

## БІОІНФОРМАЦІЙНИЙ АНАЛІЗ ПОЛІМОРФІЗМУ КОДУЮЧИХ ДІЛЯНОК ГЕНУ *bla TEM* У ШТАМІВ *ESCHERICHIA COLI*

Перетятко О.Г., Ягнюк Ю.А., Скляр Н.І.,  
Большакова Г.М., Холодна Т.В.

ДУ «Інститут мікробіології та імунології  
ім. І.І. Мечникова Національної академії медичних  
наук України», м. Харків

Одним з методів фундаментальних досліджень еволюції антибіотикорезистентності є біоінформаційний аналіз бактеріального геному з проведенням скринінгу генетичних елементів для виявлення варіабельності нуклеотидних послідовностей. Інформацію про секвенс геномів сконцентровано у міжнародних базах даних (GenBank, ENA, DDBJ та ін.). Для аналізу послідовностей нуклеотидів генів використовуються пакети прикладних програм «Vector NTI Advance», «Profiles», «Blast», «HMM» та ін., які дозволяють визначити консервативні та варіабельні ділянки генів антибіотикорезистентності. Порівняльний аналіз молекулярно-генетичної мінливості або консервативності геномного секвенсу використовується як зручний інструмент для розуміння закономірностей еволюційного процесу через визначення філогенетичних зв'язків. Еволюційна історія, встановлена за результатами філогенетичного аналізу, зазвичай, зображується у вигляді деревоподібних діаграм, які являють собою ймовірне родослів'я спадковості між явищами [1, 2].

Відомо, що основу терапії інфекцій, викликаних ентеробактеріями, складають бета-лактамі антибіотики [3, 4]. Основним фактором формування резистентності мікроорганізмів до зазначеної групи антибіотиків є продукція ферментів – β-лактамаз. Серед β-лактамаз ентеробатерій домінують ферменти груп TEM та SHV, що пов'язано з плазмідною локалізацією кодуєчих їх генів – *blaTEM*

та *blaSHV*. Представники ферментів TEM і SHV (TEM-1, TEM-2, SHV-1) загалом складають більш ніж 90 % всіх плазмідних β-лактамаз [5-8]. Вперше плазмідна β-лактамаза TEM-1 виявлена у грамнегативних бактерій ще на початку 60-х років минулого століття, а на сьогодні встановлено вже майже 100 похідних ферменту TEM-1, що обумовлено варіабельністю нуклеотидних послідовностей в кодуєчих ділянках гену *blaTEM* [9].

Метою дослідження було проведення біоінформаційного аналізу варіабельності гену антибіотикорезистентності *blaTEM* у штамів *E. coli*.

**Матеріали і методи.** Для біоінформаційного аналізу використовували пошукову базу даних нуклеїнових кислот GenBank. Пошук нуклеотидних послідовностей гену *blaTEM* проводили у fasta-format. Для аналізу послідовностей нуклеотидів гену *blaTEM* використовували пакет прикладних програм «Vector NTI Advance 11.0» Множинне вирівнювання та його статистичний аналіз для визначення консервативних і варіабельних ділянок гену *bla TEM* проводили за допомогою програмного компонентного модулю «AlignX» [1, 2].

Проаналізовано нуклеотидні послідовності 21 секвенсу гену *blaTEM*, виділених від штамів *E. coli* в 11 країнах світу та зареєстрованих у базі GenBank з 1996 по 2019 роки.

**Результати досліджень та їх обговорення.** Множинне вирівнювання послідовностей нуклеотидів досліджених секвенсів генів *blaTEM* дозволило визначити як консервативні так і варіабельні ділянки гену.

Виявлено 12 консервативних ділянок гену *blaTEM*, довжиною не менш 30 нуклеотидних послідовностей. Найбільш протяжний за гомологічністю фрагмент гену (144 нуклеотидні послідовності) знаходився у позиції 529 – 672 нн (рис. 1). Про варіабельність гену *blaTEM* свідчили точкові мутації у 41 позиції нуклеотидних послідовностей. Найбільш гетерогенна ділянка представлена у позиціях 697-717 нн та 773-813 нн (рис. 2).

