The Pharma Innovation Journal 2023; SP-12(10): 1622-1629

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277-7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2023; SP-12(10): 1622-1629

www.thepharmajournal.com
Received: 03-07-2023
Accepted: 10-09-2023

Manish Chandra Mehta
Research Scholar, Department of Entomology and Agricultural Zoology, Institute of Agricultural Sciences, BHU, Varanasi, Uttar Pradesh, India

M Raghuraman
Professor, Department of Entomology and Agricultural Zoology, Institute of Agricultural Sciences, BHU, Varanasi, Uttar Pradesh, India

Kirti Srivastava
Young Professional-II, Genome Sequencing lab, ICAR-National Bureau of Agriculturally Important microorganisms (NBAIM), Maunath Bhanjan, Uttar Pradesh, India

Corresponding Author:
Manish Chandra Mehta
Research Scholar, Department of Entomology and Agricultural Zoology, Institute of Agricultural Sciences, BHU, Varanasi, Uttar Pradesh, India

Student friendly approach of molecular data analysis for the evaluation of genetic diversity

Manish Chandra Mehta, M Raghuraman and Kirti Srivastava

Abstract
DNA fingerprinting is an acclaimed molecular technique for establishing a biological relatedness between genotypes by meticulous examination and comparison of their DNA pattern. This technique is engrossed into the polymorphic regions of the DNA, usually the minisatellites and microsatellites. It finds numerous applications in the field of medicine, agriculture, forensics and in parent confirmatory tests. The present study describes user friendly approach to examine the DNA fingerprinting image obtained through gel electrophoresis. A triumvirate approach including Adobe® photoshop® (version 23.2.2), GelAnalyzer 19.1 and NTSYS-pc 2.02e were undertaken to identify the genetic relatedness among the populations of rice yellow stem borer, *Scirpophaga incertulas*. Specimens of rice yellow stem borer were taken from 27 locations in India and a genetic diversity study was carried out across these populations.

Keywords: Genetic, DNA fingerprinting, diversity, molecular technique

Introduction
Biological data have a colossal magnitude and are high throughput in nature thus, to make it perspicuous, involvement of a multidisciplinary approach is indispensable. Bioinformatics provides a platform where smorgasbord of fields such as biology, physics, mathematics, statistics, medicine, and computational technology are integrated together to unveil the mystery of life sciences. Among many applications of DNA fingerprinting, the analysis of genomic relatedness is highly eminent for the investigation of genetic diversity. The genomic relatedness between any individual is identified through gamut of DNA finger printing techniques such as ribotyping, plasmid fingerprinting, pulse-field gel electrophoresis, repetitive sequence-based PCR, ERIC PCR, PCR agarose gel electrophoresis, Rt-PCR, AFLP, RAPD and RFLP. All the above-mentioned techniques have a common desideratum of fragmenting the biomolecules (here nucleic acid) based on molecular weight and charge using electric field (Adamson and Reynolds 1997) [1] and later the analysis of banding pattern in agarose and/or polyacrylamide electrophoresis gels is covered (Smith *et al*. 1989; Kirkpatrick *et al*. 1993) [13, 8].
The gel is observed under the UV gel docking unit and the image is captured cum stored in tiff jpeg, png, or bmp format to be processed later through various software. Most of the software are either paid or comes with sophisticated and expensive imagers provided by well-equipped laboratory (Table 1), which is either unobtainable or inaccessible to the students. Under such circumstances the students must rely on freely available software such as Dolphin1D (www.wealtec.com/products/imaging/software/dolphin-1d-software.htm), Gel Analyzer (http://www.gelanalyzer.com/), and GelClust (Khakabimamaghani *et al*. 2013) [6] which can produce effective results.
The current article discusses the methodology of preprocessing and analyzing the gel bands using a free software; GelAnalyzer version 19.1 and an inexpensive photo editing tool; Adobe photoshop version 23.2.2. Gel analysis includes the detection of lanes and bands, molecular weight analysis, comparing banding pattern and formation of similarity or dissimilarity matrix for the construction of dendrogram and the preprocessing includes enhancing the image quality, inverting colors, gamma correction, flipping, rotating, cropping and size adjustments of the elements of gel images. The flow chart for the gel image analysis is depicted in figure 1.

Gel Analyzer
It is a fully automated tool which offers a feature of lane detection through construction of densitometric-curve on a vertical line.

The local minima of this curve corresponds to a gap between bands and the local maxima is correlated with the presence of band. The vertical densitometric curve helps in lane detection while the horizontal densitometric curve ensures the detection of bands (Bailey and Christie 1994; Pavel and Vasile 2012) [11]. Normalization is achieved through molecular weight marker which is used as a reference lane to study the banding pattern within the gel (Vauterin and Vauterin 2006) [21].

Adobe Photoshop

It is an image editing tool used for graphic designing and digital art. It is not a fundamental software for the analysis of molecular biology, but it serves the purpose of gel image pre-processing, gel alignment and scoring. It is inexpensive, reliable, and user-friendly editing software to interpret the banding patterns by visual observation using guides, grids, marquee and cropping tools available along with the package.

