Using PACVr

Michael Gruenstaeudl, Nils Jenke

February 8, 2020

Contents

1   Introduction

PACVr visualizes the coverage depth of a complete plastid genome as well as the equality of its inverted repeat regions in relation to the circular, quadripartite genome structure and the location of individual genes. This vignette provides instructions for generating the necessary input files and for executing the software from within the R interpreter via function `PACVr.complete()`, and invocation from the Unix command-line shell via script `PACVr_Rscript.R`. This vignette also illustrates the operation of PACVr on an empirical dataset co-supplied with the R package.

2   Creating BAM file

For PACVr to compute and visualize coverage depth along a given plastid genome, users must provide textual information on the mapping of sequence reads on the input genome. This information is supplied via a BAM file. The following is a minimal example code for generating a BAM file using the read mapper BWA (Li and Durbin 2009).

```
bwa -index -a is <reference.fasta>
bwa aln <reference.fasta> <forward_reads.fastq> > <forwards_reads.sai>
bwa aln <reference.fasta> <backward_reads.fastq> > <backward_reads.sai>
bwa mem -t 6 <reference.fasta> <forwards_reads.fastq> <backward_reads.fastq> \
    | samtools view -Sb - > <Sample_BackMapping_bwa.bam>
samtools sort <Sample_BackMapping_bwa.bam> -o <Sample_BackMapping_bwa.sorted.bam>
samtools index <Sample_BackMapping_bwa.sorted.bam>
```

3   Installation

Prior to running PACVr, several dependencies need to be installed. Below are some strategies that have been successful in installing PACVr and all its dependencies on Linux (Arch Linux 4.18, Debian 9.9, and Ubuntu 18.10) and MacOS (HighSierra 10.13.6 and Mojave 10.14.6), respectively.

3.1   On Linux

To install mosdepth, open a shell and type:

```
system("conda install -y mosdepth")
```

To install all R dependencies, open R and type:

```
# CRAN packages
if (!require("pacman"))
    install.packages("pacman")
pacman::p_load("RCircos", "optparse", install=TRUE)
# Bioconductor packages
if (!requireNamespace("BiocManager", quietly=TRUE))
    install.packages("BiocManager")
if (!requireNamespace("genbankr", quietly=TRUE))
    BiocManager::install("genbankr")
```

3.2   On MacOS

To install mosdepth, do:

- Install brew
- Install nim via brew: `brew install nim`
- Install hts-lib via brew: `brew install htslib`
- Obtain the source code of mosdepth from its Github page
- In the unzipped mosdepth source code directory, open a terminal and type: `nimble build` (confirm when prompted) and `nim c -d:release mosdepth.nim`
- When in the unzipped mosdepth source code directory, confirm the compilation of mosdepth by typing `./mosdepth -h`

To install all R dependencies, open R and follow the instructions as given for Linux (see above).

4   Executing PACVr via R interpreter

PACVr can be executed from within the R interpreter via function `PACVr.complete()`.

```
library(PACVr)
## Specify input files
gbkFile <- system.file("extdata", "MH161174/MH161174.gb", package="PACVr")
bamFile <- system.file("extdata", "MH161174/MH161174_PlastomeReadsOnly.sorted.bam",
                       package="PACVr")
## Specify output file
```

```
outFile <- paste(tempdir(), "/MH161174_AssemblyCoverage_viz.pdf", sep="")
## Run PACVr
PACVr.complete(gbk.file=gbkFile, bam.file=bamFile, windowSize=250,
               mosdepthCmd='mosdepth', logScale=FALSE, threshold=0.5,
               syntenyLineType=3, relative=TRUE, textSize=0.5,
               delete=TRUE, output=outFile)
```

