Introduction to NGS data: experimental design

Matteo Fumagalli

Experimental design

You discovered a new species!

Experimental design

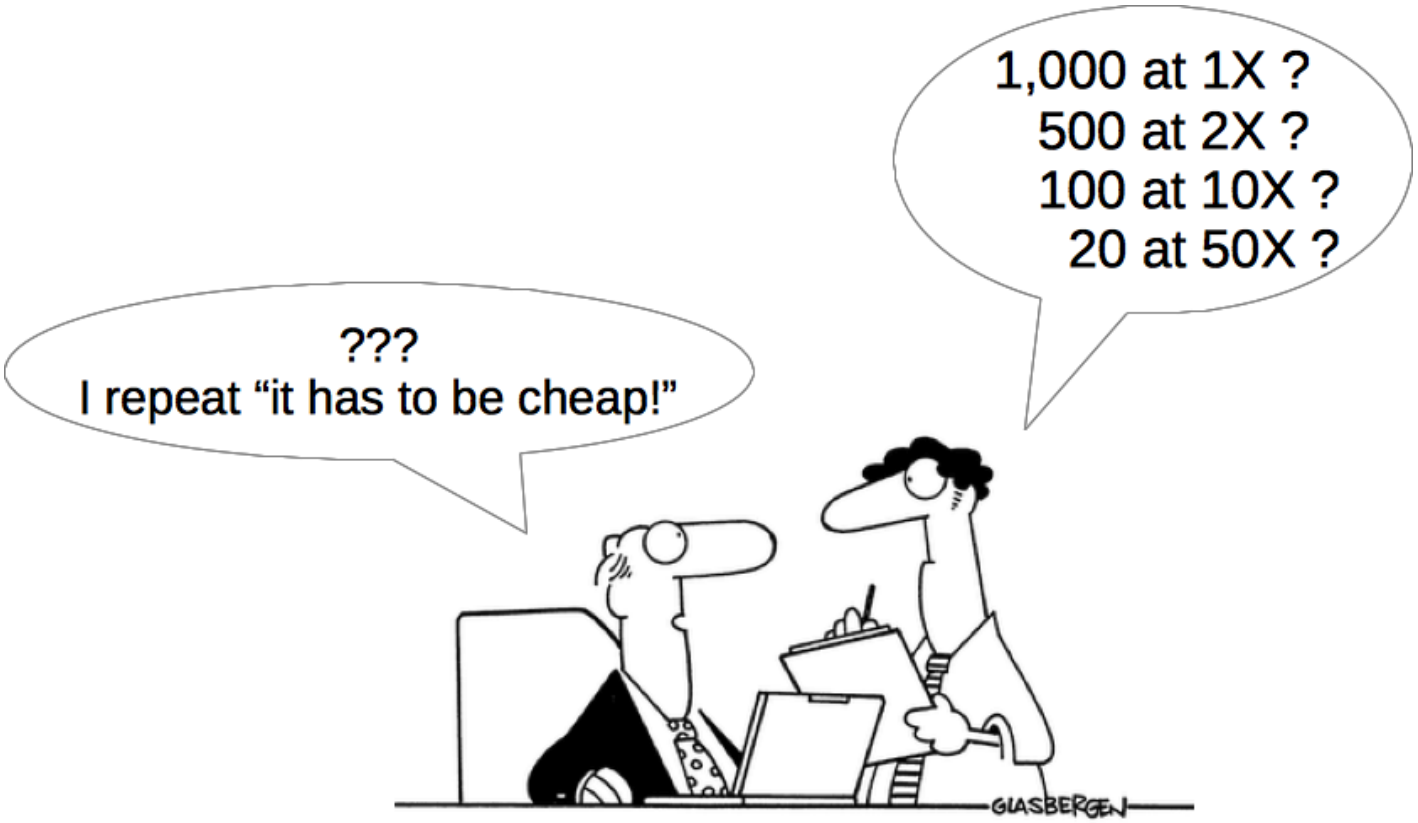
You have access to many samples!

