# A survey on gel images analysis software tools

**Mahmoud H.Alnamoly[1], Ahmed M. Alzohairy[2], Ibrahim M. El-Henawy [1]**

1 Faculty of Computers and Informatics, Zagazig University, Computer Science Department,

2 Faculty of agriculture, Zagazig University, Genetics Department

Emails: mahmoudalnamoly2014@gmail.com, ielhenawy@zu.edu.eg, amansour@zu.edu.eg

**Abstract**

One of the most severe sources of information for a molecular biologist is the gel image generated by using gel electrophoresis during the experiment of issr-pcr, sds-pages, and rapd-pcr. DNA and protein gel images are obtained through the gel electrophoresis separations techniques of DNA and protein fragments. The separation of the polymorphic bands is based on the sizes of the negatively charged DNA fragments running from the negative cathode toward the positive anode. Each gel image has some vertical lanes; each lane corresponds to one sample and has several horizontal bands. The resulting images produced by Gel electrophoresis are sometimes difficult to interpret so it was important to develop software tools to analyze the gel images to help biologists in the process of analyzing gel images as they draw their conclusions according to the results generated from the gel image analyzer software. In this article, we present a survey of some commercial and non-commercial software tools that are used for analyzing gel images. We develop a novel software for processing and analyzing the gel electrophoresis images, computing the molecular weights, saving them as excel sheets, clustering the bands based on their molecular weights using a k-means algorithm, Applying band matching using a tolerance value entered by the user, determine the similarities between samples, drawing the corresponding phylogenetic tree, saving a report of the experiment as a pdf, and printing this report. The novel software will provide the biologist with the ability of manual processing, automatic processing, and semi-automatic processing.

**Keywords:** Gel electrophoresis images; Phylogenetic tree; software tools

## 1. Introduction

Gel Electrophoresis (GE) is an essential technique used in the experiments of molecular biologists which is used for separating the DNA [1, 2]. The separation is done based on their weights in more detail DNA fragments that run from the negative cathode toward the positive anode are separated based on the size of each fragment as the smaller fragments of DNA migrate faster through the gel and occupy the lower position of the gel and the fragments with larger weights will appear on the top of the gel as the larger the size of DNA fragment the less chance of it for passing through the small pores on the gel, for this reason, the DNA fragments with large size move less slowly than other DNA fragments with small size. There are a lot of applications of Gel electrophoresis in the fields of genetics, microbiology, molecular biology, and forensics [3] but the most serious application of gel electrophoresis is molecular typing which is widely used in epidemiology as the goal of GE in the field of epidemiology is to realize from the DNA molecules that extracted from samples more accurately, comparing them with each other or with a standard sample known as a marker, determining the plant pathogenic types and specifying a genotype combined with a specific bacterium. Gel images of "DNA and protein" are produced and obtained through the separation process which is done by using gel electrophoresis. Gel images consist of vertical columns called lanes, in which every lane represents a sample, and horizontal fragments called bands that are sorted in each lane based on their molecular weights [4]. The quality of most generated gel images is low and not good because of some factors such as "the buffer



DOI: 10.5281/zenodo.3572731

chamber temperature, reorientation angle, agarose type, time, field strength, etc…." [5]. Image quality could affect the accuracy of extracting the right information from these images. Thus, the most important step is enhancing the uploaded gel image before doing any process on the gel image.

This research is organized as follows. Section 2 is the related work that presents the current gel image analysis software tools. Section 3 is the Discussion, in this section, the capabilities of our novel software are discussed. Section 4 is the conclusion, in this section, we compare our novel software with other similar non-commercial software and the missing capabilities in our software that will be founded in the next version.

## 2. Related Work

There are a few software that has been developed to analyze the gel electrophoresis image but most of them are commercial and some of them do not achieve all the requirements of the user the free software are very complex for the user and don't give many options. For example, EzQuant, Dolphin 1D, Gel-Pro Analyzer, Intelligent Quantifier, Molecular Imaging, myImage Analysis, Un-Scan-it, Gel-Quant, Image Lab, ImageStudio, LabImage, and Ultraquant are not free and can't generate a phylogenetic tree based on the gel image weights. Moreover ،there is some free software that can't generate dendrograms such as Laneruler, GelAnalyzer, GelQuant, and ImageJ [6].