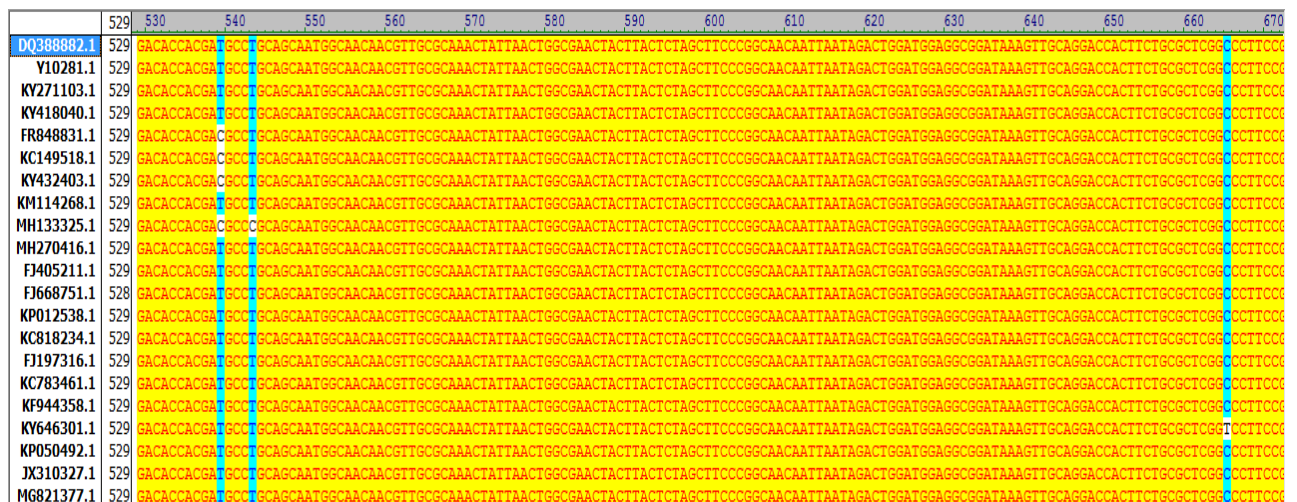


Рис. 1. Вікно візуалізації ділянки секвенсу гену *blaTEM E. coli* (позиція від 529 н до 672 н)

	695	700	710	720	730	740	750	760	770	780	790	800	810
DQ388882.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
Y10281.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KY271103.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KY418040.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
FR848831.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KC149518.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KY432403.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KM114268.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
MH133325.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
MH270416.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
FJ405211.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
FJ668751.1	694	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KP012538.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KC818234.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
FJ197316.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KC783461.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KF944358.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KY646301.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
KP050492.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
JX310327.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											
MG821377.1	695	AACTGGAGCCAAATGACGCTGGATCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACATGACGGGGAGTCAGGCAACTATGGATGAACGAAA											

Рис. 2. Вікно візуалізації ділянки секвенсу гену *blaTEM E. coli* (позиція від 695 н до 815 н)

Мутації проаналізованих генів *blaTEM E. coli* були обумовлені переважно транзиціями (заміна пуринових основ на пуринові або піримідинових на піримідинові), які склали (60,5 %). Мутації у вигляді трансверсій (заміна пуринових основ на піримідинові та навпаки) склали (37,2 %), делеції – (2,3 %).

Ступінь спорідненості проаналізованих нами секвенсів гену *blaTEM* представлено у вигляді дендрограми (рис. 3).

На філогенетичному дереві представлено три умовних генетичних кластери. Верхню позицію дендрограми займає кластер, сформований чотирма субкластерами, які включали в себе вісім секвенсів гену *blaTEM* (DQ388882.1; Y10281.1; KY271103.1; KY418040.1; FR848831.1; KC149518.1; KY432403.1; KM114268.1), виділених у країнах Європи – 2 у Франції і по одному у Польщі, Німеччині, Англії, Португалії, Італії та Нідерландах. Даний кластер характеризувався значною кількістю еволюційних подій (від 4 до 9), пов'язаних з замінами нуклеотидних послідовностей, що на дендрограмі представлено дивергенцією гілок філогенетичного дерева.

Кластер, який займає середню позицію дендрограми, представлено п'ятьма філогенетичними гілками. Даний кластер включає секвенси генів

*blaTEM* (KM114268.1, MH270416.1, FJ405211.1, FJ668751.1, KP012538.1), виділені у Кореї, Китаї, Німеччині та Індії. У переважній більшості секвенсів зазначеного кластеру встановлено 1-2 нуклеотидні заміни.

Нижня позиція дендрограми складається із секвенсів генів *blaTEM* (KC818234.1, FJ197316.1, KC783461.1, KF944358.1, KY646301.1, KP050492.1, JX310327.1, MG821377.1), що характеризувались високим ступенем спорідненості та незначною кількістю змін нуклеотидних послідовностей. Значна кількість (75 %) секвенсів виділена у США та Китаї.

## Висновки

1. За результатами проведеного біоінформаційного аналізу встановлено, що нуклеотидні послідовності гену *blaTEM E. coli* в історичному аспекті зазнали декількох еволюційних дивергенцій, та набули ознак гетерогенності, що, вочевидь, обумовлює появу великої кількості різновидів гену *blaTEM*.

2. Гомологічні послідовності нуклеотидів у досліджених секвенсах гену *blaTEM E. coli* є основою в конструюванні системи праймерів для виявлення у *E. coli* генів резистентності до β-лактамних антибіотиків методом полімеразної ланцюгової реакції.

