Electrical Penetration Graph data conversion

Description

`epgrun` converts Electrical Penetration Graph (EPG) annotation data from 'EPG Stylet +a' to EPG variables.

Usage

```
epgrun(loc, epgstart, epgstop, epgtable, epgplot, epgbin, epgname)
```

Arguments

`loc`	Vector with directory name(s) where the .ANA output files are stored from 'EPG Stylet +a' (www.epgsystems.eu). ANA files belonging to different groups (e.g. plant lines or treatments) should be saved in separate directories. The folder name will be used as group name.
`epgstart`	Number indicating the start time (in hours) of the EPG recordings. Should be > 0 if the first part of the recording should be excluded. By default 0.
`epgstop`	Number indicating the end time (in hours) of the EPG recordings. Should be larger than epgstart. This is the total duration of the actual recording, or a shorter duration if the last part of the recording should be excluded. Maximum 16.
`epgtable`	Logical scalar indicating whether a table with mean values and standard error should be made and univariate statistical tests should be performed. TRUE or FALSE. By default TRUE.
`epgplot`	Logical scalar indicating whether plots should be made. TRUE or FALSE. By default TRUE.
`epgbin`	Logical scalar indicating whether timebins should be calculated. TRUE or FALSE. By default TRUE.
`epgname`	String with project name which will appear in output file names. By default a project name will be created with the first and the last group name and the total number of groups.

Value

Maximum 7 output files will be saved to the working directory:

- rawdata.csv (raw data with user-defined end time of recording)
- probes.csv (extended information per probe, including previous and next waveform)
- variables.csv (EPG summary variables per insect, 59 variables. See detailed description below)
- epgtable.csv (if epgtable = TRUE: table with mean plus minus standard error, and outcomes of univariate statistical tests on summary variables. Chi square tests are performed on occurrence data (e.g. occ.e1). Other variables are analyzed with Wilcoxon signed rank tests "wilcox.test()" (2 groups) or Kruskal-Wallis tests "kruskal.test()" (> 2 groups). All tests are performed in default settings. Samples or groups with missing values for a variable are not included in a test, tests are not executed in case < 2 groups have data, or - in case of a chi square test- if there are no values > 0. 'test_statistic' = value of test statistic (e.g. W for Wilcoxon,), 'n_' = number of biological replicates per group for the respective variable, Pvalue is NA in case of missing or equal values, standard error is NA in case of only <= 1 biological replicate)
- boxplots.pdf (if epgplot = TRUE: boxplots of total recording)
- timebins.csv (if epgbin = TRUE: probes categorized in timebins of 1 hour. If probes are spanning several hours, they will be chopped up)
- timebinplots.pdf (if epgbin & epgplot = TRUE: lineplots)

This function is meant to facilitate EPG data processing and analysis. Correct usage and interpretation of the data and statistical tests are the responsibility of the user of this package.

Description of waveforms:

- waveform 1 - Non probing
- waveform 2 - Pathway
- waveform 3 - E1e
- waveform 4 - Salivation in the phloem
- waveform 5 - Phloem feeding
- waveform 6 - Penetration difficulties
- waveform 7 - Xylem ingestion
- waveform 8 - Potential drops
- waveform 11 - User defined
- waveform 12 - User defined

N.B. It is not required to have annotated all waveforms described above. N.B.B. Waveforms are considered to be mutually exclusive, except waveform 8. This means, that other waveforms are not considered to end when waveform 8 starts, and that their duration may include waveform 8 events. In general, waveform 8 is used to annotate potential drops during waveform 2 (pathway).

