

Auto qPCR User

Author(s): Iveta Demirova

Version	Authors/Updated by	Date	Signature
v0.1	Iveta Demirova	31 August, 2020	
V0.1	Rhalena Thomas	05 October, 2020	

The involved functions approve the document for its intended use:

Name	Function	Role	Date	Signature

GuidTable of Contents

1	Introduction.....	1
1.1	Objectives.....	1
1.2	Protocol overview.....	1
1.3	Technical considerations.....	1
2	Materials.....	2
2.1	Requirements: Local.....	2
2.2	Requirements: Online.....	2
3	Protocol.....	3
4	Appendix.....	5

1 Introduction

1.1 Objectives

This protocol describes how to use the Auto qPCR application, both online and on your local host. The Auto qPCR application was created to speed up processing of qPCR data, and minimize errors that can accumulate when data are moved between different platforms. The application is currently capable of processing qPCR data using absolute and relative methods, and assaying genomic instability, as well as performing a variety of statistical tests on the data, outputting raw, summary and statistics tables, and plotting results.

1.2 Protocol Overview

The Auto-qPCR program is made to be flexible and is capable of reading data output from various qPCR machines due to its ability to recognize synonymous column names and intelligently interpret different naming conventions. However, if you experience problems despite correctly filling out the analysis set-up form, please consult one of the example input files on the GitHub and try renaming your columns to match the format you see.

1.3 Technical and safety considerations

The following information should be read before starting:

- Technical considerations:
 - Auto qPCR can be run online at www.auto-qpcr.com or on your local host by downloading the program from the Auto qPCR GitHub and running main.py
- We also recommend looking at the examples files on the GitHub, and the other documentation there so that you may better understand how the program works.

2 Materials

2.1 Requirements: Local

The Auto qPCR program can be run on your local host, regardless of your operational system, although how you run the main.py file may be different. The local address should be 127.0.0.1 (PC) or 0.0.0.0 (Mac), and port 5000 should be accessible. Please see the requirements file on the GitHub for the most up-to-date list of Python libraries that you need installed on your machine, or below for the list at the time of publication:

- Python 3.0
- Libraries:
 - o Pandas
 - o Numpy
 - o Re
 - o Io
 - o Datetime
 - o Zipfile
 - o Psutil
 - o Os
 - o Logging
 - o Flask
 - o Mathplotlib.pyplot
 - o Pinguoin
 - o Scipy.stats

2.2 Requirements: On-line

The program was designed to run in Google Chrome so please use this browser for the best experience.

3 Protocol

3.1 Local GUI

Materials:

- Optional: Workbook (Jupyter, Spyder etc.)

Procedure:

For local GUI, first:

- i. Download auto-q-pcr-local_GUI folder from the Auto qPCR Github
- ii. Verify that all technical requirements are met (including that Python 3.0 and necessary libraries are installed)
- iii. Run main.py through your workbook, or command line/bash depending on operating system
- iv. Open link returned by the program in your browser

3.2 Running Auto-qPCR analysis

For both the local and on-line Auto-qPCR application:

1. Select model

- Selects the analysis method to be used.
- Absolute: calculates the quantity of cDNA based on the CT using a calibration curve and normalizes quantities to quantity of an endogenous control
- Relative (delta CT): calculates cDNA quantity relative to an endogenous control(s) target.
- Relative (delta delta CT): calculates cDNA quantity relative to the endogenous control target(s) and a reference sample.
- Genome Instability: uses delta delta CT method to compare quantity of DNA from one chromosome to another. The first delta is now a DNA region not a gene and the second delta is still a references/control sample.

2. Upload your data (.csv): Input csv files to be analyzed.

The files must contain:

- "Well" column (Well must be in the header of the column)
- Column defining the Task/Role/Content of each data row (ex. Unknown or Sample vs Standard and Blank)
- Column containing the Sample Names (i.e. cell lines, chromosomes, time points etc.)
- Column containing the Target/Gene Names whose quantities is being assessed. (If each file corresponds to a different gene/target and this name is in the file name but not the file itself, this is fine and names can be listed under "File information" in the form).

- Column containing the CT value for each sample+target measurement
- Column containing the Quantity

The files may also contain:

- Column identifying the dye and quencher used
- Column containing the group names (for example if different samples have undergone different treatments, or are part of biological replicate groups). This is not necessary if the group names are in the sample name (for example, for IPSC-1, IPSC-2, NPC-1, NPC-2, the groups could be IPSC and NPC, or 1 and 2).