On the other side, our software is free software that can generate a dendrogram based on molecular weights. There is some software that performs most of their tasks manually such as ImageJ and ClusterVis [7]. Moreover, there is some software do not allow users to add or delete lanes and bands manually such as PyElph [8]. On the other side, our software's tasks are carried out manually, automatically, and semi-automatic. It is more accurate in lane and bands detection than PyElph software which is mainly developed for educational uses and is not accurate in detecting lanes and bands. Another example, GelClust [9] is designed using c-sharp programming language like our software and does all that our software can do but GelClust does not give the user any privileges to detect the molecular weights of the ladder, and does not show him the molecular weights of other unknown bands, does not save weights as excel sheet, does not generate and print reports like our software.

The last program that was developed in Egypt is called Image Analyzer [10] designed using MATLB with not good GUI and its size is so large, on the other hand, our software has been developed and designed using a c-sharp programming language with smart GUI that guides the users from the first step of loading gel image to the last step of generating phylogenetic tree and its size is very small which can be downloaded by anyone and installed on windows operating system ،Unlike Image Analyze.

## 3. Discussion

Our novel software is one of the new software in the field of gel image analyzer software that has been developed using C-sharp programming language under the windows operating system. Its window contains a home page that describes all capabilities of the software, a sidebar with two options "Manual processing and Automatic processing", a header with some icons for "upload gel image, save the image, save experiment to the database, save results to pc as excel sheet or pdf and show old experiments that stored in the database" and two pages for manual and automatic processing. After the user chooses the processing type from the sidebar, he can upload the gel image. If the user chooses Manual processing from the sidebar the manual processing page will be activated and the following steps are followed during the processing of the gel image:

- Uploading the gel image, cropping extra areas from it, and remaining the interesting region which will be processed in the next steps and enhancement process "gray, complement, contrasting, performing gamma correction and some other filters to remove noise" will be executed automatically.
- Choosing lanes detection from the Detect group box and performing the detection using the right mouse click.
- After finishing detection of lanes, correction of them using left mouse-click.
- Choose bands detection from the Detect group box and perform the detection using the right mouse click.



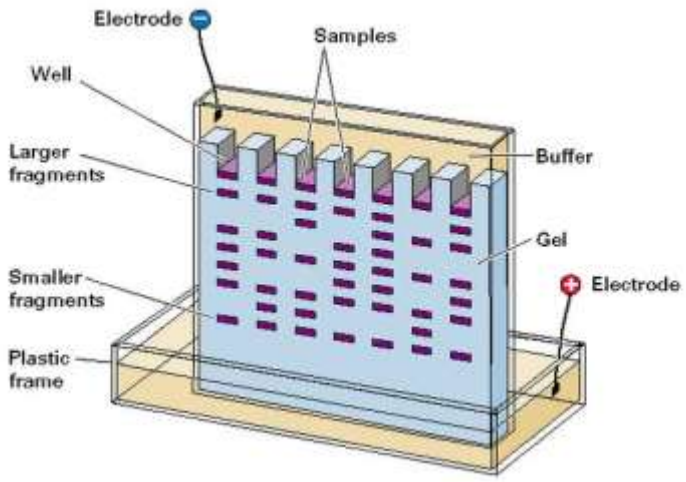
*DOI: 10.5281/zenodo.3572731*

- After finishing detection of Bands, correction of them using mouse left-click.
- Enter molecular weights of ladder or marker.
- Check weights, compute the unknown weights, and display them in the data grid view.
- Generating phylogenetic tree based on extracted weights.
- Saving the molecular weights of unknown bands as an excel sheet and printing report.