Experimental design

Experimental design

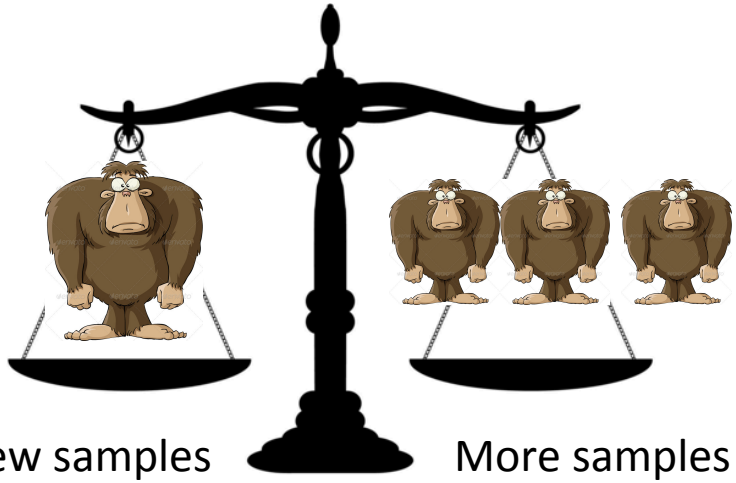
Experimental design

Experimental design



Experimental design

At a fixed budget…



Few samples
at higher quality

More samples
at lower quality

Simulations design

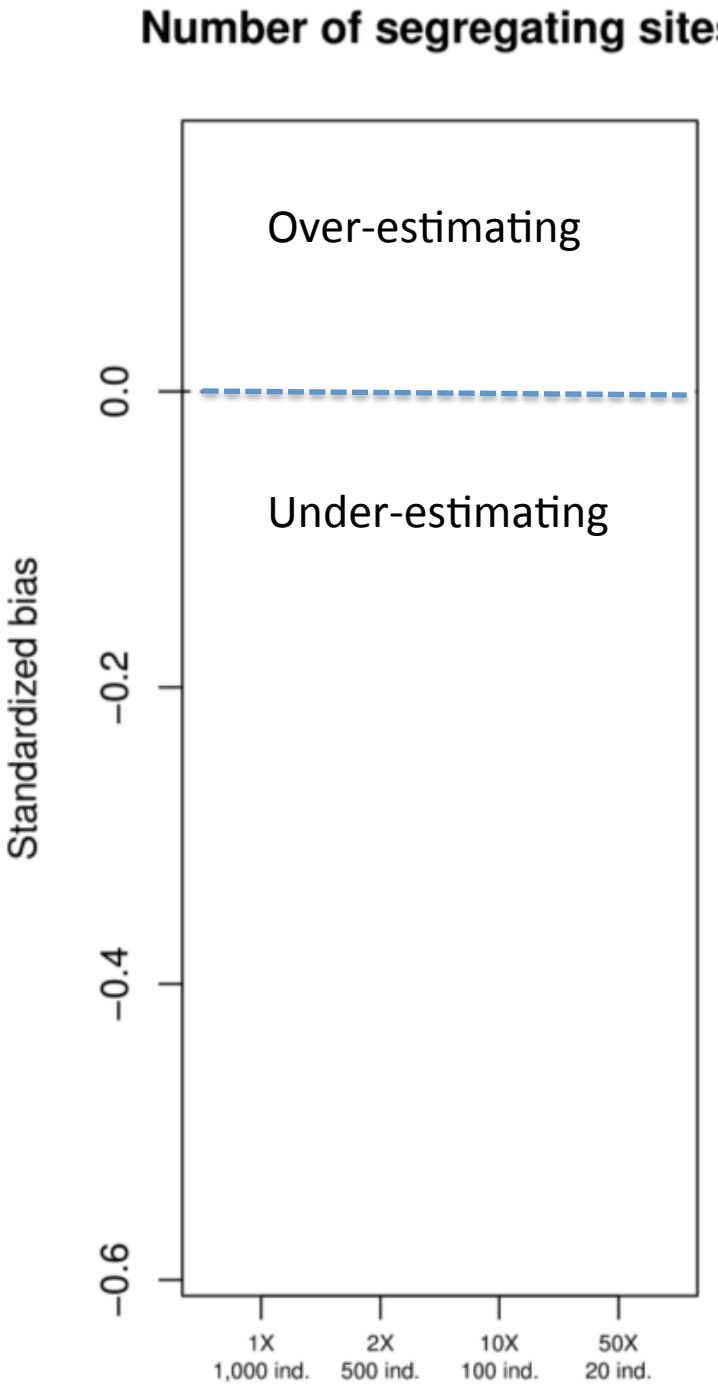
The sequencing strategy can easily be modeled in terms of the number of sequenced samples and the per-sample sequencing depth.

