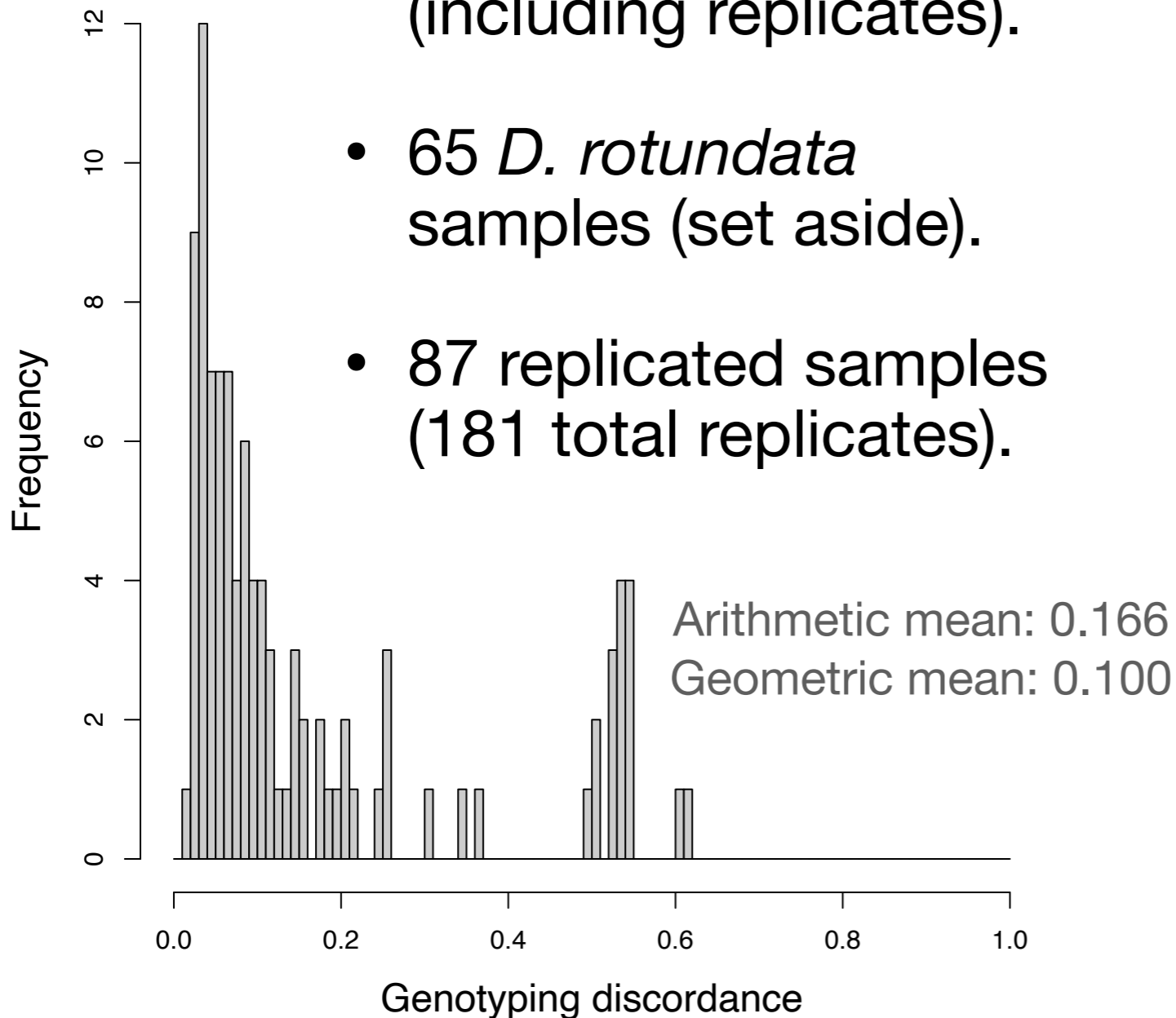


# ***D. alata* diversity**

**An update on DArTseq data analysis**

# Diversity sample summary

- 1,427 initial genotypes (including replicates).
- 65 *D. rotundata* samples (set aside).
- 87 replicated samples (181 total replicates).

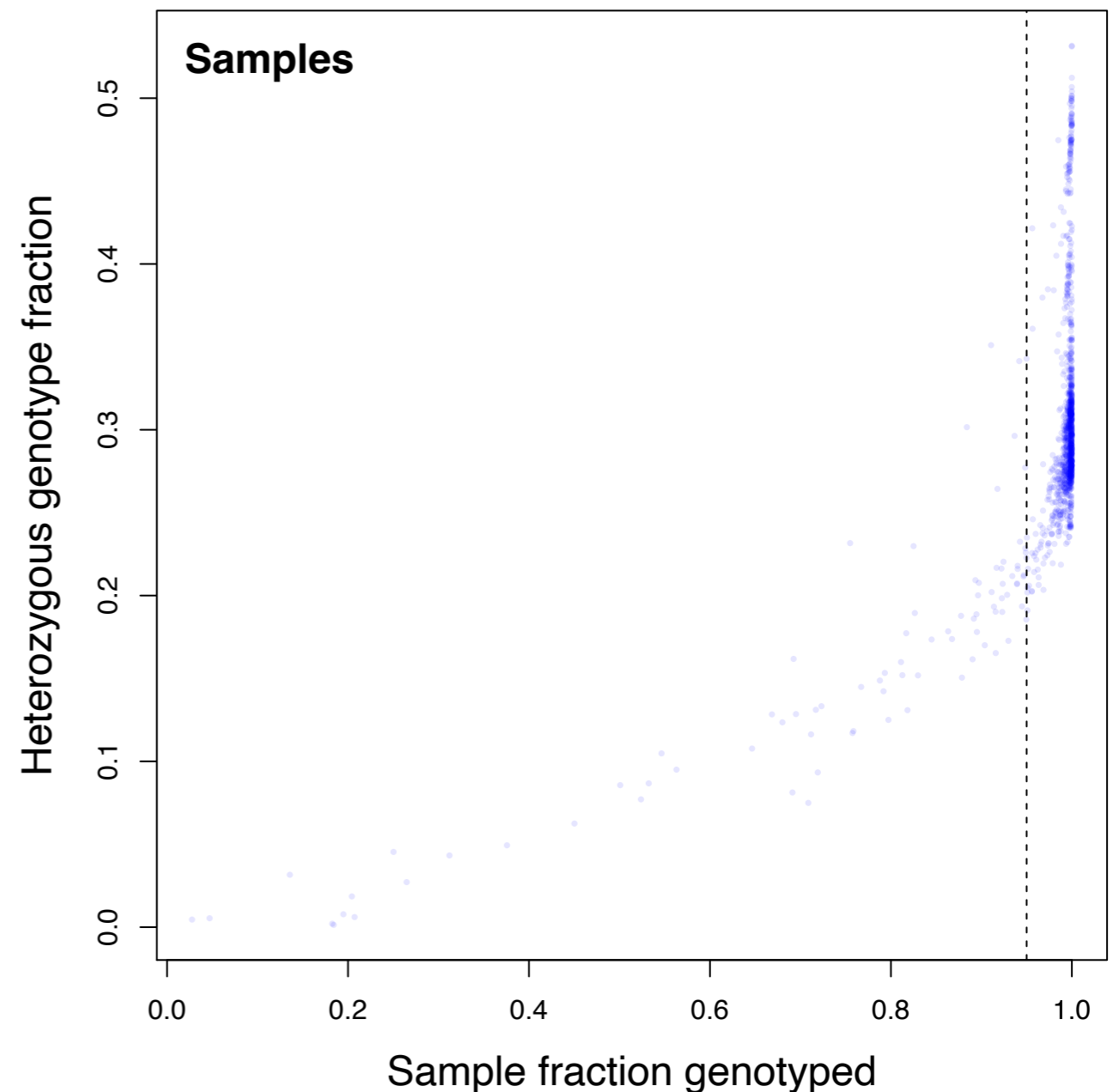
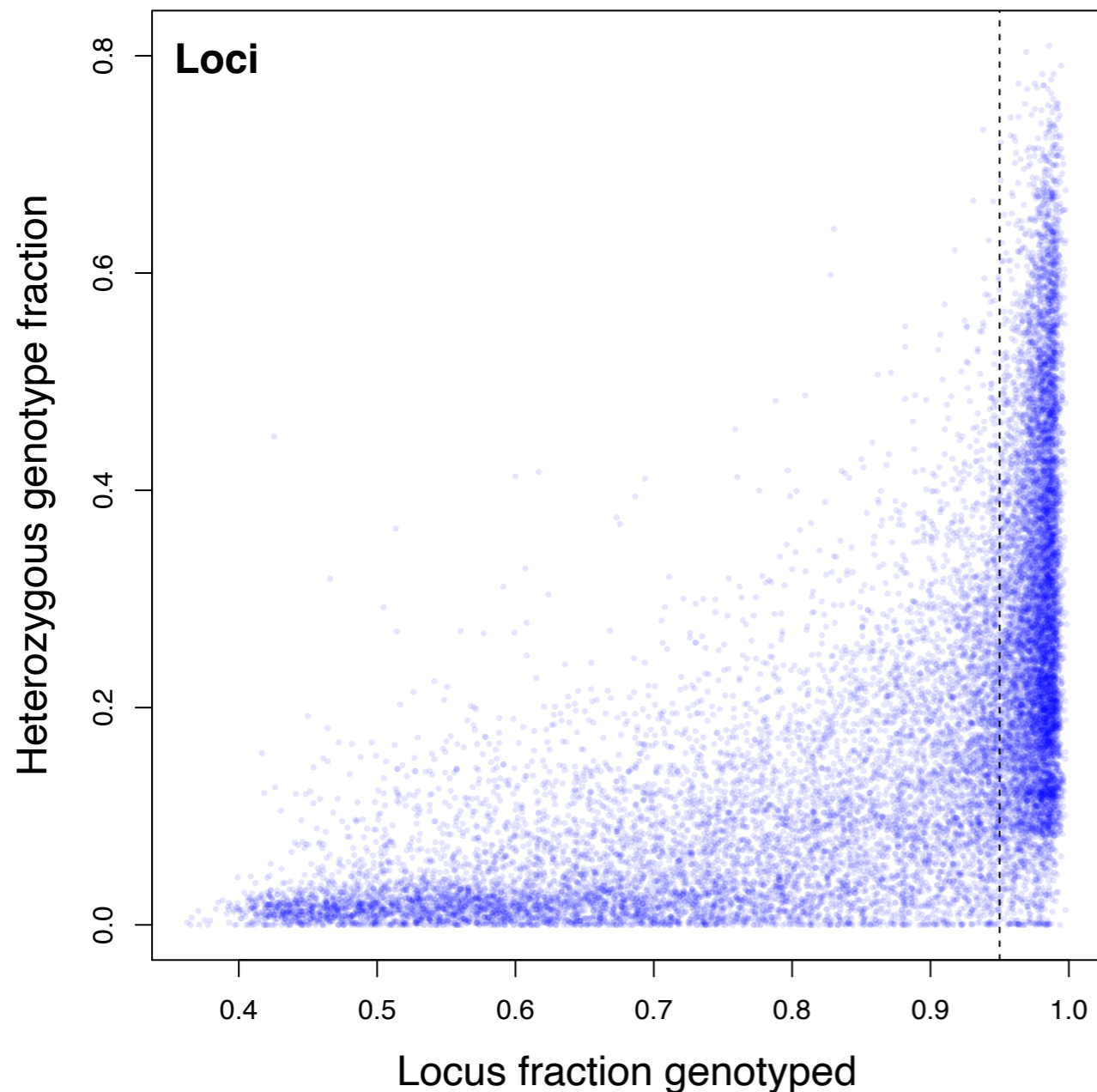


