

**QAS**  
**A software for automated quantitative comparison of data**  
**from melting analysis of *dsDNA***

**Content:**

1. Introduction	2
2. Which data formats are supported	3
3. How to install QAS	3
4. Setting general options	3
5. Setting data processing options	3
6. How to entry datafile	4
7. Browsing to path for output files	4
8. Generation of distance matrix (infile)	5
9. Generation of outtree file and outfile	5
10. How to convert sample labels in outtree and outfile	6
11. How to generate a dendrogram, cladogram etc.	6
12. Concluding remarks	6
13. References	6

---

© Vladislav Raclavský and Jiří Pavlíček  
Department of Microbiology  
Faculty of Medicine & Dentistry  
Palacký University  
Hněvotínská 3  
775 15 Olomouc  
Czech Republic  
<http://www.mikrobiologie.upol.cz>  
e-mail: [rac@tunw.upol.cz](mailto:rac@tunw.upol.cz)

# 1. Introduction

This software tool was developed for the free use by scientific community by Vladislav Raclavský, M.D., Ph.D. (concept of the comparing algorithm and user evaluation of versions under development) and Petr Pavlíček, M.Sc. (software development) at the Dept. of Microbiology, Faculty of Medicine and Dentistry, Palacký University, Olomouc, Czech Republic, under the support of Ministry of Health, Czech Republic (NR 8365-4/2005). It automates and facilitates quantification of differences in melting data or high resolution melting data from analysis of double-stranded DNA, typically acquired by PCR-amplification followed by melting analysis using a commercial instrument. However, any melting data or similar types of datasets can be compared using this tool given they meet the standards of file formats accepted by the software tool. The software is available free of charge through the web page of the Dept. of Microbiology at <http://www.mikrobiologie.upol.cz>.

Briefly, differences in melting data obtained with at least two samples are compared quantitatively. These melting data can come either in form of normalized relative fluorescence data (normalized melting curves) or in form of negative first derivation of normalized relative fluorescence data with temperature ( $-dF/\Delta t$ , so-called derivation curves). Both sets of data which are compared have to be acquired at the same scale, i.e. temperature points at the X-axis have to be the same. This is accomplished by setting a standard same interval of temperatures for acquisition of data and a standard same ramp rate (temperature increment for data acquisition) at the melting instrument. Differences between the datasets (curves) are calculated by subtracting their Y-values (normalized fluorescence or derivation) at each temperature point during melting analysis (each X-value). Then, sum of these subtracted values represented absolute numerical distance between two samples, i.e.:

$$AD_{1,2} = \sum_{t=78}^{91} |f_1(t) - f_2(t)|$$

where:

$AD_{1,2}$  is the absolute difference (distance) between datasets acquired with samples No. 1 and 2

$f_1(t)$  is the fluorescence value (normalized or derivation) of sample No. 1 measured at temperature  $t$

$f_2(t)$  is the fluorescence value (normalized or derivation) of sample No. 2 measured at temperature  $t$

When more than 2 samples (datasets) are included in the analysis, the software calculates the difference between each combination of datasets and the maximum relative distance 1.0 is assigned to the highest absolute value obtained in the most distant (numerically different) pair of datasets, abbreviated as  $AD_{\max}$ . Relative distance values for the remaining pairs of datasets are calculated as a fraction of the highest absolute value, i.e.:

$$RD_{1,2} = \frac{AD_{1,2}}{AD_{\max}}$$

A matrix of relative distances is then assembled in a file accepted e.g. by the PHYLIP free software package (infile) [1,2] to generate data for tree construction. Matrix of relative distances can also be analyzed manually, e.g. to identify the pair of datasets showing lowest quantitative difference (distance). Such analysis can be useful for instance to match an unknown sample with the best match in a database of previously processed samples.

## 2. Which data formats are supported

For the moment two data formats are supported – those generated by the High Resolution Melter HR-1 (Idaho Technology) and Rotor-Gene Q (Qiagen, formerly Rotor-Gene 6000 by Corbett Life Science). When using the HR-1 Melt Analysis Tool software, export the graph data (either normalized plot or derivative plot) as a plain type file. This can then be used as entry datafile by QAS directly. When using the Rotor-Gene Series 6000 software export the graph data (either normalized plot or derivative plot) as an \*.xml file. This can also be used as entry datafile by QAS directly.

## 3. How to install QAS

Java software (Java Runtime Environment) has to be installed on your computer to run QAS. For free download go to [www.java.com](http://www.java.com). After downloading the file „qas-1\_9.zip“ from our webpage into a folder of your choice, unzip it. One file „Qas.jar“ will be generated in your folder of choice and another file „jdom.jar“ will be generated in a subfolder designated „lib“. Simply run QAS by double-clicking the file „Qas.jar“.

