AMiGA Analyzing Microbial Growth Assays

AMiGA is a python-based program that facilitate the high-throughput analysis of microbial growth data. It applies Gaussian Process (GP) regression to infer microbial growth parameters such as maximum specific growth rate, doubling time, lag phase, and carrying capacity. It is especially useful for the analysis of Biolog Phenotypic Microarray (PM) data. The flexibility and utility of GP regression enables (1) the analysis of microbial growth data that does not follow standard logistic or sigmoidal growth, (2) inference of non-standard microbial dynamics such as diauxic shifts, and (3) hypothesis-driven statistical testing of differences in microbial growth under different environmental conditions. AMiGA is a minimalist, modular, and user-friendly program that allows for the analysis of single or multiple files in a single batch. It requires a single command line in the terminal. User arguments can be passed via the terminal or simply using the text-based parameter files described below.

Organization of folders and files in the working directory

Working Directory

- data
- data_derived
- mapping
- parameters
- figures
- summary

Icons Legend

- Folder of required input
- Folder of optional input
- Folder of AMiGA output
- Simple text file

folder name

file name

data (example)

- PRB-83_xeno.asc
- CD630_PM1-1.txt
- CD630_PM1-2.txt

mapping (example)

- meta.txt
- CD630_PM1-1.txt
- CD630_PM1-2.txt

parameters

- subset.txt
- flag.txt
- hypothesis.txt
- interval.txt

Optional files for passing meta-data on plates or wells

A meta.txt – Defines the meta-data of plates included in the data folder.

B User can instead pass individual mapping file for each plate with information on each individual well in each plate.

See section What are mapping files and how to use them? for more information.

Optional files for specifying model parameters or analysis

A subset.txt – Defines the specific conditions to be analyzed from the whole collection of data available in the working directory. For example, you can select for specific set of plates or analyze wells that meet certain criteria defined in meta-data. This is not needed if you want to analyze all data.

B flag.txt – Defines specific wells that will be ignored in the analysis. This can be used to highlight any number of wells that do not have reliable growth curves (for example due to instrumental noise) without ignoring the whole plate. This is not needed if you do not want to ignore any wells.

C hypothesis.txt – Defines the null and alternate hypothesis for detecting differential growth due to specific variables, e.g. ribotype, carbon substrate, antibiotic stress, or genetic background. This is not needed if you are not performing hypothesis testing.

D interval.txt – Defines the interval (in seconds) between time points in the specified plates. By default, the interval is set to 600 seconds but can be modified here for specific plates. This is not needed if all files have a 600 second interval.

See section What are parameters files and how to use them?" for more information.

How to format data files?

AMiGA analyzes growth curves imported from one or multiple plates. Each plate will have a corresponding data file in data.

- Each file must be encoded either in ASCII or BOM (e.g. UTF-8) and values must be tab-separated. This is likely to be the case for output by most plate readers.
- The first column must be the well IDs. The first row with raw data must start with "A1". The current version will only accept standard well IDs (i.e. A1, B1, …, G12, H12) in the format of 96-well microplates. Future versions should be a bit more flexible.
- If a text files has meta-data as headers, AMiGA will simply ignore these lines, search for A1, then read until it reaches H12 or the end of the file.
- Currently, the text reader will ignore the Time row and assume that the time interval between measurements (i.e. columns) is 600 seconds. See below section on parameter files on how to define a different time interval for each plate instead of using the default value. Future versions of AMiGA will detect Time rows.

data file example (CD630_PM1-1.txt)

Time	0	600	1200	1800	2400	…	59400
A1	0.277	0.275	0.279	0.280	0.289	…	0.474
B1	0.282	0.275	0.280	0.281	0.289	…	0.478
C1	0.279	0.276	0.278	0.282	0.289	…	0.478
D1	0.267	0.266	0.271	0.274	0.282	…	0.416
…	…	…	…	…	…	…	…
E12	0.303	0.295	0.301	0.309	0.314	…	0.713
F12	0.285	0.299	0.313	0.310	0.305	…	0.667
G12	0.278	0.297	0.313	0.333	0.349	…	0.753
H12	0.265	0.270	0.271	0.291	0.293	…	0.614

What are parameter files and how to use them?

