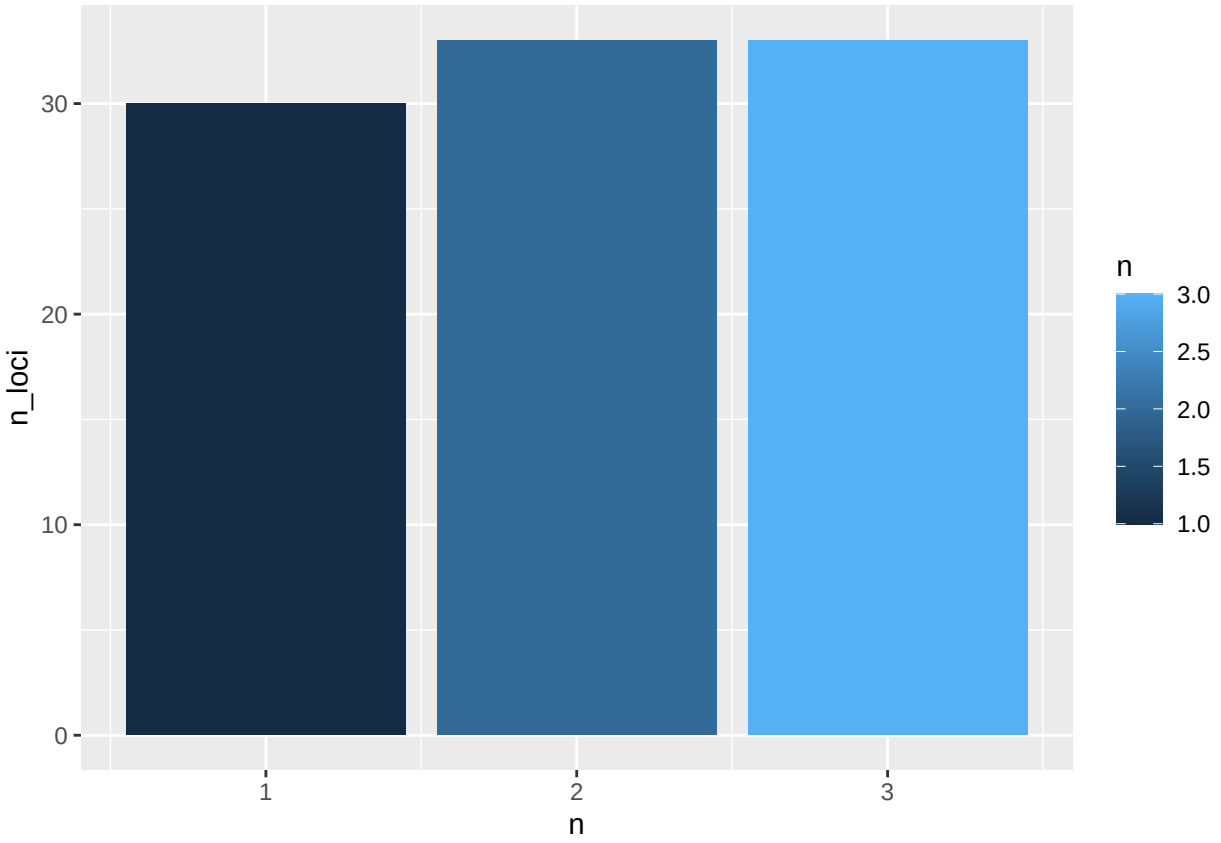
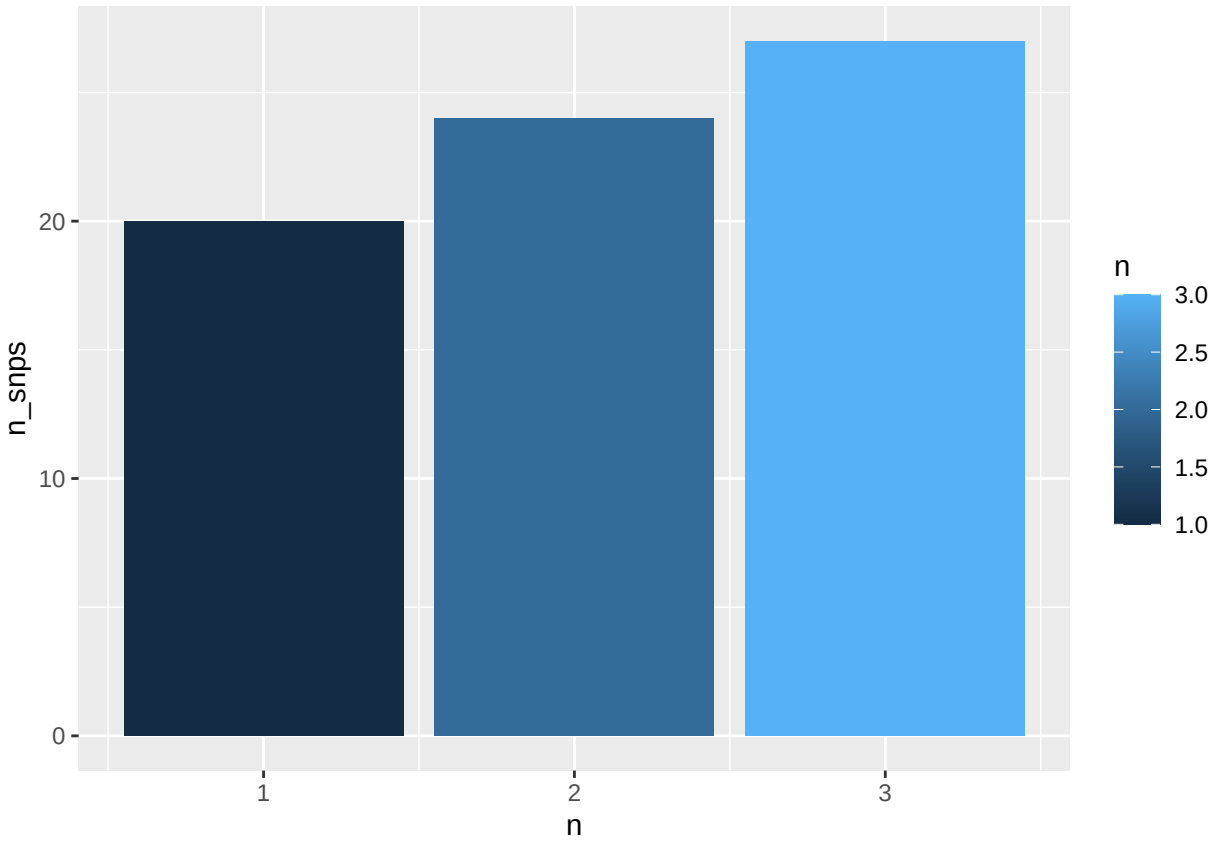
Report

