

# **SZAKDOLGOZAT**

**Burján Alex**

**2024**



**Magyar Agrár- és Élettudományi Egyetem  
Budai Campus  
Élelmiszertudományi és Technológiai intézet  
Biomérnök alapképzési szak**

**Tejsavbaktérium eredetű béta-galaktozidáz enzim  
tanulmányozása bioinformatikai eszközökkel**

**Belső konzulens:** **Dr. Bujna Erika**  
egyetemi docens  
**Dr. Nguyen Duc Quang**  
egyetemi tanár

**Belső konzulens intézete/tanszéke:**  
Élelmiszertudományi és Technológiai Intézet,  
Biomérnök és Erjedésipari Technológia Tanszék

**Készítette:** **Burján Alex**

**Budapest**

**2024**

# TARTALOMJEGYZÉK

1	Bevezetés és célkitűzések.....	4
2	Szakirodalmi áttekintés .....	6
2.1	Tejsavbaktériumok jellemzése.....	6
2.2	Béta-galaktozidáz.....	8
2.2.1	Általános jellemzők.....	8
2.2.2	Az enzim szerkezete .....	9
2.2.3	A lac operon működése .....	10
2.2.4	Az enzim forrásai.....	11
2.2.5	Az enzim felhasználási területei .....	13
2.3	Bioinformatika .....	14
2.3.1	Bioinformatika történelme.....	14
2.3.2	Bioinformatika módszerei .....	15
2.3.3	A tejsavbaktériumok bioinformatikai vizsgálata .....	17
3	Alkalmazott módszerek.....	18
3.1	Vizsgált baktériumok és kromoszóma NCBI azonosítók .....	18
3.2	Béta-galaktozidáz enzimet kódoló gének azonosítása .....	18
3.3	Vizsgált fehérjék konzervatív régiói .....	18
3.4	MSA és filogenetikai vizsgálat .....	18
3.5	Gének eloszlásának vizsgálata.....	19
3.6	Motif analízis és gén struktúra.....	19
3.7	Különböző eredetű béta-galaktozidáz enzimek jellemzése.....	19
4	Eredmények és kiértékelésük .....	20
4.1	Béta-galaktozidáz enzimet kódoló gének azonosítása .....	20
4.2	Vizsgált fehérjék konzervatív régiói .....	21
4.3	MSA és filogenetikai vizsgálat .....	27
4.4	Gének eloszlásának vizsgálata.....	29
4.5	Motif analízis és gén struktúra.....	30
4.6	Vizsgált enzimek fizikai-kémiai tulajdonságaik és sejten belüli elhelyezkedésük .....	31
5	Következtetések és javaslatok .....	35
6	Összefoglalás.....	36
7	Irodalomjegyzék.....	38

# 1 Bevezetés és célkitűzések

A  $\beta$ -galaktozidáz az egyik legfontosabb enzim az ipar, különösen a tejipar számára. Felhasználjuk a laktózmentes termékek előállítására, ami lehetővé teszi a laktózérzékeny emberek számára a tejtermékek fogyasztását. A laktóztolerancia az egyik leggyakoribb emésztőrendszeri betegség a világ lakosságának kb. 74%-át érinti (Storhaug et al., 2017). De a gyakoriság erősen különbözik etnikumok között. Ázsiában a legjelentősebb, ahol a lakosságnak megközelítőleg a 95%-a érintett, észak-európai országokban a legritkább, ahol a népességnek csupán kb. a 10%-a szenved a betegségben. Akik ezzel a rendellenességgel rendelkeznek, szervezetük nem termel  $\beta$ -galaktozidáz enzimet így nem képesek lebontani a laktózt. Emiatt az eljut a bélrendszerbe, ahol az ott élő baktériumok hasznosítják azt. Az erjesztés során széndioxid gáz keletkezik, amely bélrendszeri tüneteket okoz pl. haspuffadást és hasmenést.

Szintén fontos ipari alkalmazása a galakto-oligoszacharidok (GOS) előállítása, amelyek prebiotikumként alkalmazhatóak. Továbbá, ez az enzim képes a tejiparban keletkező melléktermékben, a tejsavóban található laktóz lebontására, így teszi könnyebben felhasználhatóvá a tejsavót, például a sütőiparban és a cukrászatban vagy lehetővé teszi a tejsavók környezetbarát-kezelését.

A laktóztolerancia azóta érdekes számomra amióta megtudtam, hogy eredetileg az emberek nem voltak képesek a laktóz lebontására csak csecsemőkorukban. De azok az emberek, akik mutációk révén képessé váltak a tejcukor lebontására számos előnyre tettek szert azokkal szemben, akik nem, emiatt a mutáció gyorsan elterjedt a népességben. A fent leírtak miatt választottam ezt a témát kutatómunkám tárgyának, és mivel a modern világban az informatika az élet szerves részévé vált, illetve folyamatosan egyre nagyobb jelentőséggel rendelkezik a biológiai kutatások során, különösen a genetikában, és gyógyszeriparban, illetve a mesterséges intelligencia megjelenésével. Talán még szélesebb és mélyebb módszereket fed fel számunkra. A bioinformatika módszerein át vizsgálva kaphatom a legátfogóbb képet erről az enzimiről.

Kutatómunkám során hat különböző tejsavbaktérium által termelt  $\beta$ -galaktozidáz enzimet terveztem össze hasonlítani bioinformatikai módszerek segítségével. A vizsgált fehérjék összevetéséhez használt információkat internetes adatbázisokból tervezem szereznem. Előre láthatólag leginkább az UniProt, az NCBI és a Brenda adatbázis kerül majd felhasználásra.

A következő feladatokat tűztem ki munkám során:

- Fehérjék aminosavszekvenciájának összehasonlítása
- A fehérjéket kódoló gének összevetése
- Gén szerkezetének vizsgálata
- Fizikai kémia tulajdonságok összevetése
- Konzervatív régiók összevetése

## 2 Szakirodalmi áttekintés

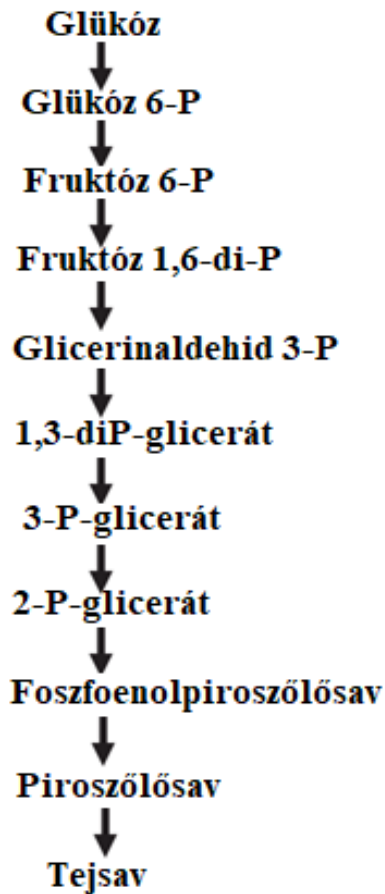
### 2.1 Tejsavbaktériumok jellemzése

A tejsavbaktériumok a *Firmicutes* osztályba tartozó pálcá vagy esetleg gömb alakú Gram-pozitív baktériumok. Az élelmiszeriparban az egyik legkiemelkedőbbben használt mikroorganizmusok, számos termék, például sajt, joghurt, savanyúságok stb. előállítására alkalmazhatók. Az emberiség már hosszú ideje készít ilyen élelmiszereket pl. Kr. e 2300-as évekből származó egyiptomi edényekben már találtak sajt maradványt (Mcgee 1984). A Bibliában is találhatunk utalásokat élelmiszer tartósítási és fermentálási folyamatokra, melyek tejsavbaktériumokkal köthetők össze, ezek alapján ezen mikroorganizmusok felhasználása legalább 4-5 ezer éves, de akár 7-10 ezer éves múltra is vezethető vissza.

Mivel a tejsavbaktériumok nem rendelkeznek működőképes citromsavkörrel, ezért energiatermelő anyagcsereként tejsavas erjedést alkalmaznak. Ennek köszönhetően savtűrő mikroorganizmusok, akár a pH 3,0 értéket is el tudják viselni. A tejsavbaktériumok két fő típusú erjesztését követnek a homofermentatív és a heterofermentatív tejsavas erjedés.

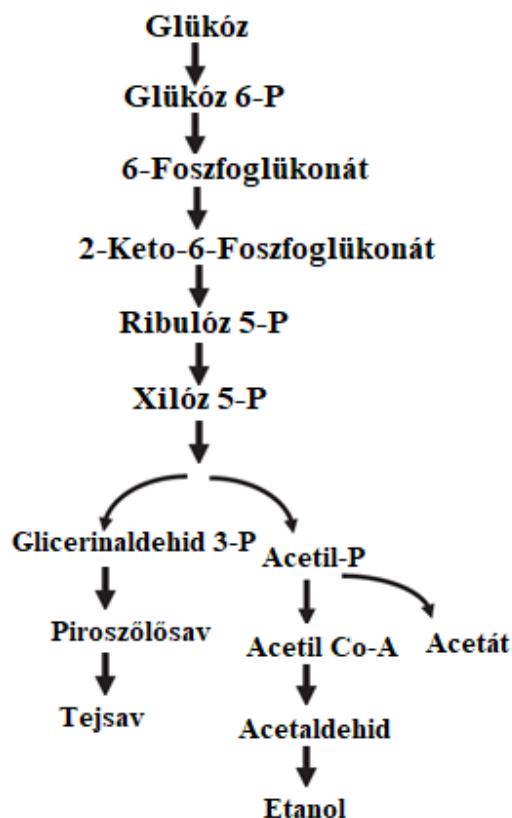
A homofermentatív tejsavas erjedés (1. ábra) során piroszőlősavból főleg tejsavat képeznek a mikroorganizmusok. Ezen az útvonalon a tejsavbaktériumok a cukrokból nagyrészt tejsavat állítanak elő, így a tejipar számára főleg joghurt és sajt gyártás során (Borsodi, 2013) kedvező, hiszen kevés az anyag vesztesége.

Ezt az utat követi például a *Lactobacillus acidophilus* és a *Lactobacillus helveticus*, amikor erjesztik a glükózt.



**1. ábra: Homofermentatív tejsavas erjedés folyamata**  
(Forrás: Kumar et al., 2015 nyomán)

A heterofermentatív útvonal (2. ábra) a xilóz-5-P intermediertől két ágra oszlik szét. Az egyik a homofermentatívhoz hasonlóan piroszőlősavon át tejsav keletkezik. A másikon pedig Acetil-P terméken keresztül Acetát, illetve Acetil Co-A keletkezik, ami végül etanollá alakul. Heterofermentatív baktériumok például a *Lactobacillus brevis* és a *Lactobacillus reuteri*.



