Results

Part I

Physico-chemical parameters of water sites

The physical and chemical analysis of water samples collected from three water resources in which some parameters were selected are presented in table (5):

1. Temperature:

There was no significant difference values of temperature recorded from the three sites of water sample collection. The temperature ranged from 23.1 to 24.9 ^{o}C.

2. pH:

The pH values for all collected samples were higher than 6.5 which is normally expected in raw water due to the presence of carbonates and bicarbonates, the highest PH value (8.21) was recorded for drainage whilst the lowest value (7.56) was for well water. The PH ranged from 7.56 to 8.21 towards slight alkalinity character.

3. Turbidity:

The water turbidity depends on pollution extend and suspended particles. There were extreme differences in turbidity values in the sites of study in the way that the highest value of drainage water was 65.14 Nephlometric Turbidity Units (NTU), while the lowest value (4.00 NTU) was extremely low as compared to that recorded for drainage water and in accordance with the limits of standard drinking water that don not exceed 5 NTU.

4. Biochemical oxygen demand (BOD):

The permissible standard limits of BOD values is (6mg/L) for fresh water bodies, in a sensible way the BOD highest value (11.00 mg/L) was recorded for drainage water and exceeds the permissible limits while the lowest value (3.00 mg/L) was shown to be for well water.

5. Chemical oxygen demand (COD):

The maximum value of COD according to the permissible standard limits is (10mg/L) for fresh water bodies. Unexpected high value was recorded for drainage water (39.00 mg/L) and exceeded the maximum value for COD, in a similar way river water (14.00 mg/L) exceeded the maximum value. The lowest value (6.00 mg/L) for well water was accepted.

6. Dissolved oxygen (DO):

DO concentrations showed variable results according to site nature and extend of pollution. So clarity of water matches with high DO concentrations as shown in the table (5) in which well water was recorded with highest DO concentration (8.91 mg/L), in contrast to drainage water (3.38 mg /L), the limits recommended DO to be not less than 5mg/L.

Table (5): Selected physico-chemical parameters of water samples collected from different sites:

Sampling site / Physico-chemical parameter	Well water	River water	Drainage water
Temperature (º c)	23.1	23.7	24.9
pH	7.56	7.98	8.21
Turbidity (NTU)	4.00	11.81	65.14
Biochemical oxygen demand(BOD)(mg/l)	3.00	6.00	11.00
Chemical oxygen demand(COD) (mg/l)	6.00	14.00	39.00
Dissolved oxygen (DO) (mg/l)	8.91	5.01	3.38

Part II

Isolation and identification of *E.coli*

1. Isolation of *E .coli*:

Tubes showing gas production with growth are considered a positive indication for *E.coli* presence; from such tubes, several loopful were streaked on macconkey agar plates. The culture showed pink to red color surrounded by red zone. Bacterial colonies suspected to be *E.coli* were chosen and picked up by streak plate method on nutrient agar medium. Pure isolates were maintained on slants of the same medium at 4°c for further examination. The well, river and drainage isolates were designated as *E.coli-EG1*, *E.coli-EG2* and *E.coli-EG3*, respectively.

2. Determination of *E.coli* count in water samples:

The blue colonies developed on membrane filters that were placed on m-FC media were counted for the three different water samples from well, river and drainage sites as in the following table:

Table (6): *E.coli* count in the three water samples filtered:

***E.coli* isolates**	***E.coli-EG1***	***E.coli-EG2***	***E.coli-EG3***
Total *E.coli* *colonies/ 100 ml*	3	900	200

3. Identification of *E.coli*:

3.1. Metabolic profiling using biolog system:

The three isolates of *E.coli* although have the same carbon source utilization pattern that they were able to utilize 46 type out of 95 carbon sources as in table (7); (Dextrin, Glycogen, N-Acetyl-D-Galactosamine, N-Acetyl-D-Glucosamine,l-Arabinose,d-Fuctose,l-Fructose,dGalactose,α-d-Glucose, α-d-Lactose,Maltose,d-Mannitol,d-Mannose,d-Melibiose,β-Methyl-d-Glucose,d-Psicose,l-Rhamnose,d-Sorbitol,d-Trehalose,Pyruvic acid Methyl Ester,Succinic acid Mono-Methyl-Ester,d-Galactouronic acid,d-Gluconic acid,d-Glucuronic acid, α-Hydroxybutyric acid,p-Hydroxy Phenylacetic acid, d,l-Lactic acid,Propionic acid,Succinic acid,Bromosuccinic acid,Glucuronamide, l-Alaninamide, d-Alanine, l-Alanyl-Glycine, l-Asparagine,l-Asparatic acid,Glycl-l-Asparatic acid,Glycyl-l-Glutamic acid,d-Serine, l-Serine, Inosine, Uridine, Thymidine, Glycerol, d,l- α-Glycerol Phosphate, α-d-Glucose1-Phosphate and d-Glucose6-Phosphate) the degree of utilization differed and influenced by geographical origin of the isolates. The pattern of utilization of carbon sources (The metabolic profile or metabolic profiling) matched with profiles of known taxa in the biolog database

using biolog microlog1 4.20 software .The pattern of carbon utilization is shown in table (7):

Table (7): Results of microplate 96 wells (according to table (1)) after incubation with the pattern of carbon sources:

No. / Symbol	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	+	+	-	-	+	+	-	+	-	-
B	-	+	+	+	-	+	-	+	-+	+	+	+
C	+	+	+	-	+	+	-	+	-	-	+	+
D	(/)	-	-	-+	-	+	+	-	+	+	-	-
E	+	-	(/)	-	-	+	-	+	-	-	-	+
F	+	-	+	+	+	+	+	+	+	-	+	+
G	-	-	-	-	-	-	-	+	+	-	-	-
H	-	+	+	+	-	-	-	-	+	+	+	+

Key: + = positive result (violet color) - = negative result (colorless)

(/) = borderline (neither positive nor negative) -+ = mismatched negative

This metabolic profile was compared and matched with 10 known taxa which are *Escherichia coli*(USP-7085),*Escherichia coli, citrobacter braaki,salmonella gp 3A, citrobacter werkmanii,citrobater murliniae,citrobacter freundii, citrobacter younge,citrobacter amalonacticus* and *salmonella gp1 (choleraesuis)*. The three isolates were identified as *Escherichia coli* (USP-7085) biotype with simple matching coefficient (SIM) of 0.84.

Cluster analysis based on metabolic profiling was performed in two steps. The first step, the plate dendrogram was established depending on distance coefficients between the test isolates and bacterial species of known profile. This dendrogram showed 3 clusters of profiles in which the three isolates matches with *E. coli* (USP-7085) in a separate cluster as in fig. (2).The second step is the creation of a three dimensional graphic showing the species interrelationships of the test isolates with the compared species of known profile as in fig. (3).

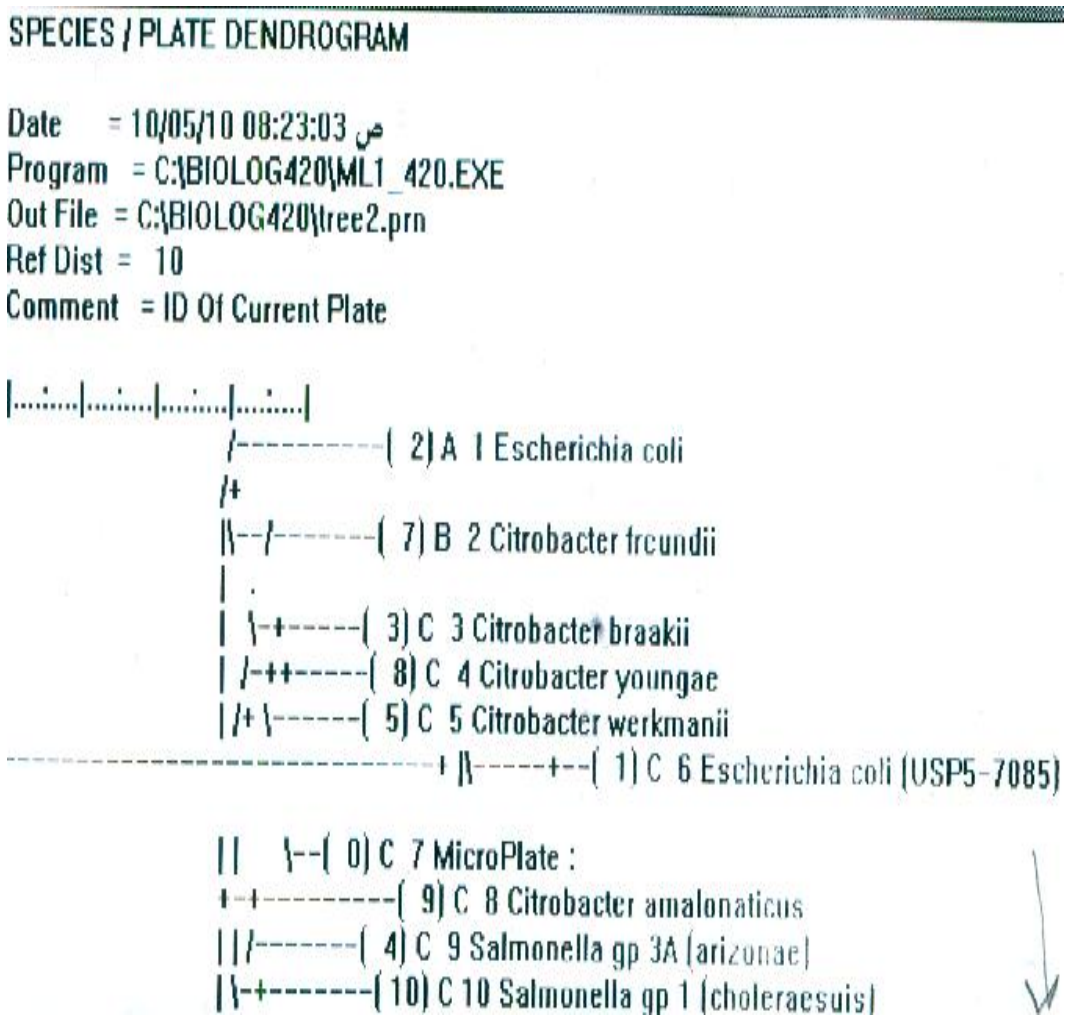


Fig. (2): Dendrogram showing the relationship between the three *E.coli* isolates and other nine bacterial species based on metabolic profiling.

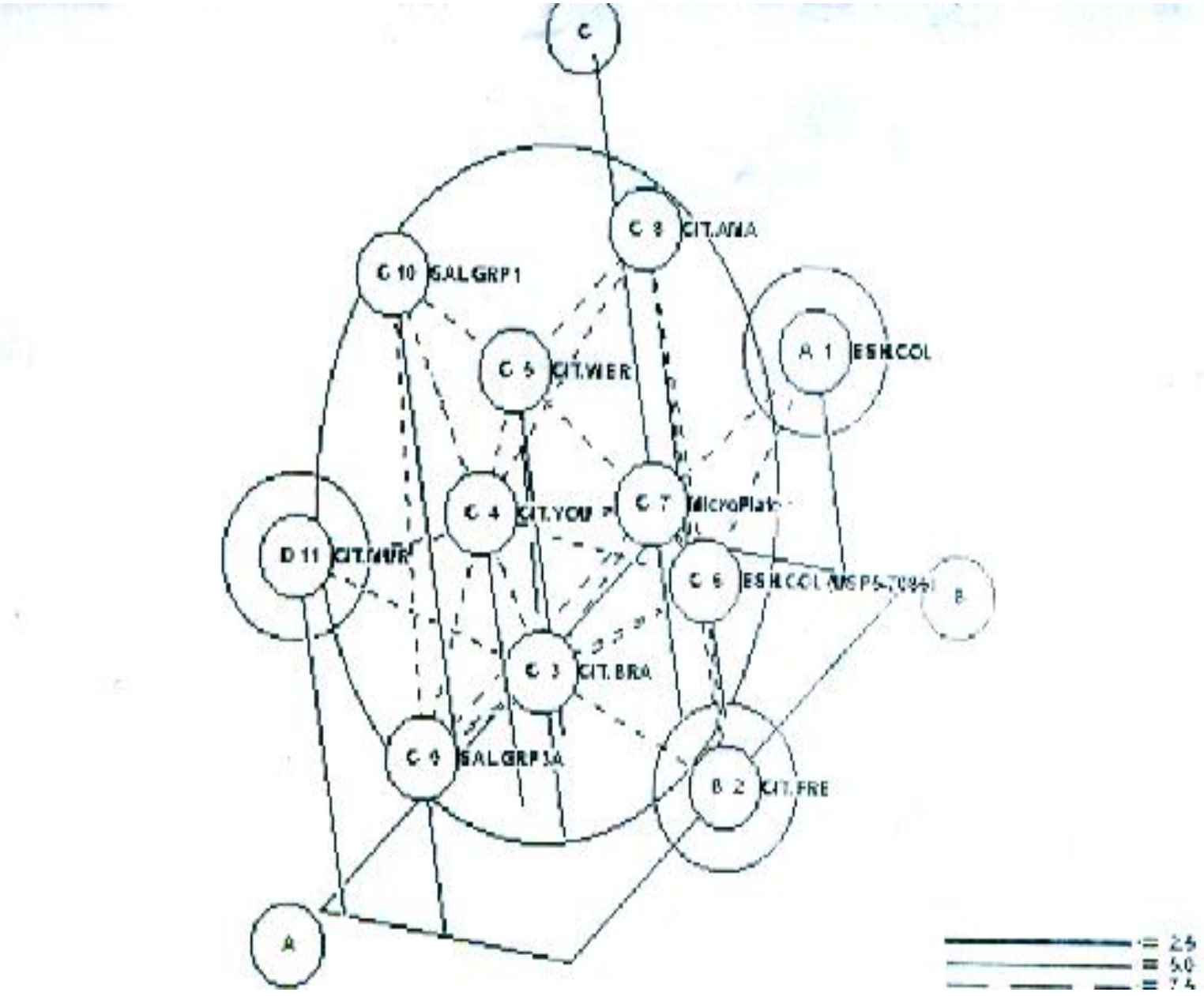


Fig. (3): Three-dimensional graph showing the interrelationships of the three *E.coli* isolates (A, B and C) with the other compared bacterial species.

