seq2lus

A R-package to convert UAS sequences to

Longest Uninterupted Stretch (LUS)
and
bracket format using LUSstrR (lusSTR)

By Øyvind Bleka

Update

- The new version utilizes the LUSstrR R-package: https://github.com/oyvble/LUSstrR

- LUSstrR is a reimplementation of **lusSTR** (Python)

- The lookup table is not used if LUSstrR is installed

- The bracket/block format (forward direction) is a new option
  FWbrack: Forward_Strand_Bracketed_form)

Installation (from github)

- Requires R (at least v3.0.0)
- Install required R-packages in R:

```
install.packages(c("gWidgets2tcltk","readxl","devtools"))
```

- Install directly from github:

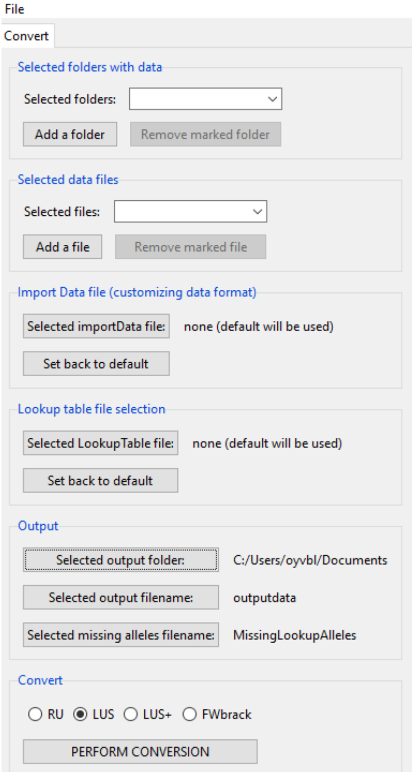
```
library(devtools)
install_github("oyvble/seq2lus")

install_github("oyvble/LUSstrR")
```

Opening the GUI

- After installation, run following in commando:

  library(seq2lus);cgui()



Selected importData file: none (default will be used)

Selected LookupTable file: none (default will be used)

Selected output folder: C:/Users/oyvbl/Documents

Selected output filename: outputdata

Selected missing alleles filename: MissingLookupAlleles

○ RU ● LUS ○ LUS+ ○ FWbrack

Save settings etc (projectfile)

Select folder(s) with datafiles

Select specific datafile(s)

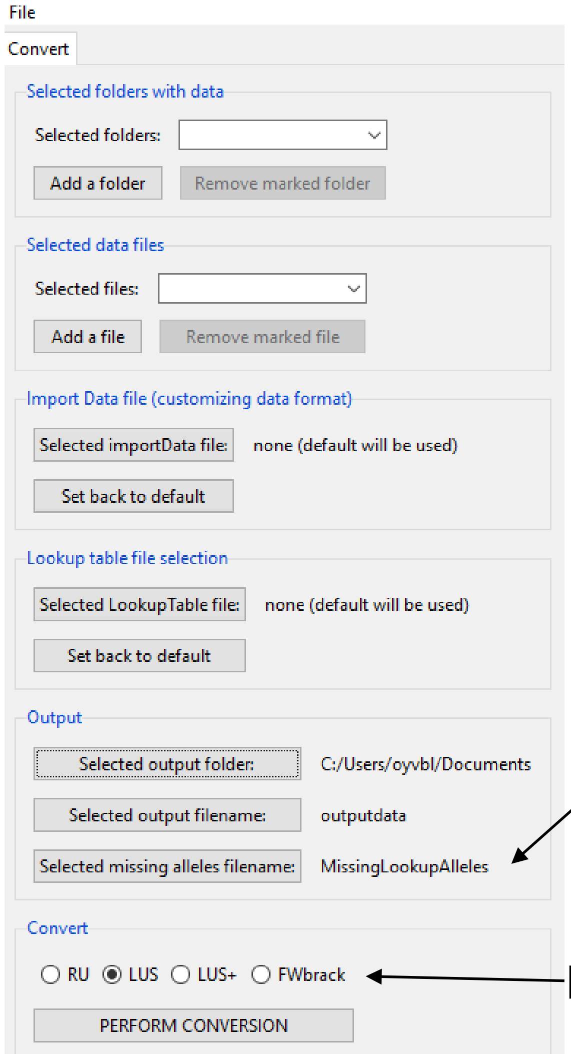
Allows data to be in another format than by default (data customization): [SampleName Marker Allele1,….]

The excel lookup-file used for conversion (can optionally be selected)

Select where data is saved after conversion (all profiles are stored in one file)

Filename for storing non-converted alleles

Select allele type to convert to



Selected importData file: none (default will be used)

Selected LookupTable file: none (default will be used)

Selected output folder: C:/Users/oyvbl/Documents

Selected output filename: outputdata

Selected missing alleles filename: MissingLookupAlleles

○ RU ● LUS ○ LUS+ ○ FWbrack

Not used if LUSstrR is installed

*All sequences **will** be converted*

New option

The converter

- Based on using the “convert” R-function
- A specific allele type is selected (RU, LUS, LUS+, others)

1) Takes sequences from the imported data
2) New version based on *LUSstrR::convert*
3) Otherwise a selected LookupTable file is used (as before)
4) Converts all sequences to specified format
   - Adding up Reads if several sequences give same allele type
5) Store “converted data” in a single file
6) Sequences not found in LookupTable are stored in “missingfile”

Important notes

- Only either evidence profiles or reference profiles can be converted at the same time:
  - This is because of different data formats
    - Evids can have many alleles (with coverage/height columns)
    - Refs only have two allele columns

- Settings will be remembered once selected
- Settings can be stored as a project file

Default data import

- The evid-data returned by the function importData must be in one of the two specific formats:
  - “Wide-format”:

    [SampleName Marker Allele1 Allele2…Height1 Height2…]

  - “Long-format”:

    [SampleName Marker Allele Height]

“Covarage” instead of “Height” is accepted as name.

Customized data import

It’s possible to create your own importData function and store it an R-script which is selected at “Select importData file”

- The function MUST return one of the formats from the previous slide.
- Example of importFiles to read from an UAS excel report is found in the installation folder:

~\R-installation\seq2lus\importData_UAS...R

References

1. R.S. Just, J.A. Irwin, Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results, Forensic. Sci. Int.: Genet. 34 (2018) 197-205.

2. R.S. Just, J. Le, J.A. Irwin, LUS+: Extension of the LUS designator concept to differentiate most sequence alleles for 27 STR loci, Forensic Sci Int.: Reports 2 (2020) 100059

3. K.B. Gettings, L.A. Borsuk, D. Ballard, M. Bodner, B. Budowle, L. Devesse, et al., STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci, Forensic. Sci. Int.: Genet. 31 (2017) 111-117.