ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com
Journal DOI: 10.21474/IJAR01

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

DNA Yield and Concentration Modeling by Age, A260/280 and A260/230 Ratios Using Multiple Linear regression (MLR), Bootstrap and Response Surface Methodology (RSM).

Siti Nazihahasma Hassan[1], Wan Suriana Wan Ab Rahman[1*], Wan Muhamad Amir W Ahmad[1], Suharni Mohamad[1], Rosline Hassan[2] and Selamah Ghazali[2].

1. School of Dental Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan.
2. School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan.

Manuscript Info

Manuscript History:

Received: 22 April 2016
Final Accepted: 17 May 2016
Published Online: June 2016

Key words:
A260/280 ratio, A260/230 ratio, DNA yield, DNA concentration and Multiple Linear Regression

Corresponding Author

Wan Suriana Wan Ab Rahman.

Abstract

Bootstrap, Multiple Linear Regression (MLR) and Response Surface Methodology (RSM) are a collection of statistical tools and techniques for analyzing data from various fields. Combining this idea was very useful for the modeling with an advanced analysis and possibly could be an alternative method for modeling options in applied statistics scope. Merging these methods was capable to handle the case of small sample size and limited data. This report supplied a comprehensive modeling of genomic deoxyribonucleic acid (gDNA) yield and concentration with age, A260/280 ratio and A260/230 ratio as independent variables using advanced statistical tools. This obtained model could be used as a noteworthy reference in genotyping and genetic epidemiological studies. The DNA yield and concentration was estimated through the Multiple Linear Regression Analysis (MLRA). Overall results showed that age, A260/280 and A260/230 ratios play an important role in predicting the DNA yield and concentration.

Introduction: -

Epidemiologists are increasingly trying to supplement observational data with biological material, including deoxyribonucleic acid (DNA). Extracted genomic DNA from different biological samples is widely used especially for mutation detection (El-Fadaly et al., 2016), DNA profiling (Vieira-Silva et al., 2015), diseases diagnosis (Liu et al., 2014) and genotyping (Chou et al., 2004; Schubert et al., 2006).

The standard nucleic acid quantitation method is ultraviolet (UV) spectrophotometry. The purity of the extracted DNA is usually confirmed by the UV spectrophotometer and calculated as the 260/280 OD ratio and 260/230 OD ratio. The ratio of absorbance at 260 nm and 280 nm was used to assess protein contamination while the ratio of absorbance at 260 nm and 230 nm was used to assess phenols, aromatic compounds, peptides and carbohydrates contamination (Nicklas & Buel, 2003). Both spectrophotometric measurements are included as standards for DNA quality assessment in molecular study (Di Pietro et al., 2011).

A 260/280 nm ratio above 1.8 is considered as standard for pure DNA (Psifidi et al., 2010). Meanwhile, the most desirable 260/230 nm ratio result is above 2.0 (Arif et al., 2010; Psifidi et al., 2010). Contamination of nucleic acid solutions makes spectrophotometric quantitation inaccurate. Falsely elevated DNA concentration could occur due to proteins, RNA, and chaotropic salts contamination (Haque et al., 2003). Meanwhile, falsely elevated A260/A280 and A260/A230 purity ratios could be resulted from buffer salts, such as Trisaminomethane (Tris), ethylenediaminetetraacetic acid (EDTA), and guanidine isothiocyanate contaminations (Wilfinger et al., 1997).

The DNA yield was affected by the age of an individual (Richardson et al., 2006; Caboux et al., 2012). Thus, we developed a DNA modeling based on age, A260/280 and A260/230 ratios using Bootstrap, Multiple Linear

Regression (MLR) and Response Surface Methodology (RSM). This paper provides a reference in estimated the DNA yield and concentration with age, A260/280 and A260/230 ratios as predictors.

Materials and Methods: -

Sample size for multiple regression analysis were calculated by using G*power with effect size = 0.15, 0.05, power of the study = 0.85 and number of predictor were 2. The minimum sample size required is 76 respondents. Namely variables are as in Table 1. Figure 1. showed the flow chart of regression method for contour plot and 3D modeling procedure for genomic DNA.

