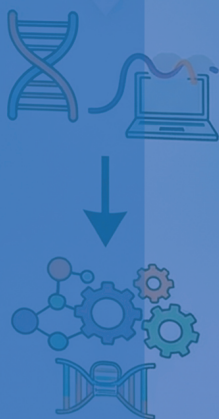
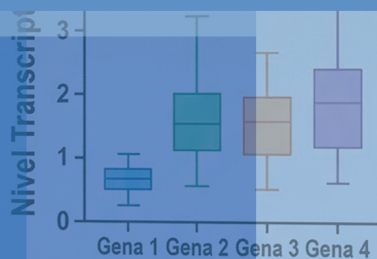


Model de analiză transcriptomică la specia *Lolium perenne* L și varianta ameliorată 'Mara'



Mihali Ciprian Valentin

Tod Monica Alexandrina

Model de analiză transcriptomică
la specia *Lolium perenne* L
și varianta ameliorată 'Mara'

Editura Trinom
Arad, 2026

Descrierea CIP a Bibliotecii Naționale a României
MIHALI, CIPRIAN VALENTIN

Model de analiză transcriptomică la specia *Lolium perenne* L
și varianta ameliorată 'Mara' / Mihali Ciprian Valentin, Tod
Monica Alexandrina. - Arad : Trinom, 2026

Conține bibliografie

ISBN 978-606-8557-89-2

I. Tod, Monica Alexandrina

63

PREFAȚĂ

Evoluția contemporană a științelor agricole este marcată de o tranziție fundamentală de la ameliorarea convențională bazată pe selecția fenotipică către strategiile avansate de ameliorare de precizie, fundamentate pe analiza moleculară de înaltă rezoluție. Lucrarea de față, intitulată „Model de analiză transcriptomică la specia *Lolium perenne* L și varianta ameliorată 'Mara' ”, se constituie într-o contribuție științifică aplicată oferind o perspectivă integrată asupra dinamicii genomice a uneia dintre cele mai relevante specii de graminee furajere.

Prezentul studiu este rezultatul unei sinergii instituționale între Stațiunea de Cercetare-Dezvoltare pentru Creșterea Bovinelor Arad și Institutul de Cercetare-Dezvoltare pentru Pajiști Brașov, reflectând necesitatea coroborării expertizei din domeniul zootehnic cu excelența în ameliorarea plantelor.

Raigrasul peren (*Lolium perenne* L.) reprezintă un taxon de o complexitate biologică remarcabilă, a cărui valoare agronomică este intrinsec legată de plasticitatea sa ecologică și profilul nutritiv superior. În contextul actual al schimbărilor climatice globale, optimizarea bazei genetice a acestei specii devine un imperativ. Analiza debutează cu o prezentare completă a caracteristicilor botanice și a cerințelor ecofiziologice, culminând cu prezentarea soiului Mara – o realizare autohtonă de prestigiu a ICDP Brașov, care demonstrează potențialul de adaptabilitate și productivitate al germoplasmei românești.

Din punct de vedere metodologic, lucrarea propune un flux de lucru (pipeline) bioinformatic riguros, centrat pe tehnologiile de secvențiere de nouă generație (Next-Generation Sequencing - NGS). Autorii detaliază etapele critice de control al calității bibliotecilor de date, alinierea la genomul de referință și cuantificarea precisă a abundenței transcriptelor, oferind astfel un model replicabil și pentru studiul altor specii de interes agricol.

Valoarea adăugată a acestei cercetări rezidă în profunzimea analizelor funcționale și structurale prezentate în secțiunile de rezultate: analiza expresiei genice diferențiale-identificarea arhitecturii moleculare care guvernează răspunsul varietății ameliorate la factorii de mediu; adnotarea funcțională (GO și KEGG)-descifrarea rețelelor metabolice și a proceselor biologice esențiale care conferă superioritate biologică soiului Mara; complexitatea structurală a transcriptomului-investigarea mecanismelor de matisare/splicing alternativ și identificarea siturilor polimorfice de tip SNP (Single Nucleotide Polymorphism) și InDel, elemente fundamentale pentru dezvoltarea viitorilor markeri moleculari în selecția asistată genomic (Genomic Selection).

Prin abordarea sa multidimensională, această studiu depășește cadrul unei simple analize de specialitate, devenind un instrument teoretic și practic

indispensabil cercetătorilor, amelioratorilor și cadrelor didactice din domeniul biotehnologiilor vegetale. Rezultatele prezentate fundamentează înțelegerea mecanismelor moleculare ale rezilienței și productivității, deschizând noi orizonturi pentru securitatea resurselor furajere într-o agricultură sustenabilă.

Arad, 18.03.2026

Autorii

CUPRINS

I	INTRODUCERE	
1.1	<i>Lolium perenne</i> L. – raigras peren – generalități	9
1.2	Caracteristici botanice și morfologice	10
1.3	Cerințe ecologice	11
1.4.	Tehnologia de cultură	12
1.5	Exploatarea și valoarea furajeră	13
1.6	Avantaje și dezavantajele cultivării speciei	13
1.7	Aspecte economice și ecologice	14
1.8	Ameliorarea raigrasului peren la ICDP Brașov	14
1.9	Soiuri de <i>Lolium perenne</i> L. realizate la ICDP – Brașov	15
1.10	Obiectivele ameliorării raigrasului peren în condițiile schimbărilor climatice	18
1.11	Metode de ameliorare	20
II	MATERIALE ȘI METODE	
2.1	Recoltarea plantelor	23
2.2	Cadru general. Controlul calității probelor. Construcția bibliotecilor de date, secvențierea	25
2.2.1	Controlul calității și pregătirea bibliotecilor	27
2.3	Clusterizare și secvențiere realizate după controlul calității bibliotecilor	27
2.4	Controlul calității datelor. Analiza bioinformatică	28
III	REZULTATE ȘI DISCUȚII	
3.1	Controlul calitativ al datelor	34
3.1.1	Datele brute	34
3.1.2	Examinarea ratei de eroare în secvențiere	35
3.1.3	Distribuția GC în conținutul secvențiat	36
3.1.4	Filtrarea datelor de secvențiere	37
3.1.5	Date statistice cu referire la calitatea secvențierii	39
3.2	Alinierea	43
3.2.1	Rezultate aliniere/„mapare”	44
3.2.2	Distribuția citirilor la genomul de referință	44

3.2.3	Rezultate cu privire la vizualizarea alinierilor ...	47
3.3	Analiza genelor nou descoperite	49
3.3.1	Identificarea/ Predicția de gene noi	49
3.3.2	Adnotarea genelor noi	49
3.4	Analiza nivelului de expresie genică	53
3.4.1	Cuantificarea expresiei genice	53
3.4.2	Distribuția nivelului de expresie genică	54
3.4.3	Analiza de corelație	55
3.4.4	Analiza componentelor principale (ACP)	58
3.4.5	Co-expresia în diagramele Venn	61
3.5	Analiza expresiei genice diferențiale	67
3.5.1	Rezultate analiză expresie diferențială	68
3.5.2	Date expresie genetică diferențială interpretate statistic	69
3.5.3	Analiza de clusterizare	71
3.6	Analiza funcțională	75
3.6.1	Analiza de îmbogățire a ontologiei genelor/GO/Gene ontology	76
3.6.2	Analiza de îmbogățire a căilor metabolice KEGG (Kyoto Encyclopedia of Genes and Genomes)	81
3.6.3	Analiza rețelelor de interacțiune proteină- proteină (PPI)	85
3.7	Analiza structurală	88
3.7.1	Analiza matisării/splicing-ului alternativ	88
3.7.1.1	Detectarea izoformelor exprimate diferențial	90
3.7.1.2	Vizualizarea evenimentelor de splicing alternativ (AS)	91
3.7.2	Analiza SNPs și InDel	93
3.7.2.1	Identificarea siturilor polimorfice	93
3.7.2.2	Statistica siturilor polimorfice	94
Anexa 1		97
Anexa 2		123
Bibliografie		127

I. INTRODUCERE

1.1	<i>Lolium perenne</i> L. – raigras peren – generalități
1.2	Caracteristici botanice și morfologice
1.3	Cerințe ecologice
1.4	Tehnologia de cultură
1.5	Exploatarea și valoarea furajeră
1.6	Avantaje și dezavantajele cultivării speciei
1.7	Aspecte economice și ecologice
1.8	Ameliorarea raigrasului peren la ICDP Brașov
1.9	Soiuri de <i>Lolium perenne</i> L. realizate la ICDP – Brașov
1.10	Obiectivele ameliorării raigrasului peren în condițiile schimbărilor climatice
1.11	Metode de ameliorare

1. *Lolium perenne* L. – raigras peren – generalități

Raigrasul peren este prima graminee furajeră cultivată, istoria ei fiind legată de introducerea pajiștilor temporare în asolament în Anglia în secolul XVII. Centrul de origine al raigrasului peren este zona temperată din Europa, Asia și nordul Africii, de unde s-a extins în nordul și sudul Americii în secolul XVIII și în secolul XIX în Australia și Noua Zeelandă.

În România, se cultivă în special în zonele colinare și de câmpie cu regim pluviometric favorabil, dar se întâlnește și în zonele montane, inclusiv în amestecuri de pajiști. Importanța sa rezidă în capacitatea de producere a unei biomase de calitate înaltă, bogată în proteine, palatabilitate, toleranța la pășunat și valoarea ecologică. Este principala specie de pășune din Transilvania, nordul Moldovei și Câmpia de Vest. Graminee de talie mică, se găsește spontan sau cultivată în pajiști din luncile râurilor, pe soluri fertile, cu aport freatic. În regiunile montane, urcă până la 1300–1400 m altitudine. Temperatura optimă de dezvoltare este de 18-20° C, **Figura 1**. Preferă zonele cu ierni blânde și zăpadă puțină și este sensibilă la ger uscat și veri secetoase. Este sensibilă la rugini (*Puccinia graminis Pers*) și la mucegaiul de zăpadă (*Fusarium nivale*). Principalele obiective de ameliorare, alături de productivitate și calitate sunt rezistența la boli și secetă care afectează perenitatea, îndeosebi în recente condiții ale schimbărilor climatice.

Din punct de vedere al valorii furajere este specia de graminee perene de pajiști cu valoarea nutritivă cea mai ridicată, având un conținut de proteină brută cuprins între 14-17% și de 24-28 % celuloză brută și un conținut ridicat de glucide solubile.

Este o specie tipică pentru pășunat, deoarece rezistă la călcat și are o bună regenerare după ce a fost exploatată fiind un bun partener al trifoiului alb dar poate fi folosită și în amestecurile pentru fâneață. Se recomandă fertilizarea

pe bază de azot. În amestecuri are o competitivitate mare, mai ales în anul al doilea de vegetație.

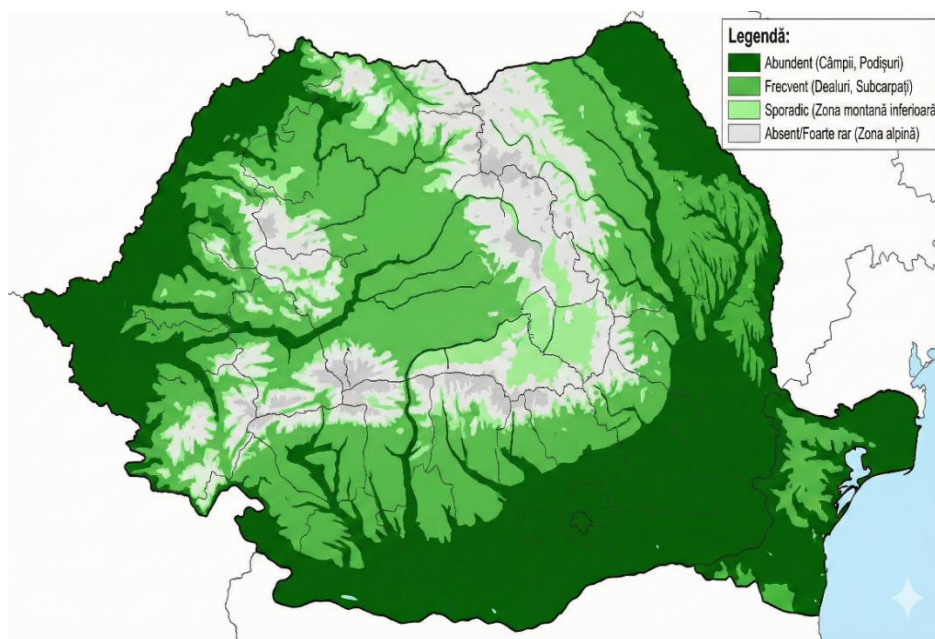


Figura 1. Distribuția geografică a speciei *Lolium perenne* L. (raigras peren) în România.

1.2 Caracteristici botanice și morfologice

Specie perenă, aparține familiei Poaceae și genului *Lolium*. Este o specie diploidă ($2n=2x=14$) dar în scopul creșterii caracterelor de interes agronomic au fost create și soiuri tetraploide ($2n=2x=28$) prin duplicarea artificială a cromozomilor, **Figura 2**.

- Tulpina: erectă, cilindrică, cu 3–6 noduri, înălțime 30–80 cm.
- Frunza: sunt plane, lung acuminate, lucioase pe fața superioară de culoare verde deschis, iar pe partea inferioară mate de culoare verde mai închis, cu lungime de 10–20 cm și 3–6 mm lățime.
- Ligula: scurtă (1–2 mm), membranoasă; auricule bine dezvoltate.
- Inflorescența: spic simplu, ax comprimat, 8–10 spiculețe cu 6–12 flori fiecare. Spiculețele sunt așezate alternativ cu muchia pe rahis, pe două rânduri. Înfloritul începe în treimea superioară a spicului și durează 4–5 zile.
- Sămânța: mică (2–3 mm), alungită, MMB 2–3 g. Durata de viață a culturii: 3–5 ani, **Figura 3**.

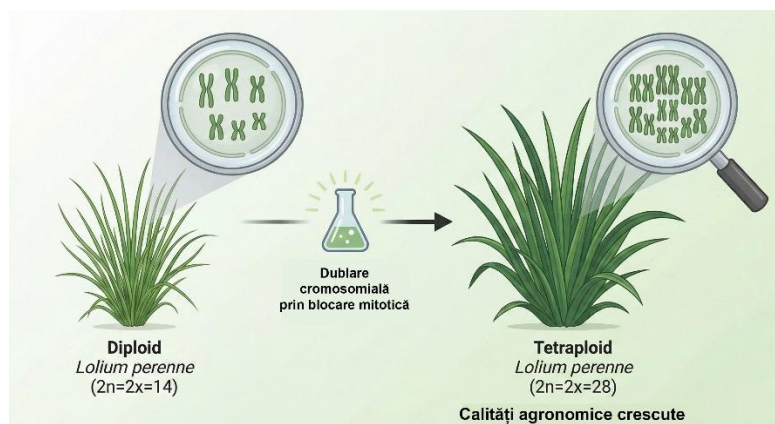


Figura 2. Strategia de ameliorare a speciei *Lolium perenne* L. .

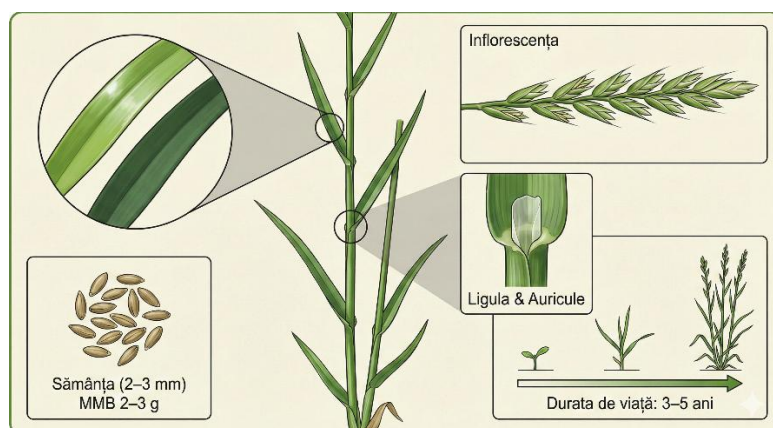


Figura 3. Aspecte morfologice, specia *Lolium perenne* L. .

1.3 Cerințe ecologice

- Temperatura: germinare la 6–8°C, optim 15–20°C; rezistă până la -15°C.
- Umiditatea: necesită 600–800 mm precipitații/an; fără irigare, producția scade cu 40–60%.
- Solul: optim pe soluri fertile, textură mijlocie, pH 5,5–7,5; pe soluri sărace sau acide necesită amendamente și fertilizare.

Specie cu o mare variabilitate privind adaptabilitatea la condițiile de mediu, **Figura 4.**

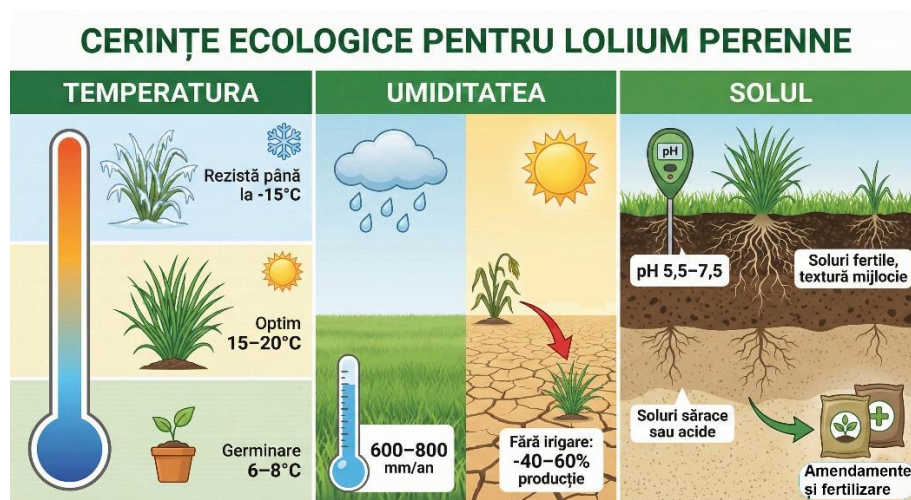


Figura 4. Condițiile pedoclimatice și regimul hidric pentru specia *Lolium perenne* L. .

1.4. Tehnologia de cultură

1.4.1 Pregătirea terenului

Arătura la 20–25 cm, grăpare și tăvălugire pentru pat germinativ uniform. Se recomandă rotația culturilor cu leguminoase pentru boabe sau prășitoare fertilizate organic.

1.4.2 Însămânțarea

- Epoci optime: primăvara (martie–aprilie) caz în care plantele se dezvoltă exclusiv vegetativ, cultura este valorificată ca furaj; sfârșitul verii (august–septembrie), în acest caz plantele întâlnesc condiții optime de vernalizare și cresc echilibrat, cu un sistem radicular bine dezvoltat
- Norma de sămânță: 25–30 kg/ha în cultură pură; 8–15 kg/ha în amestecuri.

-Adâncimea de semănat: 1–2 cm.

- Densitatea optimă: 2000–2500 plante/m² la răsărire.

1.4.3 Fertilizarea

- Azot (N): 60–100 kg/ha/an în zone secetoase, 120–180 kg/ha/an în condiții de irigare, fracționat după fiecare cosire/pășunat.
- Fosfor (P₂O₅): 40–60 kg/ha/an.
- Potasiu (K₂O): 60–80 kg/ha/an, mai ales pe soluri ușoare.

1.4.4 Lucrări de întreținere

- Irigare în perioade secetoase (25–30 mm la 2–3 săptămâni).
- Combaterea buruienilor prin erbicide selective și cosiri regulate.
- Reînsămânțarea la 3–4 ani pentru menținerea densității, **Figura 5.**

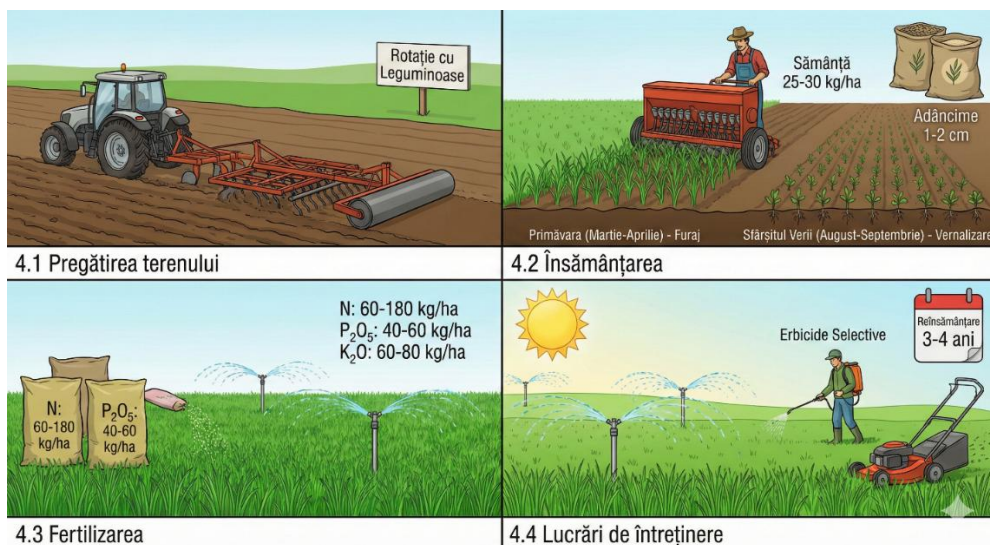


Figura 5. Tehnologia de cultură, specia *Lolium perenne* L. .

1.5 Exploatarea și valoarea furajeră

Pășunat: suportă pășunatul raționalizat cu perioade de repaus de 20–30 zile. Cosire: recomandată în faza de burduf până la începutul înspicării pentru fân și siloz de calitate.

Valoare nutritivă:

- Proteină brută: 14–18% frunză tânără, 9–12% la maturitate.
- Fibră detergent neutru (NDF): 40–50%.
- Digestibilitate: 65–75%.
- Realizează un siloz de foarte bună calitate, producția: 8–12 t s.u./ha/an

fără irigare, 14–18 t s.u./ha/an cu irigare și fertilizare optimă.

Utilizat în proporție mare în amestecuri pentru gazon.

1.6 Avantaje și dezavantajele cultivării speciei

Avantaje: instalare foarte ușoară, ritm de creștere rapid după răsărire, regenerare rapidă după exploatare, rezistență la călcat, palatabilitate bună, productivitate ridicată, valoare furajeră foarte bună, compatibilitate în amestecuri atât cu alte graminee perene de pajiști, cât și cu leguminoase furajere, cu precădere pentru constituirea amestecurilor intensive cu trifoi alb sau ghizdei. Amestecurile cu trifoi roșu sunt destinate furajării la iese sau pentru producerea fânului și semifân.

Soiurile precoce sunt general utilizate în amestecuri destinate recoltării mecanice, pentru fân sau siloz, iar cele tardive pentru exploatare prin pășunat.

Dezavantaje: sensibilitate la secetă și temperaturi extreme, la boli, durată relativ scurtă de viață, necesar ridicat de fertilizare azotată.

1.7 Aspecte economice și ecologice

Costul de înființare este moderat, dar inputurile pentru fertilizare și irigare sunt esențiale. Contribuie la reducerea eroziunii solului, îmbunătățirea materiei organice și fixarea azotului în amestec cu leguminoase.

1.8 Ameliorarea raigrasului peren la ICDP Brașov

În țara noastră primele lucrări de ameliorare la plantele furajere perene de pajiști s-au dezvoltat după anul 1950, primele soiuri “locale” fiind obținute prin selecția în masă din populații și ecotipuri recoltate din flora spontană a pajiștilor naturale, programul de selecție în masă fiind organizat la stațiunile experimentale de la Mărculești, Moara Domneasca, Măgurele, Suceava, obținând-se la obsigă și pir crestă primele soiuri locale, (Breazu, I., Grusea A., 1997).

La prima consfătuire a amelioratorilor din domeniul plantelor furajere, desfășurată la Stațiunea de la Măgurele – Brașov, în 1977, cu o participare largă a cercetătorilor de la Institutele Agronomice și Stațiunile Agricole care aveau în planul tematic ameliorarea acestor specii s-au discutat metodele de ameliorare și testare a soiurilor și s-a stabilit planul tematic al activității de ameliorare, făcând-se referiri și la activitatea de colectare, evaluare, utilizare și menținere a ecotipurilor românești, (Breazu, I., Grusea A., 1997).

În cei peste 45 de ani de activitate în domeniul ameliorării gramineelor și leguminoaselor perene de pajiști de la ICDP Brașov, s-a acumulat o importantă bază de date științifice, un bogat și valoros material biologic, care este menținut prin rezerva de semințe, dar și vegetativ, în colecția activă de clone, asigurând premisa realizării unor soiuri cu însușiri valoroase într-un interval redus de timp.

Existența în flora spontană din România a unor resurse genetice valoroase sub aspectul variabilității și adaptabilității, cercetate și selectate în condițiile noi ale schimbărilor climatice asigură un câștig genetic productiv, reproductiv și adaptativ al soiurilor viitoare, (Tod et al., 2015).

În lucrările de ameliorare se urmărește crearea de noi soiuri cu un potențial de producție ridicat, cu o calitate superioară a furajului, cu o bună adaptabilitate la condițiile de mediu, cu rezistență la boli și dăunători, competitive, cu grade diferite de precocitate, cu potențial de fructificare ridicat.

Prin utilizarea unei germoplasme cu variabilitate ridicată care a inclus și resurse autohtone cu un conținut de gene valoroase, prin selectarea unor genotipuri cu însușiri morfo-fiziologice care corespund obiectivelor de ameliorare, cu o bună capacitate combinativă generală, au fost create mai multe soiuri de raigras peren, dintre care în prezent sunt incluse în Lista Oficială de soiuri doar 2: Mara și Brevis.

Pe baza însușirilor de alogamie, anemofilie, autoincompatibilitate, heterozigoție, capacitate de înmulțire și pe cale vegetative, au fost create următoarele soiuri sintetice:

1.9 Soiuri de *Lolium perenne* L. realizate la ICDP – Braşov

Marta omologat în anul 1981, este un soi tardiv diploid,, competitiv în amestecuri simple sau complexe cu graminee şi leguminoase perene de pajişti, recomandat pe terenuri fertile, în zona colinară umedă. Realizează producţii ridicate de furaj de calitate bună. Este retras din Catalogul Oficial, **Figura 6**.



Figura 6. *Lolium perenne* 'Marta'.

Mara omologat în anul 1989 este un soi tardiv diploid, cu o bună rezistenţă la iernare, secetă şi boli, o bună capacitate de regenerare după folosinţă şi o excelentă calitate a furajului. Are în componenţă atât germoplasmă autohtonă cât şi străină, este un soi tardiv, pretabil la utilizare, în special pentru păşunat, cu o bună rezistenţă la secetă şi iernare, cu o bună perenitate şi adaptabilitate, cu productivitate ridicată (40-50 t/ha masa verde, 90-100 t/ha substanţă uscată şi 650 kg sămânţă/ha). Furajul are o bună calitate, cu un conţinut de 13,7% proteină brută şi 27% celuloză brută.

Este recomandat pentru pășune, în amestecuri simple sau complexe cu *Festuca rubra* L., *Festuca pratensis* / *Lolium pratense* (Huds.), *Poa pratensis* L. și *Trifolium repens* L. . De asemenea poate fi utilizat pentru terenuri sportive și parcuri, alte spații verzi. Se pretează la terenurile fertile și cu umiditate suficientă. Datorită adaptabilității și calităților lui deosebite acest soi este în continoare în Lista oficială a soiurilor în România și Europa. Se multiplică și în prezent în Europa, SUA și Canada, de către o firmă privată în baza unui contract de redevență, **Figura 7.**



Figura 7. *Lolium perenne* 'Mara'.

Măgura omologat în anul 1999, este un soi semitimpuriu diploid, mai precoce decât soiul Mara, pornirea în vegetație este mai rapidă decât la soiurile existente, prezintă o bună rezistență la bolile foliare, rezistență bună la iernat și secetă. În amestecuri cu ghizdei și trifoi are o capacitate sporită de competitivitate și o palatabilitate ridicată. Este retras din Catalogul Oficial. Poate fi utilizat în amestecuri simple și complexe de pajiști pentru fânează, **Figura 8.**



Figura 8. *Lolium perenne* 'Măgura'.

Brevis – la ICDP -Brașov omologat în 2025, soi sintetic semitimpuriu, diploid, obținut din selecția recurentă a soiului Măgura, are rezistență ridicată la ger, și boli (la mucegaiul de zăpadă și rugini). Producție mai mare de sămânță-până la 1000 kg /ha și talie mai mică a plantelor. Pretabil pentru gazon și pășunat, dar și pentru fâneată, **Figura 9**.



Figura 9. *Lolium perenne* X *Brevis*.

1.10 Obiectivele ameliorării raigrasului peren în condițiile schimbărilor climatice

1.10.1 Producția de furaj

Este obiectivul central al ameliorării, reprezentând principalul criteriu de evaluare agronomică. Potențialul productiv reflectă nivelul maxim de biomasă economic utilă pe care un genotip îl poate realiza în condiții optime. Acest caracter are control genetic poligenic, fiind influențat și de interacțiunea genotip–mediu. Factorii determinanți includ: vigoarea plantelor, capacitatea de înfrățire, talia, dezvoltarea radiculară, ritmul de regenerare, rezistența la stres termic și hidric, precum și toleranța la boli și dăunători. Se urmărește totodată reducerea corelațiilor negative între producția de furaj și cea de sămânță și creșterea perenității, în special la *Lolium perenne* L.

1.10.2 Potențialul de fructificare

Obținerea unor soiuri cu producție mare de sămânță este esențială pentru extinderea lor în cultură. Producția de sămânță este influențată de factori genetici, fiziologici și de mediu. Corelația negativă dintre numărul de lăstari vegetativi (furaj) și generativi (sămânță) îngreunează selecția, obiectivul fiind identificarea genotipurilor echilibrate. Reducerea scuturării semințelor și uniformizarea înfloritului sunt direcții importante pentru eficientizarea multiplicării.

1.10.3. Calitatea furajului

Calitatea nutritivă depinde de compoziția chimică, digestibilitate, palatabilitate și valoarea alimentară. Se urmărește selecția unor genotipuri cu proporție mare de frunze, conținut ridicat în proteine, glucide solubile și fibre ușor digerabile, dar sărace în lignină. Raportul frunze/lăstari este un indicator practic al calității. Momentul recoltării are o influență majoră – după înspicare, calitatea scade semnificativ.

1.10.4 Rezistența la boli și dăunători

Rezistența genetică este o componentă esențială a ameliorării durabile, mai ales în contextul încălzirii climatice, care favorizează proliferarea agenților patogeni. Bolile fungice (în special ruginile produse de *Puccinia* sp.) afectează puternic raigrasul peren și sunt prioritare în selecția pentru rezistență. Soiurile rezistente asigură stabilitatea producțiilor și reduc degradarea covorului vegetal.

1.10.5 Rezistența la stres abiotic

a) Seceta: toleranța la deficitul hidric reprezintă capacitatea de a menține o reducere minimă a producției de furaj în perioade uscate. Este determinată de factori morfologici (sistem radicular profund, frunze erecte, cuticulă groasă, densitate stomatală optimă) și fiziologici (reglarea presiunii osmotice, regenerare rapidă post-secetă).

b) Gerul și iernarea: rezistența la temperaturi scăzute este corelată cu acumularea de zaharuri și adaptări celulare care previn deteriorarea țesuturilor. Se valorifică sursele genetice provenite din zone reci și genotipuri cu conținut ridicat de glucide solubile și acizi grași nesaturați.

c) Solurile acide și saline: crearea de genotipuri tolerante la aciditate și aluminiu mobil reprezintă o soluție economică și ecologică pentru zonele podzolice. În cazul solurilor saline, selecția vizează genotipuri cu capacitate osmotică ridicată și toleranță demonstrată la germinare și creșterea plantulelor.

1.10.6 Adaptabilitatea și competitivitatea în amestecuri

Adaptabilitatea exprimă capacitatea genetică de a menține producții stabile în condiții variate de mediu și este asociată cu noțiunea de plasticitate ecologică. În culturile mixte de graminee și leguminoase, competitivitatea soiurilor se urmărește prin selecția unor genotipuri cu port vertical, creștere echilibrată și epocă de recoltare similară, pentru a favoriza cooperarea și reducerea competiției interspecifice.

1.10.7 Precocitatea

Este o însușire ereditară influențată de fotoperiodism, termoperiodism și disponibilitatea apei. În contextul schimbărilor climatice, ameliorarea se orientează către soiuri timpurii, care finalizează producția înaintea perioadelor de secetă și temperaturi extreme.

1.11 Metode de ameliorare

Soiul, în accepțiunea actuală poate fi definit ca o grupare de plante aparținând aceluiași grup taxonomic (familie, gen, specie, subspecie) care se caracterizează prin anumite însușiri ereditare comune, distincte și stabile de la o generație la alta.

Definirea soiului are la baza următoarele criterii, (*Savatti, M., 2004*) :

- **Criteriul de individualitate sau identitate**, prin care un soi se distinge de alt soi prin una sau mai multe caracteristici care sunt proprii, conferindu-i o anumită particularitate (talie, perioadă de vegetație, forma habitusului, etc).
- **Criteriul de omogenitate** are în vedere asemănarea privind principalele caracteristici ale ansamblului de indivizi care alcătuiesc soiul .
- **Criteriul stabilității** care se referă la menținerea în urma reproducerilor succesive, a caracterelor și însușirilor de identitate și omogenitate.

Utilizarea metodelor de ameliorare este în concordanță cu particularitățile genetice, biologice și botanice care caracterizează speciile de graminee perene de pajiști:

- perenitatea,
- posibilitățile de înmulțire atât prin semințe cât și pe cale vegetativă,
- autoincompatibilitatea,
- heterozigoția,
- gradul de ploidie,
- alogamia,
- polenizarea anemofilă,
- controlul poligenic al majorității caracterelor agronomice

În cazul ameliorării speciilor de plante furajere perene de pajiști se exploatează natura heterozigotă a plantelor individuale, rezultată în urma polenizării libere. Ca urmare genotipul unei plante heterozigote nu se reproduce cu exactitate în descendență, el fiind o evaluare a performanțelor rezultate dintr-o încrucișare randomizată a unui genotip matern cu origine cunoscută cu un genotip patern (polen) cu origine necunoscută.

Existența fenomenului de superioritate a heterozigoților sub aspectul producției, calității, capacității de adaptare, rezistenței la factorii de stres, respectiv vigoarea hibridă și-a pus amprenta asupra întregii activități de ameliorare, (Breazu, I., Grusea A., 1997).

- **Selecția în masă**

Cea mai veche metodă de ameliorare utilizată la plantele furajere prin simplitatea și ușurința de realizare se bazează pe recoltarea în masă a semințelor din populațiile locale stabilizate prin selecție naturală. Considerată o metodă primitivă de ameliorare, în anumite situații ea este eficientă, ca o fază pregătitoare a unei faze decisive. Astfel, înainte de încrucișarea liberă a două sau mai multe populații, pregătirea genitorilor se face prin alegerea în masă, (Varga et al., 1998). Bazată pe o mare variabilitate genetică naturală, ameliorarea constă în mărirea frecvenței unor gene care corespund obiectivelor de ameliorare, (Varga et al., 1998).

- **Selecția recurentă (selecția periodică)**

Selecția recurentă contribuie la îmbunătățirea substanțială a populațiilor care se folosesc la crearea de soiuri sintetice. Fiecare ciclu al selecției recurente necesită evaluarea plantelor valoroase și încrucișarea în toate combinațiile posibile a descendențelor acestora.

- **Metodele polycross și topcross**

Sunt un complex de metode și procedee prin care se obțin soiuri sintetice și sunt metodele folosite în general la crearea de soiuri sintetice la speciile de graminee și leguminoase perene de pajiști.

- **Soiul sintetic** este o populație de plante, rezultată din multiplicarea sexuată prin polenizare liberă a descendențelor unor încrucișări multiple între un anumit număr de constituenți, selectați pentru anumite însușiri. Componentele sinteticului trebuie să fie asemănătoare din punct de vedere fenotipic, dar deosebite din punct de vedere genetic pentru a permite exteriorizarea unui efect heterozis maxim, (Varga et al., 1998).

Procedee pentru obținerea unui soi sintetic poate fi sumarizat la următoarele:

- sinteticul este constituit din unități reproductibile, clone;
- materialul biologic ce alcătuiește soiul sintetic se alege pe baza performanțelor, sub aspectul capacității combinative sau a testelor de performanță a descendenților;
- sinteticul este constituit printr-o interpolenizare randomizată a genotipurilor componente;
- unitățile componente pot fi menținute vegetativ, prin clonare atât timp cât sinteticul este necesar să fie reconstituit.