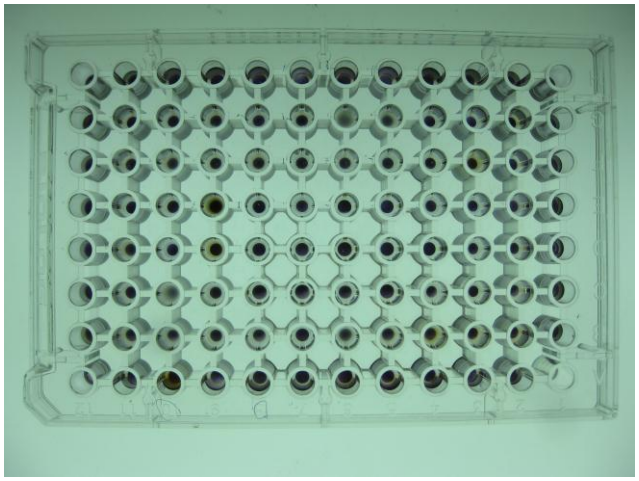


Plate (A)

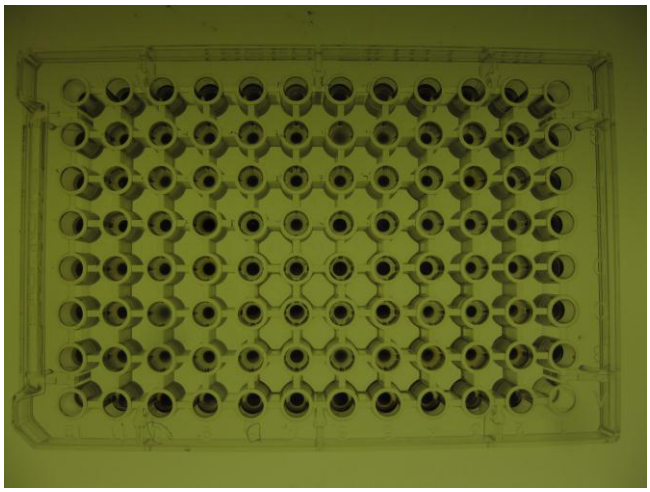


Plate (B)

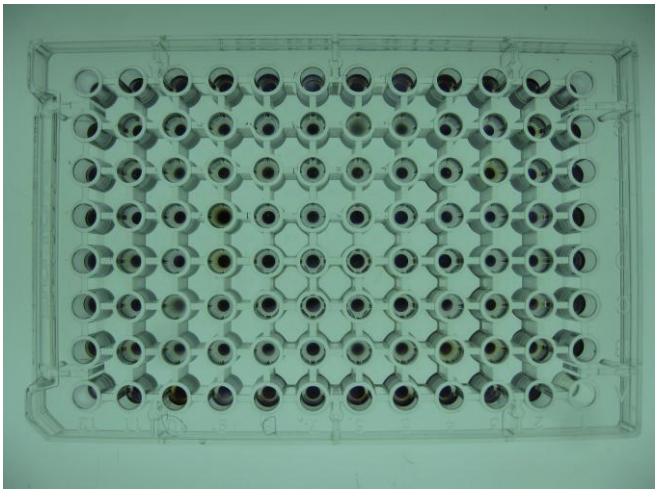


Plate (C)

Fig. (4): Biolog microplate containing 95 types of carbon sources and water as a control during identification of *E.coli* using biolog system. A, B and C represents *E.coli* from well, river and drainage, respectively.

The figure above showed the microplates inoculated with the *E.coli* isolates and incubated for 24 hours. The developing blue color indicates occurrence of carbon source utilization by *E.coli*.

3.2. 16s rRNA gene sequencing:

DNA isolated from the three *E.coli* isolates was amplified by PCR conventional method after reaching the sufficient concentration using specific primer sequences for 16s rRNA gene in *E.coli*. Amplicons from the PCR were allowed for sequencing reaction through cycle sequencing method. The DNA amplicons returned as electropherogram files. Electropherograms showed distinct peaks for each base call as well as high Q values for each call. Primers were easily identified in either the forward or reverse direction in each sequence fragment and were easily used to piece together the individual sequences as in fig (5). Sequences obtained for each primer for each isolate had sufficient overlap between them and used to form one continuous sequence (contig).

The three contig (one for each isolate) yielded and showed the partial nucleotide sequences of 16s rRNA gene. The resulted sequences were compared with bacterial species recorded on the Genbank using DNAMAN program and identified as *E.coli*.

E. coli-EG1 (530 bp)

:

```
   ١ GAGAGTTTGA TCCTGGCTCA GATTGAACGC TGGCGGCAGG CCTAACACAT GCAAGTCGAA
  ٦١ CGGTAACAGG AAGAAGCTTG CTTCTTTGCT GACGAGTGGC GGACGGGTGA GTAATGTCTG
 ١٢١ GGAAACTGCC TGATGGAGGG GGATAACTAC TGGAAACGGT AGCTAATACC GCATAACGTC
 ١٨١ GCAAGACCAA AGAGGGGGAC CTTCGGGCCT CTTGCCATCG GATGTGCCCA GATGGGATTA
 ٢٤١ GCTAGTAGGT GGGGTAACGG CTCACCTAGG CGACGATCCC TAGCTGGTCT GAGAGGATGA
 ٣٠١ CCAGCCACAC TGGAACTGAG ACACGGTCCA GACTCCTACG GGAGGCAGCA GTGGGGAATA
 ٣٦١ TTGCACAATG GGCGCAAGCC TGATGCAGCC ATGCCGCGTG TATGAAGAAG GCCTTCGGGT
 ٤٢١ TGTAAAGTAC TTTCAGCGGG GAGGAAGGGA GTAAAGTTAA TACCTTTGCT CATTGACGTT
 ٤٨١ ACCCGCAGAA GAAGCACCGG CTAACTCCGT GCCAGCAGCA AAGGTAATAA
```

E. coli-EG2 (519 bp)

```
   ١ AGAGTTTGAT CCTGGCTCAG ATTGAACGCT GGCGGCAGGC CTAACACATG CAAGGCGAGC
  ٦١ GGTAGCACAG AGAGCTTGCT CTCGGGTGAC GAGCGGCGGA CGGGTGAGTA ATGTCTGGGA
 ١٢١ AACTGCCTGA TGGAGGGGGA TAACTACTGG AAACGGTAGC TAATACCGCA TAACGTCGCA
 ١٨١ AGACCAAAGT GGGGGACCTT CGGGCCTCAT GCCATCAGAT GTGCCCAGAT GGGATTAGCT
 ٢٤١ AGTAGGTGGG GTAACGGCTC ACCTAGGCGA CGATCCCTAG CTGGTCTGAG AGGATGACCA
 ٣٠١ GCCACACTGG AACTGAGACA CGGTCCAGAC TCCTACGGGA GGCAGCAGTG GGGAATATTG
 ٣٦١ CACATGGGCG CAAGCCTGAT GCAGCCATGC CGCGTGTATG ATGAAGGCCT TCGGTTGTAA
 ٤٢١ AGTACTTCAG CGGGGTAGAA GGCGATAAGT TAATAACCTT GTCGATTGAC GTTACCCGCA
 ٤٨١ GATGAAGCAC CGGCTAACTC CGTGCCAGCA CCGGTATAA
```

E. coli-EG3 (523 bp)

```
   ١ TTGAATTTGA TCCTGGCTCA GATGAACGCT GGCGGCAGGC CTAACACATG CAAGTCGAAG
  ٦١ GTAACAGGAA GCAGCTTGCT GCTTTGCTGA CGAGTGGCGG ACGGGTGAGT AATGTCTGGG
 ١٢١ AAACTGCCTG ATGGAGGGGG ATAACTACTG GAAACGGTAG CTAATACGCA TAACGTCGCA
 ١٨١ AGACCAAAGA GGGGGACCTT CGGGCCTCTT GCCATCGGAT GTGCCCAGAT GGGATTAGCT
 ٢٤١ AGTAGGTGGG GTAACGGCTC ACCTAGGCGA CGATCCCTAG CTGGTCTGAG AGGATGACCA
 ٣٠١ GCCACACTGG AACTGAGACA CGGTCCAGAC TCCTACGGGA GGCAGCAGTG GGGAATATTG
 ٣٦١ CACATGGGCG CAAGCCTGAT GCAGCCATGC CGCGTGTAGA AGAAGGCCTC GGGTTGTAAA
 ٤٢١ GTACTTTCAG CGGGGAGGAA GGGAGTAAAG GTAATACCTT TGCTCACTTG ACTTACCCGC
 ٤٨١ AGAAGAAGCA CCGGCTAACT CCGTGCCAGC AGCAAATTTA TAA
```

Fig. (5): Partial nucleotide sequence of 16s rRNA gene for the three *E.coli* isolates.

Part III

Characterization of variability of *E.coli* isolates

1. Cultural characters:

E.coli may show distinguishing growth morphologies when cultured on different media. Isolated strains of *E.coli* were inoculated on several types of media for the purpose of observing and comparing their colonial growth characteristics. These strains of *E.coli* were inoculated on macconkey agar (MA) slants, triple sugar iron agar (TSI) slants and nutrient agar (NA) plates and growth morphologies were determined as shown in the following table (8):

Table (8): Cultural characters of the three *E.coli* isolates on different media:

	Medium type								
Colony characters	**MA**			**TSI**			**NA**		
	W	R	D	W	R	D	W	R	D
1-configration	filiform						round		
2- margin	smooth								
3- elevation	Raise-d	conve x	raised	raisd	convex	raised	raised	convex	raised
4- color of colony	Slightly red	Darkly red	Moderatel y red	Slightly yellow	white	White to yellow	white	white	white
5- transparency	+1	+3	+2	+1	+3	+2	+2	+3	+2
6- spread	+1	+3	+2	+3	+1	+2	+2	+1	+2
7- opacity	+3	+1	+2	+3	+1	+2	+1	+2	+1
8-density	Less dense	dense	Less dense	Less dense	dense	Less dense	Less dense	dense	Less dense

+1 = little +2 = moderate +3 = high

W: isolate from well water R: isolate from river water

D: isolate from drainage water

The growth and characters of *E.coli* colonies is obvious from the previous table from which macconkey agar MA, TSI and NA supported highest growth of the *E.coli-EG2* isolate, while all media supported lowest growth of the *E.coli-EG1* isolate, in addition to that, the *E.coli-EG3* isolate showed significant growth on the three types of media. Less inhibitory effect is found for all types of media to support growth of *E.coli-EG2* isolate and *E.coli-EG1*isolate on NA. These three types of media show no marked difference on colony form and margin of the three isolates. The raised elevation pattern of growth on culture media is similar for *E.coli-EG1*and *E.coli-EG3* isolates on all types of media, in contrast to the *E.coli-EG2* isolate which show a characteristic convex pattern. There is a great correlation between amount of growth exhibited by the three isolates on culture media and some characters such as direct proportionality of colony coloration, transparency and spread with amount of growth and inversely with opacity. Highest colony coloration, transparency and spread are found for *E.coli-EG2* isolates on all media, in contrast to that for *E.coli-EG1*isolate. It was obvious that the cultural pattern of *E.coli-EG1* and *E.coli-EG3* was similar on NA. There is no significant difference of the colony coloration on NA for the three isolates. The highest opacity is shown to be on MA for *E.coli-EG1*isolate, TSI and MA for *E.coli-EG2* isolate and moderate amount for *E.coli-EG3* isolate. The cultural characters were derived from the growth on the three types of media as in the following figures (6, 7 and 8).

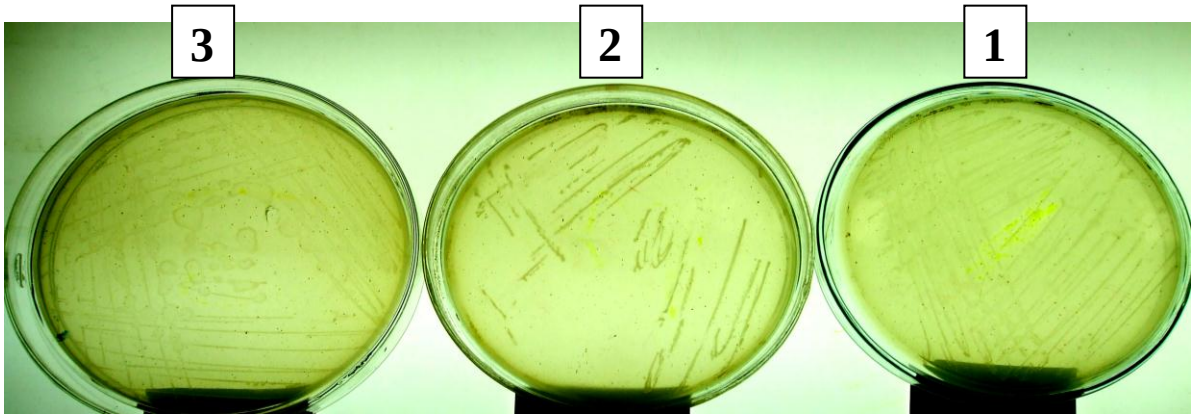


Fig. (6): Plates of nutrient agar showing growth of *E.coli*. Plate (1) represents *E.coli-EG1*, plate (2) *E.coli-EG2* and plate (3) from *E.coli-EG3*.

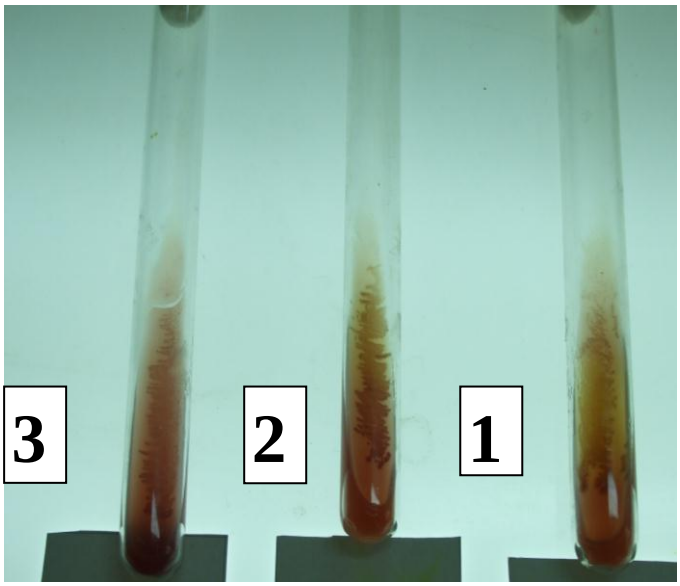


Fig. (7): Growth of *E.coli* on Macconkey agar slants. Slant (1) represents *E.coli-EG1*, slant (2) from *E.coli-EG2* and slant (3) from *E.coli-EG3*.

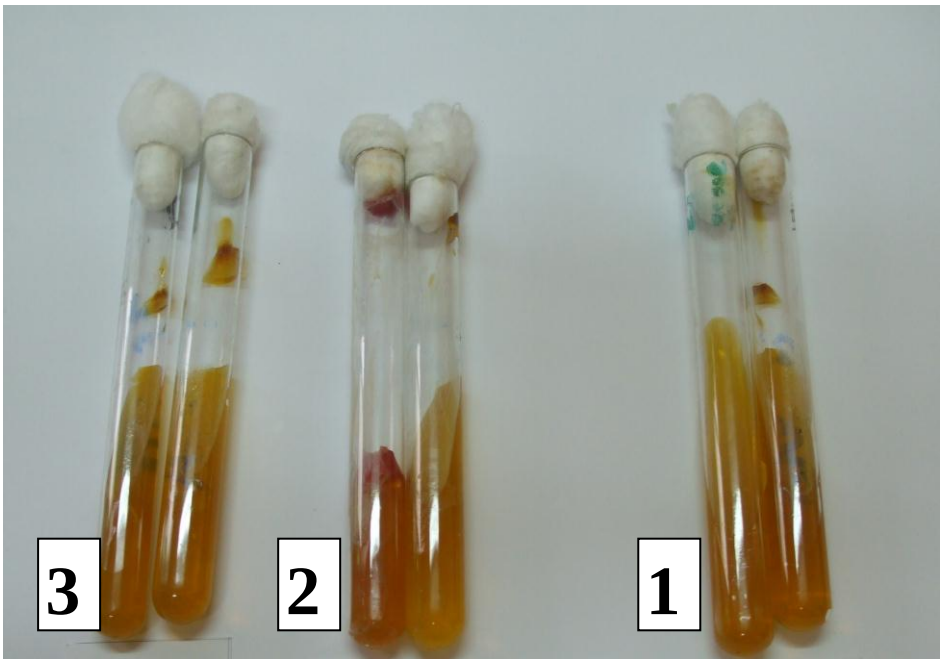


Fig. (8): Growth of *E.coli* on TSI slants. Slant (1) represents *E.coli-EG1*, slant (2) from *E.coli-EG2* and slant (3) from *E.coli-EG3*. Yellow color and fractionation of the media components due to production of acids and gas indicates growth of *E.coli* respectively.

