

README

Henrik Christiansen

January 22, 2021

Project: radseq pilot

RADseq optimization pilot experiment with several Antarctic species.

Contact: henrik.christiansen@kuleuven.be

Log:

22/01/2020: fixed xml encoding issues

20/01/2020: added new xml file for run 7, updated bioanalyzer to v0.6.2 and updated 02_empirical_digests.R

20/01/2020: added the project to a github repository

16/12/2020: added a modified read.bioanalyzer function to read in xml file with a missing lower marker in one sample

14/12/2020: small, mostly cosmetic updates in most scripts; updated project and input/output description; finalized 02_empirical_digests.R

11/12/2020: added customized plot saving function in recto_REs_and_functions.R

10/12/2020: updated 02_empirical_digests.R, added a customized ggplot plotting function in recto_REs_and_functions.R added in silico functions for comparison in recto_REs_and_functions.R

11-12/2020: updated 02_empirical_digests.R & added a customized plotting function in recto_REs_and_functions.R

11/11/2020: updated to run with R v4.0.3

07-10/2020: added script 02_empirical_digests.R to read results from bioanalyzer and analyze them in R, instead of cumbersome “manual” analysis in spreadsheets

20/07/2019: updated plot ratio scripts

10/07/2019: updated digest scripts

09/07/2019: updated the plots for test library 2

02/07/2019: updated digest scripts

28/06/2019: finalized coverage plot

27/06/2019: re-orderd scripts, included scripts for in silico digestion (previously stored in different R project)

23-27/06/2019: updated parameter plots (now based on correct de novo runs, with PE concatenating); included & updated coverage plot

16-22/06/2019: updated parameter plots; started script to plot targeted and realized coverage

01-14/06/2019: included scripts to plot `n_loci` and `n_snps_per_locus` from de novo parameter optimization series

04/12/2018: rerun due to updated input data (Trematomus mainly)

29/11/2018: created script `01_plotting.R` and ratio plots

28/11/2018: created R project & input data

Input:

Various reference genomes from target species, or related species in fasta format under:

- `/data/refgenomes`

Output files from the Agilent Bioanalyzer Software in XML format under:

- `data/bioanalyzer_results`

CSV file with all metadata related to the Bioanalyzer runs, created manually:

- `data/bioanalyzer_results/run_overview.csv`

CSV/TSV files in `data/test_libraries` with data from 5 test libraries:

- `coverage_stats.csv`: coverage for each library under different M values
- various `n_snps_per_locus.tsv`: files for each library/species and different r/M values with information about the loci and SNPs

R script that contains restriction enzyme information and various functions for analysis and plotting as used in the analyses scripts:

- `analyses/recto_REs_and_functions.R`

Additional R script to export plots consistently:

- `~/printfig.R`

Analyses:

To be run sequentially, all listed under `/analyses`.

- `01_digests.R`: script to perform in silico digestions for different target organisms
- `02_empirical_digests.R`: script to read bioanalyzer results and plot output in comparison with in silico results
- `03_plot_n_loci.R`: script based on Rochette & Catchen 2017 to plot the number of (polymorphic) loci for different M parameters
- `04_plot_n_snps_per_locus.R`: script based on Rochette & Catchen 2017 to plot the number of SNPs per locus for different M parameters
- `05_plot_coverage.R`: script to plot target and realized coverage

Output:

In /data/in_silico_results:

- various tables.csv containing in silico digest results per target taxa

In /figures:

- genome digestion curves, empirical and in silico, created with 02_empirical_digests.R
- various n_loci_Mn plots per library, created with 03_plot_n_loci.R
- various n_snps per library/species, created with 04_n_snps_per_locus.R
- coverage plots, created by 05_plot_coverage.R

Session info:

Package citations:

```
## Warning in FUN(X[[i]], ...): no date field in DESCRIPTION file of package
## 'bioanalyzer'

## [[1]]
##
## To cite package 'here' in publications use:
##
## Kirill Müller (2020). here: A Simpler Way to Find Your Files. R
## package version 1.0.0. https://CRAN.R-project.org/package=here
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {here: A Simpler Way to Find Your Files},
##   author = {Kirill Müller},
##   year = {2020},
##   note = {R package version 1.0.0},
##   url = {https://CRAN.R-project.org/package=here},
## }
##
##
## [[2]]
##
## To cite package 'SimRAD' in publications use:
##
## Olivier Lepais and Jason Weir (2016). SimRAD: Simulations to Predict
## the Number of RAD and GBS Loci. R package version 0.96.
## https://CRAN.R-project.org/package=SimRAD
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {SimRAD: Simulations to Predict the Number of RAD and GBS Loci},
```

```

##   author = {Olivier Lepais and Jason Weir},
##   year = {2016},
##   note = {R package version 0.96},
##   url = {https://CRAN.R-project.org/package=SimRAD},
## }
##
##
## [[3]]
##
## To cite seqinr in publications use:
##
##   Charif, D. and Lobry, J.R. (2007)
##
## A BibTeX entry for LaTeX users is
##
##   @InCollection{,
##     author = {D. Charif and J.R. Lobry},
##     title = {Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devot
##     booktitle = {Structural approaches to sequence evolution: Molecules, networks, populations},
##     year = {2007},
##     editor = {U. Bastolla and M. Porto and H.E. Roman and M. Vendruscolo},
##     series = {Biological and Medical Physics, Biomedical Engineering},
##     pages = {207-232},
##     address = {New York},
##     publisher = {Springer Verlag},
##     note = {{ISBN :} 978-3-540-35305-8},
##   }
##
##
## [[4]]
##
## To cite package 'bioanalyzeR' in publications use:
##
##   Joseph Foley (2021). bioanalyzeR: Analysis of Agilent electrophoresis
##   data. R package version 0.6.2.
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {bioanalyzeR: Analysis of Agilent electrophoresis data},
##     author = {Joseph Foley},
##     year = {2021},
##     note = {R package version 0.6.2},
##   }
##
##
## [[5]]
##
##   Wickham et al., (2019). Welcome to the tidyverse. Journal of Open
##   Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,