## 4. Setting general options

After starting QAS, set general option in the corresponding bookmark firstly.

Attention! New settings will become valid only after saving them by clicking the „Save“ button at the bottom of the „General options“ bookmark.

In the log option you can check all information you would like to appear in the log window („Information on processing“) in the „Distance matrix“ bookmark during processing of data.

Setting up the default file path is useful for convenient repeated processing of data. When QAS is used to normalize data for removing irregularities before further processing (useful for HR-1 melting data), name of generated file should always be specified in advance. This file is generated in \*.csv format accepted by most spreadsheets (e.g. Excel, OpenOffice Calc, TeaChart). For Czech version of spreadsheets, it can be convenient to replace decimal points in numbers by comma. Alternatively, using decimal points instead of comma can be set in language settings of your operating systems to fix this problem of interoperability between Czech and English language environment.

## 5. Setting data processing options

Attention! New settings will become valid only after saving them by clicking the „Save“ button at the bottom of the „Data processing options“ bookmark.

It is always recommended to check the „Perform validation“ button. One cycle of validation should be enough for typical melting data.

Because we have encountered problems with unexpected point irregularities in HR-1 data, we also recommend to check the „Perform normalization of temperature points“ button if HR-1 data are

processed. During this normalization, data are averaged based on the values of few adjacent temperature points starting from the temperature set manually. This starting temperature has to be higher than the first temperature point in your dataset of melting data. High resolution melting data are typically acquired at a 0.1 °C or 0.3 °C step in the melting instrument. For both cases, we recommend to set the normalization step to 0.5 °C which is enough to fix the problem with unexpected point irregularities in most cases. Setting a higher normalization step will secure better smoothing of data, but will reduce their complexity and predictive power.

Distance matrix that summarizes quantitative differences in melting data between processed samples is the main output of QAS. Distance values for each pair of isolates in the matrix are relative values calculated as described in Introduction. To establish these relative values, an absolute value has to be set. If you check the „Calculate the coefficient...“ button, this will be done automatically based on the highest absolute distance value obtained with the most dissimilar pair of dataset – this option is recommended for most cases. However, if you desire to compare e.g. several dendrograms or cladograms obtained in different experiments on the same scale, typically in a composite picture in a paper, it can be useful to set one value shared in different experiments manually. To achieve this, you first have to run all the analyses with automated calculation of the coefficient and only then opt for the coefficient which will suit all of them. This manually set coefficient has to be equal or higher than the highest coefficient obtained by automatic calculation in the most diverse dataset.

## 6. How to entry datafile

In the „Distance matrix“ bookmark browse for the entry datafile containing your dataset. Set of normalized melting curves generated by the HR-1 software can be used as well as a set of derivation data from the same software, both exported from the software using graph data export as plain type file. Alternatively, normalized melting data or derivation data in the \*.xml format exported from the Rotor-Gene Q/Rotor-Gene 6000 software can be used.

If you need to combine data from different experiments, you do not need to compile them into one entry file for QAS. Instead, simply browse step by step for all of the files containing the desired data. Then, all of them will be listed in the entry datafile line, separated by semi-colon. Once you start the generation of distance matrix (Chapter 8), all of them will be processed together. Attention! This feature helps processing of several datasets, however, if you want to start a new session with new data, you always have to erase the content of the browsing line! Otherwise, your new data will be added to the previously processed data and all of them will be processed together.

## 7. Browsing to path for output files

Before you start generation of distance matrix, always be sure to browse for proper path to output datafiles. The output datafiles generated after each QAS run are always generated under the same name, namely „infile“ and „phylipNameConversionTable.csv“. Thus, older files of the same name placed in the same folder will be overwritten without prior notice! This is because the PHYPLIP free software package recommended by us for later generation of an outtree requires a file named always simply „infile“ for input data. Thus, we recommend to create always a new folder for output data generated by each QAS data processing run. Discriminatory characteristics of a respective QAS processing run can then be conveniently specified in the name of this folder. This folder should then be browsed for output files immediately before starting the processing run.

## 8. Generation of distance matrix (infile)

After general options and data processing options were checked properly and entry datafile and path for output files were browsed correctly, click the „Start generation of distance matrix“ button and watch the content of the „Information on processing“ window for the progress of processing (in Czech only for the moment). Once the processing is finished, the message „Soubor infile pro PHYLIPa vytvořen“ (=infile for PHYLIP generated) appears at the bottom line.