2. Biotyping:

The three isolates of *E. coli* displayed the same pattern in carbon source utilization profile via biolog identification. The total number of carbon sources utilized was 48 out of 95, from these data the functional diversity was calculated from the formula: % of functional diversity =number of positive carbon source wells/ total number of carbon source wells x 100. So, this percentage was found to 50.52 % for these isolates. The 95 carbon sources are categorized into 7 groups or categories as follow:oligo/polysaccharides(3),detergents(2),Sacchrides/derivatives(23), sugar alcohols/derivatives(11),aminoacids/degradative products (22),organic acids/derivatives (31) and nucleotides(3).Based on this categorization the percentage of carbon group utilization was calculated from the formula : % of carbon group utilization = no. of positive carbon sources in a given group/ maximum no. of carbon sources in the group x

100.The percentages for all carbon groups are tabulated in the following table:

Table (9): Calculation of % of carbon group utilization for the seven carbon group by *E.coli* isolates:

Carbon source group	Total number of carbon sources	No. of utilizable carbon sources	% of carbon group utilization
1-Oligo/polysacchrides	3	2	66.66%
2-Detergent	2	0	0 %
3-Sacchrides/derivatives	23	17	73.91%
4-Sugar alcohols/derivatives	11	4	36.36%
5-Amino acids/ degradative products	22	10	45.45%
6-Organic acids/ derivatives	31	12	38.70%
7-nucleotides	3	3	100%

From the data recorded in the table, it appeared that *E. coli* at the species level favors utilization of carbohydrates in the form of monosaccharides, disaccharides and breakdown of nucleotides as carbon and nitrogen sources for energy gain and in other times for nucleic acid synthesis. Depending on the availability of efficient hydrolyases enzyme system, oligo and polysaccharides occupy a leading position in utilization capability.

The ability to utilize detergents was in significant at any way for *E.coli* isolates, so, these types of carbon sources are regarded as un utilizable hydrocarbons.

Although the same patterns of carbon source utilization, functional diversity, % of carbon group utilization were displayed for the three isolates, the degree of carbon source utilization differed among the three

isolates. The variation in color intensity for the same reaction for the three isolates accounted for the variability in carbon consumption. The highest degree of utilization always was recorded for the *E.coli-EG2* isolate whilst the lowest capability was recorded for the *E.coli-EG1* isolate. The *E.coli-EG3* isolate occupied an intermediate position and in few cases the highest degree of utilization for certain carbon sources. The following table gives an account for such differences in utilization degree of carbon sources.

Table (10): Variations among the isolates of *E.coli* in the degree of carbon sources utilization:

Carbon sources	***E.coli-EG1***	***E.coli-EG2***	***E.coli-EG3***
dextrin	+	+++ +	+ +
glycogen	+ +	+ + ++	+ + +
N-acetyl-D-galactosamine	+ +	+ + + +	+ + + +
N-acetyl-D-glucosamine	+ +	+ + +	+ + +
L-arabinose	+ + + +	+ + + +	+ + + +
D-fructose	+ +	+ + ++	+ + +
L-fructose	+ +	+ + ++	+ + +
D-galactose	+	+ + ++	+ + +
α-D-glucose	+	+ ++	+ +
α-D-lactose	+ +	+ + ++	+ + +
maltose	+ +	+ ++ +	+ +
D-mannitol	+	+ + ++	+ + +
D-mannose	+ +	+ + ++	+ + +
D-melibiose	+	+ ++ +	+ +
β-methyl-D-glucoside	+ +	+ + +	+ + +
D-psicose	+	+ ++	+ +
L-rhammnose	+ +	+ + ++	+ + +

D-sorbitol	+ +	+ + ++	+ + +
D-trehalose	+ +	+ + ++	+ + +
Pyruvic acid methyl ester	+ +	+ + ++	+ + +
Succinic acid-mono-methyl ester	+	+ + ++	+ + +
D-galacturonic acid	+ +	+ + ++	+ + +
D-gluconic	+	+ + +	+ + +
D-glucuronic acid	+ +	+ + ++	+ + +
α- hydroxybutyric acid	+ +	+ ++	+ +
p.hydroxyphenyl acetic acid	+ +	+ + +	+ + +
D,L –lactic acid	+ +	+ + + +	+ + +
Propionic acid	+	+ ++	+ +
Succinic acid	+ +	+ + + +	+ + + +
Bromosuccinic acid	+ +	+ + +	+ + +
Glucuronamide	+ +	+ + ++	+ + +
L-alanamide	+ +	+ + + +	+ + +
D-alanine	+ +	+ + ++	+ + +
L-alanine	+ +	+ + +	+ + +
L-alanyl-glycine	+ +	+ + ++	+ + +
L-asparagine	+ +	+ + ++	+ + +
L-asparatic acid	+ +	+ + ++	+ + +
Glycyl-L-asparatic acid	+ +	+ + +	+ + +
Glycyl-L-glutamic acid	+ +	+ + + +	+ + +
D-serine	+ +	+ + ++	+ + +
L-serine	+ +	+ + ++	+ + +
Inosine	+ +	+ + +	+ + +
Uridine	+ +	+ + + +	+ + + +

Thymidine	+ + +	+ + + +	+ + + +
Glycerol	+ +	+ + ++	+ + +
D,L- α- glycerol1-phosphate	+ +	+ + ++	+ + +
D-glucose-6-phosphate	+ + +	+ + + +	+ + + +
α-D-glucose-1-phosphate	+ + +	+ + + +	+ + + +

+ = low, ++ = moderate, +++ = high, ++++ = extremely high

From the previous data, the percentage of highest degree of utilization for each isolate was calculated by dividing the no. of carbon sources showing highest degree of utilization in a given group by the total no. of positive wells for the group for each isolate. *E.coli-EG2* shows 100 % of highest degree of utilization for all carbon groups, in contrast to *E.coli-EG1*which did not show the highest degree for any carbon group. These percentages were calculated for each carbon group for the three isolates in the following table (11):

Table (11): The percentage of highest degree of carbon source utilization for each carbon group among the three isolates:

Group of carbon sources	**Total no. of positive wells for each group**	**% of highest degree of carbon source utilization**		
		E.coli-EG1	***E.coli-EG2***	***E.coli-EG3***
Oligo/polysaccharides	2	0	100	0
Sacchrides/derivatives	17	0	100	35.29
Sugar alcohols/derivatives	4	0	100	0
Amino acids/ degradative products	10	0	100	20
Organic acids/	12	0	100	33. 33

derivatives				
nucleotides	3	0	100	100

The consumption of all carbon sources in all groups was found among all the isolates, but the *E.coli-EG2* isolate showed the highest degree of breakdown for all types of carbon groups and some cases the *E.coli-EG3* isolate shares that highest degree and this behavior pay an attention for the role of the geographical origin on biochemical profile of microorganisms.

3. Serotyping:

The **double diffusion agar assay (Ouchterlony technique)** is based on the principle that diffusion of both antibody and antigen (Hence, double diffusion) through agar can form stable precipitate band and easily observable immune complexes. The solutions diffuse outward, and when antigen and the appropriate antibody meet, they combine and precipitate at the equivalence zone, producing an indicator line. The visible line of precipitation permits a comparison of antigens for identity (Same antigenic determinants), partial identity (cross-reactivity), or nonidentity against a given selected antibody. Here, *E.coli-EG1* and *E.coli-EG3* showed formation of V-shaped line that means identity between these two isolates in their antigenic properties, whilst *E.coli-EG2* showed very weak line, so distant antigenic properties exhibited by that isolate as in fig. (9).

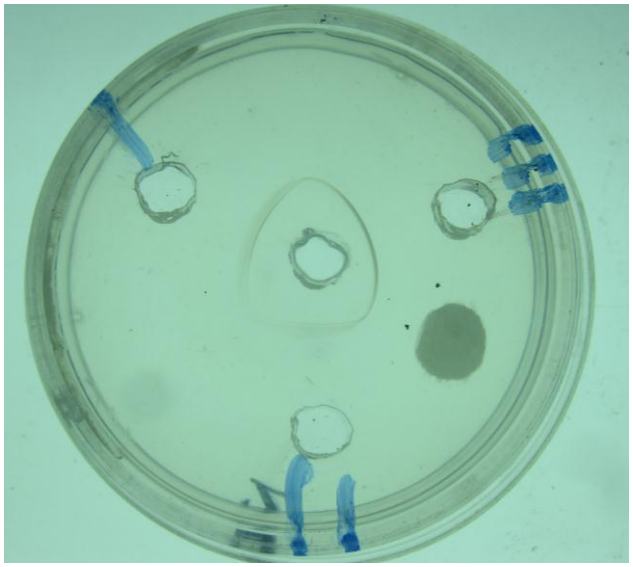


Fig. (9): Double diffusion agar assay showing stable precipitate band . Well (I) for *E.coli-EG1*, well (II) for *E.coli-EG2* and well (III) for *E.coli-EG3*.

4. Phage typing:

Bacteriophage specific for *E.coli* (coliphages) was detected in the water sample using spot test technique, but specificity was restricted to *E.coli-EG1* and *E.coli-EG2* rather than *E.coli-EG3*. Lysis of bacterial cells was demonstrated for those two isolates and plaques count also was evaluated using plaque assay that showed 10^5 and 10^3 for *E.coli-EG1*and *E.coli-EG2* as shown in the following fig. (10):

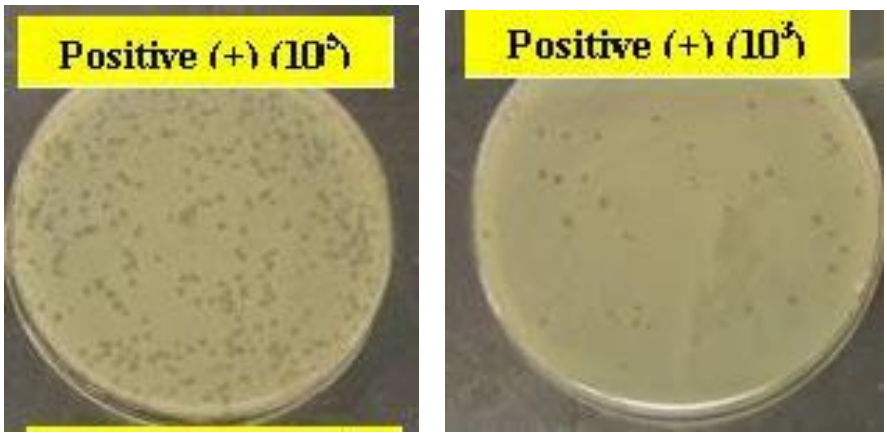


Fig.(10): Plaque assay illustrating the presence of coliphages through lysis of bacterial lawn that could be seen plaques. Plaques for *E.coli-EG1* (A) and *E.coli-EG2* (B) were shown.

Qualitative and quantitative characters that discriminate between the three *E.coli* isolates in terms of their sensitivity to coliphages were summarized in the following table (12):

Table (12): Discrimination between *E.coli* isolates through plaque assay:

***E.coli* isolates**	**Plaque assay**	
	qualitative	**quantitative**
E.coli-EG1	+	10^5
E.coli-EG2	+	10^3
E.coli-EG3	-	-

The binary image (pure black and white) for plaque assay images for both *E.coli* isolates were processed to remove image noises and analyzed numerically via histographic analysis and showed mostly similar values for *E.coli-EG1* and *E.coli-EG2* in mean (185.08 and 186.52), standard deviation (111.43 and 111.27), median (255 and 255) and no. of pixels (214452 and 214928), respectively as in table (13), in contrast to the traditional counting method which showed variation in numbers only regardless of plaque size and intensity which was of exactly great significance. The elementary map, at the same time showed no great difference in elements gray range or counts as in fig (11).

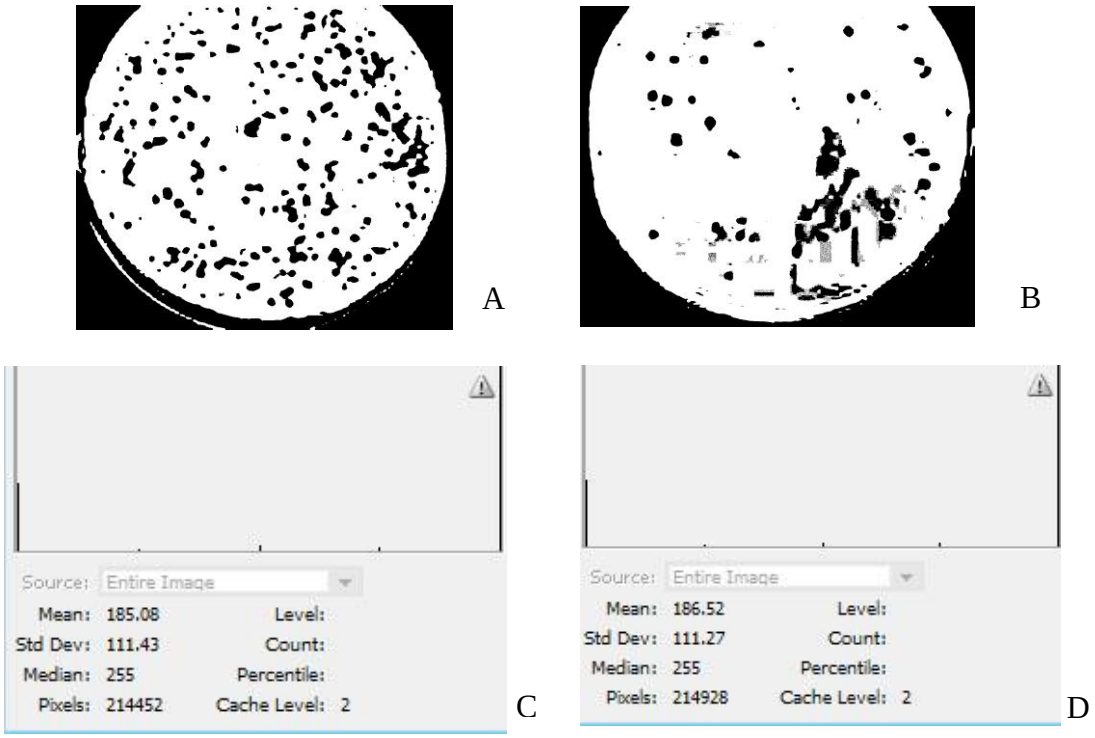


Fig (11): Binary Image processing for plaque assay concerning with *E.coli-EG1*and *E.coli-EG2* (A and B) and histographic analysis (C and D), respectively.

Table (13): Numerical analysis showing variations between *E.coli-EG1* and *E.coli-EG2* during plaque assay:

E.coli isolate	Mean	Standard deviation	Median	Pixels
E.coli-EG1	185.08	111.43	255	214452
E.coli-EG2	186.52	111.27	255	214928

5. Congo red dye agar test:

Out of the three *E.coli* isolates, two *E.coli-EG2* and *E.coli-EG3*were found Congo red binding positive as dark red and red colonies appeared for these two isolates, respectively, while *E.coli-EG1* was thereafter declared negative.

