

---

**THE APPLICATION OF BIOINFORMATICS IN THE MOLECULAR CHARACTERIZATION OF *Bacillus licheniformis*****Srdjan Tasić**College of Applied Professional Studies, Vranje, Serbia, [drsrdjantasic@ptt.rs](mailto:drsrdjantasic@ptt.rs)**Irena Tasić**College of Academic Studies „Dositej“, Belgrade, Serbia, [irena.tasic67@gmail.com](mailto:irena.tasic67@gmail.com)

**Abstract:** Bioinformatics is the application of information technology in biology and includes the processes of gathering, processing and analysing experimental results. Bioinformatics now entails the creation and advancement of databases, algorithms, computational and statistical techniques, and theory to solve formal and practical problems arising from the management and analysis of biological data. Computers are necessary in microbiology because the manual comparison of multiple sequences has become unpractical. The research subject was the characterisation of the strain ST51 isolated from the thermal water well in Vranjska Banja, south eastern Serbia. Molecular characterisation of these three strains was performed by analysis of the *tuf* gene, which encodes the elongation factor Tu. The DNA sequences were compared to those deposited in *GenBank*. data bases using the BLAST program (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn>). The biochemical characterisation was performed using the API 50 CHB system (bioMérieux) and APIWEB™ software Ver. 4.1. The molecular characterisation of the strain ST51 proved the highest level of similarity to the strain *Bacillus licheniformis* marked as ATCC 14580 (99% identical). The biochemical characterization confirmed that the strain ST51 belongs to the species *Bacillus licheniformis*. Given that all the conducted analyses yielded a substantial number of data, they were processed and compared using biostatistics methods and tools in order to achieve the highest probability of resulted taxonomic classification. Modern research contributes to the analysis of a significant number of variables which is why considerably more statistical analyses are involved in their interpretation and presentation. Our results indicate that different methods are needed for proper determination and characterisation of isolates/strains. Regarding taxonomy, molecular methods are the most precise, while for physiological specificity biochemical methods are more reliable.

**Keywords:** bioinformatics, molecular characterization, *Bacillus licheniformis*.

**PRIMENA BIOINFORMATIKE U MOLEKULARNOJ KARAKTERIZACIJI *Bacillus licheniformis*****Srdan Tasić**Visoka škola primenjenih strukovnih studija, Vranje, Srbija, [drsrdjantasic@ptt.rs](mailto:drsrdjantasic@ptt.rs)**Irena Tasić**Visoka škola akademskih studija „Dositej“, Beograd, Srbija, [irena.tasic67@gmail.com](mailto:irena.tasic67@gmail.com)

Bioinformatika predstavlja primenu informacionih tehnologija u biologiji i obuhvata prikupljanje, obradu i analizu eksperimentalnih rezultata. Bioinformatika danas podrazumeva stvaranje i razvoj baza podataka, algoritama, informatičkih i statističkih tehnika kao i teorijske osnove za rešavanje formalnih i praktičnih problema koji se javljaju u upravljanju i analizi bioloških podataka. Kompjuteri su postali neophodni i u mikrobiologiji jer se ručno poređenje višestrukih sekvenci pokazalo nepraktičnim. Predmet istraživanja bila je karakterizacija soja ST51 izolovanog iz termalnog izvora u Vranjskoj Banji, jugoistočna Srbija. Molekularna karakterizacija soja izvedena je analizom *tuf* gena koji kodira elongacioni faktor Tu. Korišćenjem BLAST programa DNK (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn>) sekvence su upoređene sa sekvencama deponovanim u bazi podataka *GenBank*. Biohemijska karakterizacija urađena je korišćenjem sistema API 50 CHB (BioMérieux) a dobijeni rezultati su interpretirani pomoću softvera APIWEB™, Ver. 4.1. Molekularna karakterizacija soja ST51 pokazala je najveću sličnost sa sojem *Bacillus licheniformis* označenim kao ATCC 14580 (99% identičnosti). Biohemijska karakterizacija potvrdila je da soj ST51 pripada vrsti *Bacillus licheniformis*. U analizi je dobijen veliki broj podataka za čija su poređenja korišćene biostatističke metode kako bi se sa najvećom verovatnoćom utvrdila pripadnost vrsti. Savremena istraživanja doprinose analizi velikog broja varijabli tako da zahtevaju sve veći udeo statistike u njihovom tumačenju i prezentaciji. Rezultati ukazuju da je za determinaciju i karakterizaciju novih izolata/sojeva neophodno koristiti veći broj metoda kako bi se potvrdila njihova taksonomska pripadnost (gde je

