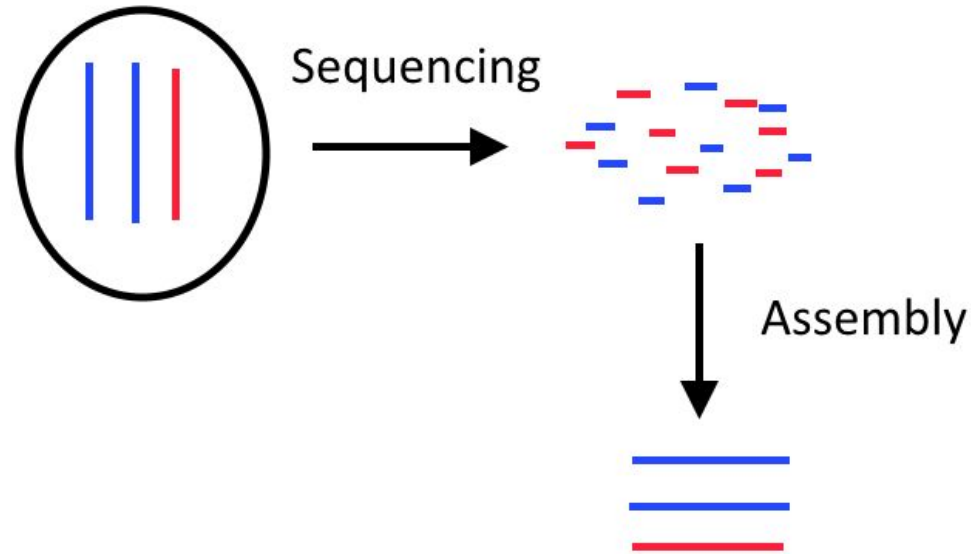# nPhase and
hybrid polyploid considerations

Omar Abou Saada
oabousaada@unistra.fr

If one of the haplotypes is very different from the others...

nPhase will struggle to distinguish between the haplotypes that are comparatively closer

SNP density strategy

The **1n** portion of the hybrid genome might account for **most** het positions

We may be able to separate 2n and 1n reads based on **SNP density**

When aligned to a reference, long reads should be long enough to accumulate:

    A low density of heterozygous positions for the **diploid** part of the hybrid
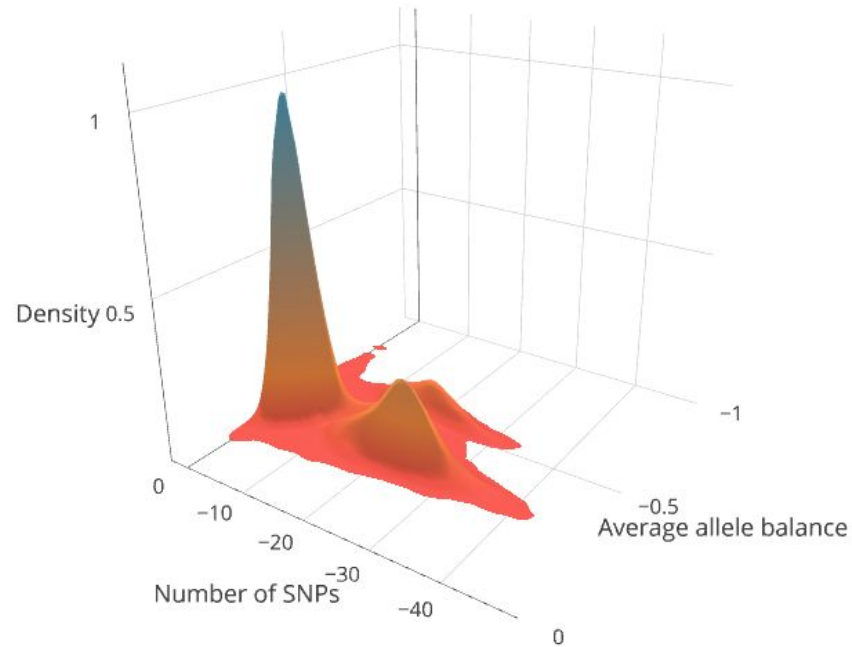
    A high density of heterozygous positions for the **haploid** part of the hybrid

(If the **1n** accounts for the least amount of het positions we simply invert the logic)

