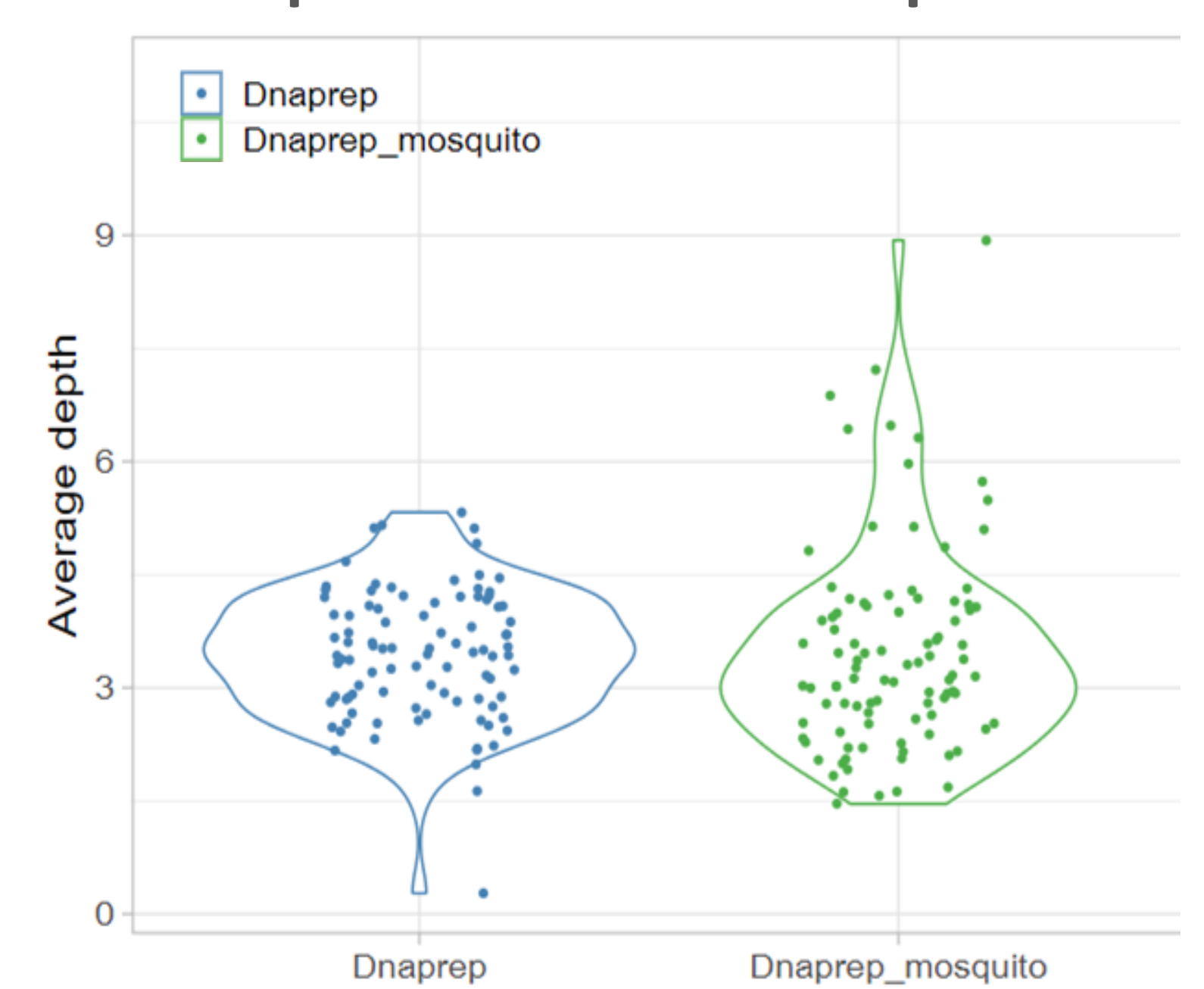
### 1A DNA Prep library workflow on Mosquito HV