But if the user chooses Automatic processing from the sidebar the Automatic processing page will be activated and the following steps are followed during processing the gel image:

- Uploading the gel image, cropping extra areas from it, and remaining the interesting region which will be processed in the next steps and enhancement process "gray, complement, contrasting, performing gamma correction and some other filters to remove noise" will be executed automatically, converting image to binary image and giving the user the control in determining the best threshold value for the image through track bar that has a range from 0 to 255.
- Automatic detection of samples or lanes: this step is done by averaging the Boolean pixel values "0 or 1" of each column of the binary image and drawing over the columns with the value 0 which represents black pixels.
- Manual correction of lanes in case there are extra lanes by using the track bar or using right-mouse-click to add new lanes and left mouse-click to remove lanes if needed.
- Automatic detection of DNA fragments or Bands: this step is done by accessing each lane "sample" row by row and grouping white pixels with 0 value until finding black pixel with 1 value then calculating the center of this group to draw a band in this location, making the same process on all white pixels into each sample "lanes".
- Manual correction of bands in case there is extra or missing bands by using right-mouse-click to add a new band, left mouse-click to remove a band if needed, and removing multiple bands at the same time by click left-mouse and dragging the mouse down.
- Insert the molecular weights of the marker or upload an existing one, check them as they must be inserted in descending order and the user can save the current ladder in his pc for later use.
- Computing the unknown weights, displaying them in the data grid view, and detecting whether the band is a primer dimmer or not.
- Generate a phylogenetic tree based on extracted unknown weights.
- Saving the molecular weights of unknown bands as an excel sheet.
- Saving the experiment results as a pdf which can be printed at any time as a report.

As we said before, the gel image generated from the gel electrophoresis is consisting of some lanes, each lane represents a sample, and horizontal fragments called bands are sorted in each lane based on their molecular weights as shown in figure 1 and figure 2. The workflow that is applied to process the gel images is summarized in figure 3.