Important! In case when data from different datasets do not share the same X-axis scale (starting with the same temperature and melting recorded at the same temperature increments), sets cannot be compared and processing will collapse (no progress apparent in the „Information on processing“ window). In such case, click the „Cancel processing“ button. To solve the problem, first check whether possibly a dataset from a previous run, which is not compatible with the newly added dataset, was maybe left in the entry datafile browsing line. When this does not help, inspect the entry datafile using a simple text processor (WordPad recommended) to check for possible incompatible formatting (compare to files processed successfully earlier). In case you cannot solve the problem, please send us the file for examination (see Chapter 12).

Once the processing is finished, the generated distance matrix appears in the „Information on processing“ window under the subheading „Matice pro PHYLIP:“ and it is also written into the infile saved into the folder specified in the path to output datafiles. Because PHYLIP software does not accept long and variable sample labels, original labels found in entry datafile are automatically replaced by labels „vz\_01, vz\_02, vz\_03...“ etc. A conversion table of the new and the corresponding original labels appears below the distance matrix in the „Information on processing“ window and is also written into the „phylipNameConversionTable.csv“ file saved into the folder specified in the path to output datafiles. Original sample labeling can be automatically recovered later after generating and outtree file (see below).

## 9. Generation of outtree file and outfile

The „infile“ file generated by QAS and saved in the folder specified in the path to output datafiles harbours distance matrix between processed datasets of melting data. This file can be used by different types of software to calculate data for a phylogenetic tree. In case of melting data, such tree of course typically cannot serve phylogenetic purposes, however, this type of plotting the distance matrix nicely visualizes the relative similarity of processed datasets and can serve their rapid and easy comparison, including identification of important relationships between groups of datasets forming closely related clusters in a „phylogenetic“ tree.

We recommend to use the widely accepted software PHYLIP [1,2] for calculating tree data, which can be downloaded free of charge at <http://evolution.gs.washington.edu/phylip.html>. For your best convenience, we recommend to copy the executable file „neighbor.exe“ found in ...phylip3.67/exe/ folder into the same folder with your output files. Once you run neighbor.exe, it will recognize your „infile“ automatically and will generate two other files in the same folder. The „outfile“ harbours simple schematic plot of a phylogenetic tree, whereas the „outtree“ harbours data suitable for tree construction by other software tools (see PHYLIP accompanying documentation for details). Both files are readable as common \*.txt files. We recommend to use WordPad or similar simple editor to view the files, if desired.

## 10. How to convert sample labels in outtree and outfile

Now, when PHYLIP has finished generation of „outfile“ and „outtree“, you can use the conversion function of QAS to revert the labeling of your datasets in „outfile“ and „outtree“ from to the „vz\_01, vz\_02, vz\_03...“ etc. labeling to the original labeling. Click the button „Apply Phylip name conversion table“ in the Distance matrix bookmark of QAS. Files named „outfile-converted.txt“ and „outtree-converted.txt“ will be generated in the same output datafiles folder.

## 11. How to generate a dendrogram, cladogram etc.

As the next step, we recommend to use another free software tool to construct a publication-suited dendrogram, cladogram or other type of tree plot using the converted data, e.g. Phylodraw ver. 0.8 software from Pusan University (<http://pearl.cs.pusan.ac.kr/phylo draw>) [3,4] or TreeView software from Glasgow University (<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>) [5,6]. Simply use the respective software open dialog to browse to the output datafiles folder and open the converted tree file.

## 12. Concluding remarks

Please contact us at [rac@tunw.upol.cz](mailto:rac@tunw.upol.cz) if you need any help with QAS or if you have any comments or suggestions. Participation in further development of the tool or cooperation in the field is of course welcome.

## 13. References

1. Felsenstein J: **PHYLIP - Phylogeny Inference Package (Version 3.2)**. *Cladistics* 1989, **5**(2):164-166.
2. **PHYLIP** [<http://evolution.genetics.washington.edu/phylip.html>]
3. Choi JH, Jung HY, Kim HS, Cho HG: **PhyloDraw: a phylogenetic tree drawing system**. *Bioinformatics* 2000, **16**(11):1056-1058.
4. **PhyloDraw: A Phylogenetic Tree Drawing System** [<http://pearl.cs.pusan.ac.kr/phylo draw>]
5. Page RD: **TreeView: an application to display phylogenetic trees on personal computers**. *Comput Appl Biosci* 1996, **12**(4):357-358.
6. **TreeView** [<http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>]