II. MATERIALE ȘI METODE

2.1	Recoltarea plantelor
2.2	Cadru general. Controlul calității probelor. Construcția bibliotecilor de date, secvențierea
2.2.1	Controlul calității și pregătirea bibliotecilor
2.3	Clusterizare și secvențiere realizate după controlul calității bibliotecilor
2.4	Controlul calității datelor. Analiza bioinformatică

Prezentul studiu vizează o analiză comparativă a profilurilor de expresie genică (transcriptom) la specia *Lolium perenne* L. (iarbă de gazon perenă) recoltată din trei medii ecologice distincte din România—Arad (câmpie), Brașov (zonă montană joasă) și Platoul Munților Bucegi (altitudine înaltă)—precum și la varietatea ameliorată *Lolium perenne* 'Mara'. Înțelegerea răspunsurilor transcriptomice la stresul abiotic impus de diferențele geografice, altitudinale și microclimatice este esențială pentru cercetarea în domeniul ameliorării plantelor. Din perspectiva bioinformatică, *Lolium perenne* L. este considerată o specie non-model, întrucât dispune doar de un genom draft parțial asamblat. Această limitare metodologică impune o rigurozitate sporită în controlul calității datelor brute (Data Quality Control – DQC), deoarece orice eroare sau bias de secvențiere ar putea compromite fie acuratețea aliniamentului la genomul de referință, fie complexitatea necesară pentru o eventuală asamblare *de novo* a transcriptomului. Controlul calității devine astfel o etapă determinantă pentru succesul cuantificării expresiei genice și pentru analiza expresiei diferențiale (Differential Gene Expression – DGE).

2.1 Recoltarea plantelor

Au fost recoltate probe la specia *Lolium perenne* L. și varietatea *Lolium perenne* 'Mara'. Probele recoltate au fost reprezentate de frunze (3 indivizi diferiți pentru fiecare specie, varietate) din 3 locatii diferite, avand urmatoarele coordonate Degrees/Minutes/Seconds (DMS), respectiv atitudini: Baza de Cercetări Pajiști Montane, Blana – Bucegi, Platoul Bucegilor, Bucegi Mts. - 45°24'36.00" N 25°26'60.00" E, 1788 m; Arad aria metropolitană - 46°10'2.37" N 21°14'28.55" E, 102m; Brașov, aria metropolitană – 45°38'11.19" N 25°30'53.39" E, 569m, **Figura 9, Tabel 1.**

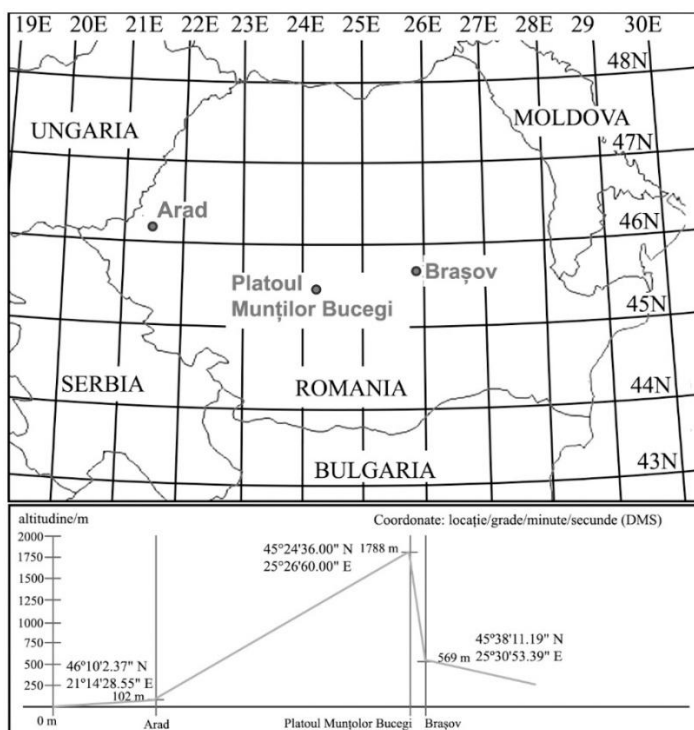


Figura 9. Reprezentare grafică a locațiilor cu punctele de recoltare pentru specia *Lolium perenne* L. și *Lolium perenne* 'Mara'.

Tabel 1. Specia/varietatea în varianta triplicată utilizate în experiment.

Nr. probă	Specia	Cantitate ADN extrasă (exprimată în ng/ μ l)	Raport 260/280	Tip probă
1	<i>Lolium perenne</i> L.	202,1	1,78	ADN
2	<i>Lolium perenne</i> 'Mara'	101,6	2,03	ADN

2.2 Cadru general. Controlul calității probelor. Construcția bibliotecelor de date, secvențierea.

Transcriptomul reprezintă un set al tuturor transcripturilor dintr-o celulă sau o populație de celule aflate într-un anumit stadiu. Analiza transcriptomului ajută la studierea identificării genelor care sunt exprimate diferențial în populații celulare distincte. Cercetătorii pot, de asemenea, obține o perspectivă mai profundă asupra identificării limitelor genelor, clivajului variabil și variației transcrierilor (*Wang et al., 2009*).

Secvențierea ARN prin intermediul platformelor Illumina, bazată pe mecanismul SBS (secvențiere prin sinteză), oferă o gamă largă de beneficii în ceea ce privește randamentul ridicat și precizia ridicată, cu un necesar redus de probe. Această metodă tehnică poate fi un instrument puternic pentru cercetarea activității transcripționale a ARN-ului. Fluxul de lucru este prezentat după cum urmează în **Figura 10**:

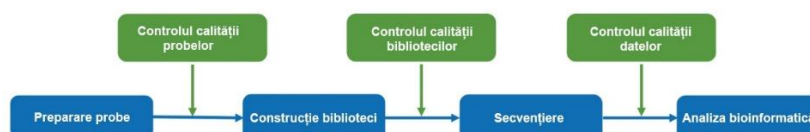


Figura 10. Fluxul de lucru în analiza transcriptomică.

ARN-ul mesager a fost purificat din ARN-ul total folosind sfere magnetice atașate la oligo poli-T. După fragmentare, ADNc-ul primei catene a fost sintetizat folosind primeri hexameri aleatorii, urmat de sinteza ADNc-ului celei de-a doua catene folosind fie dTTP pentru biblioteca nespecifică catenei, fie dUTP pentru biblioteca specifică catenei.

Biblioteca nespecifică catenei a fost finalizată după repararea capetelor, coada A, ligatura adaptorului, selecția dimensiunii, amplificare și purificare (**Figura 11. A/B**). Biblioteca specifică catenei a fost gata după repararea capetelor, coada A, ligatura adaptorului, selecția dimensiunii, amplificare și purificare.

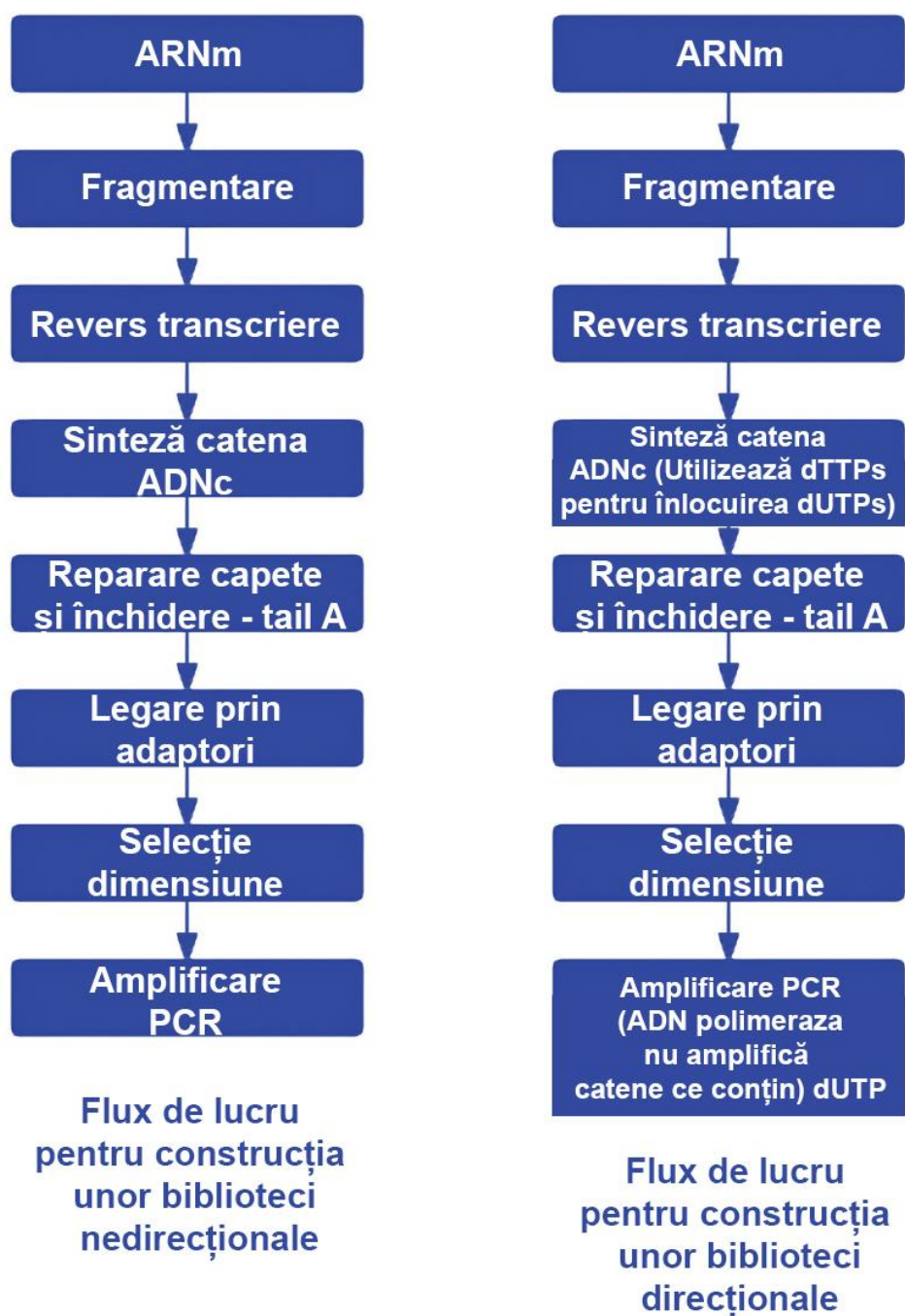


Figura 11. Protocol non-direcțional (A) și direcțional (B) de construcție al bibliotecilor.

Bibliotecile a fost verificate cu Qubit și PCR în timp real pentru cuantificare și cu bioanalizator pentru detectarea distribuției dimensionale. Bibliotecile cuantificate au fost grupate și secvențiate pe platforme Illumina, în funcție de concentrația efectivă a bibliotecii și cantitatea de date.

2.2.1 Controlul calității și pregătirea bibliotecilor.

Controlul calității probei și integritatea ARN-ului au fost evaluate utilizând sistemul Bioanalyzer 2100 (Agilent Technologies, CA, SUA).

Pregătirea bibliotecilor pentru secvențierea transcriptomului aferent bibliotecii non-specifice pentru catenă ARN-ul mesager a fost purificată din ARN-ul total folosind oligo-atașamente poli-T la "perle" magnetice. După fragmentare, prima catenă de ADNc a fost sintetizată folosind primeri hexameri aleatori urmate de cea de a 2 -a sinteza a ADNc catenar. Biblioteca a fost finalizată după sigilarea capetelor, prin atașarea porțiunii terminale "tail" A, ligatura adaptorului, selecția dimensiunii, amplificarea și purificarea. Biblioteca a fost verificată cu Qubit și RTPCR pentru cuantificare respectiv cu un bioanalizator pentru detectarea distribuției dimensionale. Biblioteca specifică pentru ARN-ul mesager a fost purificată din ARN-ul total folosind oligo-atașat poli-T perle magnetice. După fragmentare, prima catenă de ADNc a fost sintetizată folosind primeri hexameri aleatori. Apoi, a doua catenă ADNc a fost sintetizată folosind dUTP, în loc de dTTP. Direcționalitatea bibliotecii a fost gata după repararea capătului, coada A, ligatura adaptorului, dimensiune selecție, amplificare și purificare. Biblioteca a fost verificată cu Qubit și RT PCR pentru cuantificare respectiv cu un bioanalizator pentru detectarea distribuției dimensionale.

2.3 Clusterizare și secvențiere după controlul calității bibliotecilor

Diferitele biblioteci au fost grupate în funcție de concentrația efectivă și cantitatea de date vizate, apoi supuse la secvențierea Illumina. Principiul de bază al secvențierii este „Secvențierea prin „Sinteză”, unde dNTP-uri marcate fluorescent, ADN polimerază și primerii adaptorii sunt adăugați în celula de flux de secvențiere pentru amplificare. Pe măsură ce fiecare grup de secvențiere își extinde catena complementară, adăugarea fiecărui dNTP marcat fluorescent eliberează un corespondent semnal de fluorescență. Secvențiatorul captează aceste semnale de fluorescență și le transformă în vârfuri de secvențiere prin intermediul software-ului, obținând astfel informațiile despre secvența fragmentului țintă.

2.4 Controlul calității datelor. Analiză bioinformatică.

Datele brute (citirile brute) în format fastq au fost procesate mai întâi prin fastp software. În această etapă, datele curate (citiri curate) au fost obținute prin eliminarea citirilor care conțin adaptorul, a citirilor care conțin ploy-N și citiri de calitate slabă din datele brute. În același timp, conținutul Q20, Q30 și GC din datele curate au fost calculate. Toate analizele ulterioare s-au bazat pe date "curate" de înaltă calitate.

2.4.1 Citirea și "maparea"/cartografierea la genomul de referință, fișierele de adnotare a genomului de referință și a modelului genetic au fost descărcate direct de pe site-ul web al genomului. Indexul genomului de referință a fost construit folosind Hisat2 v2.0.5 și citirile curate la capătul pereche au fost aliniate la genom de referință folosind Hisat2 v2.0.5. Am selectat Hisat2 ca instrument de mapare pentru care Hisat2 poate genera o bază de date cu joncțiuni de îmbinare bazate pe fișierul de adnotare al modelului genetic și prin urmare, un rezultat mai bun al cartografierii decât la alte instrumente de mapare/cartografiere non-splicing.

2.4.2 Predicția noilor transcrieri, citirile mapate ale fiecărei mostre au fost asamblate de StringTie (v1.3.3b) într-o abordare bazată pe referințe. StringTie folosește un algoritm nou de flux de rețea, precum și un algoritm opțional *de novo* etapa de asamblare pentru asamblarea și cuantificarea transcrierilor complete reprezentând mai multe variante de splicing pentru fiecare locus genic.

2.4.3 Cuantificarea nivelului de expresie genică, s-a utilizat featureCounts v1.5.0-p3 pentru a număra numărul de citiri mapate pentru fiecare genă. Apoi, FPKM-ul (Fragments Per Kilobase of transcript per Million mapped reads/ Fragmente per Kilobază de transcript per Milion de citiri mapate) fiecărei gene a fost calculat pe baza lungimii genei și numărul de citiri mapate la această genă. FPKM, ia în considerare efectul "adâncimii" de secvențiere și lungimea genei pentru numărul de citiri în același timp și este în prezent cea mai frecvent utilizată metodă pentru estimarea nivelului de expresie genică.

2.4.4 Analiza îmbogățirii setului de gene (GSEA), este o abordare computațională pentru a determina dacă un set genetic predefinit poate prezenta o consecvență semnificativă ca diferența dintre două stări biologice. Genele au fost clasificate în funcție de gradul de expresie diferențială și apoi setul de gene predefinit a fost testat pentru a vedea dacă a fost îmbogățit la partea de sus sau de jos a listei. Analiza îmbogățirii setului de gene poate include modificări subtile ale expresiei. Am folosit versiunea locală a GSEA ca instrument de analiză <http://www.broadinstitute.org/gsea/index.jsp>, GO, KEGG. Setul de date a fost utilizat independent pentru GSEA.

2.4.5 Analiza corelațiilor, replicile biologice sunt necesare pentru orice experiment biologic, inclusiv tehnologia ARN-seq. Corelarea nivelurilor de expresie genică între probe joacă un rol important în verificarea fiabilității și a

selecției probelor, ceea ce poate nu numai demonstra repetabilitatea experimentului, dar și estima analiza expresiei genice diferențiale.

2.4.6 Analiza expresiei diferențiale DESeq2 cu replici biologice: analiza expresiei diferențiale pentru două condiții/grupuri a fost efectuată folosind pachetul DESeq2, R. DESeq2 oferă programe statistice pentru determinarea expresiei diferențiale în datele digitale de expresie genică utilizând modele bazate pe distribuțiile binomiale negative. Valoarea P rezultată este ajustată folosind metodele lui Benjamini și Hochberg pentru a controla rata de descoperire a erorilor. Rezultatul P-value a fost ajustat utilizând Benjamini & Hochberg” metod pentru controlul erorii. Valoarea corectă P-value ≤ 0.05 & $|\log_2(\text{foldchange})| \geq 1$ a fost stabilit ca valoare de prag a semnificație a expresiei diferențiale.

Analiza funcțională

2.4.7 Analizele de îmbogățire GO(ontologia genetică) și 2.4.8 KEGG(Kyoto Encyclopedia of Genes and Genomes), analiza genelor exprimate diferențial au fost implementate de pachetul clusterProfiler. Termenii GO cu valoarea P corectată mai mică decât 0,05 au fost considerați îmbogățiți semnificativ prin exprimare diferențială a genelor. KEGG este o resursă de baze de date pentru înțelegerea genelor la nivel înalt al funcțiilor și utilitățile acestora în sisteme biologice, cum ar fi celula, organism și ecosistem. Am folosit pachetul clusterProfiler R pentru a testa îmbogățirea statistică a genelor de expresie diferențială în căile KEGG.

2.4.9 Analiza PPI(Protein-Protein Interaction) a genelor exprimate diferențial a fost realizată utilizând baza de date STRING, care conține proteine cunoscute și interacțiunile dintre acestea. Pentru speciile prezente în baza de date, s-a construit o rețea prin extragerea listei de gene țintă din baza de date.

Analiza structurală

2.4.10 Analiza SNPs și InDels, un polimorfism cu un singur nucleotid (SNP) este o substituție a secvențelor de nucleotide care apar în mod obișnuit într-o populație (de exemplu, 1%), în care un singur nucleotid din genom sau alte schimburi de secvențe partajate se produc între membrii unei specii biologice sau cromozomi perechi. Două tipuri de variații apar la SNP-uri, și anume tranzițiile și transversile, cu un raport de probabilitate de 1:2. Regiunea cu apariție ridicată a SNP-urilor sunt adesea secvențele CG, rezultând tranziții de la C la T, care sunt asociate cu tendința lui C de a fi metilat. Spre deosebire de SNP-uri, InDeL este un termen de biologie moleculară pentru o inserție sau deleție de baze în genomul unui organism, cu o lungime de la 1 la 10.000 de perechi de baze, care este clasificată printre variațiile genetice mici.

2.4.11 Analiza splicing-ului alternativ/matisării alternative, splicing-ul alternativ (AS) este un proces universal de reglare genică la majoritatea eucariotelor. Genele eucariote divizate constau din secvențe intronice și exonice.

În timpul procesării ARN-ului, intronii sunt îmbinați de spliceozomi, în timp ce exonii sunt editați continuu și reținuți în ARNm-urile mature. Unele pre-ARNm pot avea modele de splicing diferit în condiții diferite și pot produce izoforme proteice diferite, ceea ce crește complexitatea biologică și adaptabilitatea eucariotelor.

Schema generală a analizei bioinformatică este prezentată în **Figura 12**.

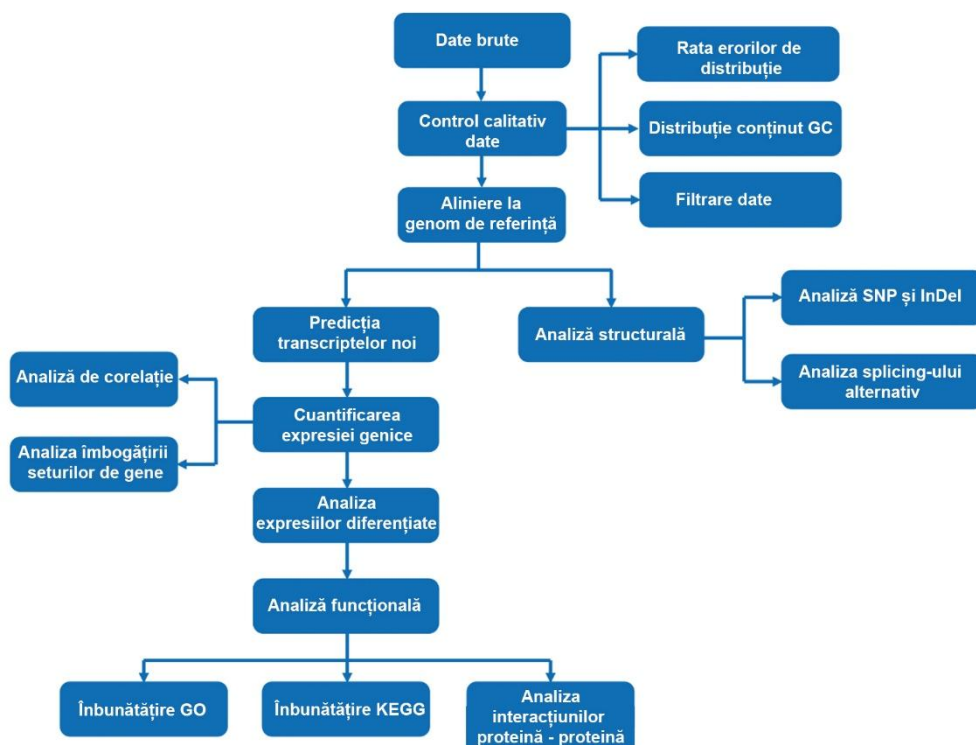


Figura 12. Fluxul tehnologic și analiza bioinformatică din RNA-seq.

Indicative pentru probele utilizate în prezentul studiu:

LP1_BV – individ nr. 1 recoltat din Brașov
 LP2_BV – individ nr. 2 recoltat din Brașov
 LP2_BV – individ nr. 3 recoltat din Brașov
 LP1_AR – individ nr. 1 recoltat din Arad
 LP2_AR – individ nr. 2 recoltat din Arad
 LP3_AR – individ nr. 3 recoltat din Arad
 LP1/LP1_B_Mts - individ nr. 1 recoltat din Platoul Munților Bucegi
 LP2/LP2_B_Mts - individ nr. 2 recoltat din Platoul Munților Bucegi
 LP3/LP3_B_Mts - individ nr. 3 recoltat din Platoul Munților Bucegi
 LP1_A – individ nr. 1 varianta ameliorată

LP2_A – individ nr. 2 varianta ameliorată

LP3_A – individ nr. 3 varianta ameliorată

III. REZULTATE ȘI DISCUȚII

3.1	Controlul calitativ al datelor
3.1.1	Date brute
3.1.2	Examinarea ratei de eroare în secvențiere
3.1.3	Distribuția GC în conținutul secvențiat
3.1.4	Filtrarea datelor de secvențiere
3.1.5	Date statistice cu referire la calitatea secvențierii
3.2	Alinierea
3.2.1	Rezultate aliniere/„mapare”
3.2.2	Distribuția citirilor la genomul de referință
3.2.3	Rezultate cu privire la vizualizarea alinierilor
3.3	Analiza genelor nou descoperite
3.3.1	Identificarea/ Predicția de gene noi
3.3.2	Adnotarea genelor noi
3.4	Analiza nivelului de expresie genică
3.4.1	Cuantificarea expresiei genice
3.4.2	Distribuția nivelului de expresie genică
3.4.3	Analiza de corelație
3.4.4	Analiza componentelor principale (ACP)
3.4.5	Co-expresia în diagramele Venn
3.5	Analiza expresiei genice diferențiale
3.5.1	Rezultat analiză expresie diferențială
3.5.2	Date de expresie genică diferențială interpretate statistic
3.5.3	Analiza de clusterizare
3.6	Analiza funcțională
3.6.1	Analiza de îmbogățire a ontologiei genelor/GO/Gene ontology
3.6.2	Analiza de îmbogățire a căilor metabolice KEGG (Kyoto Encyclopedia of Genes and Genomes)
3.6.3	Analiza rețelilor de interacțiune proteină-proteină (PPI)
3.7	Analiza structurală
3.7.1	Analiza splicing-ului alternativ
3.7.1.1	Detectarea izoformelor exprimate diferențial
3.7.1.2	Vizualizarea evenimentelor de splicing alternativ (AS)
3.7.2	Analiza SNPs și InDel
3.7.2.1	Identificarea siturilor polimorfice
3.7.2.2	Statistica siturilor polimorfice

3.1	Controlul calitativ al datelor
3.1.1	Date brute
3.1.2	Examinarea ratei de eroare în secvențiere
3.1.3	Distribuția GC în conținutul secvențiat
3.1.4	Filtrarea datelor de secvențiere
3.1.5	Date statistice cu referire la calitatea secvențierii

3.1 Controlul calitativ al datelor

Controlul calității datelor brute (FASTQ) este prima și cea mai importantă barieră de protecție împotriva artefactelor tehnice, care pot introduce variabilitate nejustificată sau pot reduce puterea statistică a experimentului.

3.1.1 Date brute

A) Standarde de Calitate și Volumetrice

Deși volumul exact de citiri per eșantion nu este furnizat, standardele academice și industriale, cum ar fi cele recomandate de Consorțiul ENCODE, sugerează că fiecare replică biologică ar trebui să atingă un minim de 10×10^6 citiri aliniate. Respectarea acestui prag volumetric (adâncime de secvențiere) este esențială pentru a asigura detectarea fiabilă a transcriptelor slab exprimate.

O adâncime de secvențiere insuficientă sau o rată mare de nealinieră cauzată de o calitate scăzută a citirilor are ca efect direct creșterea varianței tehnice. Această varianță poate afecta negativ acuratețea pachetelor statistice utilizate pentru DGE(/date de expresie genică diferențială), cum ar fi DESeq2, care se bazează pe distribuția negativ-binomială a numărului de citiri. O cuantificare eronată sau incompletă generează o matrice de numere brute (raw counts table) deficitară, ceea ce se traduce printr-o putere statistică redusă în detectarea schimbărilor reale de expresie între grupurile geografice și varietatea ameliorată.

B) Lungimea Citirilor și Consistența Protocolului

O metrică importantă a datelor brute este lungimea citirilor secvențiate. Analiza profilurilor de distribuție a ratei de eroare (Error Rate Distribution along reads) pentru toate cele 12 probe indică o lungime uniformă a citirilor de 300 de unități (baze pereche sau baze simple). Această uniformitate în lungimea citirii sugerează o consistență metodologică excelentă în etapa de preparare a bibliotecilor de ARN și în protocolul de secvențiere, un factor care reduce variabilitatea tehnică între eșantioane. Cu toate acestea, este necesară o examinare atentă a diferențelor de calitate observate între grupurile geografice (Arad, Brașov, Bucegi și Mara). Deși lungimea citirii este omogenă, se constată că eșantioanele din Bucegi (LP2) și Brașov (LP1_BV, LP2_BV) tind să prezinte rate de eroare maximă mai ridicate comparativ cu cele din Arad sau Mara. Aceste variații subtile ale metricilor QC pot fi un indiciu al efectelor de lot (batch effects)

în secvențiere. Dacă eşantioanele au fost procesate sau secvențiate în loturi diferite, variațiile în calitatea chimică a rulării pot masca sau distorsiona diferențele biologice, impunând includerea acestui factor (lotul) ca o covarianță în modelul statistic de analiză a expresiei diferențiale.

C) Pre-Filtrarea Datelor (Raw Counts Filtering)

Fluxul de lucru standard pentru DGE impune o pre-filtrare a tabelului de numere brute (raw counts table) obținut după aliniere. Obiectivul este eliminarea genelor cu număr de citiri extrem de mic, care cel mai adesea reprezintă zgomot de fond, și nu o expresie biologică validă. O strategie comună este eliminarea oricărei gene cu un număr total de citiri (row sum) sub un prag minim (de exemplu, 10 citiri pe toate eşantioanele). Această etapă nu doar că elimină „zgomotul”, dar ajută și la creșterea puterii statistice prin reducerea numărului de teste multiple care necesită corecție (Multiple Testing Correction).

3.1.2 Examinarea ratei de eroare în secvențiere

Analiza ratei de eroare (E) în funcție de poziția de-a lungul citirilor este un indicator esențial al fidelității datelor (Erlich Y, Mitra PP et al.2008; Jiang L, Schlesinger F et al.2011). Rata de eroare este direct legată de Scorul Phred (Q), unde $Q = -10 \log_{10}(E)$, (Jiang et al.). Un scor Q20 ($E \leq 0.010$ sau 1% rată de eroare) este considerat pragul minim pentru o bază de înaltă calitate, **Figura 13**, mai multe detalii specifice pentru fiecare probă în parte se regăsesc în *Anexa 1 - Figurile 1-12*.

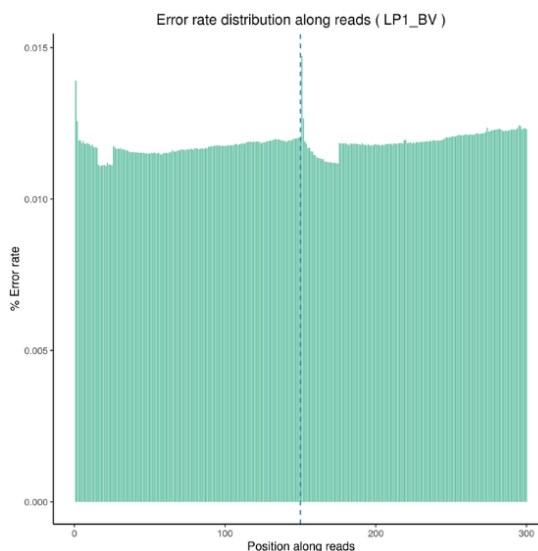


Figura 13. Distribuția ratei de eroare a datelor de secvențiere, LP1_BV.*

* Axa x reprezintă poziția bazei de-a lungul fiecărei secvențe citite (read), iar axa y indică rata de eroare a bazei.

3.1.3 Distribuția GC în conținutul secvențiat

Distribuția conținutului GC detectează o potențială **separare AT/GC**, fapt care afectează cuantificarea ulterioară a expresiei genice. Având în vedere fragmentarea aleatorie și legile biologice ale conținutului G/C-A/T, proporțiile de G și C, respectiv de A și T, acestea ar trebui să fie egale, iar conținutul ar trebui să rămână stabil pe parcursul întregului proces de secvențiere pentru o **bibliotecă non-direcțională** (Parkhomchuk et al., 2009). Dacă biblioteca este **specifică de catenă** (*strand-specific*), poate apărea fenomenul de separare AT sau GC. O variație mare a erorii de secvențiere în primele 6-7 baze este permisă, luând în considerare utilizarea **primerilor aleatorii** în construcția bibliotecii; în tehnologiile actuale de secvențiere de mare randament (*high-throughput*), este normal ca primele baze să prezinte o anumită preferință de compoziție.

Terminologie cheie utilizată:

- **GC content:** conținut GC (proporția de guanină și citozină).
- **Gene expression quantification:** cuantificarea expresiei genice.
- **Non-stranded library:** bibliotecă non-direcțională (unde nu se păstrează informația despre catenă de origine).
- **Strand-specific library:** bibliotecă specifică de catenă (sau direcțională).
- **Random primer:** primer aleatoriu (folosiți adesea pentru inițierea sintezei cDNA).
- **High-throughput sequencing:** secvențiere de mare randament (NGS).

Profilul de compoziție per bază

Așa cum se poate observa în **Figura 14** (proba LP1_BV), respectiv pentru restul probelor (*Anexa 1 – Figurile 13-24*), profilul prezintă un comportament tipic pentru bibliotecile de secvențiere de tip Illumina:

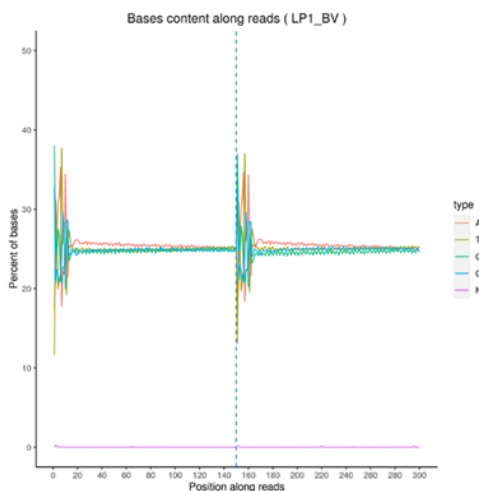


Figura 14. Distribuția în conținut GC.

- **Faza de „instabilitate” (pozițiile 1-9):** se observă oscilații evidente ale liniilor A, T, C și G. Aceasta nu este o eroare, ci rezultatul **primerilor aleatorii** utilizați la începutul sintezei. Aceștia au o preferință biochimică pentru anumite secvențe, ceea ce face ca primele baze să nu fie distribuite perfect egal.
- **Faza de echilibru (după poziția 10):** liniile devin paralele și plate. Această stabilitate confirmă faptul că, pe măsură ce citirea înaintează, compoziția nucleotidelor reflectă fidel genomul/transcriptomul țintă, fără interferențe tehnice.

Evaluarea separării AT/GC

Conform cerințelor tehnice menționate anterior:

- **Pentru biblioteci non-direcționale:** în graficele rezultate, liniile pentru **G și C** trebui să fie aproape suprapuse, iar liniile pentru **A și T** la fel.
- **Constatare:** dacă în toate cele 12 probe aceste perechi rămân apropiate, înseamnă că nu există o „separare AT/GC” problematică. Acest lucru garantează o **cuantificare corectă a expresiei genice** ulterior.

Comparația între grupuri (Reproducibilitatea)

Faptul că analiza include triplicate (LP1, LP2, LP3) pentru fiecare condiție (BV, AR, A și B_Mts) permite o observație crucială:

- **Uniformitatea probelor:** graficele pentru grupul **BV** sunt aproape identice cu cele pentru **AR** sau **A**.
- **Concluzie:** nu există variații sistematice introduse de tratamentul probelor sau de procesul de laborator. Datele sunt extrem de robuste și pot fi comparate statistic în etapele de analiză diferențială.

3.1.4 Filtrarea datelor de secvențiere

Citirile de secvențiere (citirile brute sau *raw reads*) conțin adesea secvențe de calitate scăzută sau citiri care includ adaptorii, factori ce pot afecta calitatea analizelor bioinformatiche ulterioare (*downstream*). Pentru a preveni acest lucru, este necesară filtrarea citirilor brute pentru a obține citirile curate (*clean reads*). Procesul de filtrare a citirilor brute este următorul, **Figura 15**:



Figura 15. Etapele fluxului de lucru pentru filtrarea datelor.

Criteriile de filtrare a citirilor (reads):

1. **Eliminarea citirilor contaminate cu adaptorii:** se elimină secvențele care conțin fragmente din adaptorii utilizați în construcția bibliotecii.

2. **Eliminarea citirilor cu nucleotide incerte ($N > 10\%$):** se elimină citirile în care bazele nedeterminate (notate cu „N”) constituie mai mult de 10% din lungimea totală a citirii.
3. **Eliminarea citirilor de calitate scăzută:** se elimină citirile în care nucleotidele de calitate slabă (cu un scor de calitate al bazei mai mic de 5) reprezintă mai mult de 50% din întreaga citire (Yan et al., 2011).

Secvențele adaptorilor au fost următoarele:

P5 adapter :

P5→P7'(5'→3')

AATGATACGGCGACCACCGAGATCTACAC[i5]ACACTCTTTCCCTACACGACGC
TCTTCCGATCT

P7 adapter :

P5→P7'(5'→3')

GATCGGAAGAGCACACGTCTGAACTCCAGTCAC[i7]ATCTCGTATGCCGTCTTC
TGCTTG

Rezultatele citirilor brute pe categorii din punctul de vedere al calității și locației probelor au fost generate sub forma unor grafice, așa cum se observă și în **Figura 16**(pentru proba (LP1_BV), detaliile pentru restul probelor se găsesc în *Anexa 1*, Figurile 25-36.

Classification of Raw Reads (LP1_BV)

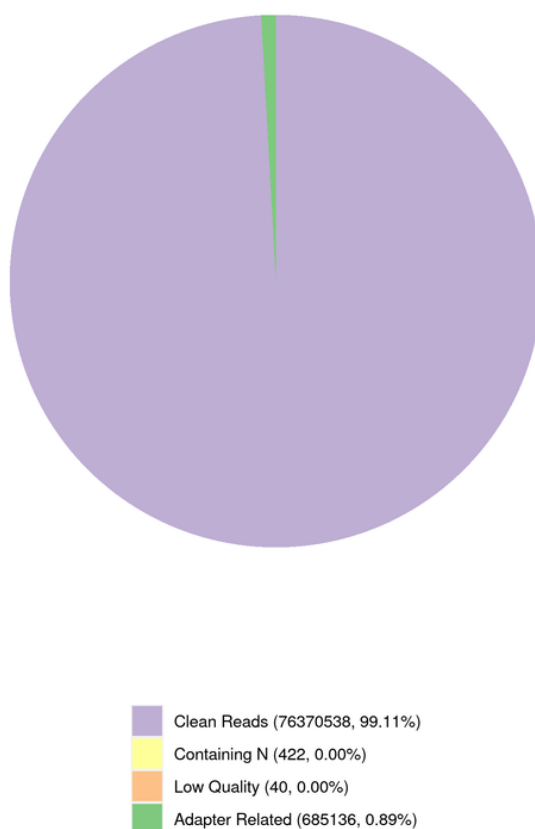


Figura 16. Clasificarea citirilor brute (LP1_BV).

