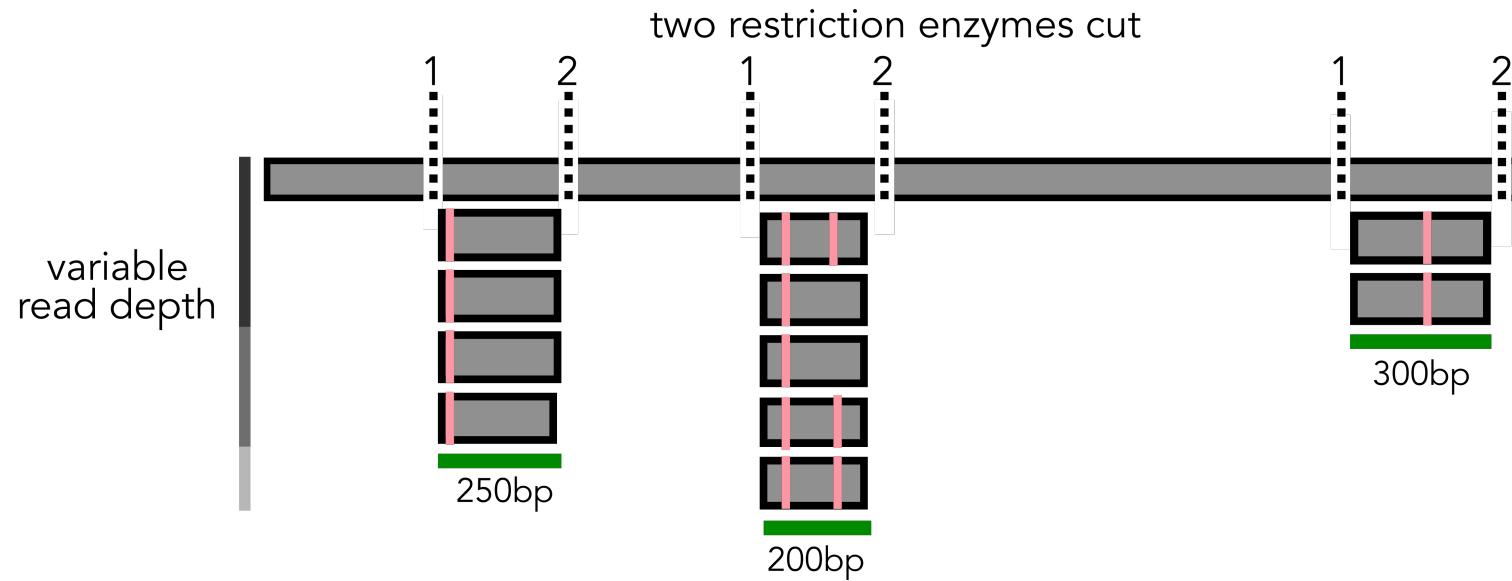
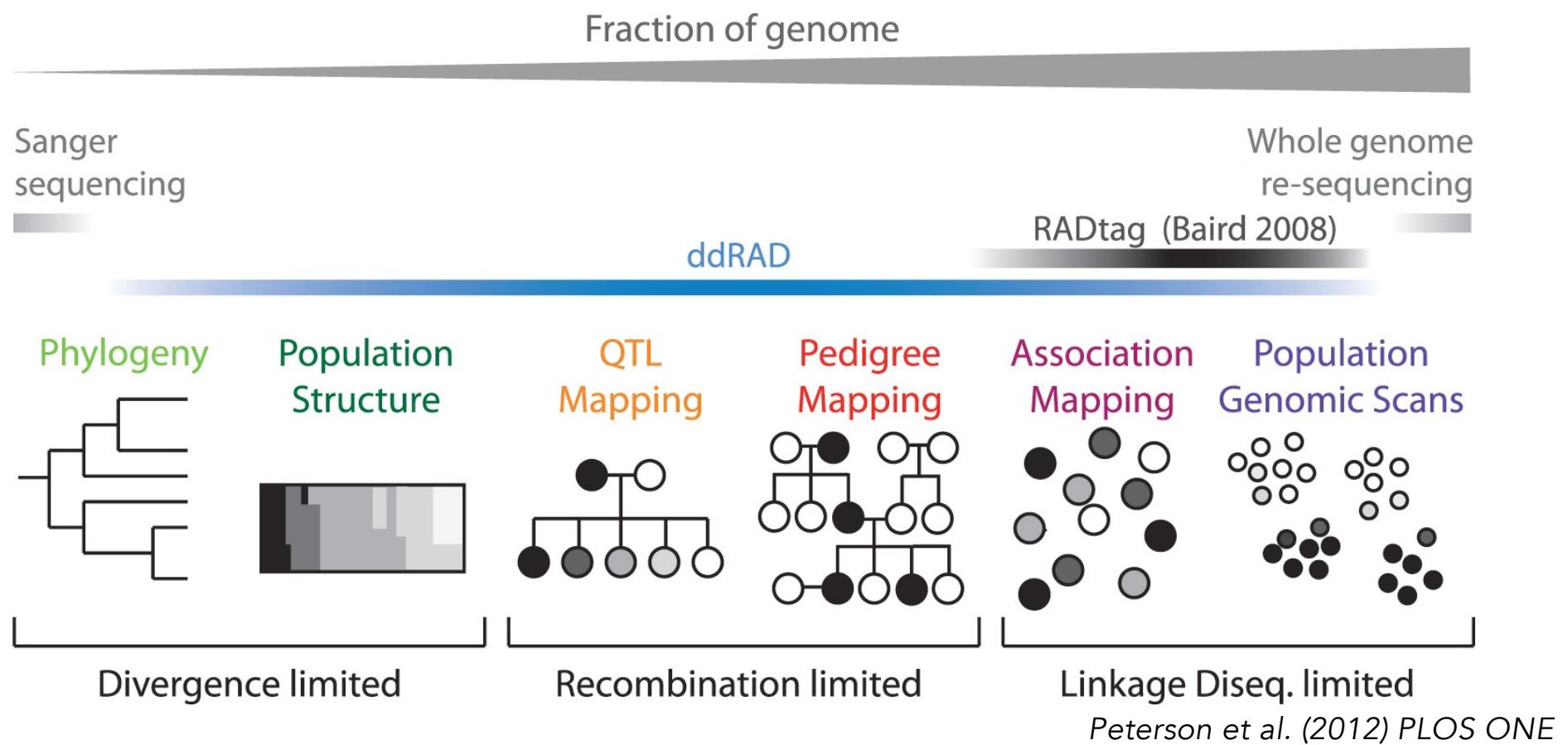


RADseq conceptual info



What is RADseq, and why do it?



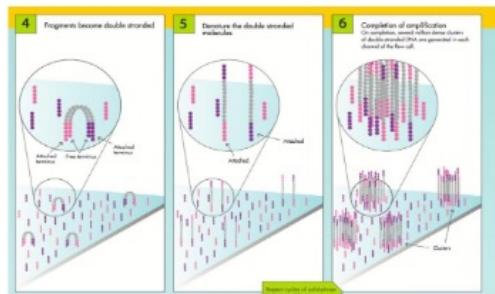
What is RADseq, and why do it?