Table 1. Description of data

Num.	Code	Explanation of user variables
1.	Age	Age of blood donors
	1	(20-29 years old)
	2	(30-39 years old)
	3	(40-49 years old)
	4	(50 years old)
2.	A260/280 ratio	Indicator for DNA purity
3.	A260/230 ratio	Indicator for DNA purity

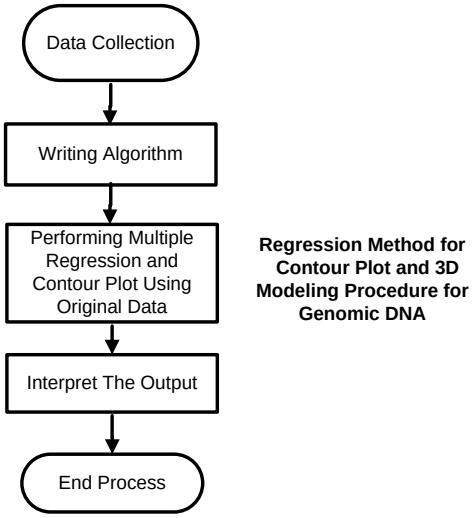

Figure 1. Flow of the modeling analysis procedure

```
Data Blood;
input Age Concentration Yield Ratio_260280 Ratio_260230;
cards;
1       49.53   1.89    2.27
1       65.38   1.87    2.75
1       32.28   1.91    2.37
1       27.27   1.89    2.40
1       26.33   1.92    2.54
4       56.34   1.89    2.33
1       40.34   1.88    2.38
7       50.25   1.88    2.37
3       35.00   1.87    2.48
1       46.21   1.89    2.53
1       26.27   1.90    2.90
4       39.72   1.85    2.35
```

```
⋮       ⋮       ⋮       ⋮
1       26.64   1.87    2.60
2       21.57   1.95    2.63
1       35.60   1.91    2.44
2       22.14   1.89    2.42
1       48.67   1.86    2.39
1       54.70   1.87    2.40
3       23.52   1.86    2.46
;
run;
ods rtf file ='robdunc0.rtf' style = journal;

/*ADDING BOOTSTRAPPING ALGORITHM TO THE METHOD */

Title "Performing bootstrap with case resampling";
Proc surveyselect data=blood out=boot1 method=urssamprate=1 outhits rep=1;
run;

/*PRINT OUT DATA FROM BOOTSTRAPT METHOD */
Data=boot1;
proc print data=boot1;
run;

ods graphics on;
procrobustreg method=ltsfwls data=boot1;
model Yield = Age Ratio_260280  / diagnostics itprint ;
output out=resids out=robout r=residual weight=weight outlier=outlier sr=stdres;
run;
ods graphics off;

procrobustreg method=ltsfwls data=boot1;
model Concentration = Age Ratio_260280 / diagnostics itprint ;
output out=resids out=robout r=residual weight=weight outlier=outlier sr=stdres;
run;
ods graphics off;

/* PLOTS = (SURFACE)*/
ods graphics on;
procrsreg data = blood plots = (surface);
model Yield = Age Ratio_260280 /lackfit;
run;
ods graphics off;

/* SURFACE(3D)*/
ods graphics on;
procrsreg data = blood plots = surface(3D);
model Concentration = Age Ratio_260280/lackfit;
run;
ods graphics off;
ods rtf close;
run;
```

Results and Discussion: -

Table 2. Parameter estimates for final model

Analysis of Maximum Likelihood Estimates							
Parameter	*DF*	*Estimate*	*Standard Error*	*95% Confidence Limits*		*Chi-Square*	*Pr > ChiSq*
Intercept	1	0.0011	0.0011	-0.0010	0.0032	1.04	0.3071
Concentration	1	0.2000	0.0000	0.2000	0.2000	1.2359	<.0001
Scale	0	0.0026					

Dependent variable: Yield

Simple linear regression

$Model_1$: Yield = 0.0011 + 0.2000 Concentration

Table 2. Showed the result of the regression modeling. The DNA concentration was a significant contribution to the DNA yield.