Sample size	Per-sample depth
1,000	1X
500	2X
100	10X
20	50X

total depth is
1,000X

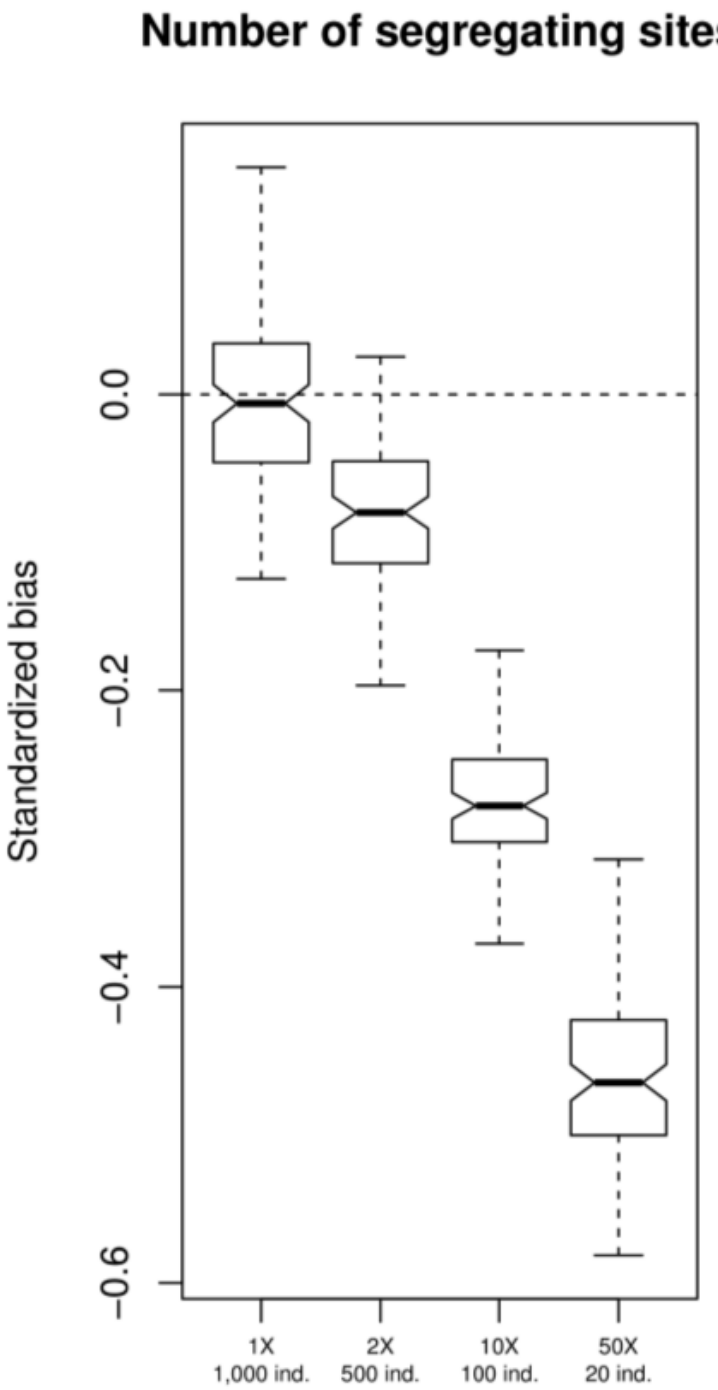