Advantages:

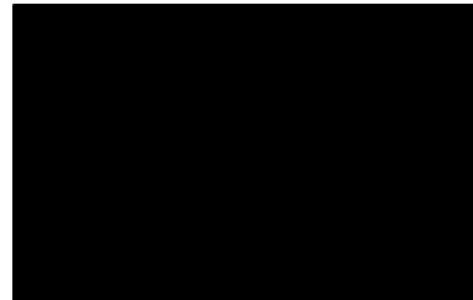
- Much lower in-lab effort
- Higher cost efficiency
- Obtain thousands of informative loci across the entire genome
- Possibility of looking at neutral and adaptive loci and their correlation with geographic or phenotypic data

What is RADseq, and why do it?

Disadvantages:



Sequences



Bioinformatics

2	3	2	2	2	2	3	2	2	2	3	3	2	4	3
2	3	2	2	4	3	2	2	2	0	3	2	4	3	
2	3	2	2	2	2	0	2	2	2	0	3	4	4	3
0	3	2	2	0	3	2	2	2	3	3	3	2	4	3
0	3	2	2	0	3	2	2	2	3	3	3	4	4	3
2	3	2	0	2	3	0	0	3	3	0	0	0	0	0
2	3	2	0	4	3	0	0	0	3	3	0	4	3	
2	3	2	0	0	3	0	2	2	3	3	0	4	3	
2	3	2	0	0	3	0	4	3	3	3	0	4	3	
2	3	2	0	0	3	0	2	2	1	3	4	0	3	
2	3	2	0	0	3	0	2	2	3	3	4	0	3	
0	0	2	2	2	3	2	2	2	3	3	2	4	3	
0	0	2	2	4	3	2	2	3	3	2	4	3		
2	3	2	2	0	3	2	2	3	3	3	4	4	0	
2	3	2	2	0	3	2	2	3	3	3	4	4	0	
2	0	2	2	2	3	2	2	3	3	2	4	3		

SNP matrix

Is RADseq appropriate for you?

Examples of bad projects for RADseq:

- Mapping functional traits (or GWAS) in an organism with no reference genome (generate genome, use markers with no ID, target sequencing)
- Finding genes that are involved in phenotypic changes and/or across developmental stages (RNAseq)
- Phylogenetics of distantly related organisms (UCEs)

Trade-offs among five RADseq methods

	Original RAD	2bRAD	GBS	ddRAD	ezRAD
Options for tailoring number of loci	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme or size selection window	Change restriction enzyme or size selection window
Number of loci per 1Mb of genome size	30–500	50–1,000	5–40	0.3–200	10–800
Length of loci	≤1kb if building contigs; otherwise ≤300bp	33–36bp	<300bp	≤300bp	≤300bp
Cost per barcoded or indexed sample	Low	Low	Low	Low	High
Effort per barcoded or indexed sample	Medium	Low	Low	Low	High
Use of proprietary kit	No	No	No	No	Yes
Identification of PCR duplicates	With paired-end sequencing	No*	With degenerate barcodes	With degenerate barcodes	No
Specialized equipment needed	Sonicator	None	None	Pippin Prep	Pippin Prep
Suitability for large or complex genomes	Good	Poor*	Moderate	Good	Good
Suitability for <i>de novo</i> locus identification (no reference genome)	Good	Poor*	Moderate	Moderate	Moderate
Available from commercial companies	Yes	No	Yes	Yes	No

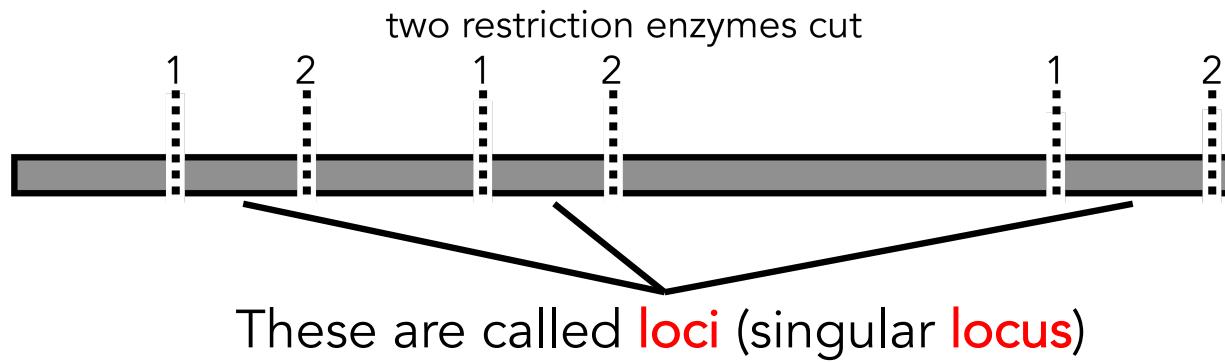
Taken from Andrews et al. (2016)

ddRAD methodology

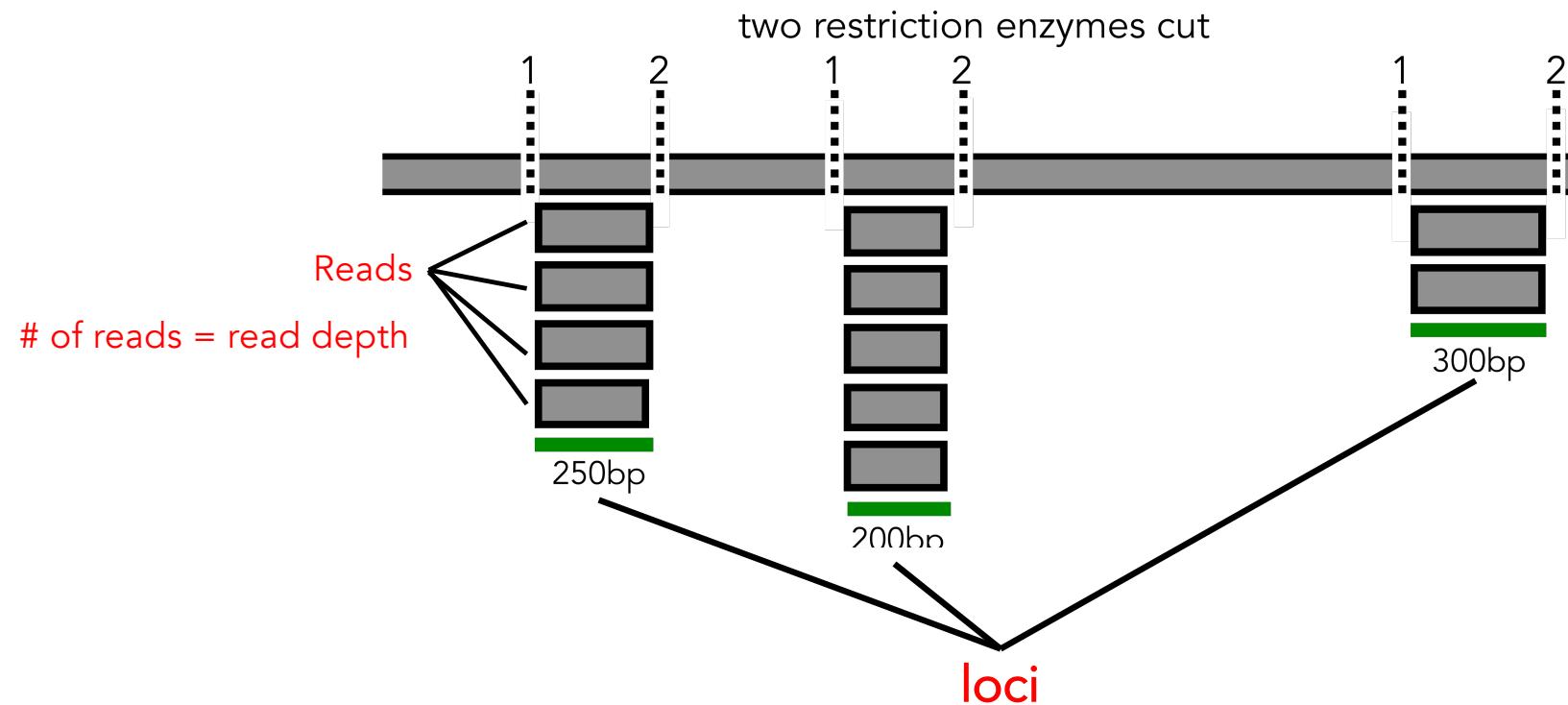
Genome (DNA)



