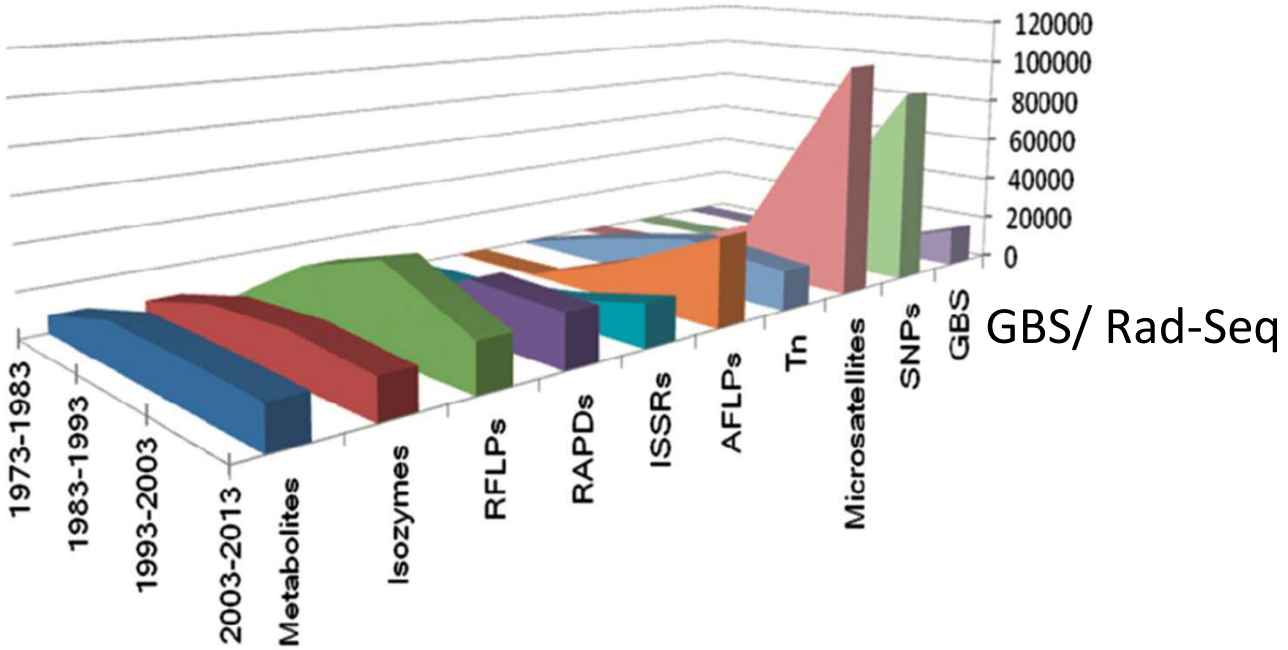
“Restriction enzyme methods for reduced genomic representation”

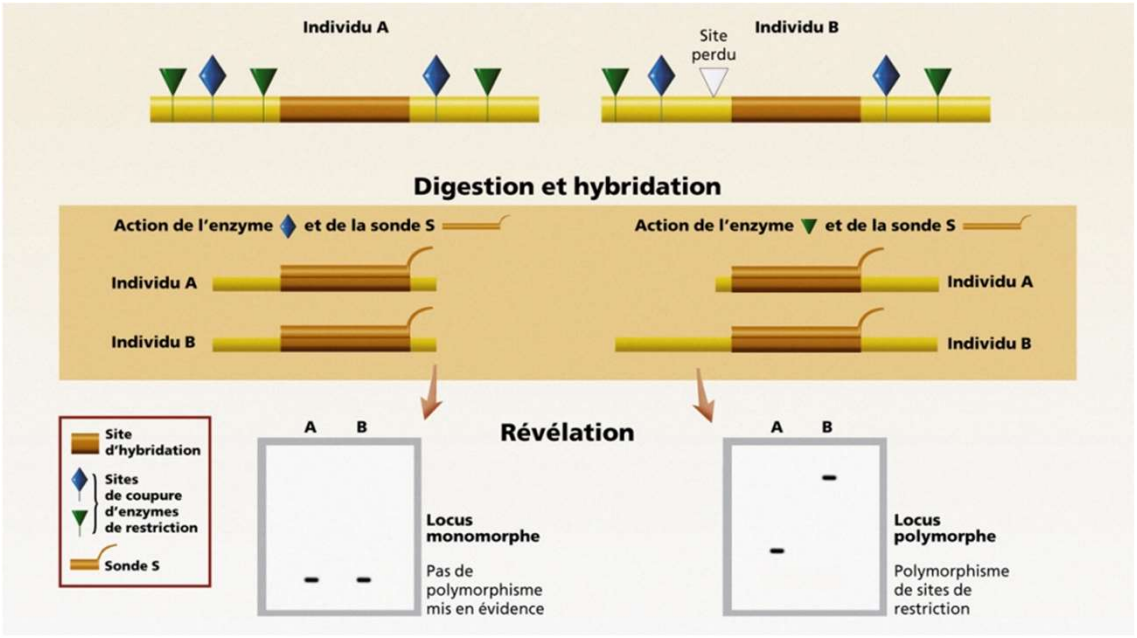
Genotyping for Population Genomics (and shallow phylogenetics / phylogeography)



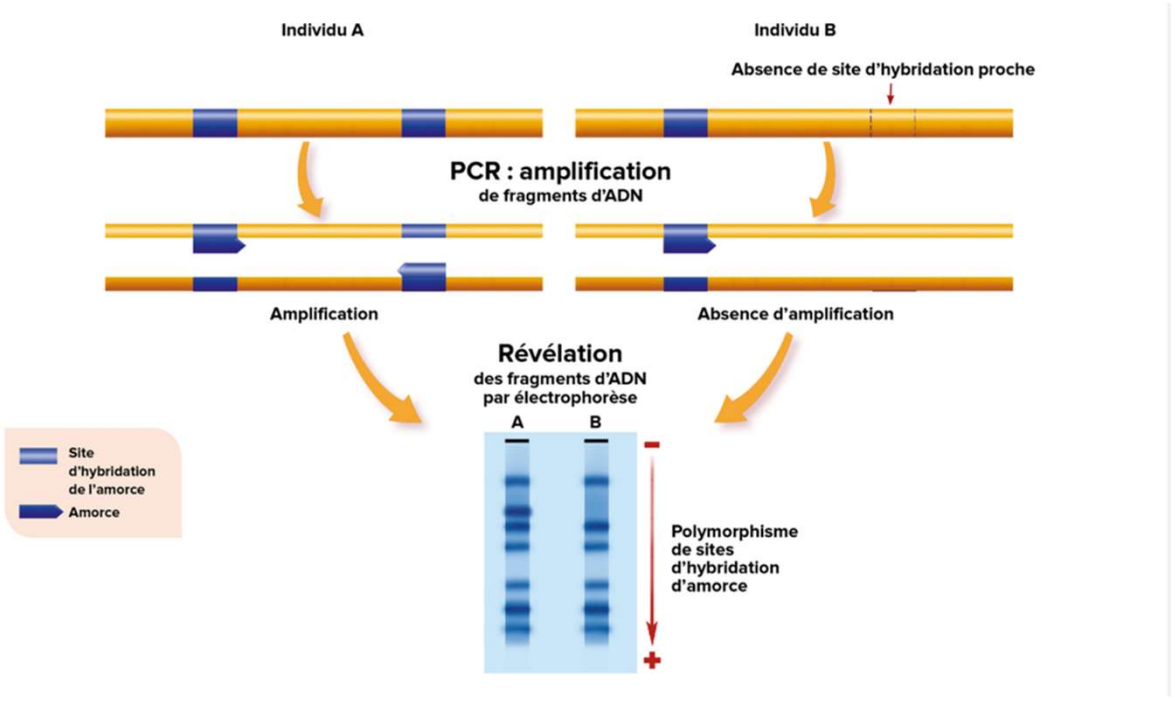
Popularity of different molecular marker systems over the past 50 years.