$$Bias(S) = \frac{\hat{S} - S}{S}$$

Compared against the whole **population!**



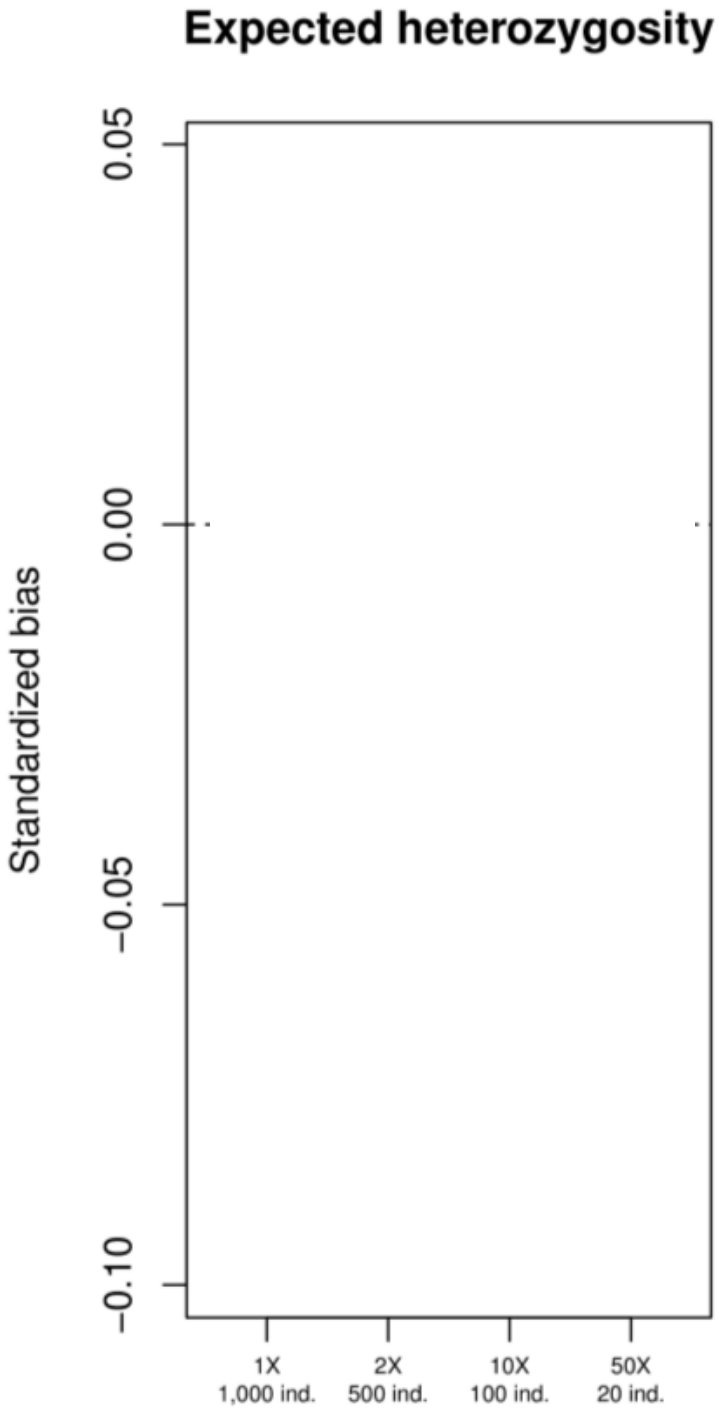