### 1B Depth distribution



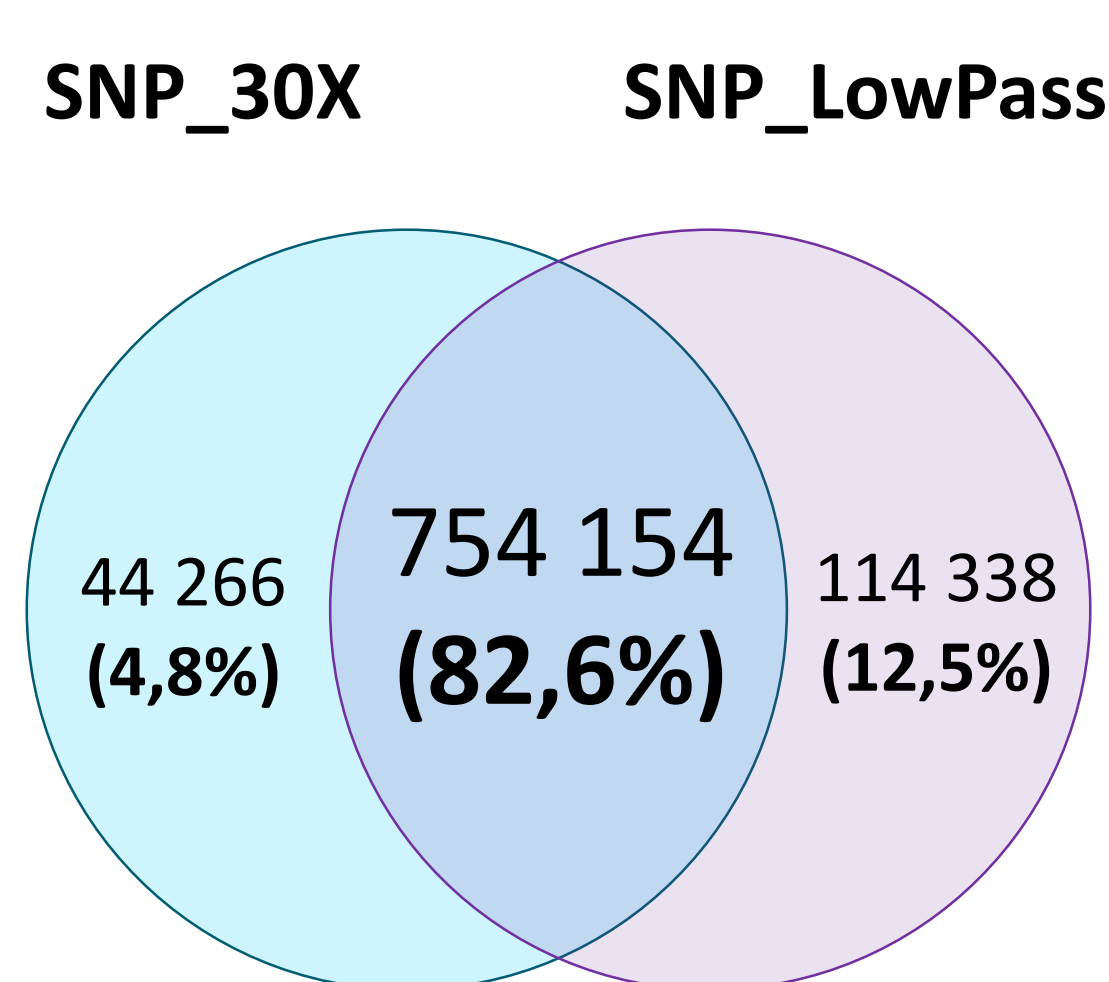
### 1C Depth distribution comparison



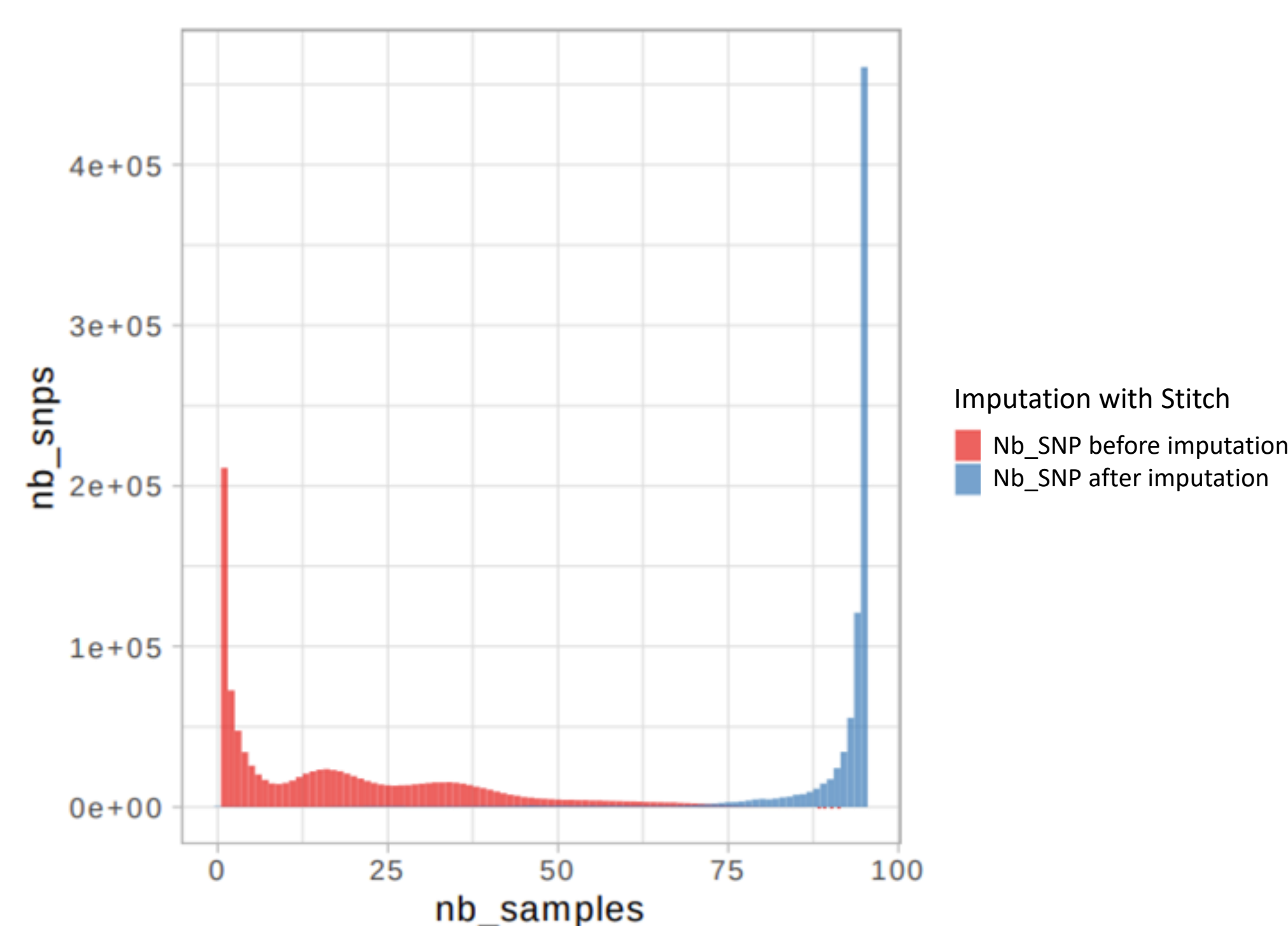
**Figure 1.** **1A.** DNA Prep library workflow. This kit uses **tagmentation chemistry**, which combines the steps of DNA fragmentation and adapter ligation. It also eliminates the need to quantify libraries prior to pooling for sequencing. This protocol has been **automated on the Mosquito HV**, with the volumes being **reduced by a factor of 5** for this project. **1B.** Distribution of coverage obtained on the cohort of 960 pigs using the DNA Prep library protocol. The mean coverage is **2.7X**. Heterogeneity is expected since balancing is based entirely on initial DNA quantification. **1C.** Graphic representation of **the average depth across the genome** (equivalent to the size of the library). This graph confirms that there is the **same variation in depth** with (green) or without Mosquito HV (blue).