*DOI: 10.5281/zenodo.3572731*

Figure 1: Gel electrophoresis

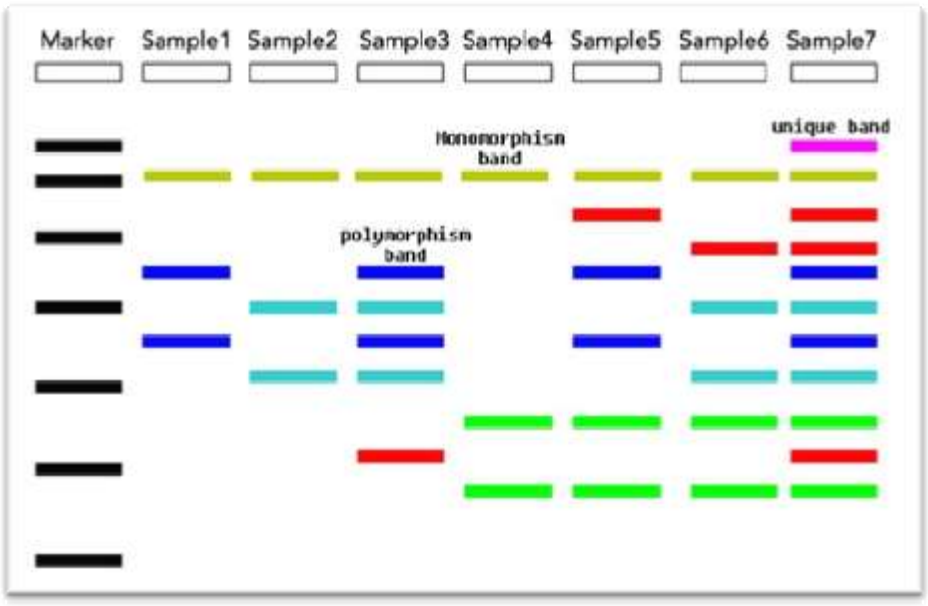



Figure 2: Gel image generated from gel electrophoresis

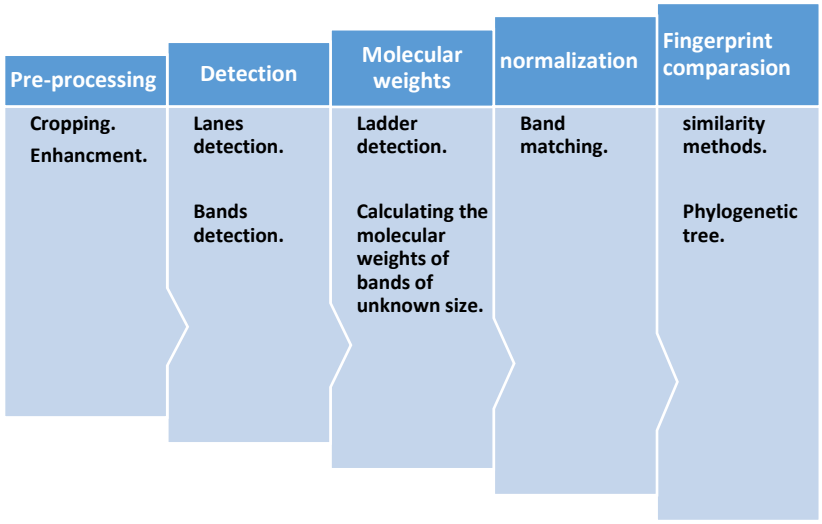



Figure 3: Workflow of analyzing gel images.

The first stage is pre-processing the gel image as we said before the quality of gel image that is produced from the gel electrophoresis during the experiment of issr-pcr, sds-pages and rapd-pcr are poor so the first stage of analyzing gel image is a pre-processing stage. In this stage, the software tool crops the gel image as required and then enhances the gel image to facilitate the analysis process.



*DOI: 10.5281/zenodo.3572731*

In the second stage, the lanes of the gel image are detected, the common idea of these methods is the construction of a 'vertical densitometric-curve' or 'histogram' as shown in figure 4, averaging the pixel values on the same vertical line. In the densitometric curve, the local minima correspond to the gap between the lanes, and this fact is used to detect the lanes of the image.

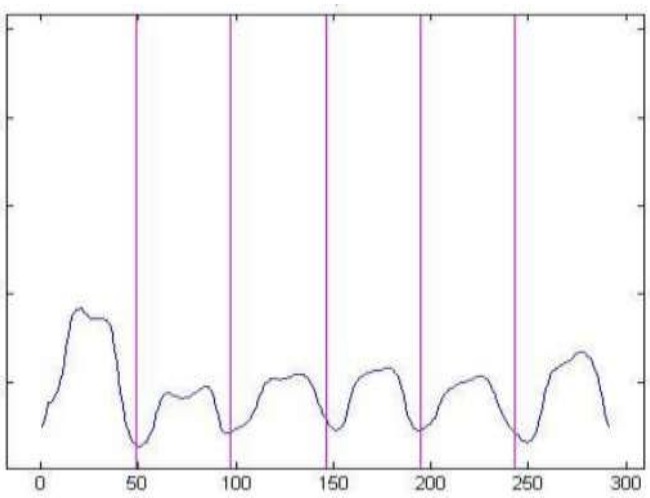



Figure 4: Vertical Densitometric curve of a gel image.

The third stage of the procedure is finding the molecular bands in each lane. The process to locate bands is almost similar to the detection of lanes: a 'horizontal densitometric curve is computed from each lane, and the local maxima in that curve specify the position of the bands as shown in figure 5.