NTSYS-pc (Numerical taxonomy System for PC)

This software is used to identify pattern and structure in multivariate data. This software uses algorithm such as UPGMA or Neighbour-joining to construct dendrograms that facilitates evolutionary study of different populations or a species. The input data for NTSYS-pc can be descriptive information about collections of objects or directly measured similarities or dissimilarities between all pairs of objects. The objects that are used depend highly on the type of applications such as morphological characters, abundances of species, presence, or absence of properties, *etc*. it allows the transformation of data, estimation of similarities or dissimilarities among objects, and preparation of summaries of the relationships using cluster analysis, ordination, multiple factors analyses and principal component analysis. The output is presented in tabular or graphical form. The resulted graph or tables can be edited through plot options which allows customization of fonts, sizes, colours, scales, line widths, background colours, margins and much more. The similarity or dissimilarity matrix obtained can be read from excel XLS and CSV files and the phylogenetic tree can be read from one type of nexus files.

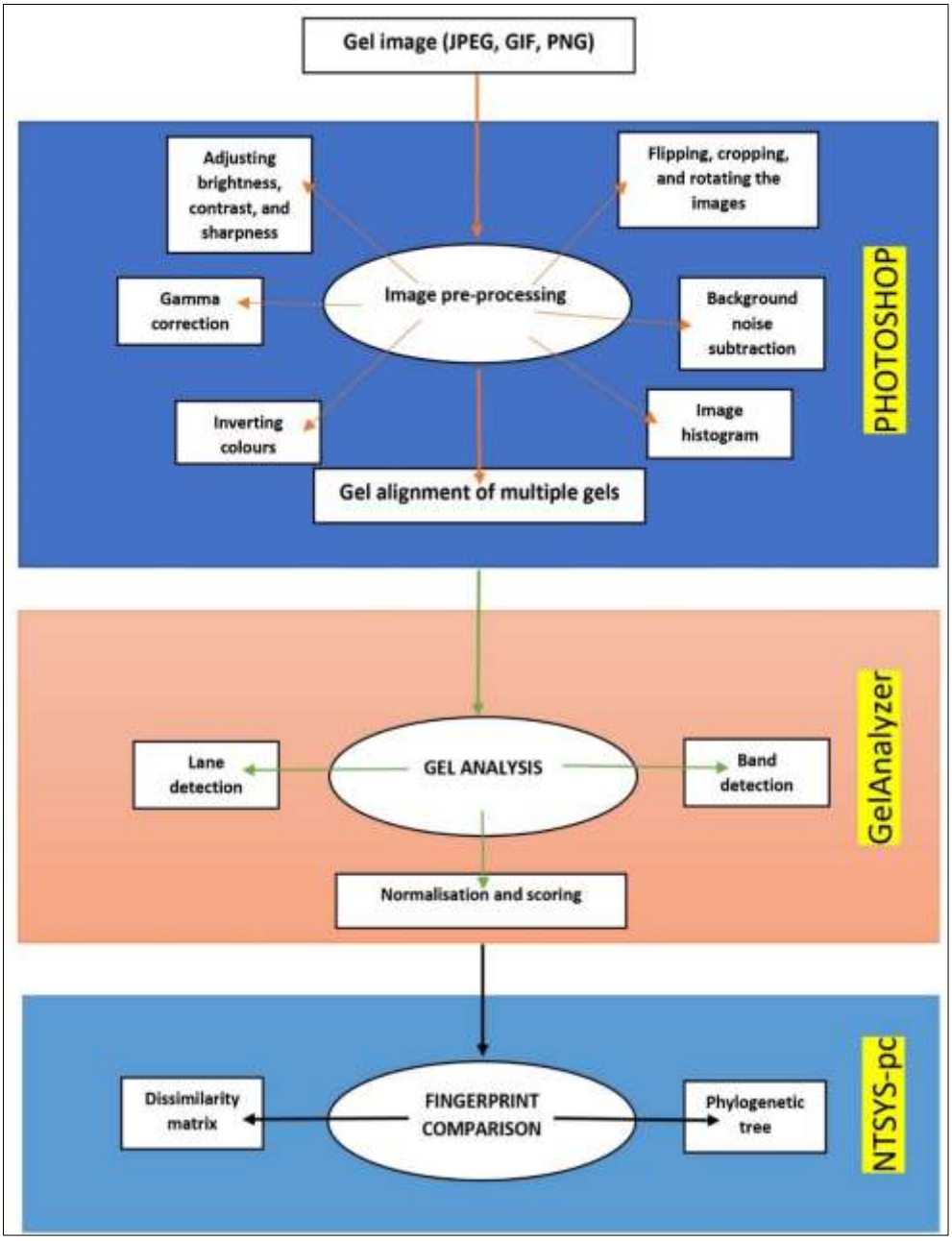


Fig 1: Flowchart for the analysis of DNA fingerprinting images.

