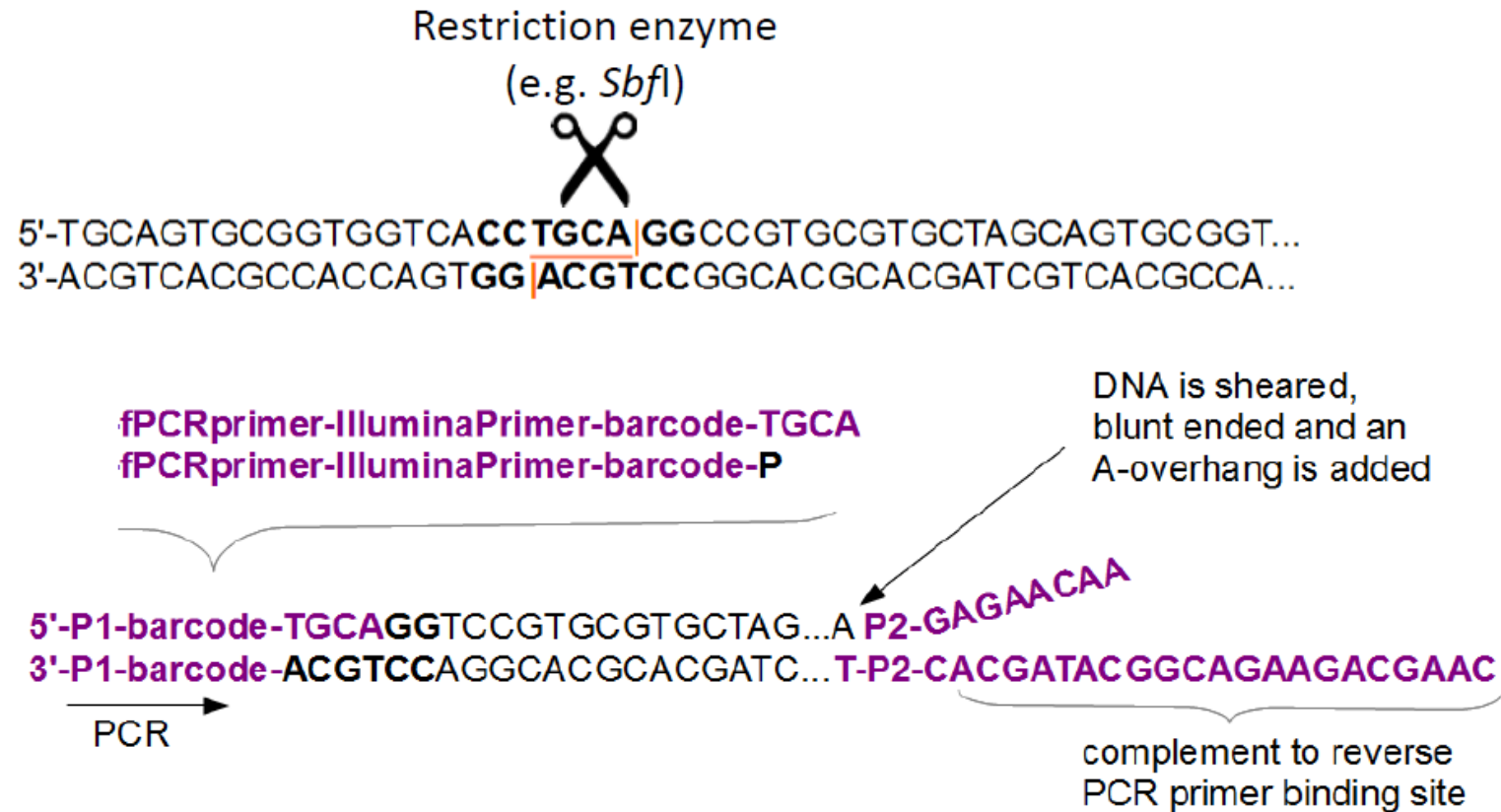


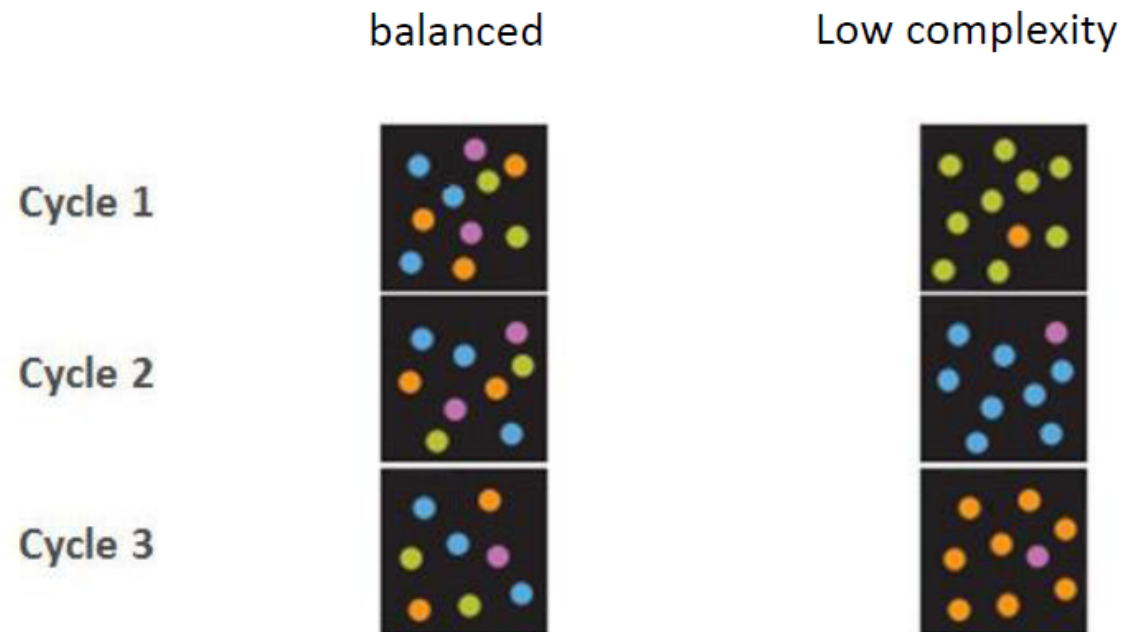
RAD/GBS



Each read: will start with the barcode, then the restriction site, then a variable sequence

Issues with cluster identification

Due to low complexity at the beginning of the sequence,
Illumina cannot distinguish if a signal comes from one or two clusters