Description of EPG variables (adapted from www.epgsystems.eu):

- sum.np - Total duration non probing (min)
- lat.c - Latency to first C from start of recording (min)
- sum.c - Total duration C (min)
- num.c - Total number of C
- num.cs - Total number of C < 3 min ('short probes')
- num.css - Total number of C < 0.5 min
- mean.c - Mean duration C (min)
- num.pd - Total number of potential drops
- rate.pd - Number of potential drops per min of C (num.pd divided by sum.c)
- occ.e1e - Does insect perform waveform E1e? (1=yes, 0=no)
- lat.e1e - Latency to first waveform E1e from start of first probe (min)
- num.e1e - Number of waveform E1e
- sum.e1e - Total duration waveform E1e (min)
- mean.e1e - Mean duration waveform E1e (min)
- occ.e1 - Does insect perform E1? (1=yes, 0=no)

- lat.e1 - Latency to first E1 from start of recording (min)
- lat.e1p - Latency to first E1 from start of first probe (min)
- num.e1 - Total number of E1
- sum.e1 - Total duration E1 (min)
- mean.e1 - Mean duration E1 (min)
- max.e1 - Maximum duration E1 (min)
- prop.e1 - Proportion of time spent on E1 in phloem phase (E1+E2, excluding single E1s) (
- occ.sing.e1 - Does insect perform single E1? (1=yes, 0=no)
- sing.e1 - Number of single E1s
- num.e1f - Number of E1 before E2 starts (excluding single E1s)
- num.e1m - Number of E1s in between E2
- num.e1l - Number of E1s at end of phloem feeding
- sum.e1f - Total duration E1 before starting E2 (min)
- sum.e1m - Total duration E1 in between E2 (min)
- sum.e1l - Total duration E1 when ending E2 (min)
- prop.e1f - Proportion of time spent on E1 before starting E2 in phloem phase (E1+E2) (
- prop.e1m - Proportion of time spent on E1 in between E2 in phloem phase (E1+E2) (
- prop.e1l - Proportion of time spent on E1 when ending E2 in phloem phase (E1+E2) (
- occ.e2 - Does insect perform E2? (1=yes, 0=no)
- occ.e2s - Does insect perform sustained E2 (>10 min)? (1=yes, 0=no)
- lat.e2 - Latency to first E2 from start of recording (min)
- lat.e2p - Latency to first E2 from start of first probe (min)
- lat.e2s - Latency to first sustained E2 (>10 min) from start of recording (min)
- lat.e2sp - Latency to first sustained E2 (>10 min) from start of first probe (min)
- num.e2 - Number of E2
- num.e2s - Number of sustained E2 (>10 min)
- sum.e2 - Total duration of E2 (min)
- sum.e2s - Total duration of sustained E2 (>10 min) (min)
- mean.e2 - Mean duration of E2 (min)
- mean.e2t - Mean total duration of E2 phase including potential e1m (min)
- mean.e2s - Mean duration of sustained E2 (>10 min) (min)
- mean.e2st - Mean total duration of E2s phase including potential e1m (min)
- max.e2 - Maximum duration E2 (min)
- occ.f - Does insect perform F? (1=yes, 0=no)
- num.f - Number of F
- sum.f - Total duration F (min)
- mean.f - Mean duration F (min)
- occ.g - Does insect perform G? (1=yes, 0=no)
- num.g - Number of G
- sum.g - Total duration G (min)
- mean.g - Mean duration G (min)
- occ.w11 - Does insect perform waveform 11? (1=yes, 0=no)
- lat.w11 - Latency to first waveform 11 from start of first probe (min)
- num.w11 - Number of waveform 11
- sum.w11 - Total duration waveform 11 (min)
- mean.w11 - Mean duration waveform 11 (min)
- occ.w12 - Does insect perform waveform 11? (1=yes, 0=no)
- lat.w12 - Latency to first waveform 12 from start of first probe (min)
- num.w12 - Number of waveform 12
- sum.w12 - Total duration waveform 12 (min)
- mean.w12 - Mean duration waveform 12 (min)
- N.B. If a waveform does not occur it will have a missing value (NA) for 'lat', 'sum' and 'mean'.

Examples

```
# Define working directory
setwd("C:/Myworking/directory/")

# Convert EPG data. In this case 2 plant lines and 8 h recordings
epgrun(loc = c("C:/folder/subfolder/plantline1", "C:/folder/subfolder/plantline2"),
epgstop = 8)

# Output files are saved to the working directory
```