Table 1: List of the commercial software used for DNA fingerprinting analysis

Sl. No.	Paid software	References
1	Advanced quantifier	www.bioimage.net/AQ.html
2	EZ Quant-Gel	http://www.ezquant.com/en/products/ezquant-gel/
3	Gel Compar II	http://www.applied-maths.com/
4	Gel Quest	http://www.sequentix.de/gelquest/inde.php)
5	Gel Scan	http://www.bioscitec.com/produkte-bioscitec/software/gelscan-6-0/
6	Image	http://www.sanger.ac.uk/resources/software/image
7	Intelligent Quntifier	http://bioimage.net/IQ.htm
8	Lane Ruler	Wong *et al*. 2010
9	Molecular bioimaging	www.molecularbioimaging.com/software.html
10	My Image Analysis	http://www.piercenet.com/product/myimageanalysis-software
11	PyElph	Pavel and Vasile 2012 [11]
12	Bio Doc Analyze	http://www.biometra.de/1163.0html
13	Gene Tools associated with Syngene systems	https://www.syngene.com/software/genetools-automatic-image-analysis/
14	GelQuant	http://www.ampl.comau/gelquant_home.htm
15	LabImage	http://www.kapelan-bioimaging.com/
16	ImageLab	http://www.sanger.ac.uk/resources/software/image
17	ImageStudio	http://www.licor.com/bio/products/software/image_studio/index.html
18	BioDocAnalyze	http://www.biometra.de/1163.0html
19	Quantity one	http://www.bio-rad.com/en-us/product/quantity-one-1-d-analysis-software

Material and Methods

The present research demonstrates analysis of the gel images obtained from the amplification product of rice yellow stem borer larva using EST-SSR marker for 27 locations of India. 20 primer pair were used to flank the SSR sequence of the extracted genomic DNA for its amplification using the technique of polymerase chain reaction. The amplification product was run in agarose gel electrophoresis with a standardized condition of 79 V for 2 hours for all the gels and the image was captured by Omega Flour™ Gel Documentation System (365 nm).

Image pre-processing

Noise reduction

Noise is a distortion in an image which is impacted by a random variation in image intensity and seen as grains. Noise is produced at the time of capturing image (Verma and Ali 2013) [10] and is influenced by external factors (Azzeh *et al*. 2018) [2]. Basically, when the pixels in the image show variable intensity values in place of the true pixel value then such condition is called noise. Many editing software provide algorithmic filters such as Standard Median Filter, Adaptive Median Filter (Puig and Angel 2001) [3], Decision Based Algorithm (Vasanth *et al*. 2012) [7], Progressive switching median filter (Beers and Kleijenen 2003) [14] and detailed preserving filter (Pratt 2007) [6], which reduces the noise by smoothening the image at the cost of the contrast details of the image. During acquisition of gel image, we usually encounter salt and pepper noise which reduces its quality and visual effects. To denoise such gel images for DNA fingerprinting analysis Adobe photoshop provides a filter option as shown in the figure 1.

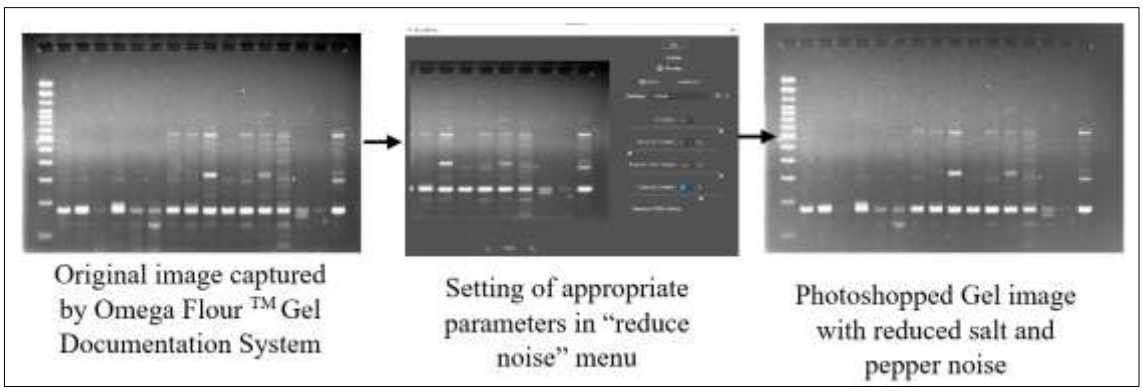
Case study 1- stary gel

The gel image that has either black or white noise pixels in the form of granular background that may be called as stary gel.

Reason: The reason for this noise could be a defective camera sensor, software failure, or hardware failure or transmission (Alazilan *et al*. 2004) [11], addition of EtBr to the molten Agarose when it is still hot, or due to presence of bubbles.