6. Sequence variability of 16s rRNA gene:

6.1. Sequence analysis based on multiple sequence alignment:

The DNA sequences were subjected to variability analysis using DNAMAN and MEGA.4 programmes. The three sequences were aligned for determining the phylogenetic relationships among the three isolates. The sequence alignment was a set of corresponding residues among nucleotide sequences for all the isolates. Aligned residues share evolutionary origin and sequence similarity to some extend. Multiple sequence alignment (MSA) was displayed in which the corresponding nucleotides occupy the same column. When a sequence has no corresponding residue due to deletion event, the position is displayed as "_" which is called a gap. Alignment of multiple genes shows the conserved sites and the percentage of conservation for each position as in fig. (12).

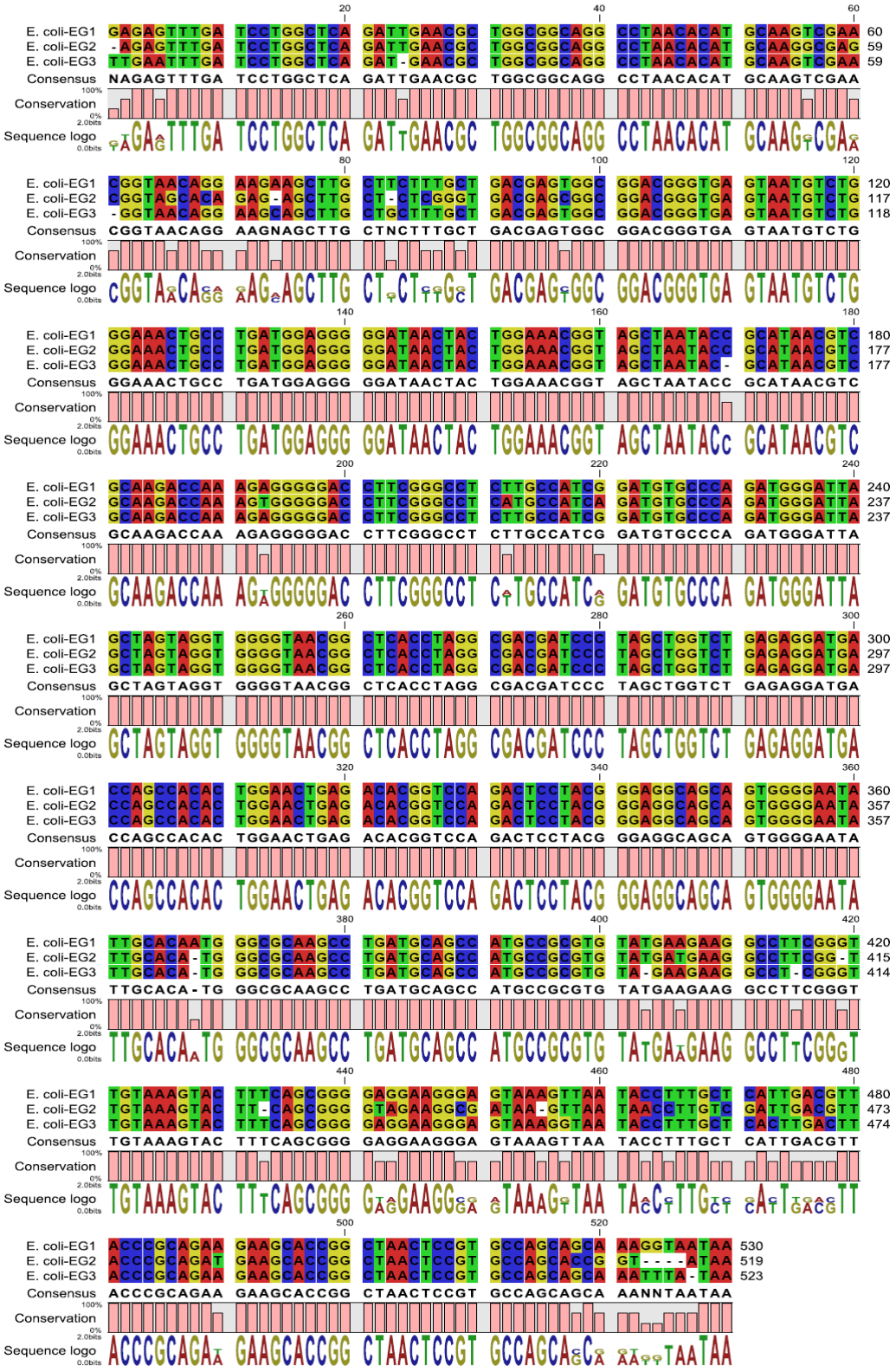


Fig.(12): Multiple sequence alignment of the *E.coli* partial nucleotide sequence of 16s rRNA gene.

Based on MSA analysis performed by MEGA.4 , the phylogenetic tree was performed and shows two clusters in which the *E.coli-EG1* and *E.coli-EG3* were found to be highly homologous with percentage (97%) while the *E.coli-EG3* showed distant homology (93%), so it was represented as a separate cluster as in fig. (13).

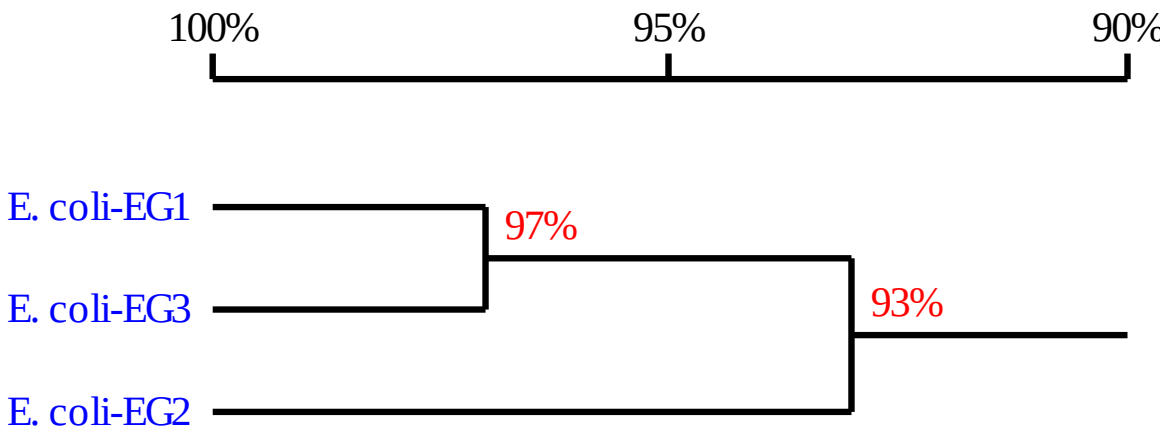


Fig. (13): Phylogenetic tree representing the relationship between the three isolates based DNA sequence homology.

Another model of MSA (performed BY MEGA.5) in which the repeated nucleotides in the same column were written one time and represented as dots "." in other positions.

The gaps and heterologous nucleotides were represented as "-" and symbols for these nucleotides, respectively. These data are represented in fig. (14).

```
E._coli-EG1 GAGAGTTTGA TCCTGGCTCA GATTGAACGC TGGCGGCAGG CCTAACACAT  [ 50]
E._coli-EG2 -......... .......... .......... .......... ..........  [ 50]
E._coli-EG3 TT..A..... .......... ...-...... .......... ..........  [ 50]

E._coli-EG1 GCAAGTCGAA CGGTAACAGG AAGAAGCTTG CTTCTTTGCT GACGAGTGGC  [100]
E._coli-EG2 .....G...G .....G..CA G..-...... ..-..CG.G. ......C...  [100]
E._coli-EG3 .......... -......... ...C...... ..G.......  ..........  [100]

E._coli-EG1 GGACGGGTGA GTAATGTCTG GGAAACTGCC TGATGGAGGG GGATAACTAC  [150]
E._coli-EG2 .......... .......... .......... .......... ..........  [150]
E._coli-EG3 .......... .......... .......... .......... ..........  [150]

E._coli-EG1 TGGAAACGGT AGCTAATACC GCATAACGTC GCAAGACCAA AGAGGGGGAC  [200]
E._coli-EG2 .......... .......... .......... .......... ..T.......  [200]
E._coli-EG3 .......... .........- .......... .......... ..........  [200]

E._coli-EG1 CTTCGGGCCT CTTGCCATCG GATGTGCCCA GATGGGATTA GCTAGTAGGT  [250]
E._coli-EG2 .......... .A.......A .......... .......... ..........  [250]
E._coli-EG3 .......... .......... .......... .......... ..........  [250]

E._coli-EG1 GGGGTAACGG CTCACCTAGG CGACGATCCC TAGCTGGTCT GAGAGGATGA  [300]
E._coli-EG2 .......... .......... .......... .......... ..........  [300]
E._coli-EG3 .......... .......... .......... .......... ..........  [300]

E._coli-EG1 CCAGCCACAC TGGAACTGAG ACACGGTCCA GACTCCTACG GGAGGCAGCA  [350]
E._coli-EG2 .......... .......... .......... .......... ..........  [350]
E._coli-EG3 .......... .......... .......... .......... ..........  [350]

E._coli-EG1 GTGGGGAATA TTGCACAATG GGCGCAAGCC TGATGCAGCC ATGCCGCGTG  [400]
E._coli-EG2 .......... .......-.. .......... .......... ..........  [400]
E._coli-EG3 .......... .......-.. .......... .......... ..........  [400]

E._coli-EG1 TATGAAGAAG GCCTTCGGGT TGTAAAGTAC TTTCAGCGGG GAGGAAGGGA  [450]
E._coli-EG2 .....T.... ....-TC... .......... ..-.......  .TA.....CG  [450]
E._coli-EG3 ..-....... ....-..... .......... .......... ..........  [450]

E._coli-EG1 GTAAAGTTAA TACCTTTGCT CATTGACGTT ACCCGCAGAA GAAGCACCGG  [500]
E._coli-EG2 A...-..... ..A.C...TC G.........  .........T ..........  [500]
E._coli-EG3 ......G... .......... ..C.TGAC.. .......... ..........  [500]

E._coli-EG1 CTAACTCCGT GCCAGCAGCA AAGGTAATAA  [530]
E._coli-EG2 .......... .......C.G GTATA.----  [530]
E._coli-EG3 .......... .......... ..TT..TA.-  [530]
```

Fig. (14): Multiple sequence alignment of the 16s rRNA gene for the three *E.coli* isolates showing the variable sites.

The total number of nucleotide positions is 530 while number of variable sites is 57. All the variable positions and their corresponding nucleotides are derived in table (14).

Table (14): Total Variable sites in the aligned sequences with their position numbers.

Position number in The sequence	Nucleotide		
	E.coli –EG1	***E.coli –EG2***	***E.coli –EG3***
1	G	-*	T
2	A	A	T
5	G	G	A
24	T	T	-
56	T	G	T
60	A	G	A
61	C	C	-
66	A	G	A
69	G	C	G
70	G	A	G
71	A	G	A
74	A	-	C
83	T	-	G
86	T	C	T
87	T	G	T
89	C	G	C
97	T	C	T
170	C	C	-
193	A	T	A
212	T	A	A
220	G	A	G
368	A	-	-
403	T	T	-
406	A	T	A
415	T	-	-
416	C	T	C
417	G	C	G
433	T	-	T
442	A	T	A
443	G	A	G
449	G	C	G
450	A	G	A
451	G	A	G
455	A	-	A
457	T	T	G

463	C	A	C
465	T	C	T
469	C	T	C
470	T	C	T
471	C	G	C
473	T	T	C
475	G	G	T
476	A	A	G
477	C	C	A
478	G	G	C
490	A	T	A
518	G	C	G
520	A	G	A
521	A	G	A
522	A	T	A
523	G	A	T
524	G	T	T
525	T	A	T
527	A	-	T
528	T	-	A
529	A	-	A
530	A	-	-

*** "-" indicates deletions of nucleotides in the corresponding position.**

The variable sites include positions in which only one nucleotide differed in the column and called singletons sites whose number is 42 as in fig. (15).

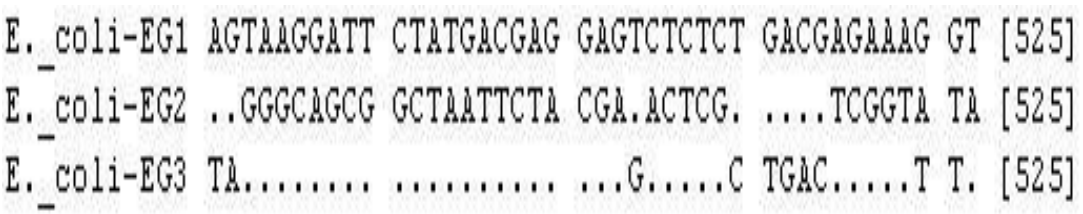


Fig. (15): Total number of singleton sites derived from the aligned sequences.

6.2. Nucleotide sequence statistics:

Sequence analysis showed variability based on sequence alignment, information could be derived by determining the sequence statistics. General information such as length and weight are found in table (15) in which there is correlation between length and weight of DNA sequence for all isolates.

Table (15): Sequence information for the three isolates:

Information	*E. coli-EG1*	*E. coli-EG2*	*E. coli-EG3*
Sequence type	DNA	DNA	DNA
Length	530 nuc	519 nuc	523 nuc
weight	163,270 kDa	159,837 kDa	161,055 kDa

The four types of nucleotides are counted for the three sequences individually and then G+C, A+T, G+C / A+T ratio and percentage of G+C were determined. A distinct similarity is found between isolates *E. coli-EG1* and *E. coli-EG3* in adenine, cytosine, thymine, A+T, G+C / A+T ratio and percentage of G+C whilst *E. coli-EG2* showed far distant relationship except in case of guanine and G+C content. A significant difference is that *E. coli-EG2* showed highest content of cytosine and lowest of thymine. These data are summarized in the following table (16).

Table (16): Counts of nucleotides individually, in combined form, G+C /A+T and % G+C for the three isolates:

Nucleotide	***E. coli-EG1***	***E. coli-EG2***	***E. coli-EG3***
Adenine	140	131	137
Cytosine	119	122	119
Guanine	168	167	165
Thymine	103	99	102
C + G	287	289	284
A+ T	243	230	239
C + G / A+ T	1.181	1.256	1.188
% C + G	54.15	55.68	54.30

The frequency of each nucleotide, G+C and A+T are calculated for each sequence through dividing the number of each nucleotide byte total number of all nucleotides as in table (17).

Table (17): Frequency of nucleotides for the three isolates:

Nucleotide	***E. coli-EG1***	***E. coli-EG2***	***E. coli-EG3***
Adenine	0.264	0.252	0.262
Cytosine	0.255	0.235	0.228
Guanine	0.317	0.322	0.315
Thymine	0.194	0.191	0.195
C + G	0.542	0.557	0.543
A+ T	0.458	0.443	0.457

The percentage of all nucleotides for the three isolates was in a comparative form as in fig. (16).