## Protocol validation

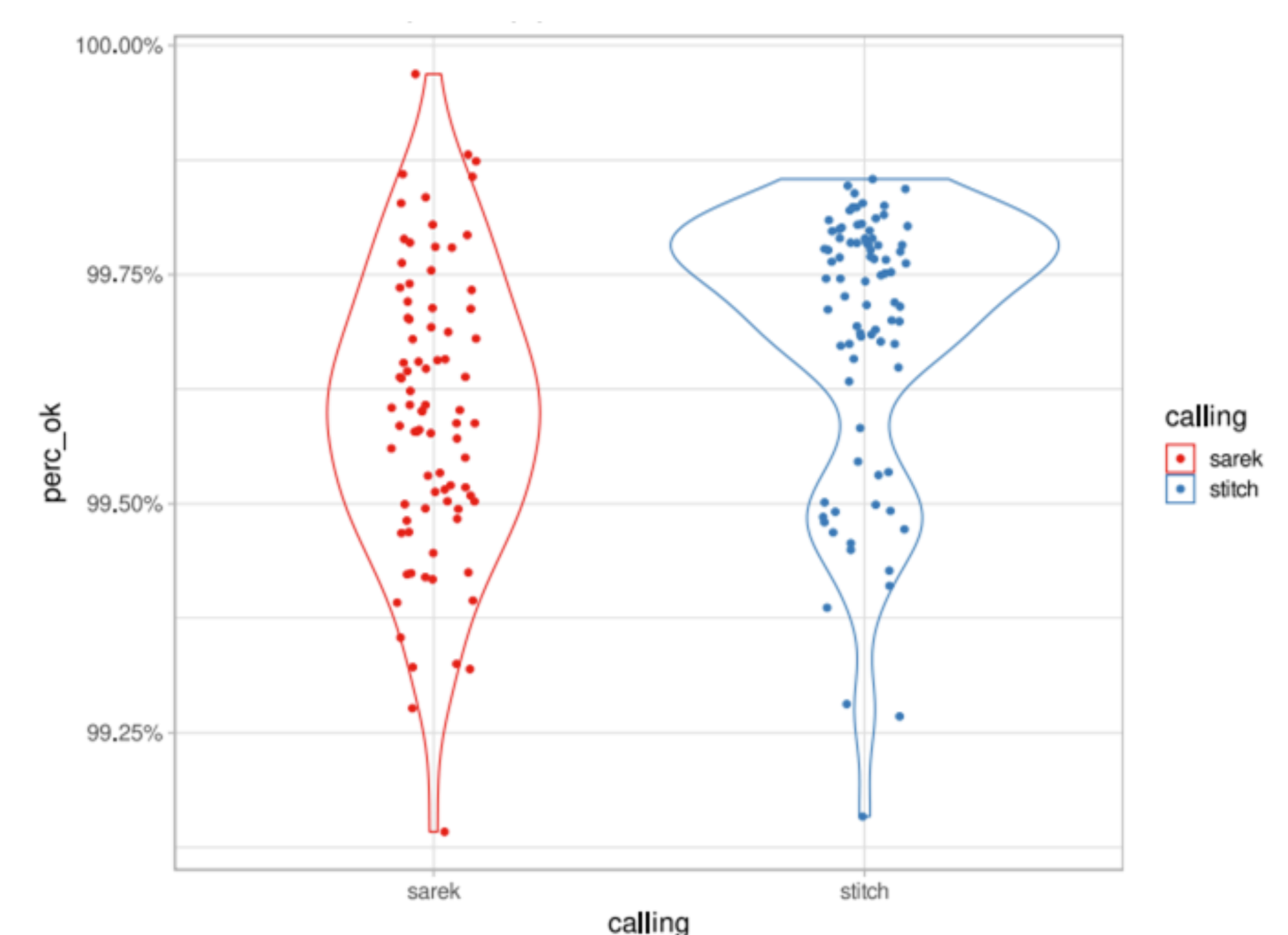
### 2A



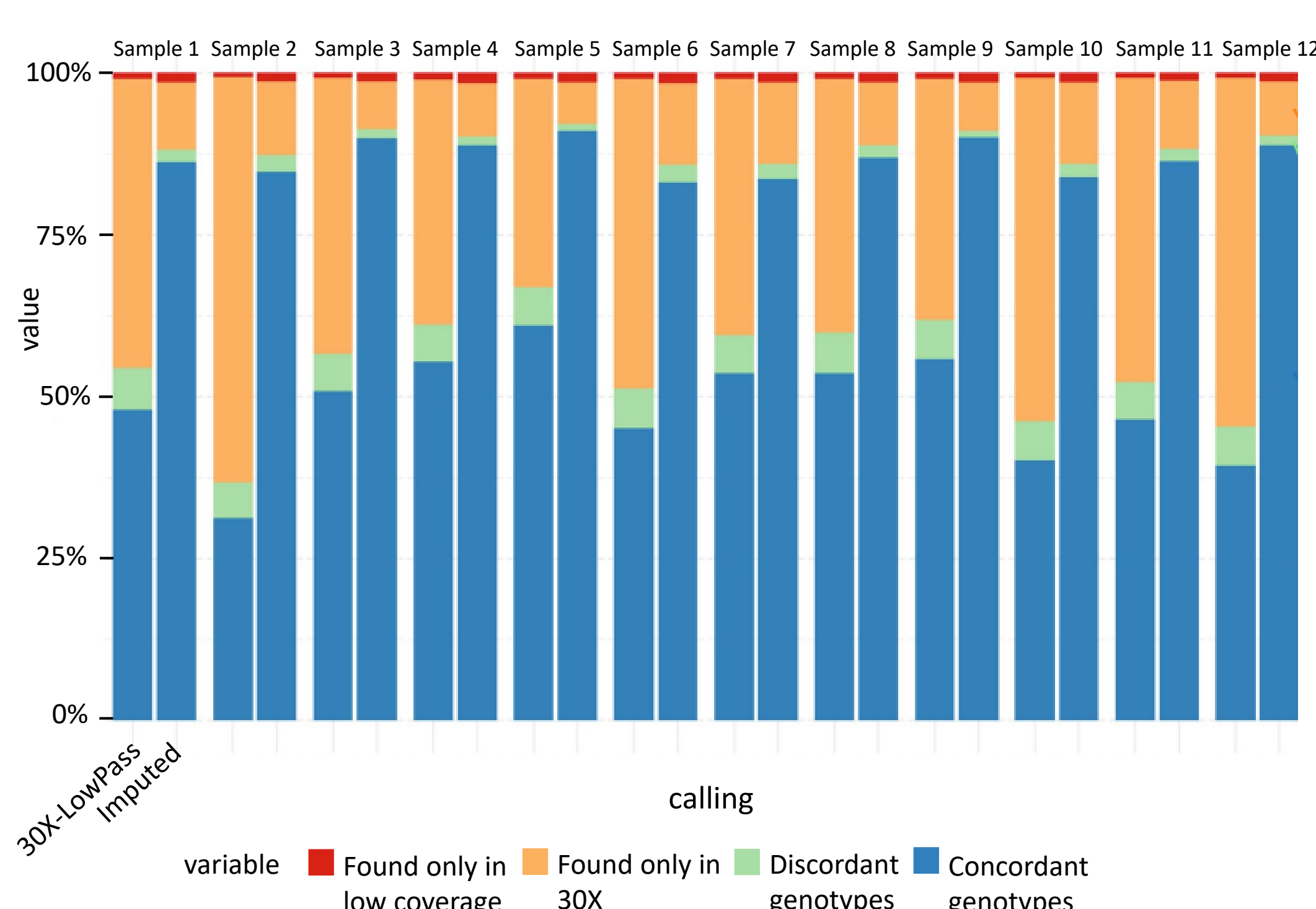
### 2C Number of genotyped samples distribution



### 2D Mendelian compatibility per trio



### 2B Concordance between low pass data and 30X reference



**Figure 2.** The parents (F0) were sequenced in 30X and the descendants (F1-F2) at 1X low pass. The tools used for **genotype imputation** are **nf-core/sarek** and **Stitch**. nf-core/sarek is a workflow designed to **detect variants** across the entire genome. The Stitch algorithm runs **without reference panels**. **2A.** Sparse allows to detect a similar set of variants compared to **high-coverage sequencing**. The phasing method on the trio detects **82.6% of variants**. **2B.** Graph of the **imputation effect** on genotyping of individuals sequenced at 30X and low pass. Imputation increases the percentage of identical genotypes between low pass and 30X (blue), corrects genotypes (green) and reduces false positives (red/orange). **2C.** The distribution of the number of variants in the same sample shows the imputation effect on the number of samples genotyped (blue). The **quality of the genotyping** is visible through the Mendelian plot. **2D.** There is **no degradation of symmetry after imputation** (blue) and genotyping is better corrected.

## Conclusion & perspectives

- Low-pass sequencing makes it possible to generate high-quality whole-genome genotyping data (pilot project results) at a cost close to that of a DNA-microarrays.
- We will develop volume reduction by automating other protocols on the Mosquito HV.