Correction

Method 1: File → Open → import image from system → Filter → Noise → Reduce noise



Method 2: File → Open → import image from system → Filter → Sharpen → Smart sharpen

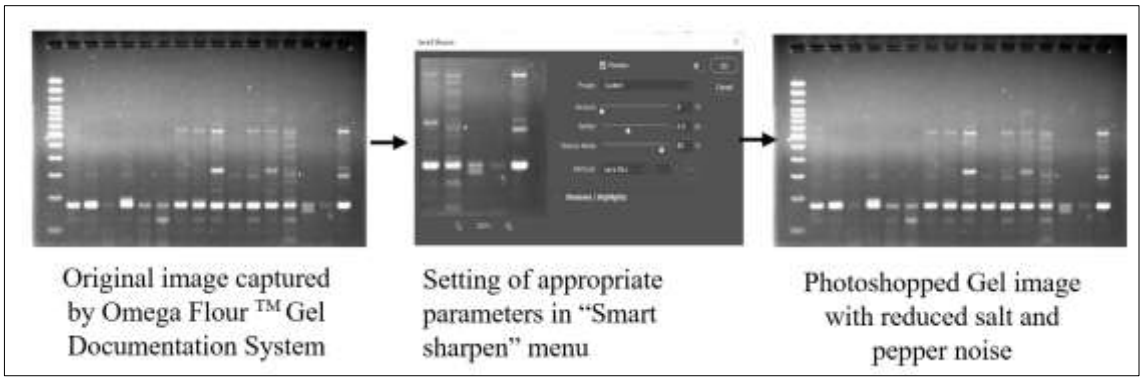


Fig 1: Step wise approach for reduction or removal of salt and pepper noise form gel image

Flipping, cropping, and rotating
case study 2- slanting DNA bands
As the DNA bands migrate towards the cathode, it fails to travel in a straight line but starts slanting which makes the lane curved or crooked affecting the scoring.

Reason: wrong placement of comb on the gel, uneven distribution of current in the buffer, repetitive use of TAE or TBE buffer, or high voltage application.

Correction
Method: File → open → import image from the system → crop tool (⌗)→ flip to straighten (Fig. 2).

Inverting image
Case study 3- inverting image
Inverting colour of an image means to produce its negative image by turning it into opposite colour.

Reason: inverted gel images are more fastidious and reveals many hidden details of the gel which facilitates manual scoring (McDaniel 2002) [9].

Method: File → open → import image from the system→ image→ adjustments→ Invert (Fig. 3).

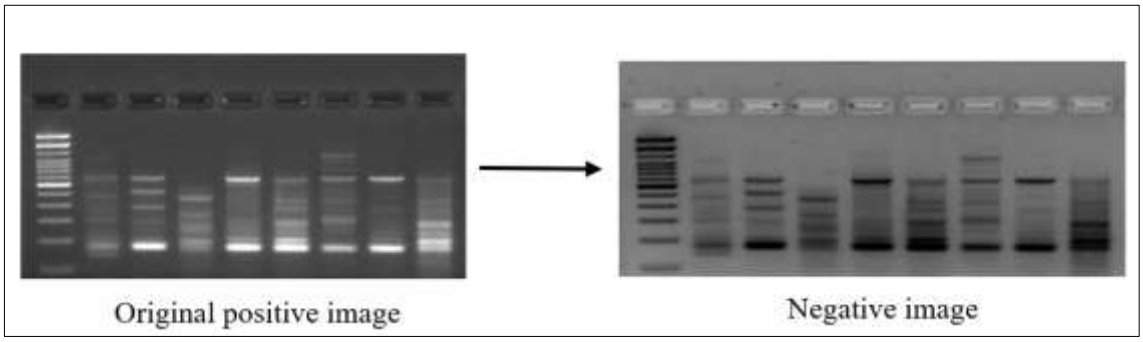


Fig 3: Step wise approach for inverting the gel image

Gel alignment
While handling many samples from various locations for the analysis of genetic diversity multiple gels were ran for the primer YSB_SSR_30F and YSB_SSR_30R, in electrophoresis unit. With multiple gels it was difficult to compare the banding pattern and hence the two gels were aligned together using photoshop. Images of the two gels were imported in the workspace of photoshop. Height of the DNA ladder was measured with marquee tool for the first gel image (base image) and under the transform option (Dimensions) of pixel section, the measured height was entered for all the images and was standardized. Once all the gel images are standardized, a new blank document was opened (ctrl + N) and the gel images were dragged/copy pasted into the blank document to get a perfectly aligned single gel image.

Method 1:
File → open → import images from the system→ marquee (tool menu)→ drag the marquee tool over the DNA ladder→ get the height value→ standardize the images with same height→ ctrl+N→ copy and paste the images into blank document→ save a copy (jpeg) (Fig. 4).