Grover & Sharma 2014

Restriction enzyme methods



RFLP: *Restriction Fragments Length Polymorphism*

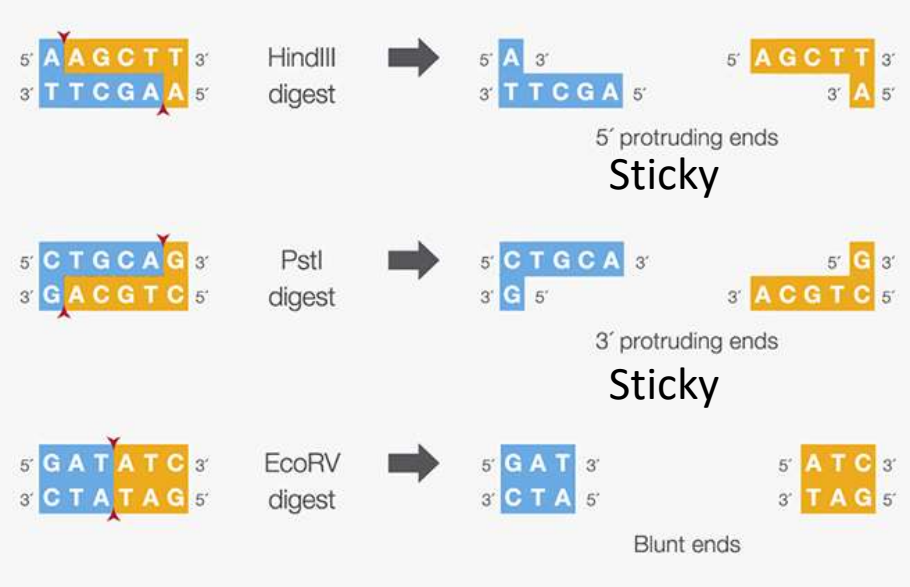


AFLP: *Amplification Fragment Length Polymorphism*

Restriction enzyme methods: RAD-Seq and derived methods

RAD-Seq = Restriction Site-Associated DNA Sequencing

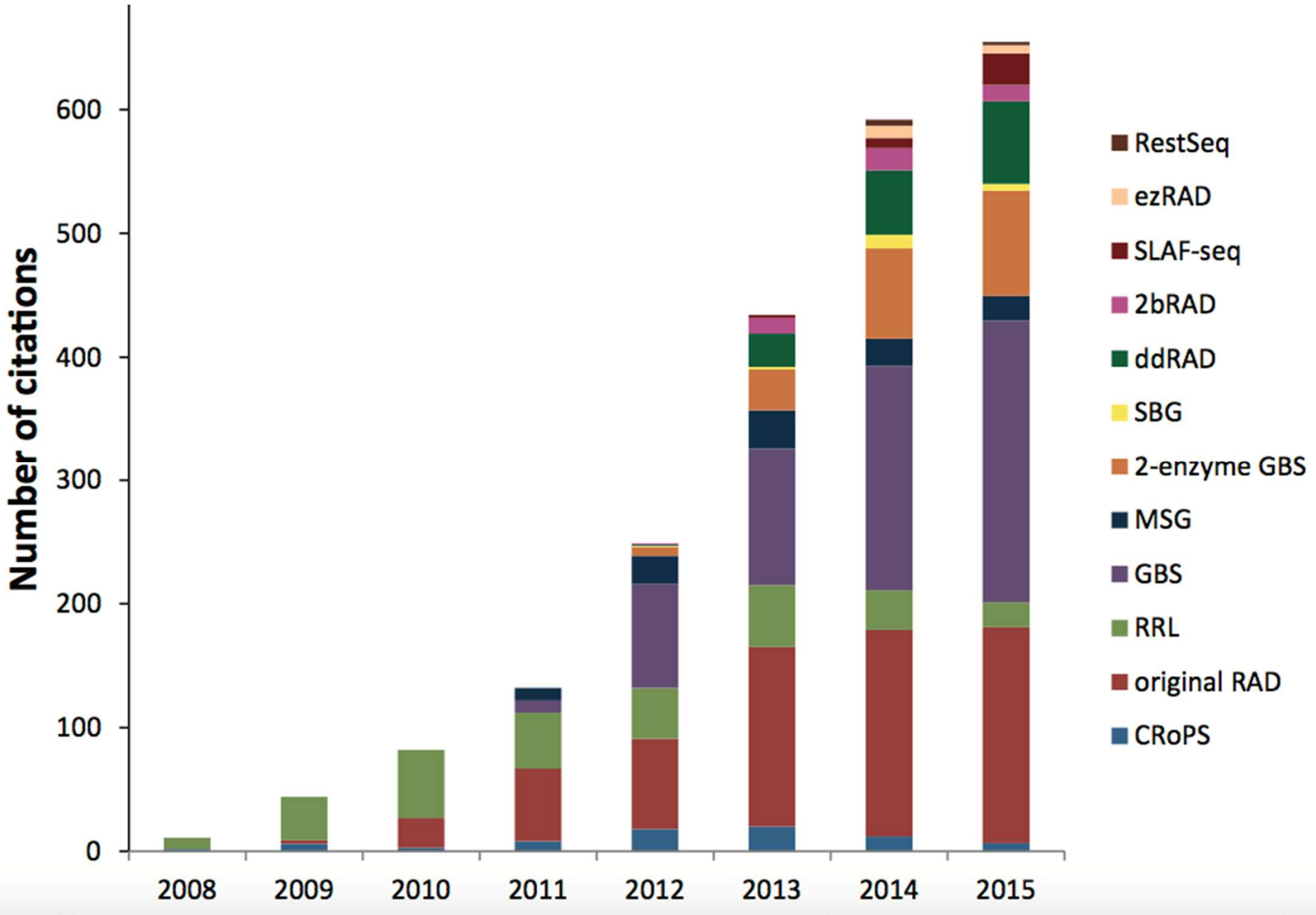
1- Genome complexity is reduced by cutting the DNA with restriction enzymes.



