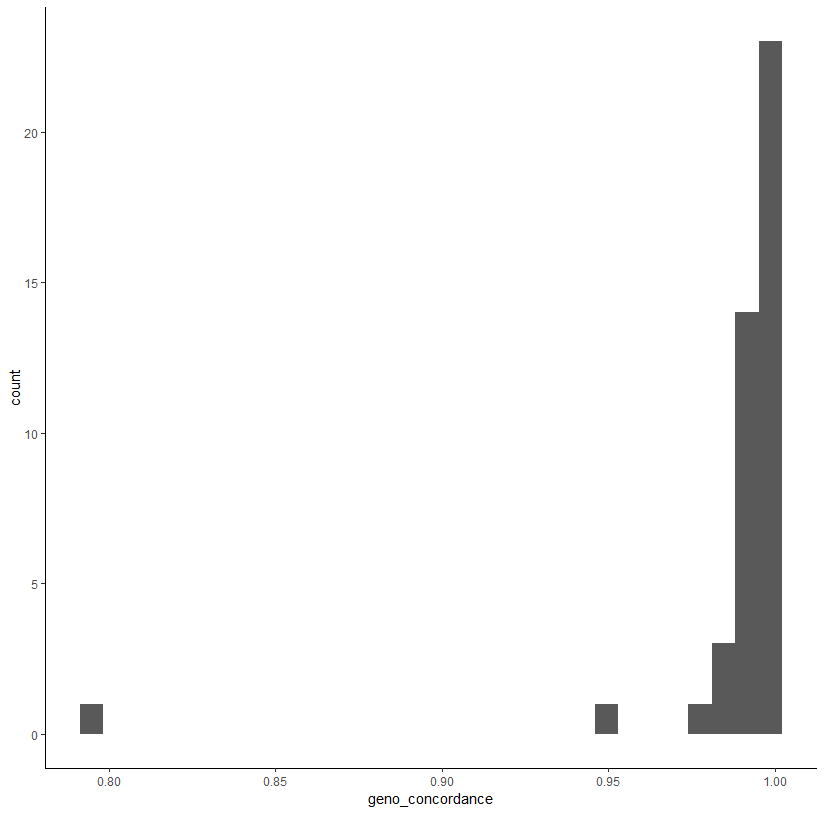
Initial dataset has 469 fastq files genotyped at 352 loci and a sex marker. 43 duplicate fastq files with repeat sample\_simple IDs are positive controls.

Look at concordance between 43 positive controls and their corresponding samples:

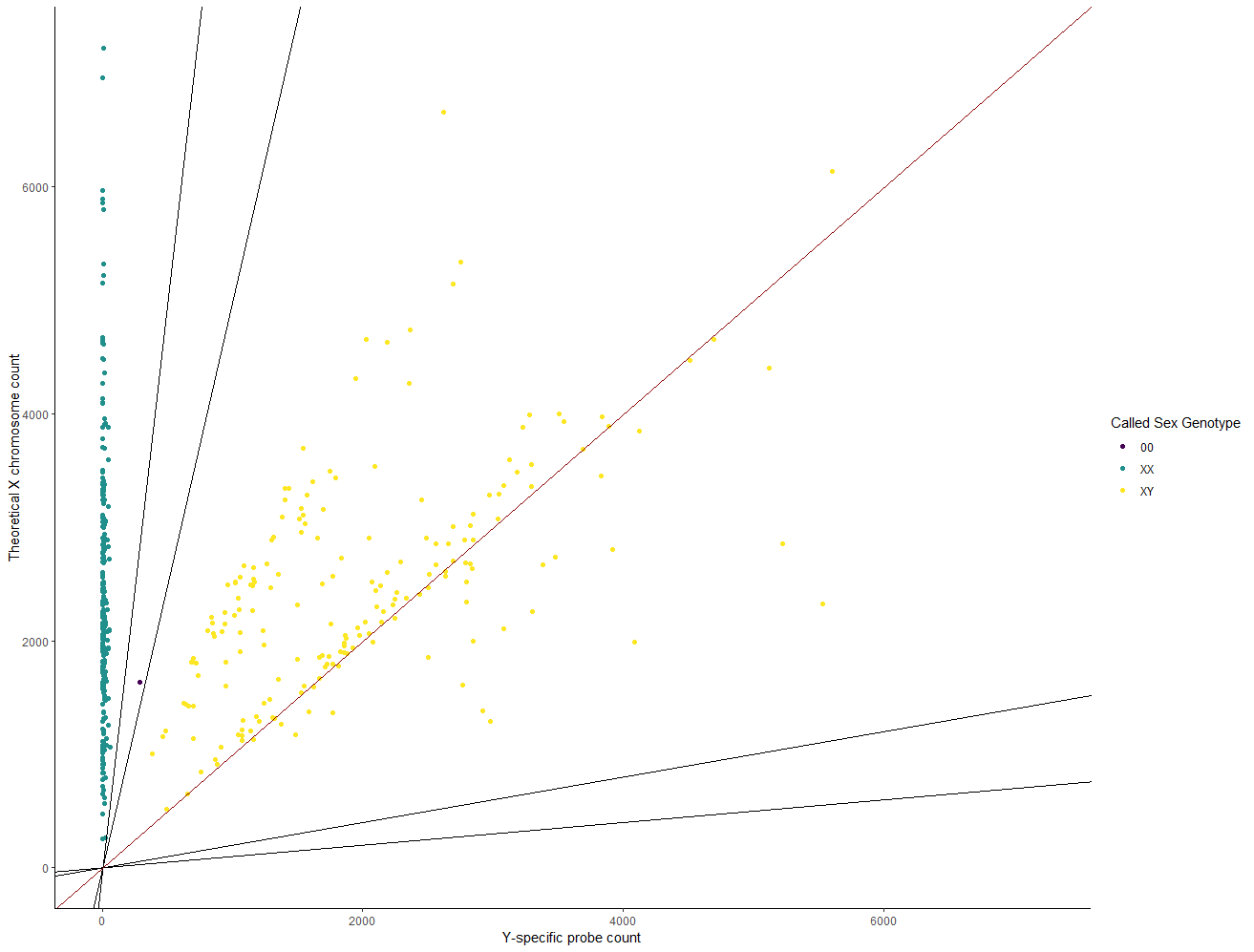


##concordance between quality controls and corresponding samples is high (0.8 - 1.0)##

Kept duplicate samples with the higher genotyping success. Removed duplicates with lesser genotyping success.