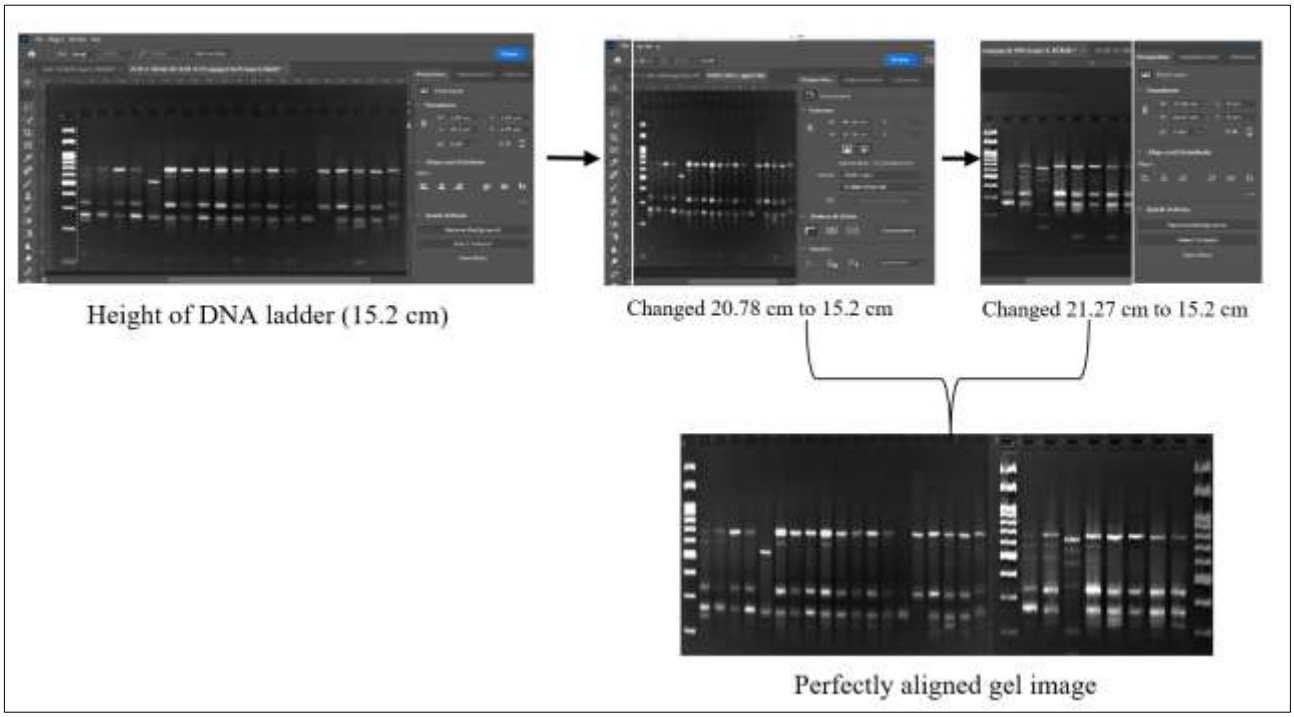


Fig 4: Step wise method of gel alignment for multiple gels

Method 2: File → open → import images from the system→ marquee (tool menu)→ drag the marquee tool over the DNA ladder of base image and second gel image→ get the height value of both →$\frac{\text{Height (H) of the base image}}{\text{Height (H)}of\ second\ image} X\ 100$→ Image→ Image size→Pixel dimensions→choose percent→input the calculated percent value of height→ Ok→ standardize both images with same pixel percent→ Ctrl+N→ copy and paste the images into blank document→ save a copy (jpeg) (McDaniel 2002) [9].

Gel analysis through GelAnalyzer

lane detection

File→New analysis→ choose document to open→ light on dark for positive image/dark on light on inverted image→crop→ Lane mode→Detect lanes/Add a new lane manually (Fig. 5).

Band detection and normalization.

Band mode→Detect bands/Add band manually→Molecular mode calibration mode→ enter the size of each band in reference lane/load MW standard on selected lane (Fig. 5).

Fig 5: Lane and band detection along with the densitometric curve from the lane DNA ladder

Scoring

The presence or absence of band for each lane was entered as '1' or '0' into the spreadsheet respectively. The binary data was entered, taking MW as a reference in a descending order i.e, biggest size band on the top while smallest band at the bottom. For example, all the bands for 700bp size were scored as '1' for lane 12, 13, 14, 17, 18, 19, 20, 23, while scored as '0' for the rest of the lanes (Fig. 6a). Scoring was also done via adobe photoshop using guides (Fig. 6b).