---

molekularna metoda pokazala najveću preciznost) kao i da bi se potvrdila njihova fiziološka specifičnost (za koju su biohemijske metode pouzdanije)

**Ključne reči:** Bioinformatika, molekularna karakterizacija, *Bacillus licheniformis*.

## 1. UVOD

Bioinformatika je naučna disciplina koja je nastala kombinacijom informatike, matematike, statistike i tehničkih nauka koja se bavi sakupljanjem, skladištenjem, obradom i analizom bioloških informacija na velikoj skali primenom kompjuterskih (*in silico*) programa i softvera [1]. Bioinformatika predstavlja praktičnu primenu informacionih tehnologija u biologiji a njen osnovni cilj je povećanje razumevanja bioloških procesa što uključuje prepoznavanje obrazaca, analizu podataka, mašinsko učenje i vizuelizaciju bioloških podataka. Analiziranje bioloških podataka, kako bi se dobile sadržajne informacije, podrazumeva pravljenje i upotrebu kompjuterskih programa koji koriste algoritme iz teorije grafova, veštačke inteligencije, algoritme sa slabo definisanim rezultatima, analize podataka, prepoznavanja obrazaca, procesiranja slika i kompjuterske simulacije. Ovi algoritmi oslanjaju se na saznanja iz diskretne matematike, kontrolne teorije, sistemske teorije, informacione teorije i statistike [2]. Bioinformatika danas podrazumeva stvaranje i razvoj baza podataka, algoritama, informatičkih i statističkih tehnika, kao i teorijske osnove za rešavanje formalnih i praktičnih problema koji se javljaju u upravljanju i analizi bioloških podataka.

Kompjuteri su postali neophodni i u mikrobiologiji jer se ručno poređenje višestrukih sekvenci pokazalo nepraktičnim. Pored kreiranja softvera za analizu dobijenih rezultata, za čuvanje dobijenih rezultata koriste se tzv. superkompjuteri. U toku je dešifrovanje milijardi baznih parova genoma na stotine vrsta. Tako obiman posao zahteva i formiranje specifičnih kompjuterizovanih baza podataka. Krajnji cilj ovih (baza)-obrada baza podataka je da rezultati rada naučnika na genomima brojnih organizama budu dostupni svim zainteresovanim istraživačima sveta posredstvom interneta [3].

Danas, američka baza podataka, poznatija pod nazivom *GenBank*, poseduje sekvence svih do danas iščitanih genoma odnosno više od 100.000.000.000 bp. Ova baza podataka je svakog meseca bogatija za nekoliko miliona bp iščitanih sekvenci [4].

## 2. MATERIJAL I METODE

Predmet molekularne i biohemijske karakterizacije bio je soj ST51 koga smo izolovali iz termalnog izvora u Vranjskoj Banji, jugoistočna Srbija. Za izolaciju su korišćene standardne bakteriološke metode.

Molekularna karakterizacija sojeva je izvedena analizom *tuf* gena koji kodira elongacioni faktor Tu. Oba lanca PCR amplikona *tuf* gena su sekvencirana prajmerima korišćenim u reakcijama PCR amplifikacije. Molekularna karakterizacija bakterija sastojala se od izolacije DNK iz bakterijskih ćelija, elektroforeze na agaroznom delu, *Touchdown PCR* – metode i metode terminacije lanaca (*Sanger*-ova metoda sekvenciranja). Za izolovanje DNK korišćene su prekonocne „overnight“- kulture (gde se većina bakterija nalazi u stacionarnoj fazi tj. u fazi kada je izolacija DNK i najuspešnija). Količina i kvalitet izolovane bakterijske DNK rutinski je proveravana na agaroznom gelu. Korišćena je *touchdown PCR* metoda u kojoj se hibridizacija prajmera odvija u svakom ciklusu na sve nižoj temperaturi (što je u konkretnom slučaju predstavljalo sukcesivno opadanje temperature anilinga prajmera za po 1°C po ciklusu i to u toku 15 ciklusa). Ovo znači da temperatura na kojoj dolazi do vezivanja prajmera posle 15 ciklusa iznosi 45°C. Sekvenciranje bakterijske DNK urađeno je metodom po *Sanger*-u na Univerzitetu u Padovi – Italija.