2. ábra: Heterofermentatív tejsavas erjedés folyamata  
(Forrás: Kumar et al., 2015 nyomán)

## 2.2 Béta-galaktozidáz

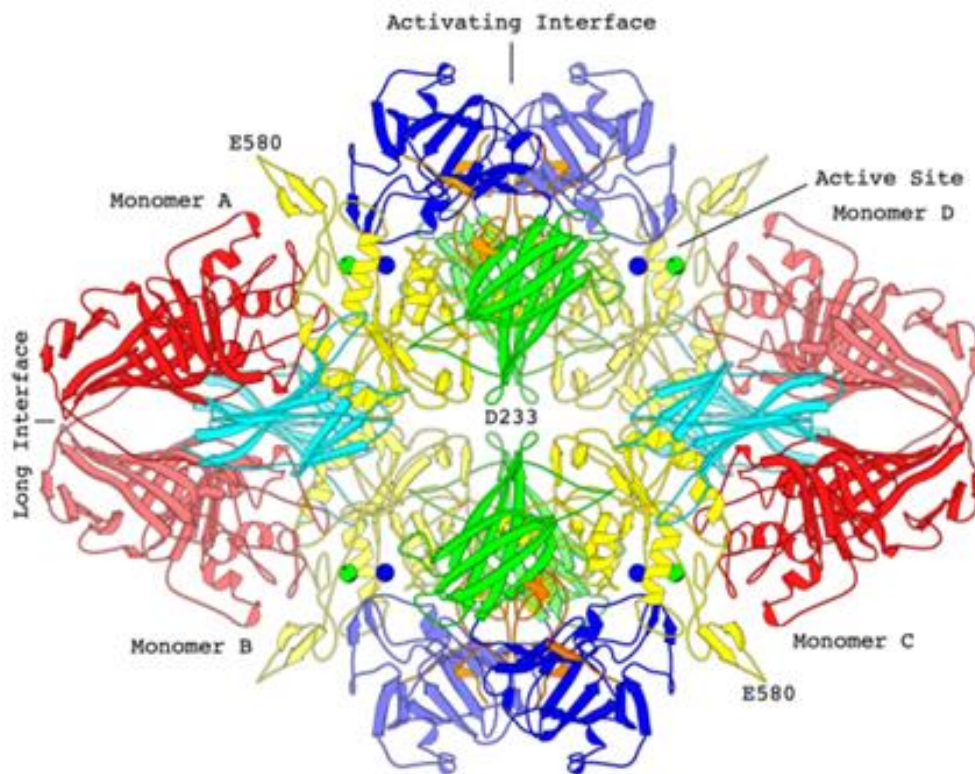
### 2.2.1 Általános jellemzők

A béta-galaktozidáz (EC 3.2.1.108), vagy ismertebb nevén laktáz, ipari szempontból, az egyik legfontosabb biokatalizátor. Az enzim felfedezése a holland Martinus Willem Beijerihez fűződik, akinek a laktóz fermentációja felkeltette az érdeklődését. Kutatása során a *Saccharomyces tyricola* és a *Kluyveromyces marxianus* nevű élesztőgombákat különítette el, amelyek képesek voltak a tejcukor bontására. Az általuk termelt enzimet nevezte el laktáznak [Rouwenhorst et al., 1989]. Enzimatis reakció típusát tekintve exoglikozidáz enzim. Természetes szubsztrátuma a laktóz, amit glükózzá és galaktózzá képes bontani. A világ lakosságának jelentős százalékában nem található meg ez az enzim, éppen ezért laktózintoleranciában szenvednek, vagyis laktóz fogyasztást követően hasi panaszok jelentkeznek náluk. Erre a problémára megoldásként szolgál, ha a laktóz tartalmú tejtermékeket előkezeljük  $\beta$ -galaktozidázzal vagy, ha az érintett személy laktóz fogyasztás előtt  $\beta$ -galaktozidáz enzimet visz a szervezetébe.



### 2.2.2 Az enzim szerkezete

A  $\beta$ -galaktozidáz enzim négy azonos alegységből épül fel, amelyeket másodlagos kötések tartanak össze, vagyis a fehérje tetramer szerkezetű. Mind a négy polipeptid lánc 1023 aminosavból épül fel, mely 5 különböző doménre osztható fel. Ezek közül kiemelendő a középső domén, ugyanis itt található az enzim aktív helye. Felépítését tekintve egy triózfoszfát-izomeráz (TPI vagy TIM) vagy egy  $\alpha_8\beta_8$  henger. A henger C-terminális végénél egy mély bemélyedésként figyelhető meg az aktív hely (Juers et al., 2012).

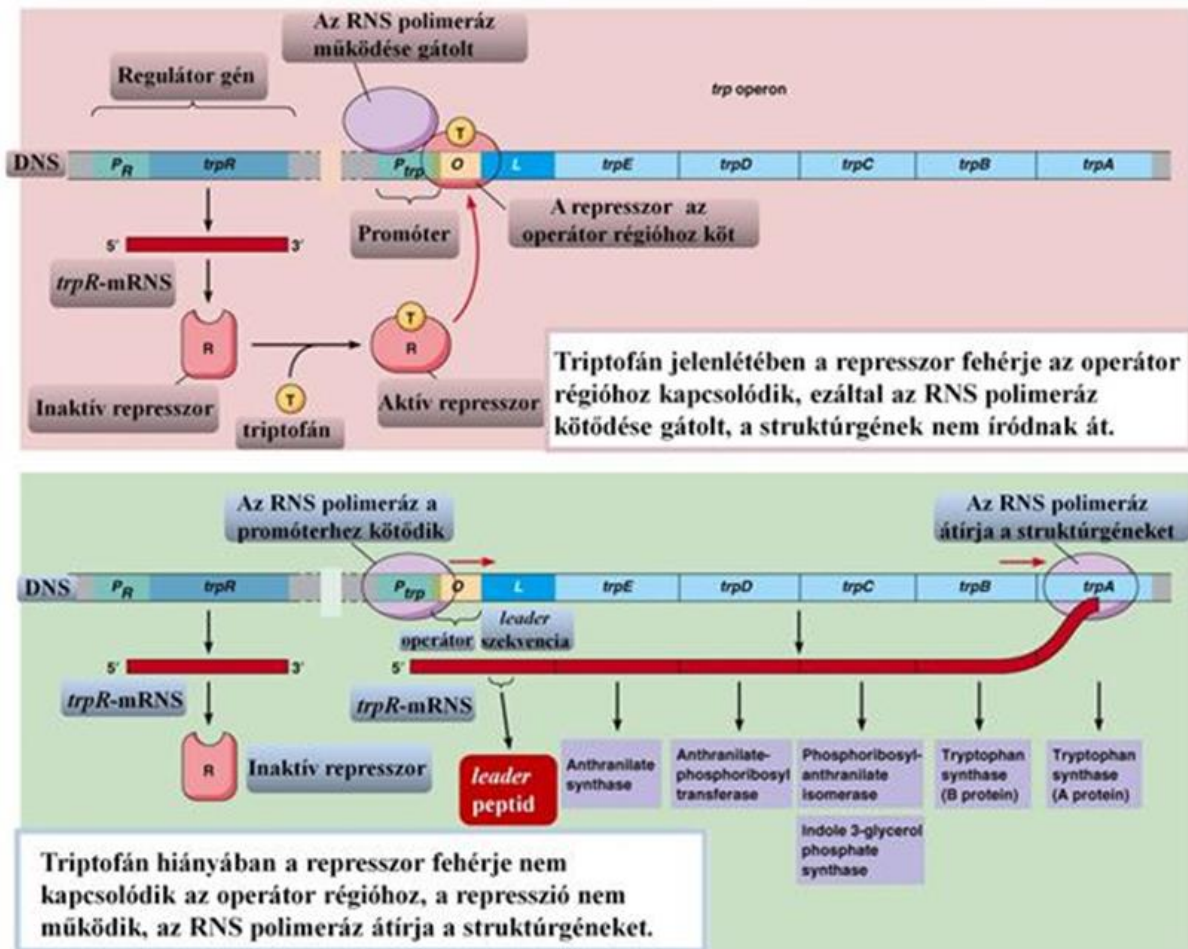


3. ábra: A  $\beta$ -galaktozidáz tetramer szerkezete (Forrás: Juers et al., 2012)  
A zöld gömbök a  $\text{Na}^+$  ionokat míg a kékek a  $\text{Mg}^{2+}$  ionokat jelölik.

A 3. ábra jól bemutatja a fehérje szerkezeti felépítését. A  $\text{Na}^+$  és  $\text{Mg}^{2+}$  ionok elengedhetetlenek a fehérje működéséhez, ugyanis, a szükséges ionok jelenléte nélkül a szerkezet könnyen destabilizálódhat és széteshet a molekula.

### 2.2.3 A lac operon működése

Francois Jacob, Jacques Monod és André Lwoff 1958-ban kezdték el kutatni, hogy a mikroorganizmusok, hogyan bontják le a szénhidrátokat monoszacharidokra (Deák, 2016). Munkásságuknak köszönhetően ismertünk meg először egy génszabályozási rendszert, vagyis a lac operon működését (4. ábrán látható).



4. ábra: Lac operon ábrázolása és működésének illusztrálása (Forrás: Deák, 2016)

A lac operon a laktóz lebontásáért felelős géneket hordozza. Felépítését tekintve három kódoló régióra osztható, a LacZ, ami a  $\beta$ -galaktozidáz enzim génjét kódolja, a LacY itt található a laktóz permeáz enzim gén és a LacA, ami a transzacetiláz fehérjét tartalmazza. Ezeken felül egy regulátor régió van jelen az operonban, ez tartalmazza a CAP kötő helyet. Alacsony glükóz koncentráció esetén a cAMP megköt rajta, így megindulhat a fehérje termelés. Az operon része még a LacI, ami a repressor fehérje génjét kódolja. A LacI gén expressziója konstitutív, de a fehérje kötődése a lac operonhoz reverzibilis, így kevés  $\beta$ -

galaktozidáz mindig termelődik. Abban az esetben, ha laktóz is jelen van a rendszerben, a  $\beta$ -galaktozidáz elkezd fele-fele arányban lebontani és átalakítani a laktózt. A lebontás során glükóz és galaktóz keletkezik, míg az átalakítás esetében transzglykolizáció játszódik le és allolaktóz keletkezik. Az allolaktóz hozzákapcsolódik a represszor részhez, így a megváltozott represszor fehérje nem tud kötődni a lac operonhoz, és megkezdődhet az enzim szintézise (Deák, 2016).

#### **2.2.4 Az enzim forrásai**

$\beta$ -galaktozidáz enzimet, növényekben, állatokban és mikroorganizmusokban is megtalálhatjuk. Az emberi szervezet is képes termelni laktáz enzimet, a vékonybélben, laktózintoleránsnak nevezünk azokat az embereket, akik nem képesek az enzim termelésére. Ipari szempontból a mikroorganizmusokból, különösen származó enzim a legjelentősebb. Az 1. táblázat összefoglal néhány  $\beta$ -galaktozidáz termelő mikroorganizmust

**1. táblázat A  $\beta$ -galaktozidáz mikroorganizmus forrásai**  
(Forrás: Panesar et al., 2010)

<b>Baktériumok</b>	<b>Penészek</b>	<b>Élesztők</b>
<i>Alicyclobacillus acidocaldarius</i>	<i>Alternaria alternate</i>	<i>Bullera singularis</i>
<i>Arthrobacter sp</i>	<i>Aspergillus foetidus</i>	<i>Candida pseudotropicalis</i>
<i>Bacillus acidocaldarius</i>	<i>A. carbonarius</i>	<i>Saccharomyces anamensis</i>
<i>B. circulans</i>	<i>Auerobasidium pullulans</i>	<i>S. lactis</i>
<i>B. subtilis</i>	<i>Curvularia inaequalis</i>	<i>Kluyveromyces bulgaricus</i>
<i>Bacteriodes polypragmatus</i>	<i>Mucor meiehei</i>	<i>K. lactis</i>
<i>Bifidobacterium bifidum</i>	<i>Neurospora crassa</i>	<i>K. marxianus</i>
<i>Clostridium acetobutylicum</i>	<i>Penicillium expansum</i>	
<i>Corynebacterium murisepticum</i>	<i>Saccharopolyspora rectivergula</i>	
<i>Enterobacter agglomerans</i>	<i>Streptomyces violaceus</i>	
<i>E. cloaceae</i>		
<i>Escherichia coli</i>		
<i>Klebsiella pneumoniae</i>		
<i>Lactobacillus acidophilus</i>		
<i>L. bulgaricus</i>		
<i>L. helveticus</i>		
<i>Leuconostoc citrovorum</i>		
<i>Pseudomonas fluorescens</i>		
<i>Streptococcus cremoris</i>		

### 2.2.5 Az enzim felhasználási területei

A laktóz megtalálható a tejben és tejtermékekben is. Mivel a világ lakosságának nagy része laktóztoleranciában szenved, így az élelmiszeripar számára a  $\beta$ -galaktozidáz enzim rendkívül értékes. A laktáz enzim segítségével kezelhetők az élelmiszerek, hogy eltávolítsuk azok laktóz tartalmát, vagy  $\beta$ -galaktozidáz tartalmú tablettá bevitelével segíthetjük a laktóz tartalmú élelmiszerek emésztését az emberi szervezet számára. Ezen felül egyéb felhasználási lehetőségek is megjelennek az élelmiszeriparban. A jégkrémek és a sűrített tejek könnyen kristályossá válhatnak nagy laktóz koncentráció esetén, ezzel rontva a termék textúráját.  $\beta$ -galaktozidáz kezeléssel a nem kívánatos kristályosodás megelőzhető, illetve mivel a reakció termékek (glükóz és galaktóz) a laktóznál édesebb ízűek, kevesebb édesítőszerrel kell hozzáadni a termékek kívánt ízének eléréséhez (Rodríguez et al., 2006).