Nominal source	Count
<b>Total</b>	1,272
<b>Nigeria</b> (1,026 IITA; 28 NRCRI)	1,054
<b>Cote d'Ivoire</b>	90
<b>Tanzania</b>	82
<b>Cameroon</b>	18
<b>Japan</b>	16
<b>Vietnam</b>	12

# Genotyping quality control

## Low genotyping rate => biased het frequencies

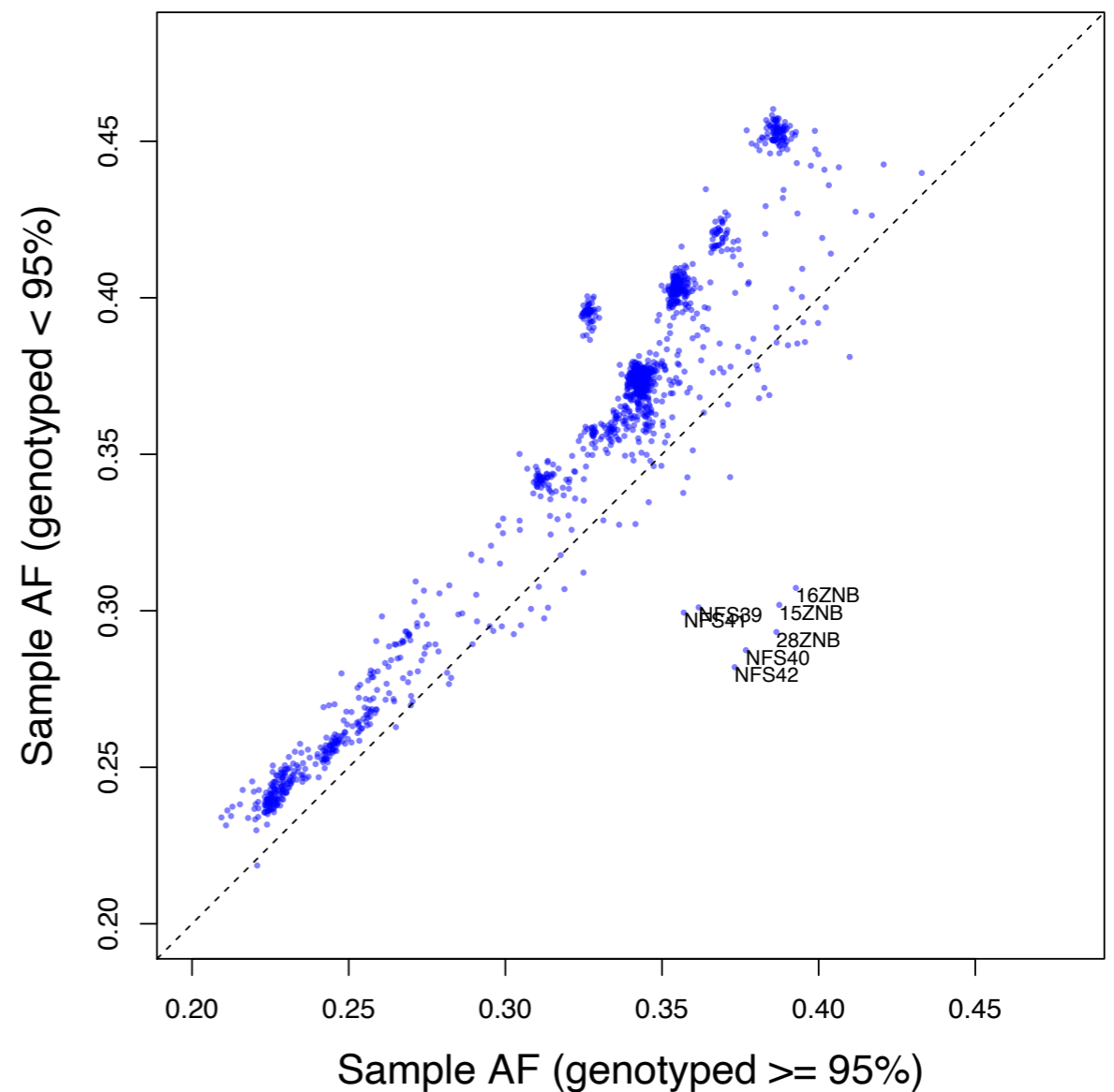
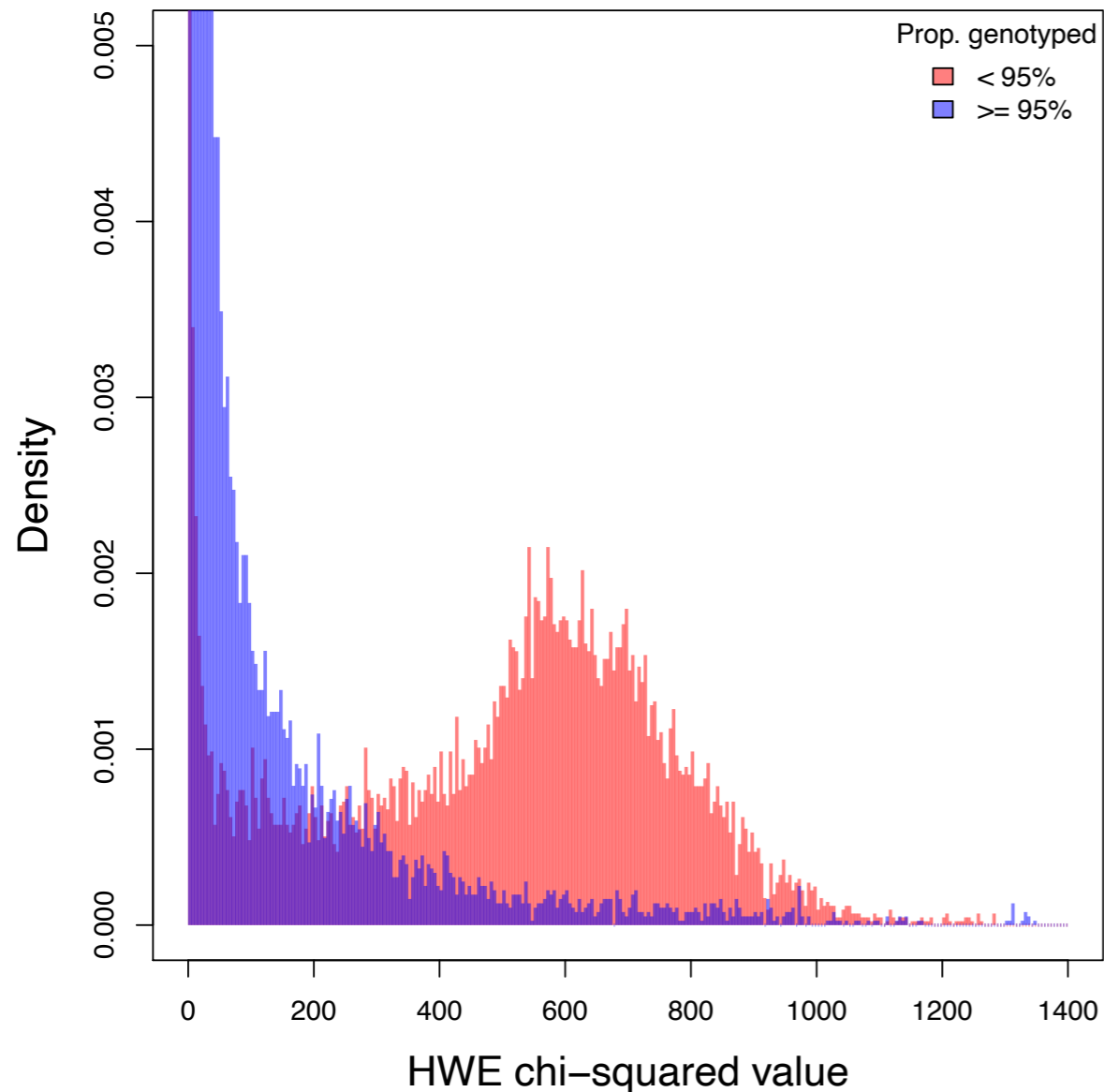
- 7,921 loci  $\geq$  95% genotyped
- 9,079 loci  $<$  95% genotyped
- 1,171 samples  $\geq$  95% genotyped
- 89 samples  $<$  95% genotyped