$$Bias(S) = \frac{\hat{S} - S}{S}$$

Compared against the whole **population!**



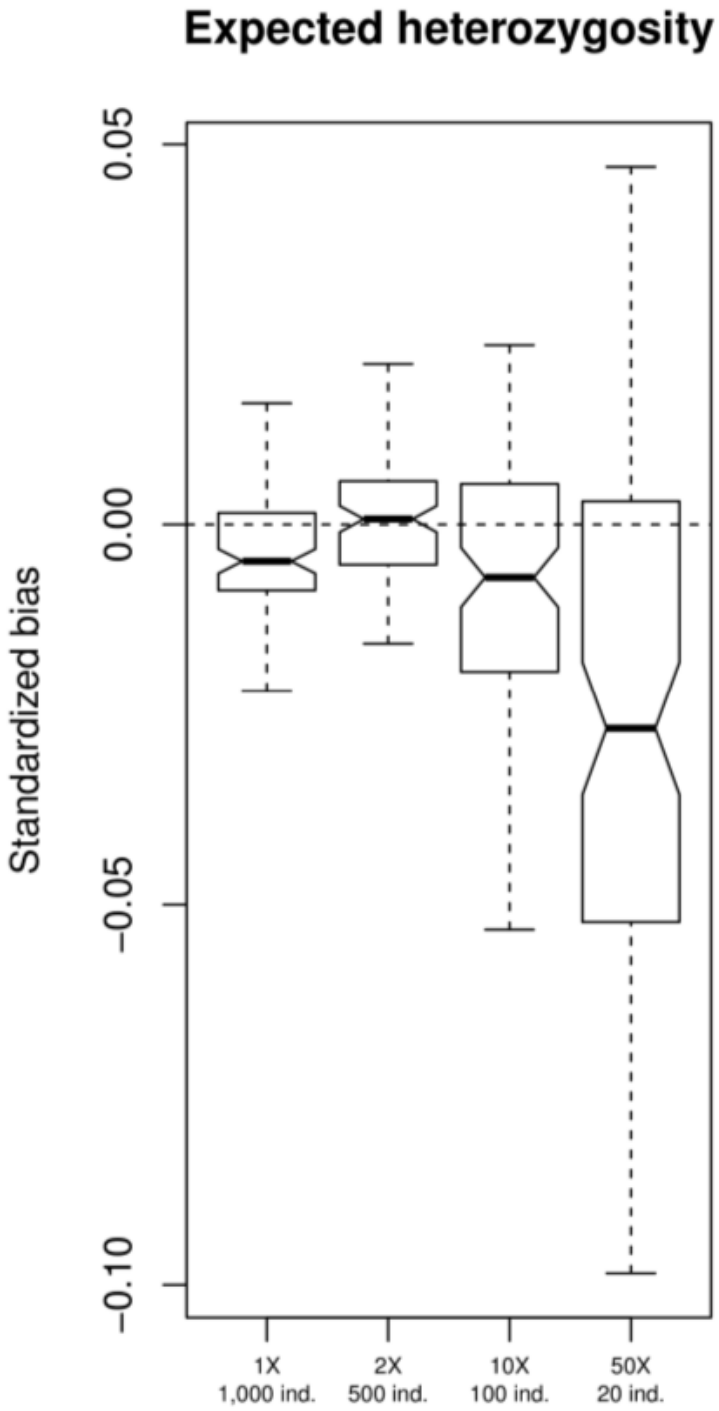