Kiemelendő felhasználási terület még a tejsavó kezelése. A sajtgyártás során nagy mennyiségű tejsavó keletkezik, ami vízbe jutva növeli a víz biológiai és kémiai oxigén igényét, ami erősen károsítja a vizek élővilágát. A tejsavó fő alkotó elemei a laktóz (3-5%), ásványi anyagok és fehérjék (1%).  $\beta$ -galaktozidázos kezelés segítségével a tejsavóban lévő laktóz lebontható, és a keletkező termék kitűnő táptalajként szolgálhat mikroorganizmusok számára.

További fontos felhasználási lehetőség a  $\beta$ -galaktozidáz enzimet illetően a galakto-oligoszacharidok (GOS) előállítása. A GOS kiválóan alkalmasak prebiotikus élelmiszerek előállítására. Mivel emésztetlenül képesek eljutni a bélrendszerhez, ahol tápanyagként tudnak szolgálni, a kedvező bélbióta számára, pozitívan változtathatják a bélrendszer mikroorganizmus populációját. Galakto-oligoszacharidok mono- és diszacharidokból állíthatók elő transzglykolizáció vagy poliszacharidok hidrolízise segítségével (Saqib et al., 2017).

## **2.3 Bioinformatika**

A bioinformatika egy interdiszciplináris tudomány, ami magába foglal biológiai ismereteket, valamint matematikai és informatika módszereket. Fő foglalkozási területe a molekuláris biológiai információk feldolgozása informatikai eszközökkel (Rana, 2012).

Ahogy egyre több és komplexebb információk váltak ismertté, úgy lett egyre lehetetlenebb a számítógépek hiányában történő kutatás, legyen szó gyógyászati vagy mezőgazdasági területről, az informatika használata elkerülhetetlenné vált. Az elektronikai eszközök által nyújtott praktikus adattárolás, illetve gyors statisztikai elemzési módszerek rendkívül hasznosnak bizonyultak a molekuláris biológia és a genetika tárgykörében. A genomok, de még a fehérjék szerkezete is túlságosan komplex, ahhoz, hogy a kutatók képesek legyenek az azokról szerzett információkat gyorsan feldolgozni.

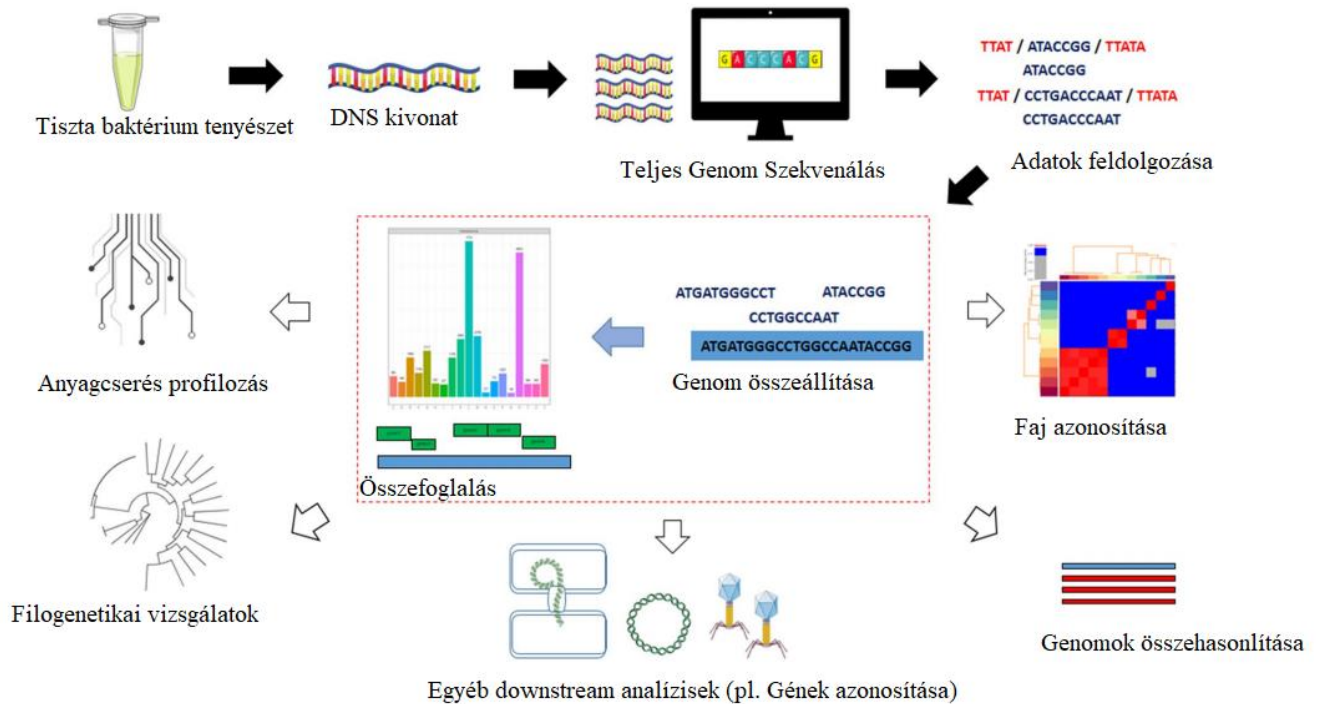
### **2.3.1 Bioinformatika történelme**

Ahhoz, hogy a bioinformatika kialakulhasson, először fontos fejlődésre volt szükség a molekuláris biológiában. Az első fontos mérföldkő 1953-ban történt, amikor James Watson és Francis Crick publikálták a DNS kettős hélix szerkezetét. Ennek következményeképp merült fel az igény a hatalmas mennyiségű biológiai adatok hatékony kezelésére és elemzésére. Az 1960-as és 1970-es években a számítógépek elterjedése és fejlesztése lehetővé tette a biológusok és a matematikusok számára, hogy közösen dolgozzanak az adatok feldolgozásán. Így az első bioinformatikai algoritmusok és szoftverek is ekkor alakultak ki. Nagyon fontos lépés volt az NCBI (National Center for Biotechnology Information) létrehozása az Egyesült Államokban, az DDBJ (DNA Databank of Japan) a Japánban és EMBL (European Molecular Biology Laboratory) a 10 ország összefogás eredményeként Európában, amelyek a létrehozott adatbázisrendszerekben gyűjtik és tárolják a biológiai információkat.

Az 1980-as és 1990-es évek elején a génszekvenálási technológiák nagy előrelépést jelentettek a biológiai kutatásban. Ennek hatására elindulhatott a humán genom projekt, amelynek célja a humán genom teljes mértékű szekvenálása volt. Ez a tudományág azóta is folyamatosan és dinamikusán fejlődik. Új módszerek és szoftverek kerülnek kifejlesztésre annak érdekében, hogy a kutatók minél hatékonyabban végezzenek kutatást többek között a genetika és a proteomika tárgykörében (Gauthier et al., 2019).

### 2.3.2 Bioinformatika módszerei

A molekuláris biológiai, genomikai technológiák robbanásszerű fejlődése a biológiai adatok mennyiségének exponenciális növekedéséhez vezetett, ennek következményeként egyre komplexebb vizsgálati módszerek lettek lehetségesek.



### 5. ábra: Általános folyamatára a genomok összehasonlítására bioinformatikai módszerekkel

(Forrás: Mendoza et al., 2022)

Ahogy az 5. ábra is szemlélteti a bioinformatika tudománya számos kutatási területet érinthet, amelyek a fehérjék vagy a DNS szekvenciájának analízisén alapulnak. A vizsgálatokból szerzett információk halmaza egyre csak növekedett, ennek hatására az 1990-es években létrejött az INSDC (International Nucleotide Sequence Database Collaboration) az-az a Nemzetközi Nukleotid-szekvencia Adatbázis Együttműködés. Az alábbi felsorolt három intézmény képezi az együttműködést:

- Az amerikai National Center of Biotechnology Information (NCBI)
- Az európai European Bioinformatics Institute (EMBL/ENA)
- A Japán DNA Databank of Japan (DDBJ)

Az INSDC próbálta a világ különböző médiákban (tudományos folyóirat, riportok, szabadalmak stb.) publikált szekvenciákhoz köthető információt feltárni, begyűjteni és feltölteni az adatbázisba. Ezek legfőképpen nukleinsavszekvencia adatbázisok, így a tárolt adatok leginkább az elsődleges és specializált DNS-szekvenciákhoz fűzhetők. Későbbiekben a fehérjeszekvenciákra is kiterjesztették. Ezeket az információkat folyamatosan szinkronizálják egymás között a konzorciumi partnerek.

A leginkább nukleinsavszekvenciákkal foglalkozó adatbázisokon túl kiemelkedően fontosak még a fehérjeszekvenciákkal foglalkozók. Ilyenek például a SWISS-PROT, a BRENDA, az ENZYME, a CATH stb. Ezeken az oldalakon a fehérjeszekvenciákon túl általános információkat találhatunk a fehérjék működéséről/funkciójáról, vagy bármi egyéb a fehérjékhez köthető adatokról.

Szintén fontosak még a térbeli szerkezetekkel foglalkozó adatbázisok. Ilyenek például a PDB (Protein Data Bank), vagy a SCOP (Structural Classification of Proteins). Ezeken az oldalakon 3D-ben ábrázolva figyelhető meg a fehérjék szerkezete, és egyéb a fehérjékhez köthető információk.

Bioinformatikai vizsgálatok során alkalmazott kiemelkedő módszerek:

- BLAST (Basic Local Alignment Search Tool) eszközt 1990-ben publikálták. A kutatók legfőbb szempontjai a fejlesztés közben a gyorsaság és az egyszerűség voltak, hiszen az ezelőtti módszerek, vagy csak nagyon lassan voltak képesek elemezni nagyméretű adathalmazokat, vagy olyan kapacitású számítógépekre volt szükség, amelyek nem könnyen elérhetőek. A BLAST módszer heurisztikus megközelítésen alapszik, és mivel nem globális, hanem lokális hasonlóságot keres a vizsgált szekvenciák között, így rövid azonosságokat vizsgál a bázispárossorrendek között, majd egy pontozási mátrix segítségével, a hasonlóság mértéke kvantitatívan kifejezésre kerül. A leggyakrabban használt mátrix a BLOSUM62 (Altschul et al., 1990). Kapott eredmények jellemzése a 2. táblázatban foglalt szempontok alapján történik.
- A klaszterezés lényege a sejtek génjeinek csoportokba osztása a szekvenciák hossza és hasonlósága alapján. Kettő fő megközelítés létezik, a bottom-up (lentől-felfelé), és a top-down (fentről-lefelé). A bottom-up lényege, hogy minden vizsgált objektum külön kategóriában kezd, és ahogy haladunk a vizsgálat fázisain át a halmazok száma, egyre csökken míg minden adat egy csoportba nem kerül. Ezzel szemben a top-down esetén



kezdetben van néhány előre meghatározott csoportunk, amikbe a vizsgált objektumok beosztásba kerülnek (Kapetanovic, 2004).

