Documentation for DNAvi

Author: Anja Hess
Date: 2025 APR 16
Contact: anja.hess@mail.de

Disclaimer

DNAvi is a research tool for DNA fragment visualization and analysis based on (capillary) electrophoresis. Its usage is voluntary, there is no guarantee for performance and results may vary depending on the accuracy of input information provided.

1. Requirements

DNAvi does not require prior installation of packages or software, since it is a fully web-based service with actual computation being performed on a remote server. However, performance may vary depending on the browser used. All tests have been majorly performed in Mozilla Firefox.

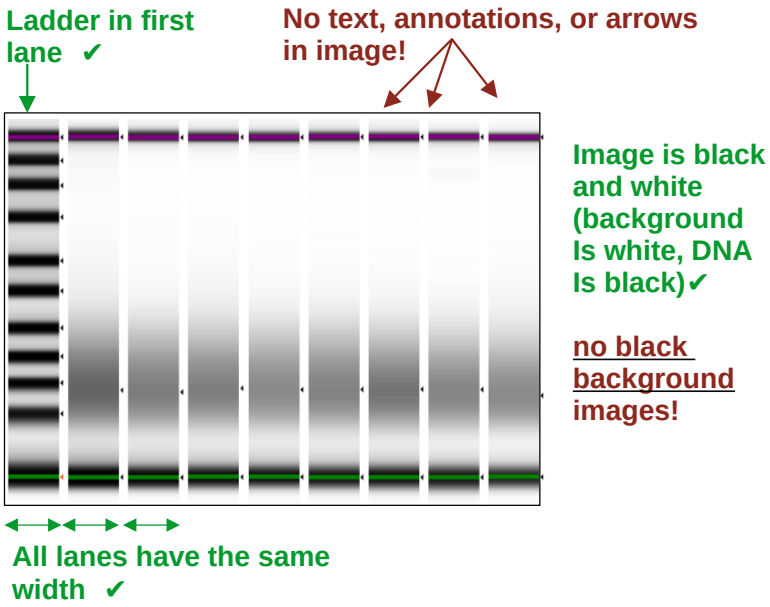
2. Table inputs

Tables have to be **tab-separated** in **.csv / .tsv / .txt** format and fulfill all criteria below:

- Columns are **tab-separated**
- The first row is the **header**
- The first column is always named **“Ladder”**
- Sample names in header (not allowed: “,;’’!.” or white space)
- Sample names must match metafile sample names (if provided)
- All **column** values are **numeric** (and refer to DNA band intensity units)

Ladder	Sample_1	Sample_2	Sample_3	Sample_4
2.56789	2.56789	2.56789	2.56789	2.56789
2.56234	2.56234	2.56234	2.56234	2.56234
1.83585	1.83585	1.83585	1.83585	1.83585
…	…	...	…	...
3.45456	2.56789	2.56789	2.56789	2.56789

2. Image inputs



The image input is naturally more difficult than giving a raw data table. The following requirements to your input image must be fulfilled:

- the format is **.png, .jpg, or .jpeg**
- the maximum file size is **16 MB**
- the gel image needs is **black&white** (white background, black DNA bands)
- the **ladder/marker is in the first lane only**
- lanes are straight and have the **same width**
- **no arrows, text, annotations, or objects are in the picture**
- **no whitespace/frame** etc is surrounding the image (crop the image if needed)
- the image has good contrast and is equally contrasted across all lanes (important to assure that bands are recognized)

Note: Inputting an inverted standard DNA agarose gel image ***may*** work, but its on ***your own risk*** and you may want to carefully check in the output folder if the bands were properly segmented. We highly recommend using only ***virtual gels from capilarry electrophoresis machines*** for optimal performance.

3. The metadata file

The metadata file is **tab-separated** in **.csv / .tsv / .txt** format and fulfills all criteria below:

- Columns are **tab-separated**
- The first row is the **header**
- The first column is always named **“SAMPLE”**
- Further variable names are put in the header (**not** allowed: “,;”!.” or “ “white space)
- Sample names must match data file sample names (if a table in .csv/.txt./.tsv is provided)

- If a gel image is provided, the **order of sample names** (top→bottom) **has to match** the order of **lane names** of the DNA gel (left→ right, no need to specify ladder here)
- The number of samples *should match* the number of given lanes in an image (or signal intensity columns in a signal table input). However, if it doesn't, you'll still be fine as long as the order and naming match.

