Full Length Article

FAILURE IN GENERATING DATA FILE DURING AXIOM™ MICROARRAY GENOTYPING: IDENTIFICATION OF CAUSES AND PROBABLE SOLUTIONS

Tejas Gohil[1]*, A. Sudhakar[2], Bhumika Parmar[3], Rutvi Rajpara[4] and Hardik Poojara[5]

National Dairy Development Board
Anand, Gujarat - 388 001
India

ABSTRACT

Genotyping of various biological samples using AxiomTM microarray genotyping technology (Thermo Fisher Scientific) generates five types of data files (.CEL, .ARR, .DAT, .JPG, and .AUDIT), among which the .CEL files are essential for further downstream data analysis. In our study, involving 63 microarray genotyping plates, we observed that the number of output data files occasionally did not match the number of samples processed, in either the 96F or 384F workflows. Although such discrepancies were rare, but its identification was crucial, considering the limited availability of sample quantities. A clear understanding of the types of output data files generated and the purpose for their generation was pivotal, which helped us in recognizing the underlying cause for failure in data file generation with .CEL extension, specifically. Broadly, the errors were categorized in to 4 types. Type one error was due to peg scan failure in Gene TitanTM multi-channel instrument's scanner, whereas second type was due to unknowingly dispensing the hyb-ready DNA or staining reagents into incorrect wells in a reagent tray. The third type of error arose due to improper clamping of the array plate with the hybridization plate. The fourth type resulted because the ligation enzyme was not included during the ligation master mix preparation. Although the majority errors may be categorized as human factors, intensive practice, rigorous training, meticulous observations and process monitoring can substantially reduce their occurrence in high-throughput genotyping laboratories.

Keywords: AxiomTM analysis suite, CEL file, error, Gene TitanTM multichannel instrument, SNP microarray genotyping

Received : 22.07.2025 Revised : 23.09.2025 Accepted : 29.10.2025

[1]Officer, Sabarmati Ashram Gaushala's Genomics Lab, D-Lab
*Corresponding Author Email : tejasgohil1990@gmail.com
[2]Manager
[3]Technician, Sabarmati Ashram Gaushala's Genomics Lab, D-Lab
[4]Junior Analyst
[5]Technician

INTRODUCTION

Single nucleotide polymorphism (SNP) microarray genotyping technology is widely used for genetic research in high-throughput genotyping laboratories (Neumann *et al*. 2021; Bernard *et al*. 2022),

offering cost-effective and efficient analysis of genetic variants. However, it is not immune to errors (Pompanon *et al.* 2005) as its accuracy and reliability depends on multiple factors such as sample processing, quality control measures, and human or instrumental errors etc.

SNP genotypes can be validated by independent SNP genotyping technologies like whole genome sequencing (Sudhakar *et al*. 2021) or PCR based methods like PCR-RFLP. Instrument induced errors are typically reflected in the raw data files generated upon completion of genotyping, which are prerequisites for further data analysis. A major concern is reduction in the number of output data files generated, affecting the percent of passing samples on the array plate. According to "AxiomTM Genotyping Solution Data Analysis User Guide", an array plate is considered high quality only if it achieves a pass rate greater than 95%.

Errors in genotype data generation can lead to null alleles (Callen *et al*. 1993; Paetkau and Strobeck 1995), allelic dropouts and false alleles (Bonin et al. 2004; Taberlet *et al*. 1996). Detection of such errors requires expertise as their origin has a variety of intricate and occasionally mysterious causes (Pompanon *et al*. 2005). However, careful execution of the experimental protocol and consistent examination may help to reduce genotyping error rates to a great extent. The source of the error may vary from human sample handling (Hoffman and Amos, 2005) to reactant or equipment failure.

The present article highlights certain human and equipment related errors encountered while executing AxiomTM microarray genotyping protocols in the laboratory for cattle and buffalo, but remained unnoticed. These errors often led to reduction in the number of output data files, leading to sample failure. Physical inspection of the hyb-ready plate and stain trays, along with visual assessment of the typical patterns of the microarray plate peg in GCC (Gene Chip Command Console), proved valuable in identifying these hidden issues.