3.1.5 Date statistice cu referire la calitatea secvențierii

Calitatea citirilor de secvențiere este cuantificată prin Scorurile Phred (Q-scores), care măsoară probabilitatea ca o bază să fie apelată incorect. Relația logaritmică este definită prin formula $Q = -10\log_{10}P$, unde P este probabilitatea de eroare.

Standardele Acurateței NGS: Q20 și Q30

Scorurile Q20 și Q30 sunt etaloane industriale pentru evaluarea calității datelor NGS.

Q20: reprezintă o acuratețe a apelării bazei de 99%, echivalentul unei erori la 100 de baze.

Q30: reprezintă o acuratețe de 99.9%, echivalentul unei erori la 1000 de baze. Atingerea unui procent ridicat de baze peste Q30 este considerată standardul pentru datele de înaltă calitate necesare în aplicațiile de citire a expresiei, cum ar fi RNA-Seq, deoarece asigură faptul că majoritatea citirilor vor fi perfecte sau vor conține erori minime.

Deși datele specifice privind procentul de baze Q20 și Q30 nu sunt furnizate în materialul de cercetare, se poate realiza o evaluare robustă a calității secvențierii utilizând metricile de filtrare deja analizate, care servesc ca indicatori de proxy.

Procentul extrem de mare de Clean Reads (între 97.44% și 99.14%) este un indicator puternic al calității excelente a bazei de date obținute. În plus, eliminarea totală a citirilor Low Quality (raportată ca 0.00%) confirmă faptul că baza de date rezultată conține doar secvențe cu scoruri Phred agregate semnificativ peste pragurile de acceptabilitate minimă. Aceste rezultate sugerează că setul de date Clean Reads are o performanță Q30 în concordanță cu standardele înalte ale platformelor Illumina.

Uniformitatea calității secvențierii între probe sălbatice (LP_B_Mts, LP_AR, LP_BV) și cele ameliorate (LP_A) este remarcabilă. Această omogenitate tehnică confirmă că metodologia de secvențiere a fost standardizată și robustă, indiferent de originea biologică sau de stresul geo-climatic suportat de planta sursă.

Deoarece calitatea datelor este uniform ridicată, erorile introduse de baze incorecte în timpul alinamentului ulterior a fost minim. Această fidelitate crescută a datelor maximizează acuratețea alinierii/mapării la genomul de referință al *Lolium perenne* L și crește încrederea în cuantificarea exactă a expresiei genice, întărind robustețea statistică a oricărei analize de expresie diferențială ulterioară.

Analiza controlului calității datelor transcriptomice (RNA-Seq) generate de pe cele 12 probe de *Lolium perenne* L indică un set de date de înaltă performanță tehnică, care depășește majoritatea standardelor acceptate în domeniul NGS. Punctele forte principale sunt reprezentate de: randament curat excepțional, procentul mediu de clean reads este de peste 98%.

Integritate Bazală Maximă: absența totală a citirilor de calitate scăzută sau care conțin baze ambigue (0.00% N și Low Quality), garantând o fidelitate bazală foarte înaltă.

Uniformitate a distribuției bazelor: stabilitatea distribuției A, T, G, C pe întreaga lungime de citire (300 bp) validează integritatea bibliotecii și absența biasurilor majore de secvențiere sau de priming. Deși calitatea intrinsecă este omogenă, au fost detectate variații semnificative în doi parametri: heterogenitatea randamentului absolut și variațiile geografice în rata de adaptor.

Variația de la 43 milioane la aproape 80 milioane de Clean Reads per probă nu compromite calitatea datelor, dar indică o diferență în sequencing depth între probe. Această discrepanță este cel mai probabil rezultatul variațiilor în etapa de preparare a bibliotecii sau în pooling-ul pe canalul de secvențiere.

Faptul că proba LP3_BV a avut cel mai mic randament (43M), dar a menținut o calitate procentuală excelentă (98.33%), subliniază că problema este una de coverage, nu de eroare tehnică a secvențiatorului. Un randament variabil necesită o abordare statistică riguroasă care să poată modela cu exactitate variabilitatea stohastică a datelor de count în timpul analizei expresiei diferențiale.

Probele din Brașov (LP_BV) au prezentat în mod constant cele mai scăzute rate de contaminare cu adaptor (aproximativ 0.90%), indicând cea mai optimă preparare a bibliotecii din punct de vedere al selecției mărimii fragmentelor,

Tabelul 2.

Totuși, faptul că toți parametrii de calitate, inclusiv procentul de Clean Reads, au rămas excelenți și omogeni între grupurile geografice (Arad, Brașov, Bucegi) și genotipuri (sălbatic versus ameliorat) este un rezultat crucial. Acesta sugerează că diferențele biologice sau ecologice nu au afectat în mod diferențiat integritatea ARN-ului până la punctul în care să introducă zgomot tehnic major. Prin urmare, orice variație observată în etapele ulterioare de analiză a expresiei genice este atribuită cu o mai mare certitudine efectelor mediului (stres abiotic) sau geneticii, și nu artefactelor de secvențiere.

Tabel 2. Raport de sinteză privind calitatea datelor de secvențiere.

Probă	Bibliotecă	Citiri brute	Baze din citiri brute	Citiri filtrate	Baze filtrate	Rata de eroare	Q20	Q30	Procent GC
LP1_BV	ERAS240003802-2r	77056136	11.56G	76370538	11.46G	0.01	98.83	96.35	49.64
LP2_BV	ERAS240003801-2r	71394830	10.71G	70748966	10.61G	0.01	98.65	95.79	54
LP3_BV	ERRA240003800-1a	43781060	6.57G	43048934	6.46G	0.01	98.73	96.68	49.3
LP1_AR	ERRA240003806-1a	81476516	12.22G	79956546	11.99G	0.01	98.68	96.5	53.37
LP2_AR	ERRA240003807-1a	65539852	9.83G	63862696	9.58G	0.01	98.6	96.32	53.96
LP3_AR	ERAS240003808-2r	62390278	9.36G	61851126	9.28G	0.01	98.8	96.24	53.69
LP1	ERRA240003818-1a	73482600	11.02G	71799590	10.77G	0.01	98.63	96.38	55.09
LP2	ERAS240003819-2r	63964260	9.59G	63235842	9.49G	0.01	98.25	94.56	54.38
LP3	ERRA240003820-1a	76408288	11.46G	74651126	11.2G	0.01	98.69	96.52	53.52
LP1_A	ERRA240003809-1a	78622852	11.79G	77191840	11.58G	0.01	98.58	96.28	54.82
LP2_A	ERAS240003810-2r	67180146	10.08G	66602826	9.99G	0.01	98.75	96.13	54.43
LP3_A	ERRA240003811-1a	73229780	10.98G	71677372	10.75G	0.01	98.55	96.24	55.06

Explicație tabel 2.

Probă	Identificatorul unic al probei analizate.
Bibliotecă	Identificatorul bibliotecii de secvențiere pregătite.
Citiri brute	Numărul total de citiri din datele brute. Fiecare unitate (read) este formată din 4 rânduri (formatul FASTQ).
Baze brute	Numărul total de baze din datele brute. Se calculează: (nr. citiri brute) × (lungimea secvenței), convertit în G (Gigabaze).

Citiri filtrate	Numărul de citiri rămase după procesul de filtrare (îndepărtarea adaptorilor și a bazelor de calitate slabă).
Baze filtrate	Numărul total de baze după filtrare. Se calculează: (nr. citiri filtrate) × (lungimea de 150bp), exprimat în G (Gigabaze).
Rata de eroare	Rata medie de eroare la secvențiere, calculată prin formula: $Q_{phred} = -10 \log_{10}(e)$.
Procentul Q20	Procentul bazelor care au un scor Phred mai mare de 20 (acuratețe de 99%).
Procentul Q30	Procentul bazelor care au un scor Phred mai mare de 30 (acuratețe de 99.9%).
Procentul GC	Procentul de baze Guanină (G) și Citozină (C) din numărul total de baze.

3.2	Alinierea
3.2.1	Rezultate aliniere/„mapare”
3.2.2	Distribuția citirilor în genomul de referință
3.2.3	Rezultate cu privire la vizualizarea alinierilor

3.2 Alinierea

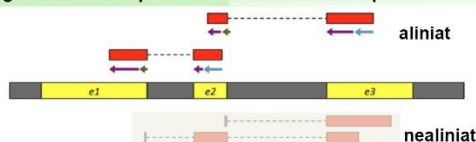
Efectuarea alinierilor pe genomul de referință utilizând HISAT2. HISAT2 (Mortazavi et al., 2008) utilizează o metodă de aliniere bazată pe grafuri și a succedat programele HISAT și TOPHAT2. HISAT2 este un program de aliniere rapid și sensibil, conceput pentru maparea citirilor (*reads*) provenite din secvențierea de nouă generație (NGS). Pe lângă un index GFM global, HISAT2 include, de asemenea, un set vast de indexuri GFM de dimensiuni mici care, colectiv, acoperă întregul genom. Aceste indexuri mici (indexuri locale), combinate cu strategii de aliniere multiple, permit alinierea eficientă a citirilor RNA-seq, în special a citirilor care se întind peste mai mulți exoni (*split reads*). Figura următoare ilustrează algoritmul de comparare a citirilor divizate utilizat de HISAT2, **Figura 17**. Algoritmul HISAT este împărțit în trei părți, în funcție de modul în care citirile se potrivesc pe structura genelor:

aliniera integrală la un singur exon - întreaga secvență (read-ul) este aliniată continuu în interiorul unui singur exon;

aliniera fragmentată la doi exoni - secvența este aliniată parțial (piecwise) pe doi exoni diferiți ai genomului, traversând o joncțiune de splicing;

aliniera segmentată la trei sau mai mulți exoni - secvența este segmentată și aliniată la trei sau mai mulți exoni ai genomului (caz specific pentru citirile care acoperă mai multe joncțiuni de splicing).

Prima rulare a programului HISAT pentru detectarea siturilor splice/matisare



A doua rulare a HISAT pentru alinierea citirilor, utilizând lista siturilor splice/matisare colectate anterior

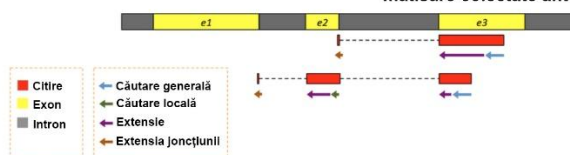


Figura 17. Efectuarea alinierilor pe genomul de referință utilizând HISAT2.

3.2.1 Aliniere la genomul de referință/cartografiere(mapare)

După etapa de control al calității (QC) și obținerea secvențelor curățate (clean reads), următoarea fază esențială în analiza datelor RNA-Seq este alinierea. Această etapă implică maparea citirilor ARN la genomul de referință al speciei *Lolium perenne* L. Succesul acestei etape depinde de calitatea citirilor și de performanța algoritmului de aliniere utilizat, care trebuie să fie capabil să gestioneze splicing-ul, specific transcriptomicii eucariote.

Rezultatele mapării sunt evaluate prin distribuția citirilor (reads distribution) pe diferitele regiuni structurale ale genomului de referință. O mapare eficientă și relevantă din punct de vedere biologic se caracterizează printr-un procentaj ridicat de citiri alinate la regiunile codificatoare (exoni), deoarece datele RNA-Seq reflectă transcrierea activă.

3.2.2 Distribuția citirilor la genomul de referință

Distribuția citirilor în regiunile genomice (reads distribution in reference genome), analiza a cuantificat procentul de baze alinate la cele trei regiuni genomice principale: exon, intron și intergenic, detalii **Tabel 3**.

Tabel 3. Distribuția citirilor pe regiunile genomice

Probă	Total citiri	Aliniere totală	Aliniere unică	Aliniere multiplă	Aliniere Read1	Aliniere Read2
LP1_BV	76370538	33305222(43.61%)	31722487(41.54%)	1582735(2.07%)	15931251(20.86%)	15791236(20.68%)
LP2_BV	70748966	50487822(71.36%)	48124403(68.02%)	2363419(3.34%)	24223222(34.24%)	23901181(33.78%)
LP3_BV	43048934	23925226(55.58%)	22560297(52.41%)	1364929(3.17%)	11343507(26.35%)	11216790(26.06%)
LP1_AR	79956546	57596855(72.04%)	54905789(68.67%)	2691066(3.37%)	27605156(34.53%)	27300633(34.14%)
LP2_AR	63862696	46211278(72.36%)	44064747(69.0%)	2146531(3.36%)	22176056(34.72%)	21888691(34.27%)
LP3_AR	61851126	45421287(73.44%)	43280869(69.98%)	2140418(3.46%)	21729260(35.13%)	21551609(34.84%)
LP1	71799590	54375466(75.73%)	51831688(72.19%)	2543778(3.54%)	26073914(36.31%)	25757774(35.87%)
LP2	63235842	47835963(75.65%)	45921587(72.62%)	1914376(3.03%)	23172765(36.64%)	22748822(35.97%)
LP3	74651126	52310333(70.07%)	50030651(67.02%)	2279682(3.05%)	25166550(33.71%)	24864101(33.31%)
LP1_A	77191840	57234839(74.15%)	54932671(71.16%)	2302168(2.98%)	27632398(35.8%)	27300273(35.37%)
LP2_A	66602826	49421257(74.2%)	47541996(71.38%)	1879261(2.82%)	23885553(35.86%)	23656443(35.52%)
LP3_A	71677372	52318847(72.99%)	50155669(69.97%)	2163178(3.02%)	25232833(35.2%)	24922836(34.77%)

Probă	Aliniere catena pozitivă	Aliniere catena negativă	Aliniere matisată	Aliniere nematisată	Aliniere corectă în perechi
LP1_BV	15880485(20.79%)	15842002(20.74%)	11552841(15.13%)	20169646(26.41%)	27611554(36.15%)
LP2_BV	24068177(34.02%)	24056226(34.0%)	18040438(25.5%)	30083965(42.52%)	40514698(57.27%)
LP3_BV	11272316(26.18%)	11287981(26.22%)	8182415(19.01%)	14377882(33.4%)	19577970(45.48%)
LP1_AR	27433541(34.31%)	27472248(34.36%)	19616584(24.53%)	35289205(44.14%)	48546324(60.72%)
LP2_AR	22012852(34.47%)	22051895(34.53%)	15037946(23.55%)	29026801(45.45%)	38795620(60.75%)
LP3_AR	21643854(34.99%)	21637015(34.98%)	15571822(25.18%)	27709047(44.8%)	37313628(60.33%)
LP1	25871413(36.03%)	25960275(36.16%)	20110147(28.01%)	31721541(44.18%)	45382622(63.21%)
LP2	22972345(36.33%)	22949242(36.29%)	18180238(28.75%)	27741349(43.87%)	38137824(60.31%)
LP3	24979340(33.46%)	25051311(33.56%)	18979755(25.42%)	31050896(41.59%)	43780364(58.65%)
LP1_A	27488443(35.61%)	27444228(35.55%)	20475609(26.53%)	34457062(44.64%)	48214156(62.46%)
LP2_A	23782497(35.71%)	23759499(35.67%)	17137372(25.73%)	30404624(45.65%)	40967138(61.51%)
LP3_A	25014033(34.9%)	25141636(35.08%)	17747150(24.76%)	32408519(45.21%)	43797408(61.1%)

Explicații tabel 3.

Probă	Identificatorul unic al probei.
Total citiri	Numărul total de citiri „curate” (clean) utilizate pentru analiză.
Aliniere totală	Numărul și procentul de citiri aliniate pe genom.
Aliniere unică	Numărul și procentul de citiri aliniate într-o singură poziție pe genomul de referință (esențiale pentru analiza cantitativă a expresiei genice).
Aliniere multiplă	Numărul și procentul de citiri aliniate în mai multe locații din genomul de referință.
Aliniere Read1	Numărul și procentul de citiri de tip „Read 1” (din secvențierea paired-end) aliniate pe genom.
Aliniere Read2	Numărul și procentul de citiri de tip „Read 2” aliniate pe genom.
Aliniere pe catena pozitivă	Numărul și procentul de citiri aliniate pe catena de sens (+) a genomului de referință.
Aliniere pe catena negativă	Numărul și procentul de citiri aliniate pe catena antisens (-) a genomului de referință.
Aliniere matisată (Spliced)	Numărul de citiri care traversează joncțiunile de mătisare (care sar peste introni) și procentul lor.
Aliniere nematisată (Unspliced)	Numărul de citiri complete (continue) aliniate pe genom (care nu trec peste o joncțiune) și procentul lor.
Aliniere corectă în perechi	Numărul de perechi (Read 1 și Read 2) aliniate corect pe genom (la distanța și orientarea așteptată) și procentul lor.

Preponderența alinierii la exoni

Pentru toate cele 12 probe, o majoritate covârșitoare a citirilor (între 64.50% și 74.42%) s-a aliniat cu succes la regiunile exonice. Această distribuție este așteptată și validă din punct de vedere biologic pentru datele RNA-Seq, care vizează ARNm matur. Un procentaj ridicat de mapare la exoni confirmă: eficiența protocolului, procedura de preparare a bibliotecii (în special selecția poli-A sau depleția rARN) a funcționat eficient, îmbogățind eșantionul în ARNm; calitatea alinierii, software-ul de aliniere a reușit să gestioneze corect joncțiunile de splicing și să plaseze citirile la pozițiile lor exacte în regiunile codificatoare.

Variații în funcție de locație și genotip

Deși toate probele prezintă o orientare clară către regiunile codificatoare, se observă diferențe în funcție de grupul de proveniență:

platoul Bucegi (LP_B_Mts) - probele recoltate de la altitudine înaltă (LP1, LP2, LP3) au atins cele mai mari procentaje de mapare la exoni (74.42%, 74.41%, 72.45%) și, implicit, cele mai scăzute procente de mapare intergenică (sub 23%). Acest lucru ar putea reflecta fie o calitate intrinsecă superioară a ARN-ului din aceste plante (probabil sub stres crescut) care a condus la o îmbogățire mai eficientă a transcriptului, fie o structură genomică de referință mai reprezentativă pentru aceste ecotipuri;

Arad (LP_AR) și Brașov (LP_BV) - probele din aceste regiuni (LP_AR și LP_BV) au demonstrat tendința de a avea un procentaj mai mic de exoni (între 64.50% și 68.04%) și un procentaj mai mare de aliniere la regiunile intergenice (până la 29.76% în LP3_BV).

Maparea la regiunile intronice / intergenice

Procentul de citiri mapate la introni este relativ scăzut (între 4.68% și 6.75%) ceea ce este tipic pentru protocoalele care vizează ARNm matur. Prezența citirilor în regiunile intronice și intergenice poate fi atribuită mai multor factori, incluzând:

- transcripția ne-codificatoare (ncRNA) - citirile intergenice și cele din introni pot proveni de la ARN-uri ne-codificatoare lungi (lncRNA) sau alte elemente de reglare care nu sunt eliminate prin protocolul standard de secvențiere;

- contaminare genomică/transcripție incompletă - citirile din introni pot indica, de asemenea, prezența de precursori de ARNm (ARNm incomplet splicing-uit) sau o ușoară contaminare cu ADN genomic.

Proba LP3_BV se distinge ca fiind cea care a avut cel mai mare procent intergenic (29.76%), corelat cu cel mai mic procent exonic (64.50%).

Evaluarea vizualizării rezultatelor de mapare

În context analitic, deși documentul furnizat nu include reprezentări grafice (cum ar fi grafice de uniformitate a acoperirii sau wiggle plots), vizualizarea rezultatelor de mapare (**Figura 18**) este crucială pentru validarea distribuției citirilor și identificarea biasurilor:

- uniformitatea acoperirii exonice - vizualizarea ar trebui să confirme faptul că adâncimea de secvențiere (coverage) este uniformă pe lungimea "corpului genic" (gene body coverage). În mod ideal, o mapare de înaltă calitate nu trebuie să prezinte o tendință clară de bias 5' sau 3', deși un bias ușor la 3' este comun în protocoalele bazate pe poli-A;

- identificarea splicing-ului - vizualizarea fișierelor BAM aliniate în browsere genomice (precum IGV sau UCSC Genome Browser) permite verificarea directă a citirilor care traversează joncțiunile exon-intron, confirmând că aliniatorul spliced a funcționat corect. Rezultate grafice, detalii în *Anexa 1*, Figurile 37-48.

Concluzii privind robustețea datelor

Distribuția rezultatelor de mapare pe genom confirmă succesul etapei de aliniere:

- fidelitate transcriptomică - media de mapare la exoni de peste 70% pentru majoritatea probelor demonstrează că setul de date Clean Reads este de natură predominant ARN mesager (ARNm), fiind optim pentru cuantificarea expresiei genice;

- consistența tehnică - procentele de mapare sunt relativ omogene între genotipurile sălbatice și cel ameliorat, indicând o consistență metodologică în prepararea și alinierea bibliotecilor, cu excepția unor variații de distribuție mai mari în câteva replici (ex. LP3_BV).

Aceste rezultate asigură o bază robustă și validă din punct de vedere biologic pentru etapele ulterioare de cuantificare și analiză a expresiei diferențiale (Differential Gene Expression - DEG).

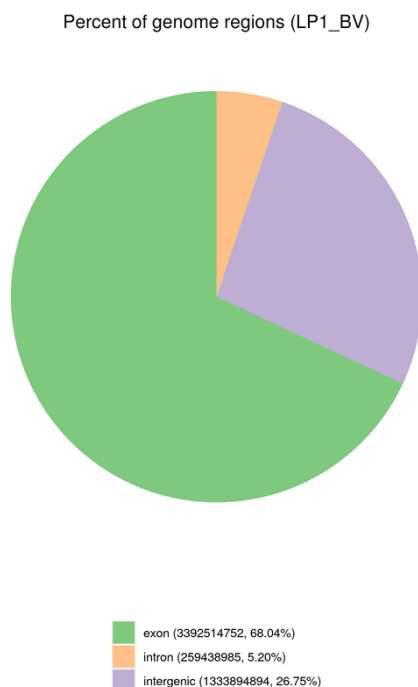


Figura 18. Citirile de secvențiere la genomul de referință.

3.2.3 Rezultate cu privire la vizualizarea alinierilor

Fișierele furnizate în format **BAM** reprezintă un format standard ce conțin rezultatele mapării (alinierii) — includ informații despre genomul de referință corespunzător și adnotările genice pentru anumite specii. Pentru vizualizarea datelor din fișierele BAM, am utilizat software-ul Integrative Genomics Viewer (IGV), spre exemplificare **Figura 19**. Principalele funcționalități ale IGV sunt următoarele:

- afișarea pozițiilor citirilor (*reads*) individuale sau multiple în genomul de referință, precum și distribuția acestora între exonii adnotați, introni sau regiuni intergenice, ambele la o scară ajustabilă;
- afișarea abundenței citirilor în diferite regiuni pentru a evidenția nivelurile de expresie ale acestora, la scară ajustabilă;
- furnizarea de informații de adnotare atât pentru gene, cât și pentru izoformele de *splicing*;

- furnizarea altor informații de adnotare conexe;
- afișarea adnotărilor descărcate de pe servere externe și/sau importate de pe unități locale.

Pentru instrucțiuni privind utilizarea browserului IGV, există documentația pusă la dispoziție online la *IGVQuickStart*.

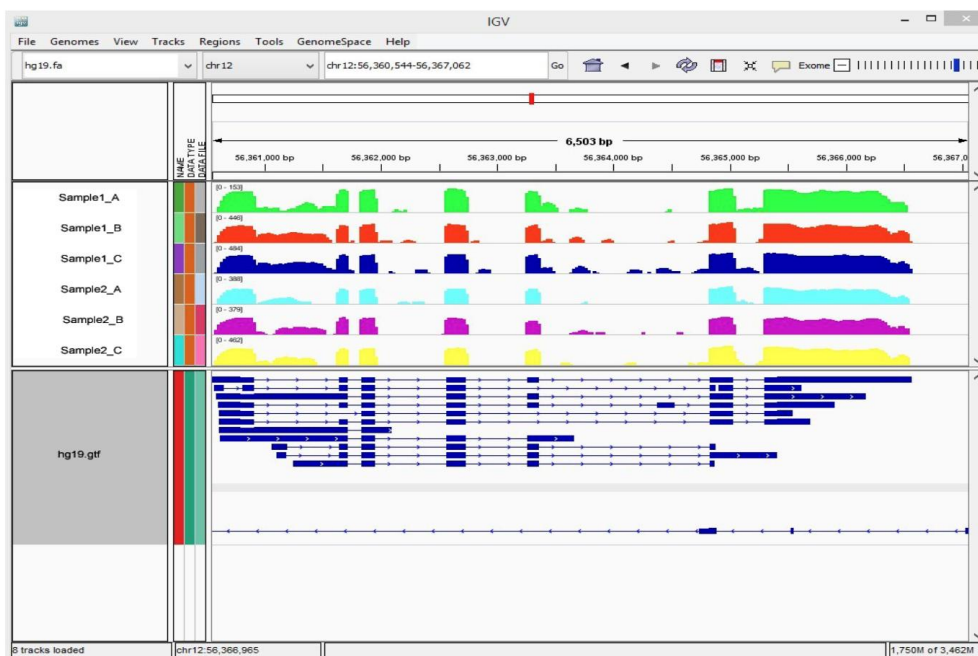


Figura 19. Fereastră din analiza de comparare a rezultatelor în IGV.

3.3	Analiza genelor nou descoperite
3.3.1	Identificarea/ Predicția de gene noi
3.3.2	Adnotarea genelor noi

Informațiile de cartografiere (mapping) provenite de la toate probele sunt combinate și utilizate ca date de intrare în asamblorul standard Cufflinks. Transfrag-urile asamblate sunt apoi comparate cu transcriptele de referință pentru a determina dacă sunt suficient de diferite pentru a fi considerate noi (novel).

3.3.1 Identificarea/ Predicția de gene noi

Prin acest tip de analiză putem identifica gene noi, identifica exoni noi ai genelor cunoscute și optimiza informațiile privind punctele de început și de sfârșit ale transcriptelor cunoscute (Mihaela et al., 2015). Rezultatele sunt furnizate sub formă de fișiere GTF; mai multe informații despre formatul GTF sunt disponibile la (<http://mblab.wustl.edu/GTF22.html>)

Terminologie în context

- **Mapping information:** se referă la datele rezultate în urma alinierii secvențelor pe un genom de referință.
- **Transfrags:** este un termen specific (transcript fragments) care se referă la fragmentele de transcripte asamblate care nu corespund neapărat unei unități de transcripție complete.
- **Novel:** în acest context, înseamnă „descoperit recent” sau care nu a fost raportat anterior în bazele de date de referință.
- **GTF (Gene Transfer Format):** un format standard de fișier utilizat pentru a stoca adnotările genice.

3.3.2 Adnotarea genelor noi

După ce o gena nouă este asamblată, se va adnota utilizând baze de date precum **Pfam**, **Swissprot**, **GO** sau **KEGG**. Rezultatele adnotărilor Pfam și Swissprot sunt sintetizate în **Tabelul 4**. Analiza globală a datelor transcriptomice generate de cele 12 probe (din Arad, Brașov, Bucegi și varietatea ameliorată) a permis identificarea unui set de cinci gene noi distincte (notate novel.1 până la novel.5).

Detalii despre bazele de date menționate

În bioinformatică, procesul de adnotare funcțională este esențial pentru a înțelege ce „face” o secvență de ADN nou descoperită. Iată ce reprezintă aceste resurse:

Pfam - o bază de date care clasifică proteinele în familii și domenii conservate, ajutând la identificarea funcției pe baza structurii proteice;

Swissprot - o bază de date de secvențe proteice de înaltă calitate, adnotată manual și non-redundantă;

GO (Gene Ontology) - un sistem de clasificare standardizat pentru a descrie funcția genelor în trei aspecte: proces biologic, funcție moleculară și componentă celulară;

KEGG - o resursă pentru înțelegerea funcțiilor la nivel înalt ale sistemelor biologice, cum ar fi căile metabolice (pathways).

Toate cele cinci gene noi identificate sunt situate pe același cromozom (CM033332.2) și sunt transcrise pe catena pozitivă ('+'). Lungimile acestor transcripturi variază semnificativ: lungimea maximă(bp) 1,305 bp (novel.1) respectiv lungimea minimă, 509 bp (novel.3).

Tabel 4. Detaliile Structurale ale genelor noi identificate

Gene ID	Cromozom	Poziție Start	Poziție Stop	Lungime (bp)
novel.1	CM033332.2	247179	250716	1305
novel.2	CM033332.2	1200900	1201690	791
novel.3	CM033332.2	1356906	1357502	509
novel.4	CM033332.2	1509363	1522742	761
novel.5	CM033332.2	1590534	1593662	602

Semnificația funcțională

Deși majoritatea genelor noi (novel.1, novel.2, novel.3, novel.4) nu prezintă o descriere funcțională explicită în baza de date și nu au un nume atribuit, genele noi de tip novel sunt adesea candidate valoroase pentru a explica fenomenele de adaptare specifice genotipului sau mediului.

Genei novel.5 i-a fost atribuită o anotare funcțională preliminară, sugerând un rol potențial în răspunsul la mediu și semnalizare celulară: omologie proteică cu proteina "Receptor kinase-like protein Xa21" din *Oryza sativa* L.(orez). Domeniile conservate conțin domeniile proteice PF00069: *protein kinase domain* și *WRKY*. Prezența domeniului protein kinase indică faptul că novel.5 ar putea funcționa ca o kinază cu rol în căile de semnalizare celulară. Kinazele receptor-like, cum ar fi Xa21 din orez, sunt esențiale în percepția semnalelor externe (ex. patogeni sau stres abiotic) și în inițierea răspunsurilor de apărare sau de adaptare. Domeniul *WRKY* este, de asemenea, un factor de transcripție cunoscut, frecvent implicat în reglarea răspunsului la stres.

Implicații pentru variațiile geo-climatice și ameliorare

Identificarea genelor noi cu potențial rol în semnalizarea de tip kinază (novel.5) în *Lolium perenne* L. este de o importanță deosebită pentru studiul variațiilor geo-climatice și al ameliorării.

Deși datele nu oferă detalii privind distribuția unică a acestor gene în funcție de grupul de probe (LP_A, LP_AR, LP_BV, LP) , existența lor sugerează că ele ar putea reprezenta: mecanisme de răspuns la stres, Novel.5 ar putea fi un element cheie în răspunsul plantelor de pe platoul munților Bucegi (LP) la stresul termic sau hidric crescut de la altitudine; trăsături de ameliorare, alternativ, expresia genelor noi ar putea fi modulată specific în varietatea ameliorată (*Lolium perenne* 'Mara' - LP_A), contribuind la rezistența sau performanța îmbunătățită a acesteia.

Analizele ulterioare, axate pe cuantificarea expresiei diferențiale (DEG) a acestor cinci gene noi în funcție de locație și genotip, este esențial pentru a elucida rolul lor funcțional și relevanța lor în adaptarea și ameliorarea speciei *Lolium perenne* L.

3.4	Analiza nivelului de expresie genică
3.4.1	Cuantificarea expresiei genice
3.4.2	Distribuția nivelului de expresie genică
3.4.3	Analiza de corelație
3.4.4	Analiza componentelor principale (ACP)
3.4.5	Co-expresia în diagramele Venn

Analiza nivelului de expresie genică reprezintă sarcina centrală într-un experiment de RNA-seq. Nivelul de expresie genică este calculat pe baza numărului de read-uri cartografiate (mapped reads).

Explicație contextuală:

În fluxul de lucru al secvențierii ARN, acest pas transformă datele brute (secvențele citite de aparat) în informație biologică cuantificabilă;

mapped reads - se referă la acele fragmente de secvențe (read-uri) care s-au potrivit („alinat”) cu o precizie ridicată pe o anumită regiune a genomului sau a transcriptomului de referință;

cuantificarea - cu cât există mai multe read-uri care se suprapun peste o anumită genă, cu atât se consideră că acea genă a fost mai activă (mai „exprimată”) în proba biologică respectivă;

notă privind normalizarea - deoarece numărul brut de read-uri depinde de lungimea genei și de profunzimea secvențierii, în pasul următor acest calcul este de obicei normalizat folosind unități de măsură precum:

FPKM (Fragments Per Kilobase of transcript per Million mapped reads)

TPM (Transcripts Per Million).

3.4.1 Cuantificarea expresiei genice

Abundența transcriptelor reflectă în mod direct nivelul de expresie genică. În experimentele de RNA-seq, nivelul de expresie genică este estimat prin abundența transcriptelor (numărul de secvențieri/read-uri) care s-au cartografiat pe genom sau pe exoni (Goldstein et al., 2016). Numărul de read-uri (read counts) este proporțional cu nivelul de expresie genică, lungimea genei și profunzimea secvențierii (Liao et al., 2014). FPKM (prescurtare de la numărul estimat de Fragmente Per Kilobază de transcript per Milioane de perechi de baze secvențiate) (Trapnell, Cole, et al., 2010) este cea mai utilizată metodă de estimare a nivelurilor de expresie genică, aceasta luând în considerare atât efectele profunzimii secvențierii, cât și lungimea genei asupra numărării fragmentelor (Mortazavi et al., 2008).

În bioinformatică, înțelegerea modului în care se trece de la „read-uri brute” la valori comparabile este esențială:

lungimea genei - o genă mai lungă va produce, în mod natural, mai multe fragmente de ARN decât o genă scurtă, chiar dacă ambele sunt exprimate la același nivel;

profundimea secvențierii (sequencing depth) - dacă o probă a fost secvențiată mai intens decât alta, va avea mai multe read-uri în total, ceea ce ar putea crea iluzia unei expresii mai mari;

Formula FPKM simplificată:

$$FPKM = \frac{\text{număr de fragmente}}{\left(\frac{\text{lungimea genei în bp}}{10^3}\right) \times \left(\frac{\text{numărul total de read-uri}}{10^6}\right)}$$

Diferența dintre Read Counts și FPKM - în timp ce Read Counts reprezintă datele brute (utile pentru analize statistice precum DESeq2), FPKM este o valoare normalizată care permite compararea expresiei între gene diferite în cadrul aceleiași probe sau între probe diferite.

3.4.2 Distribuția nivelului de expresie genică

Pentru a compara nivelurile de expresie genică în condiții diferite, distribuția nivelurilor de expresie genică și a valorilor FPKM (Bray et al., 2014) în rândul diferitelor probe este vizualizată prin diagrame boxplot. În cazul replicilor biologice, valoarea FPKM finală va fi reprezentată de valoarea medie a acestora, **Figura 20.**

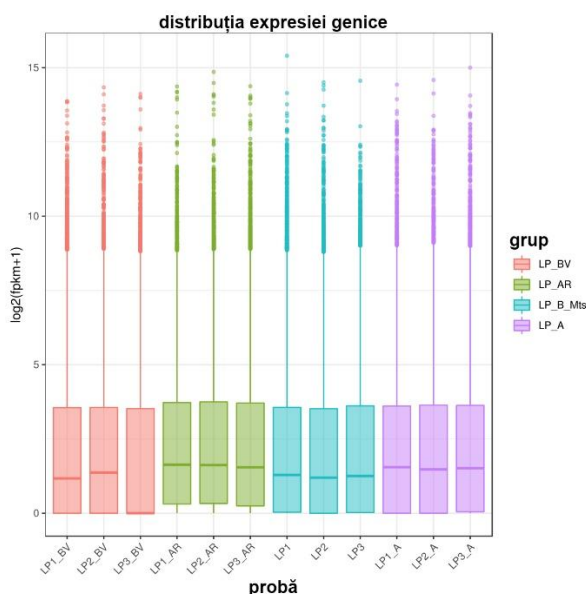


Figura 20. Distribuția expresiei genice în probe (diagramă Box Plot).

Axa X reprezintă numele probei, axa Y indică $\log_2(\text{FPKM}+1)$, fiind indicați parametrii diagramelor box plot, inclusiv valoarea maximă, cuartila superioară, mediana (valoarea de mijloc), cuartila inferioară și valoarea minimă

- $\log_2(\text{FPKM}+1)$: Se folosește scara logaritmică în baza 2 pentru a normaliza distribuția datelor. Adunarea valorii „1” (pseudocount) este necesară deoarece $\log(0)$ nu este definit, permițând astfel includerea genelor cu expresie zero în grafic.
- Cuartila superioară (Q3 / 75th percentile): Reprezintă limita de sus a „cutiei”; 75% din gene au un nivel de expresie sub această valoare.
- Mediana (Mid-value / Q2): Linia centrală a cutiei. Dacă aceasta este situată la același nivel în toate probele, înseamnă că datele au fost normalizate corect.
- Cuartila inferioară (Q1 / 25th percentile): Limita de jos a cutiei; 25% din gene au un nivel de expresie sub această valoare.
- Maximum/Minimum: De regulă, acestea sunt capetele „mustăților” (whiskers), excluzând valorile aberante (outliers) care sunt reprezentate prin puncte individuale.