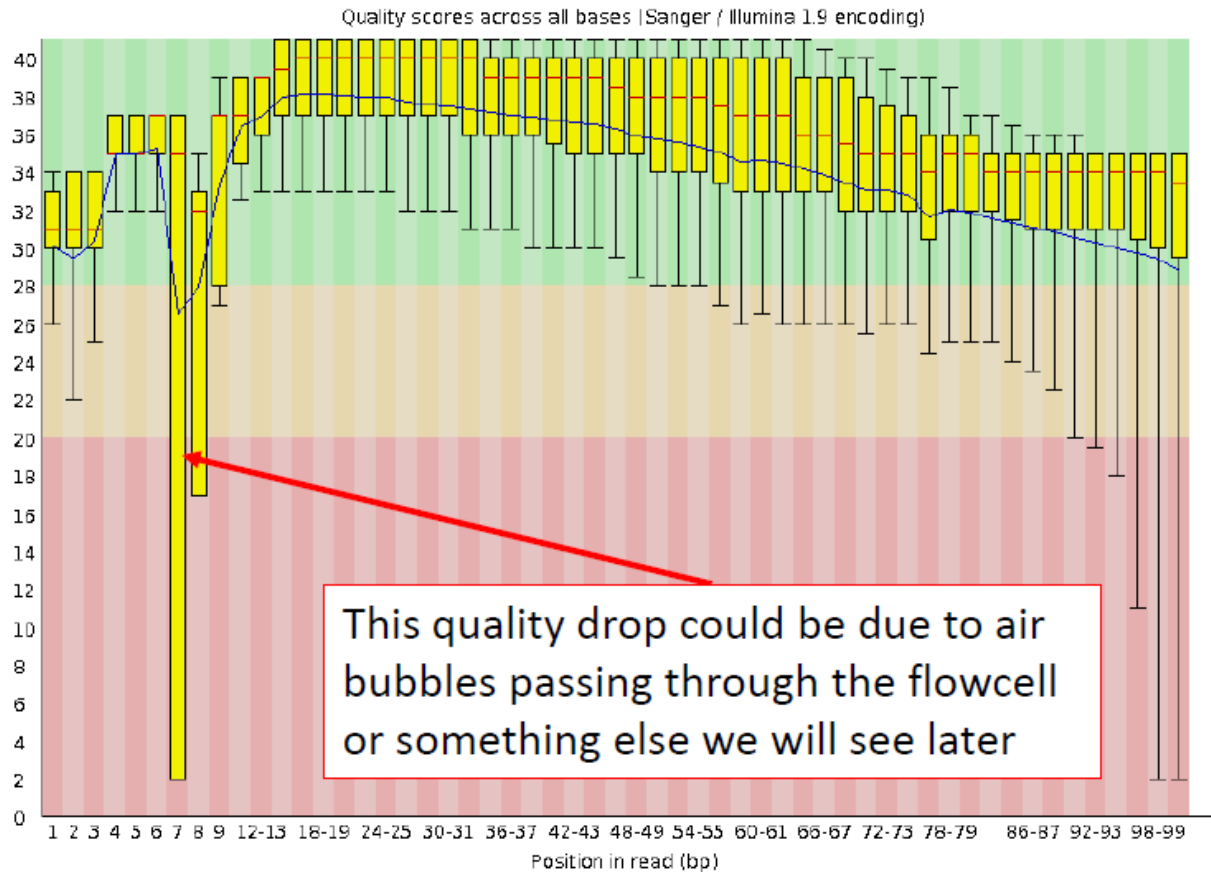


How to minimize the problem

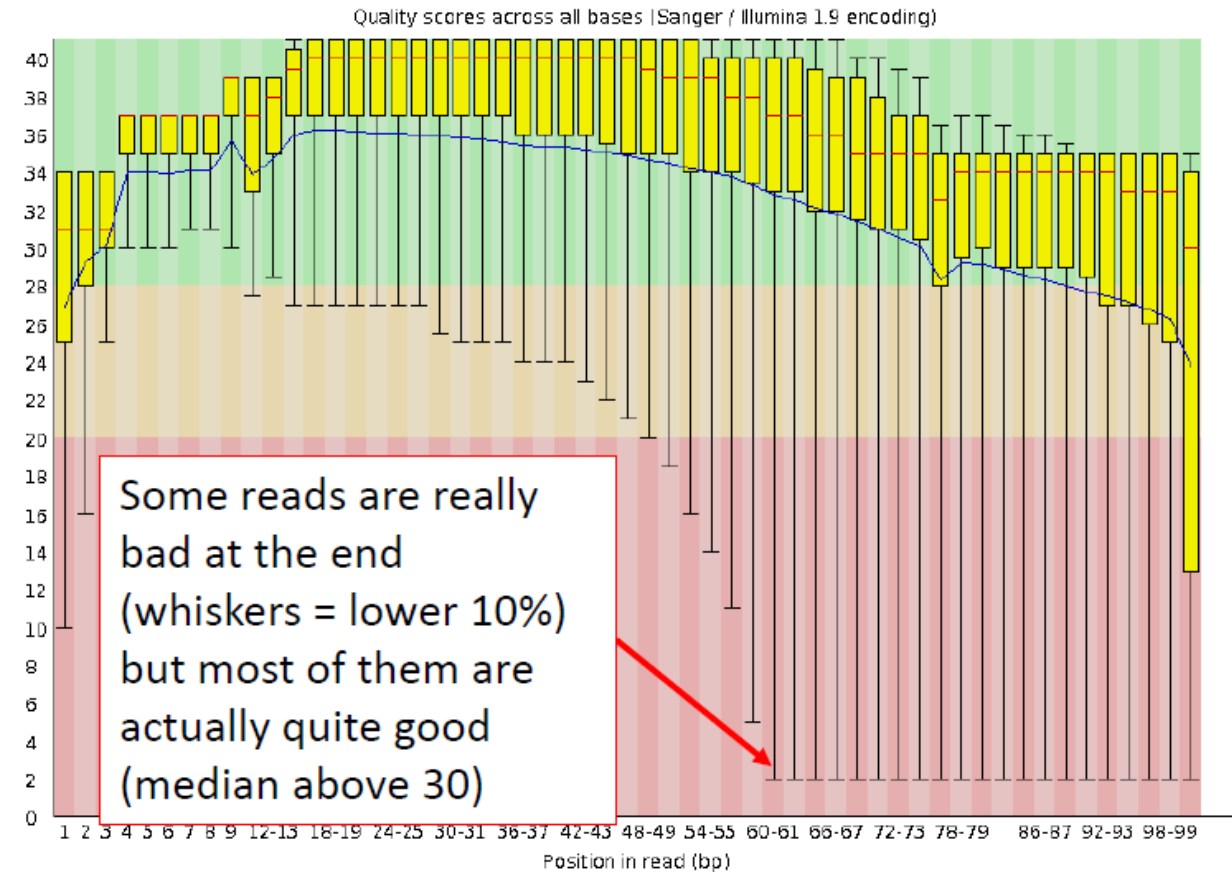
- Use barcodes of different lengths to shift the restriction enzyme cut site
- Add PhiX virus DNA to the RAD libraries to increase the complexity of reads ('spiking')
- Reduce loading concentrations of Illumina plates
- Potentially: filter out bad reads