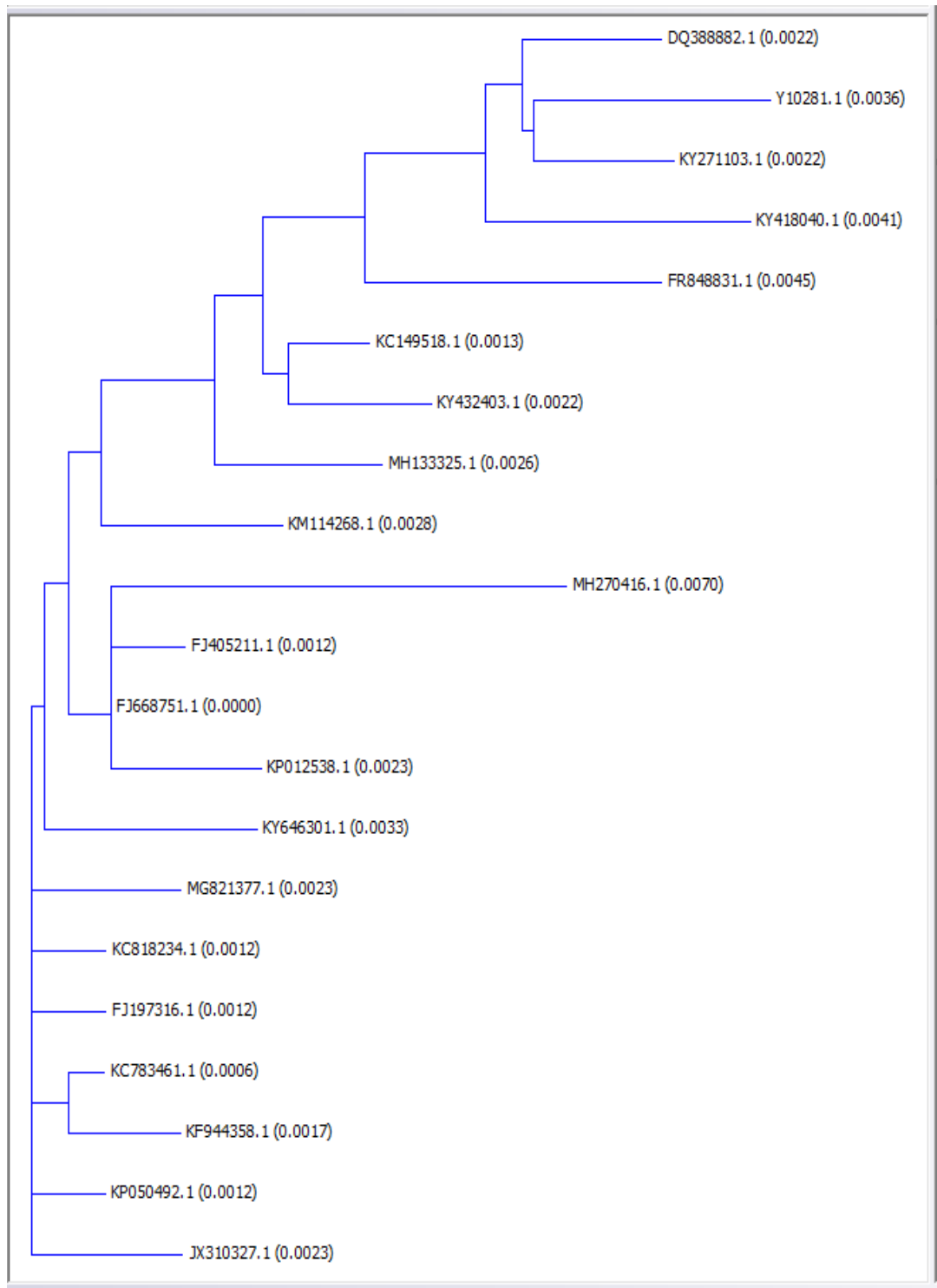


Рис. 3 – Дендрограма секвенсів гену *blaTEM E. coli*

#### References

1. Porozov Yu. Bioinformatics. SPb: NRU ITMO, 2012. 52 p.
2. Demkin VV. Bioinformational analysis of the nucleotide sequences of bacteria stored in GENBANK. Molecular Genetics, Microbiology and Virology. 2009. No 2. P. 36-39.
3. Yakovlev SV, Suvorov MP. Cefotaxime / sulbactam; important replenishment in the arsenal of inhibitor-resistant beta-lactam antibiotics. Antibiotics and

- chemotherapy. 2019.No 64 (3-4). P. 7-8. DOI: 1/24411 / 235-299-219-119
4. Bisekenova A L., Ramazanova BA., Adambekov DA., Bekbolatova KA. Molecular mechanisms of resistance of gram-negative microorganisms - pathogens of infections to beta-lactam antibiotics. Bulletin of KazNMU. 2015. No. 3, P. 223-227.
5. Eidelstein MV. Development of molecular genetic methods for the study of TEM- and SHV-type  $\beta$ -lactamases in clinical enterobacterial strains: abstract. dis. for the competition degrees of cand. biol. Sciences: 14.00.31. Moscow, 2000. 21 p.