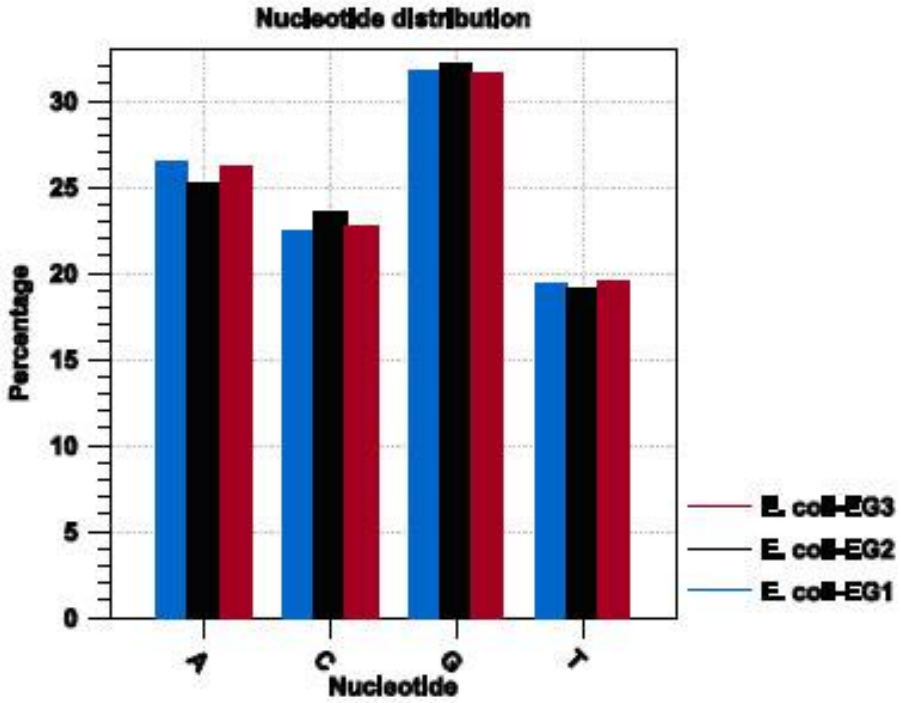


Fig. (16): Histogram of nucleotide frequencies for the three isolates.

The melting temperature (Tm) is the temperature at which the double helix separates and its value depends on G+C content of DNA sequence. In double stranded DNA, three hydrogen bonds join GC base pairs and two bonds connect AT base pairs. As a result DNA with high GC content will have high melting point. The DNA sequence dissolved in sat sodium citrate (SSC) at different concentrations, heated and Tm is recorded at each concentration. The middle value represents the Tm value for each isolate. The highest Tm was found for *E.coli-EG2* corresponding with the highest GC content as recorded before. The Tm values for *E.coli-EG1* and *E.coli-EG*3 were closely related. These data are shown in the following table (18).

Table (18): Melting temperature at different concentrations of salt sodium citrate for the three isolates:

Salt sodium citrate (SSC)	***E. coli-EG1***	***E. coli-EG2***	***E. coli-EG3***
0.1M	87.10	87.73	87.16
0.2M	92.10	92.73	92.16
0.3M	95.02	95.65	95.08
0.4M	97.10	97.72	97.16
0.5M	98.70	99.33	98.77

6.3. Restriction pattern of DNA sequence:

The three DNA sequences are subjected to digestion by restriction enzymes (four types ECONI, ECOP15I, ECO0109I and ECO57MI) that cut DNA at specific recognition sites to make a pattern of DNA fragments. ECONI showed restriction pattern for *E.coli-EG1* and *E.coli-EG3* to be similar (one undigested fragment). In a similar way ECO0109I and ECOP15I showed the same pattern (one undigested fragment and two fragments, respectively). *E. coli-EG2* was dissimilar to some extend that the patterns found for ECONI, ECOP15I and ECO0109I were one undigested, two and one fragment respectively in contrast to the other two isolates. ECO57MI recognize only the sequence of *E. coli-EG2* while had no activity on the other two sequences to support the view that *E. coli-EG2* posses a significant behavior in accordance with the previous data. The DNA band pattern was shown in fig. (17, 18 and 19) and the number and the sizes of DNA bands were recorded in tables (19) for *E. coli-EG1*, *E. coli-EG2* and *E. coli-EG3*, respectively.

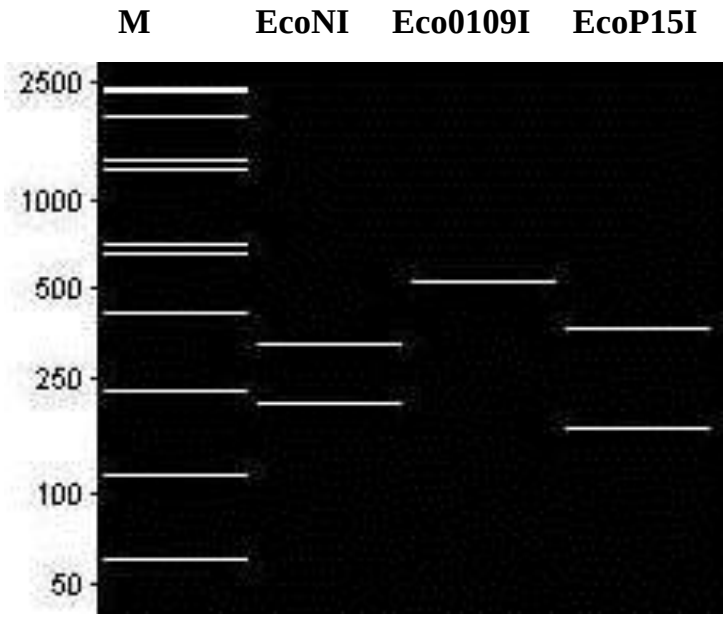


Fig. (17): Predicted electrophoretic DNA band pattern digested by EcoNI, Eco0109I and Ecop15I restriction enzymes for *E.coli-EG1*.

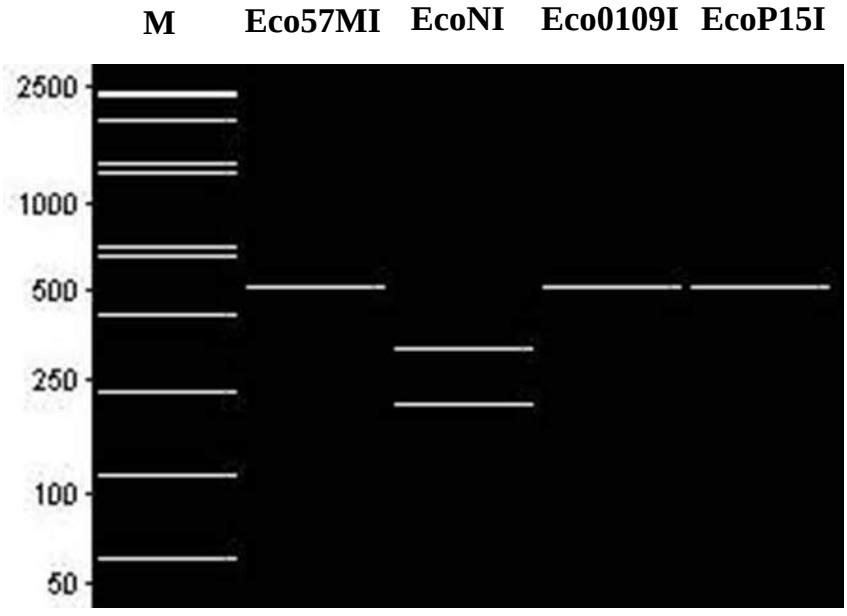


Fig. (18): Predicted electrophoretic DNA band pattern digested by Eco57MI, EcoNI, Eco0109I and EcoP15I restriction enzymes for *E.coli-EG2*.

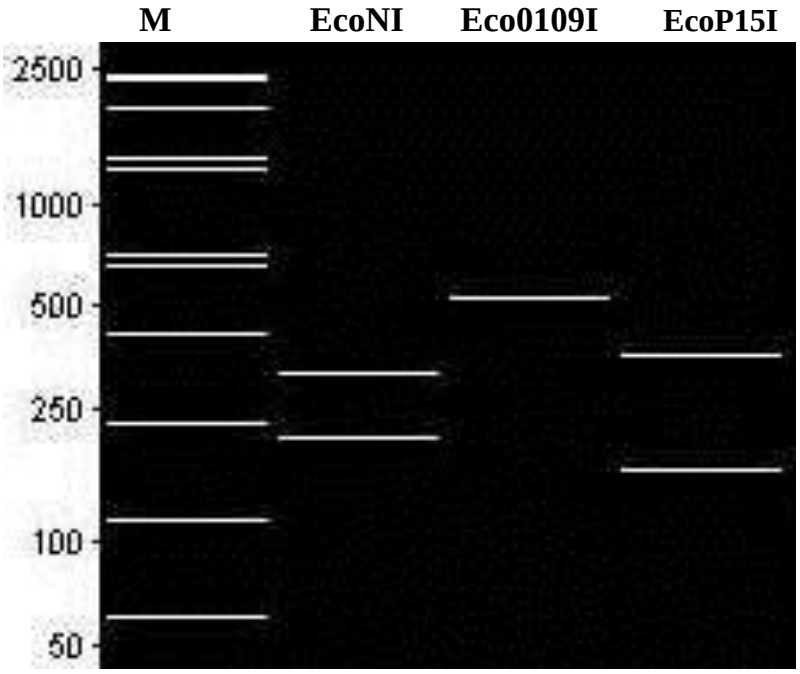


Fig. (19): Predicted electrophoretic DNA band pattern digested by EcoNI, Eco0109I and Ecop15I restriction enzymes for *E.coli-EG3*.

Table (19): Number and size of digested DNA fragments cut by restriction enzymes for the three *E.coli* isolates:

	E.coli-EG1		***E.coli-EG2***		***E.coli-EG3***	
Restriction enzyme	**No. of fragments**	**Size of fragments by base pairs**	**No. of fragments**	**Size of fragments by base pairs**	**No. of fragments**	**Size of fragments by base pairs**
Eco57MI	-	-	1	519	-	-
EcoNI	2	324 , 206	2	313 ,206	2	318 ,205
Eco0109I	1	530	1	519	1	523
Ecop151	2	363 , 167	1	519	2	359 ,164

6.4. 16s rRNA secondary structure prediction:

The 16s rRNA partial gene sequences of *E.coli* isolates were nearly identical, but some regions exhibited significant variations in their primary structure. Secondary structures were inferred on the basis of the primary sequences by using the pairs deduced from secondary-structures base pairing models that conducted by DNAMAN software. Variations in inferred secondary structure and in no. of loops differentiated the tree isolates. The 16s rRNA partial gene sequences were used for prediction of secondary structures .The dominating secondary structure composed of stems (made up of a sequence of nucleotide pairs, A-U or G-C) and loops (no base pairing) .The free energy (dG) required for folding of primary structures into secondary structures were calculated. From the figures (20,21 and 22) the similarity between *E.coli-EG1* and *E.coli-EG3* in their respective RNA secondary structure was elucidated, in contrast to *E.coli-EG2* whose RNA structure was far distant.

E. coli-EG1 **(530 bp)**

dG = -203.10 kcal/mol

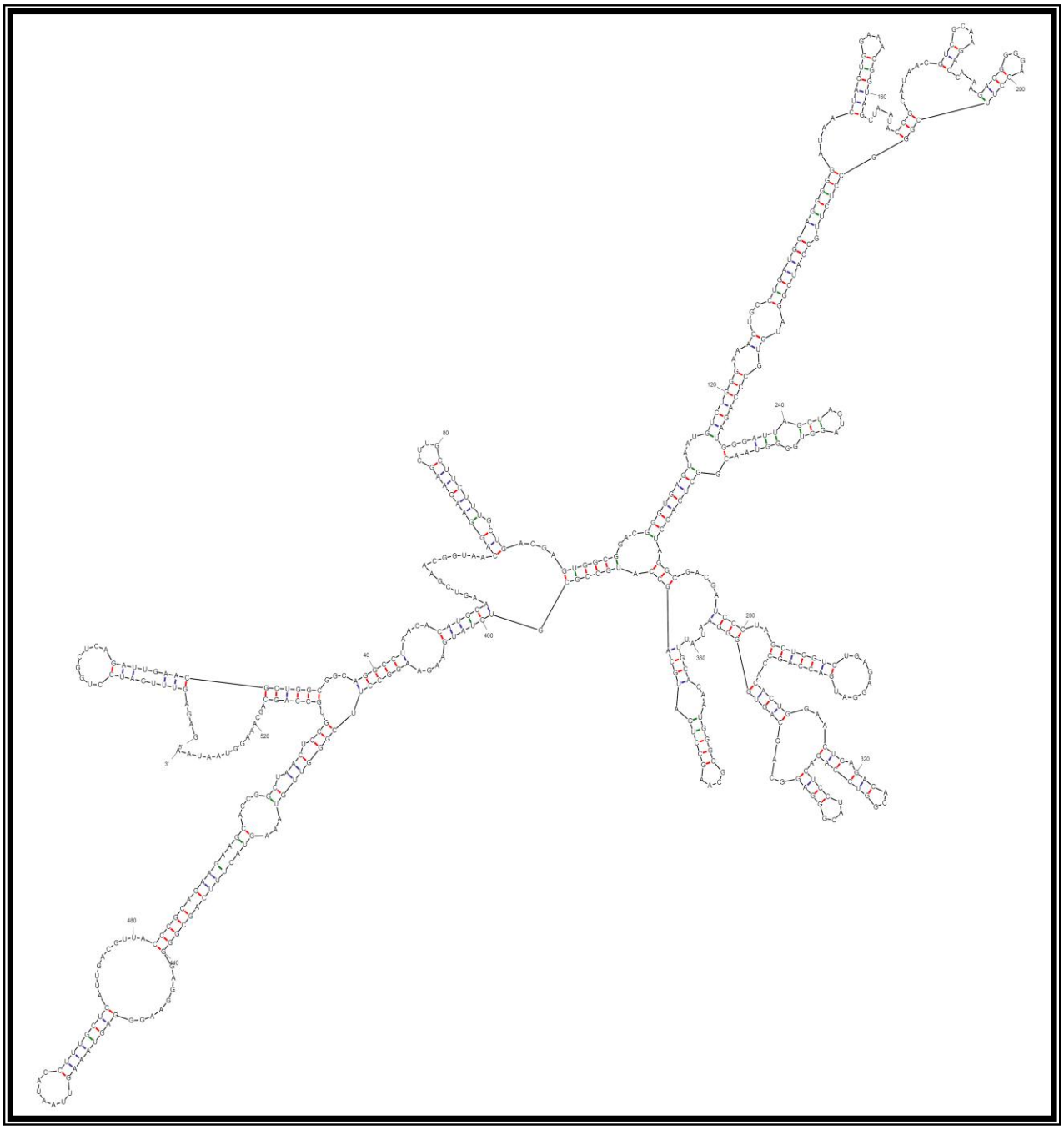


Fig. (20): Secondary structure in *E. coli-EG1* 16s rRNA partial sequence.