- MSA (Multiple Sequence Alignment) az-az többszörös szekvenciaillesztés módszerét általában 3 vagy több biológiai szekvencia összeillesztésére használjuk. Ezeknek hasonló hosszúságúaknak kell lenniük, összehasonlítás eredményeképpen információt kapunk a biológiai objektumok evolúciós kapcsolatáról (Chatzou, 2015).

**2. táblázat Fehérjék szekvenciaazonosságának mértékének jellemzése (Forrás: Ludányi, 2012)**

Szekvencia azonosság (%)	Zóna	Megjegyzés
50-100		Biztosra vehető a homológia és a nagyfokozatú szerkezeti hasonlóság erős valószínűsíthető az azonos vagy rokon funkció
20-50		Biztosra vehető a homológia és a jelentős szerkezeti hasonlóság a funkció feltételezhetően azonos vagy rokon
20 körül	alkonyzóna (twilight zone)	Kétségesse válik a homológia és a szerkezeti hasonlóság ekkora hasonlóság véletlenül is kialakulhat, egymással nem rokon fehérjék között is kifinomult módszerek szükségesek a homológia detektálásához a szekvencia sokszor nem elég, térszerkezeti információ is szükségessé válik a rokonság mértékének eldöntéséhez
0-10	éjjfélzóna (midnight zone)	Esetleg fennállhat funkcionális és szerkezeti hasonlóság a két fehérje között, de ezt a szekvencia alapján nem lehet eldönteni az esetleges rokonságot csak más információk alapján (elsősorban térszerkezet) lehet megállapítani

### 2.3.3 A tejsavbaktériumok bioinformatikai vizsgálata

A tejsavbaktériumok bioinformatikai vizsgálata egy nagyon fontos és gyakori kutatási terület napjainkban. Ezen baktériumok genomjának szekvenálása, és az általuk termelt fehérjék szerkezetének elemzése értékes információt nyújtott mind az élelmiszer-, mind a gyógyszeripar számára. Az új generációs szekvenálási technológiáknak köszönhetően a tejsavbaktériumok valamennyi nagyobb csoportjának teljes genom készlete ismertté vált. Továbbá a genom összehasonlító kutatásoknak köszönhetően jobban megérthetjük a tejsavbaktériumok alkalmazkodó képességét akár laboratóriumi akár természetes körülmények között. Ezeknek az ismereteknek köszönhetően e mikroorganizmusok élelmiszeripari alkalmazhatósága, illetve az emberi szervezetre kifejtett hatásuk jobban értelmezhető (Douillard és De Vos, 2014)

### 3 Alkalmazott módszerek

#### 3.1 Vizsgált baktériumok és kromoszóma NCBI azonosítók

Az általam vizsgált fajokat és a hozzájuk tartozó kromoszómák azonosítóját a 3. táblázatba gyűjtöttem.

3. táblázat: Vizsgált baktériumok és azok kromoszóma azonosítói

Faj neve	Kromoszóma azonosító
<i>Lactobacillus helveticus</i>	CP094500.1
<i>Lactobacillus acidophilus</i>	CP005926.2
<i>Lactocaseibacillus paracasei</i>	CP000423.1
<i>Limosilactobacillus fermentum</i>	CP124737.1
<i>Lactiplantibacillus plantarum</i>	CP028221.1
<i>Lactobacillus reuteri</i>	LN906634.1

#### 3.2 Béta-galaktozidáz enzimet kódoló gének azonosítása

Az NCBI adatbázis (<https://www.ncbi.nlm.nih.gov/>) tartalmazza az egy adott kromoszómán kódolt összes gén felsorolását. Ez alapján összeírtam az általam vizsgált kromoszómákban kódolt béta-galaktozidáz enzim kódoló géneket.

#### 3.3 Vizsgált fehérjék konzervatív régiói

NCBI adatbázis (<https://www.ncbi.nlm.nih.gov/>) segítségével összeírtam a vizsgált fehérjék konzervatív régióit.

#### 3.4 MSA és filogenetikai vizsgálat

Genomenet webserver (<https://www.genome.jp/>) segítségével többszörös szekvencia illesztést (MSA) hajtottam végre, és a kapott eredményekből filogenetikai fát készítettem.

### **3.5 Gének eloszlásának vizsgálata**

Az NCBI adatbázisból ki tudtam írni az általam vizsgált gének helyét a kromoszómákban. Ezen gének eloszlását a Phenogram webszerver (<https://ritchielab.org/software/phenogram-downloads>) segítségével vizualizáltam.

### **3.6 Motif analízis és gén struktúra**

Meme suite webszerver (<https://meme-suite.org/meme/>) segítségével motif vizsgálatot hajtottam végre. Ezt követően a Gene Structure Display alkalmazásával vizualizáltam a vizsgált gének kódoló és nem kódoló régióit.

### **3.7 Különböző eredetű béta-galaktozidáz enzimek jellemzése**

ProtParam-EXPASY webszerver (<https://web.expasy.org/protparam>) segítségével meghatároztam a vizsgált béta-galaktozidáz enzimek molekulatömegét, izoelektromos pontját, alifatikus és GRAVY értékét, továbbá az instabilitási mutatóját. Ezt követően a CELLO program alkalmazásával meghatároztam a sejten belüli elhelyezkedésüket.

## 4 Eredmények és kiértékelésük

### 4.1 Béta-galaktozidáz enzimet kódoló gének azonosítása

Az NCBI adatbázis segítségével 12 különböző béta-galaktozidáz gént sikerült elkülönítenem a hat vizsgált faj között. Ezeket a géneket a 4. táblázatban foglaltam össze. Ezek közül négy faj, a *L. helveticus*, *L. fermentum*, *L. plantarum* és a *L. reuteri* két alegységre bontva kódolja a béta-galaktozidáz enzimet. A nagyobbik egység a LacL míg a kisebbik egység a LacM génben található. Ezzel ellentétben a *Lacticaseibacillus paracasei* egyben kódolja az enzimet, míg *Limosilactobacillus fermentum*-ban csak a LacM gén volt megtalálható.

4. táblázat: A vizsgált kromozómákon azonosított béta-galaktozidáz gének

Név	Génneve	Fehérjekód	Kromoszóma azonosító	Kezdet	Vég	Gén hossza	Fehérje hossza
<i>L. helveticus</i>	LheBG1	UOE24092.1	CP094500.1	645676	646632	957	318
<i>L. helveticus</i>	LheBG2	UOE24093.1	CP094500.1	646616	648502	1887	628
<i>L. acidophilus</i>	LacBG1	AGK94524.1	CP005926.2	1335728	1337770	2043	680
<i>L. acidophilus</i>	LacBG2	AGK94619.1	CP005926.2	1442722	1444725	2004	667
<i>L. acidophilus</i>	LacBG3	AGK94623.1	CP005926.2	1449595	1451481	1887	628
<i>L. acidophilus</i>	LacBG4	AGK94624.1	CP005926.2	1451465	1452415	951	316
<i>L. casei</i>	LcaBG1	ABJ69242.1	CP000423.1	387369	388850	1482	992
<i>L. fermentum</i>	LfeBG1	WGW20706.1	CP124737.1	1199367	1200329	963	320
<i>L. plantarum</i>	LplBG1	QHM27543.1	CP028221.1	1147673	1149553	1881	626
<i>L. plantarum</i>	LplBG2	QHM27544.1	CP028221.1	1149537	1150496	960	319
<i>L. reuteri</i>	LreBG1	CUU13181.1	LN906634.1	1680169	1681128	960	319
<i>L. reuteri</i>	LreBG2	CUU13182.1	LN906634.1	1681112	1682998	1887	628

A vizsgált enzimek között a leghosszabb az LcaBG1 génen kódolt, amely a 992-es értéket éri el, mivel az *L. casei* kromozómáján a fehérje egy génen kódolva található meg. Ezzel szemben a legrövidebb enzim az *L. acidophilus* LacM génjén kódolt béta-galaktozidáz, amelynek a hossza csupán 316. Átlagosan LacL géneken kódolt fehérjék hossza 645,8, míg a LacM géneken kódolt fehérjék hossza: 318,4. A LacL és LacM gének egymás mellett vannak kódolva a vizsgált kromozómákon.

## 4.2 Vizsgált fehérjék konzervatív régiói

A lekérdezett adatokból az alábbi konzervatív régiókat kiválasztottam az értékelésre.

### *L. helveticus*

#### LheBG1:

1 mdytnnqlhi iygdatlgnv gkdfqyifsy erggleslkv hgkewlyrvp tptfwrattd  
61 ndrsgsfnlk aaqwlgadmf tkctdihlkv drhdfaelpi apfnnkfsnh eyaksaeisf  
121 tyqtlttpat nakiiynidd vghikvtmry ygkkglpplp vigirlimpt aatgfdyegl  
181 sgetypdrma gakegkfhid glpvteylvp qengmhmqtq kltinrettq nnvdrtnekf  
241 slsiqqaekp fnfsclypta eelenathie elplvrrtvl viagavrgvg gidswgt dve  
301 sayhinpeld hefsfiln

#### LheBG2:

1 mqaninwldn pevfrvnqlp ahsdhpffrd yrewqkqhss yqqslnqkwk fhfsanpmdr  
61 pqdfyqrdfd ssnfdsipvp seielsnytq nqyinvlfpw egkifrrpay aldndheeg  
121 sfskgadntv gsylkrfdls saligkdvhi kfegveqamy vwlngfhvgy aedsftpsef  
181 dltpyiqdkd nllavevfkh staswledqd mfrfsgifrs vellgipath lmdmdlkprv  
241 adnyqdgifn lklhfigkka gsfhllvkdi kghtllekne dikenvqinn ekfenvhlwn  
301 nhdpolyqll ievydeqqnl leliphqfgf rrieispek vllngkrlii ngvnrhewda  
361 krgrsitmsd mttintfke nninavrtch ypnqipwyyl cdqngiyvma ennleshgtw  
421 qkmgeiepsd nvpqgsipqwk eavidrarin yetfknhtsi lfswlgnesy agdniiamne  
481 fykshddtrl vhyegvvhrp eldkisdve scmylppkkv eeylqndppk pfmeceymhd  
541 mgnsdggmgs yiklldkypq yfggfiwdfi dqallvhdei sghdvlyrygg dfddrhdye  
601 fsgdglmfad rtpkpmqev ryyyglhk

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

*L. acidophilus*

**LacBG1:**

1 mkrelkskvf lhggdynpeq wlgepeiine dfalfknaai ntvvgifsw aklepegky  
61 dfawlddifd rvekmngyvi latpsgarpa wlarkypevl rtdfnnqkrq fggrnhhclt  
121 spiyrrkkvre intklaehfg krpslilwhi sneysgecyc dlcqqafrdw lkkyrtler  
181 nhswwntfw shtfsdwnqi hapsplsemg nkgmnlwkr fvsdqaisfi dneveplrki  
241 tseipvttnm magnplmdp f tgy nyqemak hldviswdsy plwgn dfqst eklgqnv gli  
301 h dffrslkhq nfmimentps rvnwadidra krp gmhqlas lq diahssds vlyfqlras  
361 gsaemfhgav iehrhpektr vfhdvkdvgh dleklesiys tsytakv gi vydynniwal  
421 edaegyskdk kiwqtisqy qyfyqndipv dfvspndnft qyklidpmh flmtkeymdk  
481 lesfvkkcgy vvgtyisgvv denglaymne wpkqlqsiyg iepletdsly pkqsnsiefa  
541 ghryqaydfc etifkhdakv lakyttdfys gtpaltahkc gegkgyiac rtdtdflsai  
601 ygqivkeldl lplnpikket tkislqvren ddekylfvqn fsheqqsil kqkmkemsd  
661 efeenkvivk pygtkiyqmn

Béta-galaktozidáz GanA: Ez a régió felelős a szénhidrát transzportálásáért és metabolizmusáért

**LacBG2:**

1 mtqlsrflyg gdynpdqwe etwskdihvf kkadinsati nifswallep regkynf skl  
61 dkvvqqlsda nfdivmgtat aampawmfkk ypdia rvd yq drrhvfqqrh nfcpsnsnyq  
121 rlagelvkql verykdnkhi vvwhinne y gncycencqn afrkwlnky ktveglnkaw  
181 nmnvwshtiy dwdeivvpne lgdvwgiegs etivaglsid ylrqsesmq nlfkmekkii  
241 kkydpetpvt tnfhglpnkm vdyqkwakgq diisydsypt ydapaykaaf lydlmrsikh  
301 qpfmlmesap sqvnwqpysp lkrpgqmeat efqavahgad tvqffqlkqa vggsekfhsa  
361 viahsqrtdt rvfeladlg kklknagpti lgsstkakva ivfdwsnfws yeyvdgitqd  
421 inyvsildy yrqfyernip tdiigvddd snydlv vapv lymvkhgldk kindyvengg

481 nfvttymsgm vnssdnvylg gypgplkev giwveesdav vpgqkikvln ngkdydtgli

541 cnlihpndak ilatyasefy agtpavtenq ykggrawyig trlehqgtq lfnhiifetg

601 veslvcdshk leitkrvted gkelyfvlnm sneertlpsk ftgyediltg ekahkdmkgw

661 dvqvlrn

Glyco\_hydro\_42: Ez a szakasz felelős a béta-D-galaktozidáz maradék nem redukáló végének hidrolizálásáért

Glyco\_hydro\_42M: Ez a szakasz felelős a trimelizálási folyamatban.