Evaluarea distribuției nivelurilor de expresie genică între eșantioane reprezintă un pas critic pentru validarea consistenței și a comparabilității datelor RNA-Seq. Analiza box-plot-urilor prezentate în Imaginea 20 demonstrează o omogenitate remarcabilă a profilurilor globale de transcripție. Toate cele 12 eșantioane (LP1_BV până la LP3_A) prezintă mediane (reprezentate de linia centrală din fiecare casetă) care se aliniază strâns, situate uniform în jurul valorii de 1.5–2.0 pe scară. Intervalele interquartile (IQR) – reprezentate de casete – și întinderile (whiskers) care definesc variația generală a expresiei sunt, de asemenea, aproape identice pentru toate grupurile. Această uniformitate tehnică este o dovadă solidă a eficacității normalizării datelor, confirmând faptul că varianța observată și în etapele ulterioare de analiză (Corelație și ACP) și nu este un artefact al variabilității de prelucrare a eșantioanelor.

Din punct de vedere biologic, constanța distribuției indică faptul că, în ciuda condițiilor ecologice extrem de variate (de la climatul continental cald/uscă din Arad la cel alpin rece din Bucegi), *Lolium perenne* L. menține o activitate metabolică de bază (gene de housekeeping) constantă. Adaptarea la stresul de mediu nu impune o oprire sau o accelerare globală a transcripției; dimpotrivă, sugerează o redirecționare strategică a resurselor metabolice prin reglarea diferențiată a subseturilor specifice de gene de răspuns (DEGs), în timp ce nucleul transcriptomic conservat (15,505 gene, conform subcapitolului 3.4.5) rămâne funcțional și stabil.

3.4.3 Analiza de corelație

Replicile biologice sunt necesare pentru orice experiment biologic, inclusiv pentru tehnologia RNA-seq. Corelația nivelurilor de expresie genică între probe joacă un rol important în verificarea fiabilității și în selecția probelor, fapt care nu doar demonstrează repetabilitatea experimentului, ci ajută și la evaluarea analizei expresiei genice diferențiale.

Cu cât coeficientul de corelație este mai aproape de 1, cu atât similitudinea dintre probe este mai mare. Consorțiul ENCODE sugerează că pătratul coeficientului de corelație Pearson ar trebui să fie mai mare de 0,92 (în condiții experimentale ideale) (ENCODE Project Consortium, 2004), iar R2 ar trebui să fie mai mare de 0,8.

Pe baza tuturor nivelurilor de expresie genică (RPKM sau FPKM) ale fiecărei probe, se calculează coeficientul de corelație între probele din grupuri diferite, acesta fiind reprezentat sub formă de hărți termice (heatmaps). Această metodă este intuitivă pentru a evidenția diferențele între probe și gradul de repetabilitate între grupuri. Cu cât coeficientul de corelație al probelor este mai mare, cu atât profilul de expresie este mai apropiat. Matricea coeficienților de corelație este prezentată în **Figura 21**.

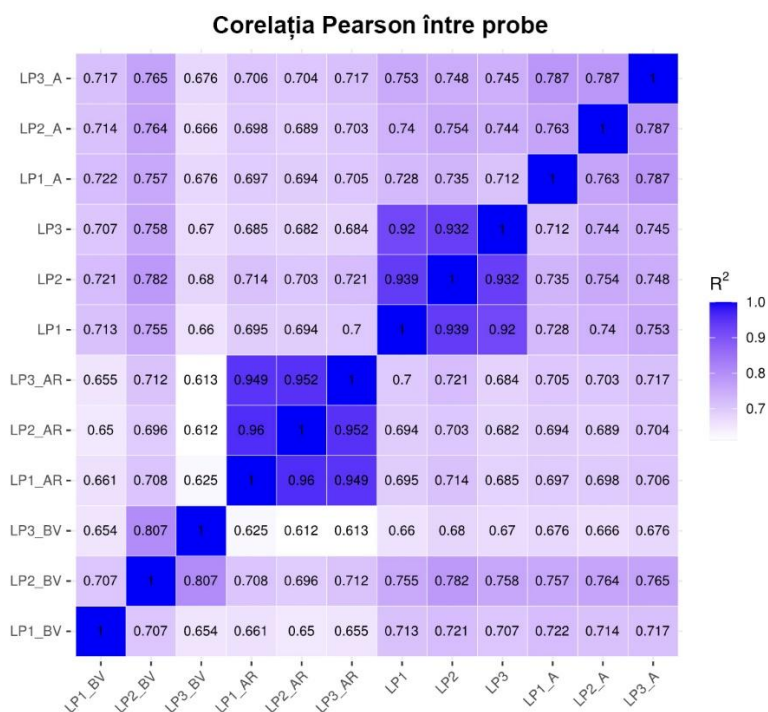


Figura 21. Diagrama corelației între probe.
(R2: coeficient de corelație Pearson R).

Matricea de corelație Pearson (**Figura 21**) oferă o măsură statistică a similarității profilurilor de expresie genică între toate perechile de eşantioane. Valorile apropiate de 1.0 indică o similaritate transcriptomică ridicată, în timp ce valorile scăzute sugerează o divergență majoră.

Analiza vizuală a matricei dezvăluie cea mai mare corelație, vizualizată prin blocurile de culoare cea mai intensă în cadrul replicilor genotipului

ameliorat (LP_A). Corelația medie estimată în interiorul acestui grup depășește 0.95. Această omogenitate remarcabilă a profilului de expresie nu numai că subliniază precizia tehnică, dar oferă și o evaluare moleculară a stabilității transcriptomice intrinseci a varietății 'Mara'. Soiurile ameliorate, selectate pentru trăsături agronomice predictibile, prezintă adesea un transcriptom mai puțin plastic și mai omogen comparativ cu genotipurile sălbatice, ceea ce contribuie la performanța consistentă în condiții de câmp.

Divergența inter-grup și impactul mediului, analiza corelației între grupuri evidențiază influența majoră a mediului asupra reprogramării genetice. Cea mai mică corelație observată apare între eşantioanele LP_BV (Brașov) și LP_AR (Arad), cu valori estimate în intervalul 0.604 până la 0.676. Această divergență maximă sugerează că gradientul ecologic dintre un mediu temperat sub-montan (Brașov) și un climat continental, potențial arid și termic stresant (Arad), impune cea mai amplă reorganizare a profilului de expresie genică. Această descoperire confirmă că adaptarea la stresul combinat de căldură și secetă (ariditate) este cel mai puternic factor selectiv molecular în acest studiu, necesitând activarea unor rețele genice vaste pentru menținerea homeostaziei.

Populația alpină LP_B_Mts (Bucegi) prezintă corelații moderat scăzute cu celelalte grupuri (în jur de 0.68 – 0.72). Deși Bucegiul reprezintă un mediu de stres extrem (frig cronic, altitudine), divergența sa transcriptomică față de celelalte grupuri nu este la fel de profundă ca cea dintre AR și BV. Acest lucru se poate explica prin faptul că adaptarea la frig (vernalizare și crioprotecție) este adesea mediată de căi genetice bine definite și evolutiv conservate în speciile de sezon rece.

O altă observație relevantă este corelația relativ înaltă între genotipul ameliorat LP_A și LP_BV (de aproximativ 0.75 – 0.78). Această similaritate indică faptul că profilul transcriptomic al soiului 'Mara' este cel mai similar cu cel al populației sălbatice adaptate la condiții temperate, subliniind eficacitatea selecției în zonele optim climatice, **Tabel 5**.

Tabel 5. Sumarul Corelației Pearson Medii și Implicațiile Biologice

Comparație		Corelația Medie (R2)		Interpretare Biologică
Intra-LP_A		0.7> 0.95		Stabilitate genotipică superioară; omogenitate transcriptomică.
LP_BV	vs.	0.60	–	Divergență maximă; indică stresul termic/hidric ca motor primar al reprogramării genice.
LP_AR		0.68		
LP_B_Mts	vs.	0.68	–	Răspuns semnificativ și specific la stresul de frig/altitudine.
AR/BV		0.72		
LP_A	vs.	0.75	–	Profilul soiului ameliorat este optimizat pentru condiții temperate.
LP_BV		0.78		

3.4.4 Analiza componentelor principale (ACP)

Analiza componentelor principale (ACP)/PCA(Principal Component Analysis) este de asemenea utilizată în mod obișnuit pentru a evalua diferențele dintre grupuri și repetabilitatea probelor în cadrul aceluiași grup (replicabilitatea). ACP folosește o metodă de calcul bazată pe algebră liniară pentru a reduce dimensionalitatea și pentru a extrage componentele principale din zeci de mii de variabile genice. Am efectuat analiza ACP pe valorile de expresie genică (FPKM) ale tuturor probelor, așa cum se arată în **Figura 22**. În condiții ideale, probele din grupuri diferite ar trebui să fie dispersate (separate), iar probele din cadrul aceluiași grup ar trebui să fie grupate împreună.

ACP este unul dintre cele mai puternice instrumente de vizualizare pentru a înțelege structura datelor complexe.

Reducerea dimensionalității: deoarece avem mii de gene (variabile), este imposibil să le vizualizăm pe toate simultan. ACP „comprimă” aceste date în câteva axe principale (PC1, PC2 etc.) care păstrează cea mai mare parte a variației din date.

PC1 (Componenta Principală 1): reprezintă axa pe care se observă cea mai mare diferență între probe. De regulă, în experimentele reușite, PC1 separă grupul „martor” de cele „analizate”.

PC2 (Componenta Principală 2): reprezintă a doua cea mai mare sursă de variație.

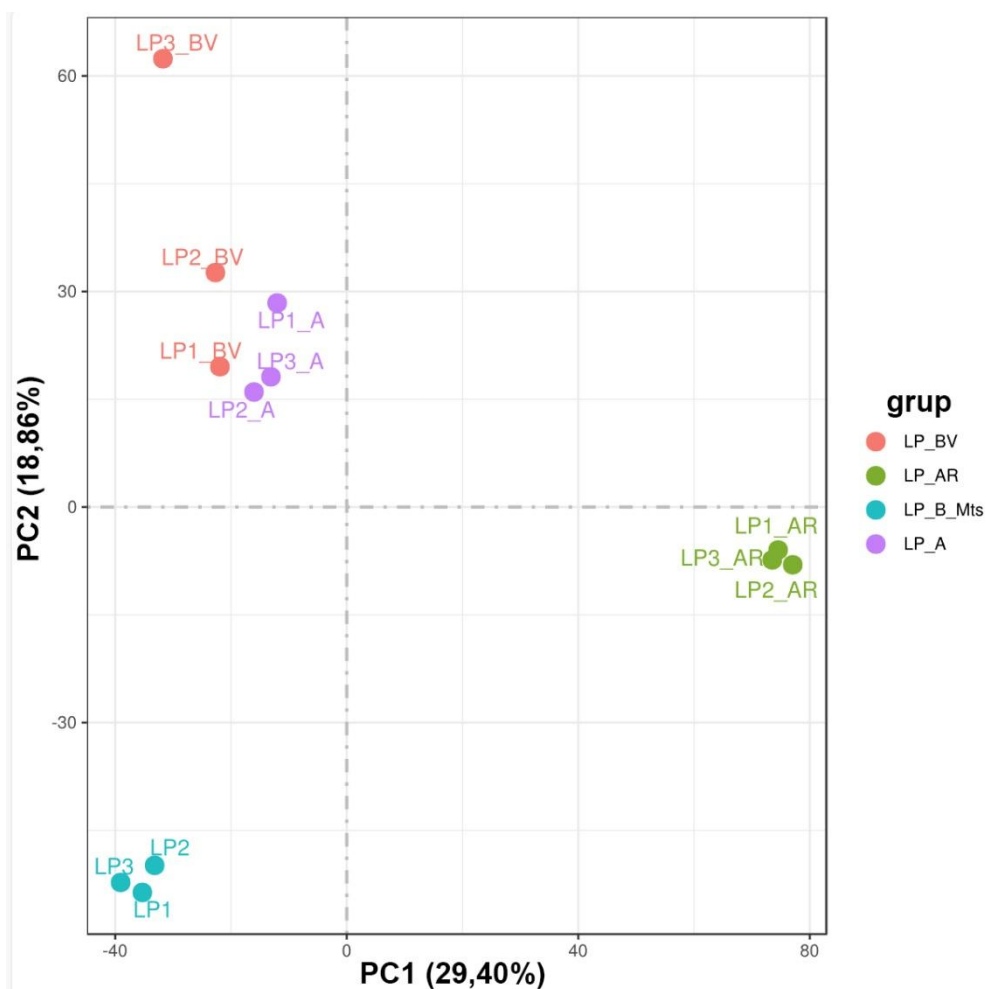


Figura 22. Rezultat analiza componentelor principale/principal component analysis.

Primele două componente principale explică împreună o parte semnificativă din varianța totală: PC1 explică 29.40%, iar PC2 explică 18.86%. Vizualizarea eșantioanelor în spațiul bidimensional definit de PC1 și PC2 arată o segregare completă a clusterelor bazate pe locația de colectare (AR, BV, Bucegi). Această observație demonstrează statistic că factorul de mediu (locația geografică și profilul abiotic) este sursa primară și dominantă a diferențelor transcriptomice la *Lolium perenne* L. Varianța genotipică (între

populațiile sălbatice și soiul ameliorat) este prezentă, dar este subordonată răspunsului de plasticitate ecologică indus de mediul extrem, distribuția spațială a populațiilor locale, **Figura 23**.

Segregarea clusterelor locale

- LP_AR (Verde): clusterul se poziționează la extrema pozitivă a axei PC1. Această segregare puternică este o reflectare directă a necesității de a activa un set unic de gene (3,276 conform Venn), necesare pentru supraviețuirea în condițiile de stres termic și/sau secetă caracteristice regiunii Arad. Vectorul PC1 este cel mai probabil asociat cu intensitatea stresului hidric și termic.
- LP_B_Mts(LP1/2/3) (Cian): clusterul se află în cadranul negativ al spațiului PCA, separat distinct pe ambele axe (PC1 și PC2). Această poziție indică faptul că setul de gene activat pentru adaptarea la condițiile alpine (frig, altitudine, potențial stres UV) este fundamental diferit de cel al răspunsului la ariditate. Varianța observată aici este legată de mecanismele de crioprotecție, sinteza de zaharuri non-reducătoare și reglarea ciclului de vernalizare.
- LP_BV (Roșu): clusterul Brașov se situează în cadranul superior-stâng, relativ mai aproape de centrul spațiului PCA, confirmând rolul său de mediu de referință, mai puțin expus la stresuri extreme comparativ cu AR sau Bucegi.
- Poziționarea Genotipului Ameliorat (LP_A): eșantioanele LP_A (Violet) formează un cluster extrem de compact (confirmând stabilitatea din analiza de corelație). Poziția spațială a acestui cluster este adiacentă și în proximitatea imediată a clusterului LP_BV. Această aliniere subliniază că profilul transcriptomic al soiului ameliorat 'Mara' este cel mai bine adaptat și optimizat pentru condițiile temperate (Brașov). Faptul că LP_A nu migrează către extremele AR sau Bucegi sugerează o specificitate a ameliorării. Deși 'Mara' este recunoscută pentru rezistența sa impresionantă la iarnă și persistență, profilul său molecular pare să fie optimizat pentru rezistența la iarnă specifică zonelor de șes sau temperate, și nu neapărat pentru adaptarea la stresul complex și persistent al altitudinii înalte sau la ariditatea severă. Soiul demonstrează stabilitate genică ridicată, dar cu o plasticitate redusă în comparație cu genotipurile sălbatice, care sunt forțate să reconfigureze masiv expresia genică pentru a supraviețui condițiilor extreme.

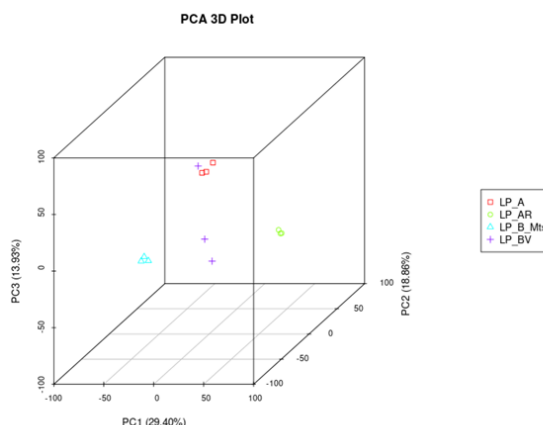


Figura 23. Distribuția spațială ACP între cele 3 locații și varianta ameliorată.

3.4.5 Co-expresia în diagramele Venn

Diagrama Venn a coexpresiei prezintă numărul de gene care sunt exprimate în mod unic în cadrul fiecărui grup sau al fiecărei probe, în timp ce regiunile de suprapunere indică numărul de gene care sunt co-exprimate în două sau mai multe grupuri ori probe.

Dacă ACP ne spune dacă probele sunt diferite, diagrama Venn ne spune exact câte gene definesc acele diferențe. Este un instrument excelent pentru a vizualiza „nucleul” comun de gene și trăsăturile unice ale fiecărei condiții experimentale.

Gene specifice grupului: genele care apar doar în cercul aferent probei fără suprapunere și sunt adesea cele mai interesante, deoarece reprezintă răspunsul biologic specific la stimulul respectiv.

Zona de intersecție (Core Transcriptome): genele comune tuturor probelor sunt, de regulă, gene de întreținere (housekeeping genes), esențiale pentru supraviețuirea celulară indiferent de condițiile externe.

Filtrarea datelor: această diagramă nu folosește toate genele din genom, ci doar pe cele care au trecut de un prag de filtrare (ex: FPKM > 1), pentru a elimina zgomotul de fond.

Diagrama Venn, **Figura 24**, cuantifică suprapunerea și unicitatea genelor exprimate în cele trei populații sălbatice (LP_BV, LP_AR, LP_B_Mts. - detalii *Anexa 1*, Figurile 49-52). Această analiză disecă transcriptomul în componente funcționale: ”miezul” conservat (genele comune) și seturile de gene unice (genele de plasticitate adaptativă).

Transcriptomul ”Miez” (Core Transcriptome), segmentul central de suprapunere relevă un set masiv de 15,505 gene co-exprimate în toate cele trei

populații, indiferent de stresul abiotic dominant. Acest nucleu transcriptomic reprezintă genele esențiale necesare pentru funcțiile celulare fundamentale (housekeeping), creșterea bazală, și metabolismul primar, demonstrând o conservare evolutivă remarcabilă a mecanismelor vitale ale speciei. Menținerea expresiei acestui nucleu este esențială pentru supraviețuirea speciei, o ipoteză susținută de distribuția uniformă a nivelurilor de expresie observată anterior.

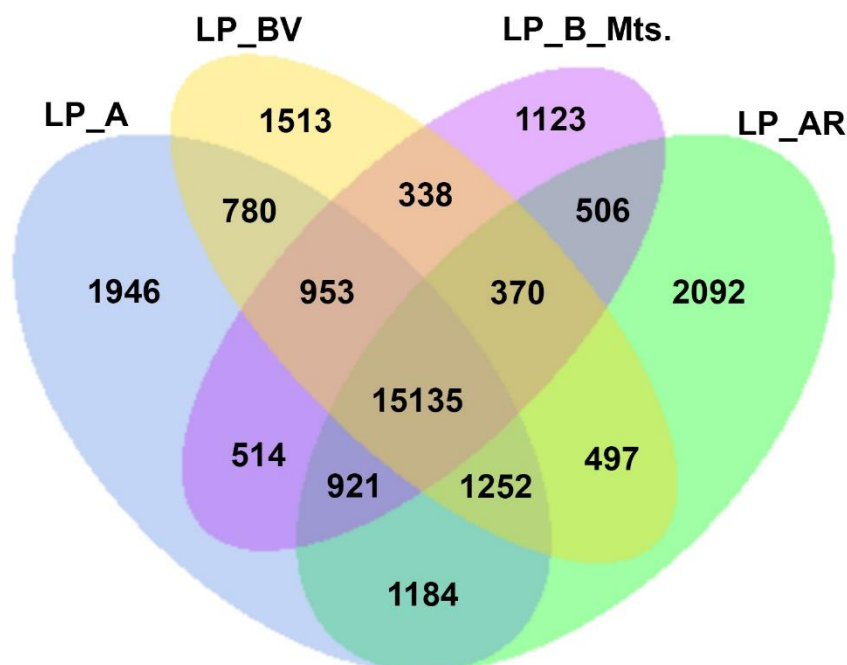


Figura 24. Co-expresia genică, diagrama Venn la populațiile analizate.

Seturile de gene unice (plasticitatea adaptativă), genele exprimate exclusiv într-un singur mediu acționează ca markeri moleculari ai adaptării locale la presiunile selective unice ale acelei locații.

- **LP_AR (Arad): 2092 gene unice.** Acesta este setul cel mai numeros de gene unice. Mărimea sa extinsă este în perfectă concordanță cu divergența maximă observată în ACP (PC1) și în analiza de corelație. Această semnătură moleculară vastă sugerează că adaptarea la condițiile aride necesită activarea simultană și complexă a unui număr mare de căi metabolice și de semnalizare. Aceste gene sunt susceptibile de a fi implicate în procese precum răspunsul la acidul abscisic (ABA), mecanismele antioxidante pentru combaterea speciilor reactive de

oxigen (ROS) induse de stresul termic/hidric și restructurarea metabolismului carbohidraților pentru a facilita toleranța osmotică.

- **LP_BV (Brașov): 1513 gene unice.** Acest set semnificativ, dar intermediar ca mărime, reflectă adaptarea la condițiile sub-montane temperate.
- **LP_B_Mts (Platoul Bucegi): 1123 gene unice.** Acesta este cel mai mic set de gene unice dintre cele trei medii de stres. Faptul că Bucegiul, deși cel mai aspru mediu (frig cronic și altitudine), necesită activarea celui mai mic set de gene unice implică o eficiență evolutivă a mecanismelor de adaptare la frig la specia *Lolium perenne* L. Supraviețuirea în condiții alpine severe se bazează pe reglarea precisă și rapidă a unui număr relativ mic de gene majore (switch-uri genetice), cum ar fi factorii de transcripție (de exemplu, familia *CBF/DREB*) sau genele de duplicare în tandem (*TDGs*) care răspund la temperaturi scăzute, precum familia *ELIP*. Sumarul seturilor de gene, **Tabelul 6**.

Tabel 6. Sumarul seturilor de gene unice și semnificația adaptativă

Grup	Număr Unice	Gene	Corelație cu Transcriptomică	Divergența	Implicații Adaptative Cheie
LP_AR	2092		Corelație minimă, PC1 dominant		Plasticitate moleculară maximă, răspuns complex la stresul termic/hidric.
LP_BV	1513		Corelație intermediară		Adaptare la condiții temperate sub-montane.
LP_B_Mts	1123		Segregare puternică pe PC2		Răspuns genetic eficient și specific la stresul de frig/altitudine.
Core Transcriptome	15,135		Consistență tehnică/biologică		Funcții metabolice de bază conservate.

Discuții și Concluzii

Analiza transcriptomică detaliată a populațiilor de *Lolium perenne* L și a varietății ameliorate 'Mara' în contextul gradientelor ecologice din România a oferit dovezi moleculare substanțiale privind mecanismele de adaptare și impactul selecției.

Implicațiile varianței induse de mediu, rezultatele ACP și ale analizei de corelație converg pentru a stabili că mediul abiotice este factorul primar care modelează profilul de expresie genică în *Lolium perenne* L. Divergența maximă între populațiile de la Brașov și Arad (corelație R2 minimă) și setul masiv de gene unice la Arad (2092) confirmă că adaptarea la ariditate și căldură necesită o reorganizare metabolică mult mai amplă decât răspunsul la frig.

În contrast, deși frigul alpin din Bucegi este sever, răspunsul său transcriptomic unic este mai restrâns (1123 gene unice). Această diferență sugerează o strategie evolutivă contrastantă: adaptarea la ariditate (AR) se bazează pe un răspuns larg de plasticitate, în timp ce adaptarea la frig (Bucegi) se bazează pe reglarea eficientă și bine orchestrată a unor rețele genetice conservate, care includ probabil genele de vernalizare și cele asociate cu crioprotecția, cum ar fi membri ai familiilor *ELIP* sau cei implicați în acumularea de carbohidrați ca sursă de energie și protecție.

Evaluarea moleculară a soiului ameliorat LP_A (Mara), demonstrează o stabilitate transcriptomică remarcabilă ($R^2 > 0.95$), indicând o omogenitate genomică superioară. Poziționarea sa în PCA în imediata vecinătate a clusterului LP_BV (Brașov) sugerează că profilul său de expresie este cel mai bine optimizat pentru un mediu temperat. Soiul 'Mara' prezintă persistența și rezistența iarna, caracteristici esențiale în regiunile temperate cu ierni blânde sau moderate. Totuși, distanța sa față de profilele de stres extreme (AR și Bucegi) sugerează că, deși stabil și performant în condiții agronomice standard, genotipul LP_A ar putea necesita o activare transcriptomică mult mai dramatică (și posibil mai puțin eficientă) pentru a aborda stresurile extreme de ariditate sau frig alpin, comparativ cu populațiile sălbatice care au evoluat local în acele nișe ecologice.

Genele identificate ca fiind unice pentru populațiile sălbatice (LP_AR și LP_B_Mts) reprezintă o resursă genomică de neprețuit pentru programele de ameliorare moleculară menite să îmbunătățească rezistența *Lolium perenne* L la schimbările climatice:

- rezistența la ariditate, cele 2092 de gene unice din populația Arad (LP_AR) au fost supuse unei analize funcționale (GO și KEGG) aprofundate. Aceste gene reprezintă candidați excelenți pentru întregirea genică, având potențialul de a conferi o plasticitate moleculară sporită, esențială pentru

dezvoltarea de soiuri de raigras peren rezistente la secetă și la temperaturi ridicate, fenomene din ce în ce mai frecvente în climatul continental;

- toleranța la frigul extrem, setul de 1123 de gene unice din Bucegi (LP_B_Mts/LP1/2/3) au fost analizate în continuare pentru identificarea genelor cheie care mediază rezistența alpină. Integrarea eficientă a acestor gene în genotipuri stabile, precum 'Mara', ar putea îmbunătăți substanțial rezistența la iernare în regiunile cu solicitări termice scăzute, severe.

3.5	Analiza expresiei genice diferențiale
3.5.1	Rezultate analiză expresie diferențială
3.5.2	Date expresie genică diferențială interpretate statistic
3.5.3	Analiza de clusterizare

După ce expresia genică este cuantificată, este necesară o analiză statistică a datelor pentru a selecta genele ale căror niveluri de expresie sunt semnificativ diferite în condiții variate. Analiza diferențială este împărțită, în principal, în trei etape:

- Normalizarea: numărul brut de citiri (readcount) este normalizat, în principal pentru a corecta diferențele de profunzime a secvențierii;
- Modelarea statistică: se utilizează un model statistic pentru a calcula probabilitatea testului cu ipoteză (p-value);
- Corecția: se aplică corecții pentru teste cu ipoteze multiple pentru a obține valorile FDR (false discovery rate – rata descoperirilor false) (Anders et al., 2010).

Pentru diferite condiții experimentale, am selectat software-ul adecvat pentru analiza diferențială a expresiei genice, conform **Tabel 7** de mai jos:

Tabelul 7. Software pentru analiza diferențială și criterii de selecție a genelor

Tip	Software	Metoda de Normalizare	Model Calcul p-value	Metodă Calcul FDR	Prag Selecție Gene Diferențiale
Replicări Biologice	DESeq2 (Anders et al., 2014)	DESeq	Distribuția Binomială Negativă	BH	$ \log_2(\text{FoldChange}) \geq 1$ & padj ≤ 0.05
Fără Replicări	edgeR (Robinson et al., 2010)	TMM	Distribuția Binomială Negativă	BH	$ \log_2(\text{FoldChange}) \geq 1$ & padj ≤ 0.05

Dacă numărul de gene selectate conform pragului de mai sus este prea mic (sub 100), este foarte probabil să nu existe rezultate semnificative în analiza ulterioară de îmbogățire funcțională (functional enrichment analysis). Prin urmare, vom reduce în mod corespunzător pragul de selecție a genelor, în funcție de condițiile specifice ale proiectului. Dacă experimentul se concentrează doar pe expresia câtorva gene (cum ar fi în cazul unui gene knockout), atunci se ignoră rezultatele îmbogățirii și se filtrează genele de interes direct din tabelul de analiză diferențială.

În general, dacă o genă prezintă o variație a expresiei mai mare de două ori între cele două seturi de probe, considerăm că acea genă este exprimată diferențial. Pentru a determina dacă diferența de expresie este cauzată de diverse

erori sau de diferențe biologice esențiale, trebuie să efectuăm un test cu ipoteză pe datele de expresie ale tuturor genelor din probele respective.

Deoarece analiza transcriptomului este efectuată pe mii de gene simultan, acest lucru duce la acumularea de rezultate fals pozitive. Cu cât numărul de gene este mai mare, cu atât gradul cumulativ de rezultate fals pozitive în testul cu ipoteză este mai ridicat. Astfel, introducerea valorii p_{adj} (p -value ajustat) calibrează testul de ipoteză pentru a controla proporția acestor rezultate eronate (Young et al., 2010).

Criteriile de selecție pentru genele diferențiale sunt extrem de importante. Standardul propus, $|\log_2(\text{FoldChange})| \geq 1$ și $p_{adj} \leq 0.05$, este o valoare empirică utilizată frecvent în proiectele reale, însă poate fi adaptat flexibil. De exemplu, pragul pentru variația de expresie poate fi setat la 1.5 sau 3 ori, iar p_{adj} poate include praguri de 0.01, 0.05 sau 0.1. Dacă numărul de gene obținute este prea mare, ceea ce ar îngreuna identificarea genelor țintă, se pot aplica criterii de prag mai stricte pentru filtrare.

3.5.1 Rezultat analiză expresie diferențială

Analiza Expresiei Genice Diferențiale (DGEA) reprezintă etapa crucială care permite identificarea genelor al căror nivel de transcripție este modificat semnificativ statistic între două sau mai multe condiții experimentale (în acest caz, populații sălbatice adaptate la mediu versus varietatea ameliorată). Această analiză dezvăluie rețelele moleculare activate ca răspuns direct la presiunile selective ale mediului (stres termic, ariditate, frig alpin) și diferențele intrinseci de genotip.

Pentru a identifica genele reglate diferențial (Differentially Expressed Genes – DEGs), s-au efectuat comparații pereche (pairwise) între cele patru grupuri principale (LP_BV, LP_AR, LP_B_Mts, și LP_A), utilizând filtre statistice stricte (de obicei, praguri de False Discovery Rate (FDR) și valoarea $\log_2(\text{Fold Change})$).

Rezultatele DGEA confirmă o reprogramare transcriptomică masivă, în special în comparațiile care implică populațiile adaptate la condiții de stres extrem. Analiza dezvăluie că cel mai mare impact asupra transcripției este determinat de gradientul ecologic maxim, adică diferența dintre mediul cald/arid (LP_AR) și cel rece/alpin (LP_B_Mts). De asemenea, varietatea ameliorată (LP_A) prezintă o distanță transcriptomică semnificativă față de populația adaptată la ariditate (LP_AR), indicând o specializare genică distinctă față de mecanismele de toleranță la secetă și căldură.

3.5.2 Date expresie genică diferențială interpretate statistic

Datele statistice a genelor diferențial exprimate statistic, respectiv a numărului de gene diferențiat exprimate (DEGs) în cele șase comparații pereche esențiale sunt rezumate în **Tabel 8** și reprezentate grafic în **Figura 25**.

Tabel 8. Numărul de gene exprimate diferențiat (DEGs) în comparațiile pereche

Grupuri Comparate	Gene Total DEGs	Gene Reglate Pozitiv (Up-regulated)	Gene Reglate Negativ (Down-regulated)
LP_AR vs LP_B_Mts	13611	7826	5785
LP_A vs LP_AR	10679	7113	3566
LP_A vs LP_B_Mts	9905	5127	4778
LP_BV vs LP_AR	6534	4035	2499
LP_BV vs LP_B_Mts	4035	2231	1804
LP_A vs LP_BV	1604	937	667

Analiza statistică cheie [Divergența ecologică maximă (LP_AR vs LP_B_Mts)], această comparație generează cel mai mare număr de DEGs (13,611), confirmând că diferența între condițiile de stres termic/ariditate (AR) și stres de frig/altitudine (B_Mts) impune cea mai profundă și extinsă reorganizare a transcriptomului.

Un număr semnificativ mai mare de gene (7,826) sunt reglate pozitiv (Up) în populația LP_AR comparativ cu LP_B_Mts, indicând o activare metabolică intensă în condițiile de stres arid.

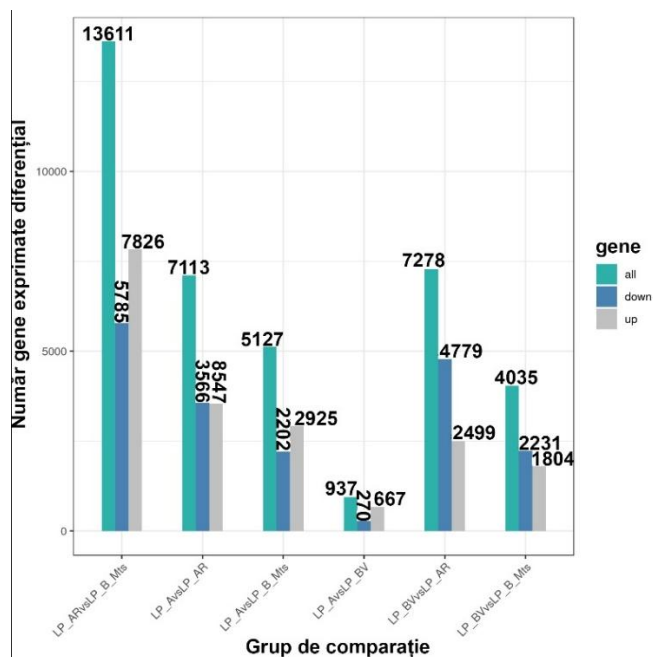


Figura 25. Histograma comparativă a numărului de gene exprimate diferențial.

Se poate realiza vizualizarea genelor exprimate diferențial și prin grafice de tip „volcano”. Graficele de tip „volcano” pot fi utilizate pentru a deduce distribuția generală a genelor exprimate diferențial. În figură, axa OX indică variația de tip „fold change” a expresiei genice între diferite probe, iar axa OY indică semnificația statistică a acestor diferențe. Punctele roșii reprezintă genele supra-exprimate (up-regulation), iar punctele verzi reprezintă genele sub-exprimate (down-regulation), **Figura 26** (pentru toate probele, detalii *Anexa 1*, Figurile 53-56).

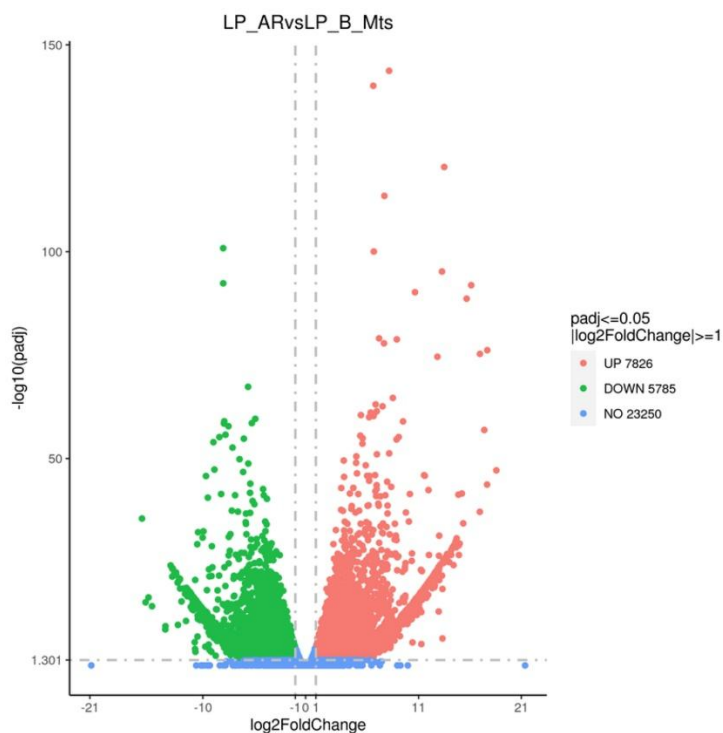


Figura 26. Graficul tip Volcano a genelor exprimate diferențial, LP_Ar vs LP_B_Mts. .