César Emiliano Escalona Prado

This exercise tested the effect of different values in the distance between catalog loci (n). The test was performed using values 1, 2 and 3. The variable n acts in the cstacks program, it uses the locis generated from the ustacks, which it merges into a catalog that must contain all the loci and alleles of the population. In the case that this parameter is set too low, there will be loci between the individuals that are independently represented in the catalog that are actually the same locus. Setting this parameter too high will allow loci to join together in the sequence space to chain together and create large and wrong loci in the catalog.

M	n	m	n_loci	n_snps
2	1	3	30	20
2	2	3	33	24
2	3	3	33	27

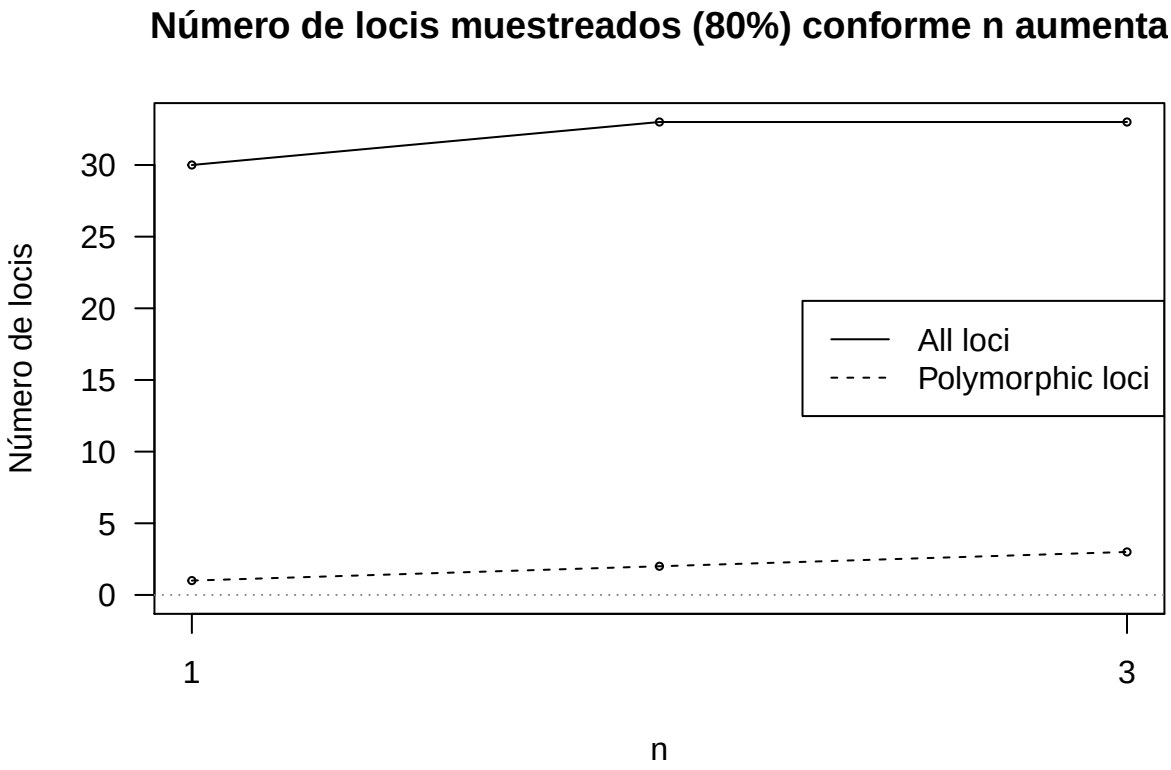


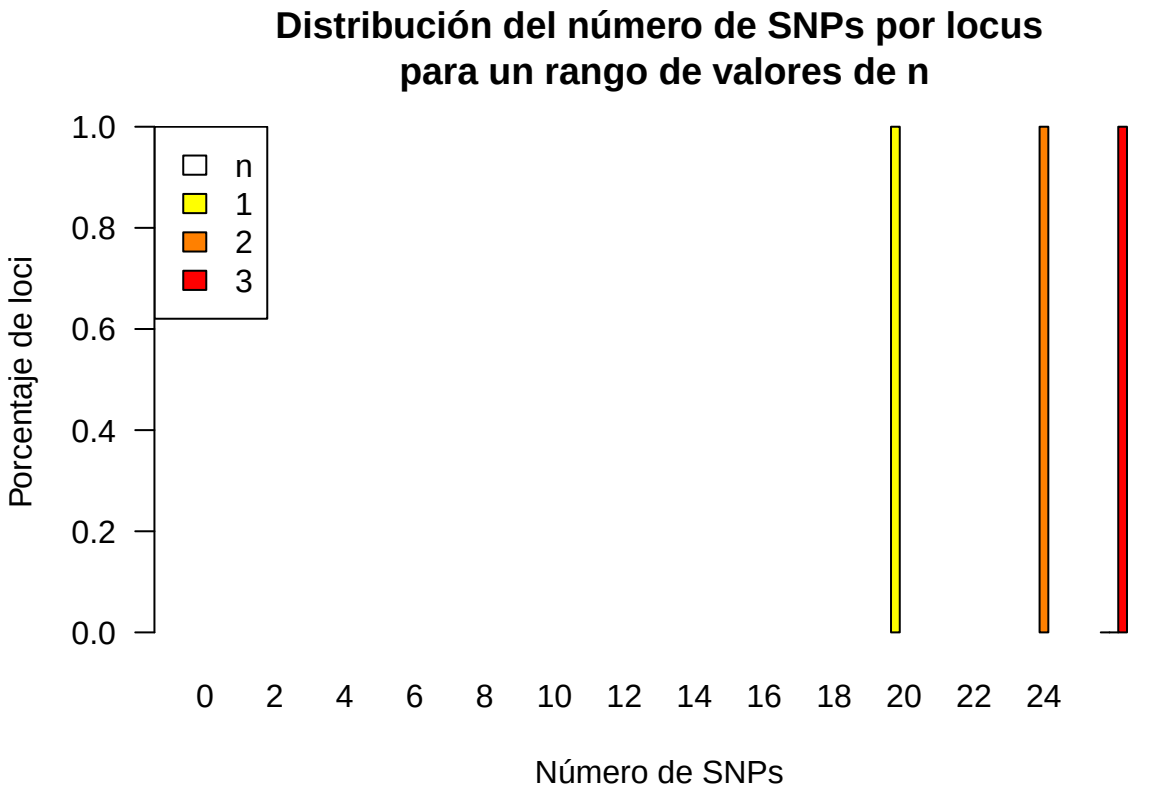



```
d= read.csv('./n_snps_per_locus.tsv') # call a variable with the data
#generate a ggplot object to show the number of loci related to n values (the plot differentiates betwe
plot(NULL,
    xlim=range(d$n),
    ylim=range(c(0, d$n_loci)),
    xlab='n',
    ylab='Número de locis',
    main='Número de locis muestreados (80%) conforme n aumenta',