Glyco\_hydro\_42C: Ez a szakasz a béta-galaktozidáz enzim C-terminális része.

### LacBG3:

1 mqanikwldd peifrvnqlp ahsdhpffkn yrewqnnhss fkqslngmwq fkfskdpqsr

61 pvdfykkdfn tssfttipvp seiennfaq nqyintlypw egkifrrpay alnksdaeeg

121 sfsegkdntv gsyikhfdln pelrdhdihi vfegaeramy vwlngghfigy aedsftpsef

181 dltkyikekd nilavevfkh staswledqd mfrfsglfrs vellafpeth lvldlkptv

241 cdnyqdgifn aelkftgsln ghvhlsvedv ngsaileqdv pldsevefts stlenihlwd

301 nhhpilyqll ievhdenghl velipyqfgf rrieinqdkv illngkclii ngvnrhewna

361 ktgrsitlnd mekdidtke nninavrtrch ypnqipwyyl cdqngiyymma ennleshgtw

421 qkmgqvpsd nvpgsvpewr eavidrarn yetfknhtsi lfwslgnesy agsnivamne

481 fykehdsrl vhyegvvhrp elkdqisdie shmylppkev aeylrnnpqk pfmeceymhd

541 mgnsdggmgs yiklideypq yvggfwdfi dqallvydpi skqnlvrygg dfddrhdye

601 fsgdglmfad rtpkpamqev ryyyglhk

lacZ super family: Béta-galaktozidáz enzimet kódoló szakasz.

### LacBG4:

1 maytnnlqii ygdatlgitg knfhylfsye rggleslnin nkewlyrvpt ptfwrattdn

61 drngnfnlka sqwlgadmft kctkielkvd drqfdelpia pinnqfsnhe yadhvqiafw

121 yqtltnpatd vkiiynidnt gcinivmhyf gkkgplpplv igmrfimpta atgfdyegls

181 getypdrmag akegkfhvdg lpvtkylvpq engmhmqtka lkitrsstln nadqesefsl

241 klkqdkqpfm fsclpytaee lenathleel plarrtvlvi agavrgvggi dswgadvekq

301 yhinpekdye fsfnln

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

### *L. casei*

LcaBG1:

1 mttaklwekp eltdinrmap rshfqtfpg enrqrhyql lngtwqfkfl dapeyapedf

61 mavdfndqdw dqipvpsnwq lqgygkmhys dlwynfpinp pfvpsenptg lyrrtftvde

121 vavneqyig fdgadsafkl ylngdfigs kgarlpsefd vtalkqgtn tiavevvqws

181 dgtyledqdm wwslglfrdv slysrpqngl ydvrvrtyll kdyragelvv tptlsgavps

241 kihyeltdkg atlidqtlst dvsldvtlnd iqawsaeapn lydltmtvlq ndaplevvrq

301 rigfrqieln gktflvngka ikfkgvnmhd ysategrvms eadfkkniiis mkrnninair

361 tahypkapyf ydlcdelgmy videtdlech gfelterydw itddprwkta yvdrmrtrlq

421 rdknhpaiim wslgnesdfg dnframaayc kaedprlvh yegdfeaevs dvystmytwl

481 ehdtkmtmad vlqktqkphi lceyahsmgn gpgnlkeyqd lfyghqqlqg gfiwewfdqg

541 vaaqqgdqty yryggdfgdq pnnsnfcidg lirpdgqpst altevkkkte pfqmtvrdlp

601 tqtitvtnrl dflssdqfnf gyeleadgkl matgkidlpt imagtttik ldielpkldp

661 eviynlhvlt elknqtswad agtvlsqtvv nlqrpqhmt hqqtalqas enattimvtg

721 gqneyrfdki kgtfslthdg hkliadgikm nfwrapidnd myllddyynk yflnlwhest

781 revqlhpqtn gdyvvnltkq vgttnsgwyy liqqqytmhq dgsfdldvig kasgkrdmap

841 emlprigvkm tlpkayqqvs ydglgpteny sdshqaayys hftssvddlf vnyvqpqeng

901 nhmtdtdqial tdgqdqlvtv makplnfsvs nyadetleaa khtidlkkssd alnlyldfrq

961 nglgtnscgq nqlkrhrckf ddfelgfnfk vn

ebgA: béta-galaktozidáz alfa alegységet kódolja.



*L. fermentum*

**LfeBG1:**

1 mdytnklhvv yddnilgldg kdfqylfsye qggpesfkik gkewlyrspr ptfwrattdn  
61 drngngfvss vqwlaadyvl pcqodialqvd gkdkklplap ktrysnqef akkvkitfty  
121 qtqtpattv qvsytkasg kikvnyhytg aqlpslpvlg wrmimtpat sfdyeglsge  
181 typdrmaggi egtyhveglp vtpylvpqen gmhmankwvq itrattlnna dpdaapfrlk  
241 feapkkgkln fsclpytsae lenathpeel paahrtvivi agevrgvggi dswgadveek  
301 yhidatvdhd fsfkivpeln

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

*L. plantarum*

**LplBG1:**

1 mqanlqwldd pevfrvnqlp ahsdhhyhd taefktsrf ikslngawrf nfaktaerp  
61 vdfyqpdfda tdfdtiqvpg hielagyqqi qyintlypwe gkiyrrpyt lnqdltpgl  
121 fsdaadntvg sylktdldd afkgqriiq fggveealyv wlnghfigya edsftpsefd  
181 ltpyiqdqn vlavrvykrv taafiedqdm frfsgifrdv nilaepashi tlddirvppn  
241 anlksgeini ttkvtgepat laltvkdhdg rvltsqtqg sgsvtfdtml fdqlhlwspq  
301 tpylyqltie vydadrqlle vipyqfgfirt velrddkviy vnnkrlving vnrhewnaht  
361 grvismddmr adiqtmlann inadrtchyp dqlpwyqlcd eagylmaen nleshgswqk  
421 mgaiepsynv pgdnphwlaa vidrarsnye wfknhpsiif wslgnesyag ediaamqafy  
481 kehddsrlvh yegvvhtpel kdridvesr myekpqniva ylednptkpf ldceymhdmg  
541 nslggmqsyn dldikypmyq ggfiwdfidq alfvhdpitd qdvlryggdf derhsdyefs  
601 gdglmfadrt pkpamqevky yyglhk

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

## LplBG2:

1 maytnnqlhv iygdgslglq ganfhylyfsy erggleslvv ndkewlyrtp tpifwrattd

61 ndhgsgfsvk saqwyaadkf stcqdietv ddqpvtpipi aplnnkytdh eiatkvsly

121 hfvttvpst ivtvytvta dgqiniathy sgqsdlpelp afglrfiip t atgfdytgi

181 sgetypdrla gathgrfhvd slpvtpylvp qecgmhmqte qvtvtrsttq nnadhdntpf

241 sltfsqadap fafsclypyta aelenathme elplarrtvl siygavrgvg gidswgtve

301 spyhipadqd idfsfnihf

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

## *L. reuteri*

### LreBG1:

1 maytnklrvi ygdatlglsg dsfhylyfsye rggleslkn gkewlyrepm ptfwrattdn

61 drgsgfnirs aqwlaadtfh kevgidltvd nqhaelpia pitnefsdpv saenvkikyt

121 fetltpatq vtvyevnrq geikvtmhyy ghedlpglpv vgmrfimptv atgfdyqgl

181 getypdrmag ategtfhvdg lpvtkylvpq engmhmdtha ltitrdstqn nadhsrepfn

241 ltikqdeqpf afsclpytae elenathiee lplarrtvlv vagavrgvgg idswgadvee

301 qyhipadrdrv efsfvlnak

béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

### LreBG2:

1 mdadikwlde petfrvnqlp ahsdhyyygn ydewrhnsr faqndgqwq fnfaenprkr

61 endfykvdyd sssfgtievp seielnnyaq nnyintlipw egkiyrrpay alsppdaqeg

121 sfsdgdntv geylkhfdle psrlgkqiri rfdgveramy vwlnghfigy aedsftpsef

181 dltpyiqdeg nvlavevfk staswiedqd mfrfsgifrs vnllaqplvh vedlhirpiv

241 tdnyqdgifn vdlqlhgekt gnvnrvidn dgntlvneth pvdstvkvqd qlfenvhlwd

301 nhdpvlyqll ieirddegnl velvpyrfgf rrieinkdhv vllngqrlil ngvnrhewna

361 krgraitmdd mtsdihtfke nninavrtch ypdqipwyyil cddngiyymma ennleshatw

421 qkmgaietsy nvpgsvpqwr dvvvdrartn yetfknhpsi lfwslgnesy agdnivkmne

481 fykkhdsrl vhyegvchtp eyrdrisdve swmylppkev eeylknnpdk pfmeceymhd

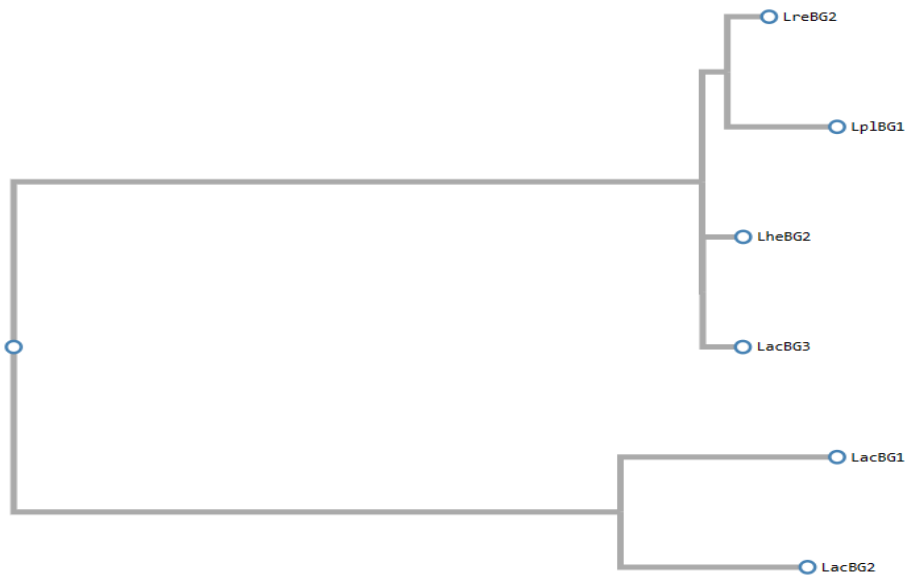
541 mgnsdggmgs yislldkypq yfggfwdfi dqallvkdpv sgqevmrygg dfddrhdye

