RADseq

BCB 504: Applied Bioinformatics

Matt Settles

University of Idaho
Bioinformatics and Computational Biology Program

April 29, 2013

1 Introduction

RAD-seq Introduction

Restriction-site associated DNA (RAD) markers are a type of genetic marker which can be useful for

- Association mapping
- QTL-mapping
- population genetics (population structure)
- ecological studies
- pedigree mapping
- evolution (phylogeny)

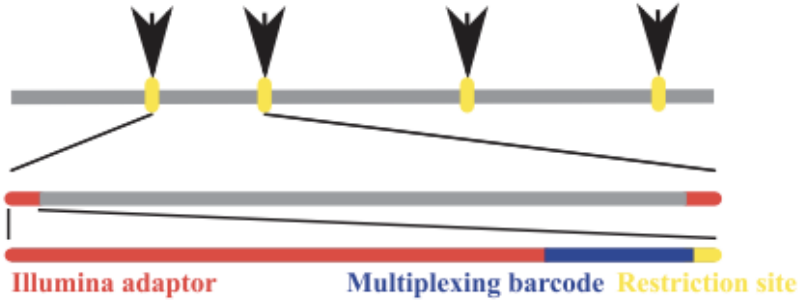
RAD-seq is a genomic reduction (reduced representation) technique, where you only sequence regions neighboring a particular restriction site(s) allowing you to (1) reduce costs associated with library preparation and (2) pool many more individuals per sequencing run.

Restriction fragment

A restriction fragment is a DNA fragment resulting from the cutting of DNA by a restriction enzyme. Each restriction enzyme is highly specific, recognizing a particular short DNA sequence (restriction site), and cutting both strands at specific points within the site. Most restriction sites are palindomic are are 4 to 8 nucleotides long. The shorter the rescriction site the higher number of probable occurances in the genome.

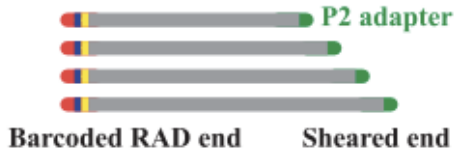
RAD-seq Library Prep

1.Restriction enzyme digestion



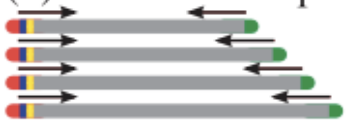
2. Ligation of P1 adapters (one barcoded adapter/individual)