2- Only sequences associated with restriction sites are sequenced

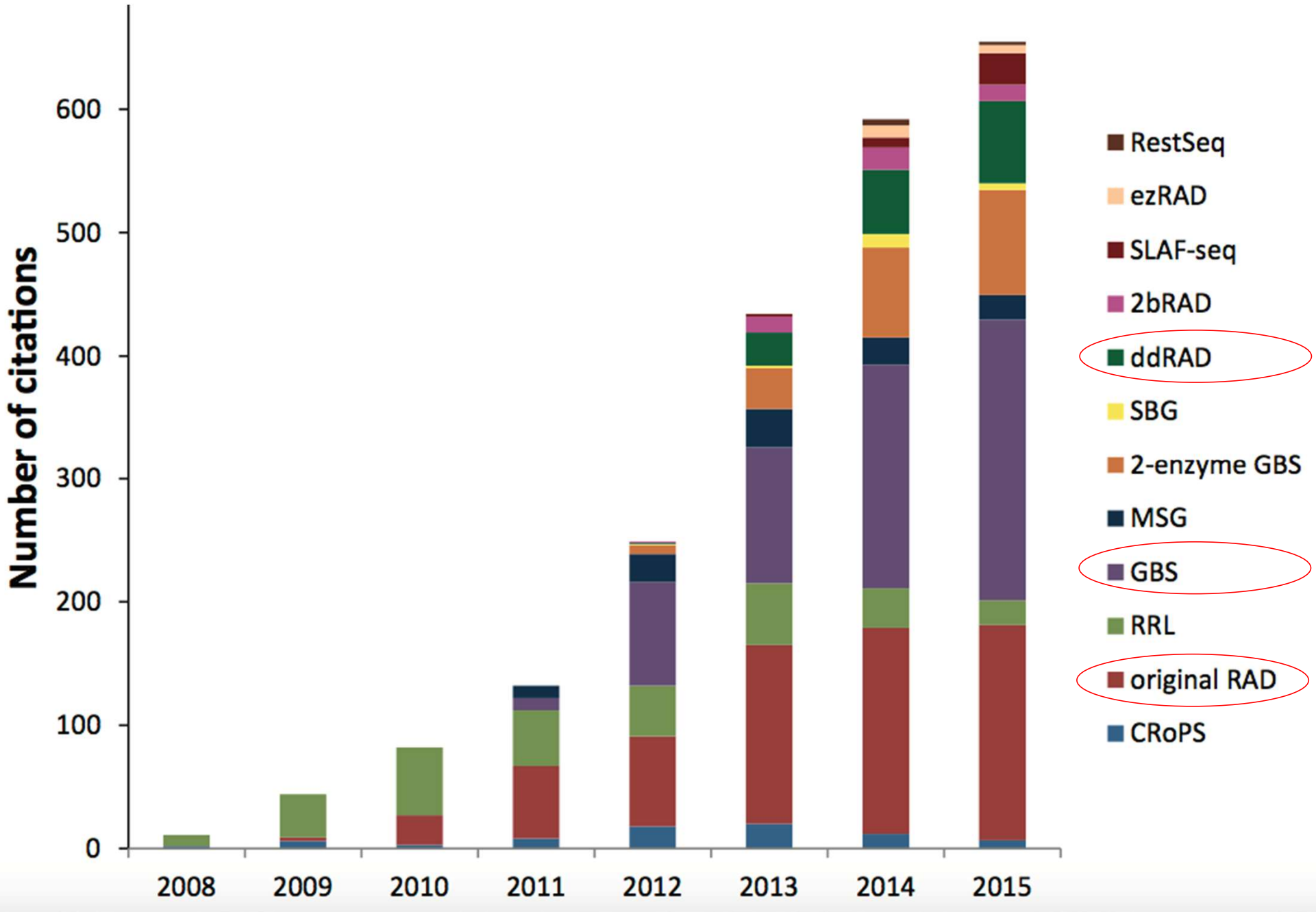
3- There is a whole series of methods derived from the initial protocol

Restriction enzyme methods: RAD-Seq and derived methods



Andrews et al. 2016

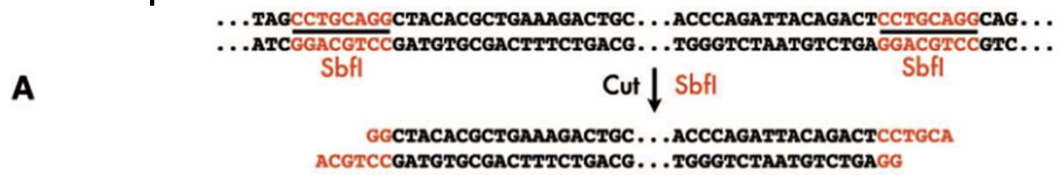
Restriction enzyme methods: RAD-Seq and derived methods



Andrews et al. 2016

Restriction enzyme methods: RAD-Seq and derived methods

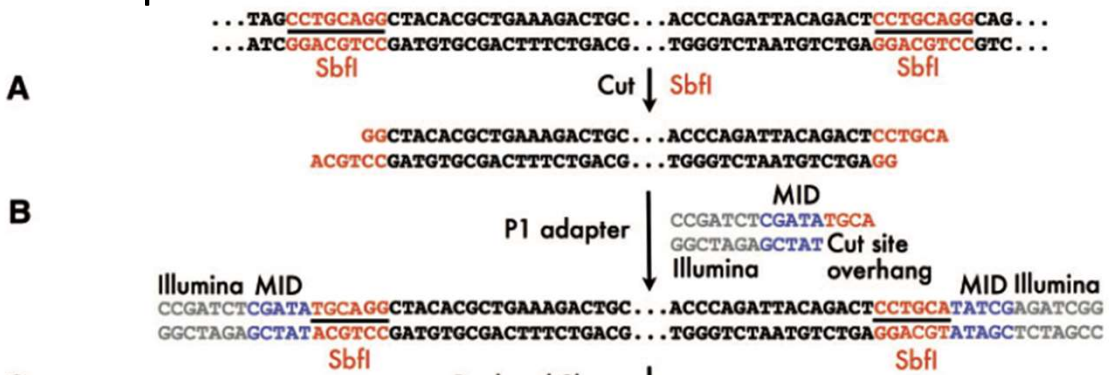
The process of RADSeq



Davey et al. 2011

Restriction enzyme methods: RAD-Seq and derived methods

The process of RADSeq



Davey et al. 2011

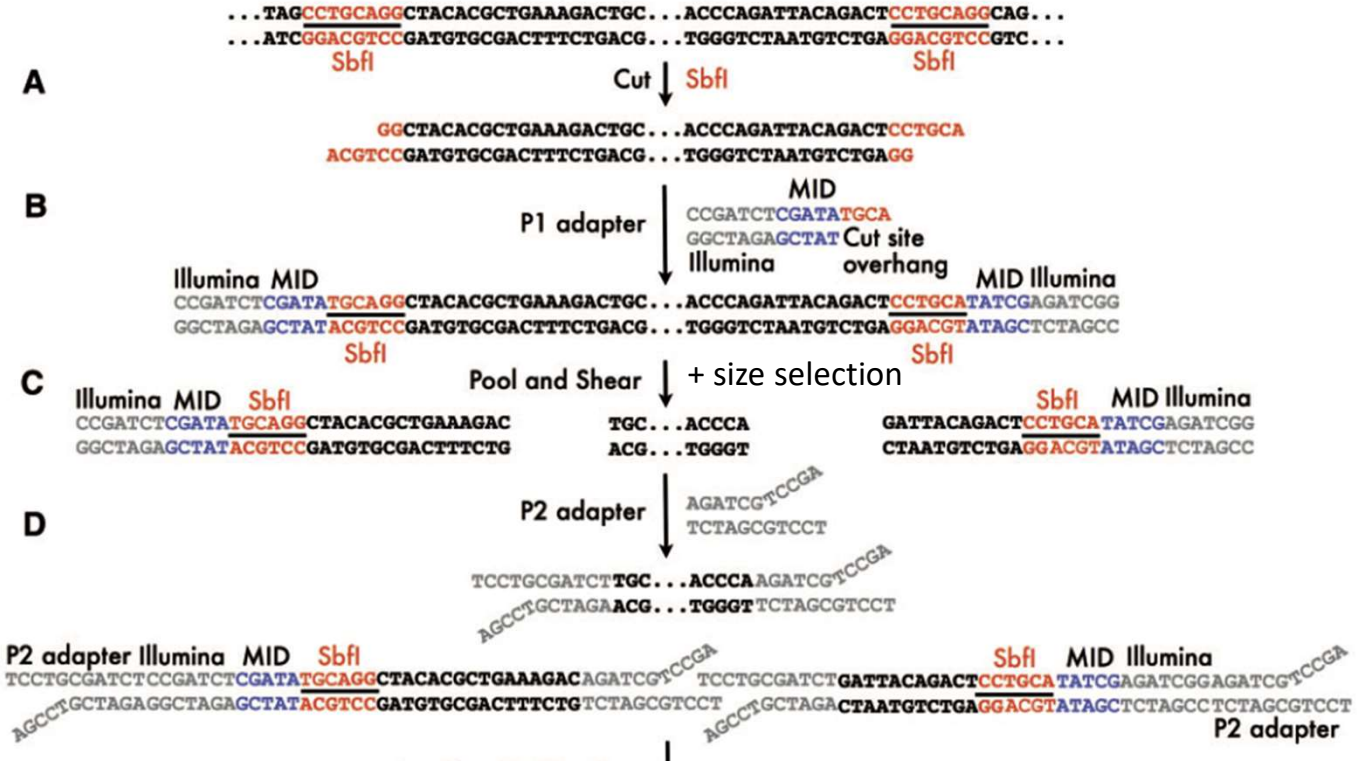
Restriction enzyme methods: RAD-Seq and derived methods

