

Introduction:

Restriction site-Associated DNA sequencing (RADseq) is a modern genomic sequencing technique that has greatly advanced the field of population genomics. As a reduced representation-sequencing method, RADseq has decreased per-individual sequencing costs, making genome-wide studies more widely accessible to the research community (Catchen, et al. 2017). Additionally, it allows for the cost-effective sequencing and genotyping of multiple individuals, which is particularly important when discerning genetic variation in natural populations (Narum, et al. 2013). Moreover, RADseq's compatibility with multiple genomic approaches, including phylogenomics, de novo population genomics, and genomic scans, have made it a key player in modern population genomics (Andrews, et al. 2016).

Although RADseq studies require relatively little previous information on the genome of interest, the selection of restriction enzymes and its subsequent effect on marker density remains one of RADseq's limiting factors (Lowry, et al. 2017). The selection of a poor enzyme can then lead to genome undersampling, resulting in low marker density, or oversampling, affecting the resulting downstream coverage. Based on this, we developed an R package pipeline for the exploratory analysis of RADseq data to guide researchers in planning their study and improving data quality. `RADseq_tools` estimates important information for the initial experimental design, including number of cutsites, and marker density and distribution for different user-defined restriction enzyme sites and reference genomes. Even when the reference of interest is not available, the user can explore with available genomes in related taxa and use this information to extrapolate the marker properties for their experiments. In addition, this package can assist with determining the cost of a RADseq experiment by providing estimates of expected coverage, number of samples per lane, and required read throughput.

Getting started:

Verifying R version

To run *RADseq_Tools*, make sure you have a recent R installation. The package was written using R **version 3.4.2**, but should work with versions above 3.1.0. To verify the installed R version, use:

```
> R.version.string  
[1] "R version 3.4.2 (2017-09-28)"
```

Installation

Install the *RADseq_Tools* package from CRAN using:

```
> install.packages("RADseq_Tools")
```

Load the package for use with:

```
> library(RADseq_Tools)
```

Alternatively, source the functions using:

```
> source(' ../R/Functions.R')
```

Tutorial:

This brief guide will provide examples of all the package's basic functions and arguments. It uses a test dataset available with the package. It consist of a small FASTA file containing the first 100Kb of sequence of six Three Spine Stickleback (*Gasterosteus aculeatus*) chromosomes.

Load and process reference sequence

To begin, we need to load the reference sequence into a `RADseq_Tools` object. This new `sequence` object can then be used to search restriction site. Create a variable containing the path to the reference sequence.

```
> fasta_dir <- ' ../inst/extdata/test_genome.fa.gz '
```

This path variable is then the input of the `process_fasta` function, which reads the file provided in the path, converting multi-line sequence objects into single line sequences. The output of the function is a vector containing the sequences strings for each of the chromosomes/scaffolds in the FASTA file. Here, we are creating the object `mySeqs` using the output of `process_fasta`, and displaying the first few nucleotides for each of the chromosomes.

```
> mySeqs <- process_fasta(fasta_dir)
> substr(mySeqs, 1, 65)
```

```
[1] "CTCTTTGTTTTCAGGTGTGGAATGTGCTTTCTACACGGCTACAAATACTACAAAGGATGTAGTA"
[2] "TGACTATTAAGGCGTTTGCAGGCTGAGAGAAGCCAGTCTTGAATGCTACCCCCTTTGAGGAGAG"
[3] "GCTAGCCTGTTTAAACAAACCATCGGTGTGTATGATTACTTGCGCCACACCCGGTCTCTGACG"
[4] "GTATTATATAGTAAATACTATACATTTTCTCTACAGATAGTACAGTGAGTTTACTCTACAACCTC"
[5] "GCTAATATTTTATGGCTGAGTGGGAGGATTCAGTCCTGAAAATGTGTCCTAAAAAACCACTGTA"
[6] "AACCCATCGCCTTATAGGCTGTACTTTATCCTCCAATGATGGAGCAGTTCTCCGCTGCGCGTCGG"
```

Other applications:

References:

Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews. Genetics*, 17(2), 81–92. doi.org/10.1038/nrg.2015.28

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R. and Allendorf, F. W. (2017), Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Mol Ecol Resour*, 17: 362–365. doi:10.1111/1755-0998.12669

Lowry DB, Hoban S, Kelley JL, et al. (2017). Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Mol Ecol Resour*. 17:366–369. doi.org/10.1111/1755-0998.12677

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R. and Hohenlohe, P. A. (2013), Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22: 2841–2847. doi: 10.1111/mec.12350

Hello, this is a test.

```
> for(i in 1:10){  
+   print(i)  
+ }
```

```
[1] 1  
[1] 2  
[1] 3  
[1] 4  
[1] 5  
[1] 6  
[1] 7  
[1] 8  
[1] 9  
[1] 10
```

This is a plot

```
> plot(1:10,1:10)
```