Method of band scoring via photoshop

Method File→Open→import pre-processed image → ctrl+R (ruler) → View → Show → Guide → guide was dragged from ruler to the centre of all the bands of DNA ladder and also between them (guide was either dragged or positioned (View→New guide→Horizontal→Position (in cm)) → a guide was drawn across the centre of all the bands→ visual interpretation of an approximate MW of the bands was drawn using DNA ladder as reference (Fig. 6b)

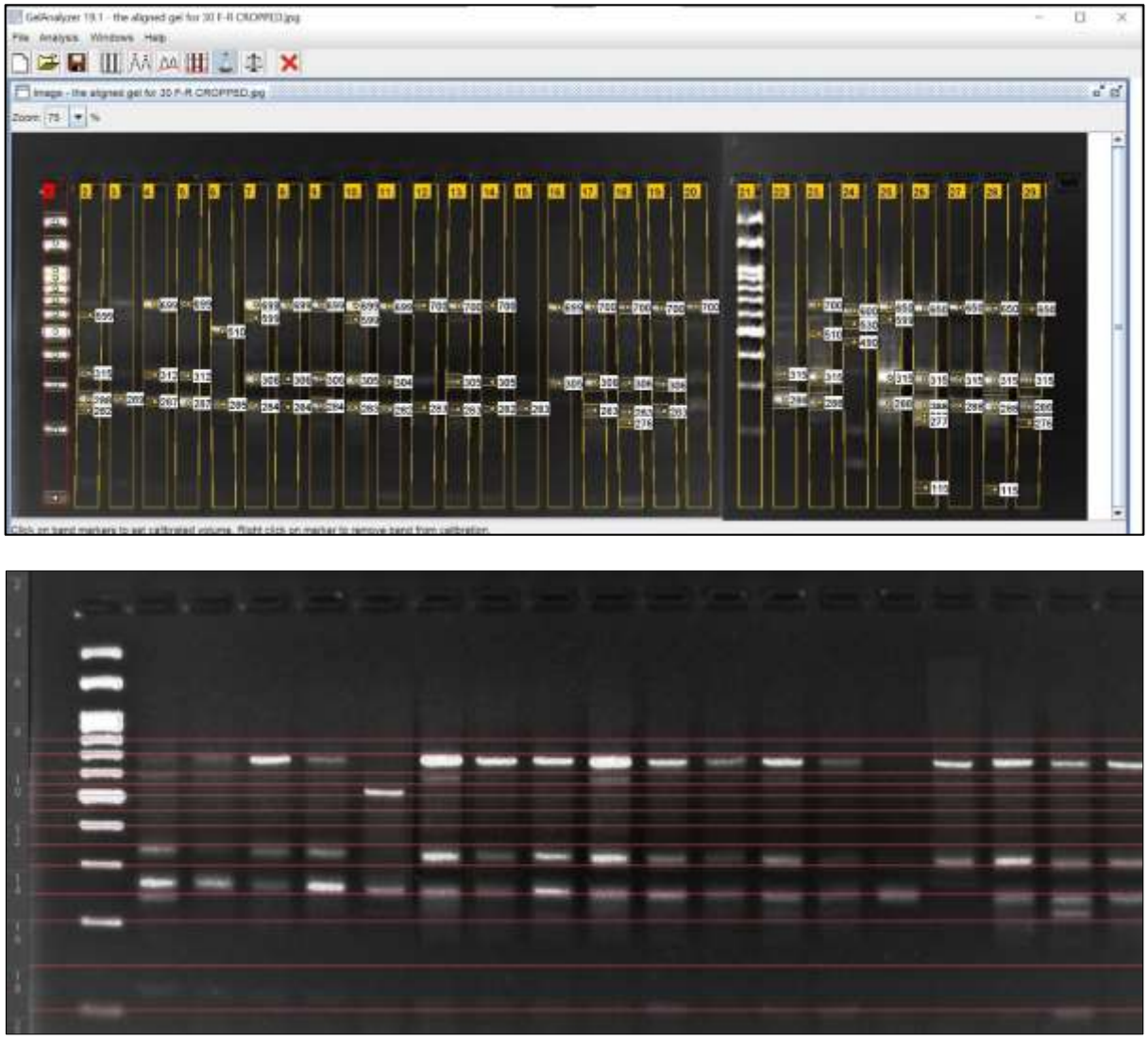


Fig 6: Step wise approach for scoring the bands via (a) GelAnalyzer and (b) Photoshop

Fingerprint comparison using NTSYS-pc

The binary data fed in Microsoft excel 97 ~ 2010 was used for fingerprint comparison. The hierarchical clustering algorithm such as UPGMA, and neighbor joining was used for the construction of dendrogram (Rohlf 1998) [4].

Molecular data set preparation and entry

Each primer's binary data was entered into different data sheet of the same file, in a manner such that A1 cell should show "1", B1 cell must have the total number of row (isolates, varieties, races, genotypes etc.), and C1 cell must have the total number of column (locus or locations). This excel sheet (.xls) is imported into NTedit program through "File→Import Excel→Using OLE". This file was saved in a specified location using "File→Save file (.NTS)"

Cluster analysis

The cluster analysis was accomplished using NTSYSpc 2.02e program in the following manner

Similarity matrix

Open NTSYS-pc application→ Similarity→SimQual (computes various association coefficients for quantitative data)→Input file (the. NTS file previously saved)→Choose desired coefficient method (Simple matching and Jaccard are most used)→Specify the output path →Compute. The above process will produce the similarity matrix.