Note terminologice utile:

Fold change: se lasă adesea ca atare în literatura de specialitate în limba română sau se traduce prin „raportul variației de expresie”.

Up-regulation / Down-regulation: deși se pot traduce prin „reglare ascendentă/descendentă”, în contextul analizei de date se preferă adesea termenii de gene supraexprimate și subexprimate.

Significance: se referă de obicei la valoarea $-\log_{10}(p\text{-value})$, motiv pentru care punctele cele mai semnificative apar în partea de sus a graficului.

3.5.3 Analiza de clusterizare

Toate genele exprimate diferențial din grupul de comparație au fost reunite pentru a forma setul de gene diferențiale. Pentru experimentele care implică mai mult de două grupuri, se poate efectua o analiză de cluster (grupare) pe diferite seturi de gene, astfel încât genele cu modele de expresie similare să fie grupate împreună.

Am utilizat metoda standard de clustering ierarhic pentru a grupa valorile FPKM ale genelor și am procedat la standardizarea pe rânduri (scorul Z). Genele sau probele cu modele de expresie similare vor apărea grupate în harta termică (heat map).

Culoarea din fiecare celulă a grilei nu reflectă valoarea absolută a expresiei genice, ci valoarea obținută după standardizarea rândurilor (situată, în general, între -2 și 2). Din acest motiv, culorile din harta termică pot fi comparate doar pe orizontală (expresia aceleiași gene în diferite probe) și nu pe verticală (gene diferite în cadrul aceleiași probe). Analiza include atât gruparea între grupuri experimentale, cât și cea între probe individuale, raportul final prezentând rezultatele grupării probelor.

Analiza de clustering ierarhic, vizualizată printr-o hartă termică (Heatmap), **Figura 27**, a fost aplicată pe setul de gene diferențiat exprimate (DEGs) la nivel global, pentru a ilustra gruparea eşantioanelor în funcție de similaritatea profilurilor lor de expresie.

Clustering-ul eşantioanelor (Clusteringul Vertical), confirmă o grupare perfectă și robustă a eşantioanelor în funcție de grupul lor de proveniență (locatie/genotip), indicând o variabilitate inter-grup mult mai mare decât variabilitatea biologică intra-grup (între replici):

Cluster LP_A: cele trei replici ale soiului ameliorat (LP1_A, LP2_A, LP3_A) formează un cluster unitar, confirmând stabilitatea și omogenitatea genetică/transcriptomică a varietății.

1Cluster LP_BV: replicile din Brașov (LP1_BV, LP2_BV, LP3_BV) se grupează împreună.

1Cluster LP_B_Mts: probele din Bucegi (LP1, LP2, LP3) formează un cluster distinct.

1Cluster LP_AR: probele din Arad (LP1_AR, LP2_AR, LP3_AR) formează cel mai izolat cluster.

Relațiile Inter-Grup (Clusteringul Ierarhic), modelul de clustering ierarhic dezvăluie relațiile evolutive și adaptative, urmând o ierarhie bazată pe divergența transcriptomică. Gruparea LP_A + LP_BV: soiul ameliorat (LP_A) și populația din Brașov (LP_BV) sunt cele mai strâns înrudite din punct de vedere al expresiei genice, ele formând un cluster de nivel superior. Acest lucru este în concordanță cu numărul minim de DEGs observat între ele (1,604).

Segregarea LP_B_Mts: populația alpină (LP_B_Mts) se alătură următorului nivel al ierarhiei, indicând o distanță moleculară moderată.

Izolarea LP_AR: populația din Arad (LP_AR), adaptată la condiții aride și termice, este cel mai izolat cluster, alăturându-se ierarhiei la cel mai înalt nivel de divergență. Această izolare subliniază impactul major al stresului combinat de căldură și secetă asupra rețelei de reglare genică a plantei, care necesită o configurație transcriptomică unică pentru supraviețuire, diferită esențial de celelalte genotipuri.

Analiza blocajelor de expresie (Clusteringul Orizontal) pe harta termică (Heatmap), Figura 27, se observă blocuri mari de gene care sunt reglate pozitiv sau negativ în mod concertat, indicând activarea unor căi metabolice specifice. Astfel, se pot identifica: un bloc de gene intens reglate pozitiv (exprimate în exces, culoare închisă) în clusterul LP_AR și reglate negativ (culoare deschisă) în LP_B_Mts, reprezentând cel mai probabil genele de răspuns la stresul hidric/termic. Blocuri de gene cu expresie ridicată în LP_B_Mts și LP_BV, dar reduse în LP_AR, sugerează gene asociate cu creșterea robustă în condiții optime și/sau toleranța la iernare (frig).

În concluzie, analiza de clustering ierarhic a DEGs consolidează ideea că variația transcriptomică în *Lolium perenne* L este dominată de adaptarea ecologică, cu populația de ariditate (LP_AR) și cea de frig (LP_B_Mts) reprezentând extremele moleculare, în timp ce soiul ameliorat (LP_A) se aliniază cu profilul genotipic temperat (LP_BV).

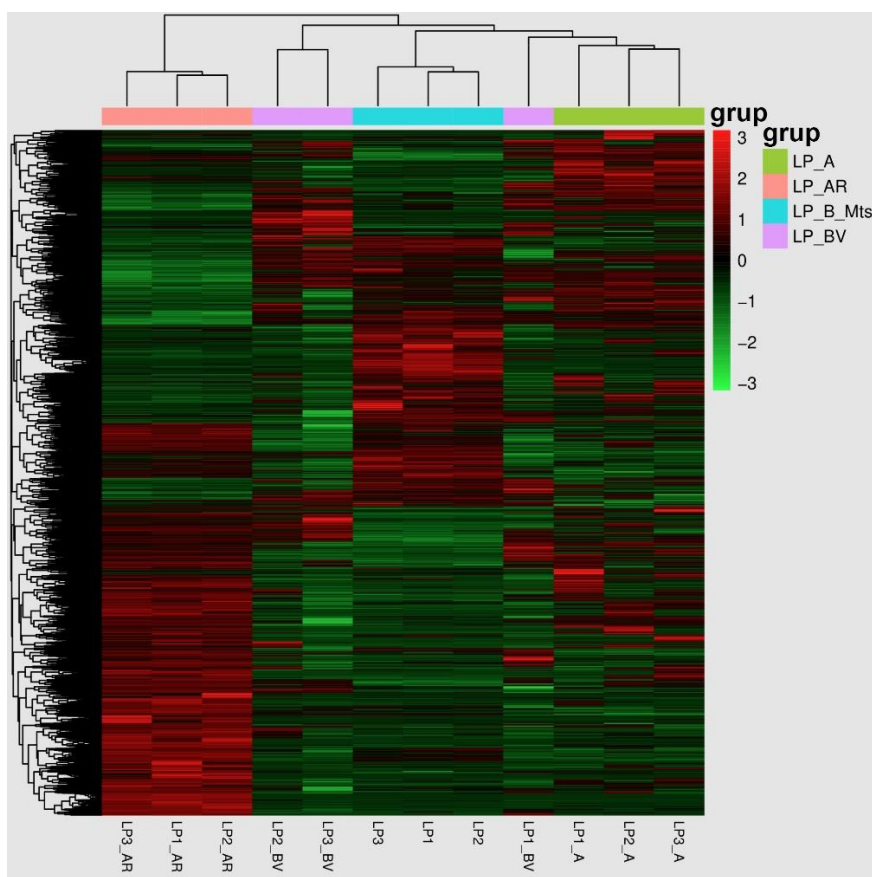


Figura 27. Hartă termică(/heat map) a grupării genelor exprimate diferențial.

Termeni cheie Eeplicați:

Clustering ierarhic: o metodă de analiză care grupează obiectele (genele sau probele) în funcție de similaritatea lor, formând o structură de tip arbore (dendrogramă).

Scorul Z (Z-score): o metodă de normalizare care transformă datele astfel încât media să fie 0 și deviația standard să fie 1. Acest lucru permite vizualizarea variațiilor relative ale unei gene, indiferent dacă aceasta are un nivel de expresie general ridicat sau scăzut.

FPKM: (Fragments Per Kilobase of transcript per Million mapped reads) – o unitate de măsură utilizată pentru a estima nivelul de expresie a genelor în secvențierea RNA-seq.

Comparație pe orizontală: esențială pentru a vedea dacă o anumită genă este mai activă într-o probă față de alta (de exemplu, ameliorat vs. spontan).

Componentele principale ale graficului:

Dendrograma (Arborele de grupare): liniile ramificate de deasupra (pentru probe) sau din lateral (pentru gene). Cu cât două ramuri se unesc mai repede, cu atât probele sau genele respective au un profil de expresie mai similar.

Scorul Z (Z-score): culorile nu reprezintă cantitatea brută de ARN, ci cât de mult deviază expresia unei gene față de media ei în toate probele.

Roșu: expresie ridicată față de medie (Up-regulated).

Verde: expresie scăzută față de medie (Down-regulated).

Axa Orizontală (Coloane): reprezintă probele individuale

Axa Verticală (Rânduri): reprezintă genele individuale selectate ca fiind semnificative.

3.6	Analiza funcțională
3.6.1	Analiza de îmbogățire a ontologiei genelor/GO/Gene ontology
3.6.2	Analiza de îmbogățire a căilor metabolice KEGG (Kyoto Encyclopedia of Genes and Genomes)
3.6.3	Analiza rețelelor de interacțiune proteină-proteină (PPI)

Prin analiza de îmbogățire (enrichment analysis) a genelor exprimate diferențial (**Figura 28**), putem identifica ce funcții biologice sau căi metabolice sunt asociate semnificativ cu acestea. Am utilizat software-ul clusterProfiler (Yu G, 2012) pentru analiza de îmbogățire, incluzând analize de tip GO (Gene Ontology), DO (Disease Ontology), precum și baze de date precum KEGG și Reactome.

Bazele de date menționate:

GO (Gene Ontology): clasifică genele în trei categorii - funcție moleculară, componentă celulară și proces biologic.

DO (Disease Ontology): face legătura între gene și afecțiunile medicale specifice.

KEGG & Reactome: acestea sunt baze de date pentru căi de semnalizare (pathways). KEGG este adesea mai schematic, în timp ce Reactome este foarte detaliat în ceea ce privește reacțiile biochimice.

Programul clusterProfiler - este standardul de aur în limbajul R pentru acest tip de analiză, deoarece integrează foarte multe baze de date și generează vizualizări clare (precum dot plots sau cnet plots), făcând datele brute mult mai ușor de prezentat într-o lucrare științifică.

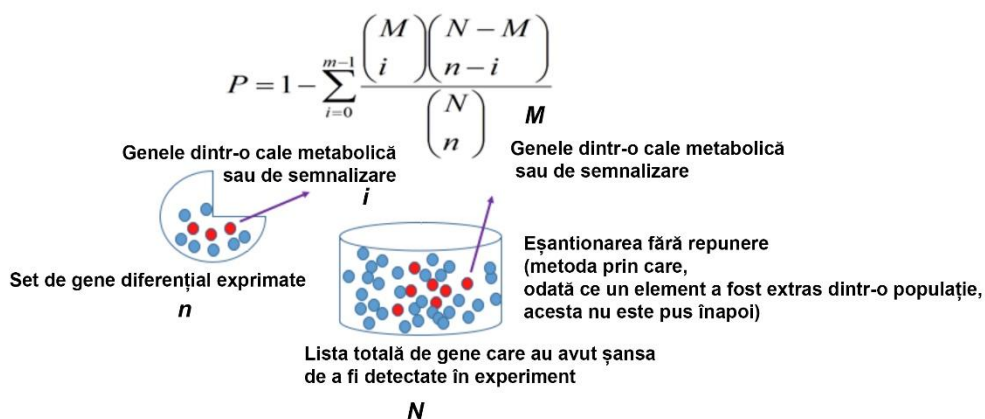


Figura 28. Schema analizei de îmbogățire a genelor.

În această etapă se analizează cele 3 genotipuri sălbatice (LP) recoltate din cele trei locații distincte geografic și climatic, comparându-le cu varietatea ameliorată (LP_A, Lolium perenne x Mara).

Arad (AR): reprezentând un mediu de câmpie cu influențe continentale puternice. Această zonă este caracterizată, în general, prin stres hidric și termic ridicat în timpul sezonului de creștere, necesitând adaptări transcriptomice centrate pe conservarea apei și termotoleranță.

Brașov (BV): situat într-o depresiune intra-montană, acest mediu prezintă o climă temperată cu umiditate relativ constantă și variații termice moderate. Ecotipurile de aici (LP_BV) pot avea o adaptare mai echilibrată, fiind folosite ca referință intermediară.

Platoul Munților Bucegi (B_Mts): un mediu alpin/subalpin, caracterizat printr-un stres sever. Acestea includ temperaturi scăzute, fluctuații rapide de temperatură, radiație ultravioletă (UV) intensă și soluri sărace. Adaptarea la altitudine implică mecanisme robuste de protecție împotriva oxidării și a destructurării celulare.

Soiul ameliorat (LP_A) este dezvoltat pentru performanțe agronomice superioare (creștere rapidă, producție mare de biomasă). Ipoteza centrală a acestui studiu este că varietatea ameliorată a dobândit o reglare genetică ce prioritizează procesele de creștere și sinteză, reușind, în același timp, să mențină sau să optimizeze anumite căi de răspuns la stres. Această optimizare este explorată prin compararea LP_A cu ecotipurile sălbatice (LP_AR, LP_BV, LP_B_Mts), care sunt forțate să aloce o parte mai mare a bugetului lor energetic pentru supraviețuire.

Analiza diferențelor de expresie genică (DEGs) a fost completată prin metode bioinformatic standard: Gene Ontology (GO) Enrichment Analysis, KEGG Pathway Enrichment Analysis și Protein-Protein Interaction (PPI) Network Analysis. Aceste instrumente permit transformarea listelor de gene reglate diferențial în procese biologice interpretabile, oferind o perspectivă sistemică asupra strategiilor de adaptare moleculară adoptate de diferite genotipuri.

3.6.1 Analiza de îmbogățire a ontologiei genelor/GO/Gene ontology

GO este abrevierea pentru Gene Ontology (Ontologia Genelor – <http://www.geneontology.org/>), un sistem major de clasificare bioinformatică utilizat pentru a unifica descrierea proprietăților genelor la toate speciile. Acesta cuprinde trei ramuri principale: componentă celulară, funcție moleculară și proces biologic. Termenii GO care prezintă o valoare $p_{adj} < 0,05$ sunt considerați ca având o îmbogățire semnificativă.

În rezultatele analizei de îmbogățire GO, au fost selectați pentru afișare cei mai semnificativi 30 de termeni, **Figura 29** pentru comparația LP_Arvs.LP_B_Mts. (*Anexa 1* - Figurile 57-60 pentru restul scenariilor). În cazul în care rezultatele conțin mai puțin de 30 de termeni, aceștia vor fi reprezentați integral, după cum se observă în figura următoare. În această figură, abscisa reprezintă termenul GO

(GO Term), iar ordonata reprezintă nivelul de semnificație al îmbogățirii acestuia. Valorile mai mari corespund unui grad de semnificație mai ridicat. Culorile diferite reprezintă cele trei subclase GO: BP (proces biologic), CC (componentă celulară) și MF (funcție moleculară).

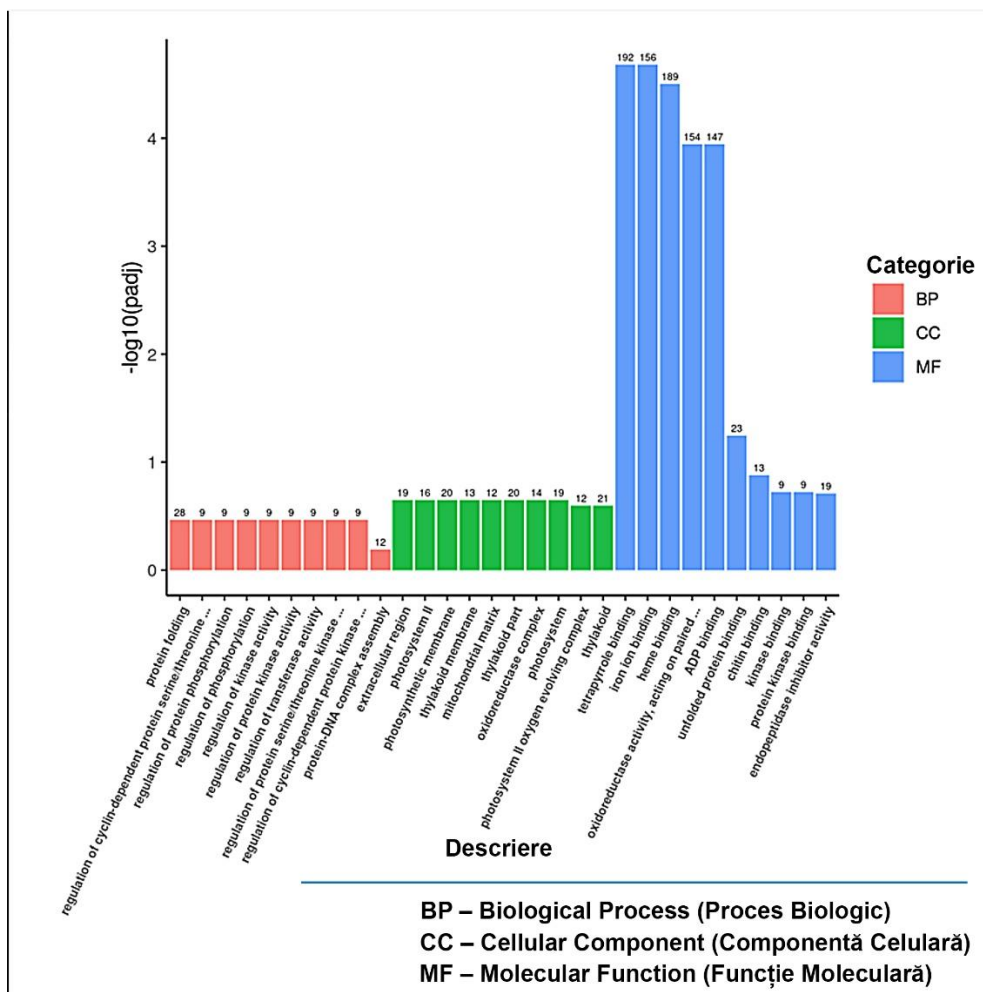


Figura 29. Histograma analizei de îmbogățire GO.

Abscisa reprezintă termenul GO (GO Term), iar ordonata reprezintă nivelul de semnificație a îmbogățirii acestuia, exprimat sub formă de $-\log_{10}(padi)$. Culorile diferite indică diverse categorii funcționale.

Din rezultatele analizei de îmbogățire GO, cei 30 de termeni GO au fost prezentați pentru o mai bună vizualizarea și interpretare a rezultatelor și sub forma unui grafic de Tip Scatter Plot, **Figura 30**, pentru scenariul LP_Arvs.LP_B_Mts. (Anexa 1 - Figurile 61-65 pentru restul scenariilor).

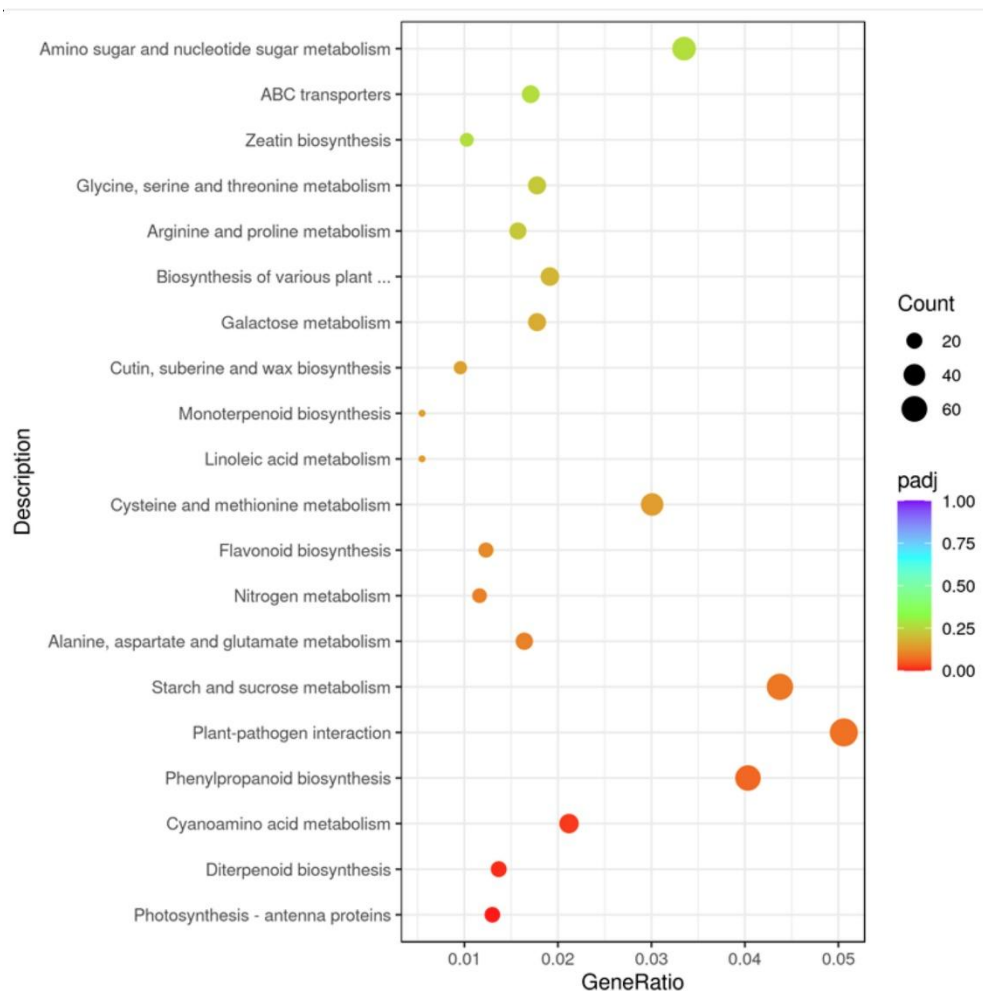


Figura 30. Analiza de îmbogățire GO, grafic de tip Scatter Plot.

În cazul în care rezultatele conțin mai puțin de 30 de termeni, aceștia vor fi reprezentați integral. În această figură, abscisa reprezintă raportul dintre numărul de gene diferențiate asociate termenului GO și numărul total de gene diferențiate, iar ordonata indică termenul GO (GO Term). Dimensiunea fiecărui punct reprezintă numărul de gene adnotate pentru un anumit termen GO, în timp ce variația culorii de la roșu la violet (mov) reprezintă nivelul de semnificație statistică a îmbogățirii. Analiza GO clasifică genele reglate diferențial (DEGs) în funcție de funcția lor moleculară (MF), procesul biologic (BP) și componenta celulară (CC) în care sunt implicate.

Modificări metabolice fundamentale (BP) în adaptarea geografică

Comparațiile între ecotipurile sălbatice dezvăluie ajustări imediate ale metabolismului primar, esențiale pentru supraviețuire în condiții de mediu variabile.

Procesul biologic carbohidrate metabolic process (GO:0005975) prezintă o valoare p de 0.00026473698779692 în comparația LP_BV vs LP_AR. Deși valoarea padj (FDR ajustat) pentru acest termen este 0.122573225349974, valoarea p extrem de scăzută reflectă o relevanță biologică puternică. Modificările în metabolismul carbohidraților sunt o piatră de temelie a adaptării plantelor, fiind adesea legate de ajustarea osmotică în fața stresului hidric, care este mai pronunțat în Arad (mediu continental) comparativ cu Brașov (mediu depresionar mai umed).

De asemenea, procesul lipid biosynthetic process (GO:0008610) a fost remarcat în comparația LP_BV vs LP_B_Mts, cu o valoare p de 0.00144552034235827. Reglarea căilor lipidice este crucială în mediile reci, cum este Platoul Bucegi, deoarece sinteza și modificarea lipidelor sunt necesare pentru menținerea fluidității membranei celulare, un aspect vital pentru toleranța la temperaturi scăzute. Se observă că adaptările fundamentale în alocarea și stocarea energiei (carbohidrați/lipide) sunt primele care se ajustează de-a lungul gradientului climatic, reflectând strategii distincte de gestionare a resurselor.

Un alt termen persistent, protein folding (GO:0006457), apare constant în comparațiile inter-sălbatice, de exemplu în LP_BV vs LP_AR și LP_AR vs LP_B_Mts, cu o valoare p de 0.00552113293394298. Această îmbogățire indică o activitate intensă a proteinelor de tip chaperon/proteine de asistare (Heat Shock Proteins - HSPs), care este un răspuns generic la stresul celular cauzat de căldură, deshidratare sau agresiuni oxidative. Faptul că acest proces este reglat diferențial în toate ecotipurile sugerează că toate sunt supuse unor stresuri semnificative care impun menținerea integrității proteice.

Semnătura răspunsului la ameliorare (LP_A) și mediu (MF & CC)

Cea mai evidentă diferență funcțională a varietății ameliorate (LP_A) față de genotipul sălbatic din Arad (LP_AR) se manifestă la nivelul mașinării de translație. Termenul de Componentă Celulară (CC) ribosome (GO:0005840) este cel mai semnificativ îmbogățit în comparația LP_A vs LP_AR, cu o valoare padj de 0.00217685213926835. Acest termen este susținut de un număr mare de gene reglate (81 de gene din 552).

Această reglare masivă a sintezei proteice, în paralel cu îmbogățirea procesului celular amide metabolic process (GO:0043603), indică o strategie a genotipului ameliorat de a prioritiza creșterea rapidă și producția de biomasă. În contrast cu genotipul sălbatic din Arad, care trebuie să economisească energie

pentru supraviețuire în condiții de stres, LP_A redirectionează resursele metabolice către o mașinărie ribozomală optimizată, stabilind astfel un echilibru care favorizează performanța agronomică în detrimentul, posibil, al flexibilității extreme de adaptare la mediu.

Pe lângă creștere, varietatea ameliorată (LP_A) demonstrează mecanisme specifice de apărare. În comparația LP_A vs LP_B_Mts, termenul de Funcție Moleculară (MF) chitin binding (GO:0008061) este semnificativ îmbogățit (padj 0.00878464230096141), cu o particularitate notabilă: toate cele 11 gene asociate sunt reglate în sus (Up), și nicio genă nu este reglată în jos (0 Down). Această reglare puternică a genelor de legare a chitinei poate fi interpretată ca un răspuns defensiv sporit împotriva patogenilor fungici sau, mai probabil, ca o interacțiune moleculară activă cu endofitul mutualist *Epichloë festucae Leuchtm., Schardl & M.R. Siegel*. Relația simbiotică cu endofitul este crucială pentru toleranța la stres (inclusiv cel alpin), iar varietatea ameliorată ar putea să fi optimizat această interacțiune, conferindu-i un avantaj ecologic.

Un alt aspect relevant este reglarea transportului. Active transmembrane transporter activity (GO:0022804, MF) este de asemenea îmbogățită în LP_A vs LP_B_Mts (padj 0.0330846658020315), dar arată un profil de reglare invers față de chitin binding, fiind predominant down-reglată (6 Up, 34 Down). Acest profil sugerează o economie energetică activă în LP_A. Deoarece transportul transmembranar este costisitor din punct de vedere energetic, reglarea generală în jos indică o optimizare a rețelei de transport pentru eficiență, în timp ce ecotipul alpin (LP_B_Mts) ar putea menține o baterie mai largă de transportori constitutiv activi pentru a gestiona o gamă mai largă de toxine sau nutrienți fluctuanți, în concordanță cu reziliența robustă necesară în medii extreme, **Tabel 9**.

Tabel 9. Top termeni GO (BP, MF, CC) semnificativi sau relevant reglați.

Comparație	Categoria	GO ID	Descriere (Termen GO)	padj (FDR)	Total Gene	Up (Count)	Down (Count)
LP_A vs LP_AR	CC	GO:0005840	ribosome	0.002177	81	-	-
LP_A vs LP_B_Mts	MF	GO:0008061	chitin binding	0.008785	11	11	0
LP_A vs LP_B_Mts	MF	GO:0022804	active transmembrane transporter activity	0.033085	40	6	34
LP_BV vs LP_AR	BP	GO:0005975	carbohydrate metabolic process	0.122573	119	-	-
LP_BV vs LP_B_Mts	BP	GO:0008610	lipid biosynthetic process	0.492922	23	-	-

3.6.2 Analiza de îmbogățire a căilor metabolice KEGG (Kyoto Encyclopedia of Genes and Genomes)

Interacțiunile dintre mai multe gene pot fi implicate în anumite funcții biologice. KEGG (Kanehisa et al., 2000) (Kyoto Encyclopedia of Genes and Genomes) reprezintă o colecție de baze de date administrate manual, care conțin resurse privind informații genomice, căi biologice și boli. Analiza de îmbogățire a căilor (pathway enrichment analysis) identifică acele căi metabolice sau de transducție a semnalului care sunt semnificativ îmbogățite în rândul genelor exprimate diferențiat, prin comparație cu fondul întregului genom. Căile KEGG care prezintă o valoare $p < 0,05$ sunt considerate ca având o îmbogățire semnificativă.

Termeni cheie utilizați:

manually curated = curatoriile manuale (termenul academic standard pentru baze de date verificate de experți);

biological-pathway = căi biologice;

signal transduction pathways = căi de transducție a semnalului (procesul prin care o celulă răspunde la substanțe din mediul său);

differentially expressed genes = gene exprimate diferențial (adesea abreviate ca DEG în engleză);

Whole genome background = fondul întregului genom (se referă la setul de referință față de care se face comparația statistică).

În timp ce GO (Gene Ontology) se concentrează pe proprietățile individuale ale genelor (unde se află, ce fac), KEGG se concentrează pe rețeaua de interacțiuni. KEGG ne arată „harta” drumurilor metabolice și cum comunică genele între ele pentru a declanșa un răspuns biologic cu referire la modificarea factorilor externi sau în cazul unei patologii.

În rezultatele analizei de îmbogățire KEGG, au fost selectate pentru afișare cele mai semnificative 20 de căi KEGG. În cazul în care rezultatele conțin mai puțin de 20 de căi, toate acestea vor fi reprezentate grafic, așa cum se prezintă în **Figura 31**, pentru scenariul LP_ArvsLP_b_Mts (pentru restul scenariilor de analiza, detalii în *Anexa 1 – Figurile 66-70*). În această figură, abscisa reprezintă calea KEGG (KEGG pathway), iar ordonata indică nivelul de semnificație al îmbogățirii căii respective. Valorile mai mari corespund unui grad de semnificație mai ridicat.

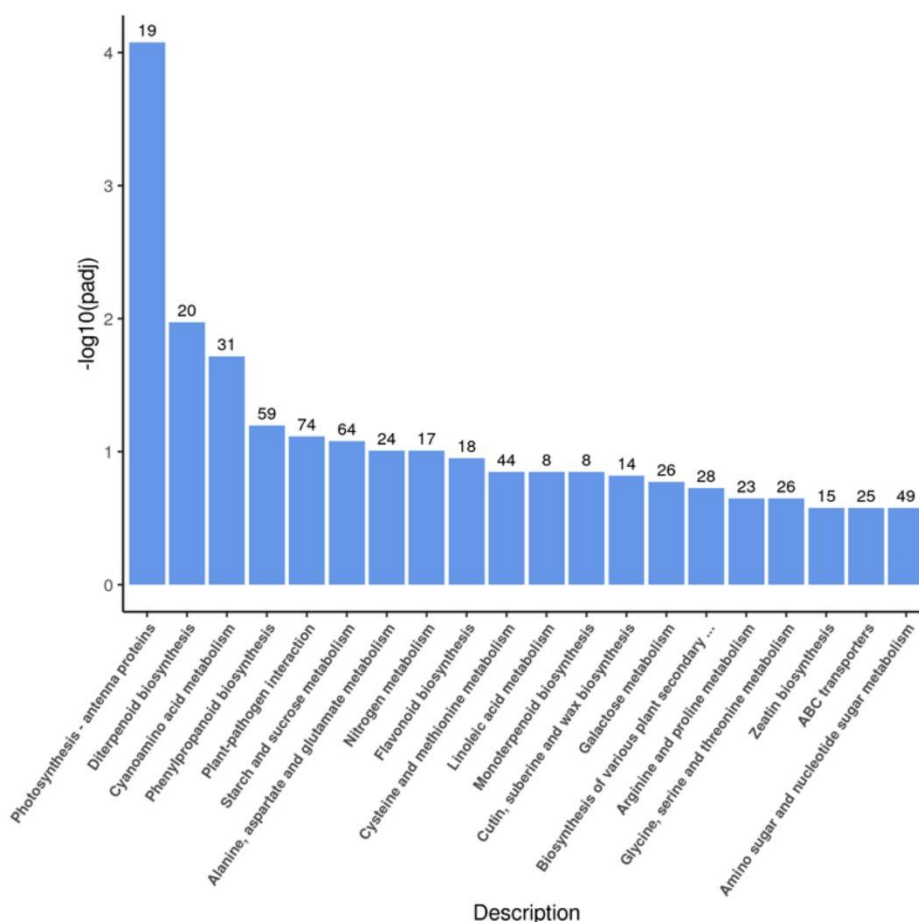


Figura 31. Histograma analizei de îmbogățire KEGG.

Abscisa reprezintă calea KEGG (KEGG pathway), iar ordonata reprezintă nivelul de semnificație al îmbogățirii căii.

Din rezultatele analizei de îmbogățire KEGG, cei 20 de termeni GO au fost prezentați pentru o mai bună vizualizarea și interpretare a rezultatelor și sub forma unui grafic de tip Scatter Plot, **Figura 32**, pentru scenariul LP_Arvs.LP_B_Mts. (Anexa 1 - Figurile 71-75, pentru restul scenariilor). Dimensiunea fiecărui punct reprezintă numărul de gene adnotate pentru un anumit termen KEGG, în timp ce variația culorii de la roșu la violet (mov) reprezintă nivelul de semnificație statistică a îmbogățirii.

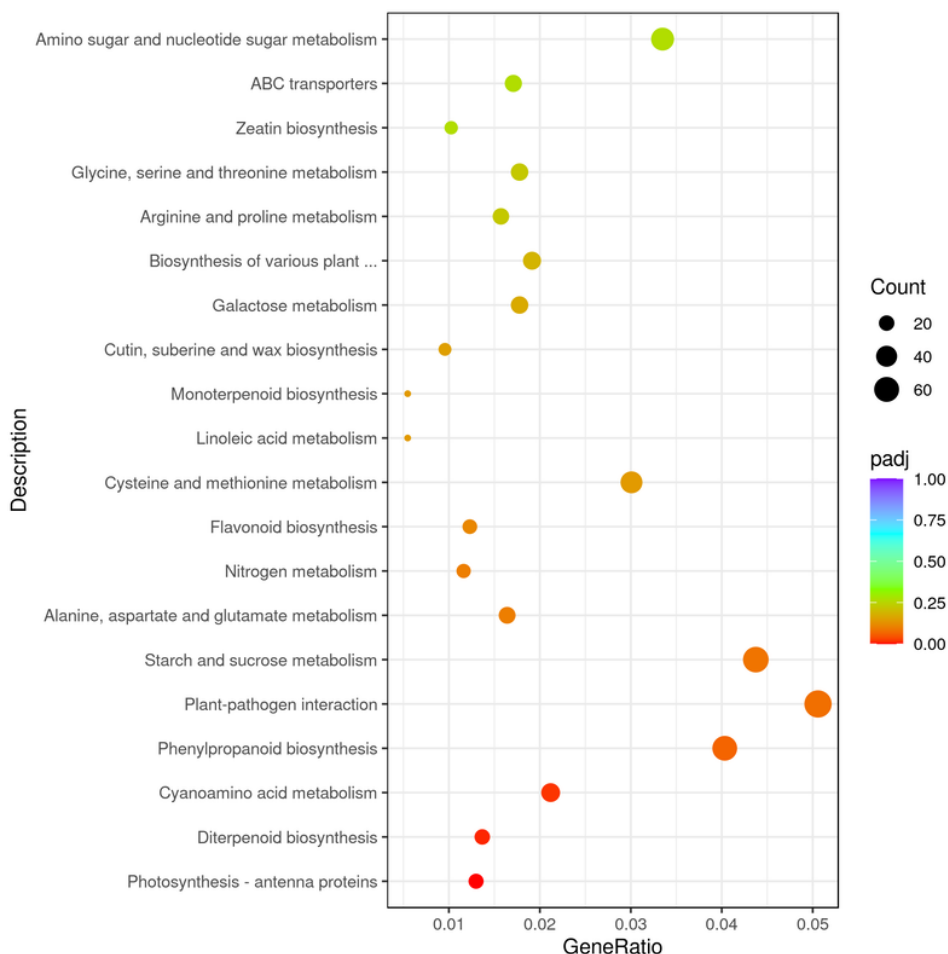


Figura 32. Analiza de îmbogățire KEGG, grafic de tip scatter plot.

În grafic, abscisa reprezintă raportul dintre numărul de gene diferențiate asociate căii KEGG și numărul total de gene diferențiate, iar ordonata indică calea KEGG (KEGG pathway).