6. Gundran, RS., Cardenio, PA., Villanueva, MA. Prevalence and distribution of blaCTX-M, blaSHV, blaTEM genes in extended- spectrum  $\beta$ - lactamase-producing *E. coli* isolates from broiler farms in the Philippines. BMC Vet Res. 2019. No 15(1):227. doi: 10.1186/s12917-019-1975-9.
7. Ulyashova MM., Presnova GV., Pobolelova Yu.I., Filippova AA., Egorov AM., Rubtsova MJ. Screening of bacterial genes responsible for resistance to beta-lactam antibiotics using microarrays with enzymatic detection. Bulletin Mosk. University. 2016. Ser. 2, Chemistry. T. 57. No. 4. P. 245-252.
8. Mironov AJ., Krapivina IV., Mudrak DE., Ivanov DV. Molecular mechanisms of resistance to  $\beta$ -lactam pathogens of nosocomial infections. Clinical laboratory diagnostics. 2012. No. 1. P. 39-43.
9. Lagoon LV. Extended beta-lactamases and their importance in the formation of the resistance of urinary tract infections to antimicrobial agents. Health and Ecology Problems. 2012. No 3 (33). P. 82-88.

**Bioinformation analysis of polymorphism of the blaTEM gene coding regions in Escherichia coli strains Peretyatko O.G., Yagnuk Y.A., Sklyar N.I., Bolshakova G.M., Cholodna T.V.**

Bioinformation analysis of the bacterial genome with screening of genetic elements to detect variability of nucleotide sequences is one of the methods of antibiotic resistance evolution fundamental research. Comparative analysis of the molecular genetic variability or genome sequence conservatism is used as a convenient tool for understanding patterns of the evolutionary process through definition of the phylogenetic relationships. **The aim of the study** was to conduct a bioinformation analysis of the variability of the antibiotic resistance blaTEM gene in *E. coli* strains. **Materials and methods.** The GenBank nucleic acid search database was used for bioinformation analysis. The search for nucleotide sequences of the blaTEM gene was performed in fasta-format. "Vector NTI Advance 11.0" software package was used to analyze the nucleotide sequences of the blaTEM gene. Multiple alignment and its statistical analysis to determine conservative and variable regions of the blaTEM gene was performed using software component module "AlignX". Nucleotide sequences of the 21 blaTEM gene sequences isolated from *E. coli* strains in 11 countries and registered in the GenBank database from 1996 to 2019 were analyzed. **Research results and discussion.** Multiple alignment of the nucleotide sequences of the studied blaTEM gene sequences allowed to determine both conservative and variable regions of the gene. 12 conservative regions of the blaTEM gene with a length of at least 30 nucleotide sequences were identified. The longest homologous region of the gene (144 nucleotide sequences) was found at the 529 - 672 nn position. blaTEM gene variability was evidenced by point mutations at 41 positions of the nucleotide sequences. The most heterogeneous region was found at the 697-717 nn and 773-813 nn positions. Mutations in the analyzed *E. coli* blaTEM genes were caused by transitions (60,5%), transversions (37,2%) and deletions (2,3%). The degree of relatedness of the

analyzed blaTEM gene sequences is presented in the form of a dendrogram. The phylogenetic tree shows three conditional genetic clusters. The top position of the dendrogram is occupied by the cluster that includes eight blaTEM gene sequences, isolated in European countries: 2 - in France and one each - in Poland, Germany, England, Portugal, Italy and the Netherlands. This cluster was characterized by a significant number of evolutionary events (from 4 to 9) associated with nucleotide sequence substitutions, which is represented on the dendrogram by the divergence of the phylogenetic tree branches. The cluster which occupies the middle position of the dendrogram is represented by five phylogenetic branches. This cluster includes blaTEM gene sequences isolated in Korea, China, Germany and India. 1-2 nucleotide substitutions were identified in the majority of the sequences in this cluster. The lower position of the dendrogram consists of blaTEM gene sequences characterized by a high degree of relatedness and a small number of nucleotide sequence substitutions, a significant number of sequences were isolated in the United States and China (75%). **Conclusions.** Results of the bioinformation analysis showed that the nucleotide sequences of the *E. coli* blaTEM gene have historically undergone several evolutionary divergences and acquired signs of heterogeneity, which apparently causes the emergence of a large number of species of the blaTEM gene.

**Key words:** bioinformation analysis, *E. coli*, blaTEM gene.