##dataset now has 426 unique samples genotyped at 352 loci and one sex marker##

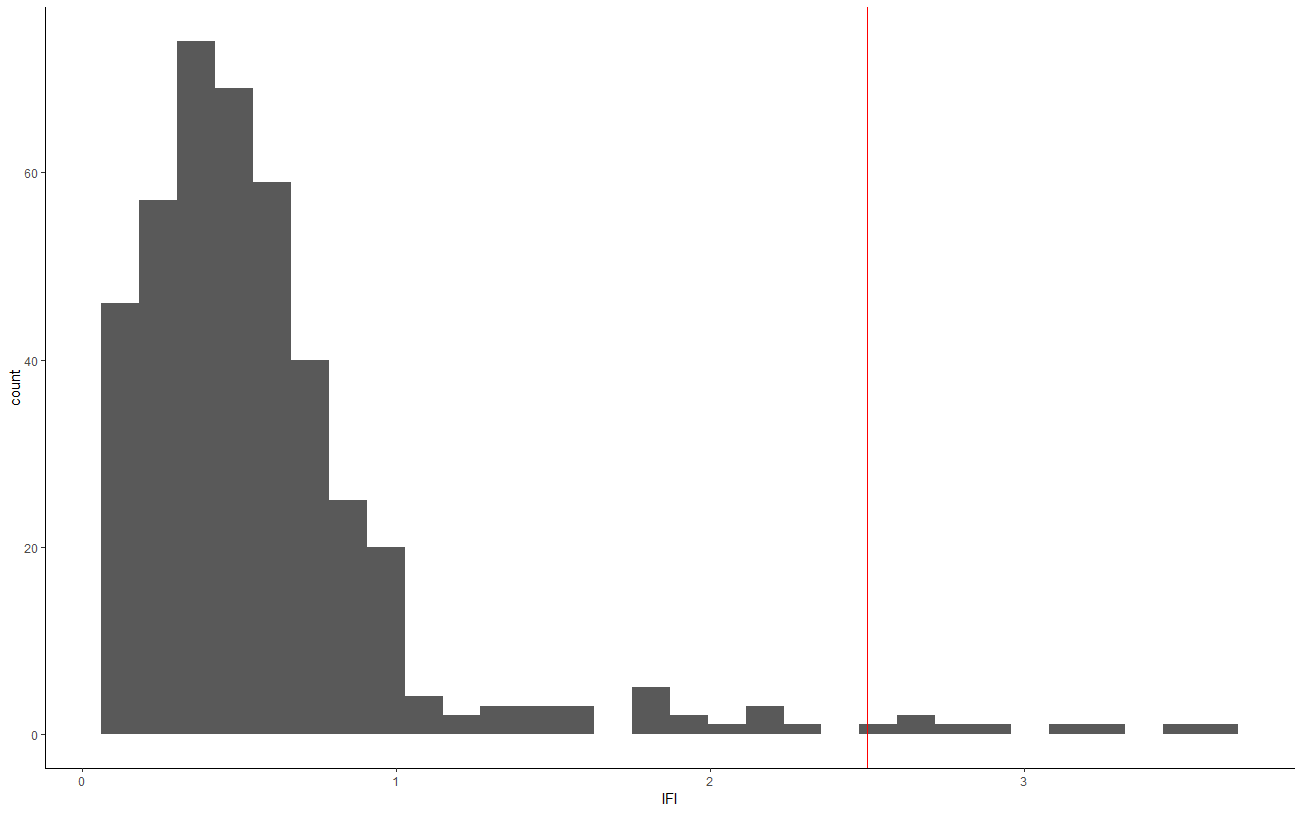