601 fsgdglmfad rtpkpamqev ryyyglhk

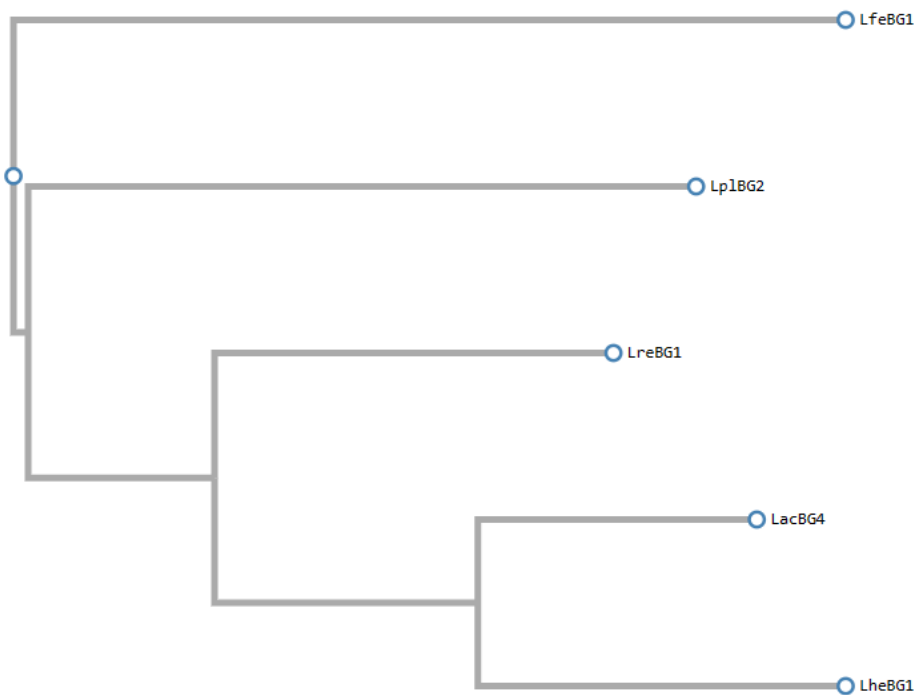
béta-galaktozidáz kis lánc: a sárga színnel jelölt konzervatív régió kódolja a béta-galaktozidáz enzim kis lánc nevű részét.

### 4.3 MSA és filogenetikai vizsgálat

A többszörös szekvencia illesztést kettő részre osztottam. Első lépésben összehasonlítottam a LacL génekben kódolt béta-galaktozidáz nagy alegységeket, majd ugyanezt a vizsgálatot elvégeztem a LacM génekben kódolt béta-galaktozidáz kis alegységekkel. A kapott filogenetikai fák a 6. és a 7. ábrákon láthatóak. A két fa erősen eltérő eredményt adott. A legnagyobb hasonlóság a két ábra között a *Lactobacillus helveticus* és a *Lactobacillus acidophilus* közelségében fedezhető fel. Különös, hogy a kapott fa alapján a LacBG1-ben és a LacBG2-ben kódolt enzim ugyan azonos ágon található, a LacBG3-ban kódolt fehérje egy eltérő (bár közeli) ágon fedezhető fel.



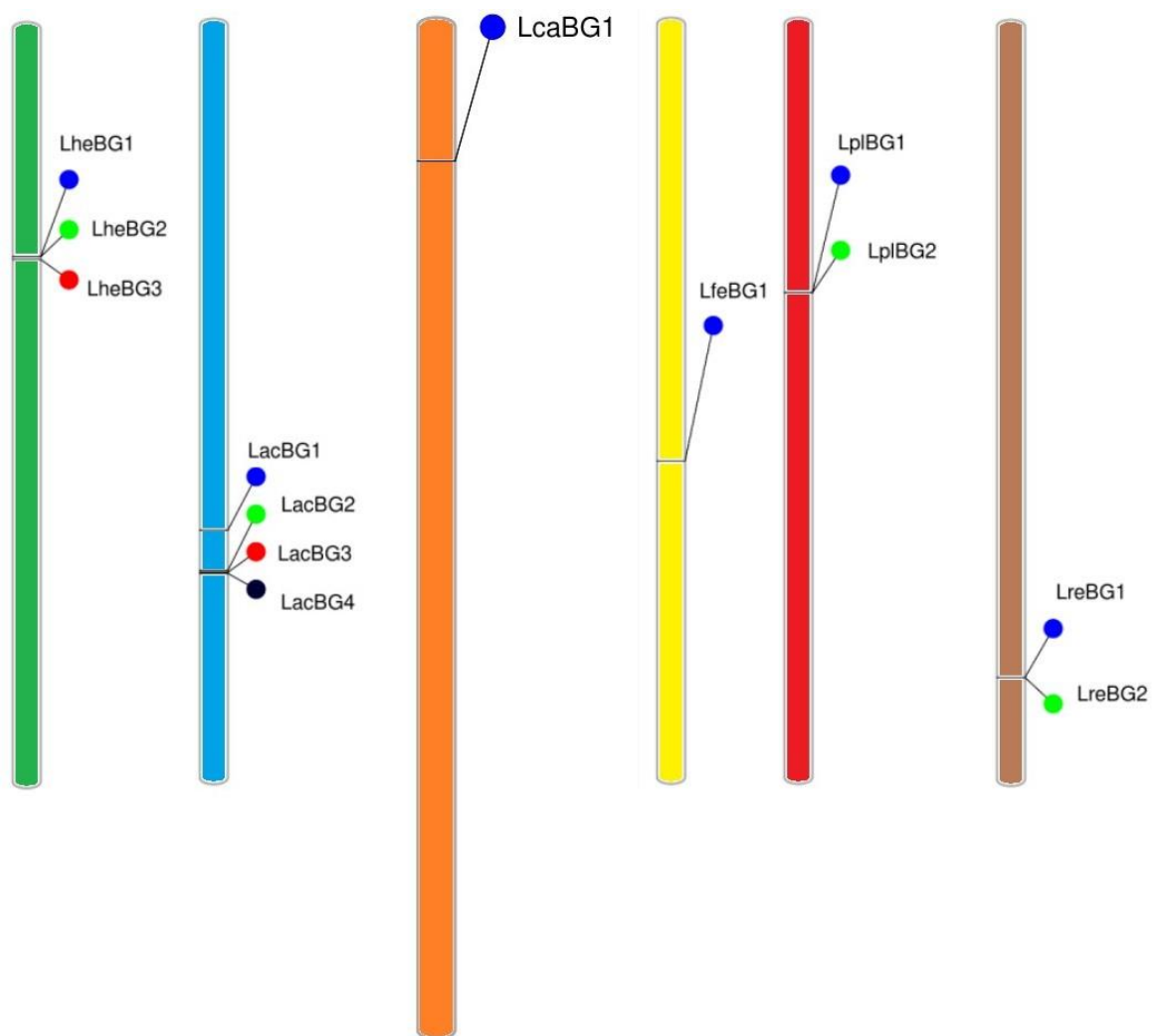
6. ábra: Filogenetikai ábra a LacL-ben kódolt fehérjék MSA eredménye alapján



7. ábra: Filogenetikai ábra a LacM-ben kódolt fehérjék MSA eredménye alapján

#### 4.4 Gének eloszlásának vizsgálata

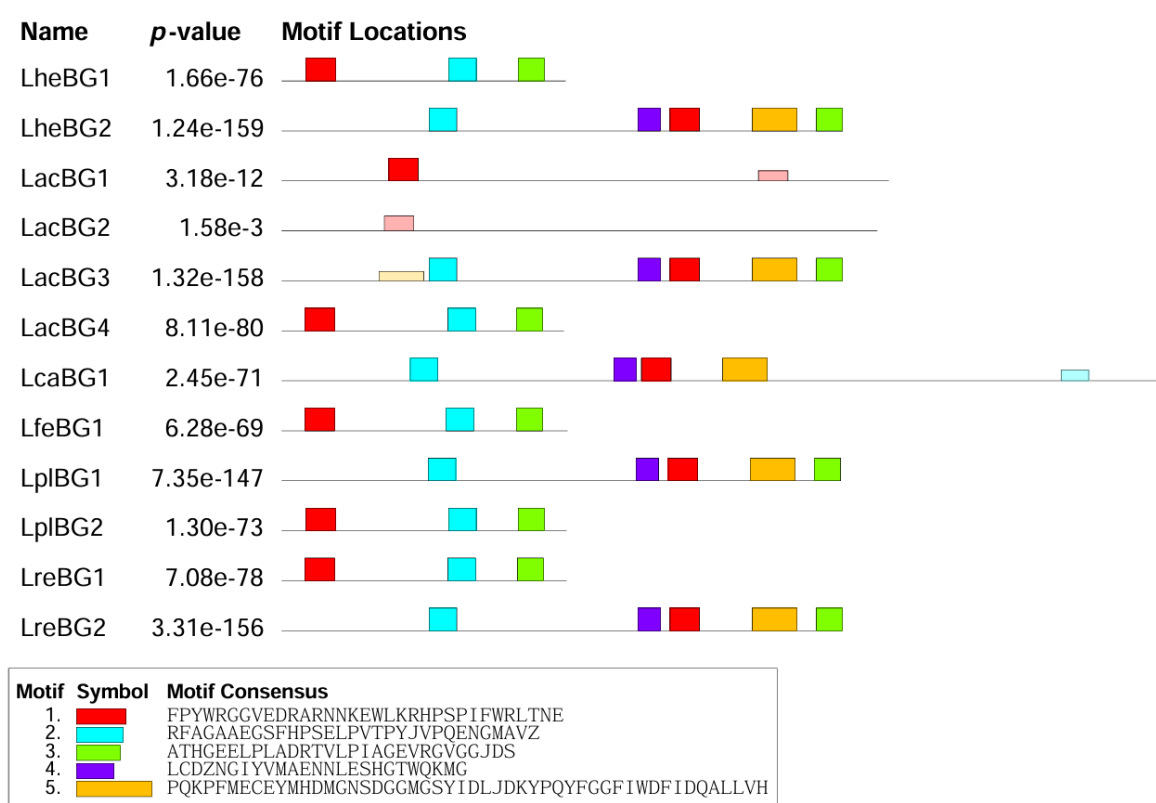
A 8. ábra illusztrálja a vizsgált fajok béta-galaktozidáz enzim génjeinek elhelyezkedését kromoszómákon. Látszólag a gének elhelyezkedése nem követ semmilyen rendszert. A vizsgált tejsavbaktériumok közül csak a *Lactobacillus acidophilus* kódolja több számban is az enzim egy bizonyos részét, a LacL gént, ugyanakkor az ismételt laktáz enzim kódolás nem egyedi ezen faj esetén. Baig és munkatársai (2023) béta-galaktozidázt termelő gombák *in silico* vizsgálata során azt találták, hogy az összes általuk vizsgált mikroorganizmus több példányban is kódolja az enzimet, de azok több kromoszómán is eloszanak.



