

# Prioritățile în dezvoltarea resurselor genomice și transcriptomice la speciile forestiere de cvercinee din România

Dragoș POSTOLACHE, Flaviu POPESCU

## 1. Introducere

Ritmul rapid al schimbărilor climatice asociate cu acțiunile antropice au condus la creșterea vulnerabilității ecosistemelor forestiere în regiunile din sudul și sud-estul Europei unde s-a observat o creștere a frecvenței temperaturilor extreme și o scădere a cantităților de precipitații (Barbu et al 2016, Olesen, 2017).

Se estimează că evenimentele climatice extreme vor avea consecințe negative asupra pădurilor cu rezistență scăzută, cum sunt majoritatea pădurilor din zona silvostepă și câmpie forestieră cu specii de cvercinee din sudul țării (Giurgiu 2010, Popa et al 2013).

Procesul de adaptare a speciilor forestiere este unul destul de îndelungat, comparativ cu plantele anuale, datorită longevității și ciclului reproductiv mare.

Ecosistemele forestiere de cvercinee din stepă și silvostepă sunt mult mai adaptate la condițiile de secetă, datorită fondului de gene rezultat în urma proceselor de evoluție (ex. selecție naturală, driftul genetic, mutațiile și fluxul de gene) care au acționat pe parcursul a numeroase generații.

Studiile efectuate în cadrul ecosistemelor forestiere de cvercinee din România cu stejar pedunculat și stejar brumăriu au identificat diverse fenotipuri cu potențial ridicat de adaptare la condiții climatice locale (Apostol et al 2017, Crăciunesc 2013, Kavaliauskas et al 2018, Konnert et al 2015).

Cea mai mare provocare în estimarea variabilității genetice adaptative o reprezintă identificarea locilor genici și a regiunilor din genom aflate sub acțiunea selecției, în ultimii 10 ani fiind înregistrate progrese semnificative (Plomion et al 2016) prin dezvoltarea accelerată a tehnologiilor de secvențiere de nouă generație (NGS – Next Generation Sequencing). Acestea au permis într-un timp relativ scurt, dezvoltarea de resurse genomice și transcriptomice, care sunt esențiale în identificarea genelor implicate în procesele de adaptare la diferiți factori de stres abiotici și biotici.

## 2. Realizările la nivel internațional în domeniul dezvoltării resurselor genomice și transcriptomice la speciile forestiere

Importante resurse genomice, cum sunt ESTs (eng. Expressed Sequence Tags), secvențe scurte de ADN din cadrul genelor, au fost dezvoltate cu ajutorul metodei de secvențiere clasice descrisă de Frederick Sanger și colegii săi (Sanger et al 1977). Secvențierea Sanger a fost una dintre cele mai utilizate metode de secvențiere ADN, timp de mai bine de 25 de ani, fiind înlocuită treptat de secvențierea de nouă generație (NGS).

Cele mai numeroase resurse genomice, de tipul ESTs, au fost dezvoltate preponderent pentru speciile forestiere cu importanță economică foarte mare, cum sunt speciile din genurile *Pinus* (~731,380 ESTs), *Picea* (~576,345 ESTs) și *Populus* (~455,294 ESTs), comparativ cu numărul redus la speciile din genul *Quercus* (~150,916 ESTs), conform resurselor genomice disponibile în baza de date NCBI-EST (<https://www.ncbi.nlm.nih.gov/nucore>).

Dacă prin metoda clasică Sanger au fost secvențiate doar anumite regiuni din genomul nuclear, prin dezvoltarea tehnologiilor noi de secvențiere NGS s-a reușit secvențierea integrală a genomului nuclear pentru speciile forestiere importante din punct de vedere economic și a căror secvențe de referință vor putea fi utilizate pentru miile de specii forestiere cu importanță ecologică semnificativă (Neale and Kremer 2011).

În general proiectele mari de secvențiere a genomului nuclear cu tehnologii NGS au fost focalizate pe specii forestiere cu importanță economică majoră, cum ar fi; *Populus trichocarpa* (Tuskan et al 2006), *Picea abies* (Nystedt et al 2013), *Eucalyptus grandis* (Myburg et al 2014), *Picea glauca* (Warren et al 2015), *Pinus taeda* (Zimin et al 2017), *Olea europaea var. sylvestris* (Unver et al 2017), *Fraxinus excelsior* (Sollars et al 2017) și *Quercus robur* (Plomion et al 2018).

Costurile de secvențiere cu tehnologiile NGS sunt foarte ridicate la speciile de conifere care au dimensiuni mari

a genomului, cum ar fi la *Pinus taeda* de aprox. 22.1 Gb, comparativ cu speciile care au dimensiunea relativ mică de 422.9 Mb cum este în cazul popului balsamifer (*Populus trichocarpa*).

Un obstacol major în dezvoltarea genomurilor de referință la speciile de cvercinee l-a constituit dificultățile întâmpinate în asamblarea secvențelor scurte generate de tehnologiile NGS (Ion Torrent și Illumina), datorat gradului ridicat de heterozigoție. Într-un final s-a reușit asamblarea genomurilor contiguu haploid cu mărirea de 814,282,569 pb (~814Mb) și a celui diploid cu mărirea de 1,455,104,916 pb (Plomion et al 2018) (<http://www.oakgenome.fr>).