Check sex genotypes:



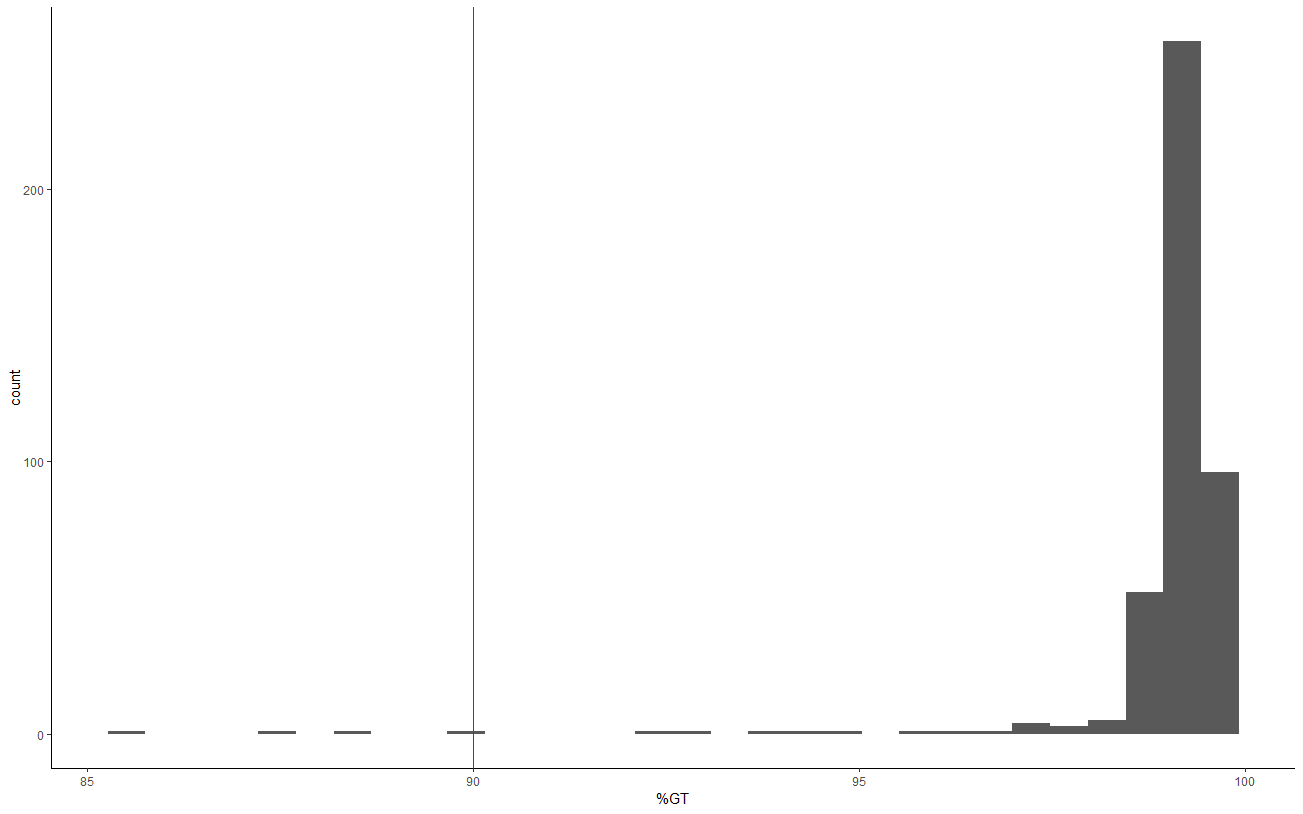
###Sex marker looks good. Leave as is without "correction" script##

