

Restriction site-associated DNA sequencing



Centurion
UNIVERSITY

Centurion University of Technology and Management

Contents:

- What is Restriction site-associated DNA sequencing?
- The process of RAD- sequencing
- Conclusion

What is Restriction site-associated DNA sequencing?

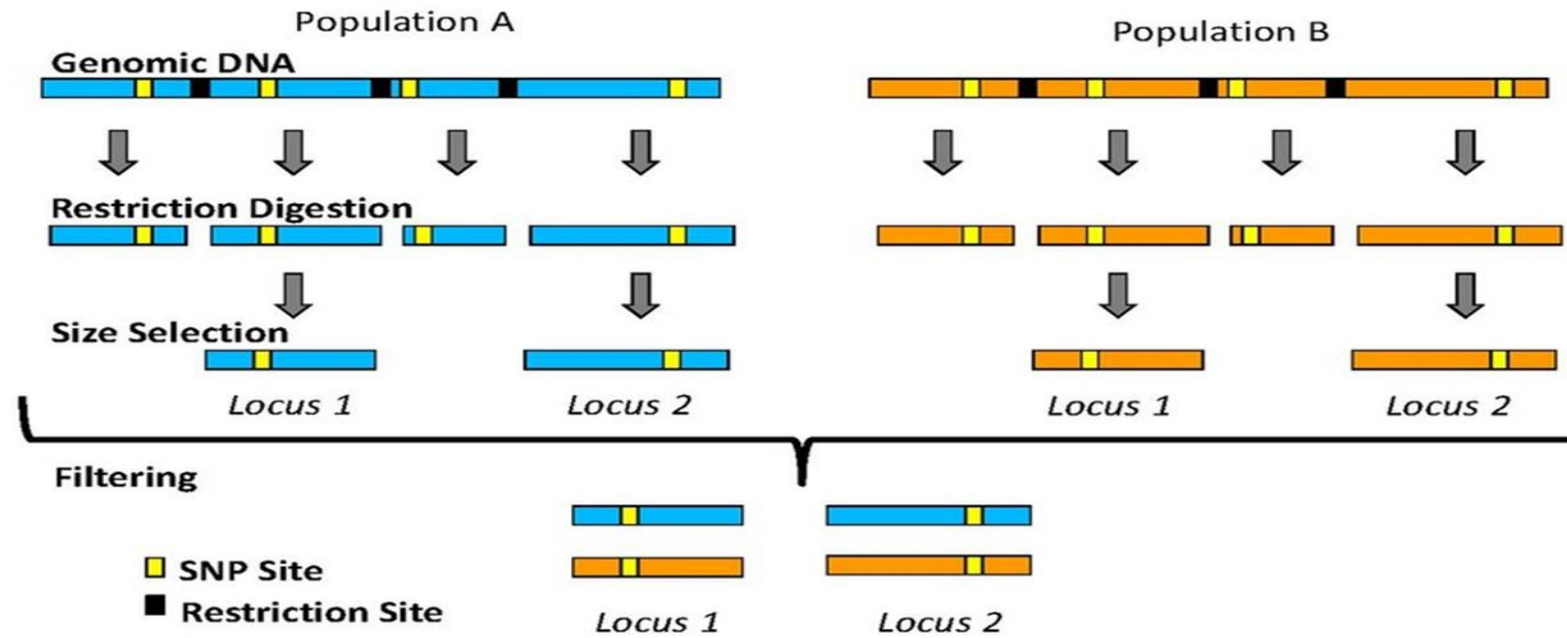
- Restriction site-associated DNA sequencing (RAD-seq or similar genotyping-by-sequencing approaches) is a method to sequence DNA next to restriction sites.
- Normally it is useful for association mapping, QTL-mapping, population genetics, ecological genetics and evolution.
- The rad sequencing essentially targets a subsets of genome for greater depth of coverage per locus yet have the ability to sequence more samples at once.

- The process is divided into different parts: -
 - a. Digestion of DNA
 - b. Ligation of adapters
 - c. Size selection
 - d. Amplification
 - e. Sequencing

The process of RAD- sequencing

- DNA from an individual is cut with the chosen restriction enzyme, producing a set of sticky-ended fragments.
- To be sequenced on an Illumina machine, these fragments are ligated to adapters that will bind to an Illumina flow cell.
- The tagged restriction fragments from a number of individuals are pooled, and then sheared randomly to generate fragments with a mean length of a few hundred base pairs.
- These sheared, sequencer-ready fragments are then size selected (approximately 200–500 base fragments are isolated) and this RADSeq library is then sequenced on the Illumina platform.

Restriction-site Associate DNA Sequencing (RADseq)



Illumina sequencing



Data analysis

Galaxy / Europe

Analyse de donnéesWorkflowVisualizeDonnées partagéesAideUtilisateur

Using 16%

MultiQCv1.6

SRR034310 Demultiplexed reads quality

General Stats

FastQC

Sequence Counts

Sequence Quality Histograms

Per Sequence Quality Scores

Per Base Sequence Content

Per Sequence GC Content

Per Base N Content

Sequence Length Distribution

Sequence Duplication Levels

Overrepresented sequences

Adapter Content

MultiQC

SRR034310 Demultiplexed reads quality

A modular tool to aggregate results from bioinformatics analyses across many samples into a single report.

Report generated on 2018-10-10, 15:36 based on data in: /data/dnb02/galaxy_db/job_working_directory/004/406/4406611/working/multiqc_wDir

General Statistics

Copy tableConfigure ColumnsPlot

Showing 16 rows and 3 columns.

Sample Name	% Dups	% GC	M Seqs
sample_CAAC	73.3%	53%	0.6
sample_CACA	65.0%	52%	0.2
sample_CAGT	75.3%	53%	0.6
sample_CATG	74.8%	52%	0.4
sample_CCAA	60.3%	52%	0.1
sample_CCCC	59.9%	52%	0.1
sample_CCGG	64.9%	52%	0.2
sample_CCTT	62.6%	52%	0.1
sample_CTAG	78.1%	53%	0.8
sample_CTCT	71.8%	53%	0.3
sample_CTGA	68.1%	53%	0.3
sample_CTTC	69.5%	53%	0.3
sample_GGAA	73.3%	53%	0.6
sample_GGCC	74.1%	53%	0.5
sample_GGGG	80.1%	52%	1.2

History

Rechercher des données

STACKS RAD: population genomics with reference genome

28 shown, 3 deleted, 51 hidden

3.15 GB

a list with 16 items

94: FastQC on collection 8: Webpage

a list with 16 items

76: Convert on data 75

lowqualityscore10score20

75: Concatenate datasets on data 70, data 69, and others

lowqualityscore10score20

74: Replace Text on data 73

score20

73: Select on data 13

score20

70: Replace Text on data 67

score20

69: Replace Text on data 66

score10

68: Replace Text on data 65

lowquality

67: Select on data 13

score20

66: Select on data 10

score10

65: Select on data 7

lowquality

13: results.log with Stacks: process radtags on data 2 and data 1: demultiplexed and cleaned reads

score20

518 lines

format: txt, génome de référence: stickleback_11chr

Processing single-end data.

Analysis of Sequenced data

Conclusion

- The Illumina platform currently permits sequencing out to 150 bases, and thus approximately 300 bases on both sides of each restriction site can be screened for polymorphisms.
- Unlike other arrays the knowledge of sequences priorly and number of SNPs markers in the given populations are not needed.
- Deep sequencing allows accurate SNPs discovery and genotyping.