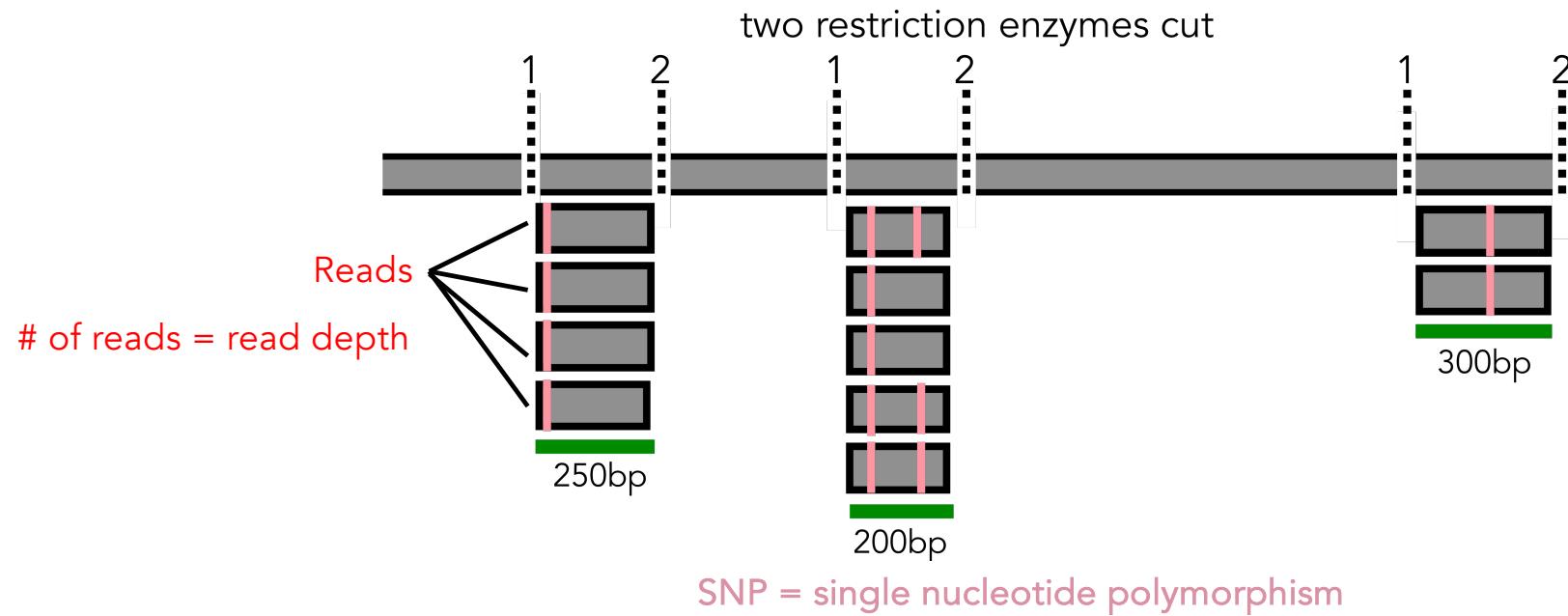
ddRAD methodology: digestion



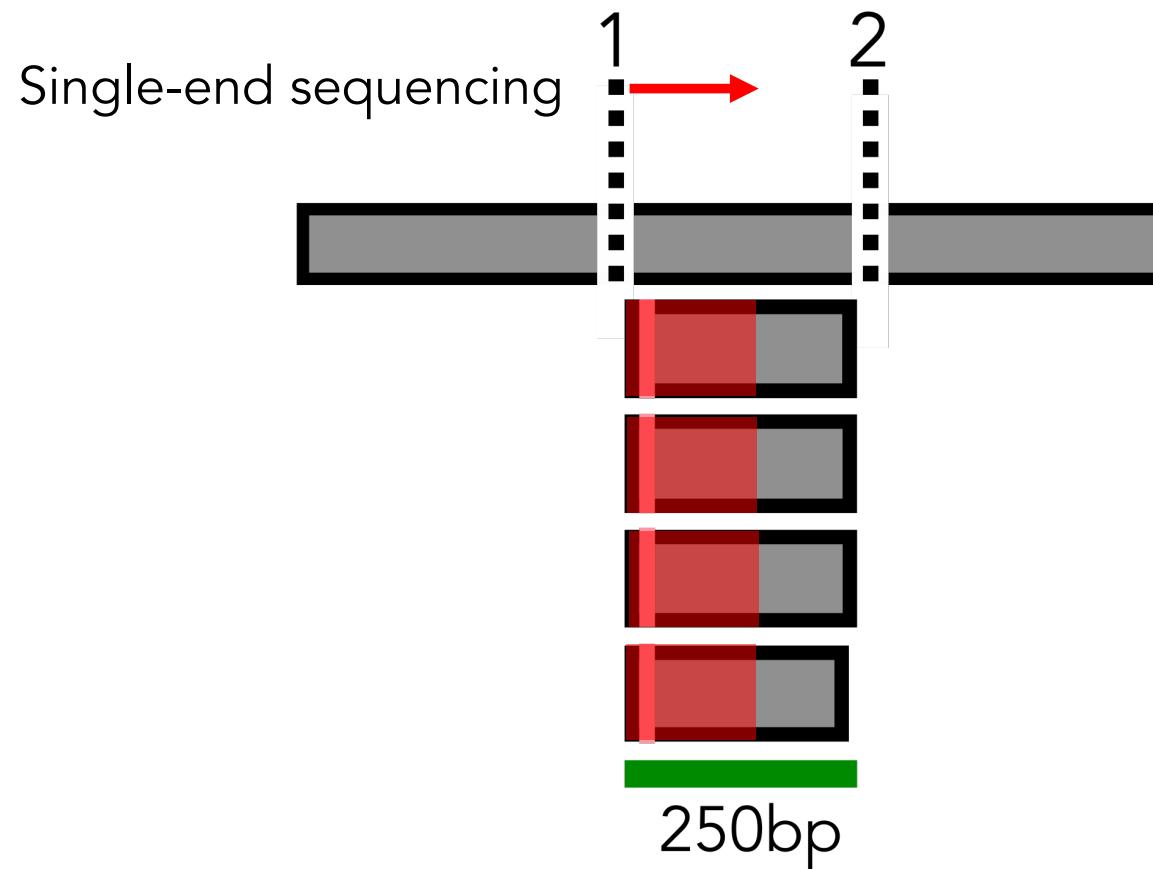
ddRAD methodology: **amplification**



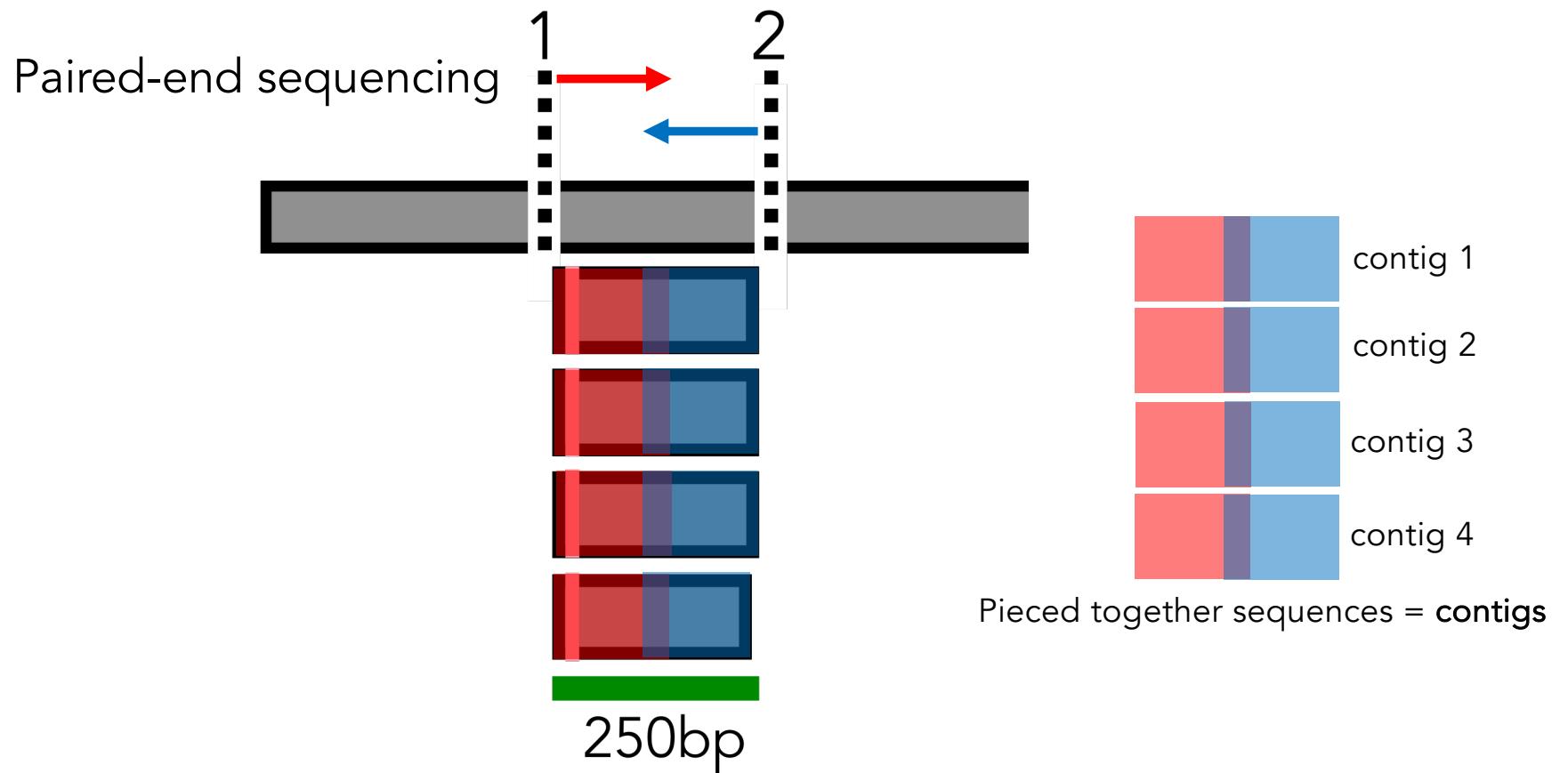
ddRAD methodology: sequencing



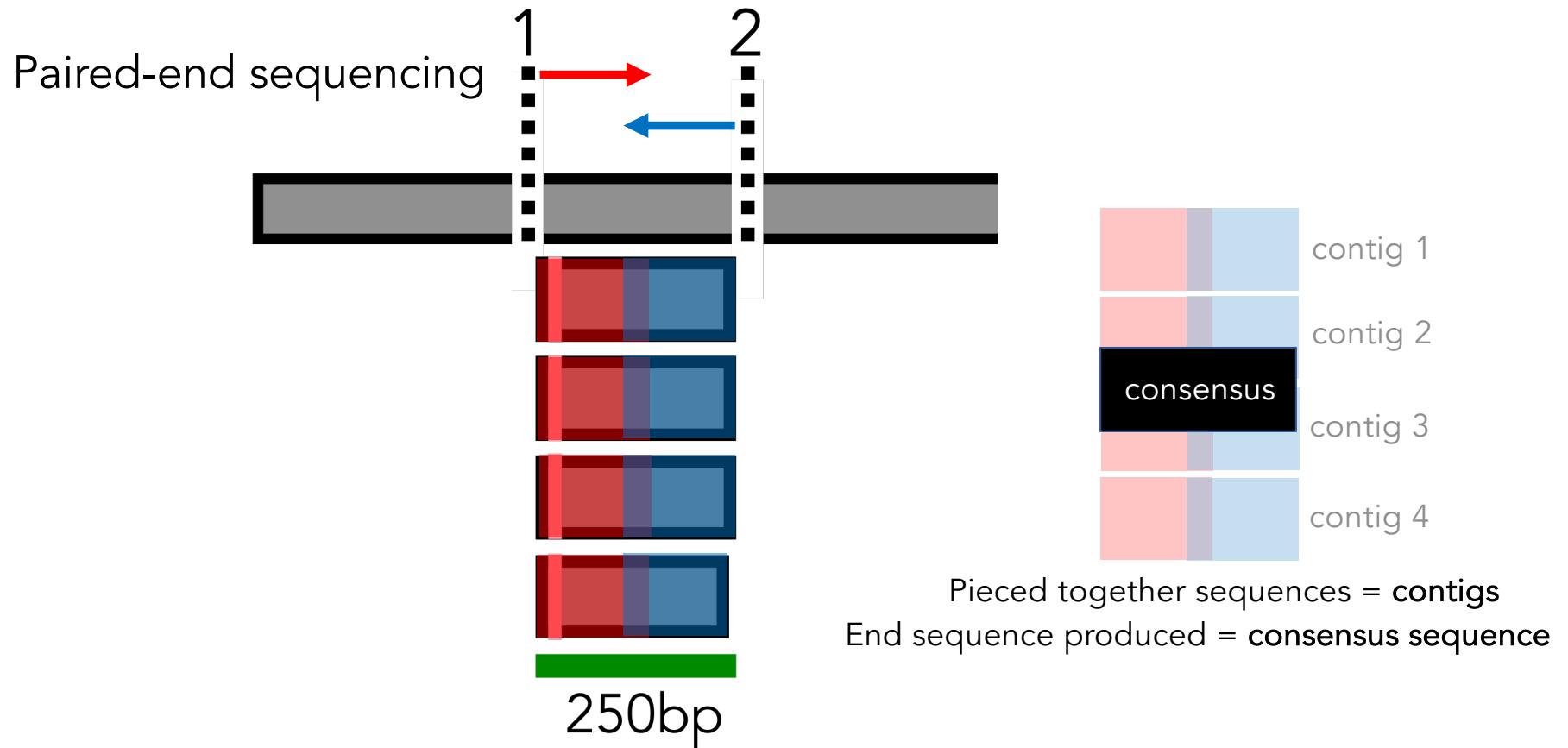
ddRAD methodology: **sequencing**



ddRAD methodology: sequencing



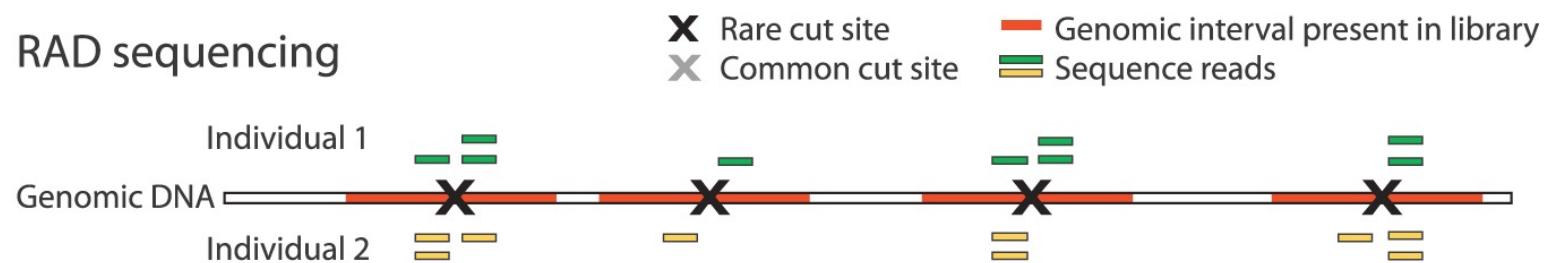
ddRAD methodology: sequencing



ddRAD methodology

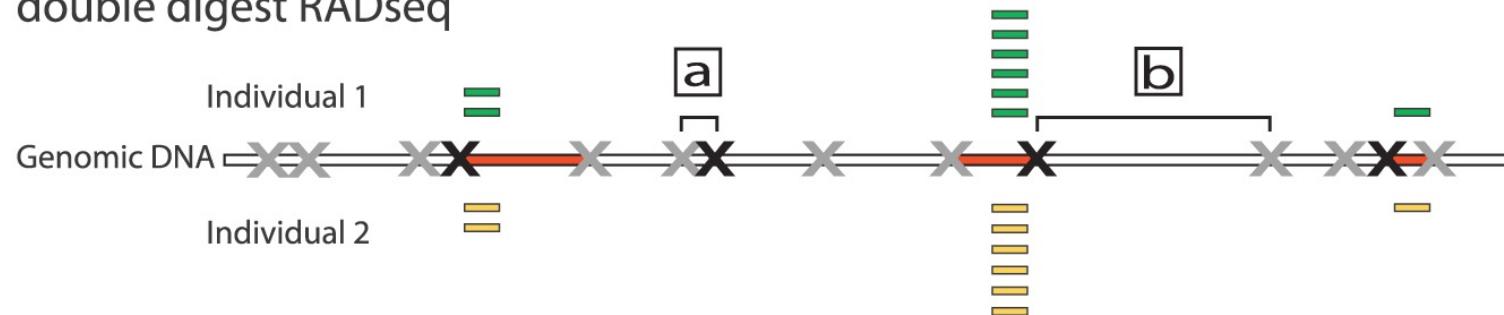
A

RAD sequencing

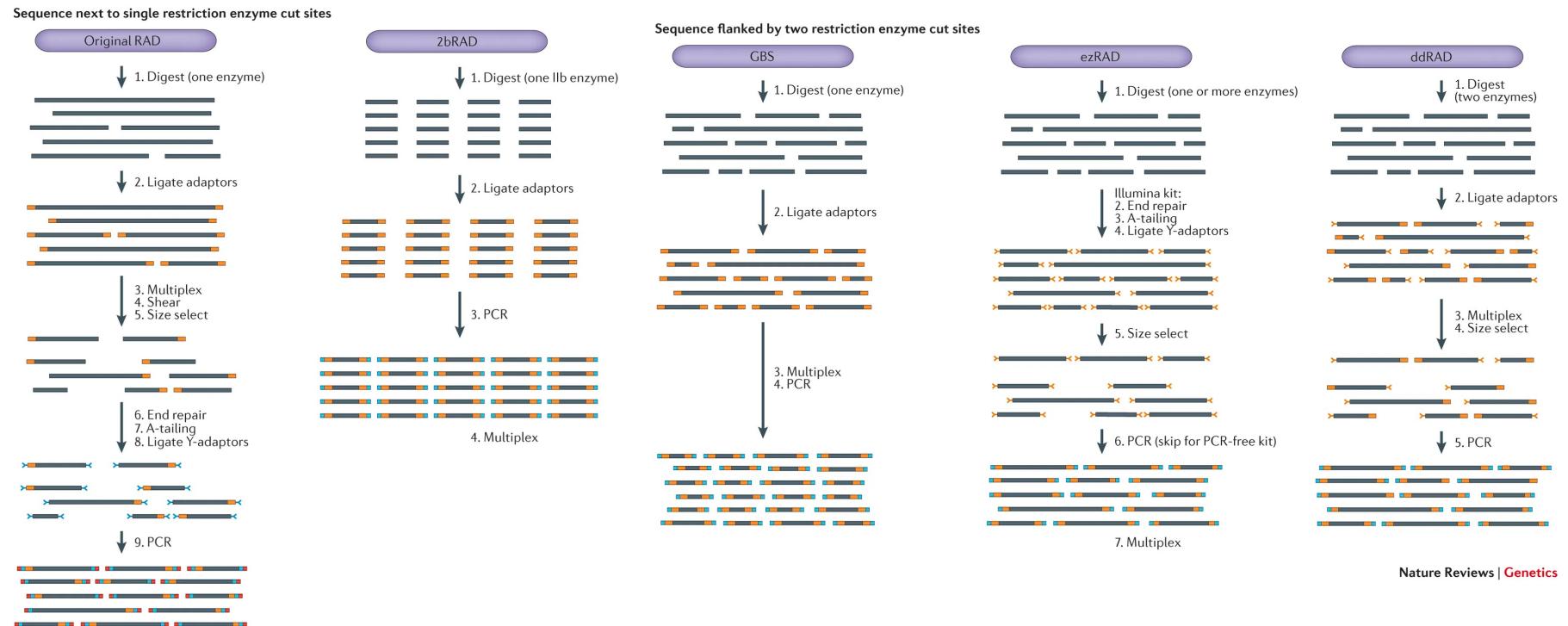


B

double digest RADseq



Comparison among RADseq methods



Andrews et al. (2016) Nature Reviews