METHODS

Protocol followed for genotyping

SNP microarray genotyping was performed using Axiom™ 2.0 assay mini 96-array format and Axiom™ propel xpress 384HT workflows (Thermo Fisher Scientific) based on the number of samples genotyped; 96 and 384 per SNP microarray plate, respectively. A total of around 75,000 DNA samples of various breeds of cattle and buffalo were genotyped during 2022-24 using these workflows. Details regarding sample collection, DNA extraction and quality control procedures followed before microarray genotyping were reported earlier (Sudhakar *et al*., 2023).

Briefly, the genotyping protocol included the following steps: isothermal amplification of genomic DNA, fragmentation, precipitation, resuspension, hybridization, staining, and scanning, as described by the manufacturer (Axiom™

2.0 Assay Mini 96-Array Format, Manual Workflow, User Guide, Publication number : 703434; Axiom™ Propel XPRES 384HT Workflow, User Guide, Publication number : MAN0018487).

In the final step, scanning of the array using the Gene Titan™ Multichannel Instrument (GTMC) produced output data files in proprietary Thermo Fisher Scientific formats, 1) .ARR, called as sample attribute file, 2) .DAT, known as image data files, 3) .JPG files, which are the copy of .DAT files in standard image format and 4) .AUDIT files that contain information on the processing of each physical array by GCC, 5) .CEL files, which contain information about the fluorescent intensities of the probe sets. Out of five, files with .CEL extension were used further for downstream data analysis.

RESULTS AND DISCUSSION

Observations on various types of errors in Genotyping

Error-1 Peg scan failure in GTMC

The most common error was failure of the GTMC scanner to capture data from certain array pegs, resulting in missing .CEL files. Overall, .CEL files were not generated for a total of 276 samples, including18 samples genotyped with Mini-96 (96F) workflowacross12genotyping batches, and 258 samples processed with Axiom™ Propel (384F) work flow across 44 genotyping batches. At the end of each genotyping batch, GCC software displayed the scan results (figure 1). The pegs that were scanned successfully were seen as green coloured squares, whereas the pegs that failed were red coloured and .CEL files were not generated for such samples. Table 1 summarizes potential causes and probable solutions.

Error-2 Misdispensing of hyb-ready DNA or stain reagents in a 384-well plate for a Mini-96 (96F) workflow.

In two separate Mini-96 (96F) genotyping batches, .CEL files were missing for the two samples in each of them. The information for the same is described in table 2.

Error-2 was identified by files with. DAT extension. These files were visualized using “GCC Viewer” of the Gene Chip Command Console Software. A .DAT file is a binary raw image data file, which contain pixel intensity values collected from the Affymetrix scanner, along with the gridding information used during feature extraction. Whereas .CEL files store the results of the intensity data on the pixel values of .DAT file. The comparison of the .DAT file image for a sample with successful .CEL file generation and for a sample with no .CEL file generation with the “GCC Viewer” is shown in figure 2.

Figure 2 shows the real time .DAT image of a scanned peg of an array. In (a), thousands of different coloured spots are seen distributed throughout the image, which represents the light emitted by laser excited fluorescent probes attached to the

oligonucleotides tiled on the peg of the array. Whereas in figure 2 (b), a uniformly dark coloured area is seen without any spots.

Error-3 Improper clamping of the array plate with hybridization plate

In one Mini-96 (96F) genotyping batch, .CEL files were missing for 21 out of 96 samples. Rescanning the array twice ruled out peg scan failure (Error 1), and visual inspection of the plates ruled out reagent misdispensing error (Error 2). The error was mysterious until all the .CEL files were analysed in the Axiom™ Analysis Suite application (AxAs). The plate view function of AxAs revealed a triangle region at the bottom left corner of the array plate (figure 3). This region contained all pegs of 21 samples for which .CEL files were not generated. Tracing this error to its root unveiled the clamping related issue. Clamping is a process that involves stacking the array plate above the hyb-plate on a flat surface of a table and giving equal pressure on 6 different positions until a clicking sound is detected. Table 3 outlines the possible reasons along with suggested solutions for error-3.