Detekcija eventualnih lažnih pozitivnih rezultata vršena je tzv. „korišćenjem kontrole bez DNK“ i „slepom probom“. Kontrolom bez DNK testira se prisustvo eventualne kontaminacije samo u reagensima za amplifikaciju dok se slepom probom testira eventualno prisustvo kontaminacije i u reagensima korišćenim za DNK ekstrakciju i purifikaciju [5].

Za molekularnu karakterizaciju koristili smo najpoznatiju bazu podataka [6] koja se nalazi na sajtu Američkog nacionalnog centra za biotehnologiju (*US National Center for Biotechnology Information, NCBI*) na web adresi <http://www.ncbi.nlm.nih.gov>. Poređenje sekvence našeg soja ST51 sa sekvencama više od 260.000 organizama, koje sadrže više od 190 milijardi nukleotida i koje se nalaze u ovoj bazi podataka [7] uradili smo pomoću računarskog programa “BLAST”.

Biohemijska karakterizacija urađena je korišćenjem sistema API 50 CHB (*BioMérieux*), a rezultati su interpretirani pomoću softvera APIWEB™, Ver. 4.1.

**3. REZULTATI I DISKUSIJA**

Temperatura vode izvora iz koga je izolovan soj ST51 iznosila je 80°C. Ovaj izvor, uz Islandske gejzire, spada u najvrelije izvore u Evropi. Voda pripada kategoriji natrijum-hidrokarbonatno-sulfatnih, fluoridnih sulfidnih hipertermnih voda sa ukupnom β-aktivnošću 475.8 mBq/L i ukupnom α-aktivnošću < 37.0 Bq/L [8].

Korišćenjem BLAST programa [9], [10] DNK sekvenca ispitivanog soja ST51 upoređena je sa sekvencama deponovanim u bazama podataka NCBI banke gena. Molekularna karakterizacija sekvence soja ST51 (slika 1) pokazala je najveću sličnost sa sojem *Bacillus licheniformis* (Weigmann, 1898) označenim kao ATCC 14580 (99% identičnosti).

Bacillus licheniformis ATCC 14580, complete genome  
 Sequence ID: ref|NC\_006270.3| Length: 4222597 Number of Matches: 1  
 Range 1: 132236 to 132994

Score	Expect	Identities	Gaps	Strand	Frame
1358 bits(735)	0.0()	754/762(99%)	6/762(0%)	Plus/Plus	