***E. coli-EG2* (519 bp)**

dG = -203.70 kcal/mol

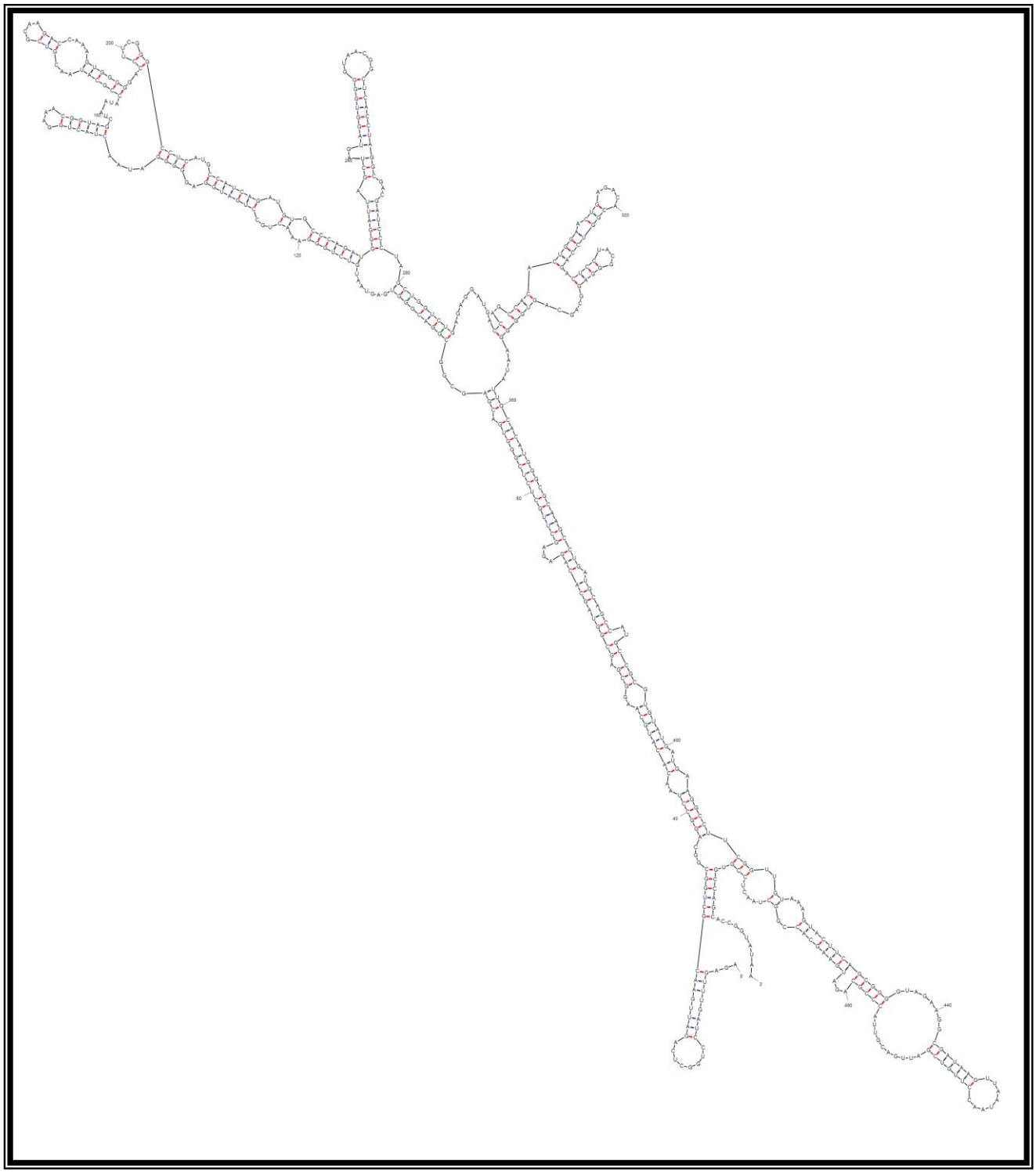


Fig. (21): Secondary structure in *E. coli-EG2* 16s rRNA partial sequence.

***E. coli-EG3* (523 bp)**

dG = -212.50 kcal/mol

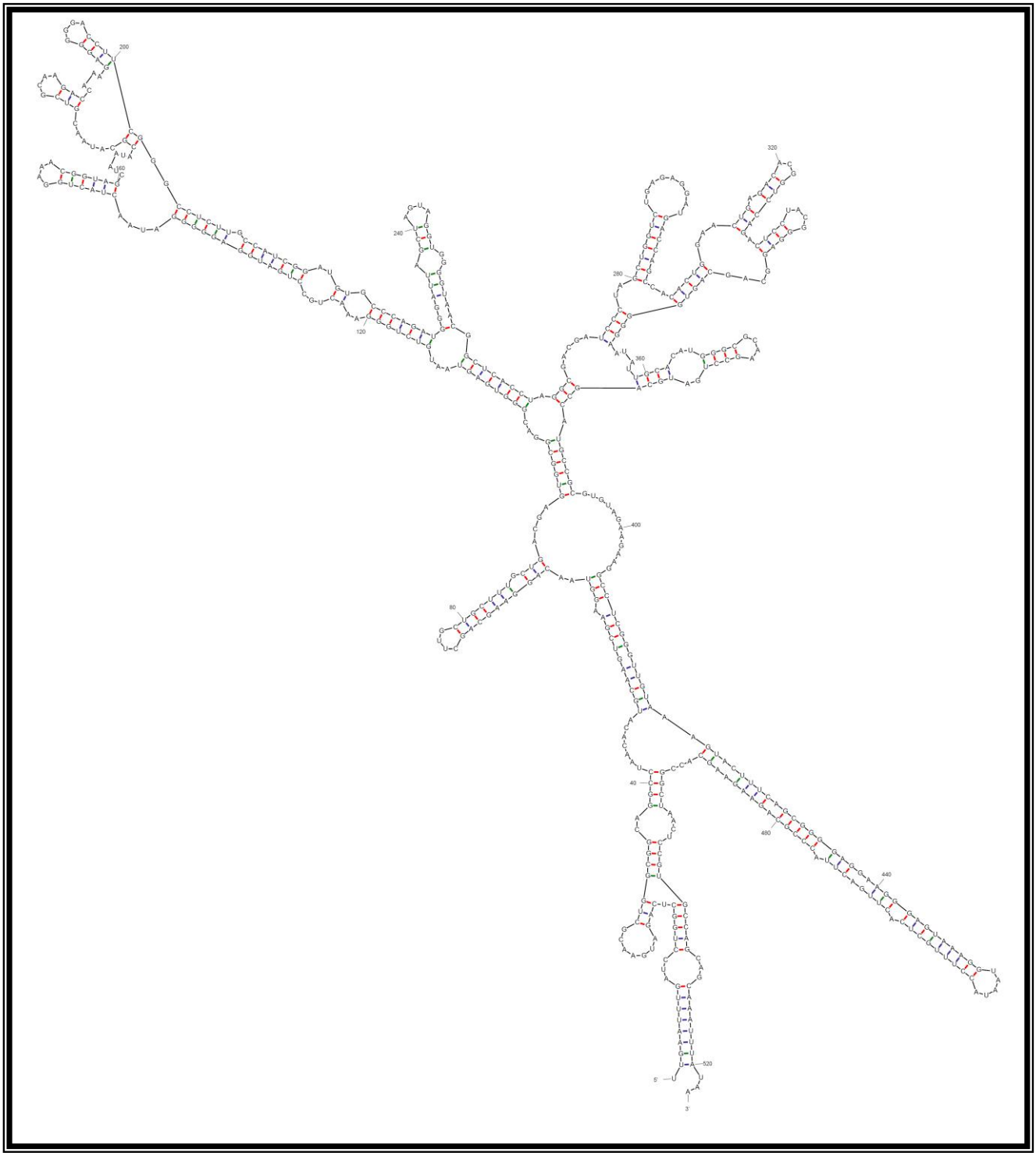


Fig. (22): Secondary structure in *E. coli-EG*3 16s rRNA partial sequence.

6.5. Population genetic analysis from an evolutionary perspective:

Population genetics has always played a central role in evolutionary biology as it deals with the mechanisms by which evolution occurs within populations and species, the ultimate basis of all evolutionary change. Population genetics is concerned with the origin, amount, and distribution of genetic variation present in populations of organisms and the fate of this variation through space and time. The analysis concerned with describing the motive forces for evolution; mutation and the genetic drift. MEGA 5 software was used for evaluating mutational rate through analysis of the three sequences together to determine substitution matrix, natural selection of codons, position by position evolutionary rate for each site, relative synonymous codon usage (RSCU) and Tajima's neutrality test. The substitution mutation among nucleotides within the sequences was described in the following table (20).

Table (20): Maximum Likelihood Estimate of Substitution Matrix:

	A	**T/U**	**C**	**G**
A	-	*5.19*	*6.32*	**15.28**
T/U	*7.10*	-	**9.63**	*8.82*
C	*7.10*	**7.92**	-	*8.82*
G	**12.29**	*5.19*	*6.32*	-

Each entry is the probability of substitution (*r*) from one base (row) to another base (column). Substitution pattern and rates were estimated. Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 100. All positions containing

gaps and missing data were eliminated. There were a total of 514 positions in the final dataset. Evolutionary analyses were conducted in MEGA5. The distance matrix based on pairwise differences among the three sequences was computed as in fig (23).

.

Fig. (23): Pairwise distances matrix for the three sequences:

	1	2	3
1.			
2.	0.065		
3.	0.020	0.080	

The pairwise differences confirmed interrelationships between the three isolates (denoted by 1, 2 and 3 to represent sequences of *E.coli-EG1*, *E.coli-EG2* and *E.coli-EG3*, respectively) in the way that *E.coli-EG1* and *E.coli-EG3* less distant (0.02) whilst *E.coli-EG2* is far distant to *E.coli-EG1* (0.065) and *E.coli-EG3* (0.080), respectively.

Analysis of natural selection codon by codon:

For each codon, estimates of the numbers of inferred synonymous (s) and nonsynonymous (n) substitutions are presented along with the numbers of sites that are estimated to be synonymous (S) and nonsyonymous (N). These estimates are produced using the joint Maximum Likelihood reconstructions of ancestral states of nucleotide substitution. The test statistic dN - dS is used for detecting codons that have undergone positive selection, where dS is the number of synonymous substitutions per site (s/S) and dN is the number of nonsynonymous substitutions per site (n/N). A positive value for the test statistic indicates an overabundance of nonsynonymous substitutions. In

this case, the probability of rejecting the null hypothesis of neutral evolution (P-value) was calculated.Values of P less than 0.05 are considered significant at a 5% level and are highlighted. Normalized dN - dS for the test statistic is obtained using the total number of substitutions in the tree (measured in expected substitutions per site). It is useful for making comparisons across data sets. Maximum Likelihood computations of dN and dS were conducted using HyPhy software package .The analysis involved 3 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 153 positions in the final dataset as in table (21). Evolutionary analyses were conducted in MEGA5.

Table (21): Analysis of natural selection codon by codon:

Codon#	Triplet	Syn (s)	Nonsyn (n)	Syn sites (S)	Nonsyn sites (N)	dS	dN	dN-dS	P-value	Normalized dN-dS
1	GAG	0	1	0.248069	2.75193	0	0.3634	0.36338	0.9173	5.02758
2	AGT	0	0	0.639785	2.04072	0	0	0	0	0
3	TTG	0	0	0.567568	2.43243	0	0	0	0	0
4	ATC	0	0	1.24807	1.75193	0	0	0	0	0
5	CTG	0	0	1	2	0	0	0	0	0
6	GCT	0	0	0.391716	2.36021	0	0	0	0	0
7	CAG	0	0	0.428298	2.26756	0	0	0	0	0
8	ATT	0	0	1	2	0	0	0	0	0
9	GAA	0	0	0	2.21657	0	0	0	0	0
10	CGC	0	0	1.3195	1.6805	0	0	0	0	0
11	TGG	0	0	0.391716	2.36021	0	0	0	0	0
12	CGG	0	0	1	2	0	0	0	0	0
13	CAG	0	0	0.248069	2.75193	0	0	0	0	0
14	GCC	0	0	0	3	0	0	0	0	0
15	TAA	0	0	0.428298	2.32363	0	0	0	0	0
16	CAC	0	1	1	2	0	0.5	0.5	0.6667	6.91777
17	ATG	1	0	0.413357	2.2825	2.4192	0	-2.4192	1	-33.4712
18	CAA	0	1	0.333046	2.5202	0	0.3968	0.39679	0.8833	5.48985
19	GTC	0	2	0.594899	2.12483	0	0.9413	0.94125	0.6104	13.0227
20	GAA	0	0	1	2	0	0	0	0	0
21	CGG	0	0	0.248069	2.43243	0	0	0	0	0
22	TAA	0.5	1.5	0.608512	2.261	0.8217	0.6634	-0.1583	0.7879	-2.18951
23	CAG	0	1	1	2	0	0.5	0.5	0.6667	6.91777
24	GAA	0	0	0.248069	2.75193	0	0	0	0	0
25	GAA	0	0	0.391716	2.30414	0	0	0	0	0
26	GCT	0	1	0.538925	1.99762	0	0.5006	0.5006	0.7875	6.92601
27	TGC	0	0	1.3195	1.6805	0	0	0	0	0

28	TTC	0	0	1	2	0	0	0	0	0
29	TTT	0	0	1	2	0	0	0	0	0
30	GCT	0	0	0.391716	2.30414	0	0	0	0	0
31	GAC	0	0	0.248069	2.43243	0	0	0	0	0
32	GAG	0	0	1.24807	1.75193	0	0	0	0	0
33	TGG	0	0	1	1.69586	0	0	0	0	0
34	CGG	0	0	0.248069	2.75193	0	0	0	0	0
35	ACG	0	0	0.248069	2.43243	0	0	0	0	0
36	GGT	0	0	1.24807	1.75193	0	0	0	0	0
37	GAG	0	0	0	3	0	0	0	0	0
38	TAA	0	0	0.391716	2.30414	0	0	0	0	0
39	TGT	0	0	1	2	0	0	0	0	0
40	CTG	0	0	0.248069	2.75193	0	0	0	0	0
41	GGA	0	0	0.248069	2.75193	0	0	0	0	0
42	AAC	0	0	0.248069	2	0	0	0	0	0
43	TGC	0	0	0	2.21657	0	0	0	0	0
44	CTG	0	0	0.428298	2.326	0	0	0	0	0
45	ATG	0	0	1.3195	1.6805	0	0	0	0	0
46	GAG	0	0	1.24807	1.75193	0	0	0	0	0
47	GGG	0	0	0.571702	2.4283	0	0	0	0	0
48	GAT	0	0	0.248069	2.75193	0	0	0	0	0
49	AAC	0	0	0.248069	2.75193	0	0	0	0	0
50	TAC	0	0	1	2	0	0	0	0	0
51	TGG	0	0	1	2	0	0	0	0	0
52	AAA	0	0	0.754302	2	0	0	0	0	0
53	CGG	0	0	1	2	0	0	0	0	0
54	TAG	0	0	0.391716	2.36259	0	0	0	0	0
55	CTA	0	1	0.42458	2.25544	0	0.4434	0.44337	0.8416	6.1343
56	ATA	0	0	1	2	0	0	0	0	0
57	CCG	0	0	1	2	0	0	0	0	0
58	CAT	0	0	0.248069	2.75193	0	0	0	0	0
59	AAC	0	0	1	2	0	0	0	0	0
60	GTC	0	0	1	2	0	0	0	0	0
61	GCA	0	1	0.692888	2.30711	0	0.4334	0.43344	0.769	5.99691
62	AGA	0	0	1	2	0	0	0	0	0
63	CCA	0	0	0.567568	2.43243	0	0	0	0	0
64	AAG	0	1	0.899649	1.82008	0	0.5494	0.54943	0.6692	7.60162
65	AGG	0	0	0.248069	2.43243	0	0	0	0	0
66	GGG	0	0	1	2	0	0	0	0	0
67	ACC	0	0	0.391716	2.36021	0	0	0	0	0
68	TTC	0	0	0	3	0	0	0	0	0
69	GGG	0	0	1	1.69586	0	0	0	0	0
70	CCT	0	0	0.676368	1.5717	0	0	0	0	0
71	CTT	0	0	1	2	0	0	0	0	0
72	GCC	0	0	0.248069	2.75193	0	0	0	0	0
73	ATC	0	0	0.71772	2.28228	0	0	0	0	0
74	GGA	0	0	0	2.21657	0	0	0	0	0
75	TGT	0	0	1	2	0	0	0	0	0
76	GCC	0	0	0.248069	2.75193	0	0	0	0	0
77	CAG	0	0	1	2	0	0	0	0	0
78	ATG	0	0	1	1.24807	0	0	0	0	0