3. File information: Select yes to see the input boxes

- Identify your target/gene names in the file name if there is no target file in the raw data
- The name of your quencher: this will filter out any data with this quencher and can be used if you have reading from two probes and want to remove artifacts. This option can also be used to analyze two probes run together. The data needs to be run twice filter out the second probe each time. Leave this blank if no filtering is needed.
- Name of the sample group in the last column. This will select your sample rows in the Task/Role/Content column to be analyzed.

4. Options: Decide on outlier identification methods and enter controls

- Endogenous control : the sample to normalize with in absolute and the first delta in the relative delta-CT model
- Cut-off: Decide on the cut off CT sample standard deviation (std) for a measurement in a group to be considered an outlier. The standard threshold is 0.3.
- Decide on maximum proportion of data points in a group that can be considered outliers and removed. The default is 0.5 which indicates that if there are three replicates only one can be removed.
- Decide whether to conserve replicates in groups with high variation where the std is over the specified cut-off but spread is great and certainty of outlier identity of further measures is less certain. If preserve outlier is set to yes and the std is over the threshold the $\text{abs}(\text{mean}-\text{median})/\text{media}$ must be greater than 0.1 for an outlier to be removed.
- Control sample: is the reference sample for the second delta in the relative delta-delta-CT model.

5. Visualizaton Options: Optionally enter the order you would like samples and targets to appear in in plots and tables.

- Target order refers to the genes or DNA regions
- Sample order: refers to the order that triplicate means should appear.

- Enter column name for statistical analysis: This must be included if you wish to perform statistics and your group names are not included in the sample names. You may insert a group column or have defined a group column with in the qPCR machine.
6. **Statistics.** Choose yes if you want to add statistics.
 - How many groups do you have.
 - Select one-way or two way anova. If you only have only two samples, select one-way ANOVA and t-test will be performed.
 - Select either 'I have a group column' or 'My group names are within sample names' to match your situation
 - Enter the names of your groups (samples, time points, treatment) to compare
 - Enter if measure are independent or dependent. Measurements are considered independent if they are repeated measure from the same samples. We consider time points in the same cell line to be repeated measures.
 - Is your data normally distributed? Normally distributed data should be analyzed with parametric test (t-test, ANOVA). Data with abnormal distributions should be analyzed with non-parametric test (MannWhitneyU test, Kruskal-Wallis) where data is ranked.
 7. Click submit wait for the files to be processed. When it is complete a window will come up to save the zip file.
 8. Save zip folder containing complete analysis and be sure to check over clean and summary data files to ensure your data was interpreted correctly.

4 Appendix

Table S4: Description of the statistical tests used each possible selection criteria The number of groups to compare, '#G' indicates the number of conditions to compare with variable. The number of variables, '#Var' indicates the number of experimental conditions to compare. The distribution of the data determines if a parametric test will be used, for normally distributed data or a non-parametric test will be used by the software. Measurement indicates if the data was collected on in depended samples or on the same samples at different time points. Test, indicates the name of the test used by the software based on the user's sections from the other 4 criteria. Auto-qPCR always uses a the same post-hoc test except when only two groups are being compared and no post-hoc test is performed.

# G	# V	Distribution	Measure	Test	Posthoc
2	1	parametric (normal)	Independent	student t-test two tailed, un-paired	none
2	1	parametric (normal)	Repeated measures (dependent)	student t-test two tailed, paired	none
2	1	non-parametric	Independent	Wilcoxon test	none
2	1	non-parametric	Repeated measures (dependent)	ManWhitney U test	none
> 2	1	parametric (normal)	Independent	one way ANOVA	pairwise t-tests with FDR correction
> 2	1	parametric (normal)	Repeated measures (dependent)	one way ANOVA	pairwise t-tests with FDR correction
> 2	1	non-parametric	Independent	Kruskal-Wallis test	pairwise t-tests with FDR correction
> 2	1	non-parametric	Repeated measures (dependent)	Friedman test	pairwise t-tests with FDR correction
> 2	2	parametric (normal)	Independent	two way ANOVA	pairwise t-tests with FDR correction
> 2	2	parametric (normal)	Repeated measures (dependent)	two way ANOVA	pairwise t-tests with FDR correction, for conditions 1,2 and

					the interaction
--	--	--	--	--	-----------------