Minor note: When parsing these parameter text files, **AMiGA** ignores white-spaces and interprets the content based on the use of colons (:) and commas (,) as delimiters or separators. Refrain from using those symbols in your plate IDs (i.e. filenames) or inside your mapping data.

subset.txt

```
Substrate : D-Trehalose, D-Sorbitol
Ribotype : R017, R027, R078
```

Formatting – Each line in **subset.txt** defines a selection criteria where the first word is the variable name (case-sensitive and should match exactly to the variable name in the meta-data), followed by a colon, which is then followed by a comma-separated list of variable values. In this example, the model will analyze wells that correspond to a bacteria belonging to either one of three ribotypes and grown on either of one of two substrates.

flag.txt

```
CD630_PM1-2 : G10, H2
PRB1950_PM2-3 : A3, D4, D5
```

Formatting – Each line in **flag.txt** defines a selection criteria where the first word is the *Plate_ID* (case-sensitive and should match exactly to corresponding data file name), followed by a colon, which is then followed by a comma-separated list of well IDs. In this case, the model will ignore two wells in *CD630_PM1-2* and three wells in *PRB1950_PM2-3*. If there are more data files beyond what is listed here, the model will not ignore any wells because none were highlighted in this file.

hypothesis.txt

```
H0 : Time
H1 : Time + Ribotype
```

Formatting – Only two lines in **hypothesis.txt** are allowed. The first line specifies the null hypothesis or model by declaring *H0*, followed by a colon, which is then followed by at least one variable such as *Time*. The second line specifies the alternate hypothesis or model by declaring *H1* followed by a colon followed by at least one variable. Multiple variables in each hypothesis are separated by plus signs (+).

interval.txt

```
CD630_PM1-2 : 300
CD630_PM2-1 : 500
```

Formatting – Each line in **interval.txt** defines the time interval between observations for a specific plate. The first word is the *Plate_ID* (case-sensitive and should match exactly to corresponding data file in data folder), followed by a colon, which is then followed by an integer that defines the time interval. By default, the program will define the time interval as 600 seconds unless otherwise noted here.

What are mapping files and how to create them?

AMiGA analyzes growth curves imported from one or multiple plates where each plate will have a data file and a corresponding mapping file, the latter is provided by the user or auto-generated by the program.

- Each row in the mapping file corresponds to a well. A minimalist mapping file (see example 1) has a column for well IDs and a column for *Plate_ID* (i.e. filename without extension). The well IDs in the mapping file must match the well IDs in the data file. A minimalist file can be auto-generated by **AMiGA** (see B below).
- For Biolog Phenotypic Microarray (PM) plates, mapping files will include the following additional columns: *Isolate*, *Name*, *PM* number, *Replicate* number, and *Substrate* (see example 2). This can also be auto-generated by **AMiGA** (see A and C below).
- The user can also add additional meta-data variables by either passing it through plate-specific mapping files or a run-specific meta-data file (see C and D below).
- Plates don't have to follow the typical 96-well format (8x12) unless they correspond to Biolog plates. A plate needs at least one row and can house more than 96 rows. Row (or index) IDs don't have to be well IDs but must match between corresponding data and mapping files.

How to pass mapping information to AMiGA?

A If you are processing Biolog PM plates, simply name your data file using this nomenclature: {isolate name}_PM{integer}-{integer} where the first integer indicates the PM number and the second indicates the replicate number (an example is CD630_PM1-2.txt). **AMiGA** will automatically generate a mapping file similar to mapping example 2 (above).

