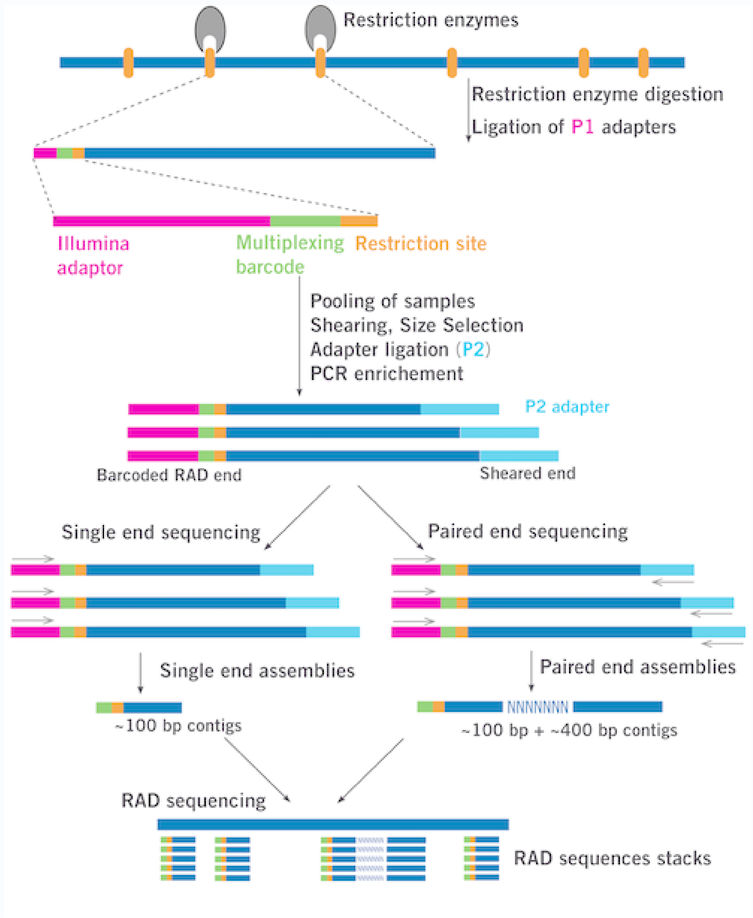
RAD-Seq: A Powerful New Tool for Next Generation Applications

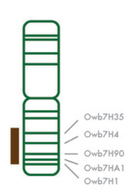


What is it? Restriction-site Associated DNA Sequencing or RAD-Seq, is a method of high-throughput DNA sequencing that are used to identify DNA sequences that flank restriction sites associated with a particular restriction enzyme throughout the genome. Once RAD tags have been isolated, they can be used to identify and genotype DNA sequence polymorphisms, such as single nucleotide polymorphisms (SNPs).

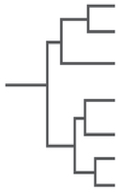
How does it work? RAD-Seq works by first fragmenting the target genome using a restriction enzyme. After digestion, a series of molecular processing steps transform the DNA into a fragment library suitable for sequencing on a NGS platform. Depending on the end goal, either single end or paired end sequencing can be used to generate the appropriate amount of information. Sequence data is then analyzed to identify and score genetic variations in the samples or population of interest.

When is it useful? RAD-Seq has many applications including identification of paralogs, linkage groups or other genomic regions under different selection pressure.

Linkage and QTL Mapping



Phylogenetic Analysis



Population Structure

Association Mapping

Benefits:

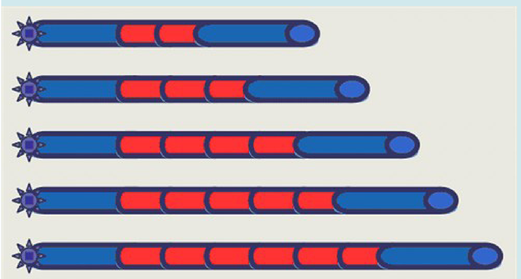
- ◊ No reference genome required
- ◊ Can be assembled de novo.
- ◊ Use multiple restriction enzymes to increase the number of markers
- ◊ Useful for organisms with reduced genetic variation such as clonal populations.
- ◊ SNPs are no longer restricted to model organisms for which sequences of full genomes are available.
- ◊ Simultaneous verification & scoring
- ◊ Compatible with MID tags for multiplexing
- ◊ Reduces complexity of genome by sub-sampling only at specific sites.
- ◊ Compatible with a random approach.

Limitations:

- ◊ Expensive
- ◊ Limitations on sample number
- ◊ Large datasets produced can be noisy

Comparing RAD-Seq with Microsattelite Approaches:

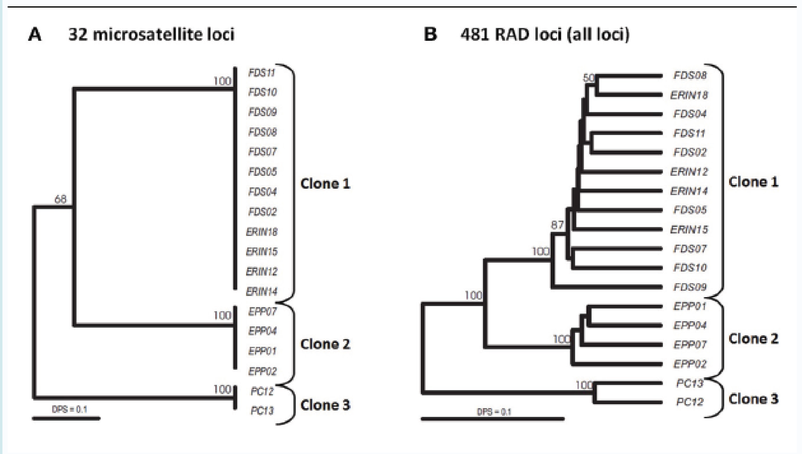
Microsatellite are tracts of DNA containing 2-8 bp motifs typically repeated 5-50 times. Also known as SSR (simple sequence repeats) or STRs (short tandem repeats), Microsatellites can be located in non-coding, regulatory or coding regions. They are highly mutable markers, and do not suffer from low heterozygosity.



Microsatellites are flanked with flourescent PCR primers and then amplified, resulting in a pair of allelic products which vary in size according to their repeat length. Traditionally, primer design was a costly-iterative process. High-throughput arrays of Single Nucelotide Polymorphisms (SNPs) remove this bottleneck but not the discovery process.

Microsatellites are also often specific to the population in which they were developed, making them unsuitable for highly divergent populations.

Case Study: Clonal patterns of a highly inbred endangered killifish



- ◊ Hundreds of SNPs from the Rad-seq analysis had better resolving power than multi-locus SSR.
- ◊ Successfully deciphered even sub-clonal populations
- ◊ SNP method could not distinguish between true heterozygotes and artificially assembled variants with similar RAD-tag loci.
- ◊ RAD-seq recovered finer population structure.

References: 1) Mesak, F. Tatarenkov, A., Earley, R.L., and J.C. Avise. (2014). Hundreds of SNPs vs. dozens of SSRs: which dataset better characterizes natural clonal lineages in a self-fertilizing fish? Frontiers in Ecology and Evolution. 2:74 DOI: 10.3389/fevo.2014.00074. 2) Davey, J.W. and M.L. Blaxter. (2010). RADSeq: next-generation population genetics. Breifings in Functional Genomics ((5-6):416-423. doi: 10.1093/bfgp/elq031. 3) Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and M.L. Baxter. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics. 12: 499-510. doi:10.1038/nrg3012.