# Genotyping quality control

## Low genotyping rate => biased allele frequencies

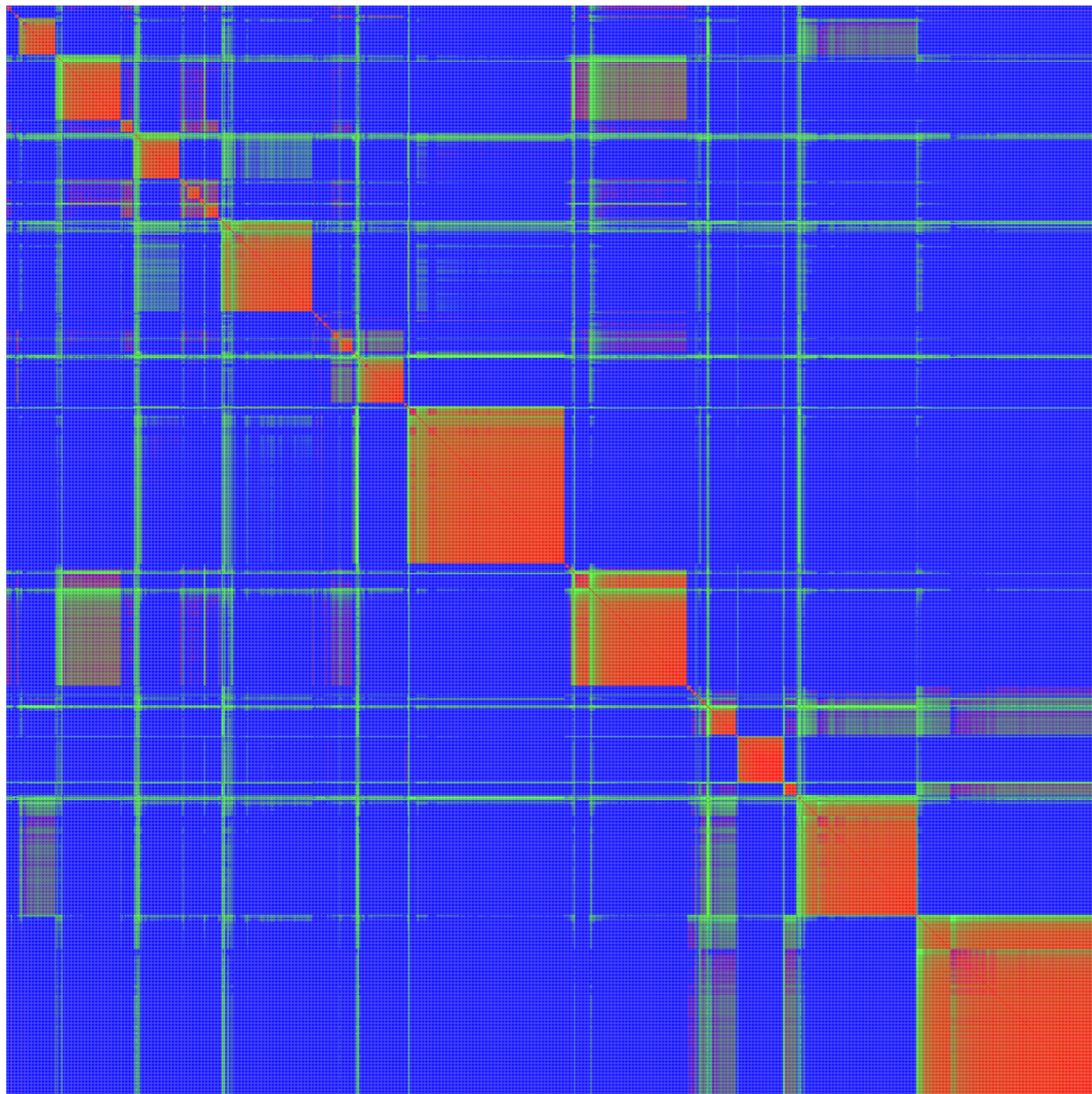
- Loci < 95% genotyping exhibit biased Hardy-Weinberg values
- Loci < 95% genotyping not due to interspecific alleles (balanced between partitions)



# Sample quality control

## Using measures of relatedness

■ IBD2    ■ IBD1    ■ IBD0

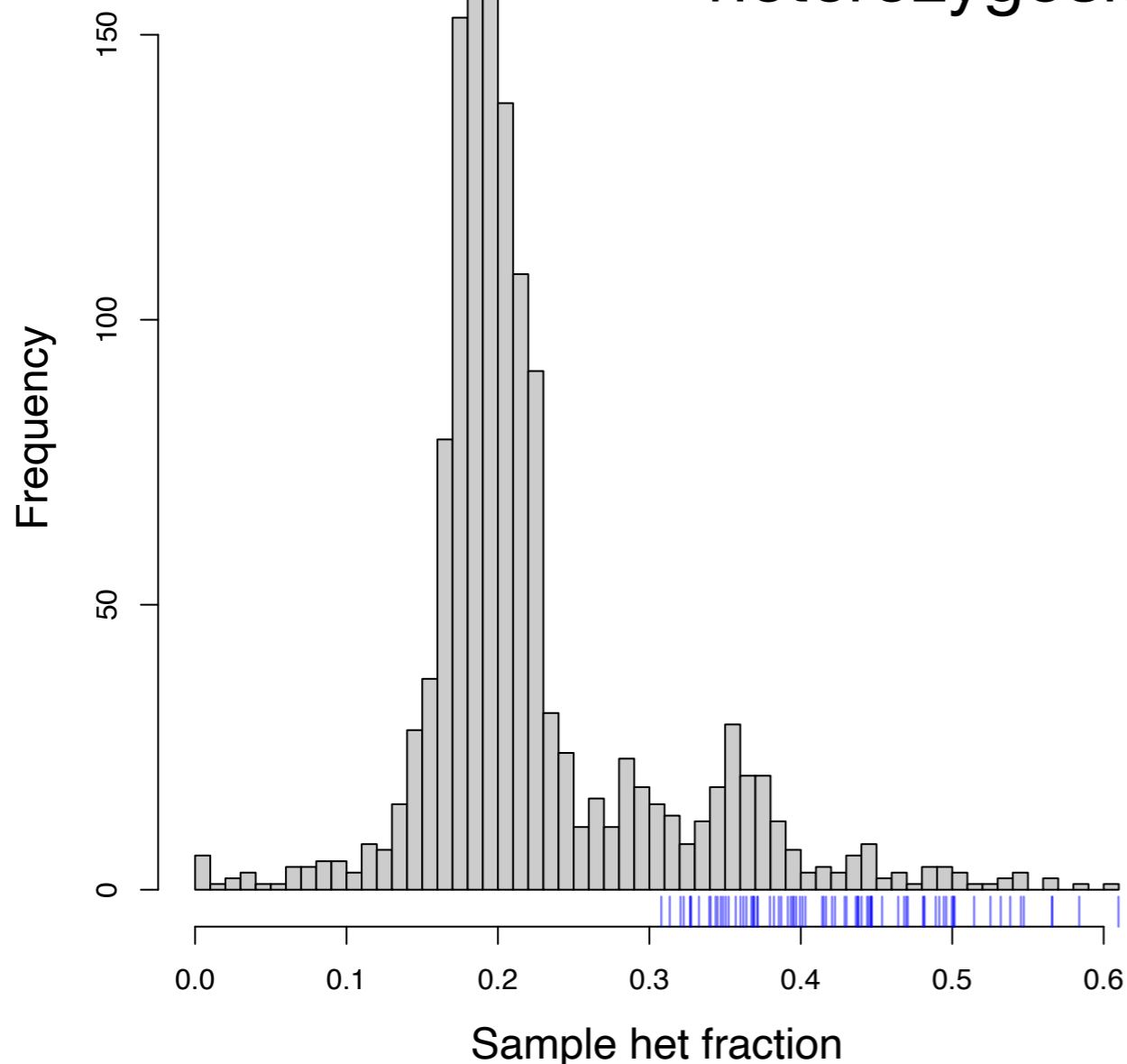


- $n = 986$  samples approx. identical (IBD2) to 35 distinct samples
- $n=102$  green horizontally/vertically "streaked" samples IBS1 with all other samples:
  - ➔ High heterozygosity