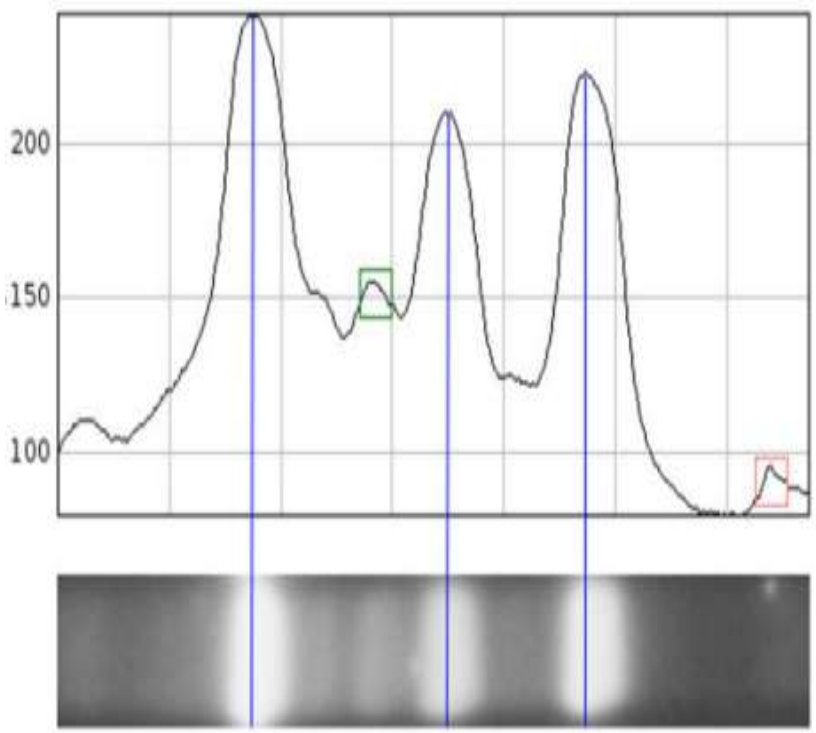



Figure 5: Horizontal densitometric curve from a lane of a gel-image.



*DOI: 10.5281/zenodo.3572731*

The fourth stage is estimating the molecular weights of unknown bands. But first, the molecular weights of the marker are inserted. The molecular weights of the marker are inserted in a descending way as the fragment or band on the top of the gel image has the largest intensity compared to the other bands on the marker. Our software gives the user the option of using a new ladder by inserting the weights in the data grid view or uploading existing molecular weights. The user inserts the molecular weights of the marker into a data grid view then it tests them, if they are not in the correct form, it pops up a message box for the user telling him that he must insert a valid weight with a descending way and then try again. Our software provides the user the ability to save the current inserted weights for later use. After that our software estimates the molecular weights of unknown bands.

The next stage is the normalization phase. Getting the matching bands between lanes, our software implements the band matching algorithm by finding the matching band in each lane to find the type of band. If the band is founded in only one lane, it means that the type of this band is unique and if the band is founded in more than one lane, it means that the type of this band is polymorphism but if the band is founded in all the lanes, it means that the type of this band is a monomorphism. Our software will mark the unique bands by a "unique" string in the gel image. Our software connects the matching bands in lanes by a line drawn in the gel image. Two bands in different lanes are matched even if their weights are not equal but close to each other, the difference between their molecular weights can be determined as a tolerance value. This tolerance value can be fixed or determined by the user, the software provides the user with the ability to determine the tolerance value. This functionality is provided by 16 of the 25 software that supports band matching.

The last step is the comparison of the similarity among the different lanes. Our software calculates the similarity matrix using two methods, by calculating the number of matching bands between the two lanes that are done in the last step using the band matching algorithm and dividing the number of matching bands by the total number of bands in both lanes, this method depends on bands. The second method that our software uses is Euclidean distance.

After that our software draws the phylogenetic tree using two methods, the first one by implementing the UPGMA clustering algorithm. The second method is drawing the tree based on the band matching algorithm that is applied in the last step.

## 4. Conclusion

After we talked more and more about the existing gel image analysis software tools and compared them with our new software which is developed mainly for computing the unknown bands of gel images, clustering the samples, drawing a phylogenetic tree, and printing a report. Based on the results we found that our software is superior to much current gel image analyzer software. Students who work in the field of genetics and molecular biology and Researchers can use it. It is free software with a smart GUI that guides the user, it has a real-time adjustment by using a track bar to manual corrections, and it has a small size that can be downloaded easily and installed on windows OS. But it still has some missing capabilities in this first version which will be founded in the next version of our software. The missing capabilities which could be added in the next versions are:

(1) More clustering algorithms.
(2) Different algorithms for image processing.
(3) Database for saving all the experiments and displaying them at any time as needed.
(4) Mobile application version.