The process of RADSeq

Davey et al. 2011

Restriction enzyme methods: RAD-Seq and derived methods

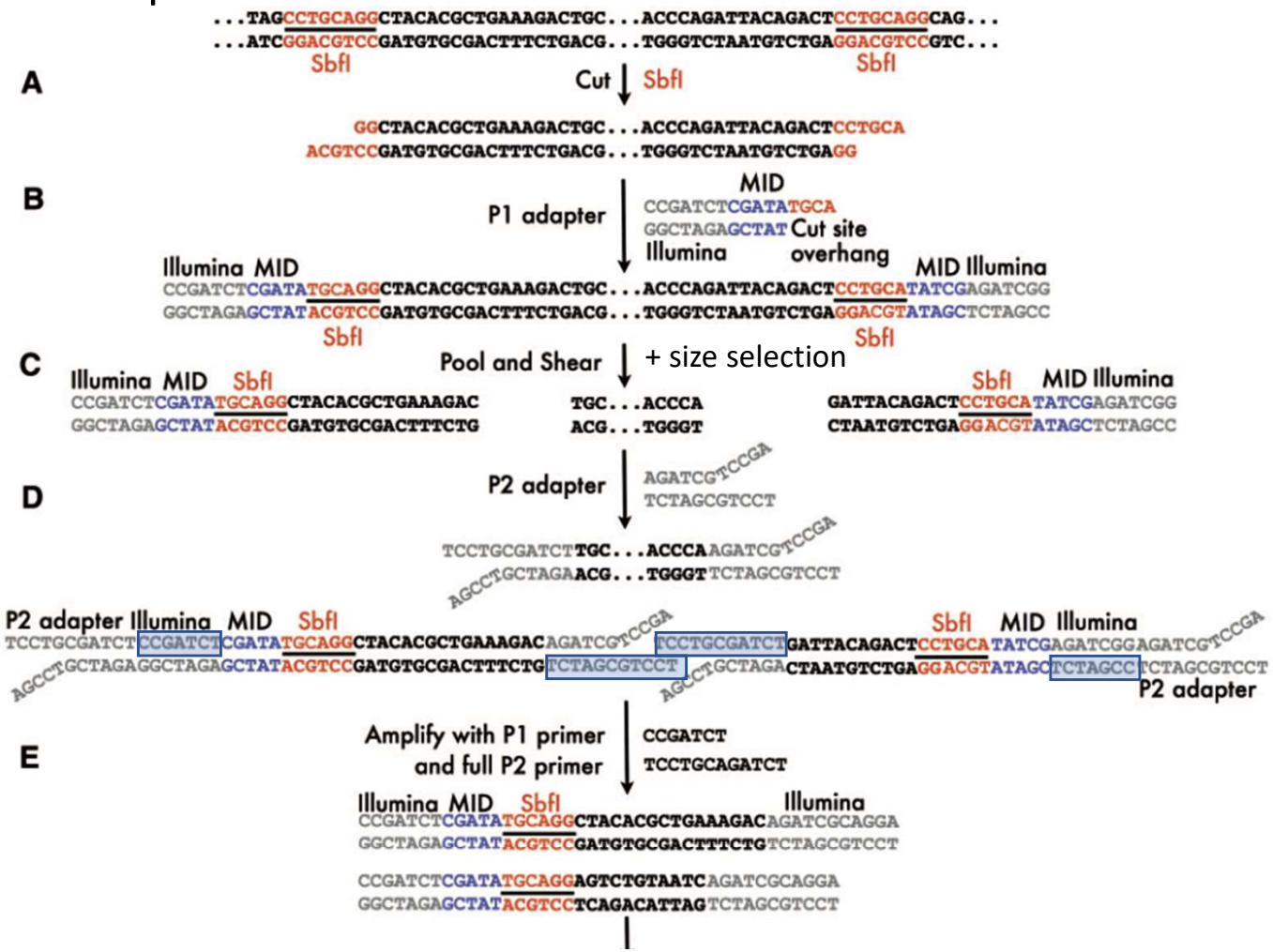
The process of RADSeq



Davey et al. 2011

Restriction enzyme methods: RAD-Seq and derived methods

The process of RADSeq

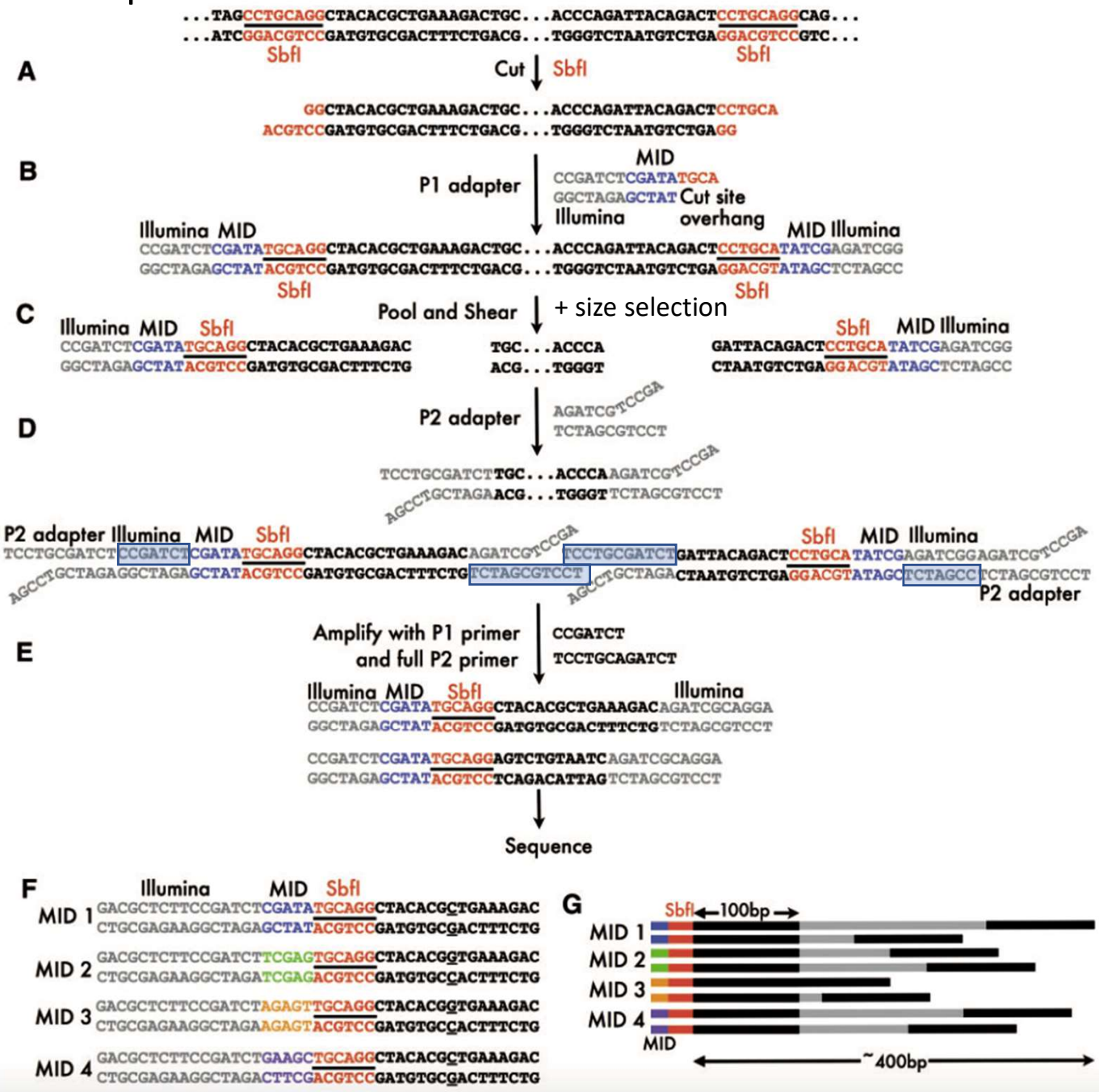


Davey et al. 2011

Restriction enzyme methods: RAD-Seq and derived methods

The process of RADSeq

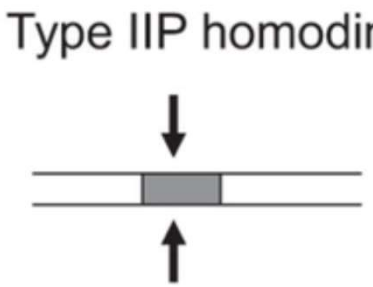
Davey et al. 2011



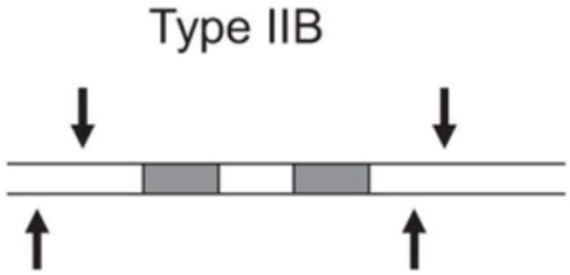
Restriction enzyme methods: RAD-Seq and derived methods

1- Sequencing of fragments adjacent to single restriction enzyme cut sites:

- original **RADseq** (Baird et al. 2008)
- **2bRAD** (Wang et al. 2012) => type IIb restriction endonucleases



'Orthodox'IIP enzymes (e.g. EcoRI, EcoRV) cut at the recognition site



Type IIB require two recognition sites and cut on the outside (e.g. BpII)