# Sample quality control

## Hybrid or contaminant?

- Streaked samples are uniformly higher heterozygosity

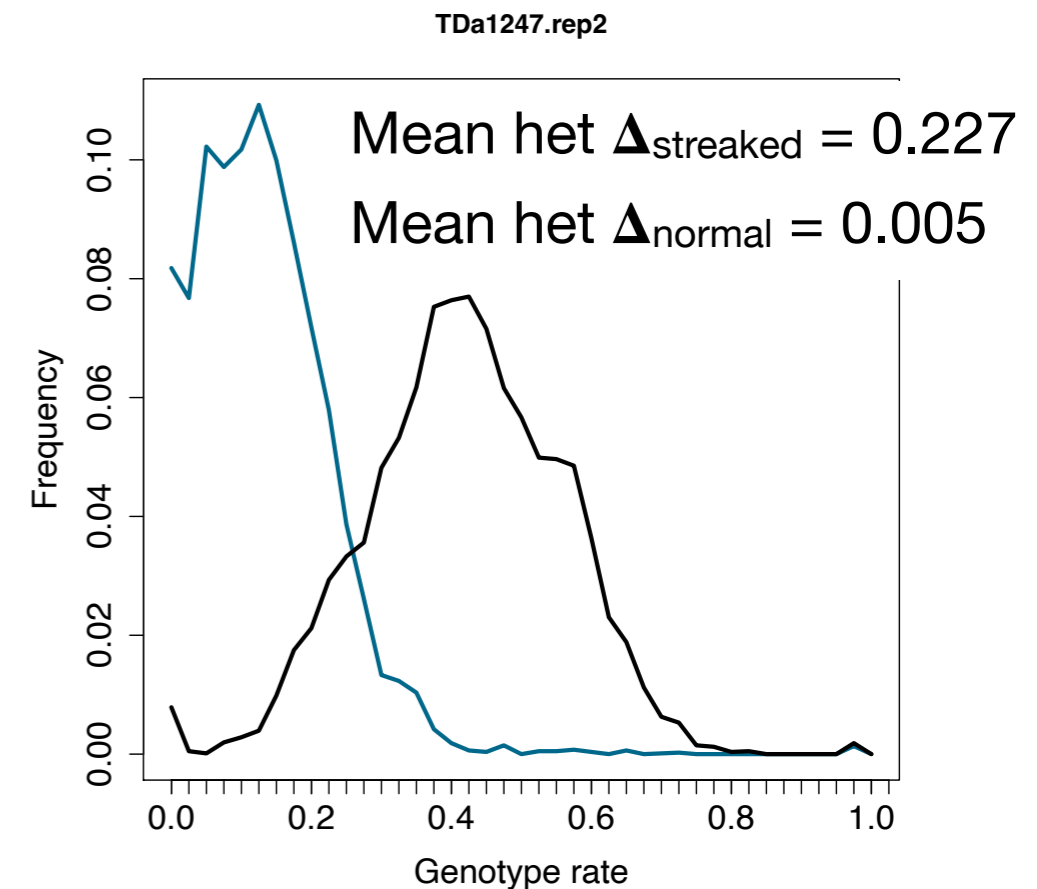
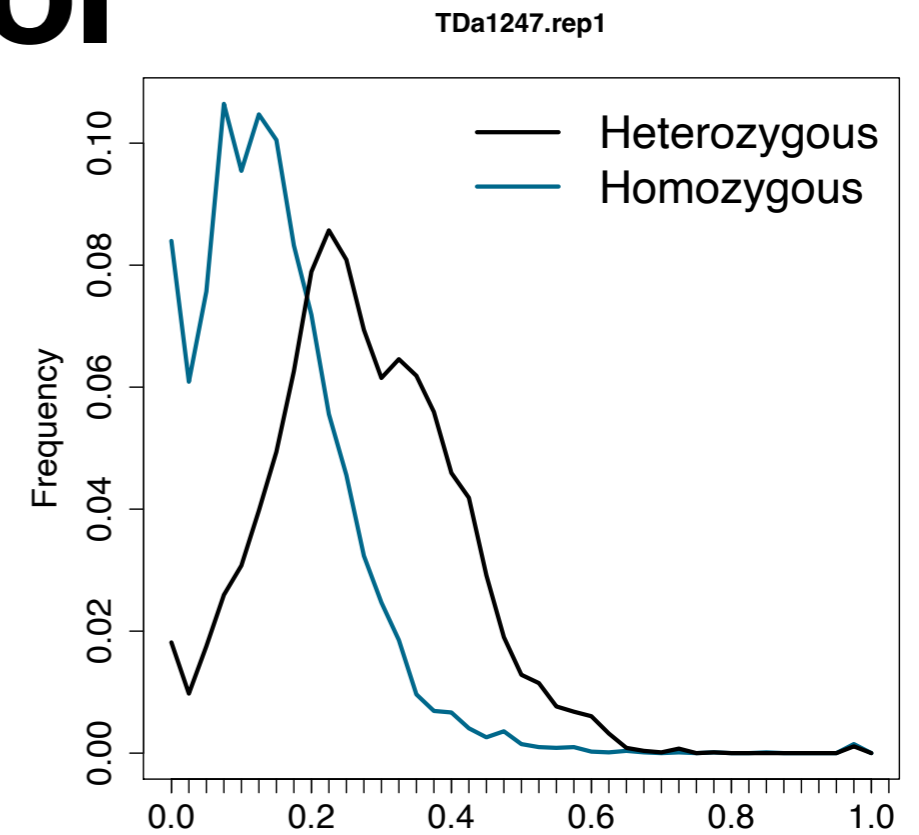
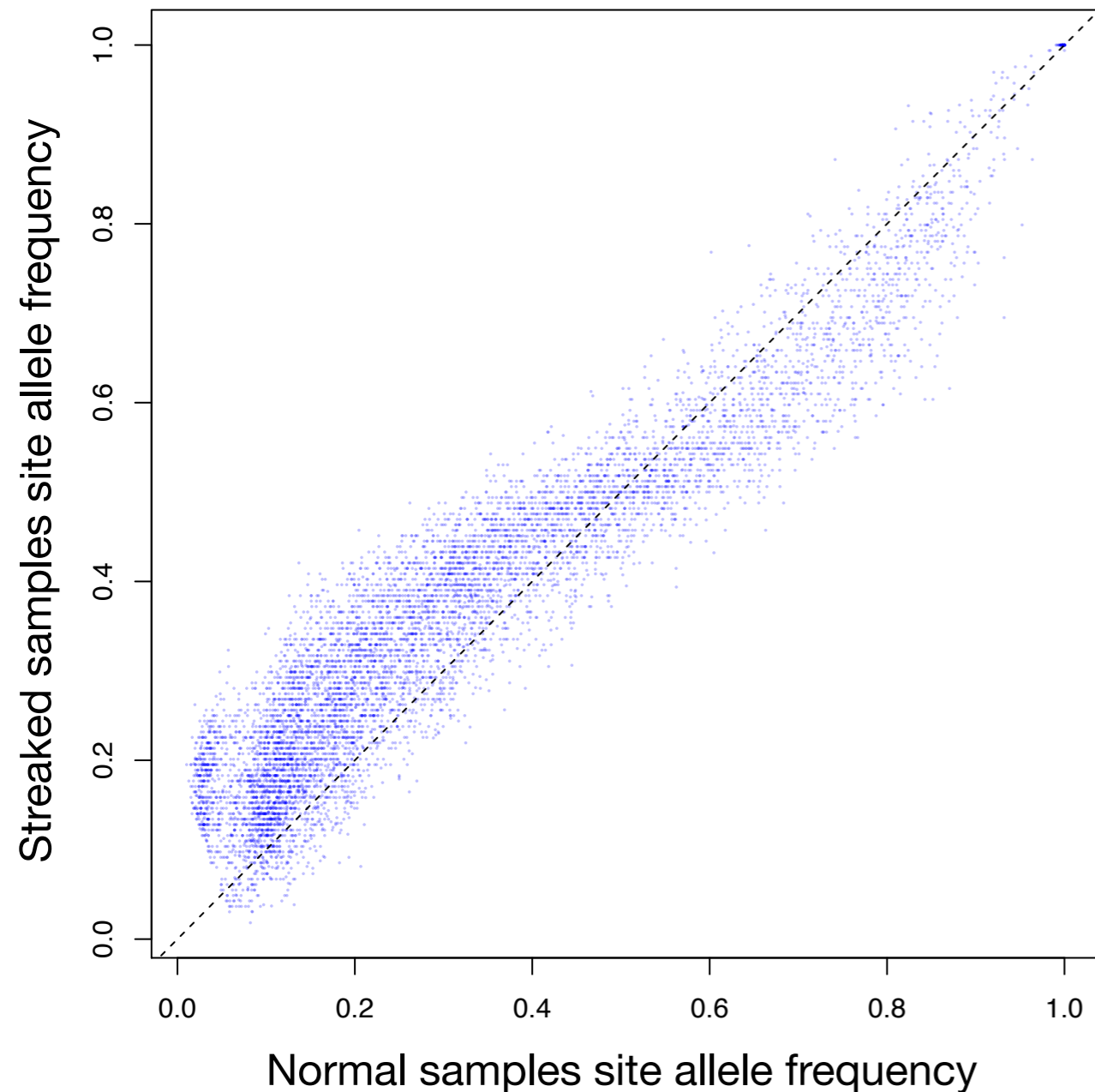


- Are these alleles:
  1. Shared with normal/non-streaked samples (i.e., the streaked samples are contaminants)?
  2. Private to/enriched in streaked samples (i.e., are they interspecific hybrid samples)?

# Sample quality control

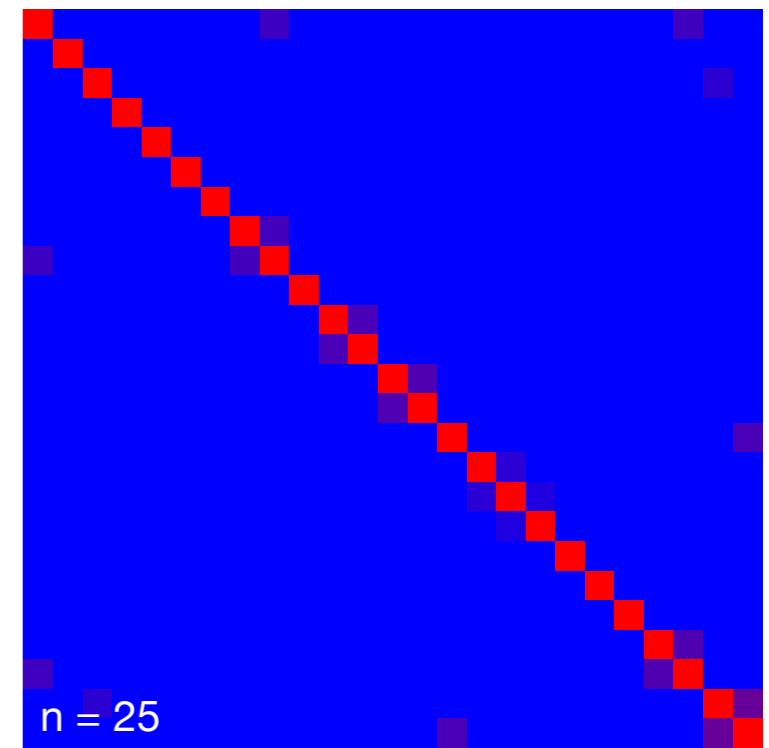
## Streaked samples are contaminated

- Streaked and normal samples overwhelmingly share alleles.
- Replicates of nominally-the-same samples exhibit different heterozygosities.