B If you simply want to process your data through **AMiGA** without appending meta-data to it, just simply deposit your data file and AMiGA will generate the minimalist mapping file similar to example 1 (above). The index column (i.e. A1, B1, … etc.) will match the well IDs in the data file.

C If you have meta-data that is file-specific, create in mapping a meta.txt that lists information on files deposited in data. If a specific plate is a Biolog PM, make sure to include a *PM* column with values between 1 and 7 and **AMiGA** will auto-generate a *Substrate* column in the mapping file. meta.txt must include a *Plate_ID* column that matches the name of the data file (without the file extension). Using meta.txt, **AMiGA** will automatically generate individual mapping files for each plate unless an individual mapping file has already been provided by the user in mapping (see D). Please save meta.txt as tab-separated text file.

D If you have meta-data that are well-specific (i.e. different wells within same file have different meta-data), you must create an individual mapping file for your plate and deposit it in mapping. This is also recommended for data files that do not correspond to a standard microplate of 96 wells. Make sure that the index column (i.e. well IDs) matches exactly the index column in the data file.

E You can pass mapping information using a mixture of the above options. You can have individual plate-specific mapping files in addition to a meta.txt. However, a plate-specific mapping file will always over-ride the plate-specific info in meta.txt.

mapping file example 1 (minimalist)

	Plate_ID
A1	CD630_PM1-1
B1	CD630_PM1-1
….	…
G12	CD630_PM1-1
H12	CD630_PM1-1

mapping file example 2 (Biolog PM)

	Plate_ID	Isolate	PM	Replicate	Substrate
A1	CD630_PM1-2	CD630	1	2	Negative Control
B1	CD630_PM1-2	CD630	1	2	D-Serine
C1	CD630_PM1-2	CD630	1	2	D-Glucose-6-Phosphate
….	…	…	…	…	…
F12	CD630_PM1-2	CD630	1	2	Inosine
G12	CD630_PM1-2	CD630	1	2	L-Malic Acid
H12	CD630_PM1-2	CD630	1	2	2-Aminoethanol

mapping file example 3 (with additional user-provided meta-data)

	Plate_ID	Isolate	Substrate	Ribotype	Comments
A1	CD_treA	CD89_wt	Negative Control	RT027	Wild-type
A2	CD_treA	CD89_wt	D-Trehalose	RT027	Wild-type
A3	CD_treA	CD89_wt	D-Trehalose	RT027	Wild-type
B1	CD_treA	CD89_ko	Negative Control	RT027	treA knock-out
B2	CD_treA	CD89_ko	D-Trehalose	RT027	treA knock-out
B3	CD_treA	CD89_ko	D-Trehalose	RT027	treA knock-out

meta.txt

	Plate_ID	Isolate	PM	Replicate	Ribotype
1	CD630_PM2-1	CD630	2	1	RT012
2	CD630_PM2-2	CD630	2	2	RT102
3	PRB-83_PM1-1	PRB-83	2	1	RT027
4	PRB-83_PM1-2	PRB-83	2	2	RT027
5	PRB-83_PM1-3	PRB-83	2	3	RT027
6	FSM-79_xeno	CD630		1	RT017
7	PRB-83_xeno	PRB-83		1	RT027

Contact Information

Firas Midani, Ph.D.
https://github.com/firasmidani
midani@bcm.edu

Lab of Robert Britton
Alkek Center for Metagenomics and Microbiome Research (CMMR)
Department of Microbiology and Molecular Virology
Baylor College of Medicine – Houston, TX

AMiGA

DATA

MAPPING

USER INPUT

TXT or ASC

TXT

Quality Control

meta.txt

USER-DEFINED PARAMETERS

flag.txt

USER INPUT

hypothesis.txt

Growth Data

Meta-Data

interval.txt

subset.txt

Growth Curve Fitting with GP Regression

OUTPUT

Bayesian Evidence

OUTPUT

Summary Statistics

OUTPUT

Growth Visualization