WHAT DOES PERSIST DO?

This programme allows fast calculation of different measures for the directional persistence of cell migration from tracking data of XY positions over time. The package contains 2 programme files and 2 test datasets that have been used in the original publication.

Please see for more information:
Theisen, U., Straube, E. and Straube, A. (2012).
Directional Persistence of Migrating Cells Requires Kif1C-Mediated Stabilization of Trailing Adhesions.
Developmental Cell 23 (6), 1153-1166.

HOW TO PREPARE A DATAFILE?

The programme requires a space-, or comma-delimited txt or ascii file with xy coordinates for each track in rows. No headers are allowed. The coordinates have to be given as

x_0 x_1 x_2 x_3 … y_0 y_1 y_2 y_3 …

The coordinates are expected to be in µm. Image Pro Analyser software can automatically produce this data structure. Remove header text and export as comma-separated file from Excel.

WHAT PARAMETERS NEED TO BE GIVEN?

The parameters are given in the command line as a row of identifiers and parameter values separated by space. The actual name used for the identifier does not matter, but it should not contain a white space, underscores are fine.

input_file give path if not same folder as programme and filename. This is case-sensitive! For the test sets use: ./datafolder/e3Control_527.dat and ./datafolder/e3Kif1C1_513.dat

nt number of tracks, i.e. the number of rows in your datafile. For the test datasets, these are 527 and 513 respectively as indicated in the filenames.

np number of points, i.e. number of columns in your datafile divided by 2. So this is the number of timepoints. It does not matter if some tracks are shorter. This value gives the maximal length of any tracks.

used_points This should be identical to np unless you don't want to consider all timepoints in the dataset. If for example, you want to compare cells early and late in the dataset use this to cut off late timepoints.

nav number of points used to average for the initial reference vector. We suggest to start with 4, which will average over 3 time periods, e.g. $xy_1 – xy_2 – xy_3 - xy_4$

start_position This should be 0 unless you don't want to consider all timepoints in the dataset. If for example, you want to compare cells early and late in the dataset use this to cut off early timepoints.

cstep bin width of accumulated distance to be used for histograms and other output files.

astep bin width of persistence length to be used for histograms and other output files. It would be recommended to use a value similar to cstep.

clmp maximum value of accumulated distance to be used for histogram. The full range of available values will be used for the output file that summarises persistence measures relative to accumulated distance. This value should be a multiple of cstep.

aminp minimum value of persistence length to be used for histograms. Note that this value can be negative.

amaxp maximum value of persistence length to be used for histograms. It does not make sense for this value to be higher than clmp. The difference between aminp and amaxp should be a multiple of astep.

cutoff this should either be 0 to show all data or give the fraction of tracks that need to contribute data to a bin of accumulated distance values to be included in the histogram. Unless clmp is limiting, few long tracks will produce islands of high probability in the plots and under those circumstances a cutoff of 0.1 or 0.05 is recommended. This will remove data when less than 5 or 10% or tracks are contributing. This is mainly important if datasets with different migration speed are plotted on the same scales.

dt time difference between frames in minutes.

HOW TO RUN PERSIST?

1. Open Terminal in Linux or MacOS. Mac Users: You will find this in your Application folder under Utilities.

2. Navigate to the folder that contains the files with the source code (persist_2d_26mmm.c and pardet_26mm.c) using cd commands. e.g.

   cd Documents

   If have trouble with folder names that contain spaces, replace spaces by underscores. If you don't know where in the file structure you are, type

   ls

   which will show you all folders and files in your current location.

3. When using it for the first time, compile the code to an executable programme by typing:

   gcc -o persist_2d_26 persist_2d_26mmm.c –lm

4. Run the programme by typing the following line. This assumes that you saved your datafiles in a datafolder next to the location of the programme. Otherwise change the path to your datafiles accordingly.

   ./persist_2d_26 input_file ./datafolder/e3Control_527.dat nt 527 np 37 used_points 37 nav 4 start_position 0 cstep 10 astep 10 clmp 250 aminp -100 amaxp 200 cutoff 0.1 dt 10

5. The programme will output many results in the terminal, the majority of which are saved in output files in the path of the datafile. In the last few lines, the programme will summarise parameters and errors occurred. Example:

   mean step length : 5.907, mean total length : 212.869

   1 undetermined start directions
   140 short input lines

   Experiment : "./datafolder/e3Kif1C1_513.dat"

   Undetermined start directions occur when the vector of the starting direction cannot be determined because nav consecutive points are identical. The analysis will just skip this starting point and continue. Short input lines are all tracks shorter than np. This is just for information, the tracks will still be included in the analysis. Wrong input lines have different x and y values and will be excluded from the analysis. Scroll up to find out which tracks cause start direction and wrong line input errors.

6. You can use the upwards arrow to recall the previous command line, modify filename and/or parameters and run another dataset. If you want to compare data for the same dataset with other parameters, be aware that results files will be overwritten. Rename or move files to a different folder to keep other versions of results files.

WHAT ARE THE NUMBERS IN THE OUTPUT FILES?

The programme will write 5 files for each dataset and append an identifier to the name of the datafile:

e3Kif1C1_513.dat_(t)
e3Kif1C1_513.dat_a(cl)
e3Kif1C1_513.dat_histc
e3Kif1C1_513.dat_histt
e3Kif1C1_513.dat_speed

These files contain only numbers, the column headers are as follows:

***_(t)**
relative time (min) | average accumulated distance (µm) | average velocity (µm/min) | ratio Euclidian distance / accumulated distance | Mean square displacement / accumulated distance (µm)

***_a(cl)**
accumulated distance (µm) | average persistence length (µm) | average Euclidian distance (µm) | ratio Euclidian distance / accumulated distance | Mean square displacement / accumulated distance (µm)

***_histc**
This is a matrix with cl (µm) in columns and a(cl) (µm) in rows. Header values are included. The matrix contains relative frequency values and can be plotted as contour diagram.

***_histt**
This is a matrix with time (min) in columns and persistence length a (µm) in rows. Header values are included. The matrix contains relative frequency values and can be plotted as contour diagram.

***_speed**
track number | average velocity (µm/min)