Dendrogram construction

Open NTSYS-pc application→Clustering→SAHN (computes Sequential, agglomerative, hierarchical, and nested clustering)→Input file (the previously saved file as output

path in similarity matrix computation)→Output file (Specify the output path)→Clustering method (UPGMA)→Maximum no. tied trees was set to 100→ Tie tolerance and Beta was set as default→Compute.

Result and Discussion

GelAnalyzer is the best free automatic tool which conveniently and effortlessly detect the lanes, bands, and the standard MW of the DNA ladder given that the acquired images are of high quality, noise free, with straight lanes, and well differentiated bands of both ladder and the sample DNA. Heras *et al*. (2016) [5] also reported the efficiency of GelAnalyzer. Most often the students fail to get such perfect gel images, and in such case, they can easily use an inexpensive handy image editing tool, Adobe photoshop to make their gel image perfect for scoring (case study 1, 2, and 3). Similar observations were made by McDaniel (2002) [9].

GelAnalyzer offered 95% accuracy for automatic lane detection, 100% accuracy for automatic band detection and 75% accuracy for MW calibration of the bands (Table 2). Since the software showed moderately low efficiency for MW calibration, Photoshop® was used for cross verification.

Table 2: GelAnalyzer offered 95% accuracy for automatic lane detection, 100% accuracy for automatic band detection and 75% accuracy for MW calibration of the bands

Sl. No.		Total number	Accuracy (%)
1	Gel images analyzed	20	
2	Images with correct lane detection	19	95%
3.	Images with correct band detection	20	100%
4	Images with correct MW calibration of bands	15	75%

The combination of this three-software resulted in a phylogenetic tree (Fig. 7) from which an inference of the genetic relatedness among different population can be deciphered.