**8. ábra: Béta-galaktozidáz enzimet kódoló gének eloszlása a kromoszómákon**  
(Balról jobbra: *Lactobacillus helveticus*, *L. acidophilus*, *Lactocaseibacillus paracasei*, *Limosilactobacillus fermentum*, *Lactiplantibacillus plantarum*, *Lactobacillus reuteri*)

## 4.5 Motif analízis és gén struktúra

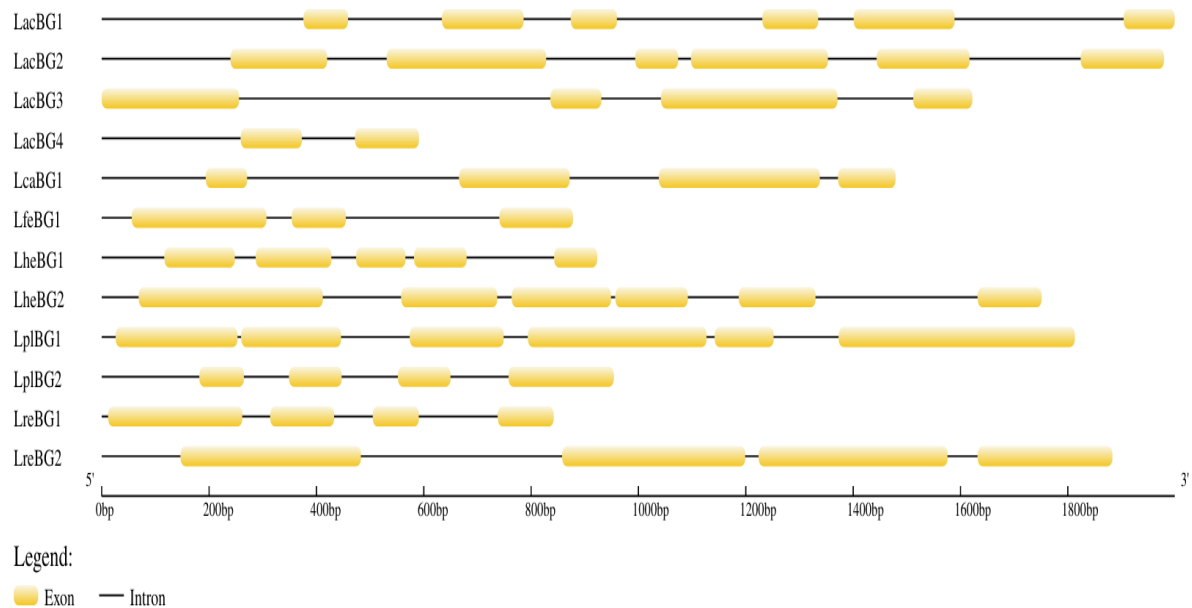
A Meme suite oldalon számos eszköz elérhető fehérje vagy gén szekvenciák vizsgálatára (<https://meme-suite.org/meme/>). Első lépésként az xstreme eszköz segítségével megállapítottam, hogy a vizsgált fehérjék öt lehetséges Motif-fal rendelkeznek, ezután a szerverre felvittem az összes vizsgált enzimet, majd MEME eszköz segítségével ezt az öt Motif-ot ábrázoltam, amely a 9. ábrán látható.



9. ábra:  $\beta$ -galaktozidáz enzimek azonos motívumai

A LacM génen kódolt fehérjék esetében három motívum figyelhető meg, gyakorlatilag pontosan ugyanazon a helyen, míg a LacL génen kódoltak esetén 5 motívum is látható, viszont ebben az esetben a LacBG1 és LacBG2 erősen eltér a többitől, illetve a LacBG3 és a LcaBG1 is mutat minimális eltérést a fő motívumoktól. LacBG1 és LacBG2 erős eltérése megmutatkozott a filogenetikai vizsgálat során is, ahol ezt a két enzim egy eltérő ágra került a többitől. A Meme suite oldal is eredetileg elrejtette ezt a két kieső értéket az ábráról.

A gének kódoló régióinak vizualizálásához először az NCBI open reading frame eszköz segítségével azonosítottam a gének lehetséges kódoló régiót, majd a kapott értékeket összevettem és GTF formátumban összeírtam. A GTF fájlnak a Gene Structure Display szerverre való feltöltését követően, megkaptam a génen lévő intronok elhelyezkedését, melyet a 10. ábrán mutatok be.



**10. ábra:  $\beta$ -galaktozidáz enzimet kódoló gének intron részeinek ábrázolása**

A kapott eredményen, láthatóak hasonló felépítések, de ennek ellenére is erős variabilitást mutatnak az intronok és az exonok váltakozása a vizsgált gének között. A LacL gének átlagosan 5 intront tartalmaznak, míg a LacM gének átlagosan 3,4 intront. Az *L casei* értékét figyelmen kívül hagytam az átlagok számítása során, mivel nem rendelkezik külön álló LacL és LacM génekkel. A LacL gének intron számainak módusza 6, míg a LacM géneknek 4. A gének hosszának átlagosan az egy harmada (Minden esetben 33,2 % vagy 33,3 %) kódolja a laktáz enzimet. Az egyetlen kivétel a *L casei*, amelynek a 66,94%-a kódoló régió.

#### **4.6 Vizsgált enzimek fizikai-kémiai tulajdonságaik és sejten belüli elhelyezkedésük**

A kapott eredmények (5.ábra) alapján is láthatjuk, hogy a különböző tejsavbaktériumokból származó laktáz enzim hasonló molekulatömegű. Minden LacM génen kódolt fehérje kb. 3500 Da, minden LacL génen kódolt fehérje valamivel 7000 Da feletti tömegű. Az egy kiugró érték a *L. casei*-ből származó enzim (LcaBG1), amelynek tömege 113640 Da). Az eltérésnek az lehet az oka, hogy itt a baktérium nem kódolja két külön részben az enzimet, hanem a kettő egyben fedezhető fel. Az izoelektromos pontjuk, pedig az összesnek a savas tartományba esik.

Egy fehérje stabilnak tekinthető, ha az instabilitási mutatója 40-es érték alatt van (Baig et al. 2023). Az esetemben a vizsgált fehérjék közül a LheBG2, LacBG3, LreBG1 és a LreBG2 instabilitási értékük: 41,66, 44,03, 41,2, 45,7. így ezek instabilnak tekinthetők, míg a többi fehérje stabilis. Az összes fehérje alifás mutatója 40%-nál magasabb, így valamennyi vizsgált enzim hőstabilis fehérje. Ezek az értékek alapján a leginstabilabb béta-galaktozidáz egység a *L. Reuterig* LreBG2 génje által kódolt enzim, mivel ennek a legmagasabb az instabilitási mutatója, és a többi enzimhez viszonyítva az LreBG2 rendelkezik a legalacsonyabb alifás mutatóval.

A GRAVY (Grandaverage of hydropathicity) mutató megmutatja számunkra, hogy egy fehérje mennyire poláris molekula. A 5. táblázat alapján LplBG2 a legkevésbé, míg a LreBG2 a leginkább poláris fehérje. Ez a két érték különbsége 0,478, ami erős diverzitást jelez a béta-galaktozidáz enzimek között. Általánosan elmondható, hogy valamennyi a vizsgált enzimek közül, poláris molekula.



**5. táblázat: Vizsgált enzimek fizikai-kémiai tulajdonságai**

<b>Fehérje</b>	<b>Molekulatömeg (Da)</b>	<b>pI</b>	<b>Negatív töltésű aminosavak</b>	<b>Pozitív töltésű aminosavak</b>	<b>Instabilitási mutató</b>	<b>Alifás mutató</b>	<b>Gravy mutató</b>
LheBG1	35830	5,57	40	30	32,43	76,38	-0,406
LheBG2	73479	5,21	95	63	41,66	72,01	-0,648
LacBG1	79347	6,11	91	80	37,09	72,99	-0,582
LacBG2	76583	5,84	85	73	33,88	75,68	-0,489
LacBG3	73253	4,98	97	56	44,03	72,47	-0,62
LacBG4	35817	5,53	38	30	38,89	76,27	-0,437
LcaBG1	113640	5,02	134	91	33,63	71,56	-0,561
LfeBG1	35776	5,93	38	33	30,74	75,25	-0,429
LplBG1	72166	4,75	93	51	38,64	74,31	-0,52
LplBG2	35237	4,73	36	15	39,42	74,92	-0,239
LreBG1	35397	4,93	42	24	41,2	75,83	-0,281
LreBG2	73602	4,68	108	58	45,7	69,95	-0,717

Az 6. táblázatban láthatjuk a fehérje aminosavszekvencia alapján jóslott lokalizáció helyét. A kapott értékek alapján megállapítható, hogy mindegyik az általam vizsgált enzimek közül intracelluláris azaz citoplazmában található. Intracelluláris termelése megnehezíti az enzim ipari kinyerését az alkalmazott mikroorganizmusokból.

6. táblázat: Vizsgált enzimek lokalizációja a sejtben

Fehérje	Citoplazma	Extracelluláris	Membrán	Sejtfal
LheBG1	4,245	0,478	0,243	0,035
LheBG2	4,583	0,203	0,202	0,012
LacBG1	3,871	0,727	0,376	0,026
LacBG2	2,658	1,952	0,33	0,061
LacBG3	4,333	0,557	0,101	0,009
LacBG4	3,945	0,534	0,488	0,033
LcaBG1	2,721	1,652	0,319	0,309
LfeBG1	2,974	1,5	0,549	0,158
LplBG1	4,802	0,151	0,038	0,009
LplBG2	3,219	0,955	0,607	0,219
LreBG1	4,577	0,226	0,166	0,031
LreBG2	4,771	0,181	0,043	0,006

## 5 Következtetések és javaslatok

### Következtetések:

- a *L. helveticus*, *L. acidophilus*, *L. plantarum* és a *L. reuteri* egy kisebb (LacM) és egy nagyobb (LacL) génen kódolja a béta-galaktozidáz enzimet, ezek a gének közösen fejeződnek ki és állnak össze. *Lacticaseibacillus paracasei* egy génen kódolja az enzimet, míg *Limosilactobacillus fermentum* kromoszómáján csak a LacM egység volt felfedezhető. Az *L. acidophilus* volt az egyetlen, ahol a vizsgált kromoszómán egy bizonyos gén többször is megtalálható.
- A filogenetikai vizsgálat során, mind a LacL mind a LacM alapján is készítette egy-egy fát. A LacL Alapján a LacBG1 és a LacBG2 egy külön ágba osztható míg a többi egy azonos ágon fut. A LacM esetében a fa nem különült el két ennyire különálló csoportba, a legközelebbi rokoni kapcsolat ez alapján a LacBG4 és a LheBG1 között lenne.
- A vizsgált mikroorganizmusok közül valamennyi intracellulárisan termeli a laktáz enzimet. Mindegyikük izoelektromos pontja a savas tartományba esik, valamennyi hőstabilis és poláris molekula. Mindegyik az általam vizsgált enzimek közül stabilis molekulának tekinthető

### Javaslatok:

- LacBG1 és LacBG2 további vizsgálatával megmagyarázhatóvá válhatnak a kiugró különbségek, más törzsekkel való összehasonlítás további információt mutathat, különösen további olyan példák esetén, ahol a LacL gén többszörösen is megtalálható a kromoszómán.

## 6 Összefoglalás

Kutató munkám során 6 mikroorganizmus (*L. helveticus*, *L. acidophilus*, *L. casei*, *L. fermentum*, *L. plantarum* *L. reuteri*) által termelt béta-galaktozidáz enzimet hasonlítottam össze in silico módszerekkel. NCBI adatbázist használva az összes fent említett baktériumnak kerestem egy-egy kromoszómát. Átnézve a kromoszómán kódolt géneket összeírtam minden olyan gént, ami laktáz enzimet kódol. Összesen 12 gént találtam a 6 mikroorganizmus között. Ebből négy (*L. helveticus*, *L. acidophilus* *L. plantarum* *L. reuteri*) egy kisebb (LacM) és egy nagyobb (LacL) génen kódolja laktáz enzimet, a *L. fermentum* kromoszómáján csupán egy LacM található, míg az *L. casei* egy génen tárolja a teljes fehérjét. Ezt követően mindegyikhez hozzáadtam egy-egy ideiglenes nevét a következő módszerrel. Mindegyik gén nevének az első karaktere nagy L betű a második és a harmadik, pedig az adott baktérium nevének első két betűje, a negyedik és az ötödik betű minden esetben BG, mint béta-galaktozidáz. Utána folytatólagosan sorszámoztam az azonos kromoszómán található géneket. Pl. az *L. helveticus* előbb kódolt laktáz génje, LheBG1

Ezeknek a géneknek először összeírtam a konzervatív régióit, leggyakrabban a béta galaktozidáz kis lánc konzervatív régió volt tapasztalható a vizsgált fehérjék esetében.