$$Bias(S) = \frac{\hat{S} - S}{S}$$

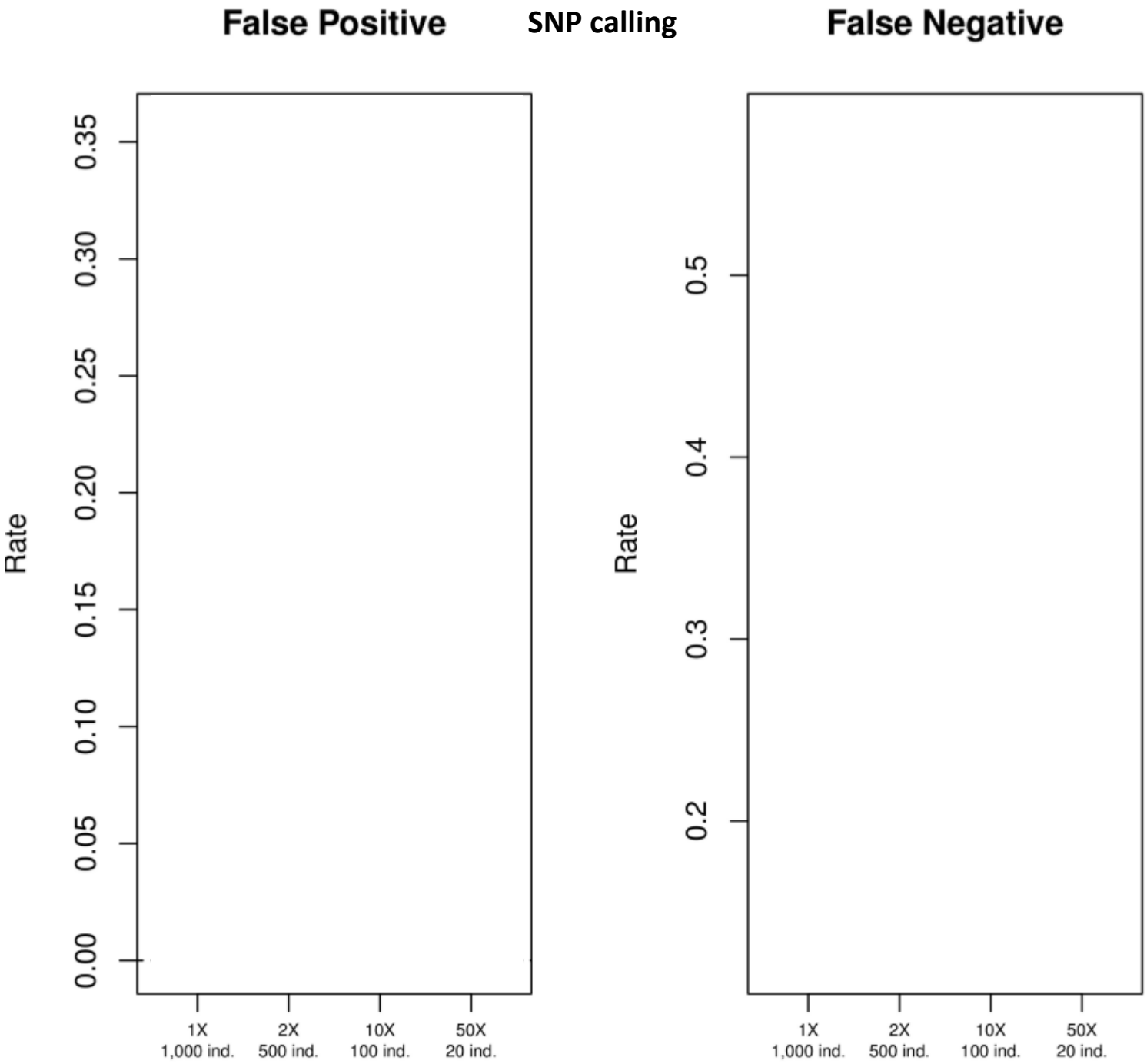
Compared against the whole **population!**