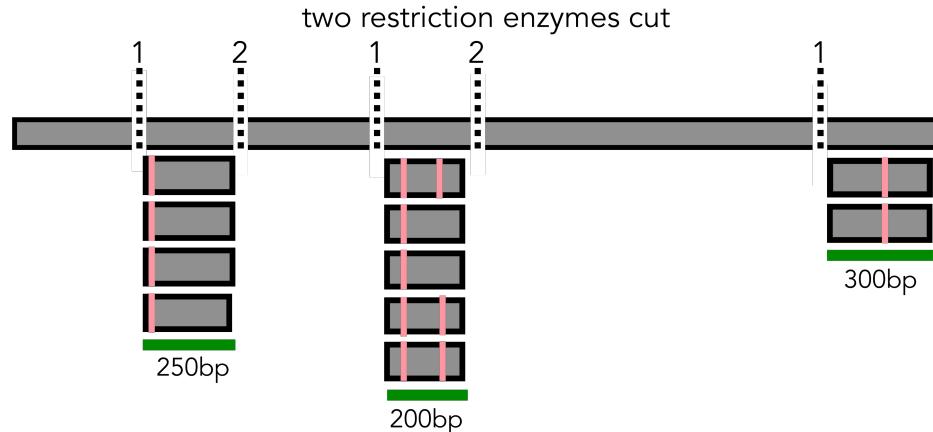
ddRAD: terminology

SNP: single nucleotide polymorphism
(variable sites, as opposed to invariant sites)

Loci (sing. locus): regions of the genome
that are amplified and sequenced

Multiple **reads** are sequenced at each locus;
how many are present per locus is the **read depth**

Sequencing is done either **single-** or **paired-end**;
resulting **contigs** (overlapping
sequence reads) are combined to form a
single **consensus read** sequence for each
locus per individual



What do the data actually look like?

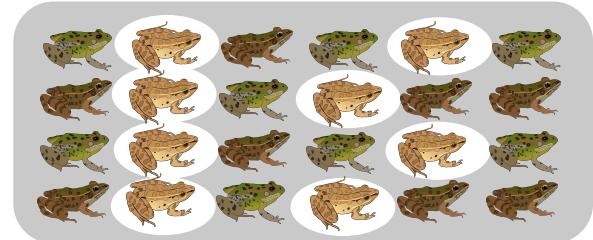
Factors to consider when doing RADseq

- (1) Available resources