Quality scores across bases: RAD datasets

RAD1

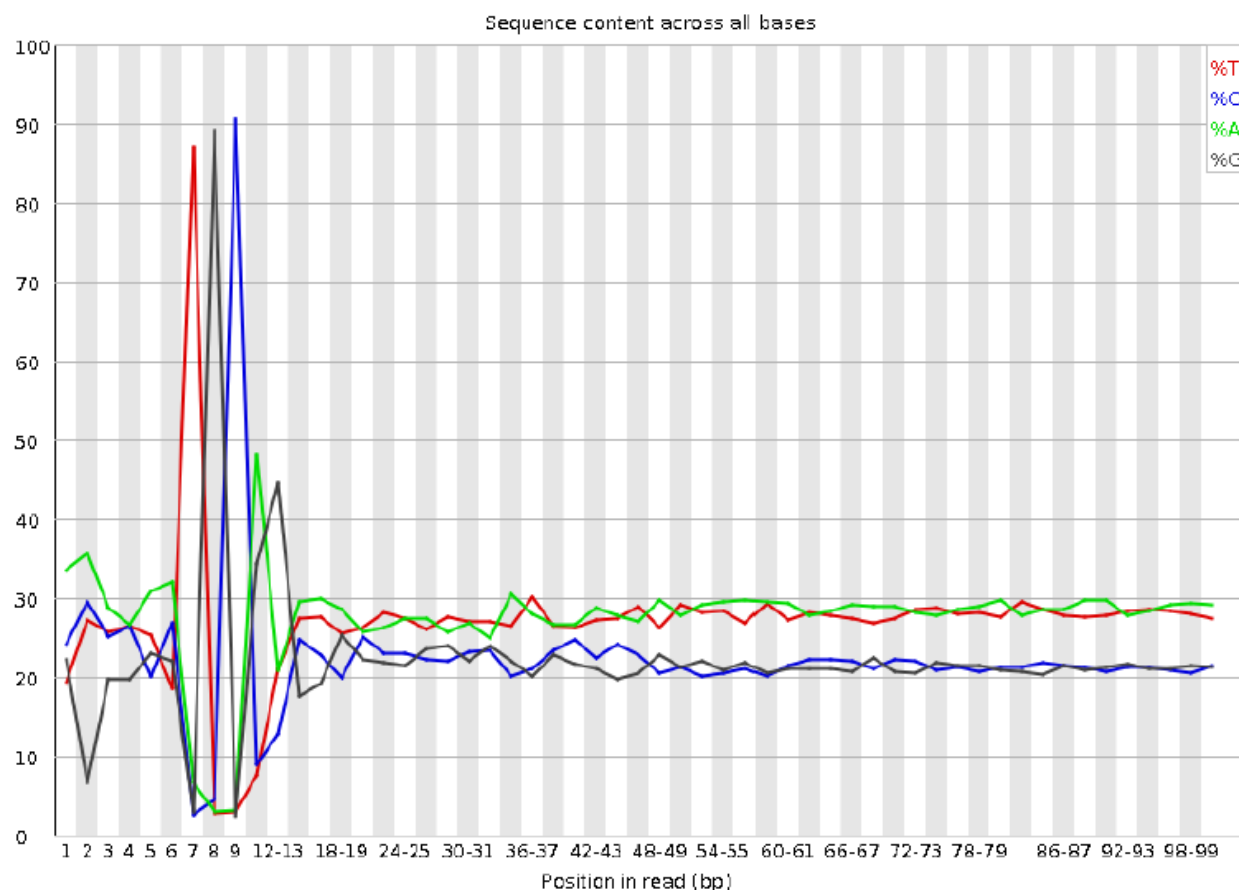


RAD2

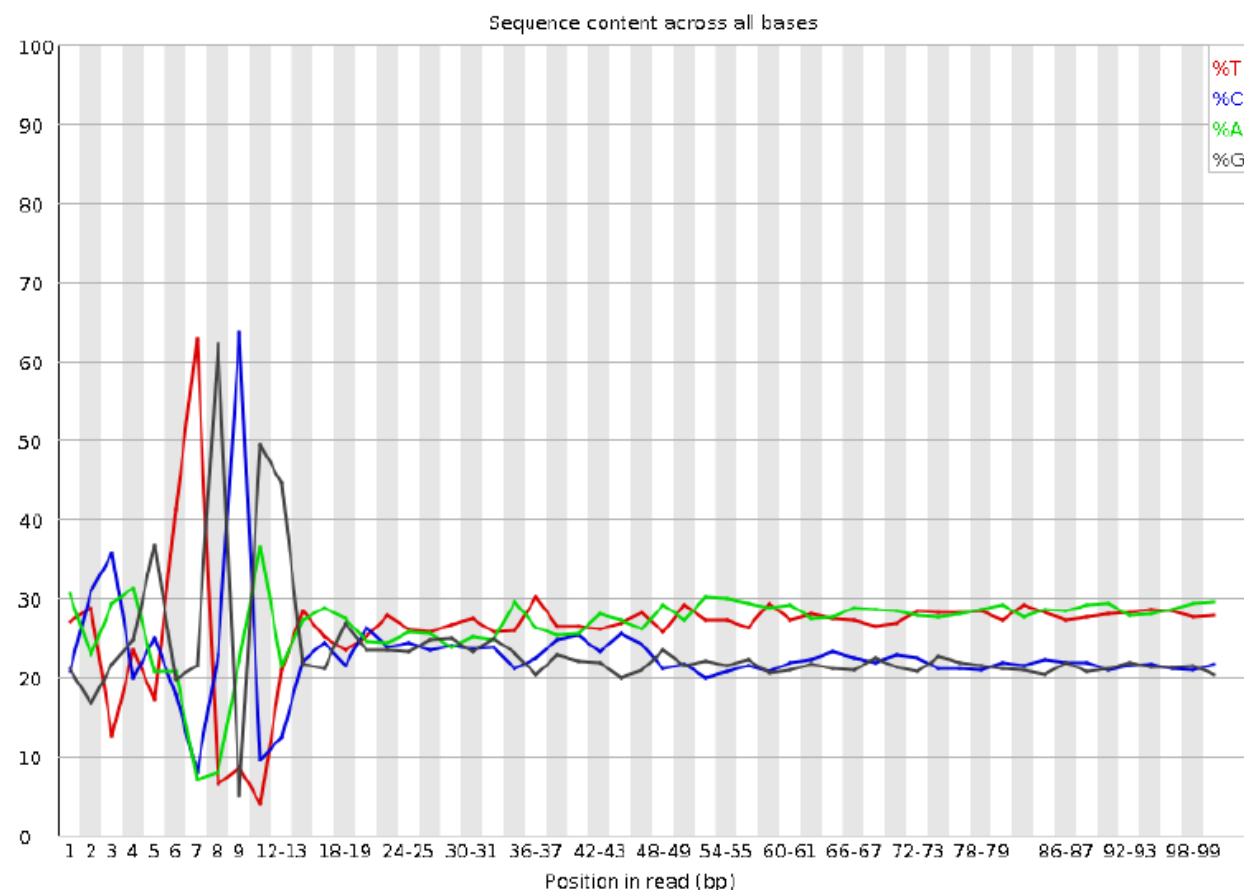