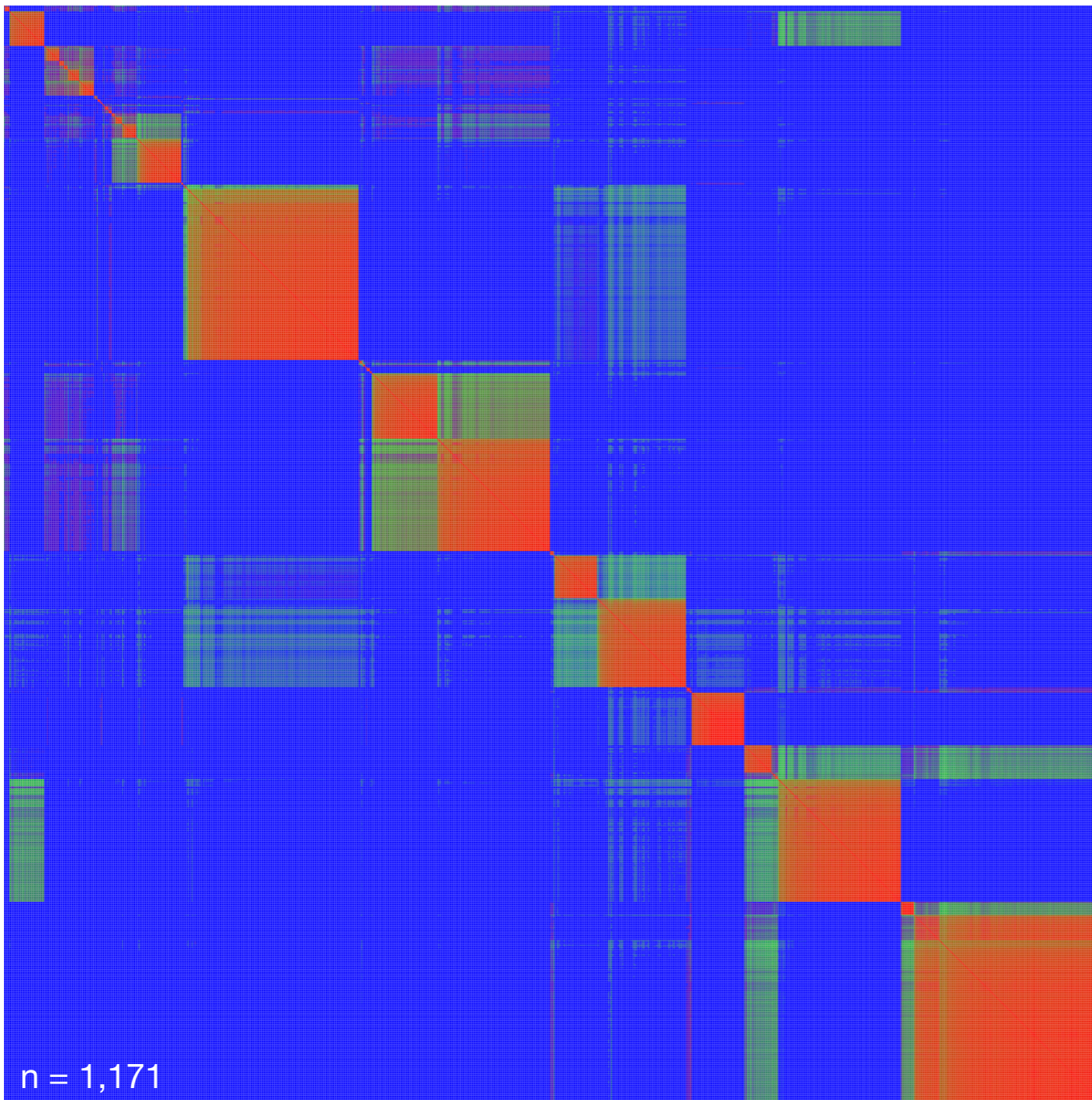
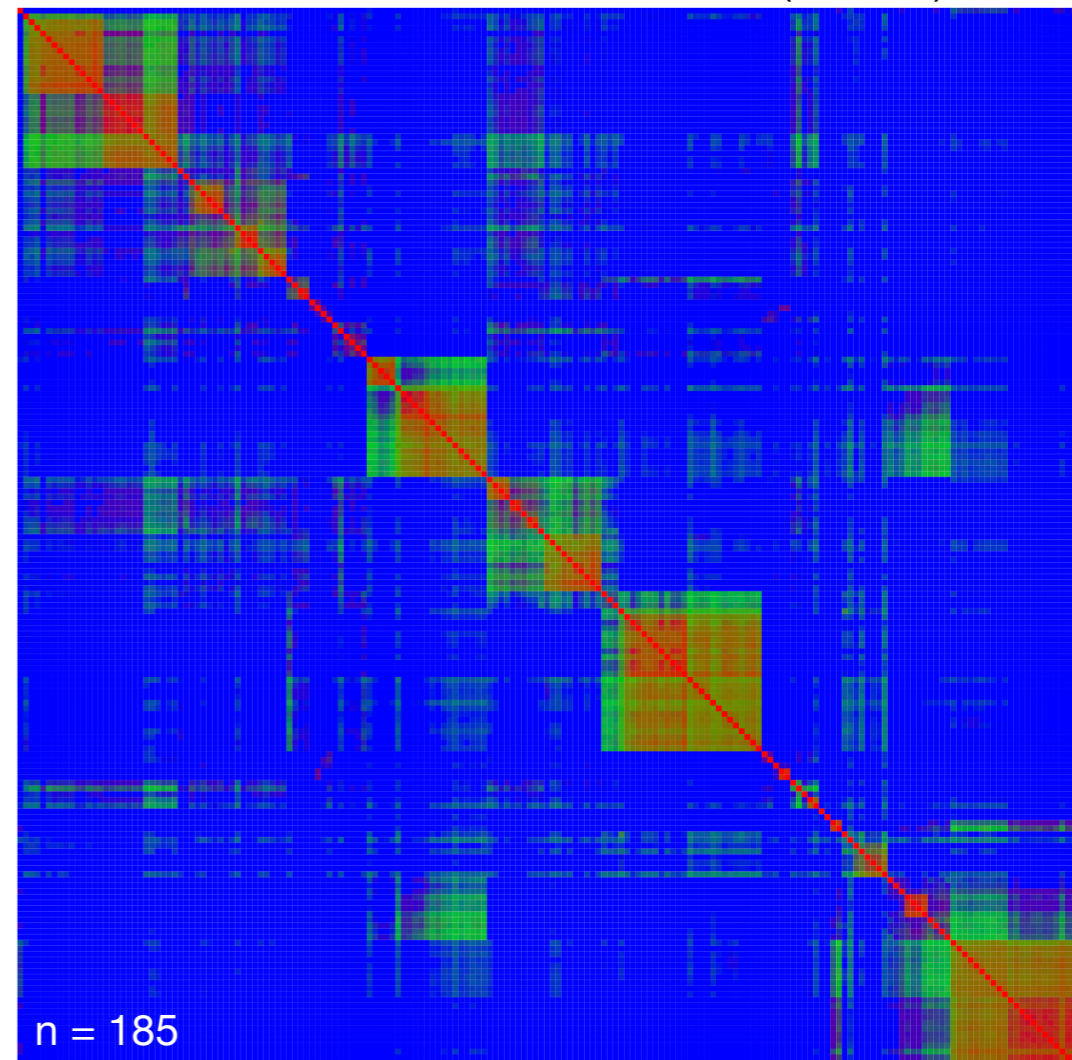
# Identifying (nearly) identical samples

Clones, replicates, heteronyms



↑ Remove  
 $\leq 1^\circ$  relatives  
(n = 160)