Filter individuals/markers based on call rates/IFI scores:

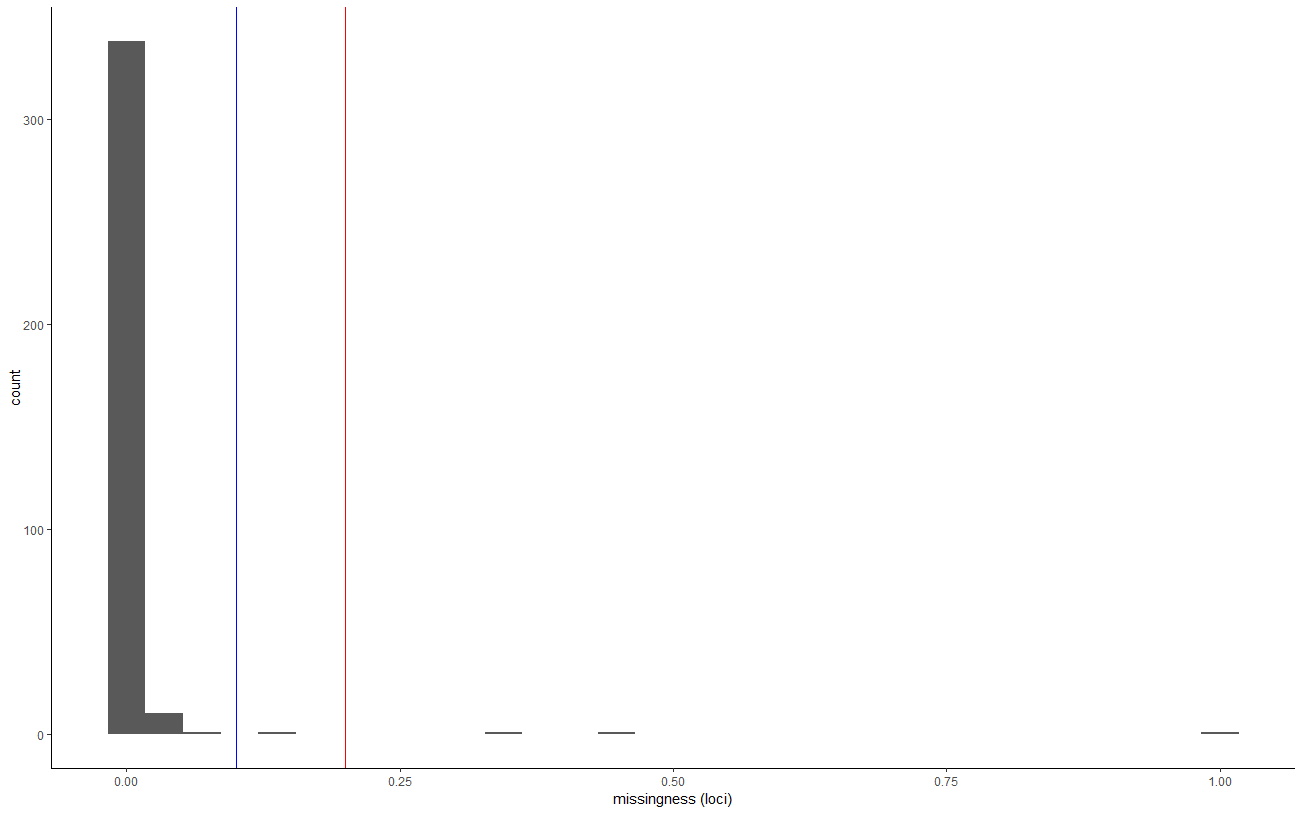
Initial IFI scores. Most ok, some poor scores.