3. Pooling of individual, shearing (300–800 bp) and ligation of P2 adapters



4(a) Single end sequencing

4(b) Paired end sequencing



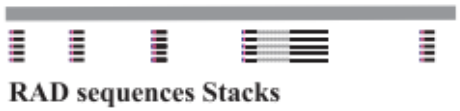
5(a) Single end assemblies

108 bp contigs

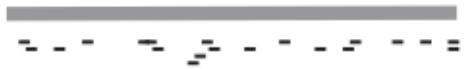
5(b) Paired end assemblies

108 + 400 bp contigs

6. RAD sequencing

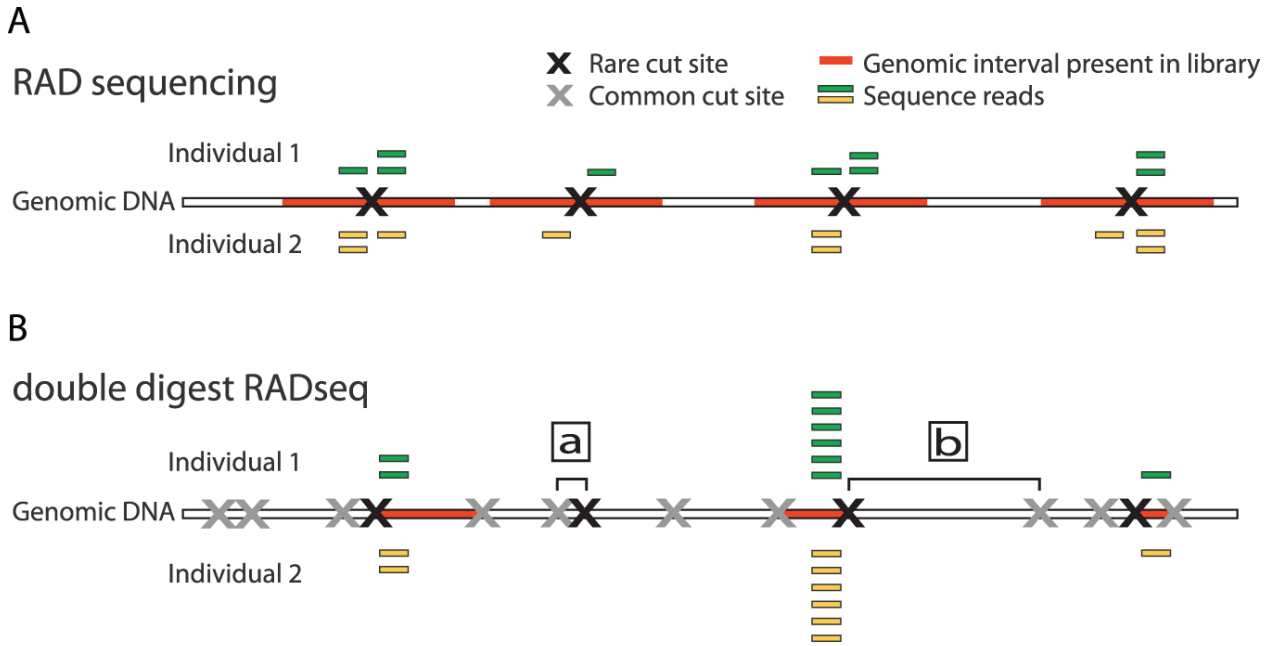


Shotgun sequencing



Double Digest RADseq

Double Digest RADseq, uses two cutters (one rare and one common) in order to further reduce the number of RAD tags and descrease the cost and complexity of the library preparation. ddRADseq removes the need to randomly shear the DNA and blunt end repair, allowing for initial DNA quantities as low as 100ng.



Other similar techniques

RRL	Reduced Represetation Library
GBS	Genotype By Sequencing
CRoPS	Complexity Reduction of Polymorhpic Sequences
MSG	Multiplex Shotgun Genotyping

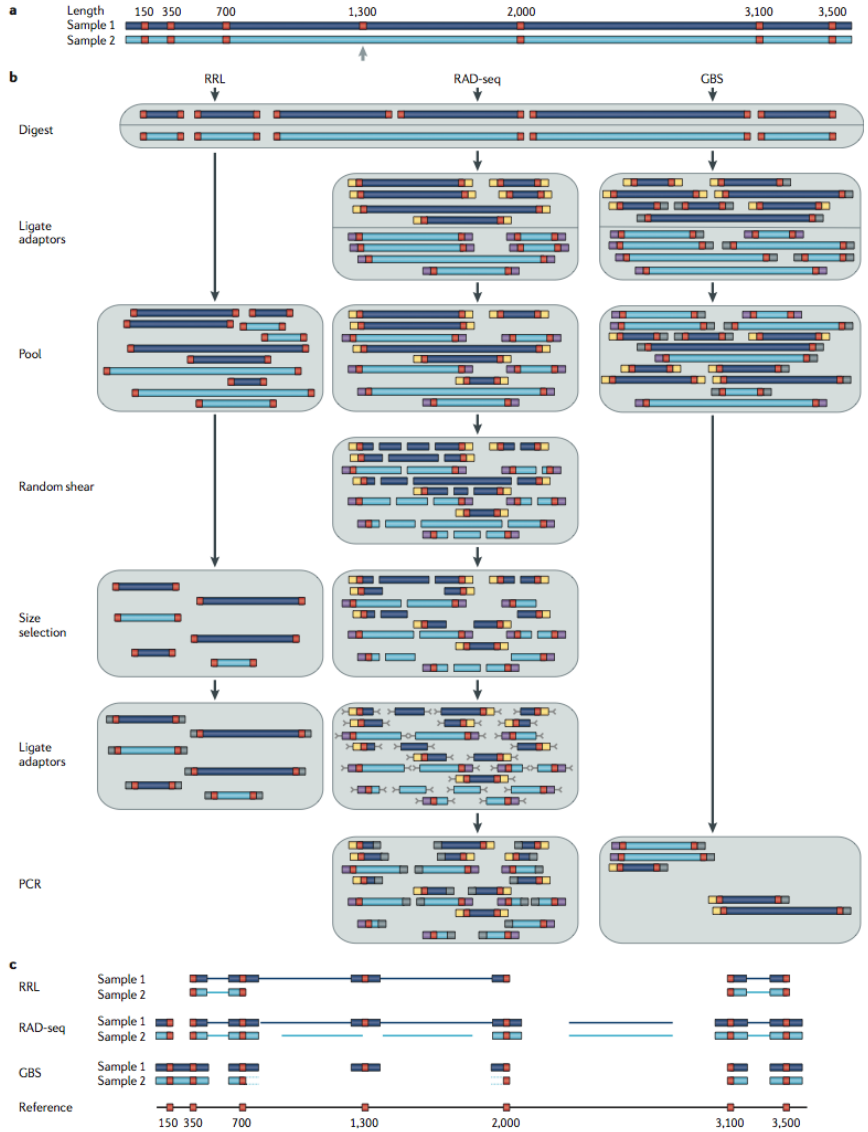
All techniques rely on restrction enzyme digestion and size selection on a gel

RRL, GBD, RAD technique comparison

Restriction-site Associated DNA tags (RADseq)

Matt Settles

Introduction



Advantages

- Requires no prior knowledge of genome (unlike SNP-chip)
- Has no ascertainment bias
- Can pool many, many individuals
- Tags are homologous across individuals

Experimental Design

Number of sites versus Depth of sequencing per site versus number of samples.

- Rescriction site and genome size correlate with the number of tags.
- Depth per tag - target mean depth of 20-40 to call genotypes
- Use number of sequencing reads and above to determine number of samples that can be pooled.

Bioinformatics

RAD-seq works because you should be sampling the same sites across many individuals. For bioinformatics you need to idendify reads which belong to the same 'tag', determine homologous 'tags' across individuals, pileup tags and call variants. If the genome is known you can additionally align to the genome and add in linkage information.