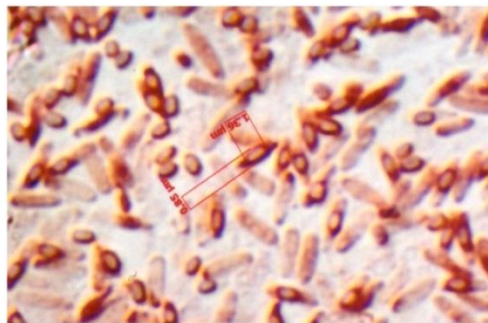
Features:  
 elongation factor Tu

Query	Score	Expect	Identities	Gaps	Strand	Frame
Query 10	66		AAA-ATGTATC-CTGAGTGTCTGCGC-AATGGACGGTGCATCCTTGTGTTTCTGCTGCT			66
Sbjct 132236			AAACATG-ATCACTG-GTCTGCGCAATGGACGGTGCATCCTTGTGTTTCTGCTGCT			132293
Query 67	126		GACGGCCAAATGCCACAACCTCGTGAGCACATCCTTCTTCTCGTAACGTAGGTGTGCCT			126
Sbjct 132294			GACGGCCAAATGCCACAACCTCGTGAGCACATCCTTCTTCTCGTAACGTAGGTGTGCCT			132353
Query 127	186		TACATCGTAGTATCTTGAACAATGCGACATGGTAGACGAAATAGCTTCTTGAGCTA			186
Sbjct 132354			TACATCGTAGTATCTTGAACAATGCGACATGGTAGACGAAATAGCTTCTTGAGCTA			132413
Query 187	246		GTTGAAATGGAAGTTCGGATCTTCTTCTGAGTATGAGTTCCTGGTGCAGATGTACCA			246
Sbjct 132414			GTTGAAATGGAAGTTCGGATCTTCTTCTGAGTATGAGTTCCTGGTGCAGATGTACCA			132473
Query 247	306		GTTATCAAAGGTTCTGCTCTTAAAGCTCTTGAAGGAGATGCTCAGTACGAAGAAAAATC			306
Sbjct 132474			GTTATCAAAGGTTCTGCTCTTAAAGCTCTTGAAGGAGATGCTCAGTACGAAGAAAAATC			132533
Query 307	366		TTTGAACGTATGGCTGCTGTTGACGAGTACATCCCACTCCAGAGCGTGAACGTGACAAG			366
Sbjct 132534			TTTGAACGTATGGCTGCTGTTGACGAGTACATCCCACTCCAGAGCGTGAACGTGACAAG			132593
Query 367	426		CCATTTCATGCTGCTGTTGAGGAGTATCTCAATCACAGCTCGTGGACAGTTGCTACT			426
Sbjct 132594			CCATTTCATGCTGCTGTTGAGGAGTATCTCAATCACAGCTCGTGGACAGTTGCTACT			132653
Query 427	486		GGACGCGTAGAGCGCGACAAGTAAAGTCGGTGCAGAGTGAATCATCGGCTCTCAA			486
Sbjct 132654			GGACGCGTAGAGCGCGACAAGTAAAGTCGGTGCAGAGTGAATCATCGGCTCTCAA			132713
Query 487	546		GAAGAAACGAAGAAAACAAGTGTACAGGTGTTGAAATGTTCCGCAAGCTCTTGACTAT			546
Sbjct 132714			GAAGAAACGAAGAAAACAAGTGTACAGGTGTTGAAATGTTCCGCAAGCTCTTGACTAT			132773
Query 547	606		GCAGAAAGCTGGAGACAACATCGGTGCACTTCTTCGCGGTGATCTCGTGAAGAAATCCAA			606
Sbjct 132774			GCAGAAAGCTGGAGACAACATCGGTGCACTTCTTCGCGGTGATCTCGTGAAGAAATCCAA			132833
Query 607	666		CGCGGTCAAGTACTTGCTCAACCGAGTACAACTCACTCCGCAAAAAATTCAAAGGTGAA			666
Sbjct 132834			CGCGGTCAAGTACTTGCTCAACCGAGTACAACTCACTCCGCAAAAAATTCAAAGGTGAA			132893
Query 667	726		GTTTACGTTCTTTCTAAAGAGAGGGTGGACGTCACACTCCATCTTCTCTAAGTACCGT			726
Sbjct 132894			GTTTACGTTCTTTCTAAAGAGAGGGTGGACGTCACACTCCATCTTCTCTAAGTACCGC			132953
Query 727	768		CCTCAGTCTACTTCCGTACAACACTGCTAAGTGGTAATCAT			768
Sbjct 132954			CCTCAGTCTACTTCCGTACAACACTGCTAAGTGGTAATCAT			132994

**Slika 1. Prikaz uporednog poravnanja dobijene sekvenciranjem dela tuf gena soja ST51 sa deponovanim sekvencama za tuf gen u svetskoj banci gena korišćenjem BLAST servisa**

Biohemijska karakterizacija izvršena je na osnovu standardnih biohemijskih testova manualnog API 50 CHB sistema. Biohemijski profil koji smo dobili tipičan je za *Bacillus licheniformis* gde je procenat identifikacije (%Id) 99,9 a T index je iznosio 0.53 (procena o tome koliko profil odgovara najtipičnijem setu reakcija za svaki takson i obrnuto je proporcionalna broju atipičnih testova).

Bojenje po Gramu pokazuje sporulišuće Gram-pozitivne štapiće dužine do 3 µm sa elipsoidnim centralno lociranim sporama prosečne veličine 1.35 x 0.86 µm (slika 2). Na hranljivom agaru raste na temperaturama inkubacije od 30°C do 55°C.