Remove  
identicals  
(n = 986)  
→



n = 1,171

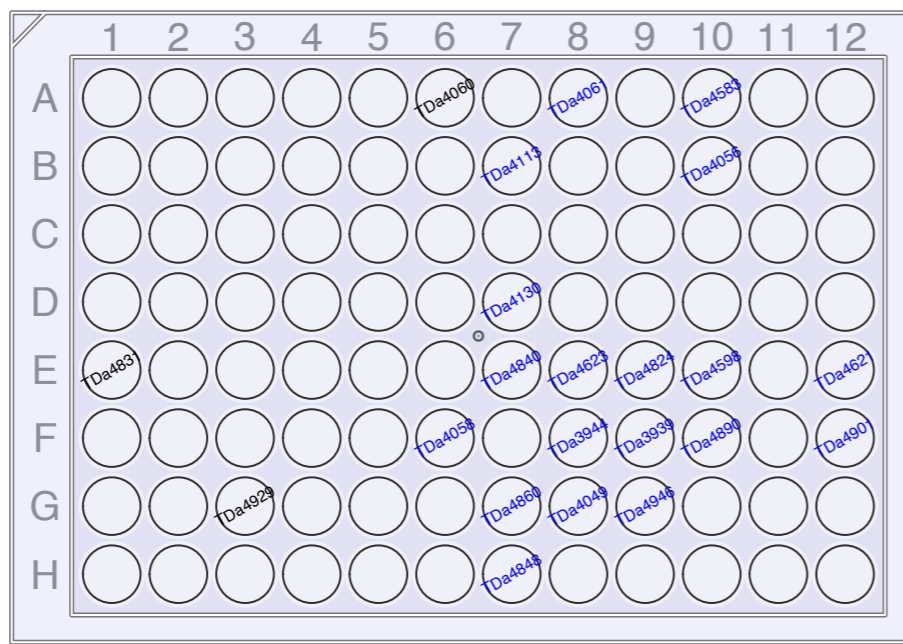
n = 185



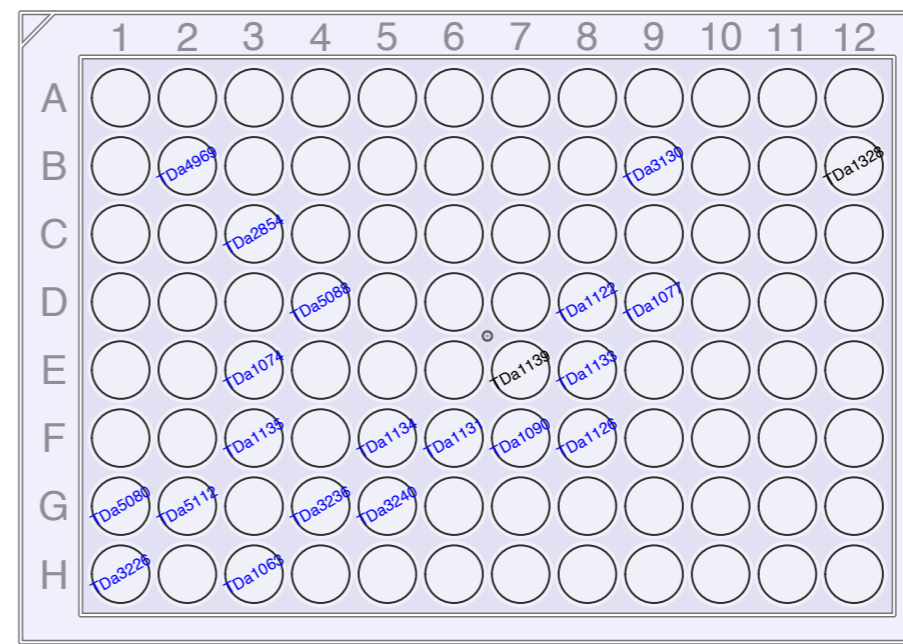
# Sample quality control summary

<b>Source report</b>	<b>Report description</b>	<b>Total Samples</b>	<b>Num. Contaminants</b>	<b>Num. Genotype Fail</b>	<b>Duplicates</b>	<b>Progenitors</b>
<b>DY17-2514</b>	Tanzania collection	82	2	8	65	2
<b>DY18-2579</b>	IITA collection	1032	96	58	740	14
<b>DY18-2634</b>	NRCRI collection	29	0	13	5	3
<b>DY18-2638</b>	IITA, Japan, Cote d'Ivoire collections	199	4	2	168	5
<b>DY19-2681</b>	Cameroon collection	20	0	8	8	1
<b>Total</b>	—	1362	102	89	986	25