Ezt követően többszörös szekvencia illesztést alkalmazva két filogenetikai fát készítettem a génekről. Az egyik fa csak a LacL és az *L. casei* géneket vette alapul, míg a másik csak a LacM géneket.

Az NCBI Phenogram adatbázis segítségével összeírtam a gének relatív elhelyezkedését a vizsgált kromoszómákon, majd azokat vizualizáltam.

A következő lépésben a Meme suite oldalon található eszközöket használtam fel, elsőként feltöltöttem az általam vizsgált gének szekvenciáját, és az oldalon használható xstreme eszköz segítségével megállapítottam, hogy a vizsgált enzimeknek várhatóan öt motívuma lehetséges. Ezt az információt tovább vittem az oldalon található MEME eszközbe, amellyel a fő motívumok elhelyezkedését ábrázolni tudtam. A legtöbb gén hasonló eredményt adott, a kiugró kivételek a LacBG1 és a LacBG2 voltak, amik már a filogenetikai vizsgálatok során is elkülönülést mutattak. Következésképpen az NCBI adatbázis open reading frame eszközét vettem használatba. Ennek segítségével azonosítani tudtam a vizsgált gének kódoló és nem kódoló régióit. Miután betápláltam ezeket az adatokat a Gene structure display szerverre ezeket ábrázoltam. Az volt megfigyelhető, hogy minden az általam vizsgált mikroorganizmusok közül a kódolt gének harmada kódoló régió, az egyetlen kivétel az *L. casei* volt hiszen ezen baktérium esetében a laktáz enzim nem két külön részben kerül kódolásra a kromoszómán.

Ezt követően az enzimek fizikai-kémiai tulajdonságait vizsgáltam meg, Ehhez ProtParam-EXPASY internetes eszközt használtam. Az eszközt felhasználva össze tudtam írni a fehérjék molekulatömegét, izoelektromos pontját, negatív- és pozitív töltésű aminosavainak számát, instabilitási mutatóját, alifás mutatóját, valamint Gravy-mutatóját. Összehasonlítva az értékeket általánosan elmondható volt, hogy a LacM géneken kódolt fehérjék átlagosan 3500 Da tömegűek, míg a LacL 7000-8000 Da közötti tömeget érik el. Valamennyi a vizsgált fehérjék közül izoelektromos pontja alapján a savas tartományba esik. Az instabilitási mutató alapján a LheBG2, LacBG3, LreBG1 és a LreBG2-en kódolt fehérjék instabilisnak, míg a többi fehérje stabilnak tekinthető. A vizsgált enzimek közül, az összes hőstabilis és poláris molekula. Végző lépésként a CELLO internete eszközbe, feltöltöttem a vizsgált enzimek szekvenciáját és az elemzést elindítva megkaptam a legvalószínűbb elhelyezkedésüket a sejten belül. Ez alapján az összes enzim intracellulárisnak mondható.

## 7 Irodalomjegyzék

**Borsodi, A., Felföldi, T., Jáger K., Makk J., Márialigeti K., Romsics Cs., Tóth E., Bánfi R., Pohner Zs., Vajna B.,** (2013) Bevezetés a prokarióták világába ELTE Eötvös Kiadó

**Danish Ilyas Baig, Zeeshan Zafar, Haris Ahmed Khan, Amna Younus, Muhammad Faraz Bhatti** (2023) Genome-wide identification and comparative in-silico characterization of  $\beta$ -galactosidase (GH-35) in ascomycetes and its role in germ tube development of *Aspergillus fumigatus* via RNA-seq analysis doi: 10.1371/ e0286428

**Deák, V.,** (2016) Általános genetika Typotex Kiadó, Budapest, 177.

**D. H., Juers, B. W., Matthews, R. E., Huber** (2012) LacZ  $\beta$ -galactosidase: Structure and function of an enzyme of historical and molecular biological importance. *Protein Sci.* 21(12): 1792-1807 DOI: 10.1002/pro.2165

**François P Douillard, Willem M de Vos** (2014) Functional genomics of lactic acid bacteria: from food to health doi: 10.1186/1475-2859-13-S1-S8

**Izet M. Kapetanovic, Simon Rosenfeld, Grant Izmirlan** (2004) Overview of Commonly Used Bioinformatics Methods and Their Applications doi: 10.1196/annals.1310.003

**Jeff Gauthier, Antony T. Vincent, Steve J. Charette and Nicolas Derom,** (2019) A brief history of bioinformatics doi: 10.1093/bib/bby063

**J.M.S Rana** (2012) BIOINFORMATICS Tools & Applications

**Kumar, Ravinder & M, Kaur & Garsa, Anita & Shrivastava, Bhuvnesh & PV, Reddy & Tyagi, Ashish.** (2015). Natural and Cultured Buttermilk CRC Press

**Ludányi A.,** (2012) A fehérjekutatás modern módszertana, Medicina Könyvkiadó Zrt.

**Maria Chatzou, Cedrik Magis, Jia-Ming Chang, Carsten Kemena, Giovanni Bussotti, Ionas Erb and Cedric Notredame** (2015) Multiple Sequence Alignment Modeling:

Methods and Applications doi: 10.1093/bib/bbv099

**McGee, H.** (1984). On Food and Cooking, Unwin Hyman Ltd, London, p. 36.

**Parmjit S. Panesar, Shweta Kumari, Reeba Panesar** (2010) Potential Applications of Immobilized  $\beta$ -galactosidase in Food Processing Industries. *Enzyme Research* ID: 473137

**Remilyn M. Mendoza, Sang Hoon Kim, Robie Vasquez, In-Chan Hwang, Young-Seo Park, Hyun-Dong Paik, Gi-Seong Moon, Dae-Kyung Kang** (2022) Bioinformatics and its role in the study of the evolution and probiotic potential of lactic acid bacteria doi: 10.1007/s10068-022-01142-8

**R.J. Rouwenhorst, J. T. Pronk and J. P. van Dijken** (1989) The discovery of  $\beta$ -galactosidase TIBS 14- October 1989

**Rodriguez A' P, Leiro RF, Trillo MC, Cerdan ME, Siso MIG, Becerra M** (2006) Secretion and properties of a hybrid *Kluyveromyces lactis*–*Aspergillus niger*  $\beta$ -galactosidase. Microb Cell Fact 5:1–8

**Stephen F. Altschul, Warren Gish, Webb Miller Eugene W. Myers and David J. Lipman** (1990) Basic Local Alignment Search Tool J. Mol. Biol. (1990) 215, 403-410

**Shaima Saqib, Attiya Akram, Sobia Ahsan Halim<sup>1</sup>, Raazia Tassaduq** (2017) Sources of  $\beta$ -galactosidase and its applications in food industry 3 Biotech (2017) 7:79

DOI 10.1007/s13205-017-0645-5

**Storhaug CL, Fosse SK, Fadnes LT.** (2017) Country, regional, and global estimates for lactose malabsorption in adults: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol. 2017 Oct;2(10):738-746. doi: 10.1016/S2468-1253(17)30154-1. Epub 2017 Jul 7. PMID: 28690131.

## Ábrák jegyzéke

1. ábra: Homofermentatív tejsavas erjedés folyamata	4. oldal
2. ábra: Heterofermentatív tejsavas erjedés folyamata	5. oldal
3. ábra: A $\beta$ -galaktozidáz tetramer szerkezete	6. oldal
4. ábra: Lac operon ábrázolása és működésének illusztrálása	9. oldal
5. ábra: Általános folyamatábra a genomok összehasonlítására bioinformatikai módszerekkel	13. oldal
6. ábra: Filogenetikai ábra a LacL-ben kódolt fehérjék MSA eredménye alapján	28. oldal
7. ábra: Filogenetikai ábra a LacM-ben kódolt fehérjék MSA eredménye alapján	29. oldal
8. ábra: Béta-galaktozidáz enzimet kódoló gének eloszlása a kromoszómákon	30. oldal
9. ábra: $\beta$ -galaktozidáz enzimek azonos motívumai	31. oldal
10. ábra: $\beta$ -galaktozidáz enzimet kódoló gének intron részeinek ábrázolása	32. oldal

## Táblázatok jegyzéke

1. táblázat: A $\beta$ -galaktozidáz mikroorganizmus forrásai	10. oldal
2. táblázat: Fehérjék szekvenciaazonosságának mértékének jellemzése	14. oldal
3. táblázat: Vizsgált baktériumok és azok kromoszóma azonosítói	17. oldal
4. táblázat: A vizsgált kromoszómákon azonosított béta-galaktozidáz gének	20. oldal
5. táblázat: Vizsgált enzimek fizikai-kémiai tulajdonságai	34. oldal
6. táblázat: béta-galaktozidáz enzimek helye a sejtben	35. oldal



## Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani konzulenseimnek, Dr. Bujna Erikának, és Dr. Nguyen Duc Quang-nak, a szakmai útmutatásukért, és a számos jó tanácsukért, amelyeket a dolgozatom készítése során kaptam.

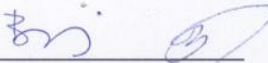
## NYILATKOZAT

Burján Alex (KE60JW) konzulenseként nyilatkozom arról, hogy a szakdolgozatot áttekintettem, a hallgatót az irodalmi források korrekt kezelésének követelményeiről, jogi és etikai szabályairól tájékoztattam.

A szakdolgozatot a záróvizsgán történő védeésre javaslom / nem javaslom.

A dolgozat állam- vagy szolgálati titkot tartalmaz: igen nem

Kelt: Budapest, 2024. április 19.



belső konzulensek

## NYILATKOZAT

### a szakdolgozat nyilvános hozzáféréséről és eredetiségéről

A hallgató neve: **Burján Alex**  
A Hallgató Neptun kódja: **KE60JW**  
A dolgozat címe: **Tejsavbaktérium eredetű béta-galaktozidáz enzim tanulmányozása bioinformatikai eszközökkel**  
A megjelenés éve: **2024**  
A konzulens intézetének neve: **Élelmiszertudományi és Technológiai Intézet**  
A konzulens tanszékének a neve: **Sőr- és Szeszipari Tanszék**

Kijelentem, hogy az általam benyújtott szakdolgozat egyéni, eredeti jellegű, saját szellemi alkotásom. Azon részeket, melyeket más szerzők munkájából vettem át, egyértelműen megjelöltem, és az irodalomjegyzékben szerepeltettem.

Ha a fenti nyilatkozattal valótlan állítottam, tudomásul veszem, hogy a záróvizsga-bizottság a záróvizsgából kizár és a záróvizsgát csak új dolgozat készítése után tehetek.

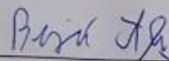
A leadott dolgozat, mely PDF dokumentum, szerkesztését nem, megtekintését és nyomtatását engedélyezem.

Tudomásul veszem, hogy az általam készített dolgozatra, mint szellemi alkotás felhasználására, hasznosítására a Magyar Agrár- és Élettudományi Egyetem mindenkori szellemi tulajdon-kezelési szabályzatában megfogalmazottak érvényesek.

Tudomásul veszem, hogy dolgozatom elektronikus változata feltöltésre kerül a Magyar Agrár- és Élettudományi Egyetem könyvtári repozitori rendszerébe. Tudomásul veszem, hogy a megvédett és

- nem titkosított dolgozat a védést követően
- titkosításra engedélyezett dolgozat a benyújtásától számított 5 év eltelté után nyilvánosan elérhető és kereshető lesz az Egyetem könyvtári repozitori rendszerében.

Kelt: **2024** év **04** hó **22** nap



Hallgató aláírása