- (2) Number and selection of samples
- (3) Number of loci
- (4) Type of read
- (5) Number of reads
- (6) Enzyme selection
- (7) Sequencing machine

(1) Available resources

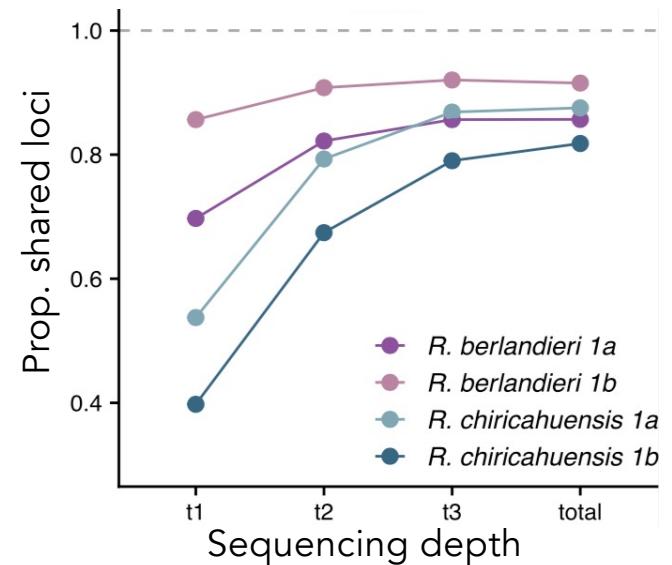
- **Genome**
 - *If available*: targeted sequencing may be more useful (Rapture, UCEs, target-bait capture, chip-seq)
 - *If unavailable*: need to carefully consider study design; some methods are particularly useful (RAD) or not (2bRAD)
- **Money**
 - *If lots*: deeper sequencing, more reliable project design
 - *If little*: careful consideration of study design and method selection
- **DNA samples**
 - Phylogenetic sampling limits questions you can ask
 - Low quality DNA – risk uneven loci recovery
 - Minimum of ~200ng of DNA
- **Supercomputer/computational support**
 - 100% necessary unless you can afford to pay someone else/collaborate

(2) Number and selection of samples

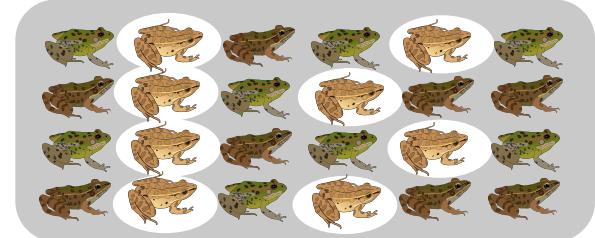


- **Number**

- *Biological replicates per species, treatment, population*
- *Technical replicates to control for bioinformatic and methodological error*
 - Reliability of locus recovery (i.e. dropout)
 - Estimate expected levels of heterozygosity
 - Detect paralogs
 - At least 3 pairs of per-sample replicates recommended
 - Consider divergence among samples