Restriction enzyme methods: RAD-Seq and derived methods

1- Sequencing of fragments adjacent to single restriction enzyme cut sites:

- original **RADseq** (Baird et al. 2008)
- **2bRAD** (Wang et al. 2012) => type IIb restriction endonucleases

2- Sequencing of fragments flanked by two restriction enzyme cut sites

A- Single enzyme, indirect size selection

- **GBS** – genotyping by sequencing (Elshire et al. 2011)
- **SBG** – sequence-based genotyping (Truong et al. 2012)

B- Double enzyme, indirect size selection

- **CRoPS** – complexity reduction of polymorphic sequences (Orsouw et al. 2007)

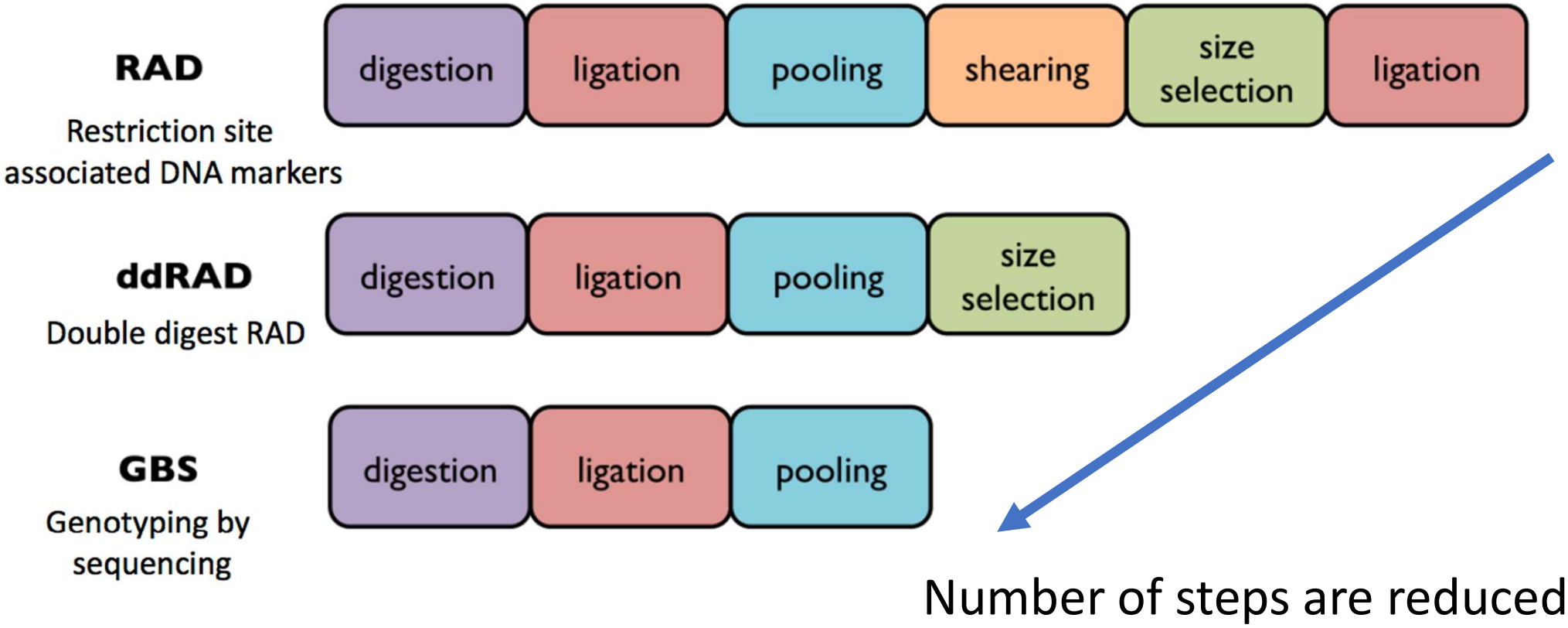
C- Single enzyme, direct size selection

- **RRLs** – reduced representation libraries (van Tassel et al. 2008)
- **MSG** – multiplexed shotgun genotyping (Andolfatto et al. 2011)
- **ezRAD** (Toonen et al. 2013)