Table 3. Parameter estimates for final model

Analysis of Maximum Likelihood Estimates							
Parameter	*DF*	*Estimate*	*Standard Error*	*95% Confidence Limits*		*Chi-Square*	*Pr > ChiSq*
Intercept	1	1921.039	492.1807	956.3822	2885.695	15.23	<.0001
Ratio_260280	1	-845.771	265.8706	-1366.87	-324.674	10.12	0.0015
Ratio_260230	1	-54.7521	27.1165	-107.899	-1.6047	4.08	0.0435
Scale	0	42.2863					

Dependent variable: Concentration

Multiple linear regression

$Model_2$: Concentration = 1921.039 - 845.771 Ratio_260280 - 54.7521Ratio_260230

Table 3. Showed the result of the regression modeling. The A260/280 ratio ($\beta_1 = -845.771$; $Se = 265.8706$; $p < 0.0015$) was a significant contribution to the DNA concentration. While A260/230 ratio ($\beta_2 = -54.7521$; $Se = 27.1165$; $p < 0.0435$) was also a significant contribution to the DNA concentration.

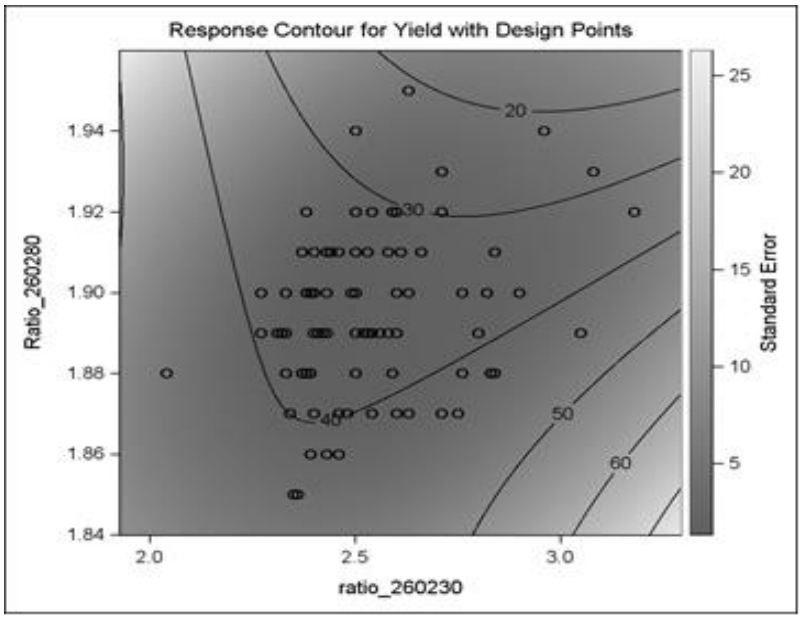

Figure 2. Response contour for yield with design points

The counter and surface plots indicate the highest value of yield (genomic DNA extract) was obtained when the reading of A260/280 ratio was low and the reading of A260/230 ratio was high (Figure 2). This area appears at the lower right corner of the plot. In addition, the shape of the response surface shows general idea of A260/280 ratio response at different setting of the A260/230 ratio reading.

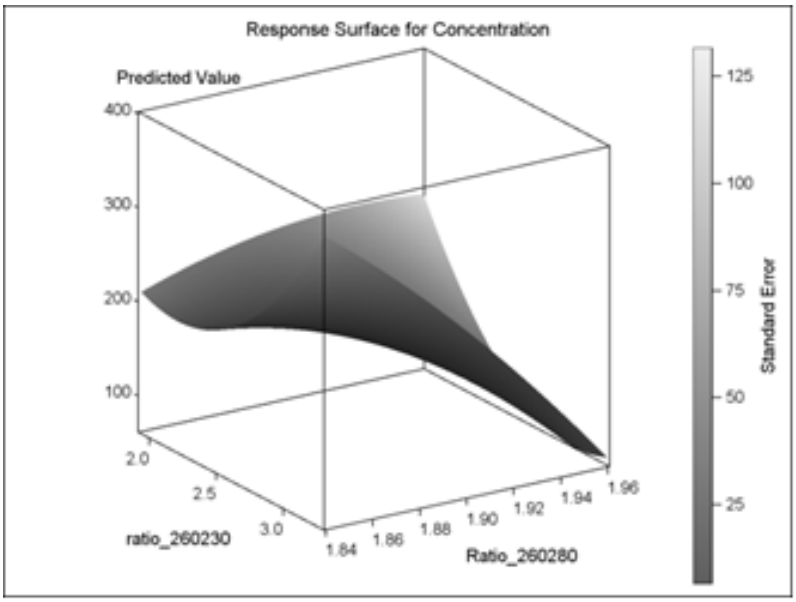

Figure 3. Response surface for concentration

The three-dimension plot (3-D) showed the behavior of reading A260/280 ratio with the reading of A260/230 ratio (Figure 3). The highest concentration appears when the A260/230 ratio was high and the reading of A260/280 ratio was low.