Per base sequence content

RAD1

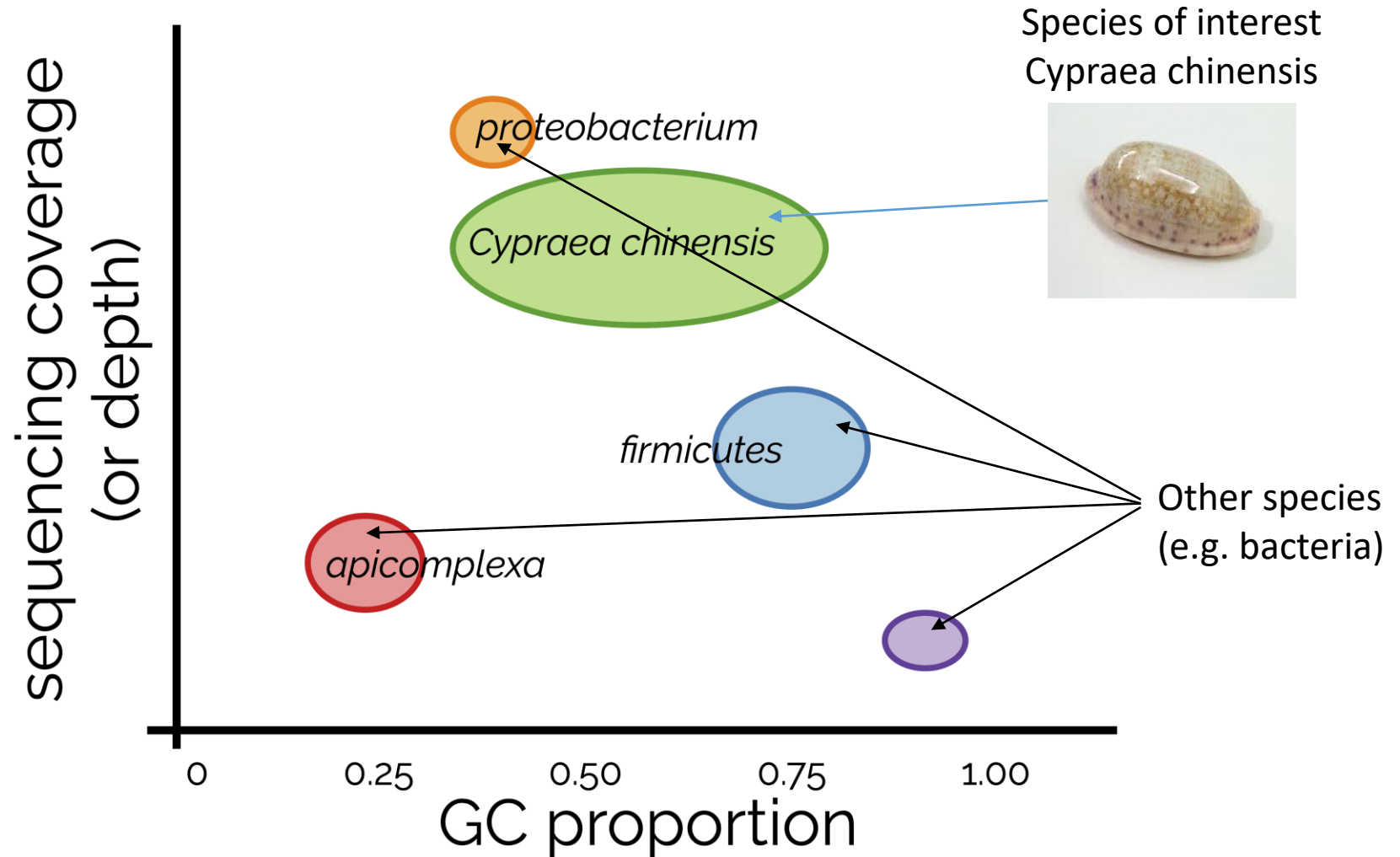


RAD2