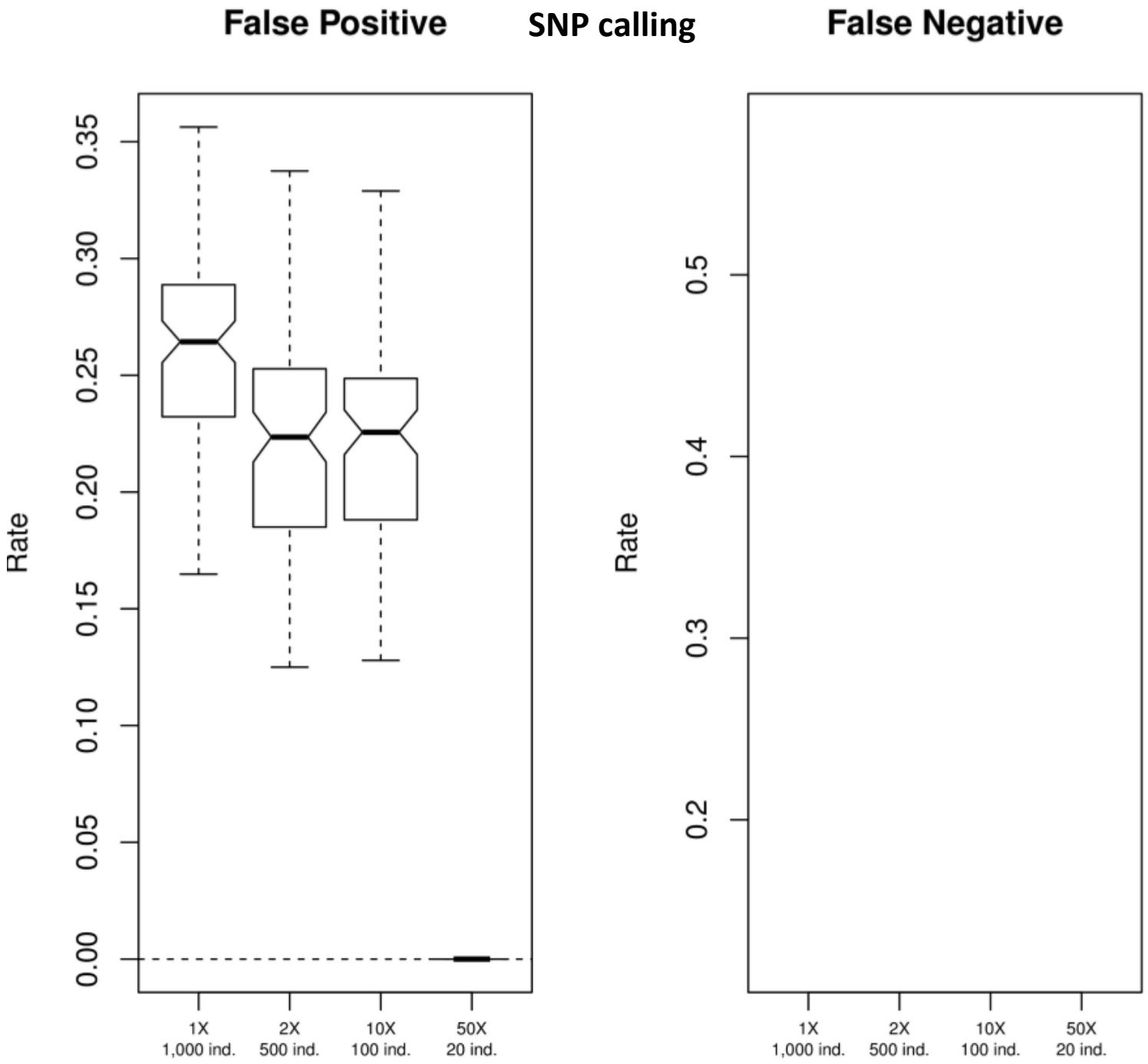


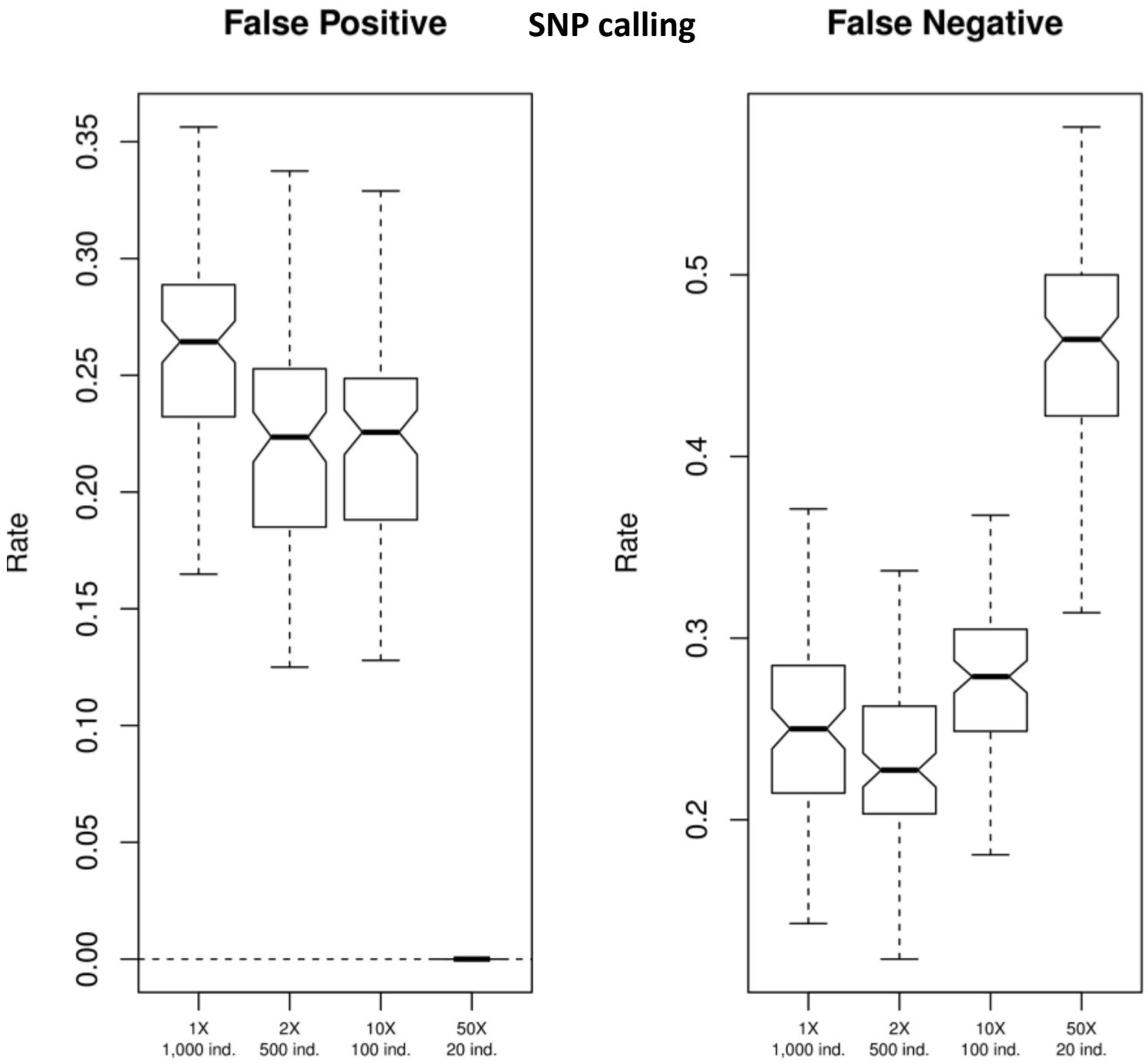
$$Bias(S) = \frac{\hat{S} - S}{S}$$

Compared against the whole **population!**



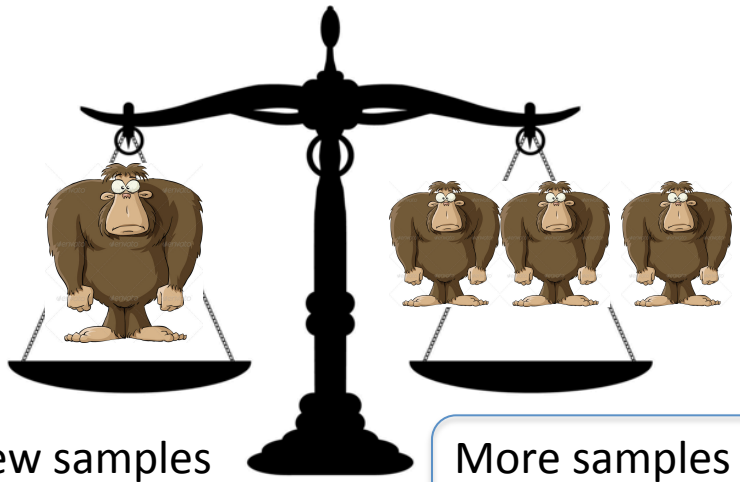

False Positive
SNP calling
False Negative
Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.
Rate
0.2
0.3
0.4
0.5
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.


False Positive
SNP calling
False Negative
Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.
Rate
0.2
0.3
0.4
0.5
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.


False Positive
SNP calling
False Negative
Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.
Rate
0.2
0.3
0.4
0.5
1X
1,000 ind.
2X
500 ind.
10X
100 ind.
50X
20 ind.

Experimental design

At a fixed budget…



Few samples
at higher quality

More samples
at lower quality

For population genetics
Sample allele frequencies will be a better representation of population variation