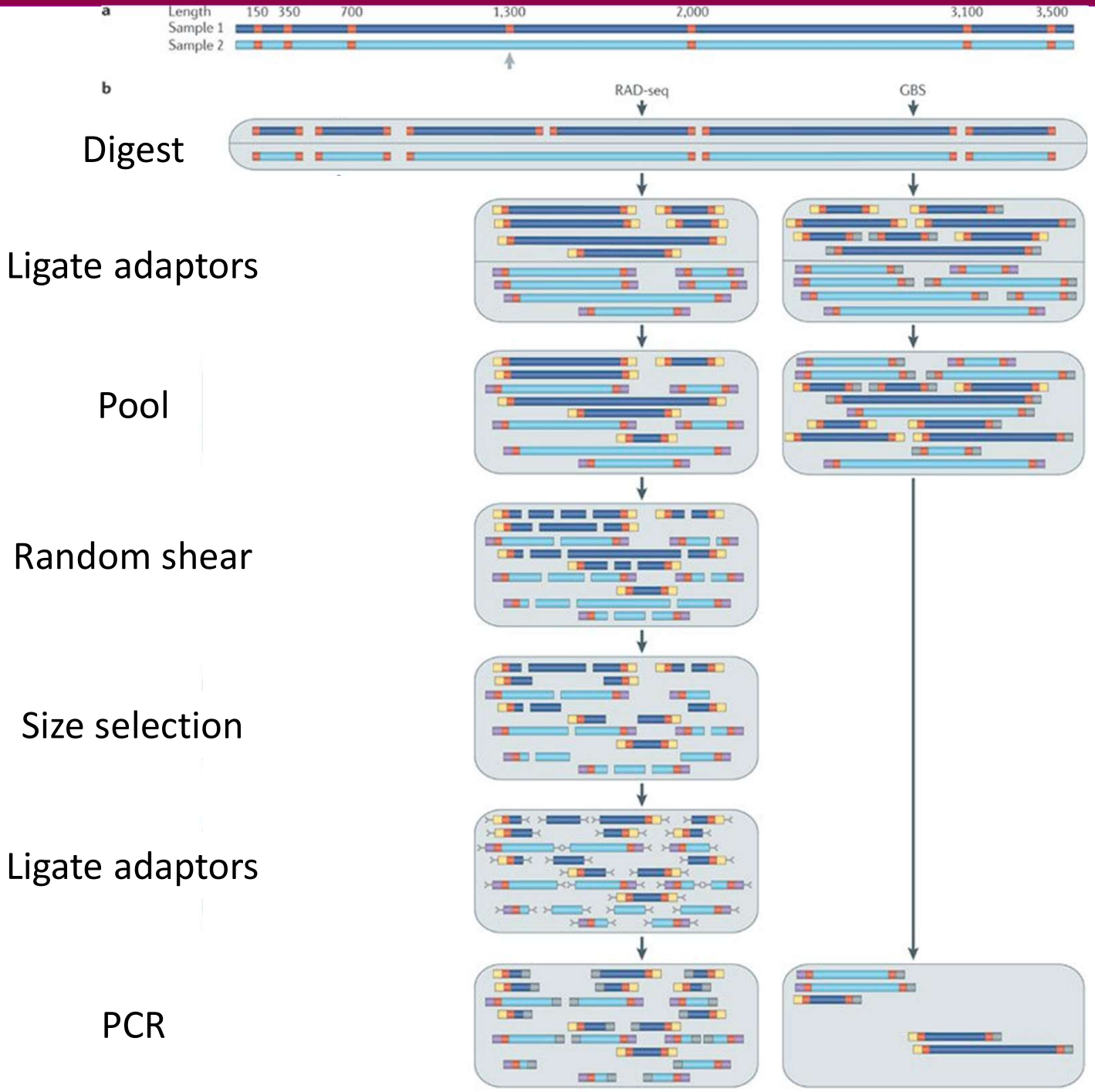
D- Double enzyme, direct size selection

- **ddRAD** – double-digest RAD (Peterson et al. 2012)

Restriction enzyme methods: RAD-Seq and derived methods



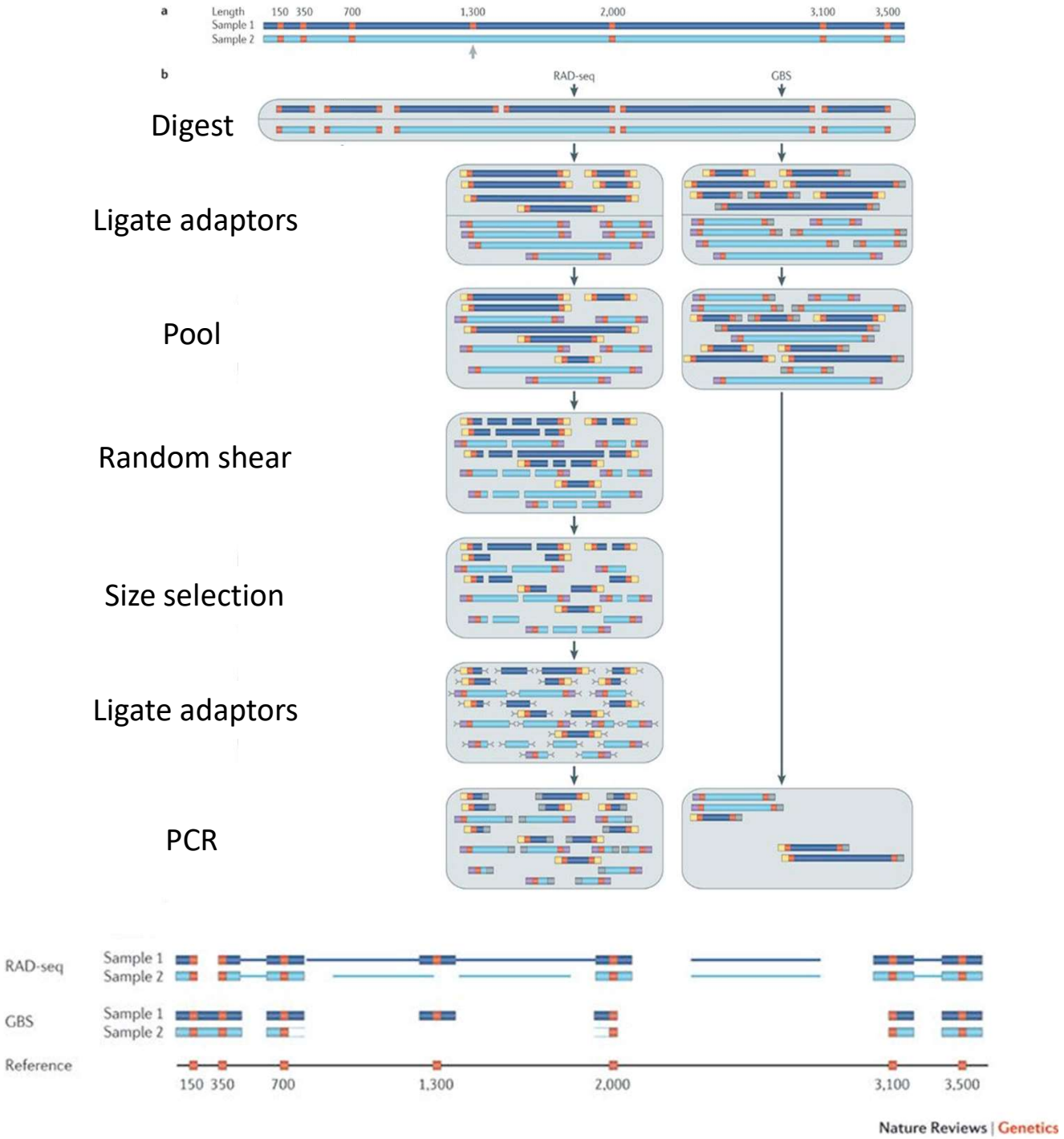
Restriction enzyme methods: RAD-Seq and derived methods



Davey et al. 2011

Restriction enzyme methods: RAD-Seq and derived methods

Davey et al. 2011



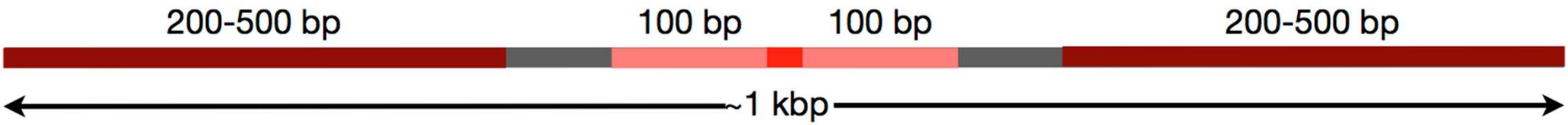
Restriction enzyme methods: RAD-Seq and derived methods