Error-4 Preparation of Ligation master mix without AxiomTM Fast Ligation enzyme

In one Axiom™ Propel (384F) genotyping batch, 218 samples failed to produce.CEL files. The array-plate was rescanned in the GTMC instrument assuming the failure was due to error-1. Even after rescanning, there were no .CEL files for the same 218 samples. Analysis of the 166 available .CEL files in AxAs revealed that there were only around 1500 best and recommended markers out of 76000, indicating major assay failure. To investigate the error further, .DAT files were visualized using GCC Viewer as discussed previously in error-2.

All .DAT file images were similar to figure 2. To dive deep into the error, all stain reagent preparation steps were traced back. The total available volume of Axiom™ Fast Ligation enzyme was 2.9 ml, out of which 1.23 ml was required to prepare Ligation master mix to process one 384F array. Usually, two 384F array plates are processed together for better utilization of the reagents using two GTMCs and a lab tracking sheet for experiment traceability. In this batch, two array plates were processed on two consecutive days, due to scheduled maintenance of one of the GTMC, but with one tracking sheet. While processing the second 384F array plate, the technician assumed that enzyme was included in the ligation master mix preparation seeing tick mark in the tracking sheet. After completing genotyping on this array plate, which generated less .CEL files, physical examination of the Axiom™ Fast Ligation enzyme vial revealed more than 1.6 ml of unused enzyme volume, indicating the enzyme was not included during the ligation master mix preparation. The following table 4 lists possible issues and their probable solutions.

It may be appropriate to state that errors do happen in high throughput genomic laboratories. Ideally annual refresher

training sessions should be organized for the laboratory technicians to ensure that all procedures described in the user guide are consistently and accurately followed. In addition, a “Lab Tracking Form” should be maintained for each genotyping batch, documenting details such as reagent volumes for the master mix preparation, reagent lot numbers, array barcodes, and any errors encountered. Furthermore, implementing a double-check system by a dedicated second technician who independently verifies each critical step - such as reagent preparation, plate loading, and checklist completion - can greatly reduce the likelihood of human error and prevent inadvertent oversights. Issues like scanner malfunction or faulty library files, require intervention from the field service engineer of the manufacturer (Thermo Fisher Scientific).

CONCLUSION

Present study investigated multiple causes of failure of.CEL file generation AxiomTM microarray genotyping, including peg scan failures, reagent misdispensing, plate clamping errors, and omission of critical enzymes.Although the occurrence of such events are less frequent, its impact may be highly significant, particularly, while dealing with limited quantities of samples or biopsied samples or irreplaceable samples or samples requiring timely genotype data. By thoroughly understanding error mechanisms and implementing corrective measures, laboratories can greatly reduce costly genotyping failures and improve data reliability.

ACKNOWLEDGEMENTS

The authors express their gratitude to National Dairy Development Board for providing technical support. The contributions of the progeny testing and pedigree selection project officers under RGM, who supplied blood samples for the development of the reference genome population, are sincerely appreciated.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest in the publication.

COMPLIANCE WITH ETHICAL STANDARDS

This study focused on the technical evaluation of errors occurring during SNP microarray genotyping procedures. All samples analyzed in this research were already available through other projects