```

```

##   title = {Welcome to the {tidyverse}},
##   author = {Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostini
##   year = {2019},
##   journal = {Journal of Open Source Software},
##   volume = {4},
##   number = {43},
##   pages = {1686},
##   doi = {10.21105/joss.01686},
## }
##
##
## [[6]]
##
## To cite package 'ggsci' in publications use:
##
##   Nan Xiao (2018). ggsci: Scientific Journal and Sci-Fi Themed Color
##   Palettes for 'ggplot2'. R package version 2.9.
##   https://CRAN.R-project.org/package=ggsci
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {ggsci: Scientific Journal and Sci-Fi Themed Color Palettes for
## 'ggplot2'},
##     author = {Nan Xiao},
##     year = {2018},
##     note = {R package version 2.9},
##     url = {https://CRAN.R-project.org/package=ggsci},
##   }
##
##
## [[7]]
##
## To cite package 'gridExtra' in publications use:
##
##   Baptiste Auguie (2017). gridExtra: Miscellaneous Functions for "Grid"
##   Graphics. R package version 2.3.
##   https://CRAN.R-project.org/package=gridExtra
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {gridExtra: Miscellaneous Functions for "Grid" Graphics},
##     author = {Baptiste Auguie},
##     year = {2017},
##     note = {R package version 2.3},
##     url = {https://CRAN.R-project.org/package=gridExtra},
##   }

```

R Session:

```

## R version 4.0.3 (2020-10-10)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 18.04.5 LTS

```

```

##
## Matrix products: default
## BLAS: /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.7.1
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.7.1
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8 LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8 LC_NAME=C
## [9] LC_ADDRESS=C LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats4 parallel stats graphics grDevices utils datasets
## [8] methods base
##
## other attached packages:
## [1] gridExtra_2.3 ggsci_2.9
## [3] forcats_0.5.0 stringr_1.4.0
## [5] dplyr_1.0.2 purrr_0.3.4
## [7] readr_1.4.0 tidyr_1.1.2
## [9] tibble_3.0.4 ggplot2_3.3.2
## [11] tidyverse_1.3.0 bioanalyzeR_0.6.2
## [13] seqinr_4.2-4 SimRAD_0.96
## [15] zlibbioc_1.36.0 ShortRead_1.48.0
## [17] GenomicAlignments_1.26.0 SummarizedExperiment_1.20.0
## [19] Biobase_2.50.0 MatrixGenerics_1.2.0
## [21] matrixStats_0.57.0 Rsamtools_2.6.0
## [23] GenomicRanges_1.42.0 GenomeInfoDb_1.26.1
## [25] BiocParallel_1.24.1 Biostrings_2.58.0
## [27] XVector_0.30.0 IRanges_2.24.0
## [29] S4Vectors_0.28.0 BiocGenerics_0.36.0
## [31] here_1.0.0
##
## loaded via a namespace (and not attached):
## [1] bitops_1.0-6 fs_1.5.0 lubridate_1.7.9.2
## [4] RColorBrewer_1.1-2 progress_1.2.2 httr_1.4.2
## [7] rprojroot_2.0.2 tools_4.0.3 backports_1.2.0
## [10] R6_2.5.0 DBI_1.1.0 colorspace_2.0-0
## [13] ade4_1.7-16 withr_2.3.0 tidyselect_1.1.0
## [16] prettyunits_1.1.1 compiler_4.0.3 cli_2.2.0
## [19] rvest_0.3.6 xml2_1.3.2 DelayedArray_0.16.0
## [22] scales_1.1.1 digest_0.6.27 rmarkdown_2.5
## [25] base64enc_0.1-3 jpeg_0.1-8.1 pkgconfig_2.0.3
## [28] htmltools_0.5.0 dbplyr_2.0.0 rlang_0.4.8
## [31] readxl_1.3.1 rstudioapi_0.13 generics_0.1.0
## [34] hwriter_1.3.2 jsonlite_1.7.1 RCurl_1.98-1.2
## [37] magrittr_2.0.1 GenomeInfoDbData_1.2.4 Matrix_1.3-2
## [40] fansi_0.4.1 Rcpp_1.0.5 munsell_0.5.0
## [43] lifecycle_0.2.0 stringi_1.5.3 yaml_2.2.1
## [46] MASS_7.3-53 grid_4.0.3 crayon_1.3.4
## [49] lattice_0.20-41 haven_2.3.1 hms_0.5.3
## [52] knitr_1.30 pillar_1.4.7 reprex_0.3.0

```

## [55]	XML_3.99-0.5	glue_1.4.2	evaluate_0.14
## [58]	latticeExtra_0.6-29	modelr_0.1.8	png_0.1-7
## [61]	vctrs_0.3.5	cellranger_1.1.0	gtable_0.3.0
## [64]	assertthat_0.2.1	xfun_0.19	broom_0.7.2
## [67]	ellipsis_0.3.1		