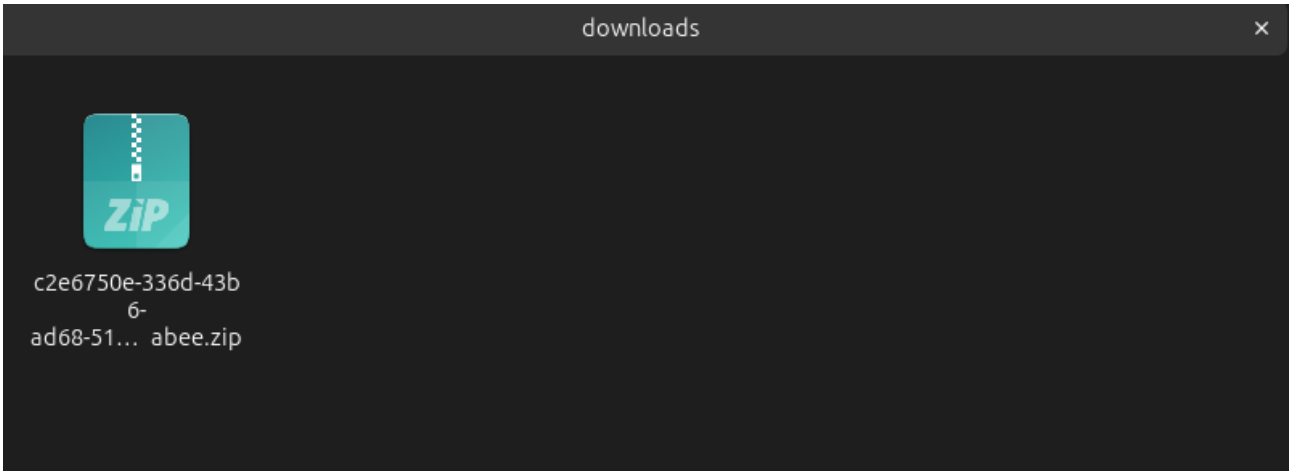
SAMPLE	YOUR_VARIABLE_1	YOUR_VARIABLE_2	...	YOUR_VARIABLE_X
Sample_1	Treated			
Sample_2	Treated			
Sample_3	Control			
…	…	...	…	...
Sample_200	Control			

Example: If a gel image with Marker+200 lanes is provided, the gel columns will be named as follows:

Ladder, Sample_1, Sample_2, Sample_3, … , Sample_200

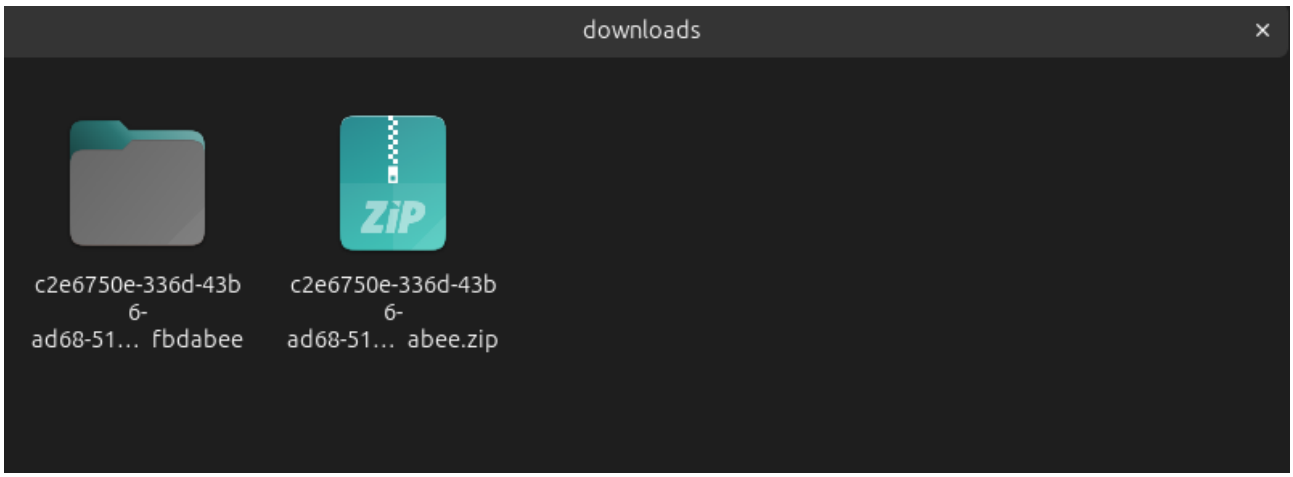
4. The outputs

After your analysis has been performed, your data will be automatically downloaded from your browser.



4.1 Analysis folder

The ZIP archive contains your unique working ID (c2e675.. in the example). Unzip the archive by double-clicking or using a software of your choice.

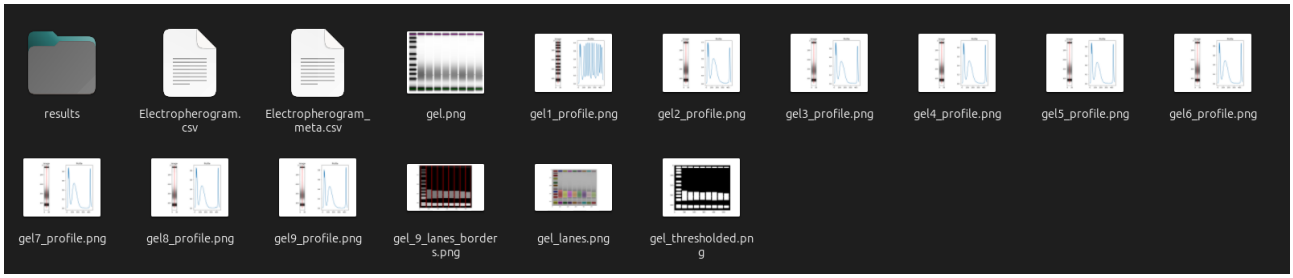


Once you have unzipped the folder, double click it to open it.

You will see the inputs you have submitted to DNAvi prior to the analysis. In the example that's the DNA intensity table file called **electropherogram.csv** and a **electropherogram_ladder.csv** file. Only if you submitted a metafile, then a third file, in this case called **electropherogram_meta.csv** will be visible. Finally there is the **results** folder which we will open now.

For gel image inputs:

In case you have uploaded a gel image instead of a table file, the folder will additionally contain some outputs from the image analysis:



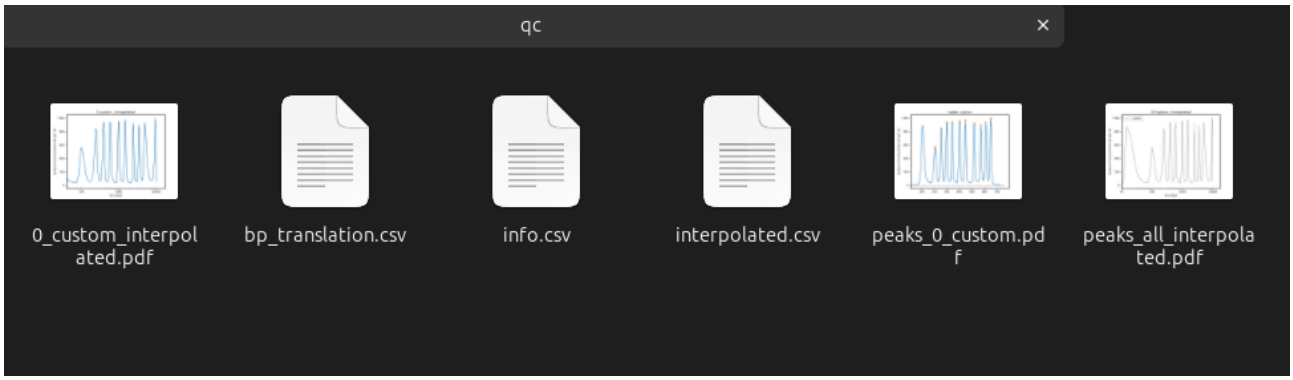
It is recommended to carefully check gel_lanes_border.png and gel_lanes.png to understand if all ladder and DNA bands have been sucessfully recongnized and in the lanes are correctly separated

before proceeding. If that's not the case, please carefully check the section on image inputs, because errors here may be likely caused by insufficient image quality.