Table.1. Possible reasons and probable solutions of Error-1

Error-1	Scan failed in GTMC
Possible reasons	1. After completion of staining step, the array plate was stacked on the scan tray filled with hold buffer, within GTMC instrument by the robotic arm, submerging all the pegs in the hold buffer. This might have created air bubbles at certain locations which may interfere with the subsequent scanning process. 2. Sometimes, any particulate matter may get adhered to the bottom surface of the scan tray, interfering with the scanning process.
Probable solution to overcome it	1. The array plate was carefully dislodged from the scan tray by placing them in the plate alignment tool. The array plate was moved upwards in a straight motion and then the scan tray was slightly jiggled. Any air bubbles were burst with a sterile micro-tip and then the array plate was again placed on the scan tray in correct orientation. 2. The bottom transparent surface of the scan tray was cleaned with a kimwipe and placed in GTMC instrument for rescan. By following the above method, 256 .CEL files were generated after rescanning 276 pegs, which were failed in initial scanning. This way 92.75% .CEL file were generated successfully. Remaining 20 pegs were rescanned 3 to 4 times, but did not generate any .CEL files. These 20 pegs did not generate any .DAT files, therefore the error investigation was not possible.

Table.2. Possible reasons and probable solutions for Error-2

Error-1	Dispensing hyb-ready DNA samples or stain reagents into incorrect wells
Possible reasons	Around 35 µl of hyb-ready DNA samples were aspirated from a 96-well PCR plate and then dispensed into a 384-well hyb plate. This transition from 96 to 384-well plate format is recommended by the manufacturer in the user guide. Dispensing 96 number of hyb-ready DNA samples into a 384-well hyb plate required extreme caution as samples were to be dispensed into every alternate wells with a 12-channel micropipette. Even a slight disorientation of a micro-tip might led dispensing samples into the incorrect wells, and remain unnoticed as hyb-ready DNA solution was colourless and sufficient light contrast was not available to detect the same. Similarly, stain reagents were also dispensed only in 96 wells in a 384-well plate. Due to light sensitive nature of stain reagents, they were prepared and dispensed to their respective plates in dark, which again led to similar event as hyb-ready DNA samples dispensing.
Probable solution to overcome it	A thorough inspection can be performed to determine whether any incorrect well loadings exist after dispensing hyb-ready DNA or stain reagents to a 384-well plate.

Table.3. Possible reasons and probable solutions for Error-3

Error-1	Improper clamping of array plate with hyb-plate
Possible reasons	While clamping, if equal pressure was not applied at all 6 positions on the array plate, then the array plate may not get attached properly to the hyb-plate.
Probable solution to overcome it	The array plate and hyb-plate were loaded separately into GTMC instrument and stacking of array plate over hyb-plate was performed by the robotic arm within GTMC instrument. At this point, the clamping must be performed manually, outside the instrument. The array plate-hyb plate combo was placed on a flat surface of the table and all 6 positions were pressed with equal and sufficient pressure to detect the clicking sound. After clamping, the combo was lifted to the eye level and observed for any space between them. If any gap was seen at any position, then it was again pressed at that point by keeping it on a flat surface.

Table.4. Possible reasons and probable solutions for Error-4

Error-1	Ligation Enzyme not included in Ligation master mix
Possible reasons	1. All the stain reagents for this AxiomTM Propel (384F) workflow were prepared to process two 384F array (for a total of 768 samples) and then divided into two equal parts, one part to be used for one AxiomTM Propel (384F) array and the other for the remaining one. 2. According to the manufacturer’s protocol, all the stain reagents can be prepared in advance up to 2 days and can be stored at 4 0C, except the ligation master mix, where the AxiomTM Fast Ligation enzyme must be added within 1 hour before dispensing it into stain tray. As one GTMC can process only one array at a time, only one part of stain reagents can be used during processing the first 384F array. While processing the other 384F array-plate on the next day, all the stain reagents were ready-made except the ligation master mix, in which the ligation enzyme needs to be added just before dispensing it in the stain plate and failure of the same resulted in the faulty .CEL files.
Probable solution to overcome it	All the chemicals that are to be used during the workflow can be listed on a genotyping batch tracking form, with check boxes next to each one. These check boxes can be marked to confirm that the stain reagents have been added step by step in accordance with their composition, avoiding any mistakes from being made. A dedicated second technician should be assigned for a genotyping batch, who can verify each crucial steps of the experimental procedures.