Căile Dominante de Adaptare și Stres

Calea Ribosome biogenesis in eukaryotes (taes03008) este de o importanță fundamentală, apărând ca fiind extrem de semnificativă în multiple comparații, inclusiv LP_A vs LP_AR (padj 0.000340776134582746), LP_AR vs LP_B_Mts (padj 0.00268117894225251) și LP_BV vs LP_AR (padj 0.00268117894225251). Această coerență indică faptul că ajustarea mașinăriei de translație este un mecanism adaptativ esențial, nu doar un efect al ameliorării. Genotipurile, indiferent de mediu, folosesc reglarea biogenezei ribozomilor pentru a ajusta rapid sinteza proteică ca răspuns la stresul nutrițional, hidric sau termic.

Un alt set de căi critice este asociat cu adaptarea chimică la mediu. Calea Phenylpropanoid biosynthesis (taes00940) este îmbogățită semnificativ în

LP_AR vs LP_B_Mts (padj 0.020797193851383) și prezintă o reglare diferențială substanțială în LP_A vs LP_B_Mts, cu 23 de gene reglate în sus (Up) și 11 reglate în jos (Down). Această cale metabolică este vitală pentru sinteza metaboliților secundari, inclusiv lignina (pentru suport structural și rezistență la secetă) și flavonoidele (pentru protecție UV și antioxidare).

Reglarea diferențială a fenilpropanoizilor în comparația LP_AR vs LP_B_Mts reflectă cerințe de adaptare diferite: în Arad (AR), accentul ar putea fi pe producția de lignină pentru a menține rigiditatea peretelui celular și a reduce pierderea de apă, în timp ce în Bucegi (B_Mts), prioritatea ar putea fi sinteza de compuși fenolici pentru a proteja împotriva radiației UV intense de la altitudine. Reglarea mixtă (23 Up / 11 Down) observată în varietatea ameliorată (LP_A) sugerează o modularitate genetică sofisticată, unde anumite brațe ale căii sunt activate pentru a sprijini obiectivele specifice de creștere sau apărare dorite prin ameliorare.

Alte căi importante de răspuns ecologic includ Photosynthesis - antenna proteins (taes00196), un semnal așteptat la altitudine (B_Mts) ca mecanism de disipare a energiei luminoase excesive (protecție împotriva fotoinhibiției). De asemenea, Monoterpenoid biosynthesis (taes00902), implicată în semnalizarea de apărare și protecția volatilă împotriva stresului termic sau biotic, este remarcată în LP_BV vs LP_B_Mts., **Tabel 10.**

Tabel 10. Căi KEGG semnificativ îmbogățite (selecție cu padj < 0.05).

Calea KEGG	KEGG ID	Comparație	padj (FDR)	Total Gene	Up (Count)	Down (Count)
Ribosome biogenesis eukaryotes	taes03008	LP_A vs LP_AR	0.000341	41	Indisponibil	Indisponibil
Ribosome biogenesis eukaryotes	taes03008	LP_AR vs LP_B_Mts	0.002681	34	Indisponibil	Indisponibil
Ribosome biogenesis eukaryotes	taes03008	LP_BV vs LP_AR	0.002681	34	Indisponibil	Indisponibil
Phenylpropanoid biosynthesis	taes00940	LP_AR vs LP_B_Mts	0.020797	34	Indisponibil	Indisponibil

3.6.3 Analiza rețelelor de interacțiune proteine-proteine

Analiza PPI vizează identificarea arhitecturii moleculare a răspunsului transcriptomic, punând accent pe nodurile centrale (Hubs) care prezintă cea mai mare conectivitate și scoruri de încredere ridicate (scor > 700, ideal > 900). Aceste Hubs sunt, de obicei, reglatori majori ai funcțiilor biologice identificate prin GO și KEGG.

Identificarea Nodurilor Centrale (Hubs), pentru comparațiile inter-ecotipice și genotipice prezintă o densitate mare de interacțiuni cu scoruri de încredere maxime (970-999), subliniind conexiuni funcționale extrem de sigure între proteinele reglate diferențial.

În comparația LP_AR vs LP_B_Mts, care reflectă contrastul maxim de mediu (câmpie fierbinte vs. alpin rece), au fost identificate mai multe interacțiuni de vârf. De exemplu, gena *gene_SNK15_020256* formează o interacțiune cu *gene_SNK15_013644* la un scor de 999. Această interacțiune, având cea mai mare încredere observată, sugerează că *SNK15_020256* acționează ca un nod central critic, probabil un factor de transcripție sau o kinază, care coordonează adaptarea fundamentală, cum ar fi gestionarea stabilității membranei în fața frigului sau reglarea homeostaziei antioxidative necesare la altitudine. Alte interacțiuni Hub notabile includ *gene_SNK15_006132* interacționând cu *gene_SNK15_034213* (scor 995) și *gene_SNK15_016302* interacționând cu *gene_SNK15_001473* (scor 976). Aceste gene coordonatoare sunt esențiale pentru strategiile de supraviețuire în mediile extreme.

În comparația LP_A vs LP_AR (ameliorare vs. mediu stresant), s-a observat că gena *SNK15_003780* este un Hub proeminent, formând multiple interacțiuni puternice, cum ar fi cea cu gena *SNK15_004285* (scor 987). Prezența acestui Hub cu conectivitate maximă în genotipul ameliorat indică faptul că *SNK15_003780* ar putea fi un reglator al proceselor de creștere rapidă și de eficiență metabolică, mecanisme care definesc superioritatea agronomică a LP_A.

De asemenea, s-a observat o genă cu un rol potențial transversal în adaptare: gena *SNK15_037705* interacționează puternic cu gena *SNK15_035616* (scor 993) în LP_BV vs LP_AR și apare frecvent și în alte comparații. Acesta ar putea reprezenta un modul de reglare esențial și conservat, responsabil pentru gestionarea tensiunii ecologice generale, indiferent de tipul specific de stres (termic, hidric, altitudinal).

Integrarea funcțională a rețelelor PPI cu GO și KEGG, oferă o imagine completă: procesele funcționale (identificate prin GO/KEGG) sunt orchestrate de o ierarhie de gene reglatoare (Hubs).

De exemplu, reglarea intensă a căii Ribosome biogenesis in eukaryotes (KEGG) și a termenului ribosome (GO) este direct controlată de factori de transcripție și kinaze care acționează ca Hubs în rețeaua PPI (e.g., SNK15_003780 în LP_A vs LP_AR). Astfel, îmbunătățirea performanței la LP_A se datorează nu doar reglării genelor ribozomale individuale, ci și eficienței sporite a acestui regulator Hub.

Similar, căile de biosinteză a fenilpropanoizilor, vitale pentru apărarea la altitudine și secetă, sunt coordonate de Hubs care integrează semnalele de stres (UV, deshidratare). Genele PPI identificate sunt, prin urmare, candidați puternici pentru a fi principal-regulatoare care ajustează alocarea resurselor metabolice între creștere, sinteza proteică și producția de metaboliți de apărare.

Discuții și interpretarea biologică

Mecanismele transcriptomice ale răspunsului ecotipic la stresul ecologic

Analiza a demonstrat că adaptarea ecologică a genotipurilor sălbatice de *Lolium perenne* L de-a lungul gradientului geografic este mediată de ajustări modulare ale căilor metabolice.

Ecotipul de altitudine (LP_B_Mts) prezintă o necesitate puternică de a regla căile de protecție, evidențiată prin îmbogățirea căilor metabolice secundare (fenilpropanoizi și monoterpenoizi) și gestionarea energiei fotosintetice (Photosynthesis antenna proteins). Această semnătură transcriptomică reflectă o strategie de supraviețuire axată pe protecția chimică împotriva daunelor UV și oxidării, precum și pe eficiența energetică în condiții termice dificile. Modificările în sinteza lipidelor (LP_BV vs LP_B_Mts) subliniază necesitatea ajustării structurale a membranelor pentru a face față temperaturilor scăzute.

În schimb, ecotipurile de câmpie și depresiune (LP_AR și LP_BV) se diferențiază prin ajustări în metabolismul primar, cum ar fi carbohidrate metabolic process. Deși procesul de pliere a proteinelor (protein folding) este activ în toate mediile (indicând stres general), specificitatea răspunsului la stresul hidric și termic între AR și BV se manifestă în ajustările structurale și hidrice, necesare pentru a menține turgorul și a preveni deteriorarea celulară în fața variațiilor de umiditate și căldură.

Semnătura genetică a ameliorării: prioritizarea creșterii și răspunsul modular

Varietatea ameliorată (LP_A) afișează o reorientare distinctă a resurselor metabolice. Dominanța absolută a termenilor legați de sinteza proteică (ribosome GO:0005840 și Ribosome biogenesis in eukaryotes KEGG

taes03008) în LP_A vs LP_AR sugerează că selecția a favorizat acele gene care maximizează rata de translație. Acest lucru justifică performanța agronomică superioară, dar implică o întrebare fundamentală în genetica plantelor: există un compromis energetic, unde LP_A își asumă un risc mai mare în fața factorilor stresori imprevizibili prin devierea resurselor de la mecanismele robuste de apărare către creștere rapidă?

Un alt element cheie al ameliorării este modularea sistemului de apărare. Semnalul puternic și unidirecțional (11 Up, 0 Down) al chitin binding în LP_A vs LP_B_Mts este un indiciu al unei strategii optimizate. Aceasta poate fi fie o apărare biotică specifică, fie o interacțiune simbiotică eficientă. Având în vedere rolul endofitului *Epichloë festucae* **Leuchtm., Schardl & M.R. Siegel** în toleranța la stres la *Lolium perenne* **L**, reglarea ascendentă a genelor de legare a chitinei ar putea reflecta o interacțiune gazdă-endofit specifică și avantajoasă, care conferă LP_A o rezistență indirectă la stresul ecologic.

Eficiența energetică a LP_A este sugerată și de reglarea rafinată a transportului transmembranar activ (majoritar down-reglat cu 34 de gene). Această restrângere a profilului transportorilor indică o reducere a cheltuielilor energetice inutile, un mecanism de economie care eliberează energie pentru procesele de sinteză și creștere.

Interconectarea căilor metabolice și rețelele de reglare

Analiza integrată a permis trecerea de la identificarea proceselor (GO/KEGG) la identificarea controlului ierarhic (PPI). De exemplu, deși calea phenylpropanoid biosynthesis este crucială pentru adaptarea la stres (KEGG), reglarea sa finală este orchestrată de Hub genes.

Genele nodale identificate în rețelele PPI (e.g., SNK15_020256, SNK15_003780) funcționează ca integratori de semnal. Aceste gene, cel mai probabil factori de transcripție sau componente ale cascadelor de kinaze (reflectând termenul GO regulation of protein kinase activity), primesc semnale ecologice (lumină, temperatură) și apoi direcționează alocarea resurselor fie către creștere (prin reglarea biogenezei ribozomilor), fie către apărare chimică (fenilpropanoizi).

Diferențele în scorurile de încredere PPI reflectă puterea corelației funcționale: interacțiunile cu scoruri de 999 (e.g., SNK15_020256 în adaptarea alpină) sunt dovezi clare ale unor module de reglare esențiale care definesc fenotipul ecotipic. Identificarea acestor Hubs oferă oportunități de a decoda complexitatea adaptării la stres, arătând că răspunsul nu este o sumă de căi izolate, ci un sistem interconectat, coordonat de un număr restrâns de regulatori cheie.

3.7 Analiza structurală

3.7.1 Analiza matisării/splicing-ului alternativ

3.7.1.1 Detectarea izomorfelor exprimate diferențial

3.7.1.2 Vizualizarea evenimentelor de splicing/matisare alternativă(AS)

3.7.2 Analiza SNPs și InDel

3.7.2.1 Identificarea siturilor polimorfice

3.7.2.2 Statistica situri polimorfice

3.7.1 Analiza matisării/splicing-ului alternativ

Matisarea alternativă (AS) este un proces universal de reglare genetică întâlnit la majoritatea eucariotelor. Genele segmentate ale eucariotelor sunt alcătuite din secvențe **intronice** și **exonice**. În timpul procesării ARN-ului, intronii sunt eliminați (matisați) de către **spliceozomi**, în timp ce exonii sunt editați continuu și reținuți în moleculele de **ARNm matur**, **Figura 33**. Anumite molecule de pre-ARNm pot prezenta tipare de matisare diferite în condiții variate, generând **izoforme proteice** diferite, ceea ce sporește complexitatea și adaptabilitatea biologică a eucariotelor.

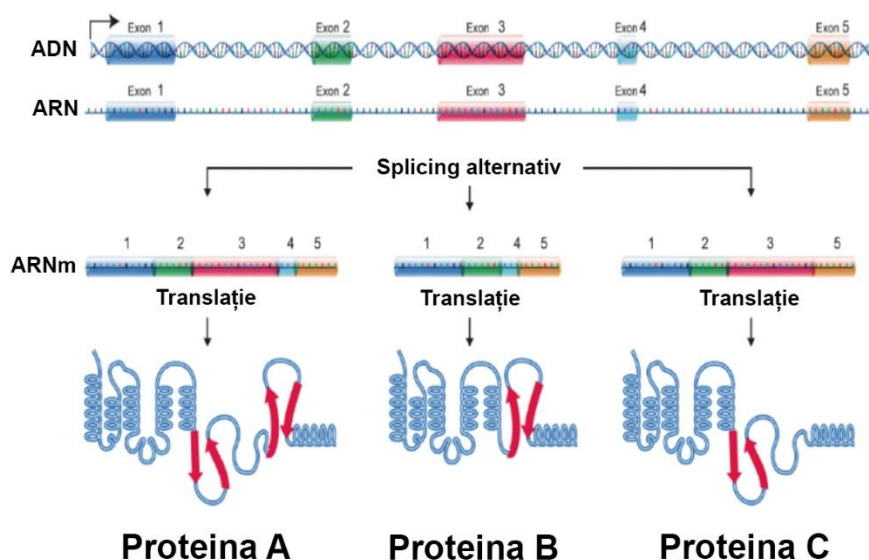


Figura 32. Eliminarea intronilor în procesul de matisare/splice-ing cu formarea ARN matur.

rMATS este un software adecvat pentru analiza matisării alternative a datelor de tip RNA-Seq (Shen et al., 2014), care nu numai că poate clasifica evenimentele de matisare variabilă, dar poate efectua și o analiză diferențială a acestora între diferite probe. Evenimentele de matisare alternativă sunt prezentate mai jos în **Figura 33**.

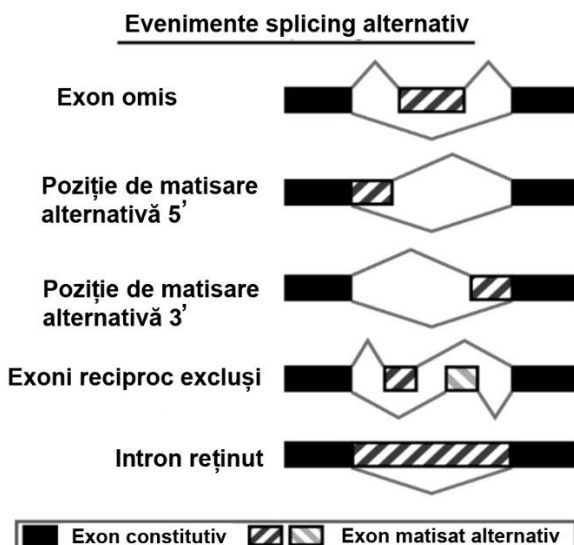


Figura 33. Clasificarea evenimentelor de matisare variabilă.

Clasificarea evenimentelor de matisare variabilă (matisare alternativă), utilizată în mod obișnuit în bioinformatică și genetică pentru a descrie modul în care pre-ARNm-ul este procesat:

1. Sărirea exonului (Skipped Exon - SE)

Este cel mai frecvent tip de matisare alternativă. În acest caz, un exon central (numit exon casetă) poate fi fie inclus în transcriptul final, fie exclus (sărit) împreună cu intronii care îl flanchează.

Rezultat: Două izoforme de ARNm, una mai lungă (cu exonul inclus) și una mai scurtă.

2. Sit alternativ de matisare 5' (Alternative 5' Splice Site - A5SS)

Acest eveniment are loc atunci când există două sau mai multe situri de matisare la capătul 5' (donator) al unui exon.

Rezultat: Schimbarea joncțiunii de matisare modifică lungimea exonului din amonte, adăugând sau eliminând o porțiune din secvența acestuia.

3. Sit alternativ de matisare 3' (Alternative 3' Splice Site - A3SS)

Similar cu A5SS, dar procesul are loc la capătul 3' (acceptor). Se folosesc situri de matisare diferite pentru a finaliza un exon.

Rezultat: Exonul din aval va avea o lungime variabilă în funcție de situl ales.

4. Exoni reciproc excluși (Mutually Exclusive Exons - MXE)

Acesta este un mecanism mai complex în care, dintr-un grup de doi (sau mai mulți) exoni adiacenți, doar unul singur este reținut în ARNm-ul matur, niciodată amândoi în același timp.

Rezultat: Izoformele produse vor conține secvențe proteice distincte, dar care ocupă aceeași poziție structurală în proteină.

5. Retenția intronului (Retained Intron - RI)

În acest scenariu, o secvență care în mod normal este considerată intron nu este eliminată, ci este păstrată în fragmentul final de ARNm.

Rezultat: De multe ori, acest lucru duce la introducerea unui codon stop prematur, ceea ce poate declanșa degradarea ARNm-ului sau poate produce o proteină cu o funcție complet diferită.

3.7.1.1 Detectarea izomorfelor exprimate diferențial

Fiecare eveniment de matisare alternativă corespunde a două izoforme, care sunt denumite Izoforma cu Incluziunea Exonului (Exon Inclusion Isoform) și Izoforma cu Omiterea Exonului (Exon Skipping Isoform), **Figura 34**. Statistica volumului de expresie este realizată prin raportarea la lungimea efectivă de expresie în momentul calculării raportului dintre Izoforma cu Incluziune și expresia totală a celor două izoforme, urmată de analiza semnificației diferențiale. Pragul de semnificație pentru analiza matisării alternative (AS) diferențiale este stabilit la $FDR < 0,05$.

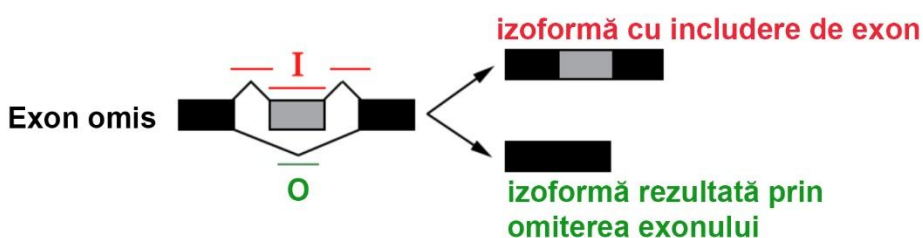


Figura 34. Tipul de matisare alternativă SE (skipped exon / sărirea exonului).

Caseta gri din figură indică exonul care prezintă o tăiere variabilă (matisare variabilă), iar cea neagră reprezintă exonul constitutiv.

3.7.1.2 Vizualizarea evenimentelor de splicing/matisare alternativă(AS)

O modalitate de vizualizare a datelor generate prin procesul de matisare alternativă se poate face prin garficul de tip Sashimi plot, **Figura 35**, (*Anexa 1* - Figurile 76-80 pentru restul matisărilor).

Caracteristicile graficului de tip Sashimi plot sunt următoarele:

- densitățile de citiri RNA-Seq de-a lungul exonilor sunt reprezentate sub formă de histograme, codificate prin culori în funcție de proba analizată.
- densitățile RNA-Seq sunt aliniate cu izoformele desenate în partea de jos a graficului; acestea sunt citite automat din adnotarea GFF a evenimentelor furnizate ca date de intrare.
- citirile de jonctiune (junction reads) sunt vizualizate sub formă de arcuri care conectează perechea de exoni pe care jonctiunea îi mărginește.
- grosimea arcului este proporțională cu numărul de citiri de jonctiune prezente în probă, însă numărul exact al acestora poate fi, opțional, afișat pe grafic.

Coordonata verticală (axa Y) indică o valoare RPKM modificată.

Termeni tehnici explicați:

Sashimi plot: un instrument de vizualizare standard în genomica ARN care combină histogramele de acoperire a exonilor cu arcuri ce reprezintă jonctiunile de matisare.

GFF (General Feature Format): un format de fișier utilizat pentru a descrie genele și alte caracteristici ale secvențelor de ADN, ARN și proteine.

RPKM (Reads Per Kilobase of transcript, per Million mapped reads): o metodă de normalizare a datelor RNA-Seq care permite compararea nivelurilor de expresie între diferite probe și gene de lungimi diferite. În Sashimi plots, valorile sunt adesea "modificate" pentru a reflecta densitatea locală.

Junction reads: acele citiri care se suprapun exact peste locul unde doi exoni au fost legați, oferind dovada directă a unui eveniment de matisare specific.

CM033338.2:19537877:19537963:-@CM033338.2:19537199:19537299:-@CM033338.2:19537028:19537127:-

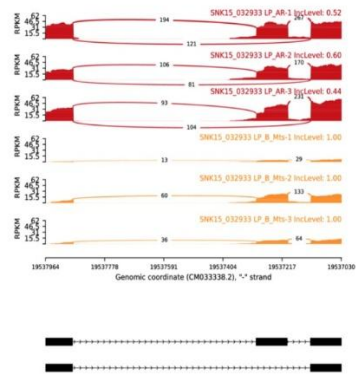


Figura 35. Vizualizarea matisării alternative de tip SE (Skipped Exon).

Culorile roșu și portocaliu din figură reprezintă probe (eșantioane) diferite, iar valoarea nivelului de incluziune (inclusion level) reprezintă raportul dintre expresia totală a izoformei cu incluziunea exonului și expresia totală a celor două izoforme; această valoare poate indica direct diferența de expresie a evenimentelor de matisare alternativă între diferite probe.

3.7.2 Analiza SNPs și InDel

Un polimorfism nucleotidic unic (SNP) reprezintă o substituție a secvențelor nucleotidice care apare frecvent în cadrul unei populații (de exemplu, într-un procent de peste 1%), în care un singur nucleotid din genom sau o altă secvență partajată este schimbată între membrii unei specii biologice sau între cromozomii pereche.

În cazul SNP-urilor, apar două tipuri de variații — tranzițiile și transversiunile — care se produc cu un raport de probabilitate de 1:2. Regiunile cu o incidență ridicată a SNP-urilor sunt adesea secvențele CG, rezultând tranziții de la C la T, care sunt asociate cu tendința citozinei (C) de a fi metilată.

Spre deosebire de SNP-uri, InDel este un termen din biologia moleculară care desemnează o inserție sau o deleție de baze în genomul unui organism, cu o lungime cuprinsă între 1 și 10.000 de perechi de baze, fiind clasificată în categoria variațiilor genetice mici.

Analiza siturilor de mutație este o componentă importantă a analizei structurale RNA-seq, incluzând detectarea mutațiilor congenitale (germinale) și a mutațiilor somatice dobândite, ambele fiind esențiale în cercetarea oncologică. Am utilizat pachetul software GATK pentru a efectua analiza siturilor de mutație pe datele probelor și software-ul SnpEff pentru adnotarea siturilor variante (McKenna A et al., 2010).

3.7.2.1 Identificarea siturilor polimorfice

Siturile variante sunt împărțite în SNP-uri și INDEL-uri. Tabelul următor prezintă rezultatele SNP, genotipurile și adnotările pentru fiecare sit, **Tabelul 11**.

Tabelul 11. Tipul sitului variant și adnotarea acestuia.

CHROM	POS	ID	REF	ALT	QUAL	DP	AD	GT	GeneID	GeneName
CM033332.2	295368	-	G	C	526.02	13	0.13	0/1	gene_SNK15_000003	SNK15_000003
CM033332.2	295575	-	T	C	248.6	19	12.7	0/1	gene_SNK15_000003	SNK15_000003
CM033332.2	1201456	-	A	T	6800.03	154	0.154	0/1	gene_SNK15_000008-gene_SNK15_000009	SNK15_000008-SNK15_000009
CM033332.2	1242139	-	A	G	460.02	11	0.11	0/1	gene_SNK15_000010	SNK15_000010
CM033332.2	1242154	-	C	A	317.6	11	3.8	0/1	gene_SNK15_000010	SNK15_000010

CHROM	FeatureID	Biotype	HGVS_C	HGVS_P	EFFECT	IMPACT
CM033332.2	gnt_WGS_JAEVFB_mma_SNK15_000003	protein_coding	c.463-269C>G	-	intron_variant	MODIFIER
CM033332.2	gnt_WGS_JAEVFB_mma_SNK15_000003	protein_coding	c.462+90A>G	-	intron_variant	MODIFIER
CM033332.2	gene_SNK15_000008-gene_SNK15_000009	-	n.1201456A>T	-	intergenic_region	MODIFIER
CM033332.2	gnt_WGS_JAEVFB_mma_SNK15_000010	protein_coding	c.3255A>G	p.Thr1085Thr	synonymous_variant	LOW
CM033332.2	gnt_WGS_JAEVFB_mma_SNK15_000010	protein_coding	c.3270C>A	p.Thr1090Thr	synonymous_variant	LOW

1. Identificarea Variantei (Pozitie și Nume)

- POS (Pozitie): locația exactă (coordonată numerică) a variației pe cromozom. Este adresa „stradală” a mutației.
- ID: codul de identificare din baza de date publică dbSNP (de obicei începe cu "rs", ex: rs12345). Dacă este o mutație nouă, acest câmp rămâne adesea gol (.).

2. Tipul de Schimbare (REF vs. ALT)

- REF (Referință): nucleotida (A, C, G sau T) care se găsește în mod normal în genomul standard la acea poziție.
- ALT (Alternativă): nucleotida detectată în proba analizată, care diferă de referință. Aceasta este „mutația”.

3. Calitate și Profunzime (Fiabilitatea Datelor)

Aceste valori îți spun câtă încredere poți avea că mutația este reală și nu o eroare de aparat.

- **QUAL (Calitate):** un scor numeric (Phred-scaled). Cu cât este mai mare, cu atât e mai puțin probabil ca varianta să fie o eroare de citire. (De exemplu, un scor de 30 înseamnă o probabilitate de eroare de 1 la 1000).
- **DP (Depth / Profunzime):** numărul total de „citiri” (fragmente de ADN) care au acoperit acea poziție. Mai multe citiri înseamnă o precizie mai mare.
- **AD (Allele Depth / Profunzimea Alelelor):** defalcare a parametrului DP. Arată exact câte citiri susțin varianta de referință și câte susțin varianta mutantă.
 - *Exemplu:* AD = 10, 15 înseamnă că 10 citiri spun „e normal” și 15 citiri spun „e mutație”.

4. Genotipul (GT)

- **GT (Genotype):** determină starea genetică a probei la acel punct:
 - 0/0: homozigot referință (ambele alele sunt ca în șablon).
 - 0/1: heterozigot (o alelă normală, una mutantă).
 - 1/1: homozigot variantă (ambele alele sunt mutante).

GenelD / GeneName: ID-ul / Numele Genei unde este localizat site-ul variantei.

FeatureID: ID-ul elementului genetic (de obicei ID-ul transcriptului) unde se află varianta.

Biotype: biotipul (tipul) transcriptului respectiv (ex: protein coding, lincRNA).

HGVS_C: adnotarea HGVS la nivel de ADN. Descrierea variantei folosind nomenclatura standard la nivel de secvență de ADN.

HGVS_P: adnotarea HGVS la nivel de proteină. Dacă varianta se află într-o regiune codificatoare, acest câmp descrie modificarea la nivelul aminoacizilor. (Deoarece ID-ul transcriptului este deja menționat, acesta poate fi omis aici).

EFFECT: efectul presupus - descrierea detaliată a impactului funcțional prezis de algoritmul SnpEff (ex: missense variant, stop gained).

IMPACT: impactul - categorie prestabilită pentru a ajuta la prioritizarea variantelor semnificative. Acestea sunt: {high (ridicat), moderate (moderat), low (scăzut), modifier (modificator)}.

3.7.2.2 Statistica situri polimorfice

După utilizarea GATK pentru detecția variantelor, am cuantificat fiecare site de variantă conform informațiilor de adnotare furnizate de SnpEff, incluzând în principal statistici privind localizarea acestora, **Figura 36** (*Anexa 1 – Figurile 81-83, cu privire la localizare, funcție și impact*).

Funcția variantei (SNP_function): este reprezentată grafic sub trei aspecte: mutații sinonime (nu schimbă aminoacidul), mutații missense (schimbă un aminoacid cu altul) și mutații nonsense (introduc un codon STOP prematur).

Regiunea variantei (SNP_region): este analizată statistic pe baza regiunilor structurale ale genei, cum ar fi EXON (regiuni codificatoare), INTRON (regiuni non-codificatoare intracelulare) și INTERGENIC (spațiul dintre gene).

Impactul variantei (SNP_impact): este clasificat pe patru niveluri de severitate:

HIGH (ridicat) - variante care au un impact disruptiv major (ex: pierderea startului sau stopului);

MODERATE (mediu) - variante non-disruptive care ar putea schimba eficiența proteinei (ex: missense);

LOW (scăzut) - variante de obicei inofensive (ex: sinonime);

MODIFIER (modificator) - variante în regiuni non-codificatoare unde predicția impactului este mai dificilă fără alte date suplimentare.

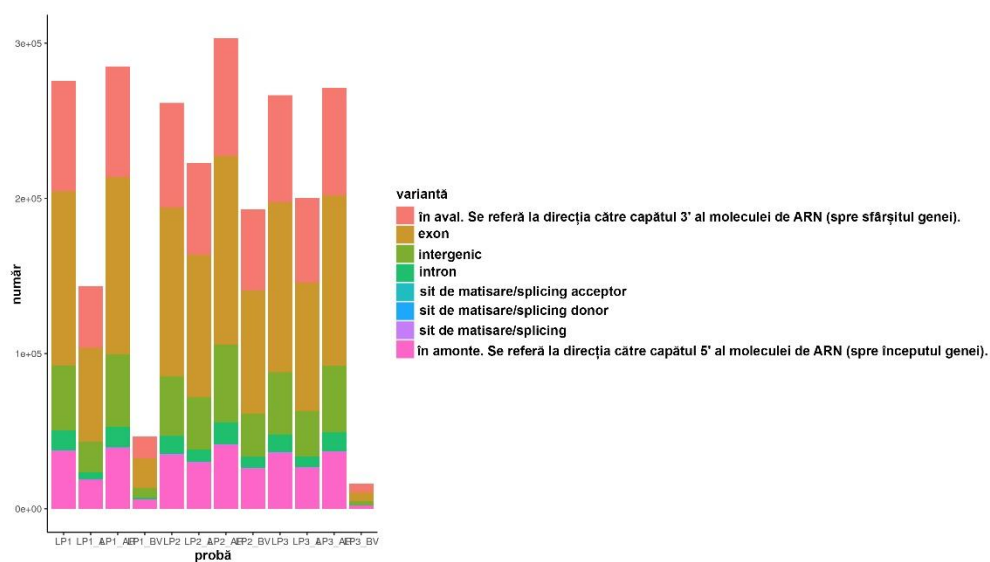


Figura 36. Analiza statistică a siturilor polimorfice în funcție de regiune.

IV. ANEXA 1

Rata erorilor de secvențiere

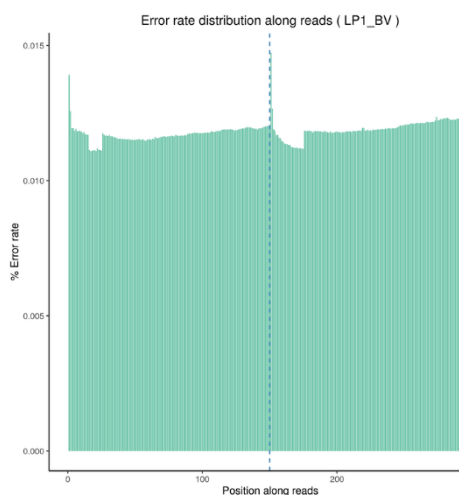


Fig. 1 - Distribuția ratei de eroare a datelor de secvențiere, LP1_BV.

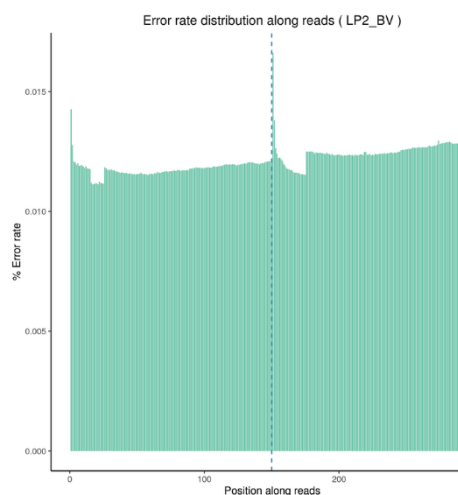


Fig. 2 - Distribuția ratei de eroare a datelor de secvențiere, LP2_BV.

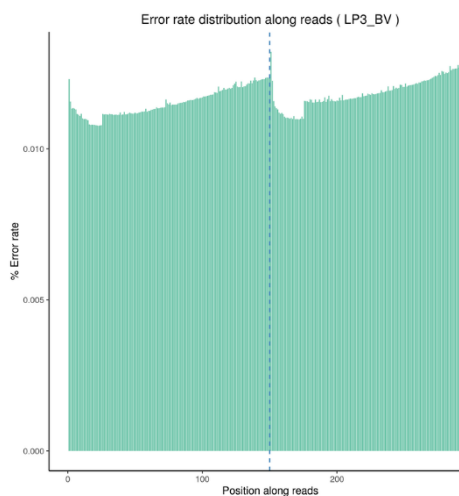


Fig. 3 - Distribuția ratei de eroare a datelor de secvențiere, LP3_BV.

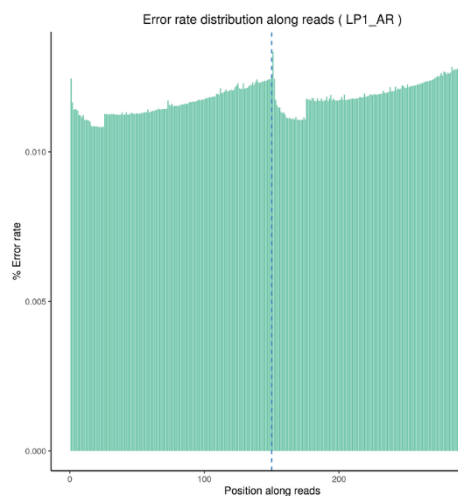


Fig. 4 - Distribuția ratei de eroare a datelor de secvențiere, LP1_AR.

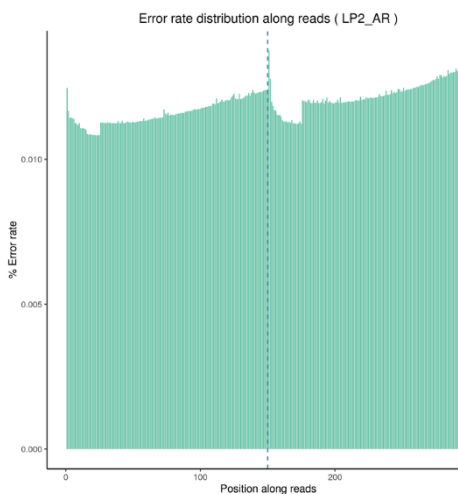


Fig. 5 - Distribuția ratei de eroare a datelor de secvențiere, LP2_AR.

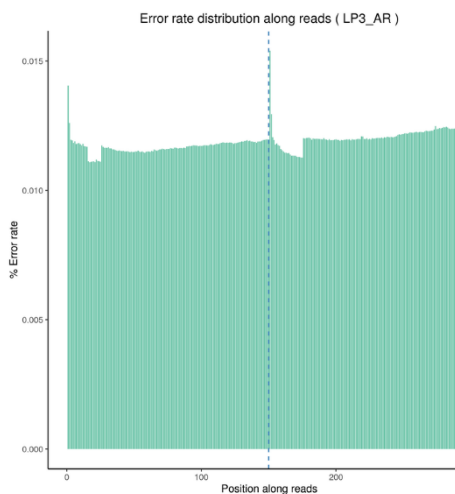


Fig. 6 - Distribuția ratei de eroare a datelor de secvențiere, LP3_AR.

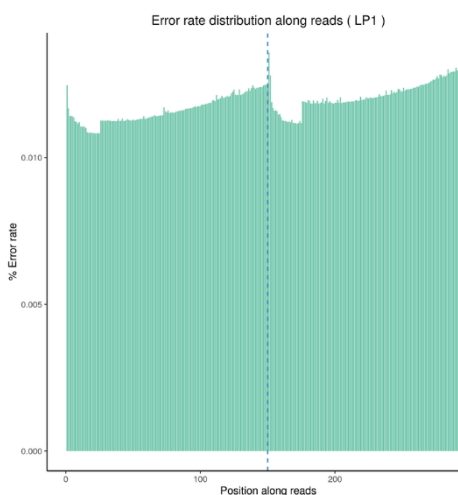


Fig. 7 - Distribuția ratei de eroare a datelor de secvențiere, LP1.