79	GGA	0	0	1	2	0	0	0	0	0
80	TTA	0	0	0.71772	2.28228	0	0	0	0	0
81	GCT	0	0	1.3195	1.43243	0	0	0	0	0
82	AGT	0	0	1.3195	1.43243	0	0	0	0	0
83	AGG	0	0	1	2	0	0	0	0	0
84	TGG	0	0	1.24807	1.75193	0	0	0	0	0
85	GGT	0	0	1	2	0	0	0	0	0
86	AAC	0	0	1	2	0	0	0	0	0
87	GGC	0	0	1.24807	1.75193	0	0	0	0	0
88	TCA	0	0	0.754302	2	0	0	0	0	0
89	CCT	0	0	1	1.69586	0	0	0	0	0
90	AGG	0	0	1	2	0	0	0	0	0
91	CGA	0	0	1	2	0	0	0	0	0
92	CGA	0	0	1	2	0	0	0	0	0
93	TCC	0	0	1.24807	1.75193	0	0	0	0	0
94	CTA	0	0	0.428298	2.26756	0	0	0	0	0
95	GCT	0	0	1.24807	1.75193	0	0	0	0	0
96	GGT	0	0	0.754302	2	0	0	0	0	0
97	CTG	0	0	0.248069	2.75193	0	0	0	0	0
98	AGA	0	0	1	2	0	0	0	0	0
99	GGA	0	0	1	2	0	0	0	0	0
100	TGA	0	0	0.248069	2.75193	0	0	0	0	0
101	CCA	0	0	1	2	0	0	0	0	0
102	GCC	0	0	0.248069	2	0	0	0	0	0
103	ACA	0	0	1	2	0	0	0	0	0
104	CTG	0	0	0.71772	2.28228	0	0	0	0	0
105	GAA	0	0	0.391716	2.36021	0	0	0	0	0
106	CTG	0	0	0.391716	2.36021	0	0	0	0	0
107	AGA	0	0	0	2.21657	0	0	0	0	0
108	CAC	0	0	1	1.69586	0	0	0	0	0
109	GGT	0	0	0.571702	2.4283	0	0	0	0	0
110	CCA	0	0	0.639785	2.04072	0	0	0	0	0
111	GAC	0	0	0.248069	2.75193	0	0	0	0	0
112	TCC	0	0	1	2	0	0	0	0	0
113	TAC	0	0	1	2	0	0	0	0	0
114	GGG	0	0	0.391716	2.36259	0	0	0	0	0
115	AGG	0	0	1	2	0	0	0	0	0
116	CAG	0	0	0.248069	2.75193	0	0	0	0	0
117	CAG	0	0	1	2	0	0	0	0	0
118	TGG	0	0	1	2	0	0	0	0	0
119	GGA	0	0	0	3	0	0	0	0	0
120	ATA	0	0	1	2	0	0	0	0	0
121	TTG	0	0	1	2	0	0	0	0	0
122	CAC	0	0	1	2	0	0	0	0	0
123	AAT	0	0	0.71772	2.28228	0	0	0	0	0
124	GGG	0	0	1	2	0	0	0	0	0
125	CGC	0	0	0.248069	2.43243	0	0	0	0	0
126	AAG	0	0	0.428298	2.326	0	0	0	0	0
127	CCT	0	0	1	2	0	0	0	0	0
128	GAT	0	0	1	2	0	0	0	0	0
129	GCA	0	0	1	2	0	0	0	0	0

130	GCC	0	0	1	2	0	0	0	0	0
131	ATG	0	0	0.391716	2.36259	0	0	0	0	0
132	CCG	1	1	1	1.86149	1	0.5372	-0.4628	0.8779	-6.40302
133	CGT	0	1	0.825069	2.17493	0	0.4598	0.45979	0.725	6.36137
134	GTA	0	1	0.700201	1.60324	0	0.6237	0.62374	0.696	8.62973
135	TGA	0	0	0.571702	2.4283	0	0	0	0	0
136	AGA	1	1	1	2	1	0.5	-0.5	0.8889	-6.91777
137	AGG	0	0	0.639785	2.04072	0	0	0	0	0
138	CCT	1	2	0.941181	1.97182	1.0625	1.0143	-0.0482	0.7543	-0.666955
139	TCG	0	1	0.599413	2.40059	0	0.4166	0.41657	0.8002	5.7634
140	GGT	0	1	1	2	0	0.5	0.5	0.6667	6.91777
141	TGT	0	0	1	2	0	0	0	0	0
142	AAA	0	0	1	2	0	0	0	0	0
143	GTA	0	0	0.754302	2	0	0	0	0	0
144	CTT	0	0	0.248069	2.75193	0	0	0	0	0
145	TCA	0	0	1	2	0	0	0	0	0
146	GCG	0	0	1	2	0	0	0	0	0
147	GGG	0	0	1	2	0	0	0	0	0
148	AGG	0	0	1	2	0	0	0	0	0
149	AAG	0	0	1	2	0	0	0	0	0
150	GGA	0	0	0.248069	2.75193	0	0	0	0	0
151	GTA	0	1	0.555182	2.44482	0	0.409	0.40903	0.8149	5.65913
152	AAG	0.3333	2.66667	0.569352	2.27734	0.5855	1.171	0.58549	0.64	8.10062
153	TTA	1	3	0.769132	2.23087	1.3002	1.3448	0.0446	0.7275	0.617087

S = synonymous n= nonsynonymous S= synonymous site
N = nonsynonymous site dS = s/S dN =n/N

Position by position evolutionary rates:

Mean (relative) evolutionary rate are shown for each site next to the site number. These rates are scaled such that the average evolutionary rate across all sites is 1. This means that sites showing a rate < 1 are evolving slower than average and those with a rate > 1 are evolving faster than average. These relative rates were estimated .A discrete Gamma (+G) distribution was used to model evolutionary rate differences among sites (5 categories). The probability of classification of a site in each discrete rate category in Gamma is shown. Mean relative evolutionary rates in discrete Gamma categories are shown in the column headers. The analysis involved 3 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 514 positions in the final dataset.

Evolutionary analyses were conducted in MEGA5. The polymorphic sites were categorized in group 5 whilst all other sites were in 5 (see appendix II).

Relative synonymous codon usage (RSCU):

Due to the degeneracy of genetic code, most amino acids are coded by more than one codon (synonymous codon). Studies of the synonymous codon usage can reveal information about the molecular evolution of individual genes, but here the partial sequence of 16s rRNA gene did not code for a protein, but for rRNA, so the RSCU affect secondary structure and so function of such gene.

RSCU of different codons in each gene sample have been calculated. The relative synonymous codon usage value of the jth codon for the ith amino acid is calculated as below **(Sharp and Li, 1986**).

$$RSCUij = \frac{(obsij)/(\_SUM\ of\ all\ nij\ obsij)}{1/ni}$$

In formula, obsij is the observed number of the jth codon for the ith amino acid, which has ni type of synonymous codons. The highest characteristic codons used in *E-coli-EG1* partial gene sequence were CUG, ACC, GUA, CCU and UAC while for *E-coli-EG2* UUC, CUG, GUA, ACC, UAC and AAC were used and for *E-coli-EG3* CUG, GUA, CCU, ACC, UAC and AGC were also used.

Table (22): Relative synonymous codon usage for the three partial sequences:

Codon	*E-coli-EG1*		*E-coli-EG2*		*E-coli-EG3*	
	Count	**RSCU**	**Count**	**RSCU**	**Count**	**RSCU**
UUU	1	0.67	0	0	2	1.33
UUC	2	1.33	1	2	1	0.67
UUA	2	0.71	2	0.8	1	0.33
UUG	3	1.06	3	1.2	4	1.33
CUU	2	0.71	1	0.4	3	1
CUC	2	0.71	1	0.4	2	0.67
CUA	2	0.71	2	0.8	2	0.67
CUG	6	2.12	6	2.4	6	2
AUU	2	0.86	2	0.67	0	0
AUC	2	0.86	2	0.67	2	1.2
AUA	3	1.29	5	1.67	3	1.8
AUG	4	1	4	1	4	1
GUU	1	0.67	1	1	0	0
GUC	2	1.33	1	1	2	1.33
GUA	3	2	2	2	4	2.67
GUG	0	0	0	0	0	0
UCU	0	0	0	0	0	0
UCC	2	1.2	2	1.2	2	1.5
UCA	2	1.2	1	0.6	2	1.5
UCG	1	0.6	3	1.8	0	0
CCU	5	2	4	1.78	5	2.22
CCC	0	0	0	0	0	0
CCA	3	1.2	3	1.33	3	1.33
CCG	2	0.8	2	0.89	1	0.44
ACU	0	0	0	0	1	0.67
ACC	3	2.4	5	2.86	3	2
ACA	1	0.8	1	0.57	1	0.67
ACG	1	0.8	1	0.57	1	0.67
GCU	5	1.43	4	1.14	5	1.33
GCC	6	1.71	6	1.71	6	1.6
GCA	2	0.57	2	0.57	3	0.8
GCG	1	0.29	2	0.57	1	0.27
UAU	0	0	0	0	0	0
UAC	2	2	2	2	2	2
UAA	4	1.71	3	0.9	4	1.71
UAG	1	0.43	3	0.9	1	0.43
CAU	1	0.5	2	0.67	1	0.5

CAC	3	1.5	4	1.33	3	1.5
CAA	1	0.29	1	0.33	1	0.29
CAG	6	1.71	5	1.67	6	1.71
AAU	2	0.67	0	0	1	0.4
AAC	4	1.33	4	2	4	1.6
AAA	3	0.86	2	0.8	3	0.86
AAG	4	1.14	3	1.2	4	1.14
GAU	2	0.8	2	0.8	2	1
GAC	3	1.2	3	1.2	2	1
GAA	5	1.11	2	0.67	4	1.14
GAG	4	0.89	4	1.33	3	0.86
UGU	3	1.2	3	1.2	3	1
UGC	2	0.8	2	0.8	3	1
UGA	2	0.86	4	1.2	2	0.86
UGG	5	1	5	1	5	1
CGU	2	0.52	2	0.55	2	0.55
CGC	3	0.78	3	0.82	3	0.82
CGA	2	0.52	2	0.55	2	0.55
CGG	4	1.04	5	1.36	3	0.82
AGU	2	1.2	2	1.2	1	0.75
AGC	3	1.8	2	1.2	3	2.25
AGA	6	1.57	6	1.64	6	1.64
AGG	6	1.57	4	1.09	6	1.64
GGU	6	1.2	6	1.26	5	1.05
GGC	2	0.4	3	0.63	2	0.42
GGA	6	1.2	4	0.84	6	1.26
GGG	6	1.2	6	1.26	6	1.26
Average codons	176		168		169	

Tajima's Neutrality Test:

The evolution of living organisms is the consequence of two processes. First, evolution depends on the genetic variability generated by mutations (which were described) which continuously arise within populations. Second, it also relies on changes in the frequency of alleles within populations over time. That is why the Tajima's test was performed for inspection of whether occurrence of neutrality between such two processes performed the forcing mechanisms for evolutionary change or it would be the role of natural selection by mutation.

Table (23): Results from Tajima's Neutrality Test:

m	***S***	***p_s***	***Θ***	***π***	***D***

3	40	0.077821	0.051881	0.052529	3332176.416214

Abbreviations: m = number of sites, S = Number of segregating sites, ps = S/m, Θ = the population genetics parameter, π = nucleotide diversity, and D is the Tajima test statistic.

The Tajima's D test is a widely used test of neutrality in population genetics. This statistic illustrates the allele frequency distribution of nucleotide sequence and is based on the difference between two estimators of (the population mutation rate): (1) Tajima's estimator, which is based on the average number of pairwise difference between sequences, and (2) Watterson's estimator, which is based on the number of segregating sites in the sample. Tajimas estimator takes into account allele frequency when comparing pairwise differences between sequences whereas Watterson's estimator does not. A segregating site counts as any point where there are differing nucleotides between sequences in the data set, independent of the actual number of differences and hence independent of allele frequency. The difference between these two estimators of is scaled by the standard deviation of their difference. A positive value of D as in our study indicates an excess of intermediate frequency (polymorphic) alleles, while a negative value indicates an excess of rare alleles. There was equivalence (confirmed by nearly equivalence between Θ (0.051881) and π (0.052529)) between genetic drift and mutation. The equivalence between such two values and the positive value of D considered the null hypothesis to be held.

7. Nanoscale evaluation of biodiversity:

7.1. Using image processing:

The three *E. coli* isolates were subjected to TEM imaging (which showed morphological variations concerning with shape in size such away that *E. coli- EG2* was elongated than others with intense external

appendages) and showed white and black photo micrograph constituting the internal cellular structures. The image at such mode "white and black" showed graduation of color intensity which was expressed as Numerical Data represented by Pixels, Mean, standard deviation, and Median and such mode referred to grey level. The individual components of the cell were analyzed via histographic analysis to convert the image into Digital Data. Those organelles were flagella, nuclear body, cytoplasm, bacterial envelope (Three – layered) and fimbrae. The Numerical analysis refers to the quantity and frequency of each component in each bacterial isolate. The peak in each histographic analysis corresponds to an individual component in such way that its quantity and frequency can be provided through the Numerical data belongs to such peaks.

For better visualization of the image, the grey level is transformed into color image via using RGB mode that is consists of Red, Green and blue filters, these three filters are combined into single composite color image.This mode allows color range for Pixels from O- 016777215 (numbers of colors: 256 x 256 x 256).

The RGB mode of Image allows specification of colors as colors are related to structures as shown in colorized images that flagella (Fl) , fimbrae (Fm), Envelope (Env) which is three layered (Om (outer membrane), Cw (cell wall) and (inner membrane) Im) , cytoplasm (Cyt) and Nuclear body (Nu) took green , green to yellow , (composite color) blue and red respectively .The numerical expression of each organelle was provided By Pixels values as Pixels are image elements. The quantity of all components divided by thus total number gives mean values.