**Slika 2. Bacillus licheniformis, Bojenje po Gramu (foto: S. Tasić)**

#### 4. ZAKLJUČCI

Kompjuteri su postali neophodni i u mikrobiologiji jer se ručno poređenje višestrukih sekvenci pokazalo nepraktičnim. Bioinformatika danas podrazumeva stvaranje i razvoj baza podataka, algoritama, informatičkih i statističkih tehnika, kao i teorijske osnove za rešavanje formalnih i praktičnih problema koji se javljaju u upravljanju i analizi bioloških podataka. Danas, američka baza podataka, poznatija pod nazivom *GenBank*, poseduje sekvence svih do danas iščitanih genoma odnosno više od 100.000.000.000 bp.

Predmet molekularne karakterizacije bio je soj ST51 izolovan iz termalnog izvora, čija temperatura vode iznosi 80°C. Molekularna karakterizacija soja izvedena je analizom *tuf* gena koji kodira elongacioni faktor Tu. Pomoću računarskog programa "BLAST" uradili smo poređenje sekvence našeg soja ST51 sa sekvencama više od 260.000 organizama, koje sadrže više od 190 milijardi nukleotida i koje se nalaze u bazi podataka *GenBank*. Ispostavilo se da soj ST51 pokazuje najveću sličnost sa sojem *Bacillus licheniformis* označenim kao ATCC 14580 (99% identičnosti). Biohemijaska karakterizacija, urađena korišćenjem sistema API 50 CHB (BioMérieux), potvrdila je da soj ST51 pripada vrsti *Bacillus licheniformis* tako što su rezultati interpretirani pomoću softvera APIWEB™, Ver. 4.1

U analizi soja ST51 dobijen je veliki broj podataka za čije su poređenje korišćene biostatističke metode kako bi se sa najvećom verovatnoćom utvrdila pripadnost vrsti. Savremena istraživanja doprinose analizi velikog broja varijabli tako da zahtevaju sve veći udeo statistike u njihovom tumačenju i prezentaciji.

Rezultati ukazuju da je za determinaciju i karakterizaciju novih izolata/sojeva neophodno koristiti veći broj metoda kako bi se potvrdila njihova taksonomska pripadnost (gde je molekularna metoda pokazala najveću preciznost) kao i da bi se potvrdila njihova fiziološka specifičnost (za koju su biohemijske metode pouzdanije). Metodi sekvenciranja, kao svojevrsnog „molekularnog hronometra“, u ovom slučaju treba dati prednost i treba je smatrati pouzdanijom od biohemijske karakterizacije. Razlike u rezultatima identifikacije ove dve metode mogu da ukazuju i na autohtoni karakter ispitivanih sojeva.

#### LITERATURA

- [1] Stevanović, M. (2016). *Osnovi manipulisanja genima*. Beograd: Biološki fakultet.
- [2] <https://sh.wikipedia.org/wiki/Bioinformatika>
- [3] Brajušković, G. (2012). *Molekularna biologija 2*. Beograd: Savremena administracija.
- [4] Lesk, A. (2013). *Introduction to Bioinformatics*. New York: Oxford University Press Inc.
- [5] Romac, S., Vukosavić, S., Stojković, O., Čuljković, B. (1999). *PCR u kliničkoj dijagnostici*. Beograd: Biološki fakultet.
- [6] Baxevanis, A.D., Ouellette, B.F. Francis. (2004). *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*. New Jersey: Wiley-Interscience.
- [7] Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. (2008). *Nucleic Acids Research*, 36 (1), 25–30.
- [8] Tasić, S. (2010) Ecological biochemistry of the *Bacillus thermophilus vranjensis* in the warm water springs of Vranjska Banja spa, Power of microbes in industry and environment - central European symposium on industrial microbiology and microbial ecology, Book of abstracts, Croatian microbiological society, pp. 133 - 133, Malinska, Hrvatska, 22. - 25. sep, 2010.
- [9] <https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn>
- [10] Rangel, P., Giovannetti, J (2001). *Genomes and Databases on the Internet: A Practical Guide to Functions and Applications*. Berkley: Horizon Scientific Press.
- [11] Tasić, S. (2007). *Molecular and biochemical characterisation of Gram-negative aerobic and facultatively anaerobic bacteria in oligomineral waters of wells in the Jakov Do on the territory of Vlasina*. Doctoral thesis. Belgrade: Faculty of Agriculture, University in Belgrade.