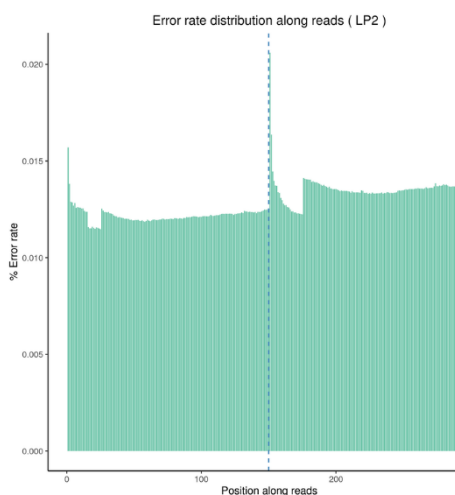


Fig. 8 - Distribuția ratei de eroare a datelor de secvențiere, LP2.

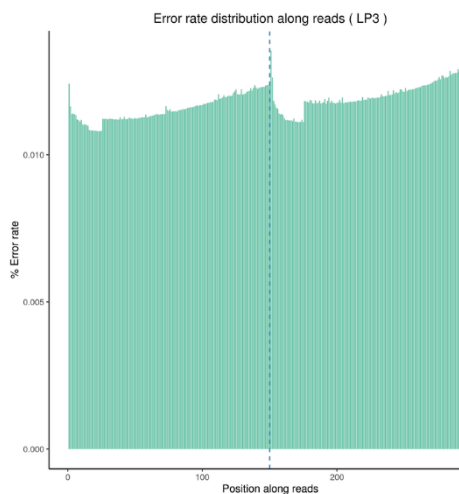


Fig. 9 - Distribuția ratei de eroare a datelor de secvențiere, LP3.

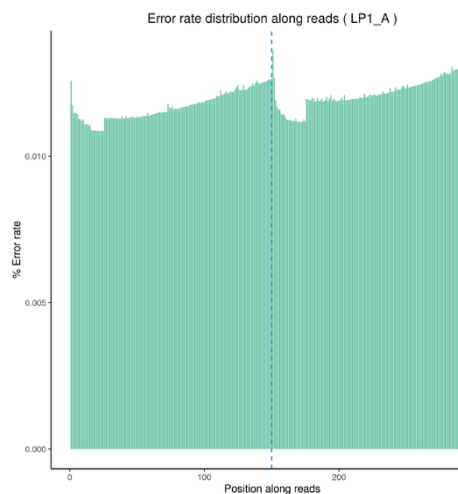


Fig. 10 - Distribuția ratei de eroare a datelor de secvențiere, LP1_A.

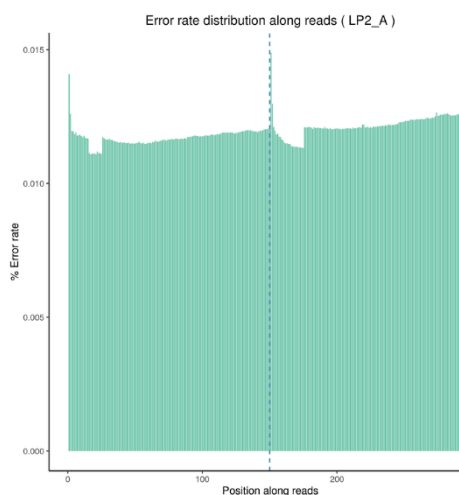


Fig. 11 - Distribuția ratei de eroare a datelor de secvențiere, LP2_A.

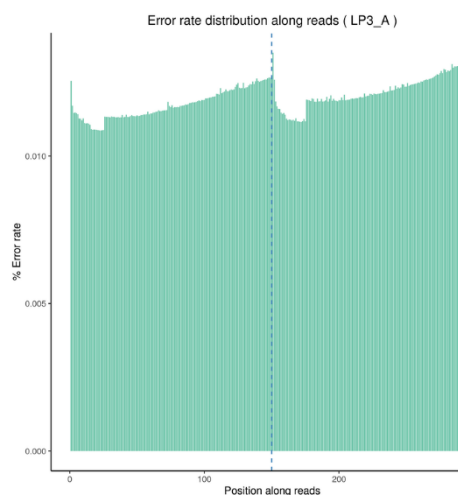


Fig. 12 - Distribuția ratei de eroare a datelor de secvențiere, LP3_A.

Distribuția conținutului GC

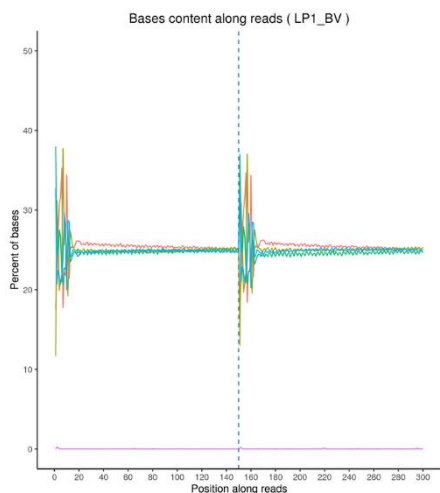


Fig. 13 - Distribuția conținutului GC, LP1_BV.

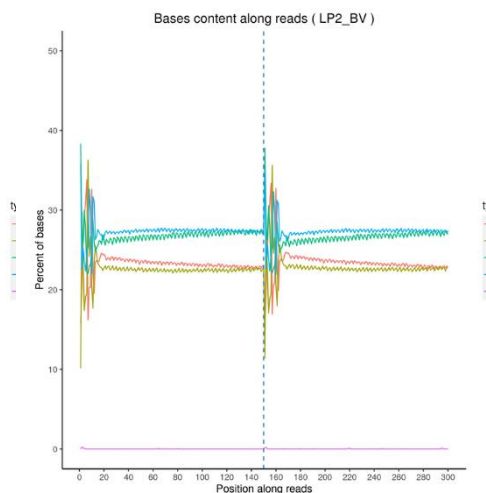


Fig. 14 - Distribuția conținutului GC, LP2_BV.

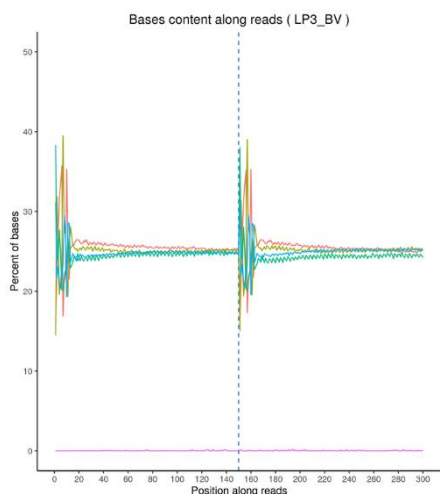


Fig. 15 - Distribuția conținutului GC, LP3_BV.

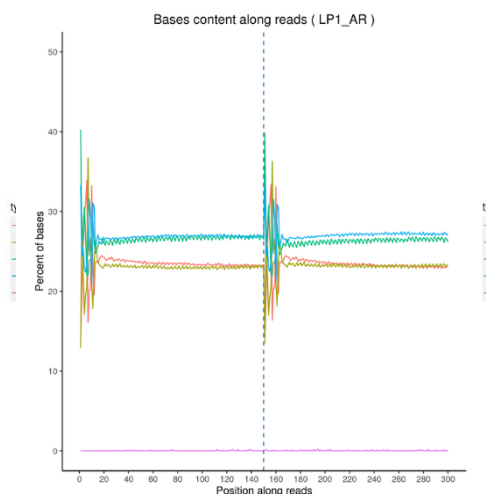


Fig. 16 - Distribuția conținutului GC, LP1_AR.

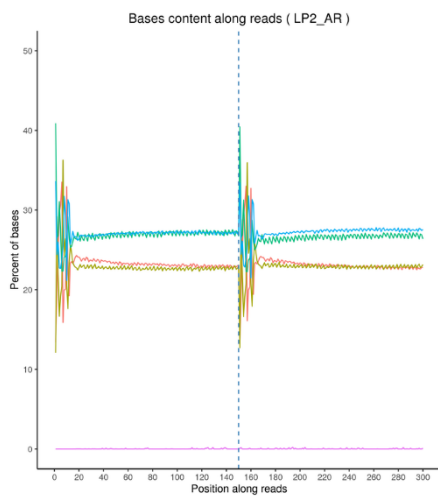


Fig. 17 - Distribuția conținutului GC, LP2_AR.

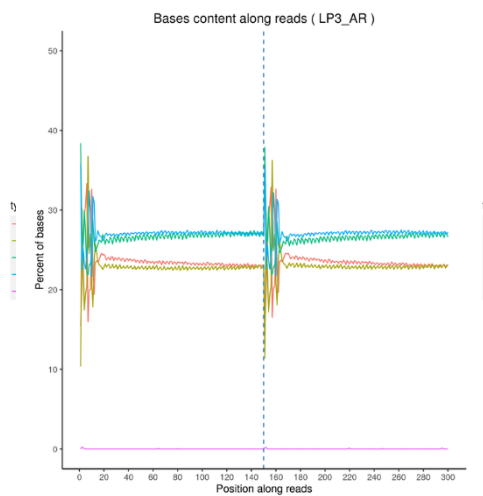


Fig. 18 - Distribuția conținutului GC, LP3_AR.

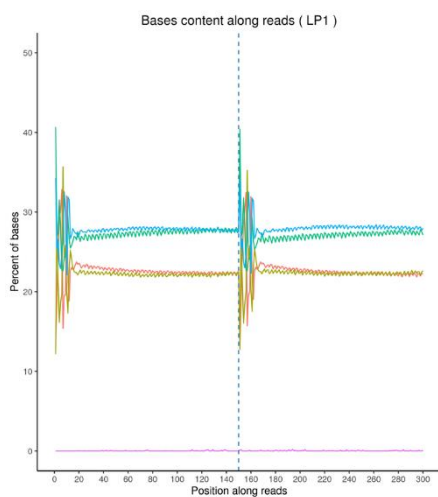


Fig. 19 - Distribuția conținutului GC, LP1.

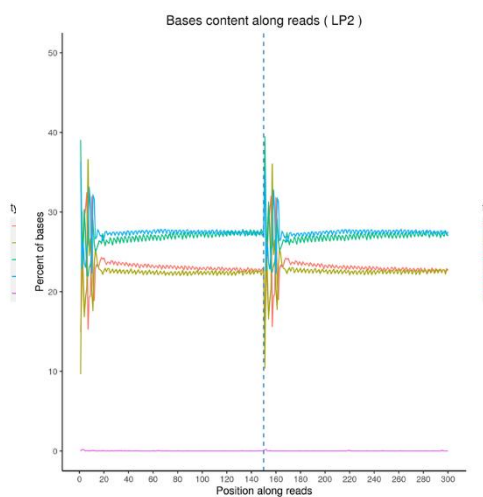


Fig. 20 - Distribuția conținutului GC, LP2.

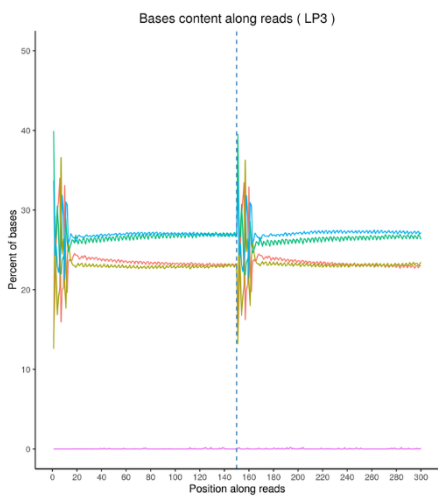


Fig. 21 - Distribuția conținutului GC, LP3.

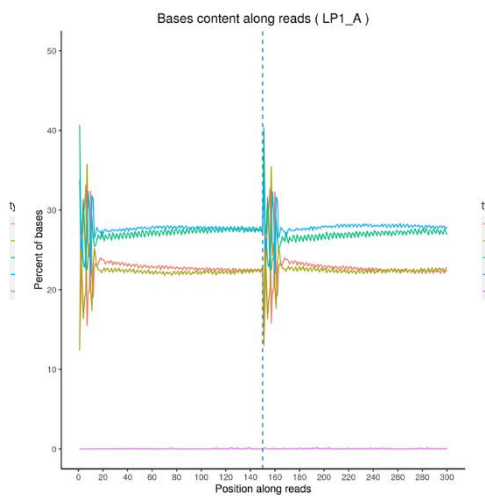


Fig. 22 - Distribuția conținutului GC, LP1_A.

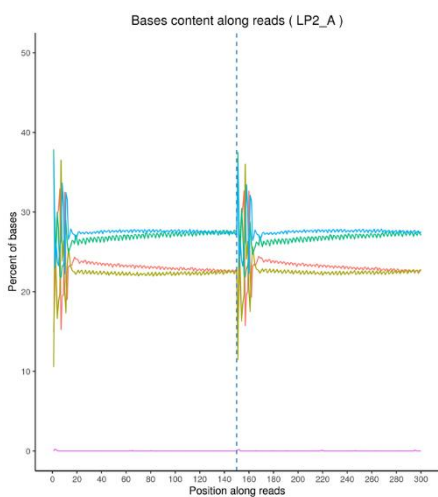


Fig. 23 - Distribuția conținutului GC, LP2_A.

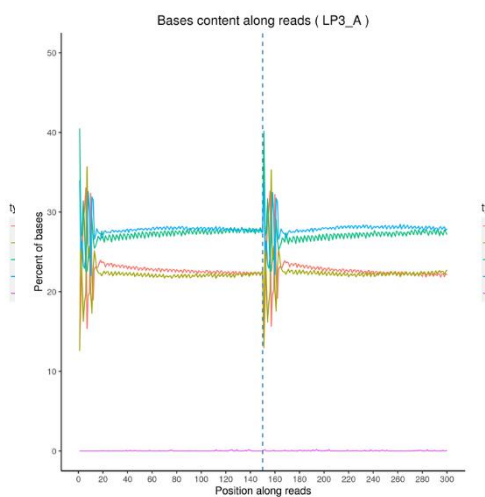
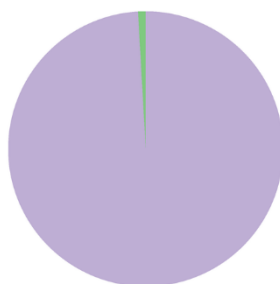


Fig. 24 - Distribuția conținutului GC, LP3_A.

Clasificarea citirilor brute - filtrarea datelor de secvențiere

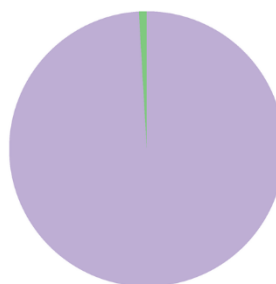
Classification of Raw Reads (LP1_BV)



■ Clean Reads (76370538, 99.11%)
■ Containing N (422, 0.00%)
■ Low Quality (40, 0.00%)
■ Adapter Related (685136, 0.89%)

Fig. 25 - Filtrarea datelor de secvențiere, LP1_BV.

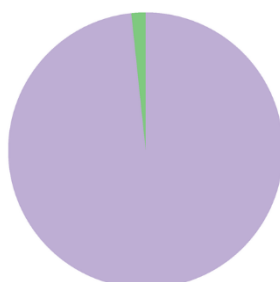
Classification of Raw Reads (LP2_BV)



■ Clean Reads (70748966, 99.10%)
■ Containing N (352, 0.00%)
■ Low Quality (58, 0.00%)
■ Adapter Related (645454, 0.90%)

Fig. 26 - Filtrarea datelor de secvențiere, LP2_BV.

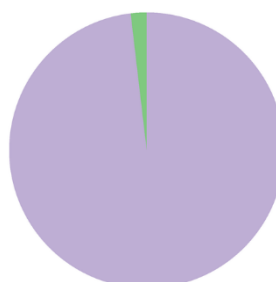
Classification of Raw Reads (LP3_BV)



■ Clean Reads (43048934, 98.32%)
■ Containing N (1128, 0.00%)
■ Low Quality (128, 0.00%)
■ Adapter Related (730972, 1.67%)

Fig. 27 - Filtrarea datelor de secvențiere, LP3_BV.

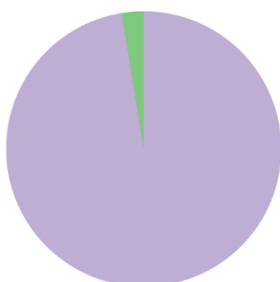
Classification of Raw Reads (LP1_AR)



■ Clean Reads (79956546, 98.13%)
■ Containing N (1788, 0.00%)
■ Low Quality (270, 0.00%)
■ Adapter Related (1517932, 1.86%)

Fig. 28 - Filtrarea datelor de secvențiere, LP1_AR.

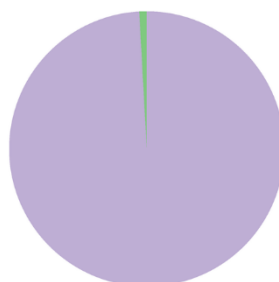
Classification of Raw Reads (LP2_AR)



Clean Reads (63862696, 97.44%)
Containing N (1432, 0.00%)
Low Quality (210, 0.00%)
Adaptor Related (167524, 2.56%)

Fig. 29 - Filtrarea datelor de secvențiere, LP2_AR.

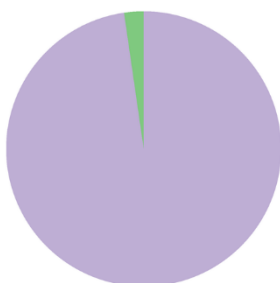
Classification of Raw Reads (LP3_AR)



Clean Reads (61651126, 99.14%)
Containing N (330, 0.00%)
Low Quality (48, 0.00%)
Adaptor Related (538774, 0.86%)

Fig. 30 - Filtrarea datelor de secvențiere, LP3_AR.

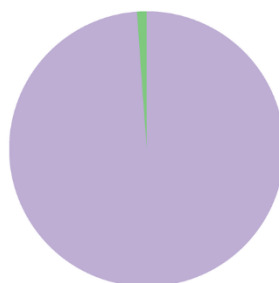
Classification of Raw Reads (LP1)



Clean Reads (71799590, 97.71%)
Containing N (1774, 0.00%)
Low Quality (264, 0.00%)
Adaptor Related (1680972, 2.29%)

Fig. 31 - Filtrarea datelor de secvențiere, LP1.

Classification of Raw Reads (LP2)



Clean Reads (63235842, 98.86%)
Containing N (340, 0.00%)
Low Quality (136, 0.00%)
Adaptor Related (727942, 1.14%)

Fig. 32 - Filtrarea datelor de secvențiere, LP2.

Classification of Raw Reads (LP3)

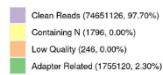
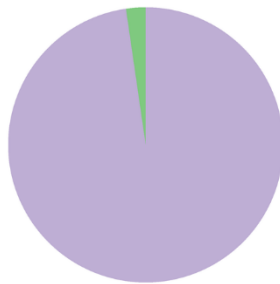


Fig. 33 - Filtrarea datelor de secvențiere, LP3.

Classification of Raw Reads (LP1_A)

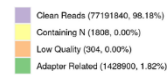
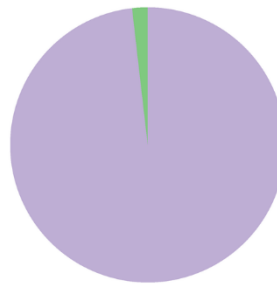


Fig. 34 - Filtrarea datelor de secvențiere, LP1_A.

Classification of Raw Reads (LP2_A)

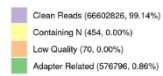
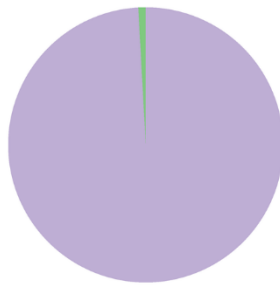


Fig. 35 - Filtrarea datelor de secvențiere, LP2_A.

Classification of Raw Reads (LP3_A)

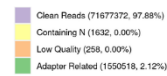
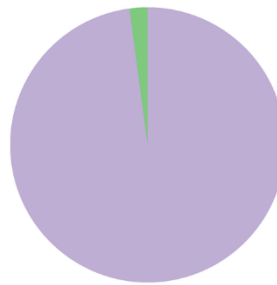
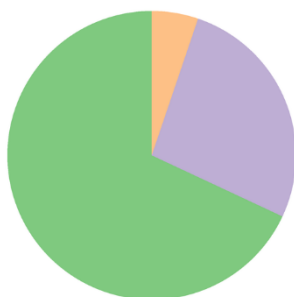


Fig. 36 - Filtrarea datelor de secvențiere, LP3_A.

Distribuția citirilor la genomul de referință

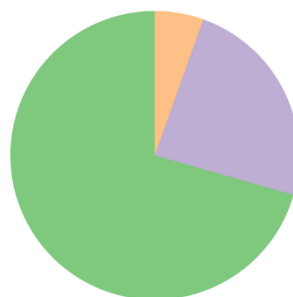
Percent of genome regions (LP1_BV)



exon (3392514752, 66.04%)
intron (259438965, 5.20%)
intergenic (1333894894, 26.75%)

Fig. 37 – Procent din regiunile genomice, LP1_BV.

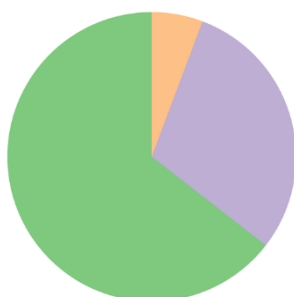
Percent of genome regions (LP2_BV)



exon (5335986075, 70.58%)
intron (413889800, 5.47%)
intergenic (1809944680, 23.94%)

Fig. 38 - Procent din regiunile genomice, LP2_BV.

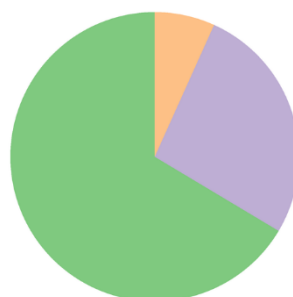
Percent of genome regions (LP3_BV)



exon (2209044629, 64.50%)
intron (205370059, 5.74%)
intergenic (1065668060, 29.76%)

Fig. 39 - Procent din regiunile genomice, LP3_BV.

Percent of genome regions (LP1_AR)



exon (5729951661, 66.43%)
intron (581896483, 6.75%)
intergenic (2313337546, 26.82%)

Fig. 40 - Procent din regiunile genomice, LP1_AR.

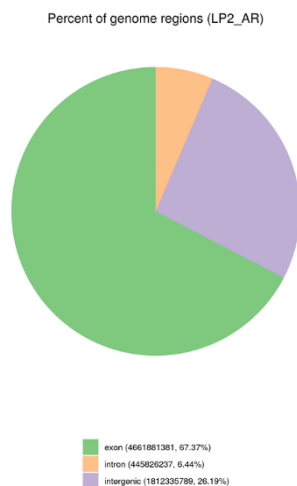


Fig. 41 - Procent din regiunile genomice, LP2_AR.

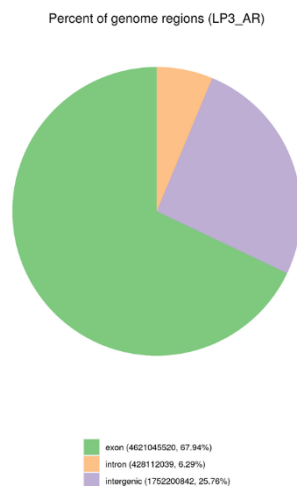


Fig. 42 - Procent din regiunile genomice, LP3_AR.

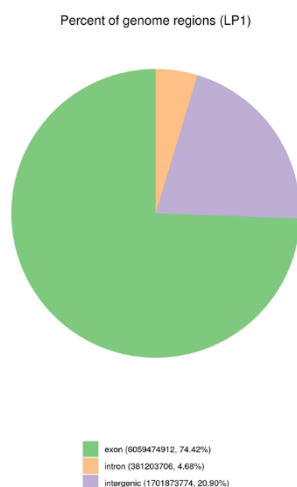


Fig. 43 - Procent din regiunile genomice, LP1.

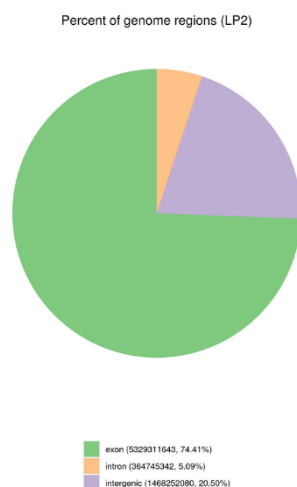


Fig. 44 - Procent din regiunile genomice, LP2.

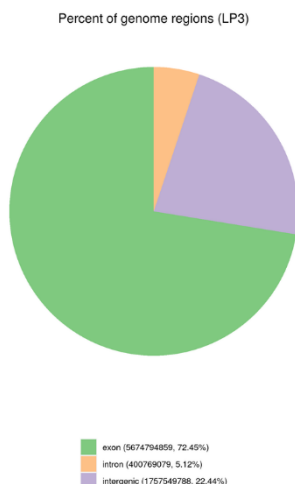


Fig. 45 - Procent din regiunile genomice, LP3.

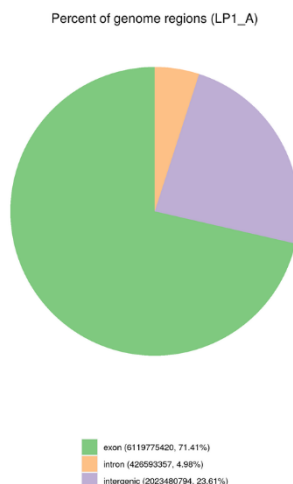


Fig. 46 - Procent din regiunile genomice, LP1_A.

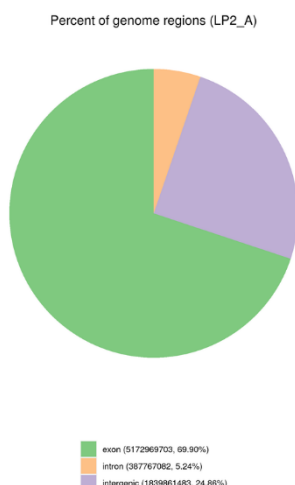


Fig. 47 - Procent din regiunile genomice, LP2_A.

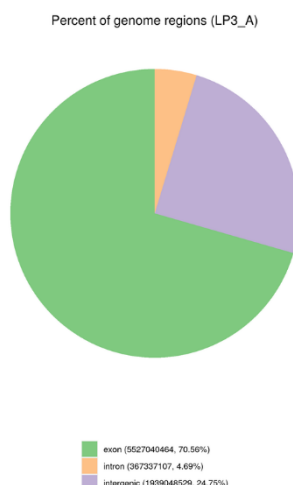


Fig. 48 - Procent din regiunile genomice, LP3_A.

Exon - numărul de baze aliniate în regiunile exonice ale genomului și proporția acestora față de numărul total de baze aliniate pe genom.

Intron - numărul de baze aliniate în regiunile intronice ale genomului și proporția acestora față de numărul total de baze aliniate pe genom.

Intergenic - numărul de baze aliniat în regiunile intergenice (spațiile dintre gene) ale genomului și proporția acestora față de numărul total de baze aliniat pe genom.

Co-expresia diagrame Venn

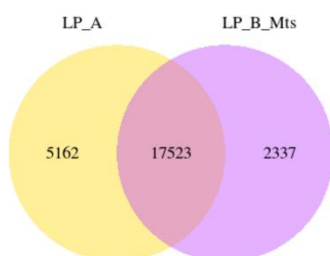


Fig. 49 – Co-expresia genică, LP_A vs. LP_B_Mts.

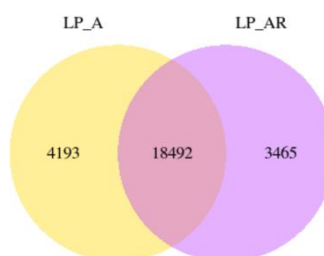


Fig. 50 - Co-expresia genică, LP_A vs. LP_AR.

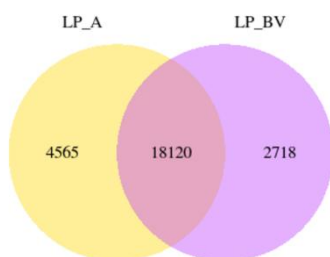


Fig. 51 - Co-expresia genică, LP_A vs. LP_BV.

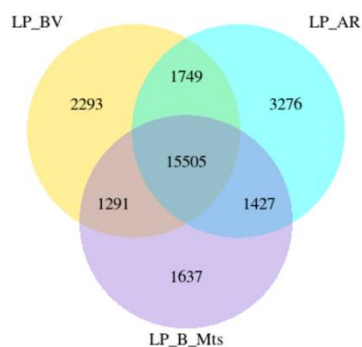


Fig. 52 - Co-expresia genică, LP_BV vs. LP_AR vs. LP_B_Mts.

Distribuția generală a genelor exprimate diferențial

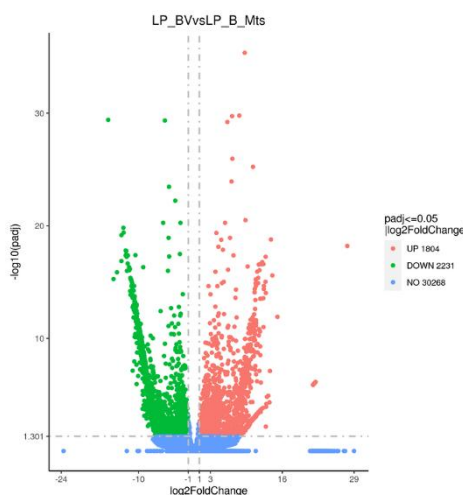


Fig. 53 – Graficul de tip «volcano» al genelor exprimate diferențiat, LP_BV vs. LP_B_Mts.

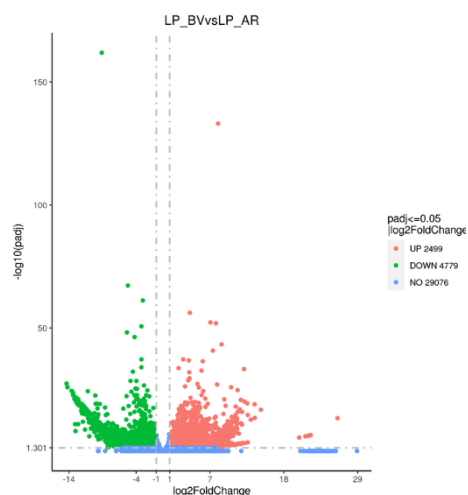


Fig. 54 - Graficul de tip «volcano» al genelor exprimate diferențiat, LP_BV vs. LP_AR.

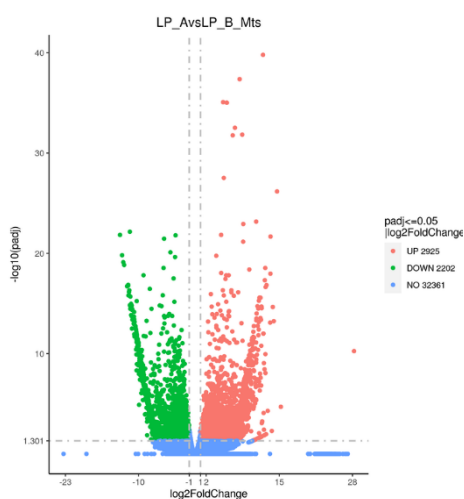


Fig. 55 - Graficul de tip «volcano» al genelor exprimate diferențiat, LP_A vs. LP_B_Mts.

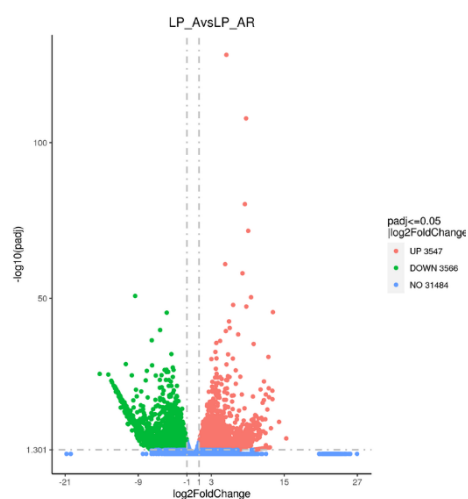


Fig. 56 - Graficul de tip «volcano» al genelor exprimate diferențiat, LP_A vs. LP_AR.

Analiza de îmbogățire GO (Gene Ontology)

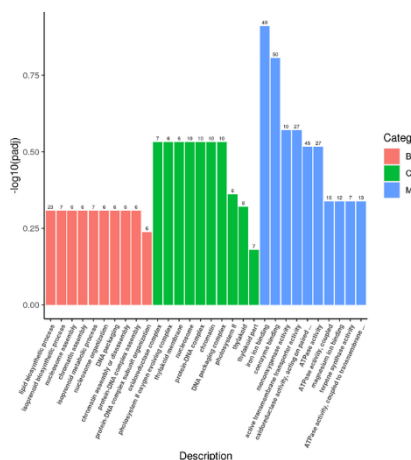


Fig. 57– Histograma analizei de îmbogățire GO, LP_BV vs. LP_B_Mts.

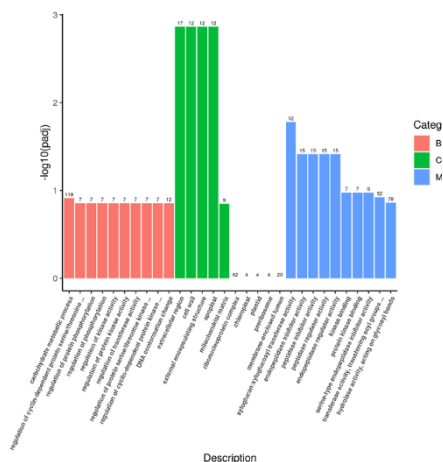


Fig. 58 - Histograma analizei de îmbogățire GO, LP_BV vs. LP_AR.

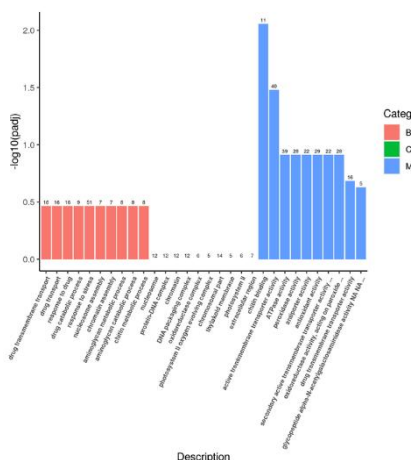


Fig. 59 - Histograma analizei de îmbogățire GO, LP_A vs. LP_B_Mts.

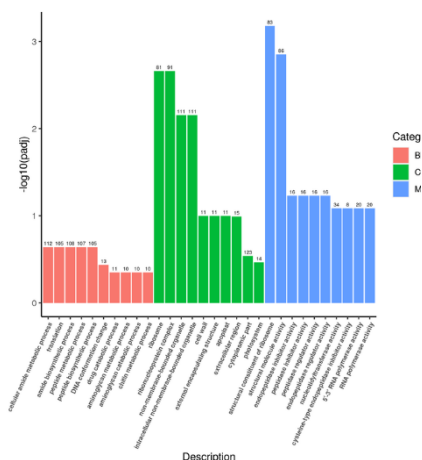


Fig. 60 - Histograma analizei de îmbogățire GO, LP_A vs. LP_AR.

Analiza de îmbogățire GO, grafic de tip Scatter Plot

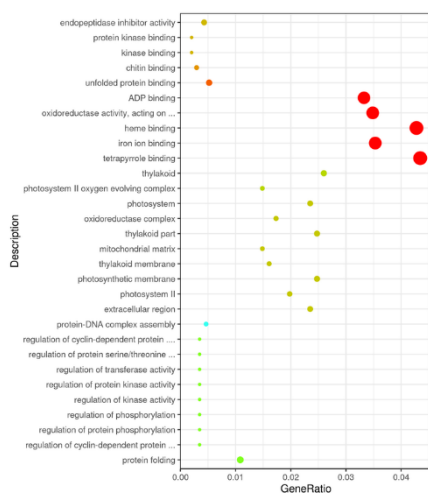


Fig. 61 – Analiza de îmbogățire, scenariu LP_ARvsLP_B_Mts.

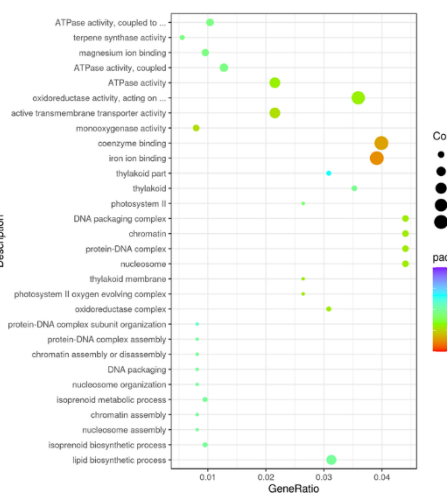


Fig. 62 - Analiza de îmbogățire, scenariu LP_BVvsLP_B_Mts.