    xaxt='n',
    las=2
    )
abline(h=0:20*5000, lty='dotted', col='grey50')
axis(1, at=c(1,3,5,7,9))
legend('right', c('All loci', 'Polymorphic loci'), lty=c('solid', 'dashed'))

# Total number of loci.
lines(d$n, d$n_loci)
points(d$n, d$n_loci, cex=0.5)

# Number of polymorphic loci.
lines(d$n, d$n_loci_poly, lty='dashed')
points(d$n, d$n_loci_poly, cex=0.5)
```





Results

n1) Removed 50508 loci that did not pass sample/population constraints from 50538 loci. Kept 30 loci, composed of 2852 sites; 2 of those sites were filtered, 20 variant sites remained. Mean genotyped sites per

locus: 95.07bp

n2) Removed 49225 loci that did not pass sample/population constraints from 49258 loci. Kept 33 loci, composed of 3146 sites; 2 of those sites were filtered, 24 variant sites remained. Mean genotyped sites per locus: 95.33bp

n3) Removed 48511 loci that did not pass sample/population constraints from 48544 loci. Kept 33 loci, composed of 3146 sites; 2 of those sites were filtered, 27 variant sites remained. Mean genotyped sites per locus: 95.33bp

Using n = 3 gives the greatest number of variant sites, keeping the greatest number of loci