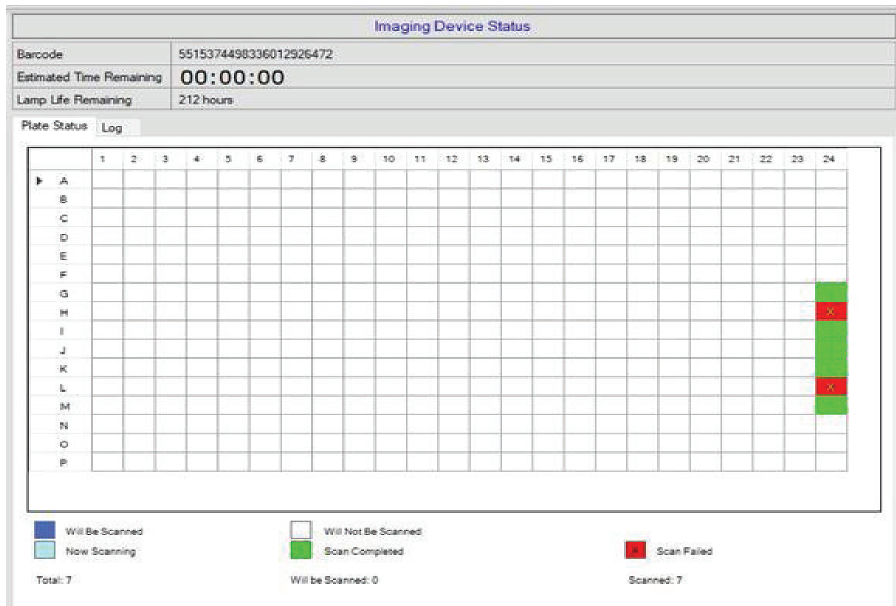
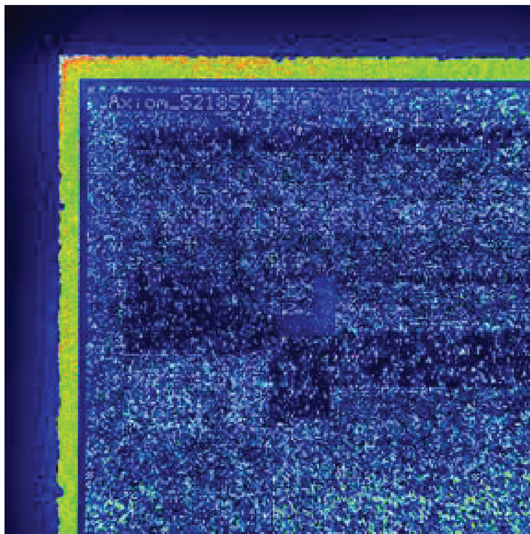


Fig.1. GCC software Scanner Status : Red coloured squares with "X" mark are samples that failed in scanning (H24 and L24) and green colored squares with "S" means these pegs were scanned successfully with a .CEL file generation (G24, I24, J24, K24, and M24).



(a)

(b)

Fig.2. Representative image of a peg of the Affymetrix microarray plate generated in the Gene Chip® Command Console software (.DAT format). (a) The image displays a raw scan of the microarray peg surface, where each small fluorescent dot corresponds to an individual oligonucleotide probe set immobilized on the array. The fluorescent signal intensity of each dot reflects the degree of hybridization between the target nucleic acids and the probe, forming the basis for subsequent genotyping and analysis. The whole image represents successful .CEL file generation for a single sample, and (b) a raw scan image of a microarray plate peg surface for a sample with no .CEL file generation indicating complete absence of fluorescent probe signals. The uniformly dark appearance indicates a lack of hybridization.