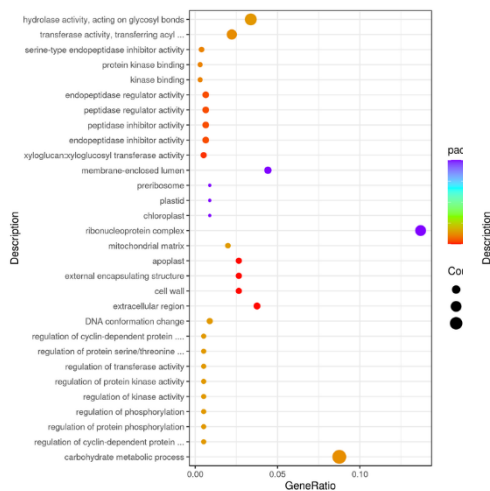


Fig. 63 - Analiza de îmbogățire, scenariu LP_BVvsLP_AR.

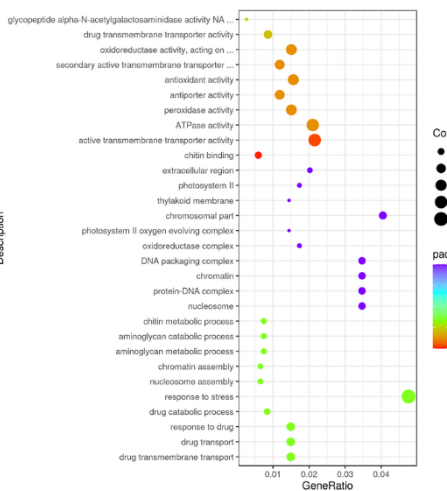


Fig. 64 - Analiza de îmbogățire, scenariu LP_AvsLP_B_Mts.

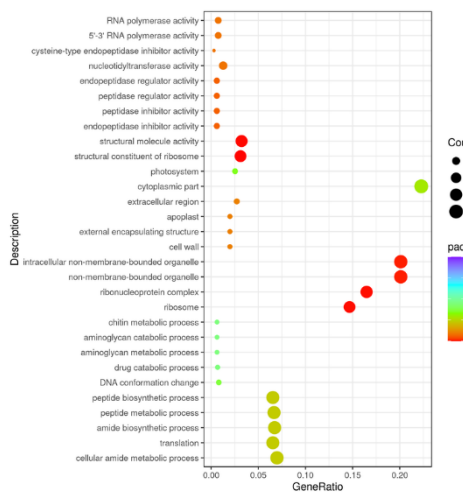


Fig. 65 - Analiza de îmbogățire,
scenariu LP_AvsLP_AR

Analiza de îmbogățire KEGG (Kyoto Encyclopedia of Genes and Genomes)

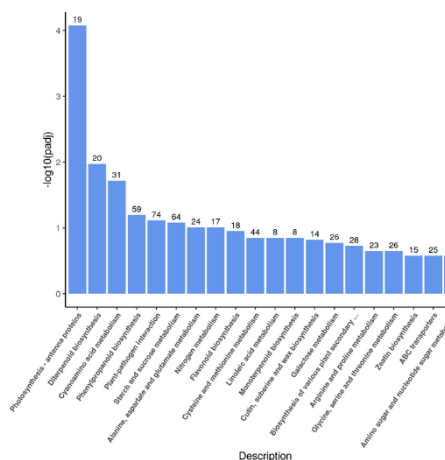


Fig. 66 – Analiza de îmbogățire, scenariu LP_ARvsLP_B_Mts.

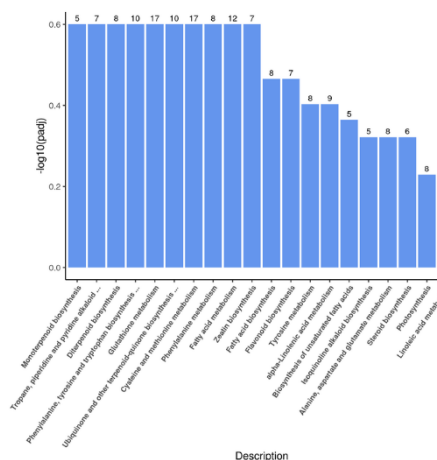


Fig. 67 - Analiza de îmbogățire, scenariu LP_BVvsLP_B_Mts.

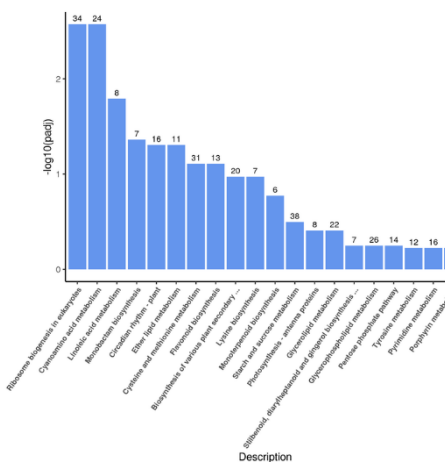


Fig. 68 - Analiza de îmbogățire, scenariu LP_BVvsLP_AR.

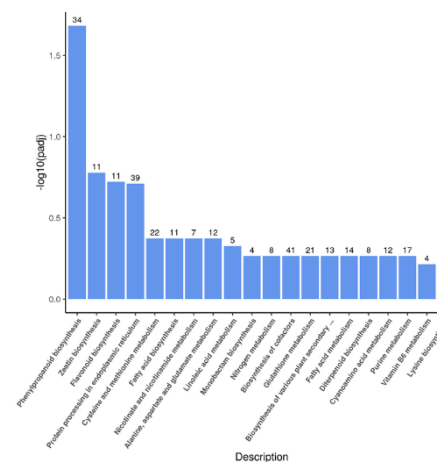


Fig. 69 - Analiza de îmbogățire, scenariu LP_AvsLP_B_Mts.

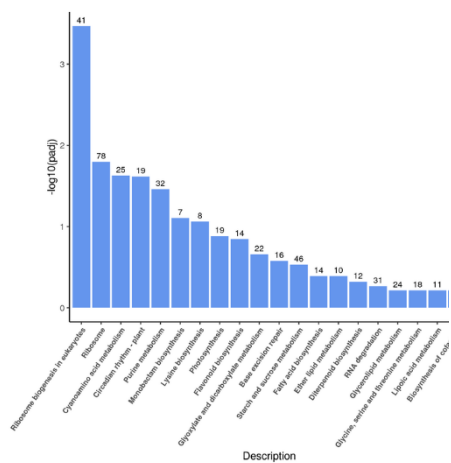


Fig. 70 - Analiza de îmbogățire,
scenariu LP_AvsLP_AR

Analiza de îmbogățire KEGG, grafic de tip Scatter Plot

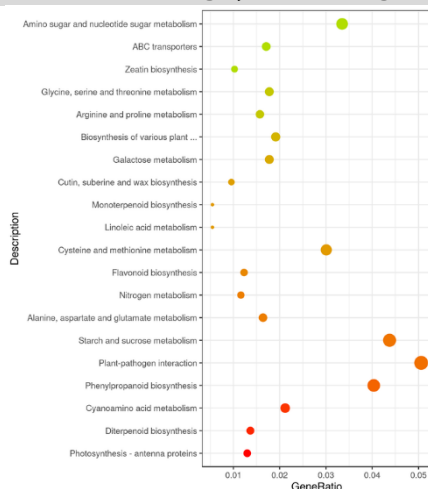


Fig. 71 – Analiza de îmbogățire, scenariu LP_ARvsLP_B_Mts.

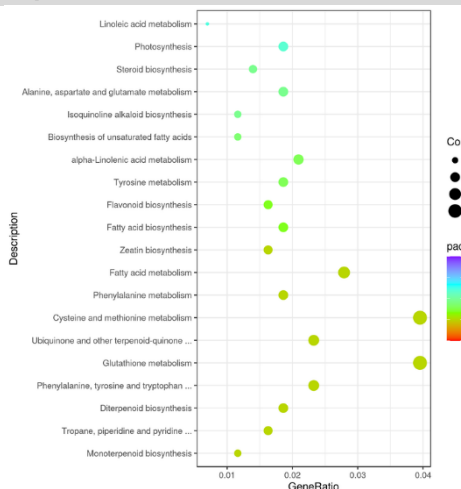


Fig. 72 - Analiza de îmbogățire, scenariu LP_BVvsLP_B_Mts.

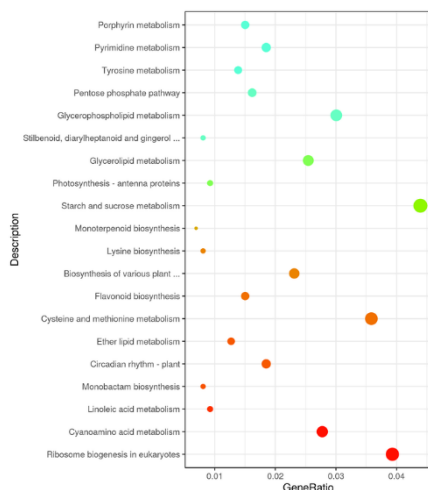


Fig. 73 - Analiza de îmbogățire, scenariu LP_BVvsLP_AR.

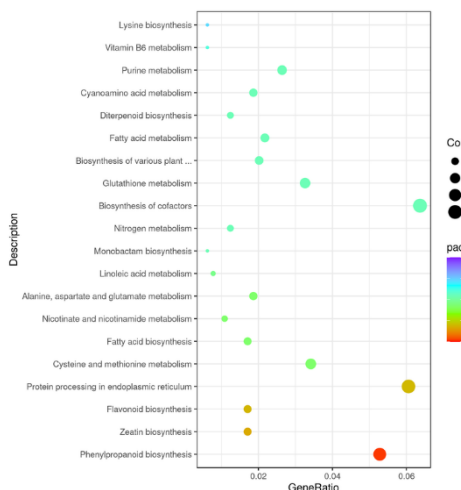
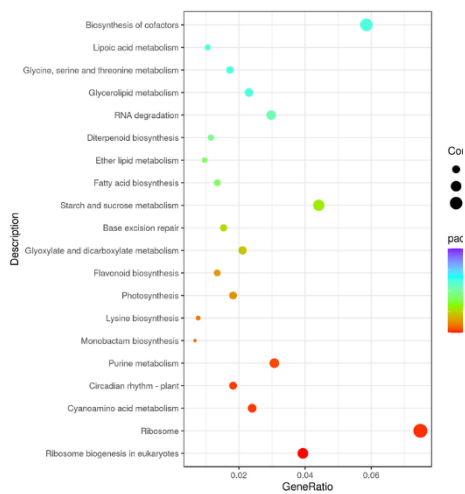


Fig. 74 - Analiza de îmbogățire, scenariu LP_AvsLP_B_Mts.



**Fig. 75 - Analiza de îmbogățire,
scenariu LP_AvsLP_AR**

Splicing/matisare alternativă

CM033338.2:19537877:19537963:-@CM033338.2:19537199:19537299:-@CM033338.2:19537028:1

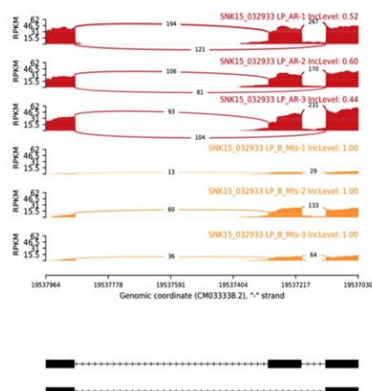


Fig. 76 – Evenimentul de tip SE2 (Skipped Exon) diagrama tip Sashimi plot.

CM033335.2:16777182:16777280:+@CM033335.2:16777333:16777431:+@CM033335.2:16

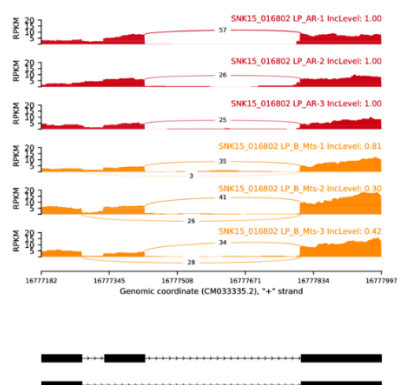


Fig. 77 - Evenimentul de tip SE3 (Skipped Exon) diagrama tip Sashimi plot.

CM033333.2:20351967:20352027:+@CM033333.2:20352294:20352443:+@CM033333.2:20

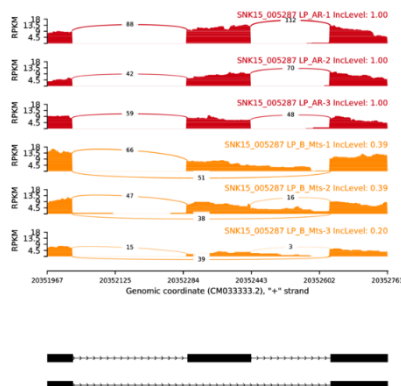


Fig. 78 - Evenimentul de tip SE1 (Skipped Exon) diagrama tip Sashimi plot.

CM033334.2:10426980:10427080:+@CM033334.2:10427320:10427378:+@CM033334.2:10

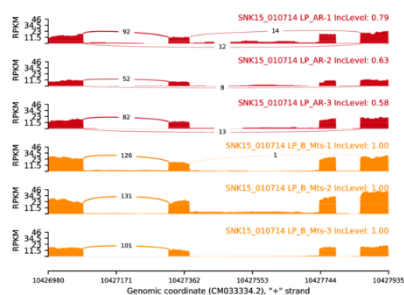


Fig. 79 - Evenimentul de tip SE5 (Skipped Exon) diagrama tip Sashimi plot.

CM033336.2:20441120:20441248:+@CM033336.2:20441891:20442043:+@CM033336.2:20

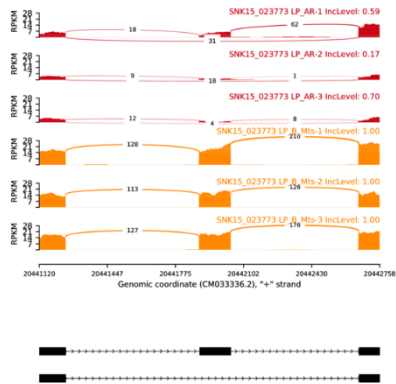


Fig. 80 - Evenimentul de tip SE4
(Skipped Exon) diagrama tip
Sashimi plot.

Analiza SNP&InDel - statistica variabilității la nivelul sit – urilor.

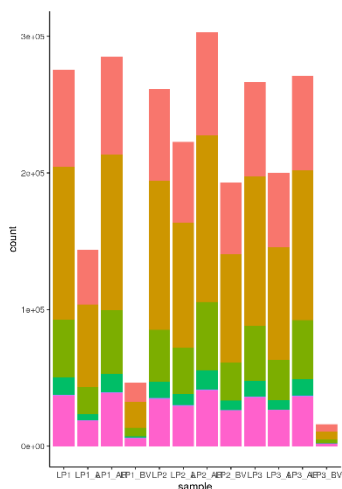


Fig. 81 – Statistica siturilor de variabilitate, pe regiuni.

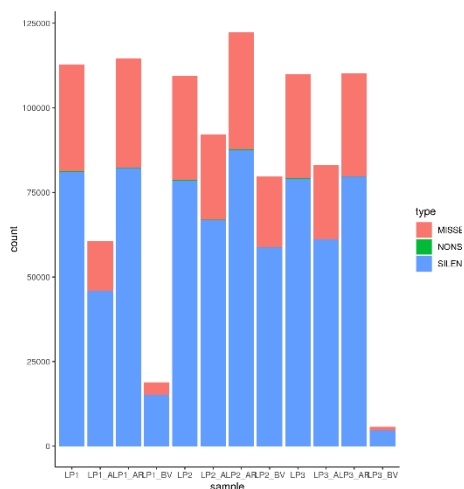


Fig. 82 - Statistica siturilor de variabilitate, raportat la funcție.

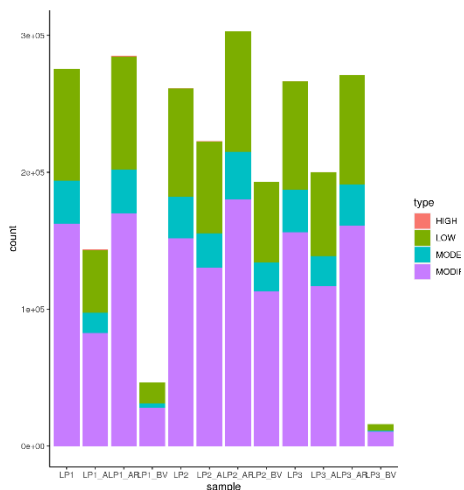


Fig. 83 - Statistica siturilor de variabilitate, în relație cu impactul asupra expresiei.

V. ANEXA 2

Programe utilizate în analiza transcriptomică

Analiza	Software	Versiune	Parametru	Observații
Aliniere	hisat2	2.0.5	Setări predefinite	Aliniere la o referință
Asamblare	Stringtie	1.3.3b	Setări predefinite	
Cuantificare	featureCounts	1.5.0-p3	Setări predefinite	
Analiza diferențială	DESeq2	1.20.0	$ \log_2(\text{FoldChange}) \geq 1$ & $\text{padj} \leq 0.05$	Pentru probe cu bio replicate
	edgeR	3.22.5	$ \log_2(\text{FoldChange}) \geq 1$ & $\text{padj} \leq 0.05$	Pentru probe fără bio replicate
Analiza de îmbogățire	clusterProfiler	3.8.1	$\text{padj} < 0.05$	Pentru analiza de îmbunătățire GO, KEGG
Analiza GSEA	gsea	v4.3.2	Setări predefinite	
Analiza Interacțiune proteină-proteină	diamond	0.9.14	e-value = $1e-10$	Utilizate blast, baza de date String.

Matisare/spl icing alternativ	rMATS	4.1.0	Setări predefinite	
Analiza SNP/InDel	GATK	4.1.4.1	MQ < 40.0 și QD < 2.0 și FS > 30.0 și DP < 10 și QUAL < 20	Detecția variantelo r SNP/InDe l
	snpEff	4.3q	Setări predefinite	Anotarea variantelo r de tip SNP/InDe l

Bibliografie

1. Anders S, Huber W. Differential expression analysis for sequence count data[J]. *Genome biol*, 2010, 11(10): R106.
2. Armstead, I. P., Turner, L. B., Marshall, A. H., Humphreys, M. O., King, I. P., & Thorogood, D. (2008). Identifying genetic components controlling fertility in the outcrossing grass species perennial ryegrass (*Lolium perenne*) by quantitative trait loci analysis and comparative genetics. *New Phytologist*, 178(3), 559–571. <https://doi.org/10.1111/j.1469-8137.2008.02380.x>;
3. Atanasiu L., 1984, *Ecofiziologia plantelor*, Ed științifică și enciclopedică, București, pp 111-160;
4. Breazu, i., Grusea A., 1997, Aspecte privind ameliorarea gramineelor perene de pajisti la ICPCP Brasov, *Lucrări științifice ale ICPC – Brasov*, Volum jubiliar 50 de ani de cercetare in domeniul culturii pajistilor: 33-48;
5. Breazu, I., Oprea, G., Chiper, C., 1987, Contribuția trifoiului alb la aprovizionarea pajiștilor cu azot, *Lucrări științifice ale ICPC – Brasov*, vol XII:67-83;
6. Bray N, Pimentel H, Melsted P, et al. Near-optimal RNA-Seq quantification[J]. *arXiv preprint arXiv:1505.02710*, 2015.
7. Buxton, D.R., Marton, G.C., 1989, Forage quality pf plant parts of perennial grasses and relationship to phenology, *Crop Science* 29:429-435;
8. Byrne, S., Asp, T., & Jensen, C. S. (2025). Genomic prediction of agronomic traits in perennial ryegrass. *bioRxiv*. <https://doi.org/10.1101/2025.03.18.642199>
9. Casler, M.,D., Eduard von Sante, 2010, *Breeding Objectives in Forages, Fodder Crops and Amenity Grasses*, Springer:115-136;
10. Chen, Y., et al. (2024). An improved chromosome-level genome assembly of *Lolium* spp. *Gigabyte Journal*. <https://doi.org/10.46471/gigabyte.112>;
11. Conaghan, P., Casler, M., D., 2011, A theoretical and practical analysis of the optimum breeding system for perennial ryegrass, *Irish Journal of Agricultural and Food Research* 50: 47-63;
12. Dologa G, Balan, M., 1997, Resurse genetice de graminee si leguminoase perene de pajisti, *Lucrări științifice ale ICPC – Brasov*, Volum jubiliar 50 de ani de cercetare in domeniul culturii pajistilor: 25- 32;
13. Dracatos, P. M., Cogan, N. O., Sawbridge, T. I., Gendall, A. R., Smith, K. F., Spangenberg, G. C., & Forster, J. W. (2009). Molecular characterisation and genetic mapping of candidate genes for qualitative disease resistance in perennial ryegrass (*Lolium perenne* L.). *BMC Plant Biology*, 9, 62. <https://doi.org/10.1186/1471-2229-9-62>;
14. Dragomir Neculai, Vasile Adrian Blaj, Teodor Marușca, Vasile Mocanu, Marinela Horablaga, Dorin Rechițean, Marcela Dragoș, Monica Tod, Tudor

- Ene 2018, Pajiștile de câmpie – cauzele degradării vegetației și măsuri tehnologice de îmbunătățire, Editura Artpress, nr.pagini 60;
15. Dragomir, N., Breazu, I., Grusea, A., Oros, A., 1997 Cercetări și realizări în domeniul ameliorării leguminoaselor perene de pajiști la ICPCP Brasov, Lucrări științifice ale ICPCP – Brasov, Volum jubiliar 50 de ani de cercetare în domeniul culturii pajistilor:49-55.
 16. Dragoș, M. M. (2019). Erosion of genetic resources in perennial ryegrass (*Lolium perenne* L.). Research Journal of Agricultural Science, 51(1), 55–61.
Disponibil la:
<https://www.cabidigitallibrary.org/doi/pdf/10.5555/20203389923>
 17. Elbersen, H.W.,1996, Growth and water relations of field-grown tall fescue as influenced by drought and endophyte, Grass and Forage Science 51 (4), 333-342
 18. Elgersma, A., 1990, Spaced-plant traits related to seed yield in plots of perennial ryegrass (*Lolium perenne* L.) Euphytica 51: 151-161.
 19. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science. 2004;306(5696):636-640.
 20. Garber M, Grabherr M G, Guttman M, et al. Computational methods for transcriptome annotation and quantification using RNA-seq[J]. Nature methods, 2011, 8(6): 469-4.Order_AS.
 21. Gay A.,P., 1994, Breeding for leaf water conductance, its heritability and its effect on water use in *Lolium perenne*, Aspects of Applied Biology 38: 41-46
 22. Goldstein L D , Cao Y , Pau G , et al. Prediction and Quantification of Splice Events from RNA-Seq Data.[J]. Plos One, 2016, 11(5):e0156132.
 23. Guțu A.; Ababii A., ; Andreoiu A., Tod M. Doroftei V. ; Nazare A. ; Țiței V. ; Gadibadi M.,2025, The evaluation of the biomass quality of *Lolium perenne* ‘Măgura’ and *Phleum pratense* ‘Tirom’ grown under the climatic conditions of Moldova , the Agriculture For Life International Conference- Bucuresti, BOOK OF ABSTRACTS Section 1: AGRONOMY, pp.245.
 24. He Z , Zhao X , Lu Z , et al. Comparative transcriptome and gene co-expression network analysis reveal genes and signaling pathways adaptively responsive to varied adverse stresses in the insect fungal pathogen, *Beauveria bassiana*[J]. Journal of Invertebrate Pathology, 2017:S0022201117304391.
 25. Jensen K., B., Assaz, K., H., Waldron, B., L., 2001, Dry Matter Production of Orchardgrass and Perennial Ryegrass at Five Irrigation Levels, Crop Sci, 41, pp479-487.
 26. Jităreanu, G., Săulescu, N. N., & Vinătoru, C. (2014). Evaluation and characterization of perennial ryegrass (*Lolium perenne* L.) genotypes collected from natural flora. Annals of NARDI Fundulea, 82, 31–39,
 27. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes[J]. Nucleic acids research, 2000, 28(1): 27-30.(KEGG)

28. Katz Y, Wang E T, Airoidi E M, et al. Analysis and design of RNA sequencing experiments for identifying isoform regulation[J]. Nature methods, 2010, 7(12): 1009-1015.
29. Kovacs, S., J., 1982, Germoplasmă autohtonă la gramineele perene de păjiști, Lucrări științifice ICPCP Brașov, vol. VIII
30. Liao Y1, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014 ,30(7):923-30.(featureCounts)
31. Love M I, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2[J]. Genome biology, 2014, 15(12): 1-21.(DESeq2)
32. Malinowska, M., et al. (2025). The value of early root development traits in breeding programs for biomass yield in perennial ryegrass (*Lolium perenne* L.). *Theoretical and Applied Genetics*, 138(3), 489–502. <https://doi.org/10.1007/s00122-024-04797-5>
33. Marușca T. 2016, Praticultura pe înțelesul tuturor, Tipografia Coprint, pp. 98-99
34. Marușca T., Mocanu V, Haș E. , Monica Alexandrina Tod, Andreea Andreoiu, Marcela Dragoș, Blaj V. , Ene T., Doina Silistru, E. Ichim, P. Zevedei, C. Constantinescu, S. Tod 2014, Ghid de întocmire a amenajamentelor pastorale, Editura Capolavoro, Brașov, ISBN 978-973-98711-8-1, 179-195;
35. Marușca T., Tod Monica, Silistru Doina, Dragomir N., Maria Schitea, 2011, Principalele soiuri de graminee și leguminoase perene de păjiști, Ed Capolavoro, Brașov
36. Marușca T., Tod Monica, Zevedei P., 2016 Varieties of perennial grasses and legumes made in research and development institute for grassland Brașov, Romanian Journal of Grassland and Forage Crops, no.5, The Romanian Society for Grassland, Cluj Napoca
37. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. Genome research, 2010, 20(9): 1297-1303.(GATK)
38. Mihaela Perteau, Geo M Perteau, Corina M Antonescu1, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads[J]. *Nat Biotechnol*. 2015 March ; 33(3):290-295(StringTie)
39. Mocanu, V., Dragomir, N., Blaj, V.A., Ene, T.A., Tod, M., A., Mocanu, V., 2021, Pajiștile României, Resurse, strategii, de îmbunătățire și valorificare, 129-133;
40. Moga, I., Schitea, M., Mateias, M., 1986, Plante furajere, Ed Ceres.
41. Moga, I., Schitrea, M., 2000, Cultura plantelor furajere pentru sămânță;
42. Mortazavi A, Williams B A, McCue K, et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq[J]. Nature methods, 2008, 5(7): 621-628.

43. Munns R. 2002. Comparative physiology of salt and water stress. *Plant, Cell and Environment* 25: 239-250.
44. Oprea, G., Razec, I., 1997 , Cercetări privind calitatea furajelor de pajisti, Lucrari stiintifice ale ICPC – Brasov, Volum jubiliar 50 de ani de cercetare in domeniul culturii pajistilor: 133-146
45. Parkhomchuk D, Borodina T, Amstislavskiy V, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA[J]. *Nucleic acids research*, 2009, 37(18): e123-e123.
46. Patro R, Mount S M, Kingsford C. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms[J]. *Nature biotechnology*, 2014, 32(5): 462-464.
47. Reheul, D., Benny de Cauver, Cougnou, M., 2010, The role of forage crops in multifunctional agriculture, Fodder crops and amenity grasses, Springer, 1-12.
48. Robinson M D, McCarthy D J, Smyth G K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data[J]. *Bioinformatics*, 2010, 26(1): 139-140.(edgeR)
49. Sampoux, J. P., Barre, P., & Studer, B. (2024). Genetic variation for seed retention in accessions and populations of *Lolium perenne*. *Crop Science*, 64(3), 912–925. <https://doi.org/10.1002/csc2.20837>
50. Savatti, M., Nedelea, G., Ardelean M., 2004, Tratat de ameliorarea plantelor, Ed. Marineasa, Timisoara.
51. Shen S., Park JW., Lu ZX., Lin L., Henry MD., Wu YN., Zhou Q., Xing Y. rMATs: Robust and Flexible Detection of Differential Alternative Splicing from Replicate RNA-Seq Data.(rMATs).
52. Sojka, R., E., 1984, Field Evaluation of drought response in small-grain cereals. In *Progress in Plant Breeding*, G. E. Russel (ed), pp: 165-191
53. Soussana. J.,F., Luscher, A., 2007, Temperate grasslands and global atmospheric change: a review, *Grass and Forage Science*, vol 62(2), 127-134.
54. Studer, B., et al. (2021). Identification of candidate genes for self-compatibility in perennial ryegrass (*Lolium perenne* L.). *Frontiers in Plant Science*, 12, 707901. <https://doi.org/10.3389/fpls.2021.707901>
55. Tod M., Balan, M., Zevedei, P., Andreoiu, A., Ene, T., Taulescu, E., 2020, Effect of pH Medium on Germination and Seedling Growing on Some Perennial Grasses *Journal of Mountain Agriculture on the Balkans* 23 (2)
56. Tod M., Bălan M., Andreoiu A., Zevedei P, 2024, Evaluarea unor noi surse de germoplasmă la unele specii de graminee perene de pajiști in vederea includerii in programul de ameliorare , The International Scientific Conference „Genetics, Physiology and Plant Breeding”, 7-8 octombrie, Chisinau, electronic version of the Proceeding materials ,pp 430-437, <https://doi.org/10.53040/gppb8.2024.75>

57. Tod M., Bălan M., Țiței V.,, Guțu A., 2025, Evaluation of meadow fescue (*Festuca pratensis* Huds.) germplasm for breeding purposes, Rumanian Jurnal of Grassland and Forrage Crops (2025), no 31, Cluj-Napoca, pp.77-84. https://sropaj.ro/documente/ro/revista/RJGFC_31.pdf
58. Tod M., Bălan M., Nițu S., Zevedei P., 2025, Evaluation Of Tall Fescue (*Festuca arundinacea*) Genetic Resources For Breeding Activity, Agriculture For Life International Conference Bucuresti, Scientific Papers. Series A. Agronomy, Vol. LXVIII, No. 1, 2025 ISSN 2285-5785; ISSN CD-ROM 2285-5793; ISSN Online 2285-5807; ISSN-L 2285-5785, pp.655-660.
59. Tod Monica, Blaj V.A., Marusca T., Mocanu V., 2015 The romanian varieties of perennial grasses productivity and quality in terms of climatic changes. Journal of Mountains Agriculture on the Balkans, Vol 18 , no.1, Conferince, Rimsa, Troyan , Bulgaria, pp.101-111
60. Tod Monica, Marusca T., Mocanu V., Ciopata Cristina, 2012, The Research Concerning Red Clover Overseeding of Some Perennial Grassland on Brașov Depression, Romanian Journal of Grassland and Forage Crops, no.5, The Romanian Society for Grassland, Cluj Napoca
61. Tod, M., Anghelus-Olenici, G., Balan, M., Andreoiu., A., Zevedei, P., 2021, Effect of salinity on seed germination and seedling growing on some perennial grasses, Journal of Mountain Agriculture on the Balkans.
62. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-515.
63. Varga, P., Moisuc, A., Savatti, M., Schitea M., Olaru, C., Dragomir, N., Savatti jr M., 1998 Ameliorarea plantelor furajere și producerea semințelor, ed. Lumina, pp. 206-282.
64. Vogel, K., P., Pedersen J., F., 1993, Breeding Systems for Cross-Pollinated Perennial Grasses, , Plant Breeding reviews, vol. 11:251-274
65. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics[J]. Nature Reviews Genetics, 2009, 10(1): 57-63.
66. Yan L, Yang M, Guo H, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nat Struct Mol Biol. 2013;20(9):1131-1139.
67. Young M D, Wakefield M J, Smyth G K, et al. Method Gene ontology analysis for RNA-seq: accounting for selection bias[J]. Genome Biol, 2010, 11: R14.
68. Zhu, J., et al. 2025, Estimation of ryegrass (*Lolium*) dry matter yield using genomic prediction considering genotype by environment interaction across south-eastern Australia. Frontiers in Plant Science. <https://doi.org/10.3389/fpls.2025.1579376>



Absolventă a Liceului Teoretic „Liviu Rebreanu”, din Târgu-Lăpuș, specializarea matematică-informatică (2003), apoi a Facultății de Agricultură, Specializarea Biologie – Agricultură, din cadrul Universității de Științe Agricole și Medicină Veterinară a Banatului Timișoara (2007); Master specializarea – Biologie aplicată în agricultură (2010) și Teza de doctorat (2008-2011), sub îndrumarea Prof. dr. Alexandru Moiscu (USAMBVT); Doctor în agronomie (2011). Stagiul de pregătire de 8 luni (2010) la Institutul de Montanologie și Creșterea Animalelor – RIMSABulgaria, Departamentul de Cultura plantelor furajere, sub îndrumarea domnului prof.dr. Tsvetoslav Mihovsky. Asistent cercetare (2007), Cercetător Științific (2010), Cercetător Științific gr.III (2013), Șef laborator (2020) în cadrul Institutului de Cercetare-Dezvoltare pentru Pajiști Brașov, în cadrul laboratorului de Resurse Genetice și Ameliorare, Director Adjunct Științific din 2024; A publicat: 7 cărți ca și coautor, 61 lucrări științifice indexate în diferite baze de date, din care 17 în calitate de prim autor, 44 – coautor, și 3 lucrări științifice indexate în Web of Science de Clarivate. Este autoare a 5 soiuri de graminee perene de pajiști și anume la 2 coautor : ierbăluță - Minier (*Phalaris arundinacea* L) (2022), timoftică-Carpatica (*Phleum pratense* L) (2022); 3 autor principal: raigras peren - Brevis (*Lolium perenne* L) (2025), păiuș înalt- Bârsa (*Festuca arundinacea* Schreb.) (2025), timoftică-Dona (*Phleum pratense* L) (2026).

Absolvent al Liceului „Emil Racoviță” Baia Mare (1993), apoi al Facultății de Biologie și Geologie, Universitatea Babeș-Bolyai (UBB) Cluj-Napoca (2002); Master 2005 (Biotransformare celulară); Teză de doctorat (2005-2010) cu Prof. Emerit Constantin Crăciun (Centrul de Microscopie Electronică) și Prof. Rakosy Laszlo, UBB; Doctor în Biologie (2010); Specializare în Microscopie Electronică (ME) (Curs Tecna), Curs de Tomografie Electronică, ME de Transmisie și ME de Scanare la Academia FEL, FEL NanoPort, Eindhoven, Olanda (2011) și specializare postdoctorală în caracterile ultrastructurale ale dermatofitelor (la Academia Română, Centrul de Studii și Cercetări pentru Diversitate Agrobiologică „Acad. David Davidescu”, 2012); Cursuri la Școala Avansată Comună, „Receptori și „Semnalizare” (2016), Insula Spetses, Grecia, sub auspiciile Uniunii Internaționale de Biochimie și Biologie Moleculară, Uniunii Internaționale pentru Biofizică Pură și Aplicată și Uniunii Internaționale de Științe Fiziologice; din 2024 Cercetător Științific II în cadrul Stațiunii de Cercetare și Dezvoltare pentru Creșterea Bovinelor din Arad, Laboratorul de Cercetare, Compartimentul Genetică, Biologie Moleculară și Bioinformatică; a publicat 41 de lucrări indexate în Web of Science de Clarivate Analytics și 34 de lucrări indexate în diferite baze de date internaționale; 804 de citări (conform Web of Science, 2026) și 1432 citări (conform Google scholar).



**Institutul de Cercetare - Dezvoltare
pentru Pajiști Brașov**



**Stațiunea de Cercetare-Dezvoltare
pentru Creșterea Bovinelor Arad**

Această lucrare a fost realizată cu sprijinul Ministerului Agriculturii și Dezvoltării Rurale (MADR) și al Academiei de Științe Agricole și Silvici (ASAS), în cadrul proiectului de cercetare ADER 15.1.1/18.07.2023: „Cercetări privind obținerea de noi soiuri de plante furajere la speciile de graminee perene (*Festuca arundinacea*, *Festuca pratensis*, *Festuca rubra*, *Dactylis glomerata*, *Lolium perenne*, *Phleum pratense*, *Bromus inermis*) și leguminoase perene pentru pajiști (*Trifolium repens*, *Lotus corniculatus*, *Onobrychis viciifolia*), adaptate la schimbările climatice prin care să se îmbunătățească furajarea animalelor și crearea de noi posibilități de reconstrucție ecologică prin înierbarea terenurilor degradate”.

ISBN 978-606-8557-89-2