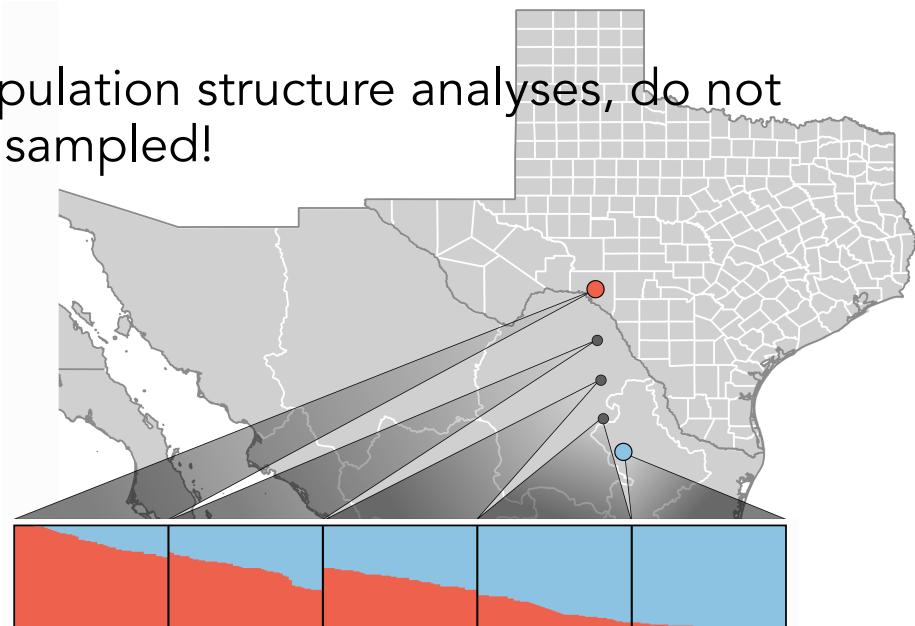
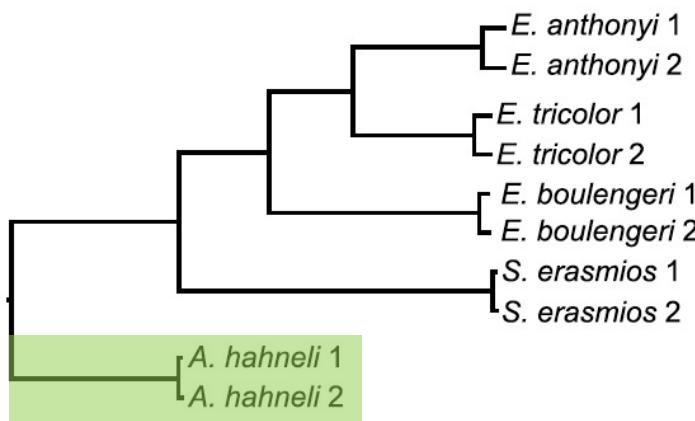


(2) Number and selection of samples



- **Selection**

- Be sure to include outgroups
- For species delimitation and population structure analyses, do not leave large geographic areas unsampled!



(3) Number of loci

- 100s – 1,000s: phylogenetic inference and population genetics
- 1,000s – 100,000s: functionally important SNP identification (mapping)
- Bottlenecks/recent admixture: low genetic diversity; more loci required
- Problematic genomes (polyploids and genomes with high levels of repeat sequences) require more loci and coverage

(4) Type of reads

- **Single-end (SE) sequencing**
 - Single read in one direction from fragment
 - Cheaper, but less data
 - Only option from some sequencing technologies

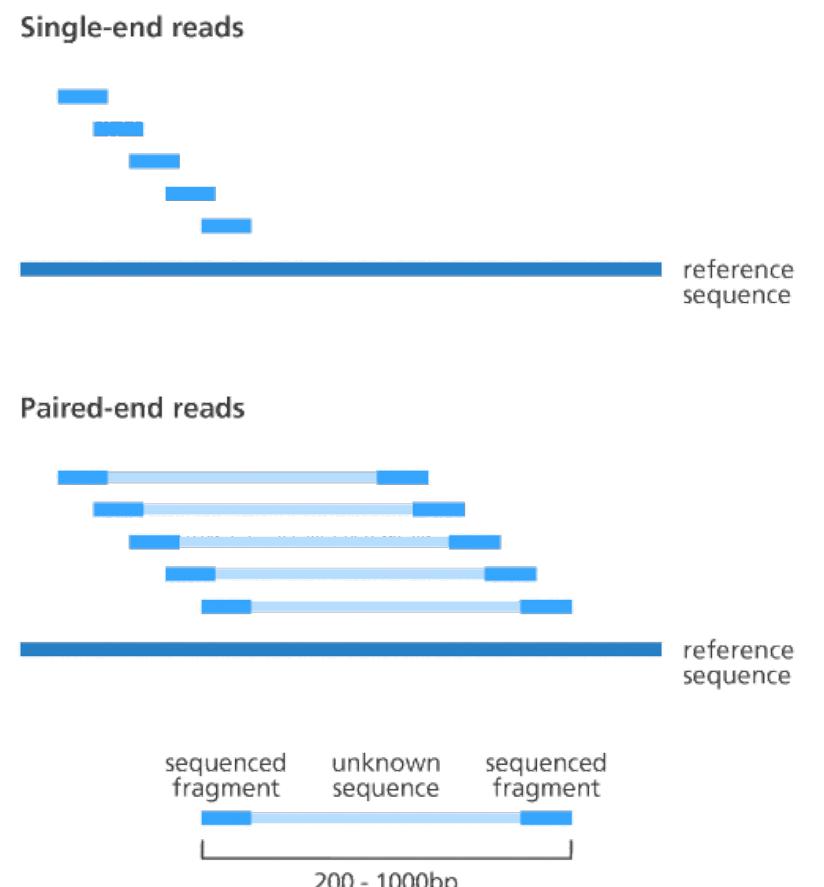


https://ivanek.github.io/analysisOfGenomicsDataWithR/05_Overview_mRNAseq.html.html

Rebecca Tarvin, https://rdtarvin.github.io/RADseq_Quito_2017/

(4) Type of reads

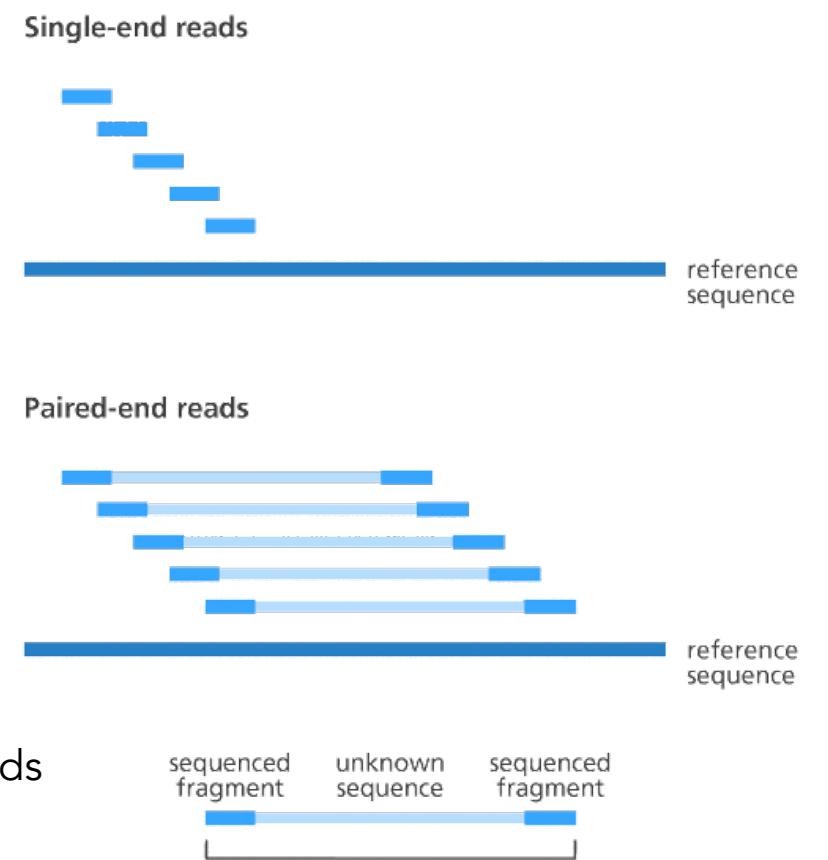
- **Single-end (SE) sequencing**
 - Single read in one direction from fragment
 - Cheaper, but less data
 - Only option from some sequencing technologies
- **Paired-end (PE) sequencing**
 - Two reads from opposite ends of fragment
 - Longer loci, helps with reliable loci reconstruction
 - More data per library prep, but more \$



https://ivanek.github.io/analysisOfGenomicsDataWithR/05_Overview_mRNAseq.html.html
Rebecca Tarvin, https://rdtarvin.github.io/RADseq_Quito_2017/