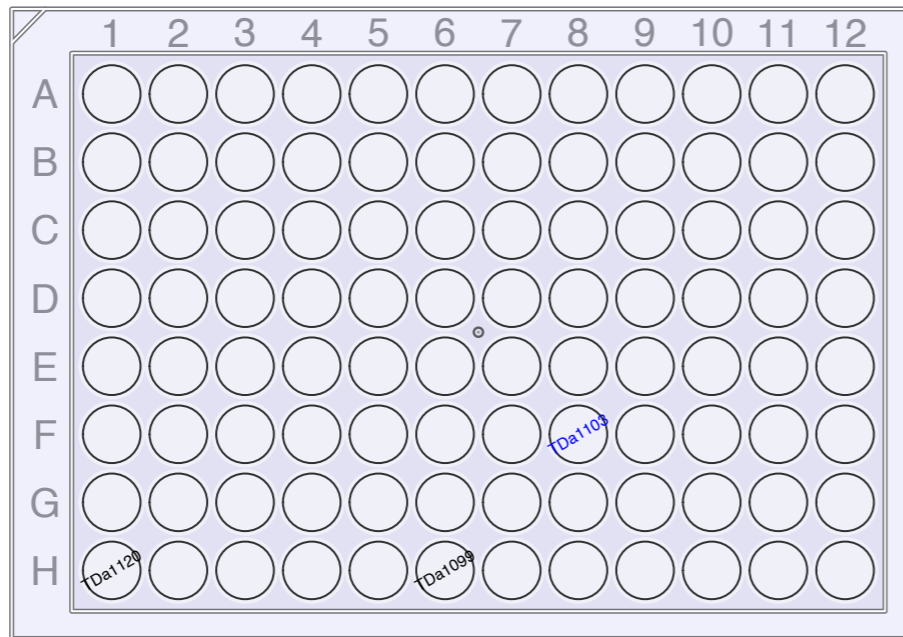
**Contaminated**  
**Genotyping fail**



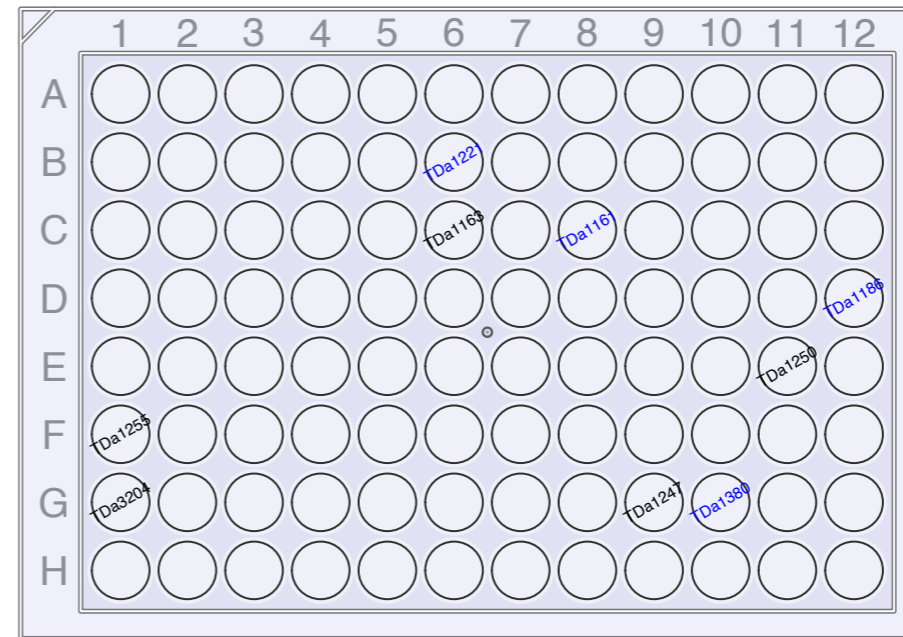
IITA DY18-2579  
Plate910418176001



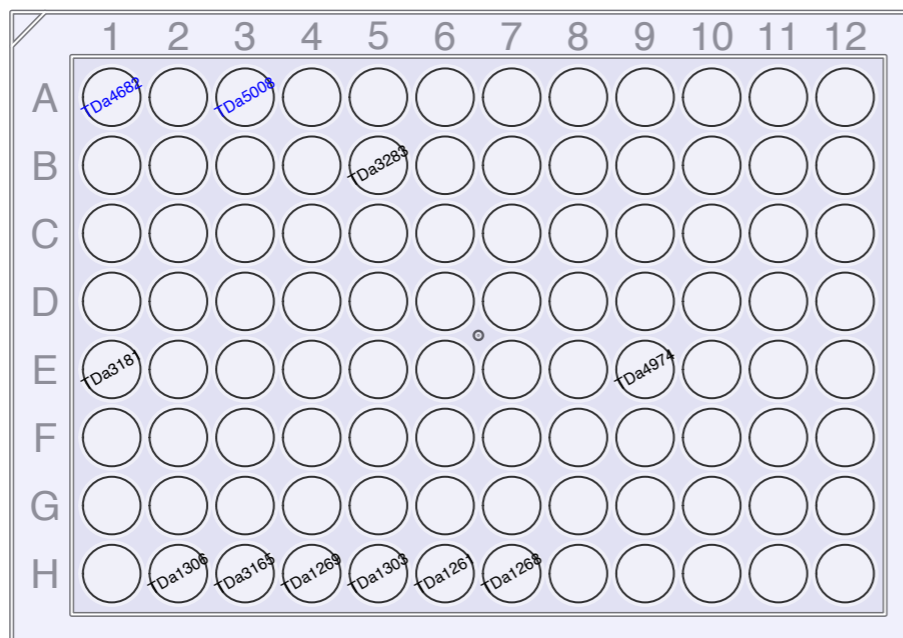
IITA DY18-2579  
Plate910418176004



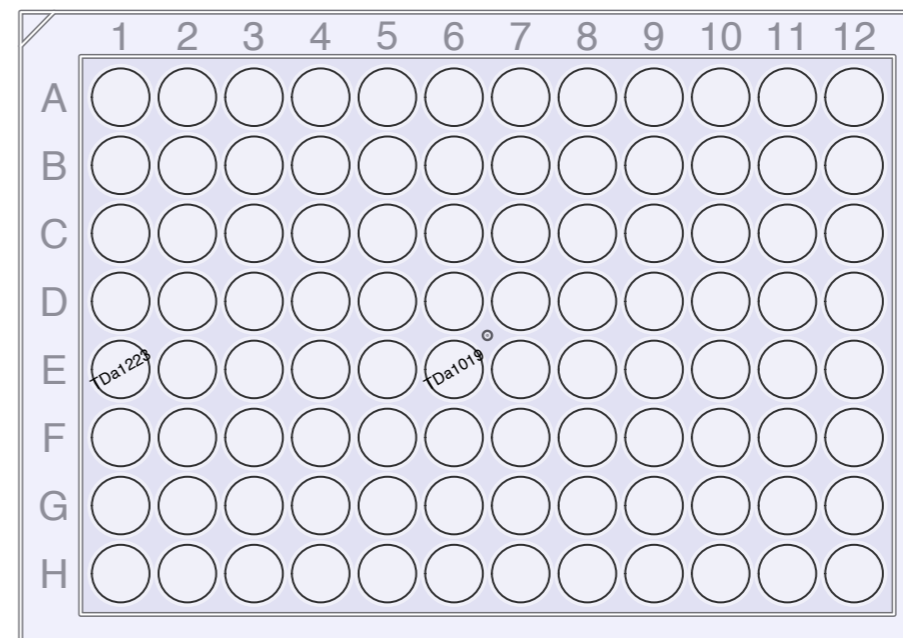
IITA DY18-2579  
Plate910418176002



IITA DY18-2579  
Plate910418176005

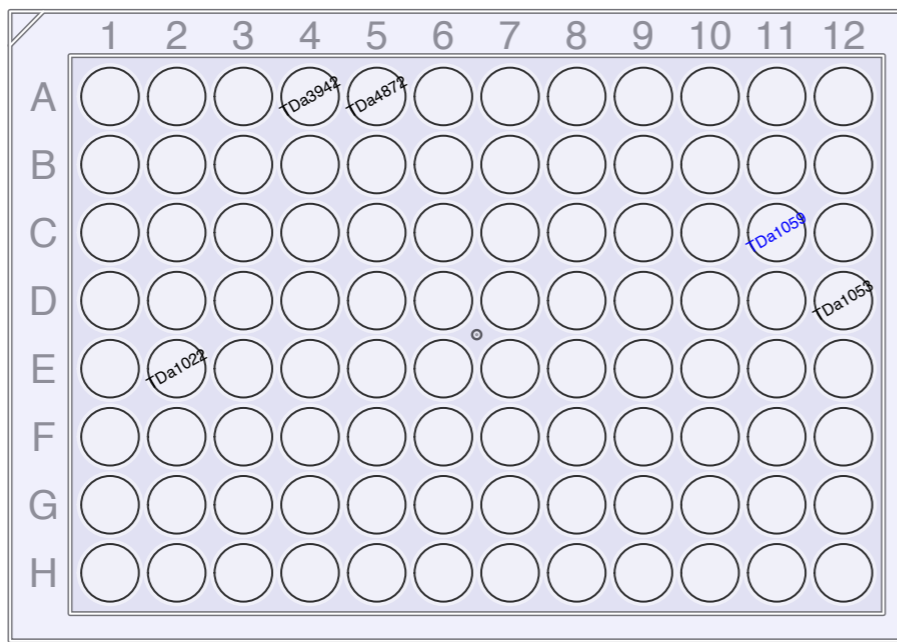


IITA DY18-2579  
Plate910418176003

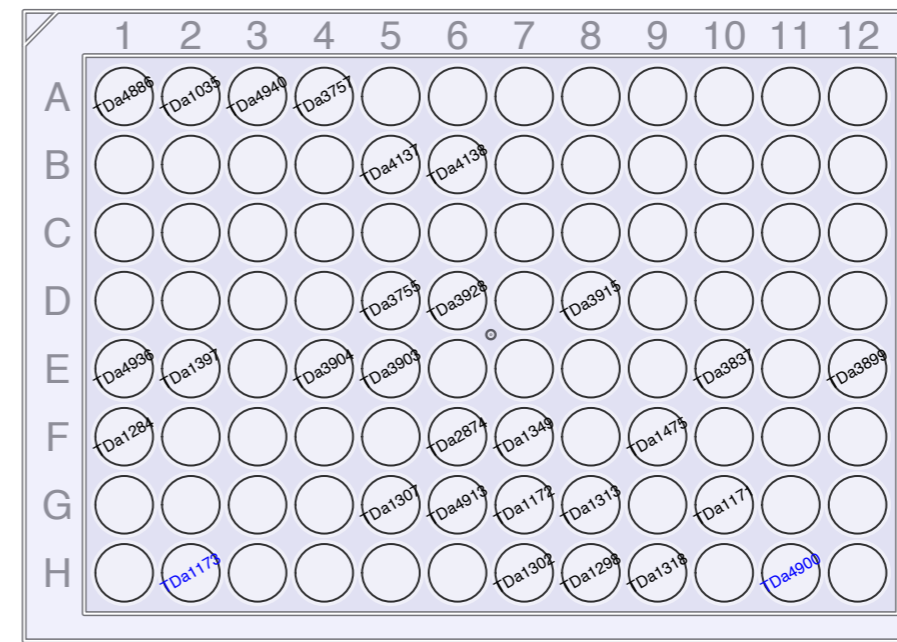


IITA DY18-2579  
Plate910418176006

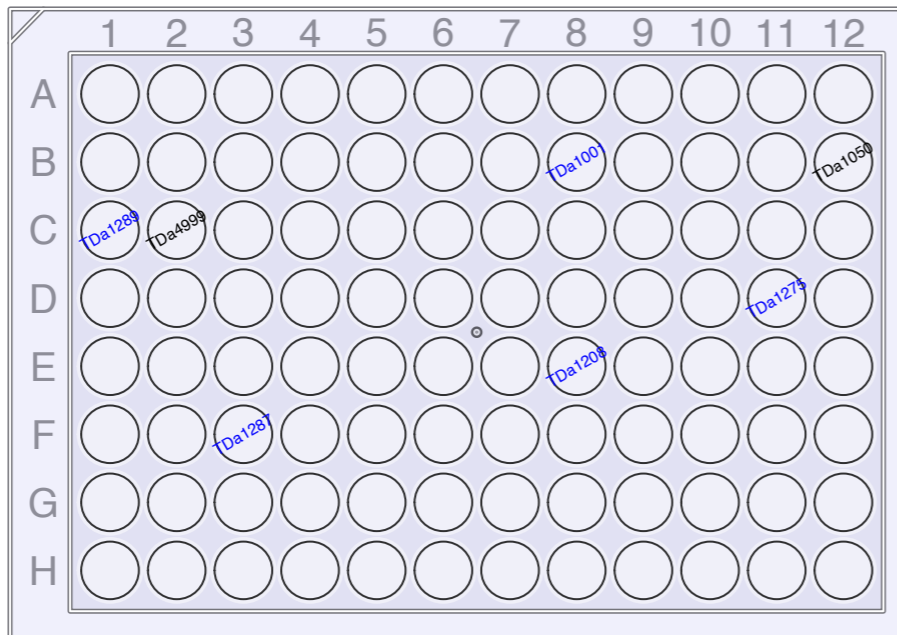
**Contaminated**  
**Genotyping fail**



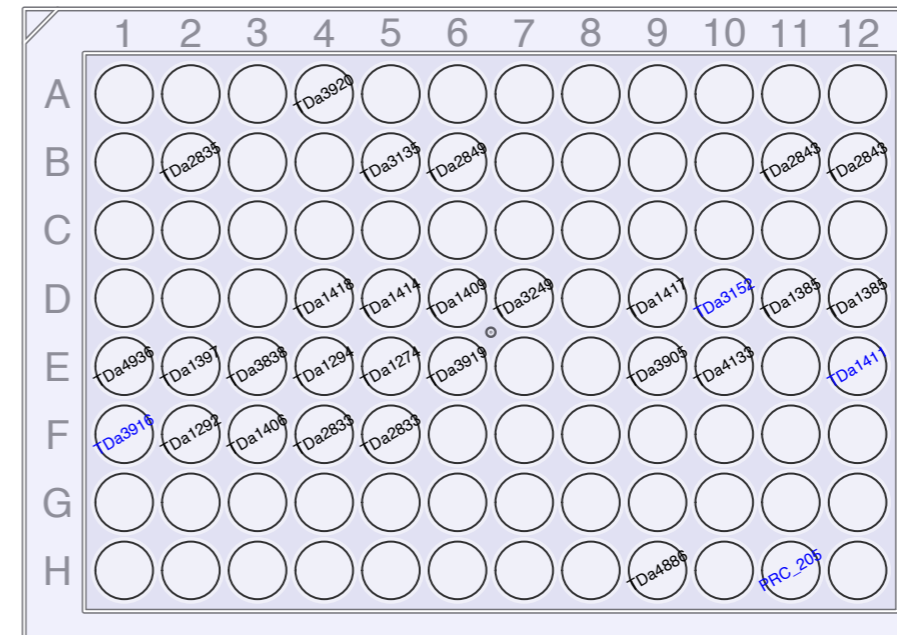
IITA DY18-2579  
Plate910418176007