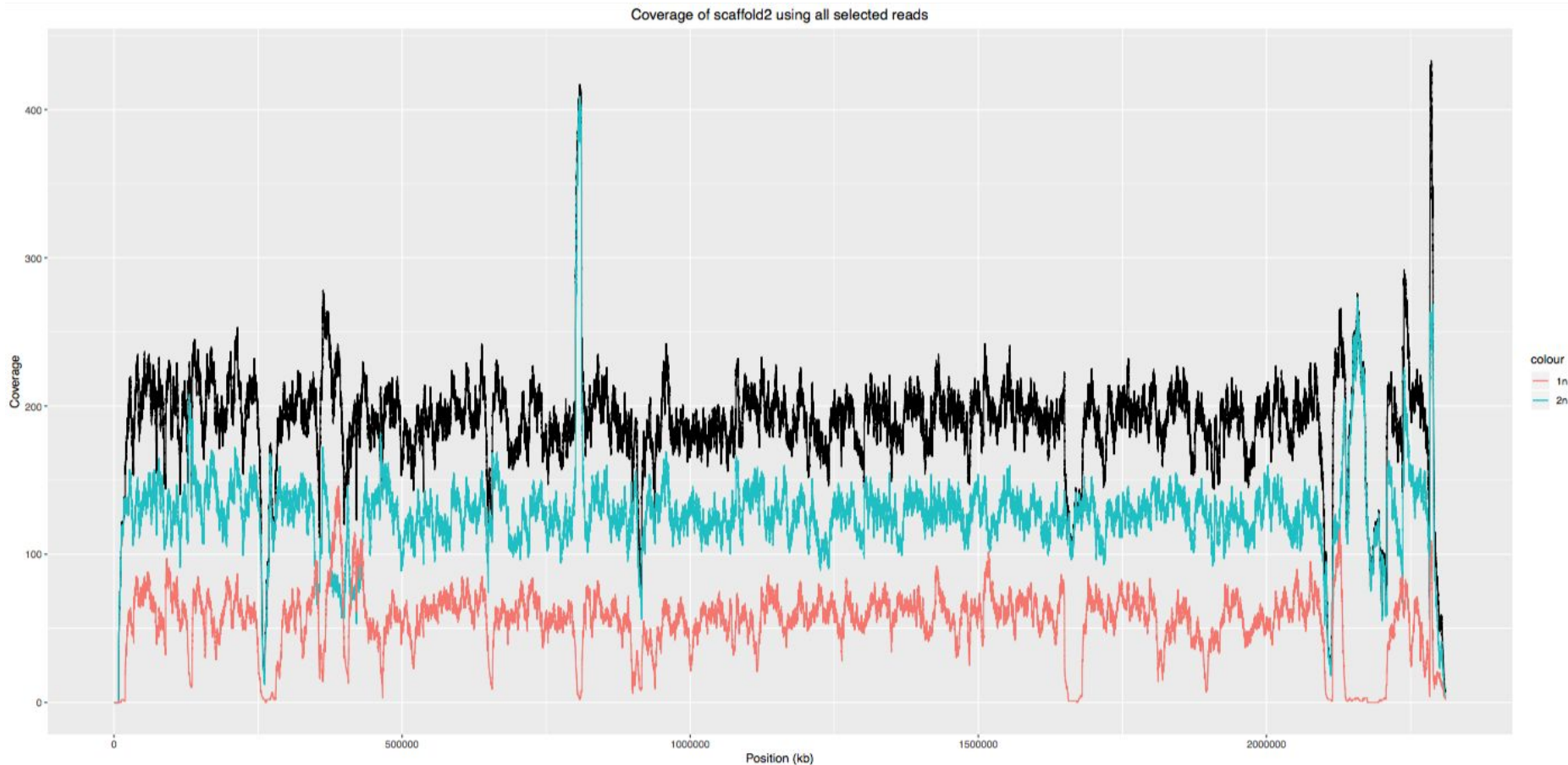**Distribution of SNP density in 30350 reads > 3500 bp**