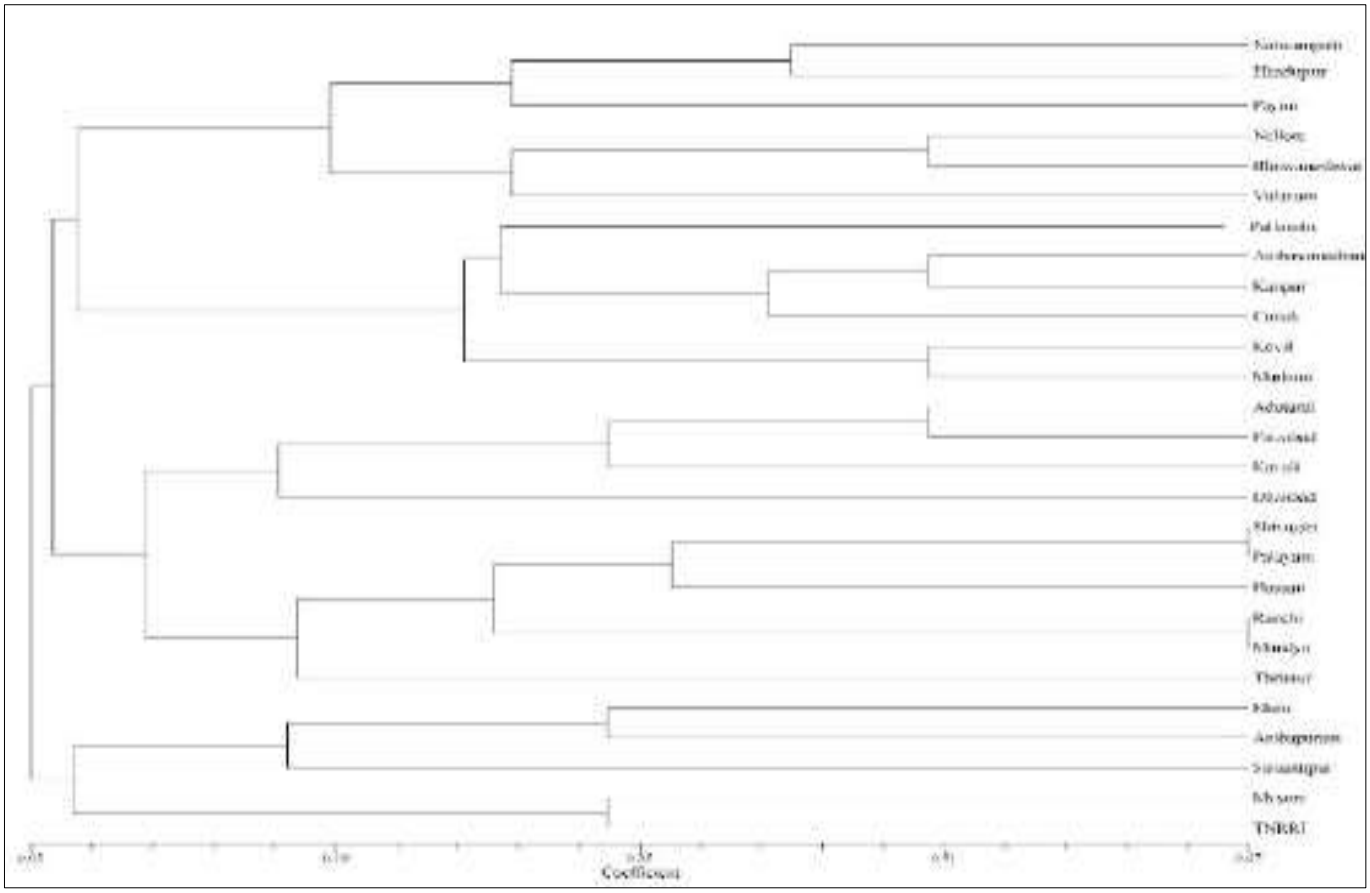


Fig 7: Phylogenetic tree obtained from 27 populations of yellow stem borer, *Scirpophaga incertulas*

Conclusion

Indubitably, commercial software provides very useful and striking functionality, but this does not rule out the worth of free software. Comparatively, free software may provide less functionality or accuracy of the result, but they are equally reliable. For students, working on small molecular projects with moderate amount of biological data can easily utilize these software for their analysis and evaluation of genetic studies.

References

1. Adamson NJ, Reynolds EC. Rules relating electrophoretic mobility, charge and molecular size of peptides and proteins. Journal of Chromatography B Biomed. 1997;699(1-2):133-147.
2. Alajlan N, Kamel M, Jernigan E. Detail Preserving impulse noise removal. International journal on Signal processing: image communication. 2004;19(10):993-1003.
3. Azzeha JA, Zahran B, Alqadi Z. Salt, and pepper noise: Effects and removal. International Journal on Informatics Visualization. 2018;2(4):252-256.
4. Celebi EM, Hassan A, Kingravi, Y. Aslandogan Nonlinear vector filtering for impulsive noise removal from color images. Journal of Electronic Imaging. 2007;57(16):1-21.

5. Heras J, Dominguez C, Mata E, Pascual V, Lozanon C, Torres C, *et al*. A survey of tools for analysing DNA fingerprints. Briefings in Bioinformatics. 2016;17(6):903-911.
6. Khakabimamaghani S, Najafi A, Ranjibar R, Raam M. GelClust: A software tool for gel electrophoresis images analysis and dendrogram generation. Computer methods and programs in biomedicine. 2013;111(2):512-518.
7. Kirkpatrick FH, Dumais MM, White HW, Guiseley KB. Influence of the agarose matrix in pulsed-field electrophoresis. Electrophoresis 1993;14(1):349-354.
8. Labyed Y, Kaabouch N, Schultz RR, Singh BB. Gel electrophoresis image segmentation and band detection based on the derivative of the standard deviation. Artificial intelligence and pattern recognition; c2007. p. 31-35.
9. McDaniel J, Pillai SD. Gel alignment and band scoring for DNA fingerprinting using adobe photoshop. 2002;32(1):120-122.
10. Parminder K, Jagroop S. A Study Effect of Gaussian Noise on PSNR Value for Digital Images International journal of computer and electrical engineering. 2011;3(2):1793-8163.
11. Pavel AB, Vasile CL. PyElph: A software tool for gel images analysis and phylogenetics. BMC Bioinfomatics. 2012;13(1):1-6.
12. Pratt K. Digital Image Processing, Fourth Edition, John Wiley & Sons, Inc., Publication; c2007.
13. Puig D, Angel García M. Determining optimal window size for texture feature extraction methods. IX Spanish Symposium on Pattern Recognition and Image Analysis, Castellon, Spain. 2001;2:237-242.
14. Rohit Verma and Ali J. A comparative study of various types of image noise and efficient noise removal techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2013;3(10):617-622.
15. Rohlf FJ. NTSYSpc Numerical Taxonomy and Multivariate Analysis System Version 2.0 User Guide. Applied Biostatistics Inc., Setauket, New York; c1998. p. 37.
16. Smith SB, Aldridge PK, Callis JB. Observation of individual DNA molecules undergoing gel electrophoresis. Science. 1989; 243(4888):203-206.
17. Beers VWCM, Kleijnen JPC. Kriging for interpolation in random simulation. Journal of the Operational Research Society. 2003;54:255-262.
18. Vasanth K, Karthik S, Divakaran S. Removal of Salt & Pepper Noise using Unsymmetrical Trimmed Variants as Detector. European Journal of Scientific Research. 2012;70(3):468-478.
19. Smith SB, Aldridge PK, Callis JB. Observation of individual DNA molecules undergoing gel electrophoresis. Science. 1989;243(4888):203-206.
20. Beers VWCM, Kleijnen JPC. Kriging for interpolation in random simulation. Journal of the Operational Research Society. 2003;54:255-262.
21. Vauterin L, Vauterin P. Molecular identification, systematics, and population structure of prokaryotes. Springer, Berlin; c2006. p. 141-217.