(4) Type of reads

- **Single-end (SE) sequencing**
 - Single read in one direction from fragment
 - Cheaper, but less data
 - Only option from some sequencing technologies
- **Paired-end (PE) sequencing**
 - Two reads from opposite ends of fragment
 - Longer loci, helps with reliable loci reconstruction
 - More data per library prep, but more \$
- **Length: 50–350 bp (\$ increases with size)**
 - Generally longer is better
 - BUT, high quality short reads > low quality long reads

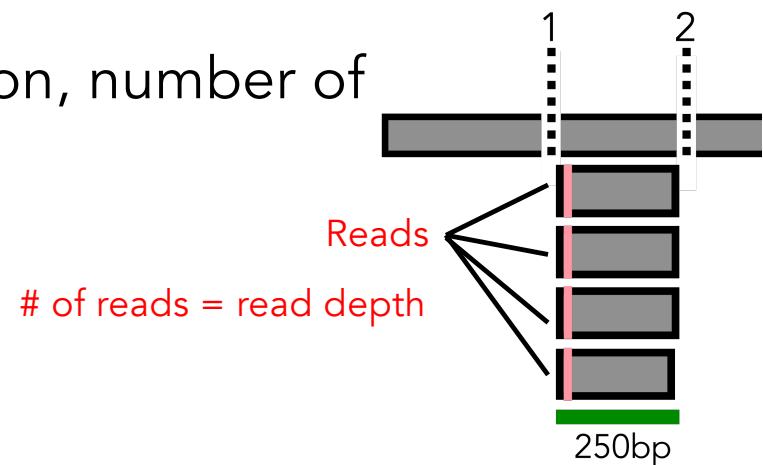


https://ivanek.github.io/analysisOfGenomicsDataWithR/05_Overview_mRNAseq.html.html

Rebecca Tarvin, https://rdtarvin.github.io/RADseq_Quito_2017/

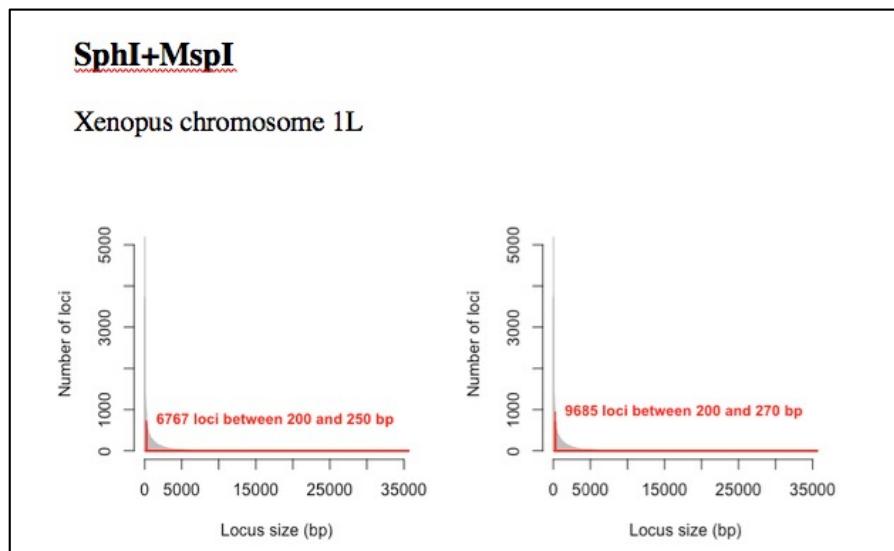
(5) Number of reads

- **Coverage or depth:** number of reads per fragment (1 SE read = 1 set of PE reads)
- More reads are better
 - Recover more loci (need ~5/fragment to generate clusters, ideally more)
- Depends on genome size, enzyme selection, number of expected loci



(6) Enzyme selection

- Search literature to see what has been used in your system
- Check genome size, do *in silico* digestion if possible



Example *in silico* digest using simRAD package in R

(6) Enzyme selection

- Search literature to see what has been used in your system
- Check genome size, do *in silico* digestion if possible
- Longer recognition sequence = fewer cuts = fewer loci (better for large genomes!)
- For ddRAD, mix 6-bp ("rare") and 4-bp ("frequent") cutters

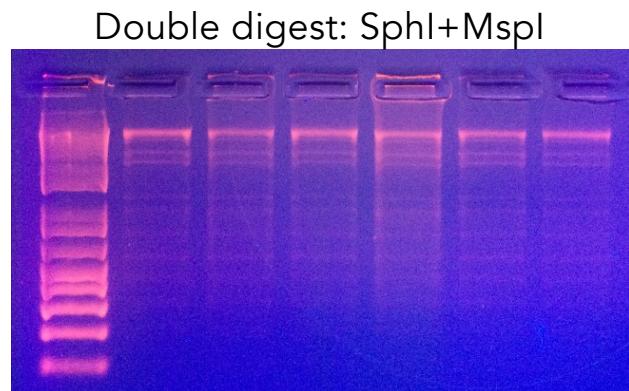
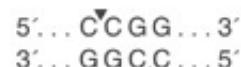


SphI



be INSPIRED
drive DISCOVERY
stay GENUINE

MspI



Rebecca Tarvin, https://rdtarvin.github.io/RADseq_Quito_2017/

(6) Enzyme selection

- Search literature to see what has been used in your system
- Check genome size, do *in silico* digestion if possible
- Longer recognition sequence = fewer cuts = fewer loci (better for large genomes!)
- For ddRAD, mix 6-bp ("rare") and 4-bp ("frequent") cutters
- Perform double digests, run them on Bioanalyzer

(7) Sequencing machine

- Illumina HiSeq (do not use HiSeq4000):
 - 50 – 150bp (up to 300bp total for PE fragments)
- Illumina MiSeq:
 - <300bp (600 for PE)
- Ion Torrent:
 - <400bp (only SE)
- *Not usually used for RAD: PacBio (<20,000bp, very long sequences); La Roche (<750bp, high error rate)*

How to save money

- Be smart in your study design (reduce PCR duplicates, enzyme selection, etc.)
- Collaborate with other groups at your institution to share more expensive supplies like adaptors
- Make your own beads (\$1/mL vs. \$0.03/mL)
- Research prices at different sequencing facilities