Methods comparison

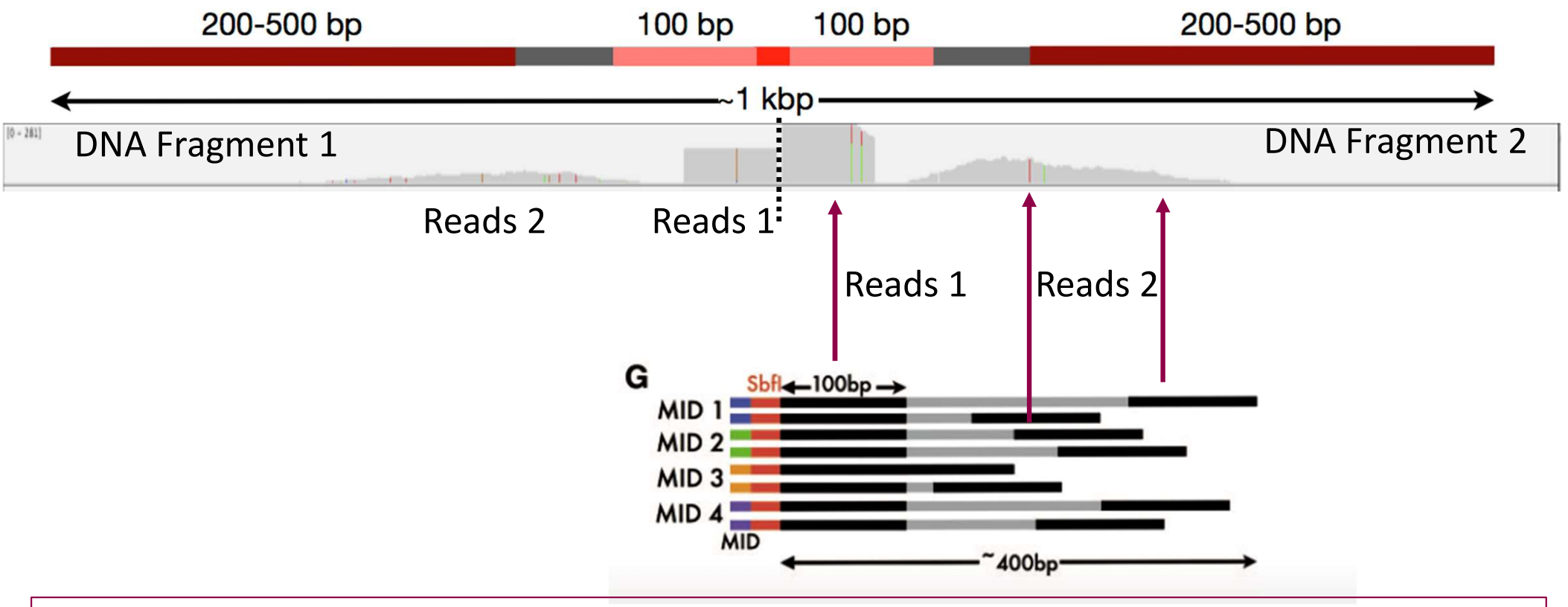
	Original RAD	2bRAD	GBS	ddRAD
Options for tailoring number of loci	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme	Change restriction enzyme or size selection window
Number of loci per 1 Mb of genome size*	30–500	50–1,000	5–40	0.3–200
Length of loci	≤1kb if building contigs;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

Andrews et al. 2016

RAD-Seq bioinformatics



RAD-Seq bioinformatics

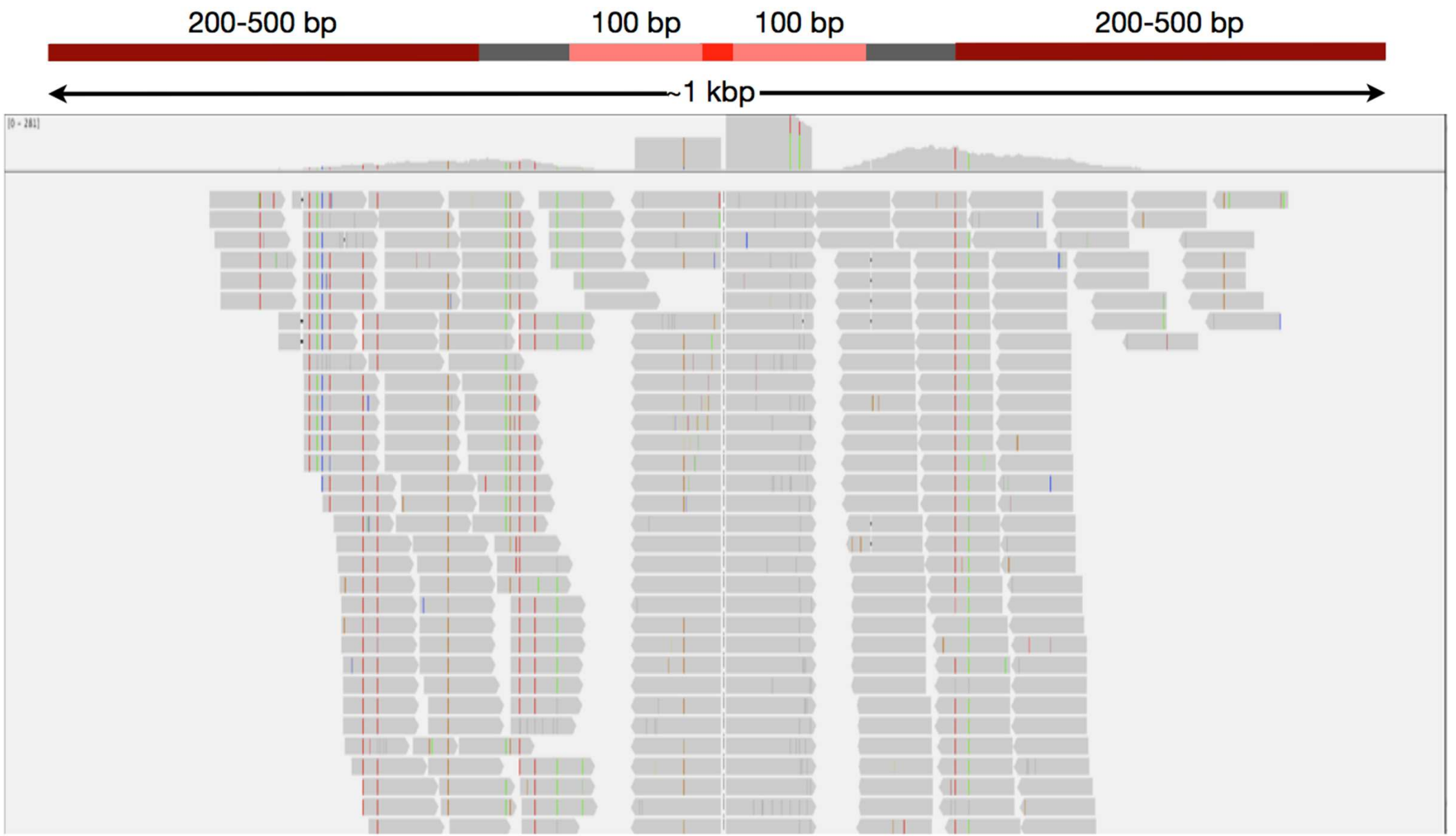


The coverage in DNA sequencing is the number of unique reads that include a given nucleotide in the reconstructed sequence.