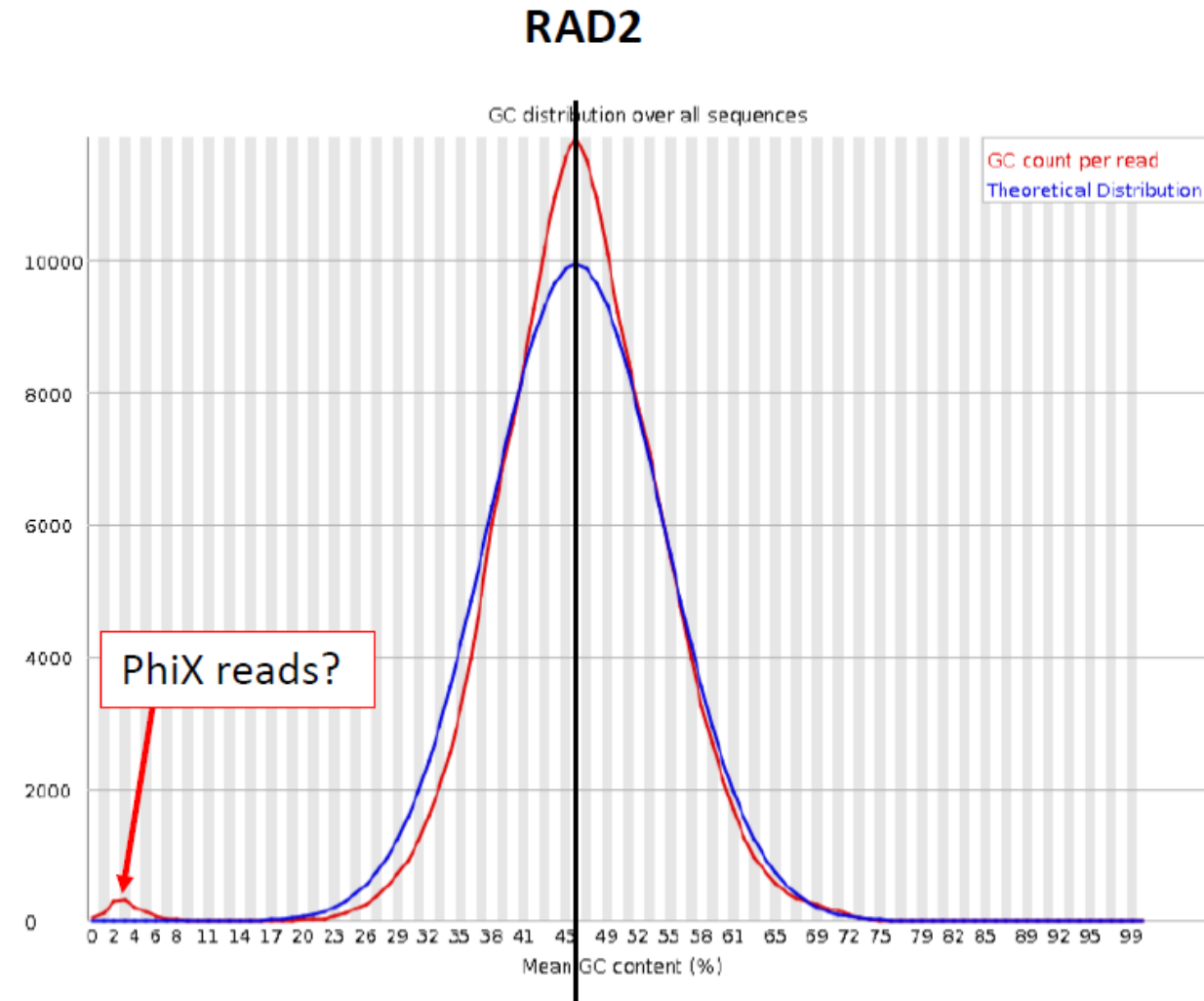
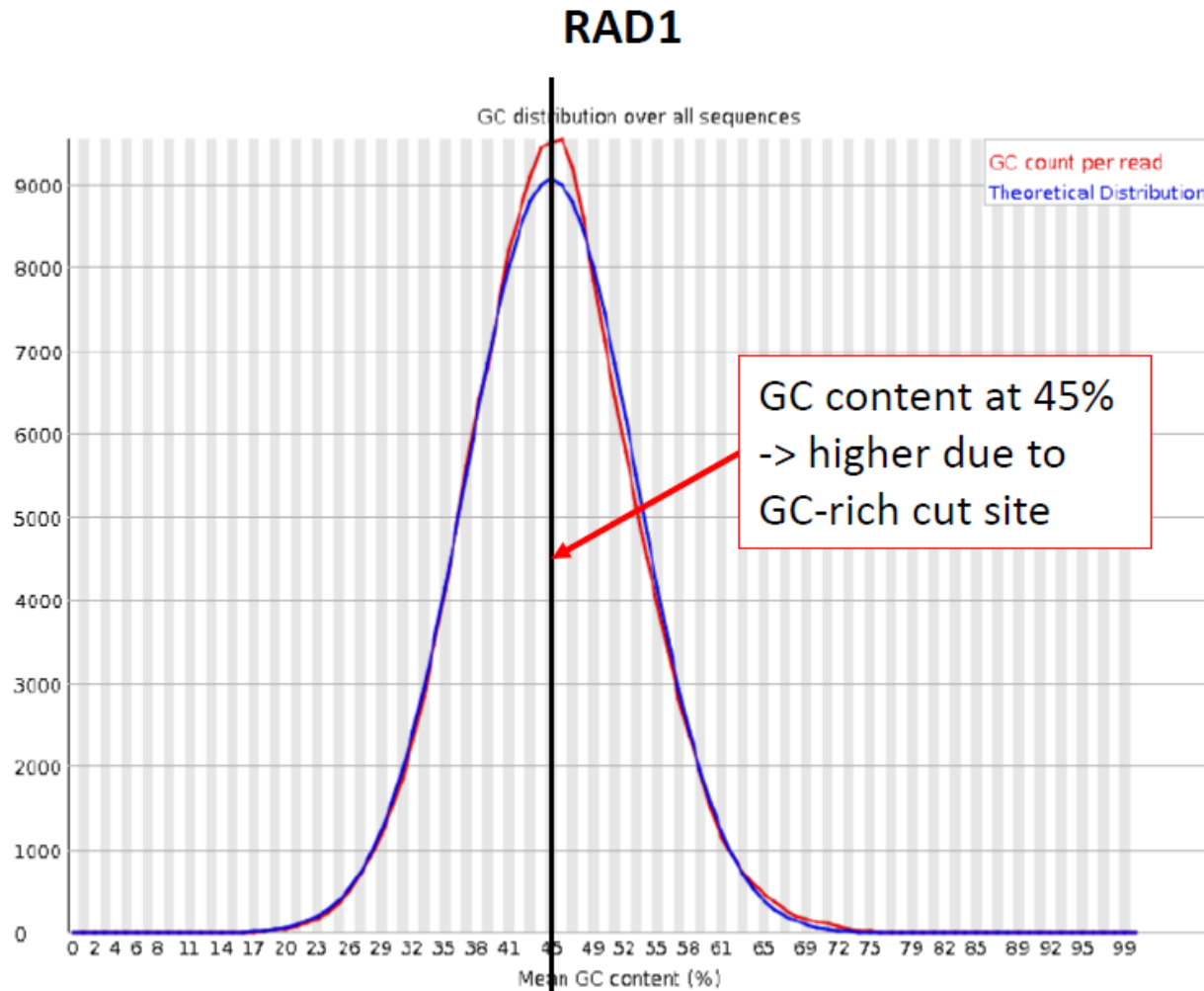


Blobtools kit uses the GC versus AT proportion to identify contamination from bacteria or other organisms

(<https://blobtoolkit.genomehubs.org/>)



GC distribution over all sequences



Examples of what you do not want to see:

Fastqc: Per sequence GC content

! Per sequence GC content

