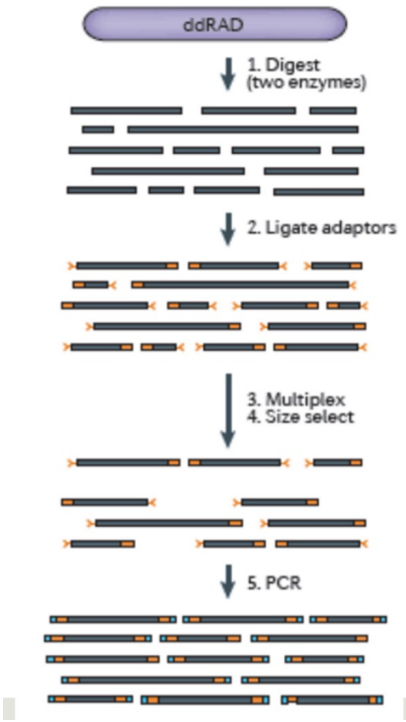
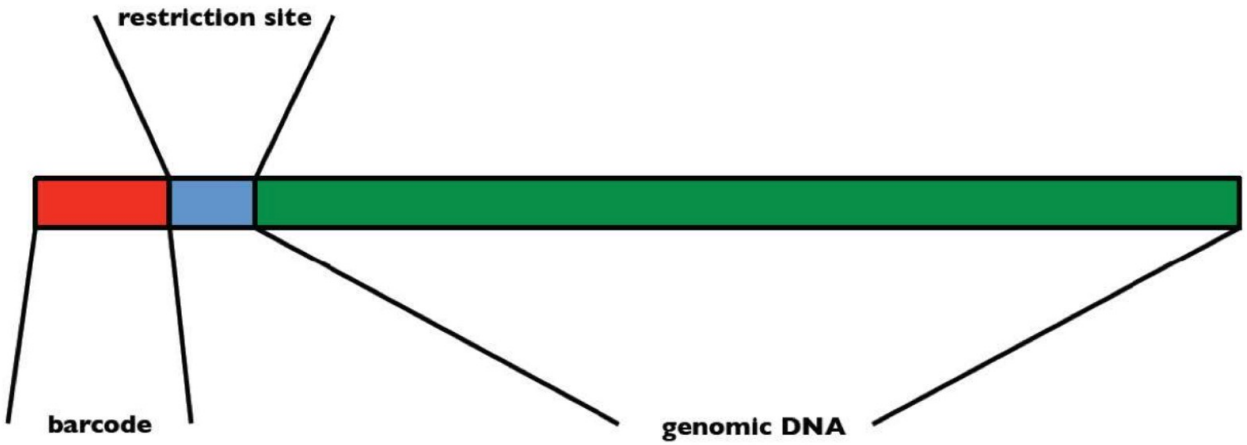


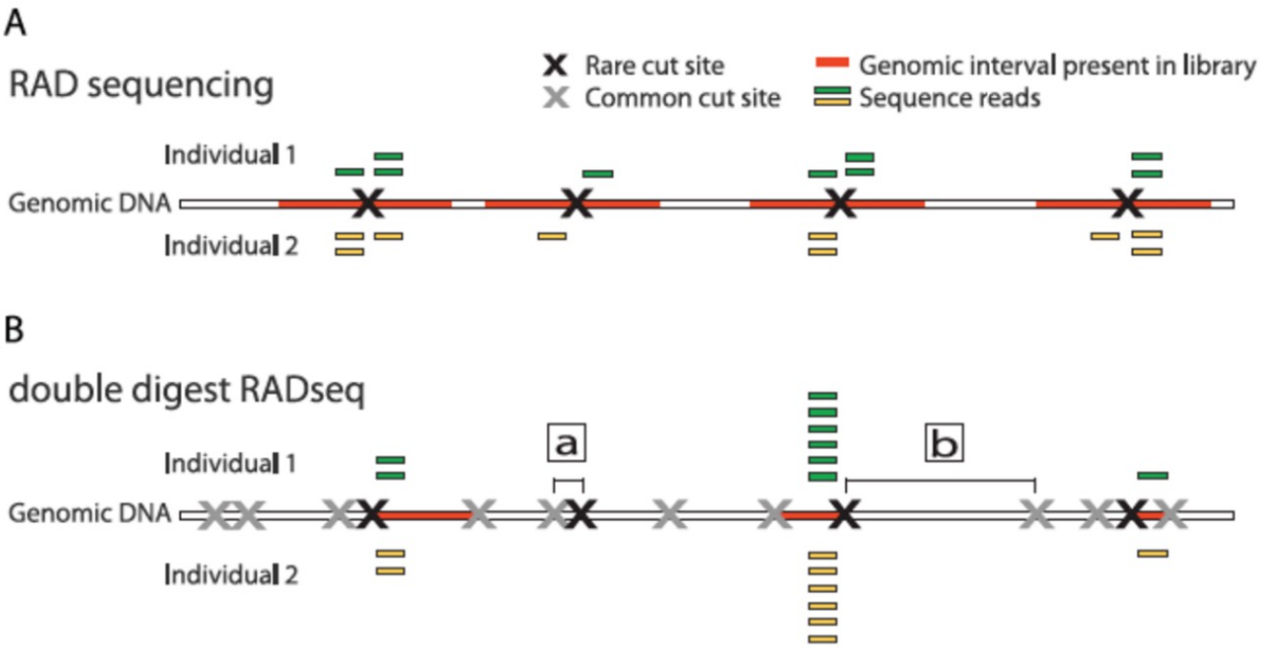
There are DOZENS of kinds of genomic library preparations, and they each have their pros and cons. In class, we will be analyzing Illumina data generated using ddRAD libraries, so here's a bit of background.

- "***Reduced representation libraries***" aim to have less than 5% of the total genome
- **Genotyping By Sequencing** is a group of protocols that are based on not knowing much about the genome and targeting 'haphazard' regions throughout.
- double-digest Restriction Associated or **ddRAD**, is a protocol which uses two enzyme cuts and a size selection step to further reduce the genomic representation, which is especially good for eukaryotes with large repetitive genomes

The data within your files will look like what you see below; each color indicates the specific part of the sequence – the **green is the target sequence**. Illumina pre-parses your data and eliminates the known primer sequences that were used for hybridizing the oligos.



@M00689:44:000000000-A1N97:1:1101:11642:2590 1:N:0:1
CTGATGCTTGCAGGACGCACCTCCCCGCGGTGCGGCTAATGTCCTCGCAGC
+
AAAAABBBDDDDDDDDGGGGGGIIIHHHHHEHHHHHBHHIIIIIIIHHH@E


A
RAD sequencing
Rare cut site
Common cut site
Genomic interval present in library
Sequence reads
Individual 1
Genomic DNA
Individual 2
B
double digest RADseq
a
b
Individual 1
Genomic DNA
Individual 2