*DOI: 10.5281/zenodo.3572731*

Table 1: Comparing our novel software with other similar non-commercial software

| Software name | Capability code | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **c1** | **c2** | **c3** | **c4** | **c5** | **c6** | **c7** | **c8** | **c9** | **c10** | **c11** |
| **Our software** | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | × | ✓ | ✓ |
| **GelClust** | ✓ | ✓ | × | × | × | × | × | ✓ | ✓ | × | × |
| **GelAnalyzer** | ✓ | × | × | ✓ | × | × | × | × | × | × | ✓ |
| **PyElph** | × | × | × | × | ✓ | × | × | ✓ | ✓ | × | ✓ |

**c1:- Accuracy of column detection**
**c2:- Accuracy of band detection**
**c3:- Insert names of the samples by the user**
**c4:- Display the molecular weights of bands of unknown size to the user and save them as a spreadsheet.**
**c5:- Band matching algorithm**
**c6:- Detect the number of primer dimmer bands in each lane**
**c7:- Clustering bands using the k-means algorithm**
**c8:- Fingerprint comparison**
**c9:- Variety in clustering methods & similarity coefficient**
**c10:- Save a report as a pdf & Print a report**
**c11:- Save molecular weights of the marker as a .txt file for later use and export an existing ladder**

Furthermore, the main specific capability of our software that is embedded only in this software is grouping the bands based on their molecular weights and labeling each band with its group number in the gel image using the k-means algorithm.

## References

[1] Kaabouch, N., Schultz, R. R., Milavetz, B., & Balakrishnan, L. (2007, May). An analysis system for DNA gel electrophoresis images based on automatic thresholding an enhancement. In 2007 IEEE International Conference on Electro/Information Technology (pp. 26-31). IEEE.

[2] Kaufmann, M. E., & Pitt, T. L. (2018). Pulsed-field gel electrophoresis of bacterial DNA. In Methods in practical laboratory bacteriology (pp. 83-92). CRC Press.

[3] Heras, J., Domínguez, C., Mata, E., Pascual, V., Lozano, C., Torres, C., & Zarazaga, M. (2015). GelJ–a tool for analyzing DNA fingerprint gel images. BMC bioinformatics, 16(1), 270.

[4] Lin, C. Y., Ching, Y. T., & Yang, Y. L. (2007). Automatic method to compare the lanes in gel electrophoresis images. IEEE Transactions on information technology in biomedicine, 11(2), 179-189.

[5] Murchan, S., Kaufmann, M. E., Deplano, A., de Ryck, R., Struelens, M., Zinn, C. E., ... & Cuny, C. (2003). Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant Staphylococcus aureus: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. Journal of clinical microbiology, 41(4), 1574-1585.

[6] Heras, J., Domínguez, C., Mata, E., Pascual, V., Lozano, C., Torres, C., & Zarazaga, M. (2015). A survey of tools for analysing DNA fingerprints. Briefings in bioinformatics, 17(6), 903-911.

[7] Ferreira, T., & Rasband, W. (2012). ImageJ user guide. ImageJ/Fiji, 1, 155-161.

[8] Pavel, A. B., & Vasile, C. I. (2012). PyElph-a software tool for gel images analysis and phylogenetics. BMC bioinformatics, 13(1), 9.

[9] Khakabimamaghani, S., Najafi, A., Ranjbar, R., & Raam, M. (2013). GelClust: a software tool for gel electrophoresis images analysis and dendrogram generation. Computer methods and programs in biomedicine, 111(2), 512-518.



*DOI: 10.5281/zenodo.3572731*

[10] Alawdi, R. M., Amer, R. B., Alzohairy, A. M., & Khedr, W. M. The Computational Techniques Developed to Analyze DNA Gel Images. International Journal of Advanced Engineering Research and Science, 3(7).



*DOI: 10.5281/zenodo.3572731*