Table 4. Parameter estimates for final model

Analysis of Maximum Likelihood Estimates							
Parameter	*DF*	*Estimate*	*Standard Error*	*95% Confidence Limits*		*Chi-Square*	*Pr > ChiSq*
Intercept	1	1.7895	0.0361	1.7188	1.8602	2460.12	<0.0001
Age	1	-0.0104	0.0042	-0.0186	-0.0022	6.22	0.0126
Ratio_260280	1	0.0492	0.0140	0.0217	0.0767	12.29	0.0005
Scale	0	0.0146					

Dependent variable: Yield

Multiple linear regression

Model$_3$: Yield $= 1.7895 - 0.0104\,\text{Age} + 0.0492\,\text{Ratio\_260280}$

Table 4. Showed the result of regression modeling. The A260/280 ratio $(\beta_2 = 0.0492;\ Se = 0.0140;\ p < 0.0005)$ was a significant contribution to the DNA yield. Age $(\beta_1 = -0.0104;\ Se = 0.0042;\ p < 0.0126)$ also was a significant contribution to the DNA yield.

Table 5. Parameter estimates for final model

Analysis of Maximum Likelihood Estimates							
Parameter	*DF*	*Estimate*	*Standard Error*	*95% Confidence Limits*		*Chi-Square*	*Pr > ChiSq*
Intercept	1	112.9941	12.2993	88.8879	137.1004	84.40	<0.0001
Age	1	-4.6571	1.5880	-7.7695	-1.5447	8.60	0.0034
Ratio_260280	1	-27.8635	4.8240	-37.3184	-18.4087	33.36	<0.0001
Scale	0	6.4658					

Dependent variable: Concentration

Multiple linear regression

$Model_4$: $\text{Concentration} = 112.9941 - 4.6571\,\text{Age} - 27.8635\,\text{Ratio\_260280}$

Table 5. Showed the result of regression modeling. The A260/280 ratio ($\beta_2 = -27.8635$; $Se = 4.8240$; $p < 0.0001$) was a significant contribution to the DNA concentration. Age ($\beta_1 = -4.6571$; $Se = 1.5880$; $p < 0.0034$) also was a significant contribution to the DNA concentration.

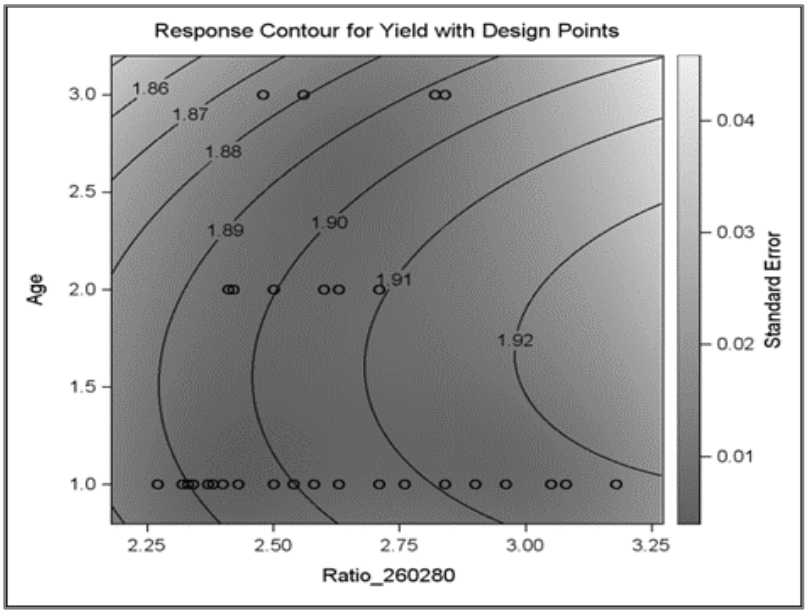

Figure 4. Response contour for yield with design points

The counter and surface plots indicate that high value of yield (genomic DNA extract) was obtained when the number of age decrease and the reading of A260/280 ratio increase (Figure 4). This area appears at the bottom right corner of the plot. In addition, the shape of the response surface gives a general idea of response age at various setting of reading A260/280 ratio.