We have a clear ⅓ ⅔ distribution here by which to segregate the reads
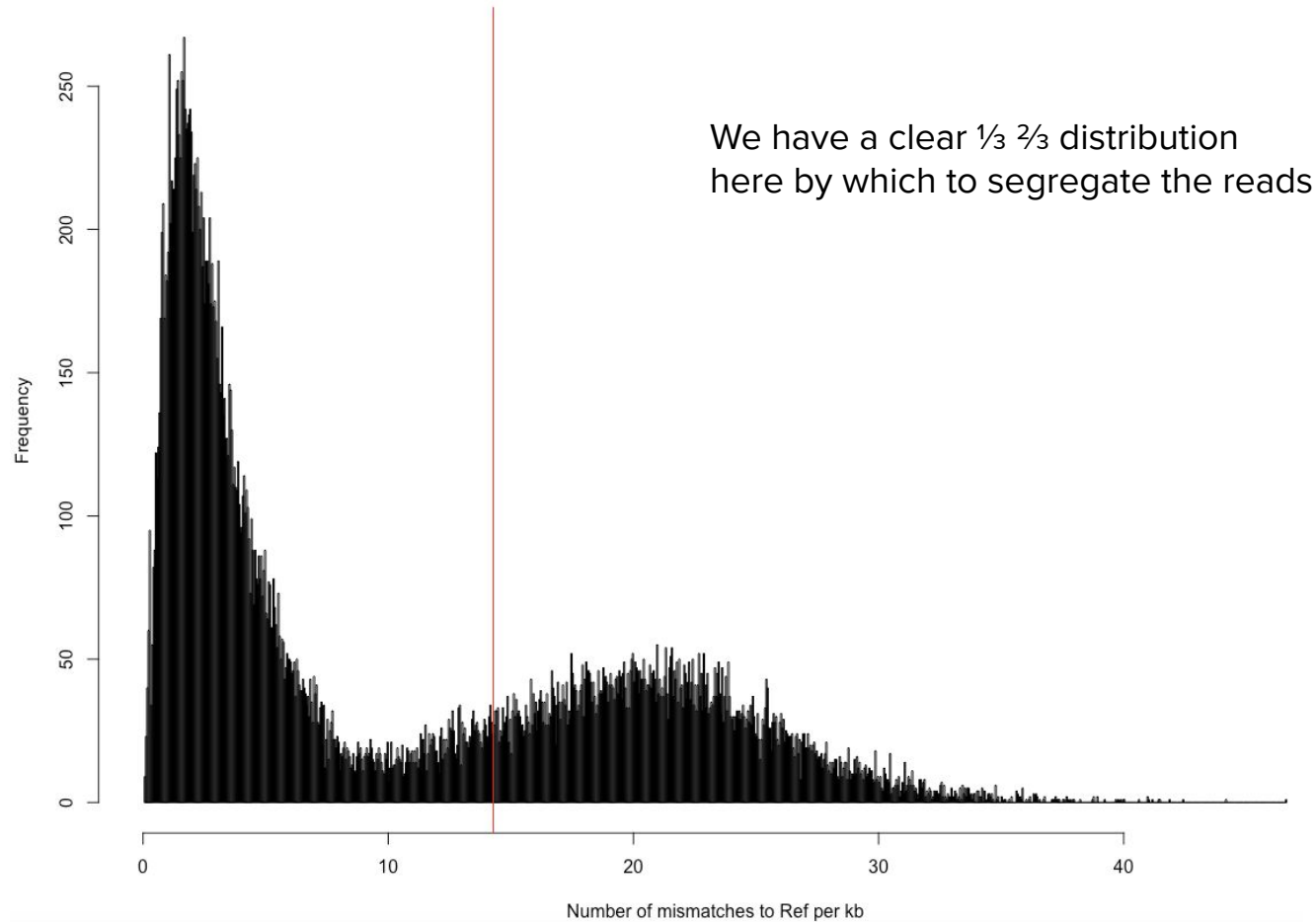
SNP density and average allele balance 3D plot