5 Executing PACVr via Unix shell

PACVr can also be executed from the Unix command-line shell via script `PACVr_Rscript.R`.

```
Rscript ./inst/extdata/PACVr_Rscript.R \
    -k ./inst/extdata/MH161174.gb \
    -b ./inst/extdata/MH161174_PlastomeReadsOnly.sorted.bam \
    -o ./inst/extdata/MH899017_AssemblyCoverage_viz.pdf
```
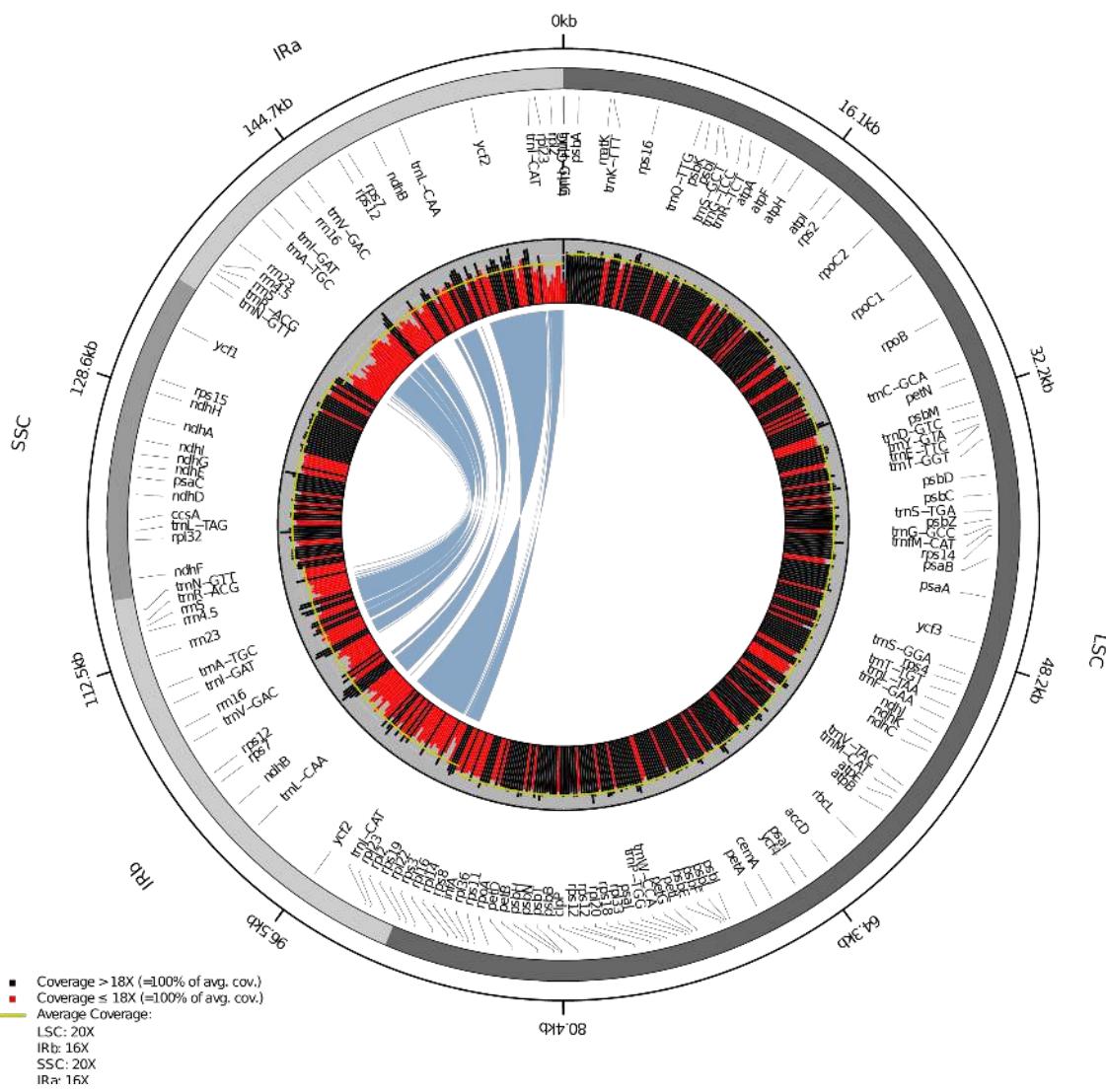
Figure 1: File `MH161174_AssemblyCoverage_viz.pdf` as generated via `PACVr.complete()`