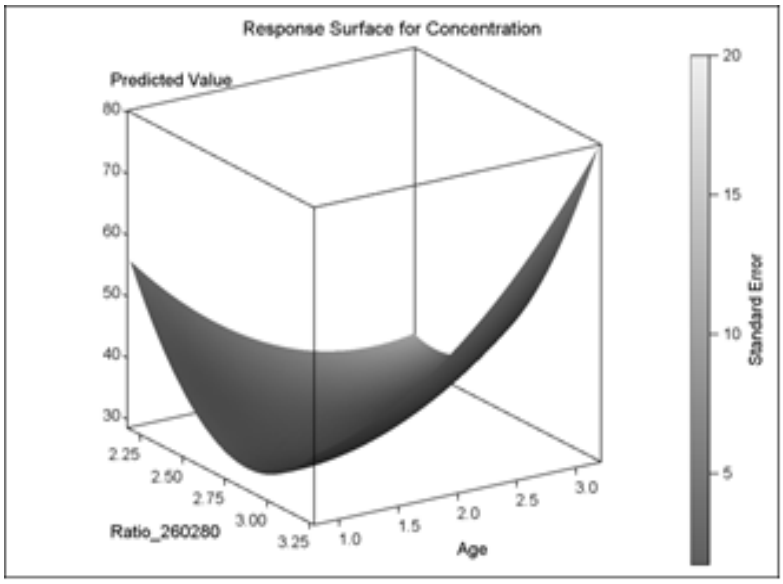

Figure 5. Response surface for concentration

The three-dimension plot (3-D) showed the behavior of reading A260/280 ratio with the reading of age (Figure 5). The concentration was affected by the A260/280 ratio and age. This (3-D) plot showed that the concentration was increased when the A260/280 ratio and the number of age were increased.

Summary and Conclusion: -

DNA modeling was performed where age, A260/280 and A260/230 ratios affects the DNA yield and concentration. DNA yield was directly proportional to DNA concentration. The higher the concentration, the higher the yield. According to the Nano drop measurement, good-quality DNA will haveA260/A280 ratio of 1.8-2.0 and A260/A230 ratio of greater than 2.0 (Arif et al., 2010; Psifidi et al., 2010). The study by Erkelleryuksel et al., (1992) found a statistically significant in leukocyte count from birth to adults. While, Richardson et al., (2006) found that age group 20-50 years old produced higher DNA yield as compared to age above 50 years old. This study found that DNA yield was affected by the age. Thus, this study emphasizes on regression modeling for modeling of age, A260/280 and A260/230 ratios on DNA yield and concentration using Multiple Linear Regression (MLR), Bootstrap and Response Surface Methodology (RSM). In addition, it provides comprehensive information and general idea of how the curve of the dependent variables moves with the two independent variables.

Acknowledgments: -

The authors would like to express their gratitude to Universiti Sains Malaysia for providing the research funding (Grant no.304/PPSG/61312118), the support of USM fellowship, My Brain15 (Ministry of Higher Education Malaysia) and Hematology Department staff, School of Medical Sciences, USM.

Conflict of interest: -

The authors declare there is no conflict of interest regarding the publication of this paper.

References: -

1. Alonso, A., Martin, P., Albarran, C., Garcia, P., Primorac, D., Garcia, O., de Simon, L. F., Garcia-Hirschfeld, J., Sancho, M., & Fernandez-Piqueras, J. (2003). Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies. *Croat Med J*, *44*(3), 273-280.
2. Arif, I. A., Bakir, M. A., Khan, H. A., Ahamed, A., Al Farhan, A. H., Al Homaidan, A. A., Al Sadoon, M., Bahkali, A. H., &Shobrak, M. (2010). A Simple Method for DNA Extraction from Mature Date Palm Leaves: Impact of Sand Grinding and Composition of Lysis Buffer. *Int J MolSci*, *11*(9), 3149-3157.