The sequencing depth (also known as read depth) is the ratio of the total number of bases obtained by sequencing to the size of the genome or the average number of times each base is measured in the genome

Coverage and sequencing depth are basically the same thing

RAD-Seq bioinformatics



RAD-Seq bioinformatics

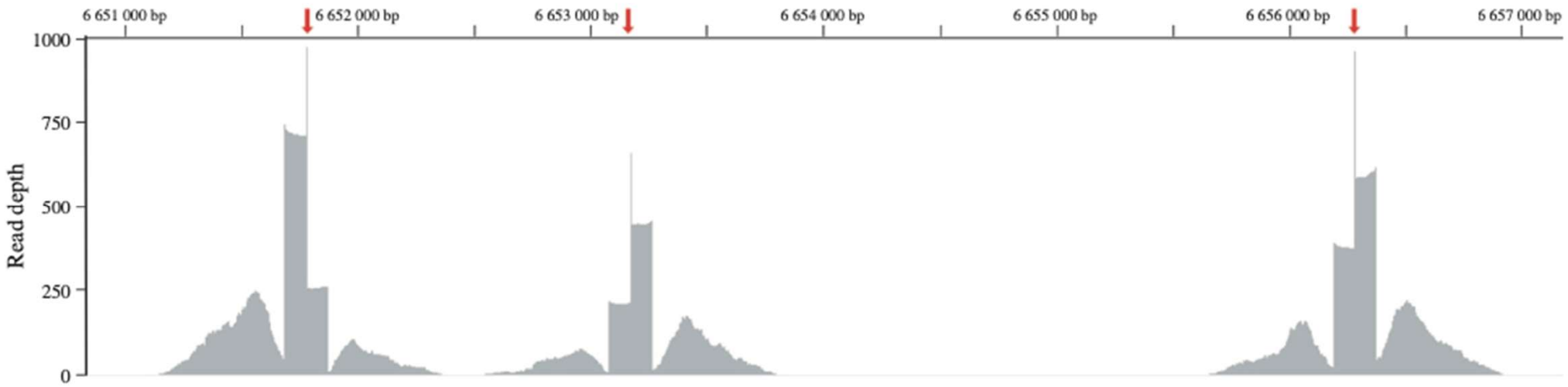


Fig. 1 Characteristic pileup of RAD-Seq data. Three PstI restriction sites (red arrows) in *Caenorhabditis elegans* chromosome I (6.651 Mb–6.657 Mb) are covered both upstream and downstream by RAD-Seq raw reads (dark grey). Read 1 sequences are piled up in stacks either side of each restriction site; read 2 sequences are spread out in heaps up to 700 bp beyond the restriction site. The restriction site overhang TGCA is covered by reads belonging to both upstream and downstream RAD loci, producing narrow peaks of read coverage at the restriction sites. RAD loci on either side of a restriction site have different read depths; however, loci from the same restriction fragment have similar read depths. Bases in the read 2 regions are covered at much lower depth overall; read 2 sequences also partially cover the read 1 regions, as seen by the increase in read depth at bases towards the end of read 1 regions.

The coverage in DNA sequencing is the number of unique reads that include a given nucleotide in the reconstructed sequence.

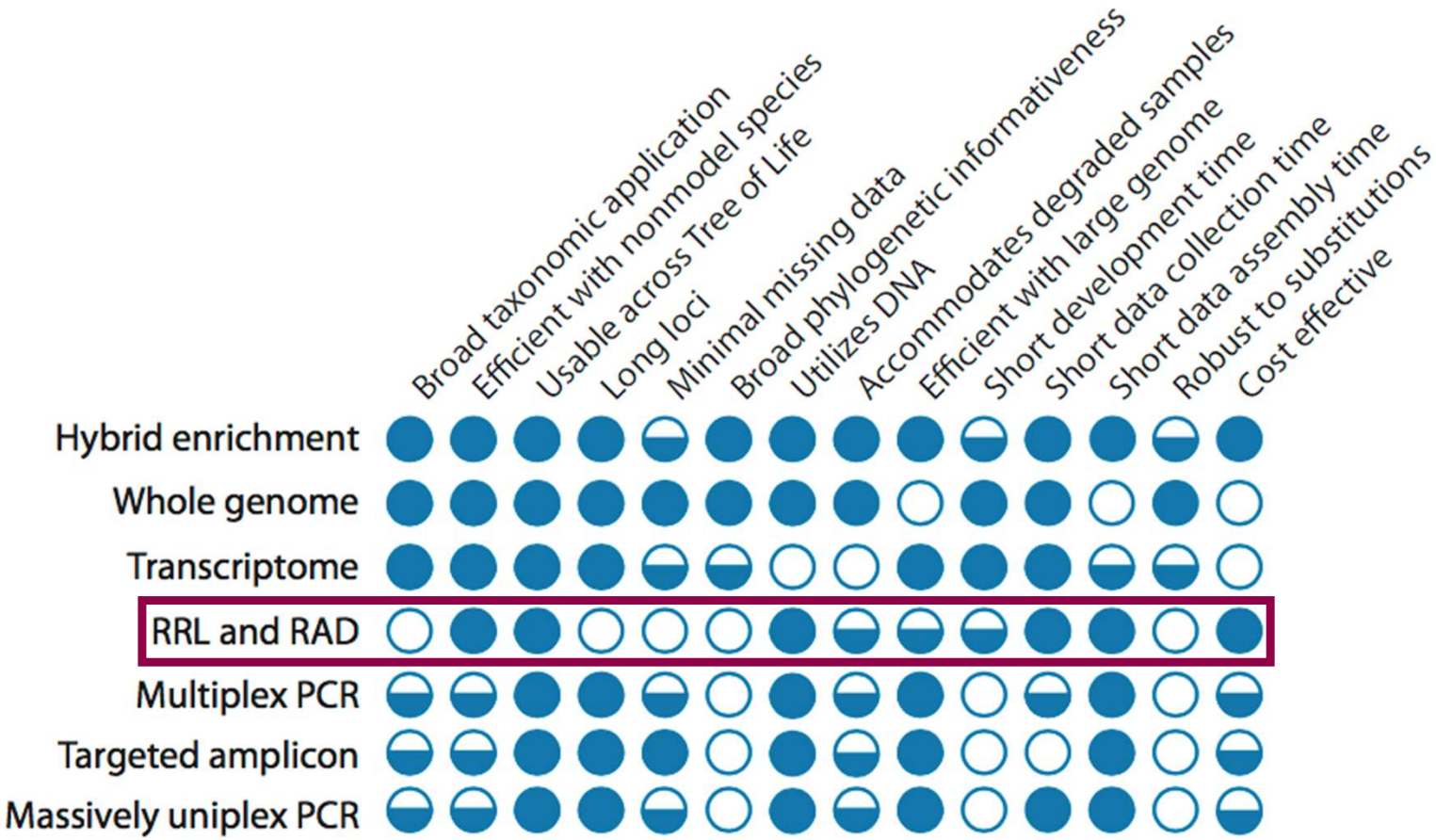
The sequencing depth (also known as read depth) is the ratio of the total number of bases obtained by sequencing to the size of the genome or the average number of times each base is measured in the genome

Coverage and sequencing depth are basically the same thing

Davey et al. 2013

Properties of RAD-Seq data

- wide genomic distribution
- relatively short loci
- allelic dropout/null alleles
- large proportion of missing data
- orthology/paralogy – bioinformatic assessment



Lemmon & Lemmon 2013

Applications des données de RAD-Seq

- shallow phylogenomics
- population structure, phylogeography
- population genomics
- evolution of recently radiated groups
- hybridization, introgression
- genetic mapping
- …