The histogram for each *E.coli* isolate occupied most possible grey level values as in (plate I, II and III), in addition to that, the histogram at

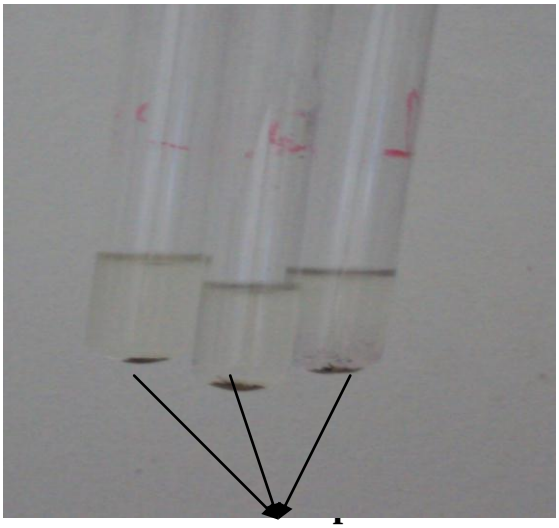
RGB level showed nearly uniformity in the range, but at different numerical characters. The flagellar elementary map for all isolates were at lighter side differentially in such away that *E. coli- EG3* mean values (209.63) was larger than *E. coli- EG1*(192.01) and *E. coli- EG2* (204.53).at RGB level, the composite color image always showed separation of structuring elements (SE) and occupied wider range than that of monochromatic grey level image. For flagella, at RGB level, higher mean values for *E. coli- EG1* (142.17) as compared to *E. coli- EG2* (136.79) and *E. coli- EG3* (139.90) were found. The elementary map of NU, ENV and CYT at grey level showed highest mean values for *E. coli- EG3*(30.91, 142.74 and 73.72) as compared to *E. coli- EG1*(3.55, 89.29 and 33.17) and *E. coli- EG2*(1.85, 51.53 and 13.92),respectively as in (plate I, II and III) and table (24). At RGB level, the mean values were found in the same order for NU, but for ENV, *E. coli- EG2* exhibited highest level (136.50) and for CYT, *E. coli- EG1*(160.27) shown to be the highest. *E. coli- EG2* was characterized by highest mean values (203.83, 148.16 at grey and RGB level, respectively) for FM as compared with other isolates. The intensities of SEs were analysed for standard deviation and median characterization as in table (25 and 26). The total number of pixels constituting SEs in each elementary map was counted as in table (27).The color intensity (level), counts and percentile changes from peak to peak in the same map.

7.3. Biosynthesis of gold nanoparticles (GNPs) by *E.coli* isolates:

7.3.1. EM characterizations of the bio-nanocomposites:

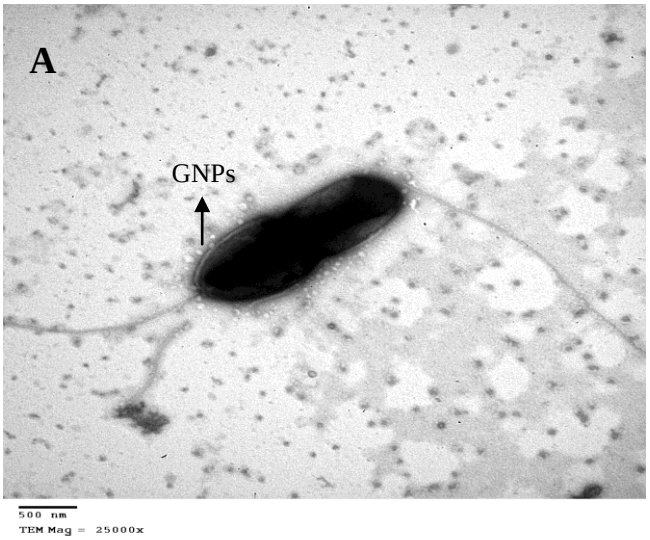
The bioreduction of gold ions and its accumulation into nanoparticles on the surface of *E.coli* cells was investigated by TEM. The suspended yellow gold ions incubated with the three *E.coli* isolates showed changes in color and solubility. During te time course of incubation, successive events including bioreduction, nucleation and growth of GNPs occurred and ended with the formation of nanocomposite between NPs and bacterial surfaces or envelope extensions as in fig (25). The size and morphology of NPS and nature of the role of bacteria in synthesis and interaction with GNPs were revealed by TEM and image processing operations. As seen from fig (26. A and 27.A) the bright spots represented GNPs were accumulated on bacterial surface, a matter that was not found for fig (28.A). The accumulation on bacterial surface observed reflected the nature of NPs binding and chemical composition of bacterial envelope to direct the synthetic process. Based on previous studies biosynthesis of GNPs might occur within envelope or intracellular, multi-step image processing operations were performed to reveal such event via transforming the grey level image into RGB level. A great significance had been made by such operations in such away that the blue NPs integrated within bacterial envelope while another red ones accumulated on the surface. This color variation of GNPs depended on size or otherwise on the stage of NPs growth. The larger particles exhibited red color while blue ones were smaller in size and integrated in the envelope. The correlation between particle size and its color is a good property that might be utilized to solve the problem of controlling particles size.

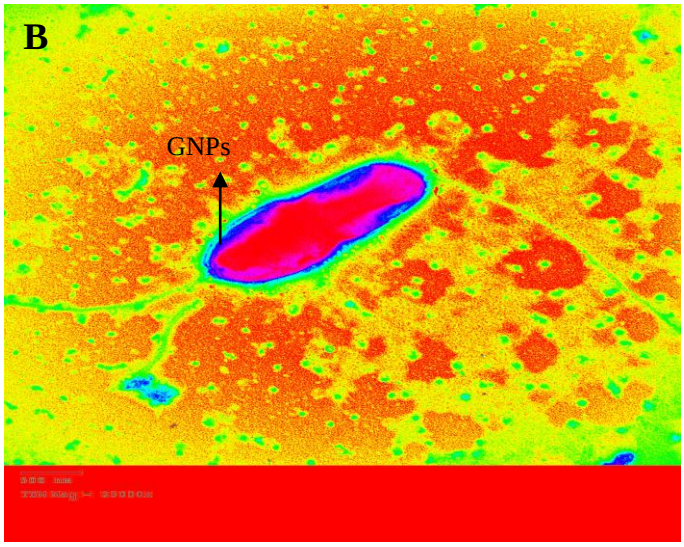
The biotransformation extend mediated by *E.coli* isolates is variable and different particles shapes can be revealed in TEM. The GNPs appeared in spheres, irregular and seldom triangles. The gold nanoparticles are mostly spherical. The actual sizes of such particles could be derived via the scale bar. The mean range of size for GNPs synthesized by *E.coli-EG1* was 20-30 nm, 10-20 nm for *E.coli-EG2* and 20-25 nm for *E.coli-EG3*. The smaller sized particles, the better the electrical-optical properties they acquire, a fact that achieved by *E.coli-EG2* bioreduction process. By applying Edge detection operation, the intercalation of the gold nanospheres and the surrounding envelope of the bacterial isolates appeared and were valuable in understanding the nature of the synthetic process. The grey level, the RGB level and edge detection of *E.coli-EG1*, *E.coli-EG2* and *E.coli-EG3* are shown in fig. (26, 27 and 28) respectively.



Bio – nanocomposite

Fig. (25): The three *E.coli* isolates incubated with distilled water containing gold ions revealed bio-nanocomposites that are settled as slight red precipitates.





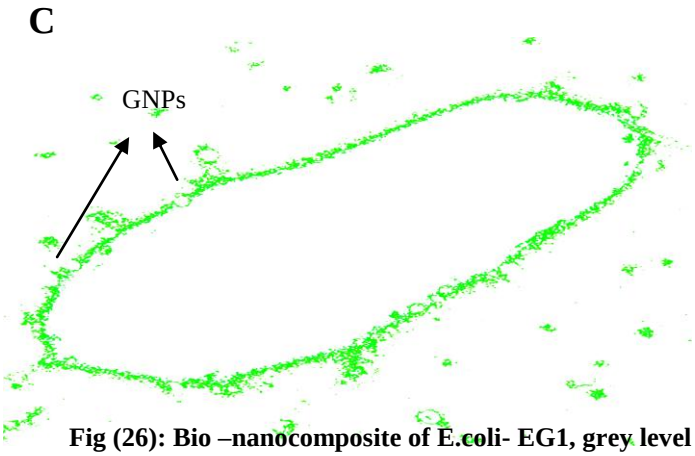
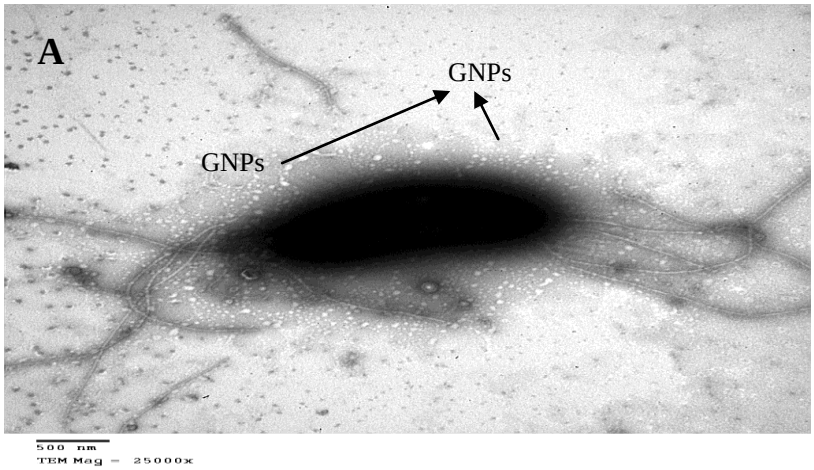
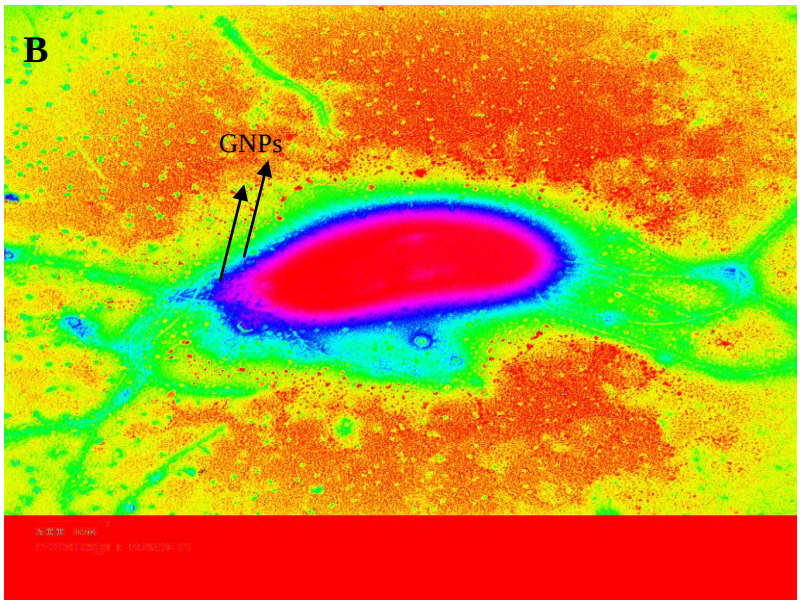


Fig (26): Bio –nanocomposite of E.coli- EG1, grey level (A), at RGB level (B) and Edge detection (C)





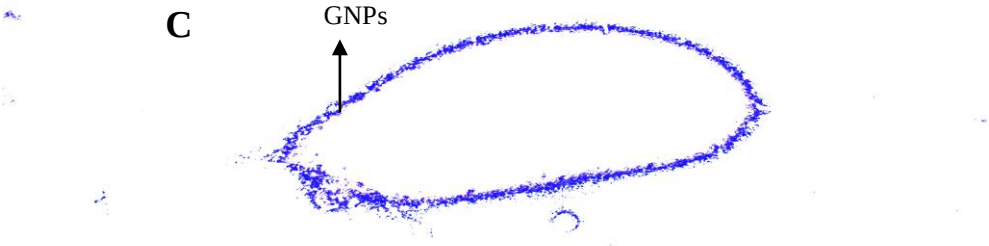
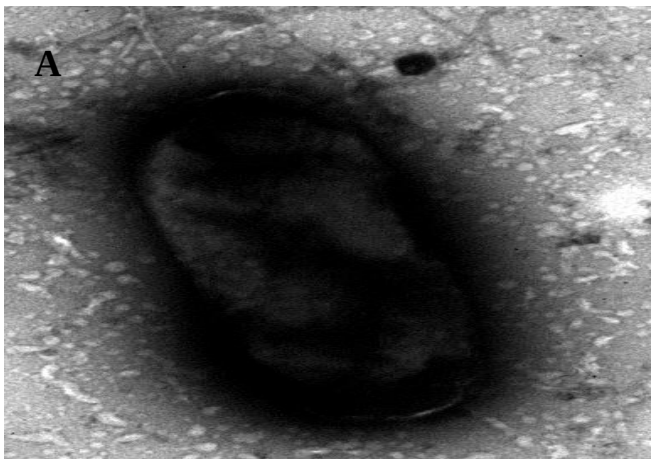
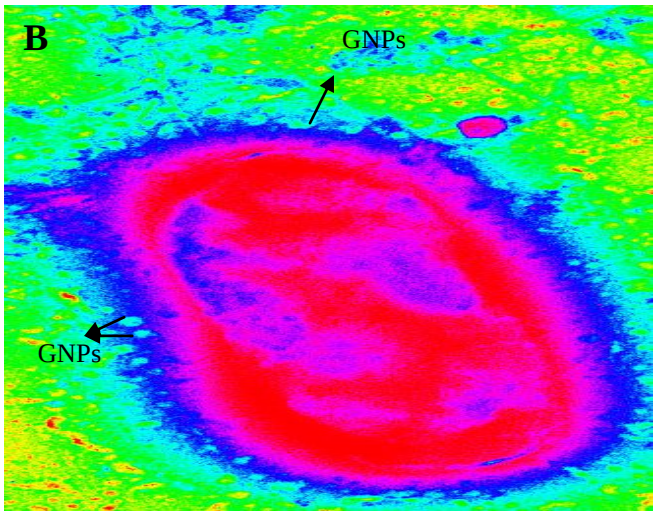


Fig (27): Bio –nanocomposite of E.coli- EG2, grey level (A), at RGB level (B) and Edge detection (C)





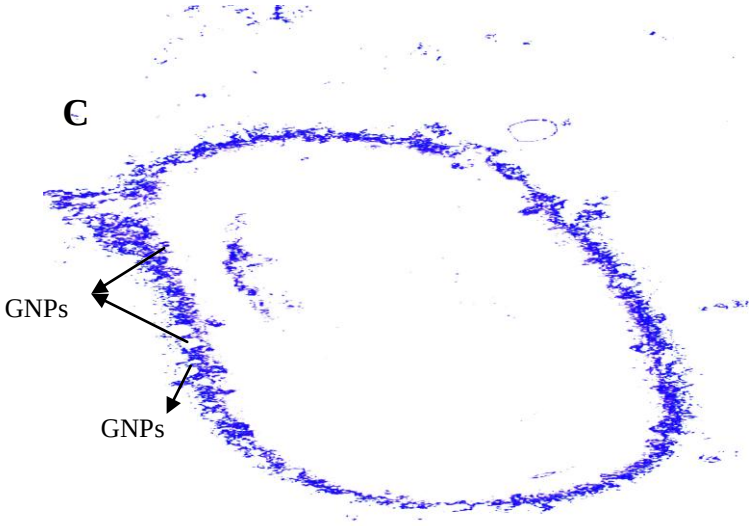


Fig (2): Bio –nanocomposite of E.coli- EG3, grey level (A), at RGB level (B) and Edge detection (C)