3. Chou, Y. Y., Chen, C. C., Kuo, C. L., Tsai, W. H., & Lin, S. J. (2004). Russell-Silver syndrome: Molecular diagnosis of maternal uniparental disomy of chromosome 7 using methylation, specific polymerase chain reaction assay and single nucleotide polymorphisms genotyping. *J Formos Med Assoc, 103*(10), 797-802.
4. Caboux, E., Lallemand, C., Ferro, G., Hemon, B., Mendy, M., Biessy, C., Sims, M., Wareham, N., Britten, A., Boland, A., Hutchinson, A., Siddiq, A., Vineis, P., Riboli, E., Romieu, I., Rinaldi, S., Gunter, M. J., Peeters, P. H. M., van der Schouw, Y. T., Travis, R., Bueno-de-Mesquita, H. B., Canzian, F., Sanchez, M. J., Skeie, G., Olsen, K. S., Lund, E., Bilbao, R., Sala, N., Barricarte, A., Palli, D., Navarro, C., Panico, S., Redondo, M. L., Polidoro, S., Dossus, L., Boutron-Ruault, M. C., Clavel-Chapelon, F., Trichopoulou, A., Trichopoulos, D., Lagiou, P., Boeing, H., Fisher, E., Tumino, R., Agnoli, C., & Hainaut, P. (2012). Sources of Pre-Analytical Variations in Yield of DNA Extracted from Blood Samples: Analysis of 50,000 DNA Samples in EPIC. *Plos One*, 7(7).
5. Di Pietro, F., Ortenzi, F., Tilio, M., Concetti, F., & Napolioni, V. (2011). Genomic DNA extraction from whole blood stored from 15- to 30-years at -20 degrees C by rapid phenol-chloroform protocol: A useful tool for genetic epidemiology studies. *Mol Cell Probes, 25*(1), 44-48.
6. Erkelleryuksel, F. M., Deneys, V., Yuksel, B., Hannet, I., Hulstaert, F., Hamilton, C., Mackinnon, H., Stokes, L. T., Munhyeshuli, V., Vanlangendonck, F., Debruyere, M., Bach, B. A., &Lydyard, P. M. (1992). Age-Related-Changes in Human Blood Lymphocyte Subpopulations. *J Pediatr, 120*(2), 216-222.
7. Eeles. R.A., & Stamps, A.C. 1994. Polymerase Chain Reaction (PCR). The technique and its Applications.R.G. *Landes Company,* 2nd Printing, USA.
8. El-Fadaly, N., Abd-Elhameed, A., Abd-Elbar, E., & El-Shanshory, M. (2016). Accuracy of Reverse Dot-Blot PCR in Detection of Different Beta-Globin Gene Mutations. *Indian J Hematol Blood Transfus, 32*(2), 239-243.
9. Haque, K. A., Pfeiffer, R. M., Beerman, M. B., Struewing, J. P., Chanock, S. J., & Bergen, A. W. (2003). Performance of high-throughput DNA quantification methods. *BMC Biotechnol, 3*, 20.
10. Liu, Y. F., Gao, J. L., Yang, Y. F., Ku, T., &Zan, L. S. (2014). Novel extraction method of genomic DNA suitable for long-fragment amplification from small amounts of milk.*J Dairy Sci, 97*(11), 6804-6809.
11. Nicklas, J. A., &Buel, E. (2003). Quantification of DNA in forensic samples. *Anal BioanalChem, 376*(8), 1160-1167.
12. Psifidi, A., Dovas, C. I., &Banos, G. (2010). A comparison of six methods for genomic DNA extraction suitable for PCR-based genotyping applications using ovine milk samples. *Mol Cell Probes, 24*(2), 93-98.
13. Richardson, A. J., Narendran, N., Guymer, R. H., Vu, H., & Baird, P. N. (2006). Blood storage at 4 degrees C - factors involved in DNA yield and quality. *J Lab Clin Med,147*(6), 290-294.
14. Schubert, K., von Bonnsdorf, H., Burke, M., Ahlert, I., Braun, S., Bemer, R., Deichmann, K. A., &Heinzmann, A. (2006). A comprehensive candidate gene study on bronchial asthma and juvenile idiopathic arthritis. *Dis Markers, 22*(3), 127-132.
15. Vieira-Silva, C., Afonso-Costa, H., Ribeiro, T., Porto, M. J., Dias, M., &Amorim, A. (2015). Quantifiler (R) Trio DNA validation and usefulness in casework samples. *Forensic SciIntGenetSuppS,5*,246-247.
16. Wilfinger, W.W., Mackey, K., &Chomczynski, P. 1997. Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *Biotechniques*, 22, 474-481.