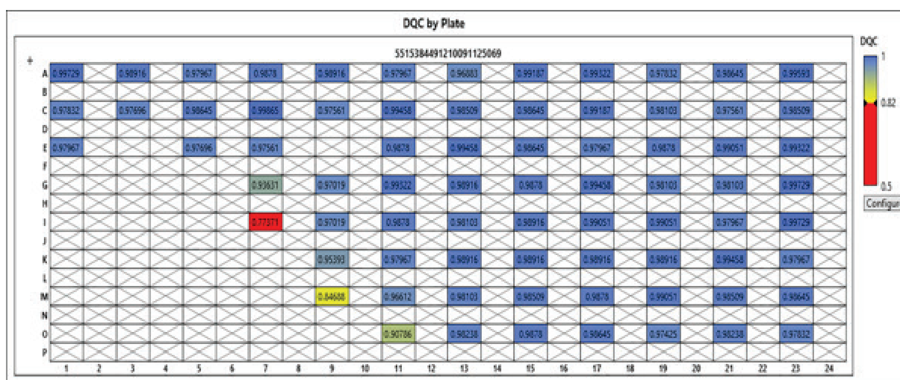


Fig.3. Plate View of Dish Quality Control (DQC) values generated after analysis of .CEL files from a mini-96 Affymetrix microarray plate in the AxiomTM Analysis Suite software. Each well corresponds to an individual sample, with the color scale representing DQC scores ranging from 0.5 (red, poor quality) to 1.0 (blue, high quality) with 0.82 being the threshold value. DQC is a key QC parameter that reflects quality of input DNA samples. Most wells across the plate exhibit high DQC values (>0.97), indicating robust performance. However, 20 wells located in the bottom-left corner of the plate show missing data and 3 samples with low DQC values, suggesting localized failure. This clustered pattern may be attributed to technical factors such as pipetting or dispensing errors or uneven plate clamping issue.

REFERENCES

Bonin, A., Bellemain, E., Bronken, P., Pompanon, F., Brochmann, C. and Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**: 3261-3273.

Bernard, M., Dehaullon, A., Gao, G., Paul, K., Lagarde, H., Charles, M., Prchal, M., Danon, J., Jaffrelo, L., Poncet, C., Patrice, P., Haffray, P., Quillet, E., Dupont-Nivet, M., Palti, Y., Lallias, D. and Phocas, F. (2022). Development of a High-Density 665 K SNP array for rainbow trout genome-wide genotyping. *Front Genetics,* **13**: 941340. doi: 10.3389/fgene.2022.941340.

Callen, D.F., Thompson, A.D., Shen, Y., Phillips, H.A., Richards, R.I., Mulley, J.C. and Sutherland, G.R. (1993). Incidence and origin of “null” alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics*, **52**(5): 922-927.

Hoffman, J.I. and Amos, W. (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecualr Ecology*, **14**(2): 599-612. doi: 10.1111/j.1365-294X.2004.02419.x.

Neumann, B., Korkuć, P., Arends, D., Wolf, M., May, K., Reißmann, M., Elzaki, S., König, S. and Brockmann, G. (2021). Design and performance of a bovine 200 k SNP chip developed for endangered German Black Pied cattle (DSN). *BMC Genomics,* **22**: 905.

Paetkau, D. and Strobeck, C. (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, **4**(4): 519-520. doi: 10.1111/j.1365-294x.1995.tb00248.x.

Pompanon, F., Bonin, A., Bellemain, E. and Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics,* **6**(11): 847-859. doi: 10.1038/nrg1707.

Sudhakar, A., Khade, A., Nayee, N., Chandrasekhar, R. and Maurya, B. (2021). Novel mutation in UMPS gene leads to false positive result of DUMPS (genetic disorder) in buffaloes: need for gene sequencing before confirming results of RFLP in new species. *Journal of Genetics*, **100**: 55.

Sudhakar, A., Nayee, N., Saha, S., Donthula, S.K., Poojara, H.V., Gohil, T., Patel, A.C. and Maurya, B.K. (2023). Role of genetic introgression in introducing mutant alleles in *Bos indicus* cattle and prevalence of lethal genetic disorders in *Bos taurus* × *Bos indicus* and *Bos indicus* cattle in India. *Tropical Animal Health Production*, **955**(6): 399. doi: 10.1007/s11250-023-03798-8.

Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N. and Bouvet, J. (1996). Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**(16): 3189-3194. doi: 10.1093/nar/24.16.3189.