Remapping reads reveals a consistent ⅔ ⅓ distribution

Coverage of scaffold2 using all selected reads

If all reads have the same SNP density...
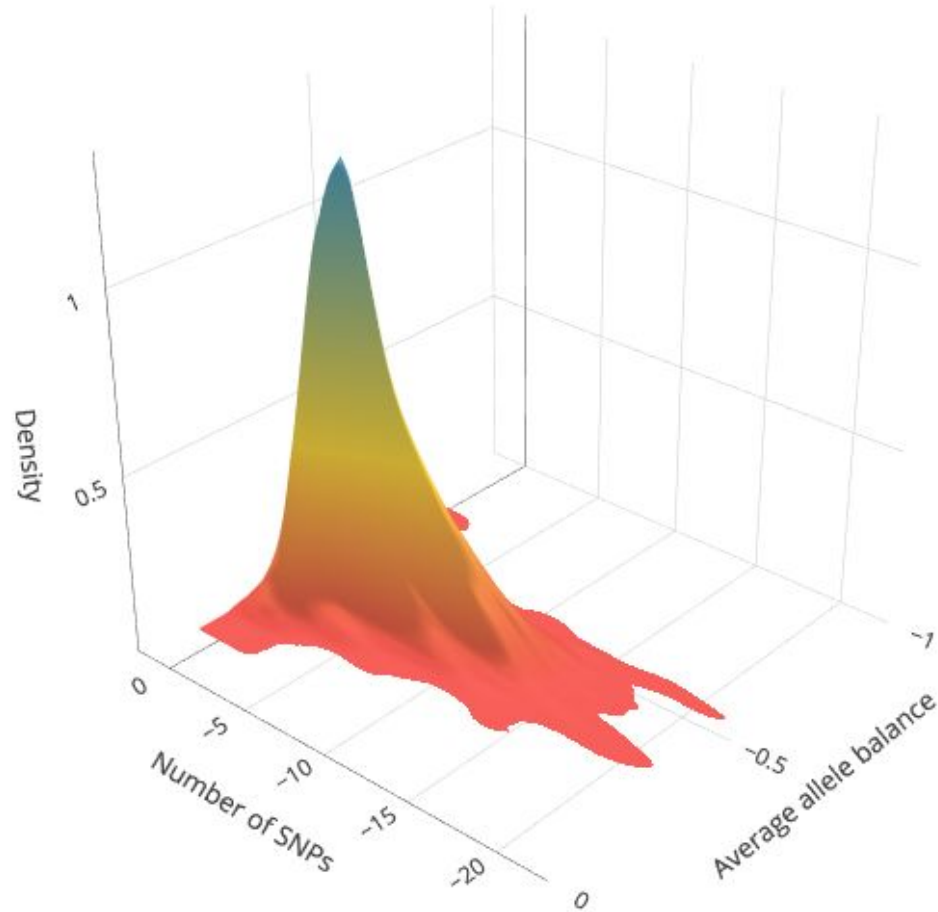
Then use nPhase: Polyploid phasing pipeline
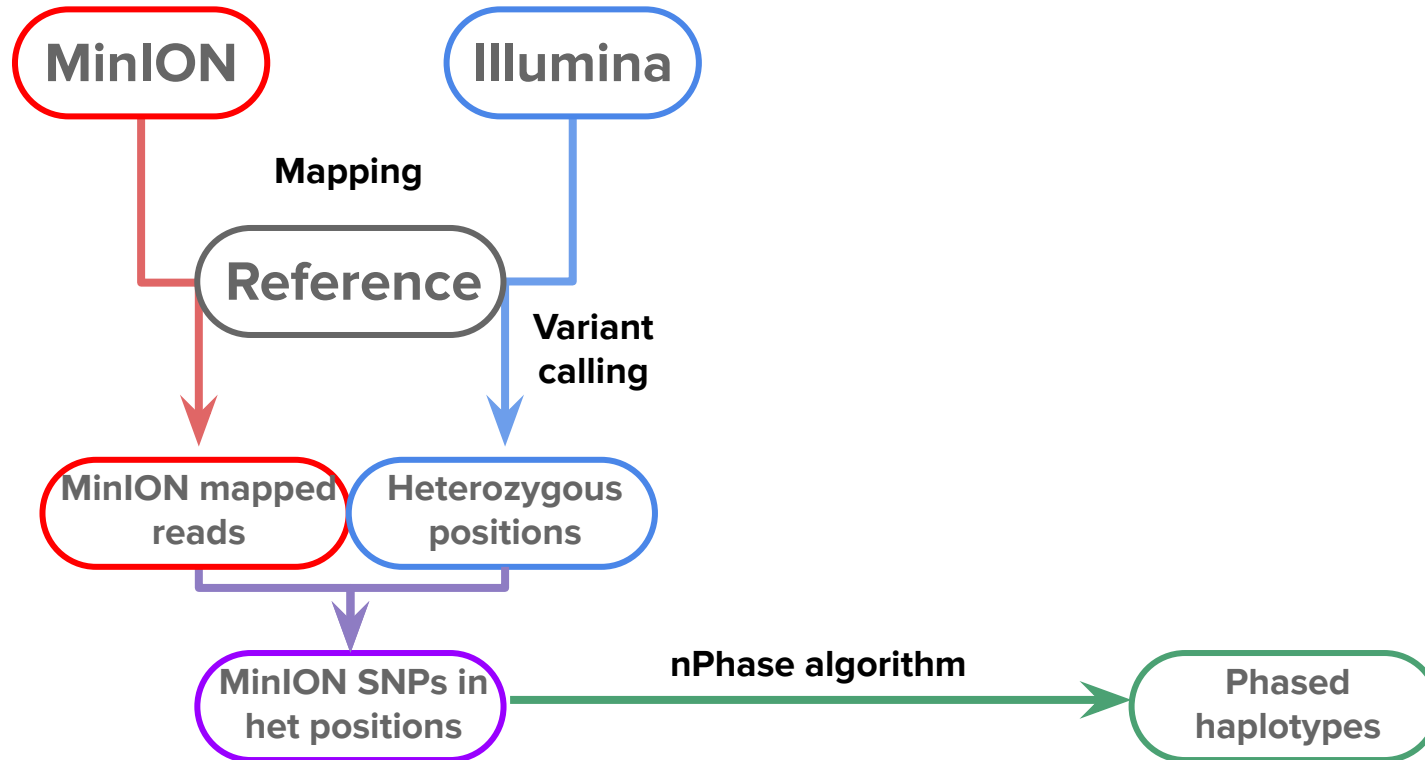
Thank you for your attention

Université de Strasbourg

NIH

gmgm
Génétique Moléculaire ∎ Génomique ∎ Microbiologie

cnrs

Oxford NANOPORE Technologies