sangeranalyseR: Simple and Interactive Processing of Sanger Sequencing Data in R

Kuan-Hao Chao*, Kirston Barton, Sarah Palmer, Robert Lanfear*

Department of Ecology and Evolution, Australian National University;
The Garvan Institute of Medical Research; The University of Sydney School of Medicine

Australian National University

https://github.com/roblanf/sangeranalyseR

https://sangeranalyser.readthedocs.io/

https://doi.org/doi:10.18129/B9.bioc.sangeranalyseR

https://doi.org/10.1093/gbe/evab028

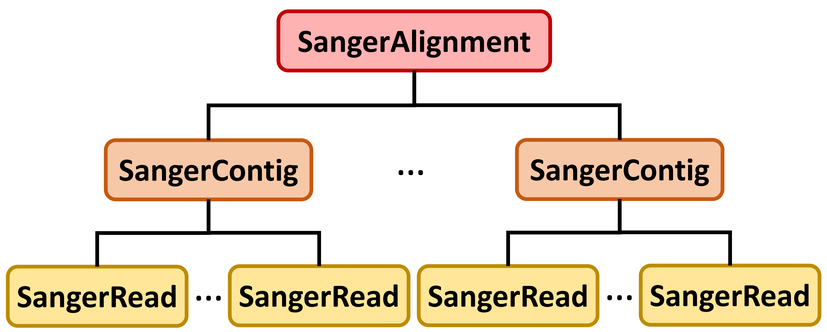
Introduction

sangeranalyseR is an R package that provides fast, flexible, and reproducible workflows for processing your Sanger sequencing data. It allows users to go from loading reads to saving aligned contigs in ***5 lines of R code*** by using sensible defaults. Most importantly, it's free, it's open source, and it's in R.

Main Features

- **Pure R environement**
- **Supporting AB1 and FASTA input**
- **Automated data analysis**
- **Regular Expression or CSV matching**
- **Interactive Shiny apps**
- **Thorough report output**

Design Concept: 3-level S4 class



Analysis Workflow

Step 1: Prepare input files

AB1

[*Consensus Read Name*] + _ + [*index*] + _ + [F,R] + .ab1

FASTA

[*Consensus Read Name*] + _ + [*index*] + _ + [F,R]

```
# Loading sample files in sangeranalyseR to the R environment
rawDataDir <- system.file("extdata", package = "sangeranalyseR")
parentDir <- file.path(rawDataDir, 'Allolobophora_chlorotica')
```

Step 2: Load and analyze data

Input: AB1 / FASTA **Output**: SangerAlignment S4 object

Trimming method:

- M1 => modified Mott's trimming (Phred/Phrap and Biopython)
- M2 => sliding window method (trimmomatic)

AB1 + REGEX

```
align <- SangerAlignment(inputSource = "ABIF",
                         processMethod = "REGEX",
                         ABIF_Directory = parentDir,
                         REGEX_SuffixForward = "_[0-9]*_F.ab1$",
                         REGEX_SuffixReverse = "_[0-9]*_R.ab1$")
```

FASTA + REGEX

```
align <- SangerAlignment(inputSource = "FASTA",
                         processMethod = "REGEX",
                         FASTA_File = fastaFN,
                         REGEX_SuffixForward = "_[0-9]*_F$",
                         REGEX_SuffixReverse = "_[0-9]*_R$")
```

CSV file example

```
"reads","direction","contig"
"Achl_ACHLO006-09_1_F.ab1","F","Achl_ACHLO006-09"
"Achl_ACHLO006-09_2_R.ab1","R","Achl_ACHLO006-09"
```

AB1 + CSV

```
csv_names <- file.path(rawDataDir, "ab1", "SangerAlignment",
                       "names_conversion_all.csv")
align <- SangerAlignment(inputSource = "ABIF",
                         processMethod = "CSV",
                         ABIF_Directory = parentDir,
                         CSV_NamesConversion = csv_names)
```

FASTA + CSV

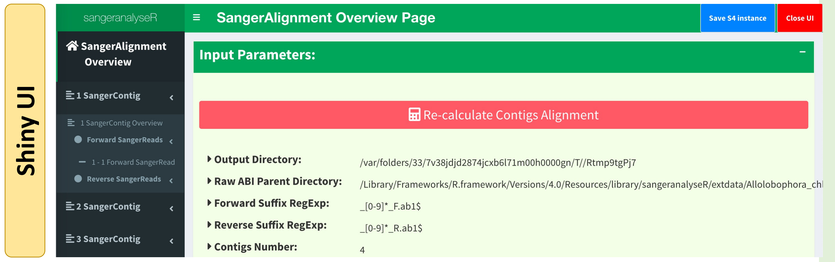
```
csv_names <- file.path(rawDataDir, "fasta", "SangerAlignment",
                       "names_conversion.csv")
align <- SangerAlignment(inputSource = "FASTA",
                         processMethod = "CSV",
                         FASTA_File = fastaFN,
                         CSV_NamesConversion = csv_names)
```

Step 3:Explore data

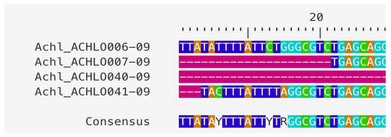
Input: SangerAlignment S4 object **Output**: Shiny UI object

```
# Launching a Shiny app to check SangerAlignment S4 object
launchApp(align, outputDir = tempdir())
```

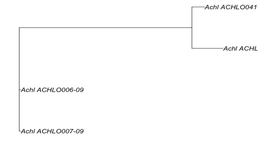
Shiny UI



alignment

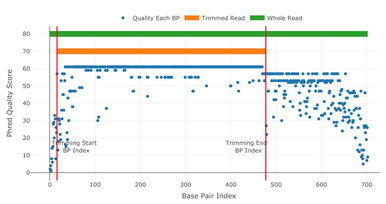


tree

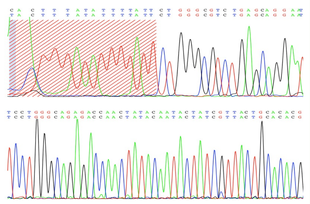


read info

trimming plot



chromatogram



Step 4: Output aligned contigs

Input: SangerAlignment S4 object **Output**: FASTA file

```
writeFasta(align, outputDir = tempdir(),
           compress = FALSE, compression_level = NA)
```

Step 5: Create interactive report

Input: SangerAlignment S4 object **Output**: HTML files

```
generateReport(align, outputDir = tempdir(),
               includeSangerRead = FALSE,
               includeSangerContig = FALSE)
```

Contact

Kuan-Hao Chao: kuanhao.chao@gmail.com; https://www.kuanhaochao.com
Robert Lanfear : rob.lanfear@anu.edu.au; https://www.robertlanfear.com

Kuan-Hao Chao, Kirston Barton, Sarah Palmer, and Robert Lanfear (2021). "*sangeranalyseR: simple and interactive processing of Sanger sequencing data in R*" in Genome Biology and Evolution. DOI: doi.org/10.1093/gbe/evab028