4.1.1 Results folder

The results folder contains two sub-directories, called QC and plots.

4.1.1.1 QC



The QC folder is all about the DNA marker and detecting its peaks. It makes sense to check it and make sure your DNA ladder has been recognized correctly, and that the base pairs assigned make sense. You will find the following files:

- **info.csv** – a simple table giving information on your ladder type

Column type:

	Standard	Standard
1		0
2	Ladder	custom

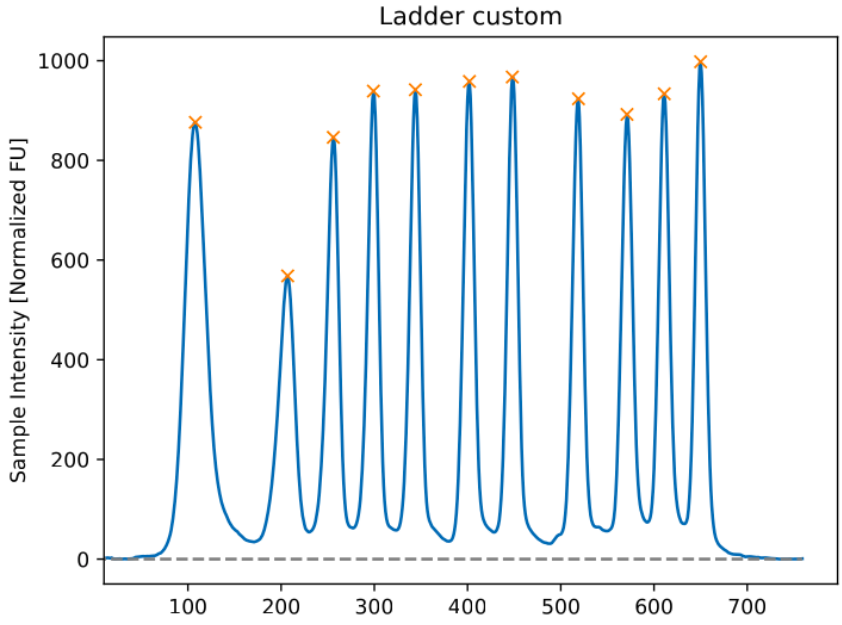
- **interpolated.csv** – your input data with missing intensity values interpolated

	Ladder	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7
0	2.989603	2.42713	0.7146179	6.35804	2.991041	1.130515	0.9038349	0.2525618
1	3.360477	2.020639	0.6151214	6.315273	2.731391	0.9291756	1.303182	0.172832
2	3.430417	1.893378	0.4197658	5.906331	2.643009	0.3953043	1.732408	0.0627502
3	3.303449	1.909102	0.239225	5.269081	2.614673	0.0980783	2.106465	0
4	3.102744	1.923925	0.1669339	4.50062	2.445065	0.3804498	2.460003	0.06304434
5	2.748271	2.036593	0.1851551	3.647917	2.159403	0.9875571	2.615517	0.3277496
6	3.560105	2.393621	0.116698	3.304178	2.076624	1.227253	2.856802	0.9612923
7	3.546266	2.70818	0.02938752	3.756241	1.825265	0.9384552	2.69857	0.913372
8	2.772796	2.508052	0.1692096	4.833201	1.390767	1.773986	2.378411	0.814907