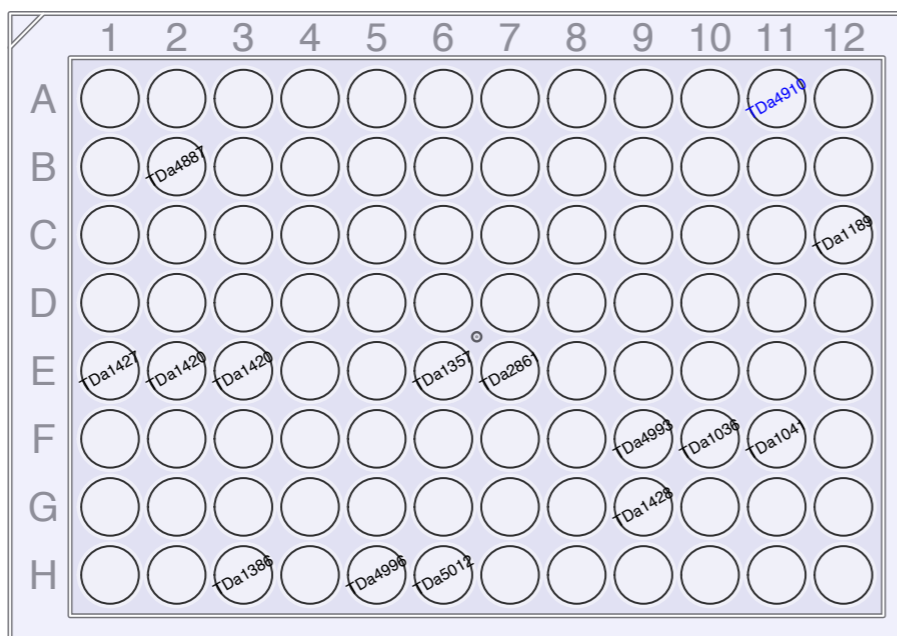
IITA DY18-2579  
Plate910418176010



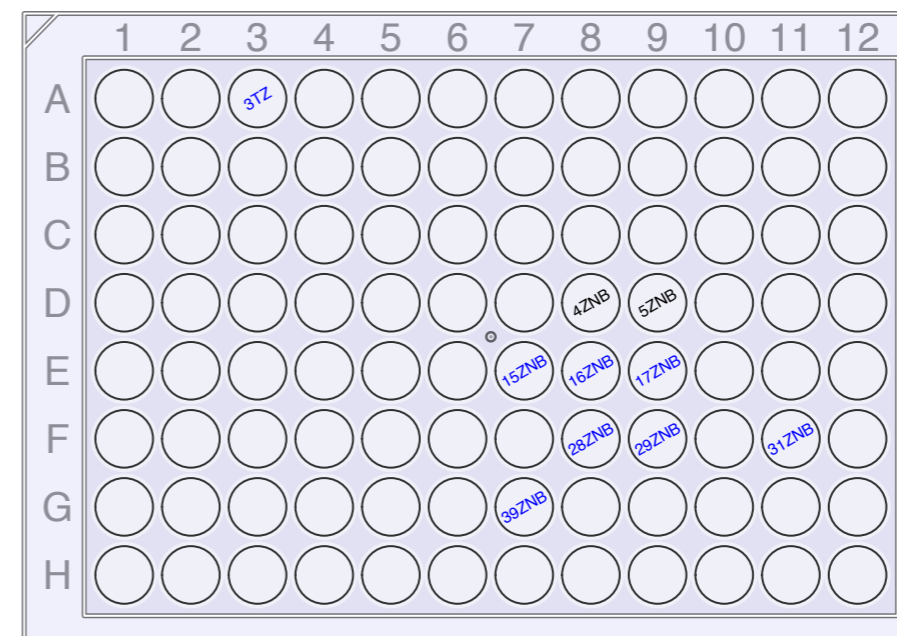
IITA DY18-2579  
Plate910418176008



IITA/VNM DY18-2579  
Plate910418176011

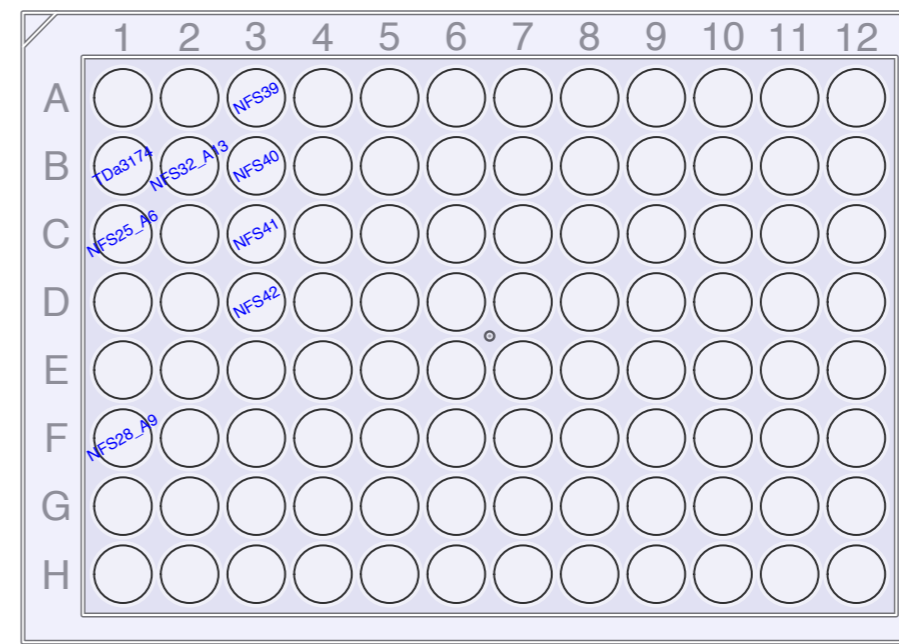
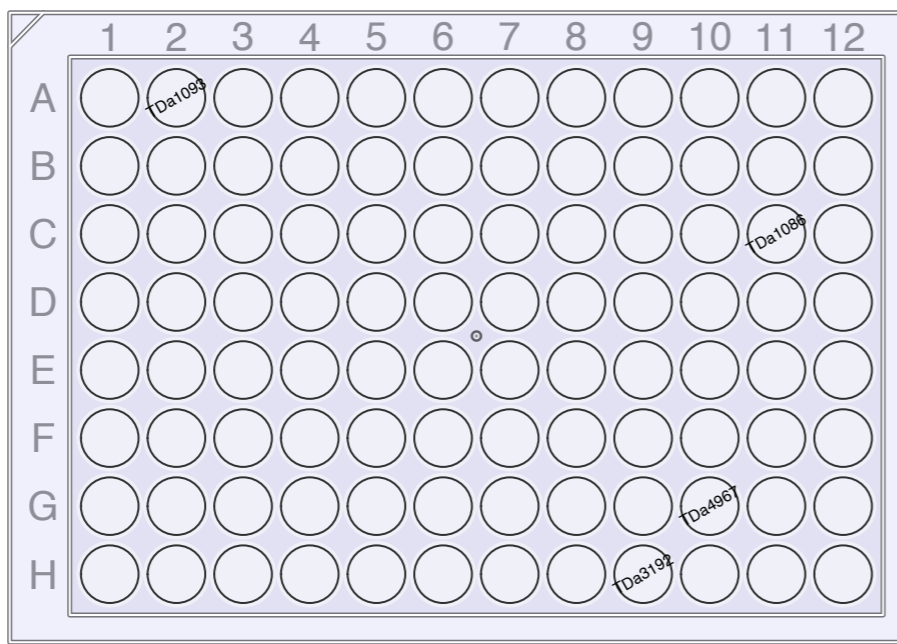


IITA DY18-2579  
Plate910418176009



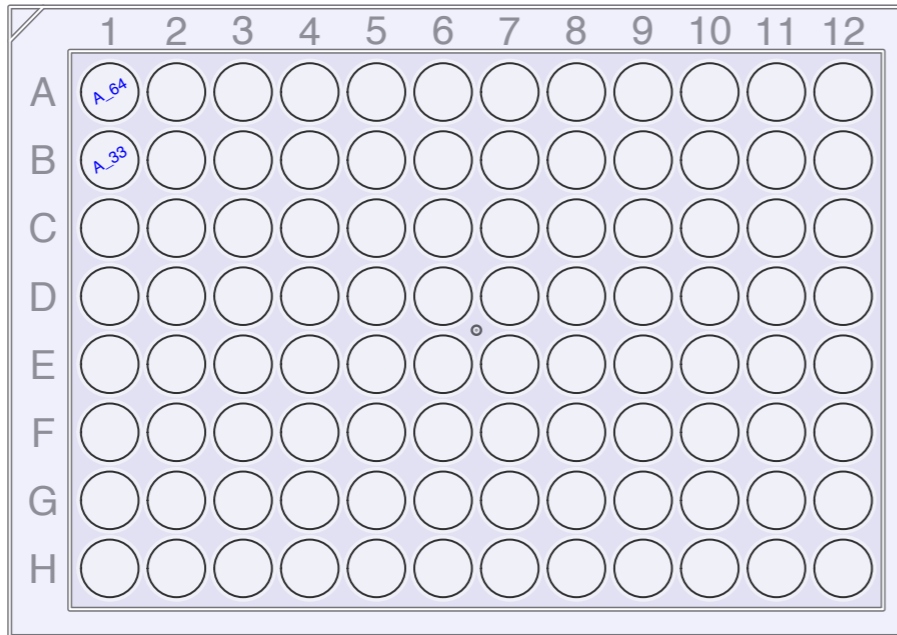
TZA DY17-2514  
Plate910918047001

**Contaminated**  
**Genotyping fail**

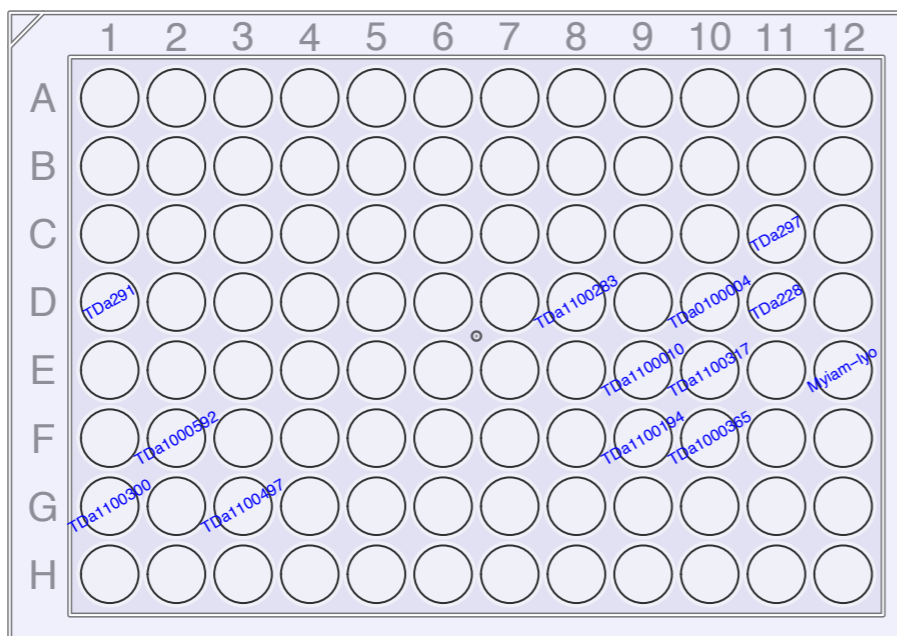


IITA/ETH/CMR  
DY19-2681  
Plate910919119002

CIV/JPN DY18-2638  
Plate910918340002



NRCRI DY18-2634  
Plate910918345002



# Chromosomes are well sampled

- 7,441 QC-passing loci @ MAF  $\geq 5\%$