Most individuals are well genotyped. A few with less than 90% call rates



Most loci are well genotyped. A few missing in more than 20% of individuals. One (Ots\_wenYhap\_33126) not genotyped in any individual.



Removed locus not genotyped in any individual. Recalculate individual call rates.

3 individuals have call rates below 90%

OtsAC22COQR\_0020 OtsAC23COQR\_0177 OtsAC23COQR\_0239

Removed these 3 individuals. Recalculate locus level call rates.

2 markers missing in more than 20% of individuals

(Ots17\_1066109\_C6 & Ots17\_1486479\_C6).

Removed these 2 markers. Recalculate IFI scores.

5 individuals have IFI scores greater than 2.5

(OtsAC22COQR\_0039 OtsAC22COQR\_0062 OtsAC22COQR\_0096 OtsAC22COQR\_0114 OtsAC23COQR\_0193)

Removed them.

##At this point the dataset has 418 Individuals and 349 loci and a sex marker##

##Individual call rates range 94 - 100%; Marker call rates range 86 - 100%

##IFI ranges 0.09 - 2.25##

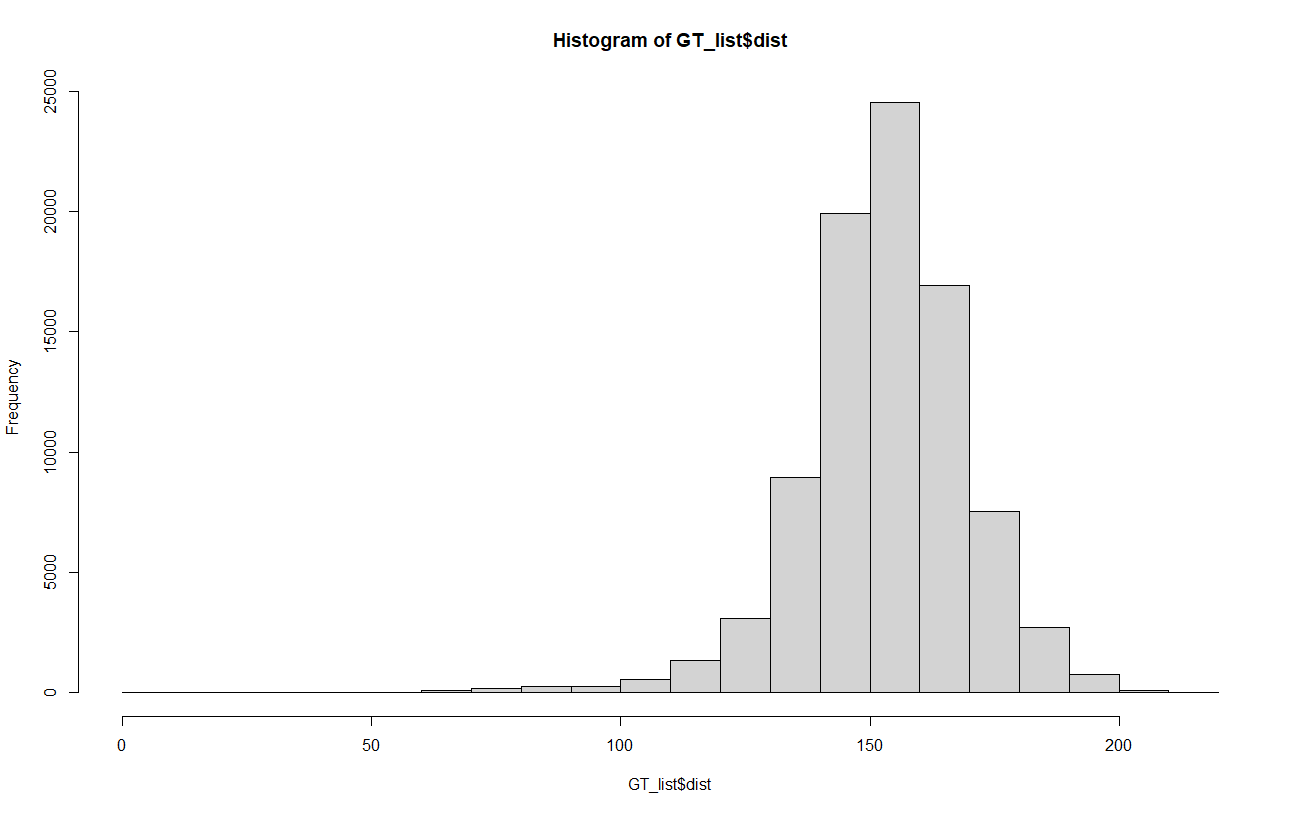
Evaluate for potential PSVs:

##6 markers are potential paralogs

"Ots\_wenYhap\_106664\_9", "Ots28\_11143508", "Ots37124-12281207", "Ots\_TLR3", "Ots\_afmid-196", "Ots19\_46172427"

Removed them.

Quantify pairwise genetic distance among all individuals to identify repeat sampling.



11 pairs of individuals have genetic distances below 50.

X1 X2 dist

4991 OtsAC22COQR\_0013 OtsAC22COQR\_0069 0

14029 OtsAC22COQR\_0037 OtsAC22COQR\_0068 0

27487 OtsAC22COQR\_0076 OtsAC22COQR\_0095 0

50313 OtsAC23COQR\_0038 OtsAC23COQR\_0055 0

56775 OtsAC23COQR\_0104 OtsAC23COQR\_0107 0

59659 OtsAC23COQR\_0116 OtsAC23COQR\_0117 0

62573 OtsAC23COQR\_0128 OtsAC23COQR\_1215 0

64951 OtsAC23COQR\_0139 OtsAC23COQR\_1217 0

67219 OtsAC23COQR\_0150 OtsAC23COQR\_1230 0

73473 OtsAC23COQR\_0186 OtsAC23COQR\_0202 0

76663 OtsAC23COQR\_0207 OtsAC23COQR\_1214 0

For each pair grabbed the sample with higher call rate.

16 monomorphic markers

Ots3\_57055518 Ots\_105401-325 Ots\_GH2 Ots\_IGF-I\_1-76 Ots\_IL8R\_C8 Ots\_IsoT Ots\_NFYB-147 Ots\_crRAD23631-48 Ots\_crRAD26081-28 Ots\_crRAD46751-42 Ots\_tpx2-125 Ots\_txnip-321 Ots\_u07-64\_221 Ots\_wenYhap\_25067\_92 Ots\_wenYhap\_71572 Ots\_zn593-346

At this point the dataset has 407 individuals (2022 n = 106 ; 2023 n = 301) and 327 markers and a sex marker.

Depth summary:

mean median sd

1684 977 2181