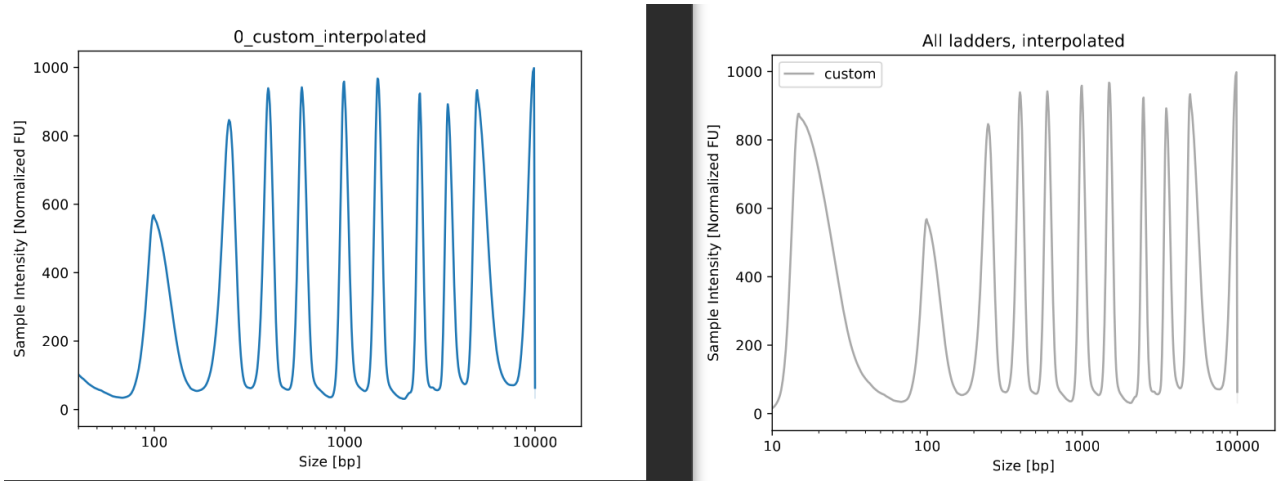
- **bp_translation.csv** – your input, but instead of ladder intensity values now with the assigned base pair position.

	Ladder	Sample_1	Sample_2	Sample_3	Sample_4	Sample_5	Sample_6	Sample_7
0	0	2.42713	0.7146179	6.35804	2.991041	1.130515	0.9038349	0.2525618
1	0.137614678899083	2.020639	0.6151214	6.315273	2.731391	0.9291756	1.303182	0.172832
2	0.275229357798165	1.893378	0.4197658	5.906331	2.643009	0.3953043	1.732408	0.0627502
3	0.412844036697248	1.909102	0.239225	5.269081	2.614673	0.0980783	2.106465	0
4	0.55045871559633	1.923925	0.1669339	4.50062	2.445065	0.3804498	2.460003	0.06304434
5	0.688073394495413	2.036593	0.1851551	3.647917	2.159403	0.9875571	2.615517	0.3277496
6	0.825688073394496	2.393621	0.116698	3.304178	2.076624	1.227253	2.856802	0.9612923

- **Peaks_N_ladder-name.pdf** – A line plot that will show you the detected peaks as yellow crosses. Make sure all peaks that you consider important are correctly detected. The x-axis will only give you positional values at that stage.

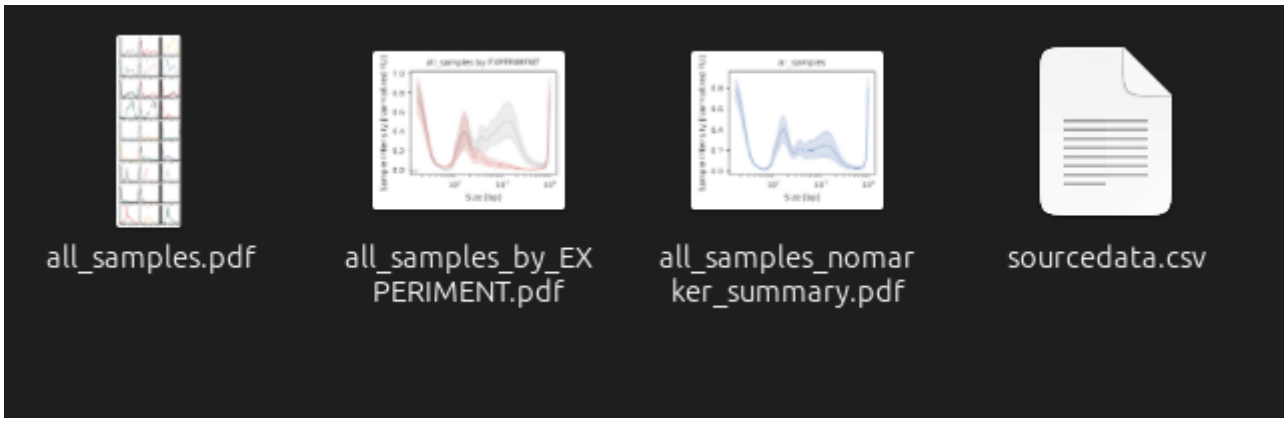


- **peaks_all_interpolated.pdf** & **N_ladder-name_interpolated.pdf** – Similar visualizations of ladder with base pairs already annotated.



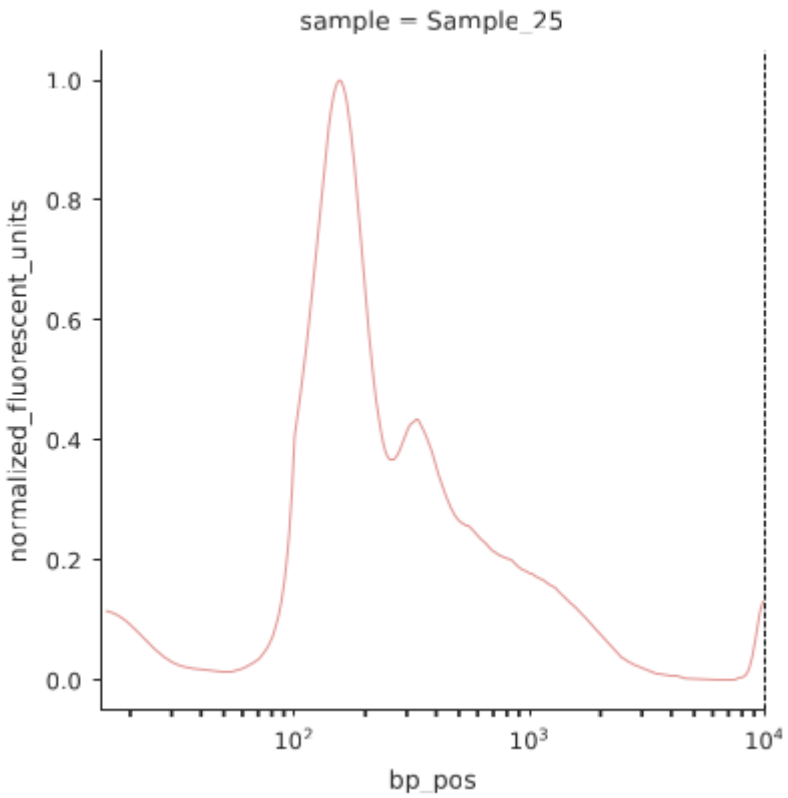
If you have checked the QC outputs and found your ladder correctly annotated, you may proceed to the plots folder and check your samples.

4.1.1.2 Plots



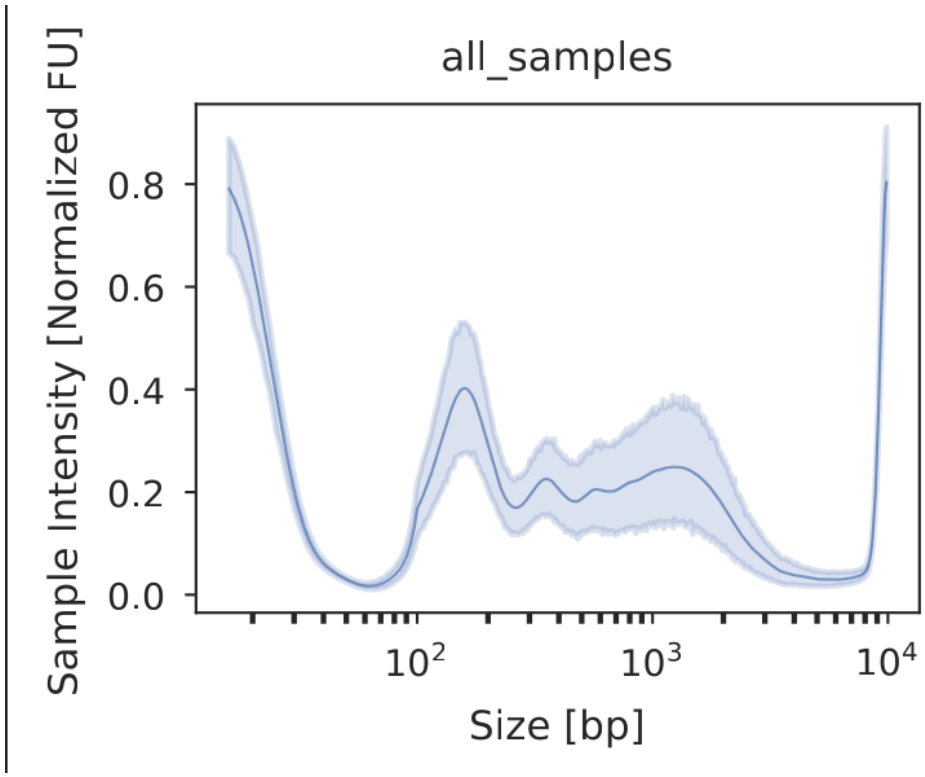
In its simplest form, the output may consist of only three files, and expand with an additional plot yor each variable you specified in the metadata:

- **all_samples.pdf** – a grid plot showing each DNA sample as an individual line plot

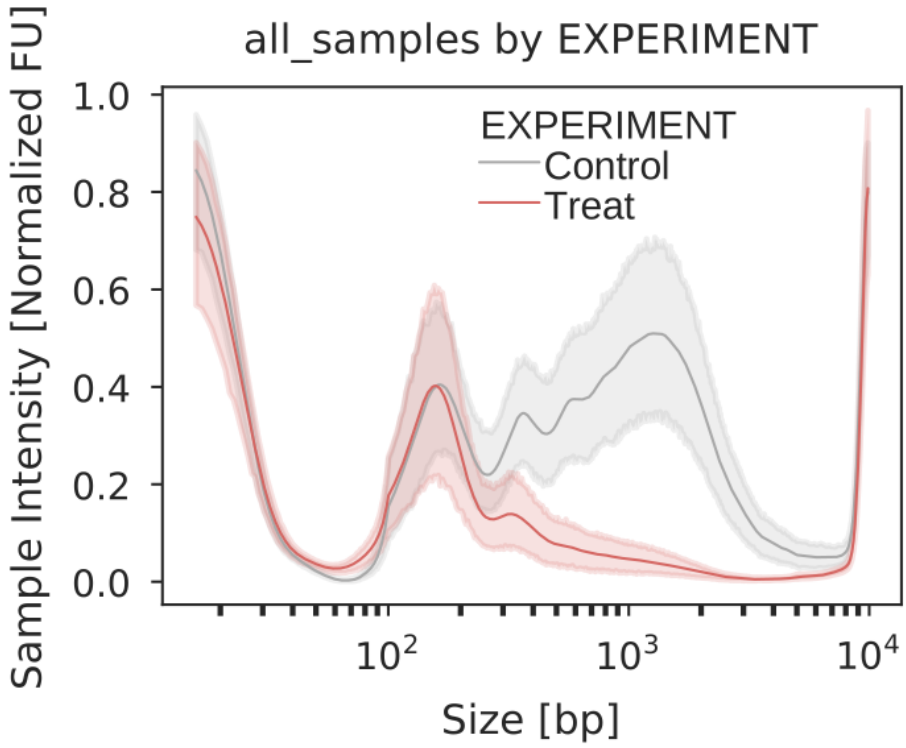


The **y-axis** shows normalized fluoresence signals to fit a value between 0 and 1. This way fragment profiles become comparable irrespective of sample concentration. The **x-axis** shows the basepair position based on the values submitted for your ladder. This is displayed in **log scale**. Each subplot is titled by the sample name you specified either in the table or meta file.

- **all_samples_nomarker_summary.pdf** – a line plot summarizing all samples including the lower and upper marker signal



- **all_samples_by_YOURVARIABLE.pdf** – for each variable specificd in the metafile, the summary line plot will stratify the samples into your groups, each group highlighted in another color.



- **sourcedata.csv** – this table provides the source data used for generating above plots and is very helpful in case you want to process / visualize the data later in another program (R, pyhton, GraphPad) or if you need to upload them for a publication.