Resursele transcriptomice dezvoltate pentru speciile forestiere, cum ar fi *Fagus sylvatica* (Lesur et al 2015) sau *Quercus robur* și *Q. petraea* (Lesur et al 2015) sunt mult mai numeroase comparativ cu genomurile de referință dezvoltate. Numărul resurselor transcriptomice dezvoltate este mai mare deoarece mărirea transcriptomului este mult mai redusă, deoarece include doar moleculele de ARN (ARNm – ARN mesager, miARN – micro ARN, ARNt – ARN de transfer și ARNr – ARN ribozomal) dintr-o celulă sau țesut la un moment dat.

Astfel, dacă în cazul stejarului pedunculat (*Q. robur* L.) mărirea genomului (2C) este de aproximativ 1,45Gb pb (Plomion et al 2018), mărirea transcriptomului este de doar ~199,2Mb (199,278,344 pb) (Lesur et al 2015) sau 12,23%.

O strategie fiabilă pentru reducerea costurilor în dezvoltarea resurselor genomice la specii forestiere o reprezintă secvențierea exomului, ce include doar regiunile de codificare a genelor ce sunt transcrise în ARN-mesager matur

În cazul stejarului pedunculat (*Q. robur*) dimensiunea exomului este de aproximativ 75Mb (variante haploidă) și include 25,808 gene (<http://www.oakgenome.fr>).

Tehnologiile de secvențiere de nouă generație (NGS) permit de asemenea secvențierea integrală a genomului cloroplastic ADNcp, care în cazul speciilor de cvercinee este de aproximativ 132Mb (Pham et al 2017).

Resursele genomice dezvoltate pe baza secvențierii transcriptomului și exomului au adus contribuții importante pentru identificarea genelor implicate în diverse procese fiziologice, cum sunt procesul de fotosinteză, procesul de respirație, procesul de transpirație, dar și la descoperirea unor gene implicate în reglarea unor mecanisme fiziologice cum este cel al stării de dormanță a mugurilor (Lesur et al 2015).

Noile tehnologii de secvențiere (NGS) au contribuit decisiv la creșterea exponențială a numărului markerilor moleculari de tip SNPs (Single Nucleotide Polymorphism), mai ales după lansarea pe piață, în anul 2005, a primei platforme de secvențiere de tip NGS (454 Life Sciences cu modelul GS20).

Astfel, dacă prin secvențierea Sanger a 700 gene candidate s-au generat 7,576 de noi markeri SNPs, prin aplicarea tehnologiei NGS s-au identificat 191,281 de noi markeri SNPs în 297 arbori proveniți dintr-o populație mixtă cu *Q. robur* și *Q. petraea*. (<https://doi.org/10.15454/OMO2JC>)

Pe baza polimorfismului markerilor moleculari de tip SNPs, observat în cadrul genelor s-a reușit evidențierea repartizării geografice a variabilității genetice adaptative prin cuantificarea corelației dintre frecvența alelelor și diverși factori climatici (ex. Indicele de ariditate de Martonne), de-a lungul unor gradiente latitudinale, longitudinale sau altitudinale (Brousseau et al 2016, Eckert et al 2010, Roschanski et al 2016)

Genomica forestieră a beneficiat considerabil în ultimii 10 ani de progresele tehnologiei de secvențiere de nouă generație (NGS), care a făcut posibilă investigarea rolului diferitor gene, dezvoltarea de markeri noi și studierea evoluției speciilor forestiere (Plomion et al 2016).

### 3. Prioritățile naționale în dezvoltarea resurselor genomice și transcriptomice la speciile forestiere din România

La nivel național există un interes foarte mare pentru dezvoltarea de resurse genomice și transcriptomice la speciile forestiere, care să permită obținerea de cunoștințe noi privind mecanismele moleculare implicate în procesul de adaptare a speciilor la factorii biotici și abiotici.

Până de curând, accesul limitat la tehnologiile de secvențiere de nouă generație (NGS) și costurile relativ mari de secvențiere, au împiedicat dezvoltarea de resurse genomice și transcriptomice la nivel național.

În scopul dezvoltării de resurse genomice și transcriptomice la nivel național, cercetători din cadrul Institutului Național de Cercetare – Dezvoltare în Silvicultură “Marin Drăcea” (INCDS), au efectuat un stagiul de specializare în domeniul aplicării tehnologiilor NGS, la Institutul de Științe și Resurse Biologice din cadrul Consiliului Național de Cercetare din Florența, Italia (CNR-Firenze).

Utilizarea tehnologiilor de secvențiere de nouă generație (NGS) în identificarea genotipurilor valoroase, cu potențial productiv și adaptativ ridicat, deschide noi oportunități în selectarea populațiilor și ecosistemelor cu plasticitate ridicată, având avantajul reducerii dimensiunilor temporale necesare testării valorii resurselor genetice selecționate.

În cadrul stațiunii INCDS Simeria a fost implementat un sistem bioinformatic integrat, bazat pe tehnologiile de secvențiere de nouă generație Ion Torrent, care asigură un flux de lucru automatizat, începând cu prepararea librăriilor genomice de ADN și încărcarea automată a cipurilor semiconductoare cu ajutorul echipamentului IonChef system și terminând cu secvențierea librăriilor genomice ADN la secvențiatorul Ion Torrent GeneStudio S5 NGS System.

Traducerea informației genetice în informație digitală genomică este realizată de Ion Torrent Server preinstalat pe sistemul GeneStudio S5, iar analiza analitică se realizează cu softuri bioinformatic specializate (CLC Genomics Workbench 11), pentru detectarea variabilității în cadrul secvențelor de ADN (ex. inserții, deleții, inversii, mutații).

Sistemul bioinformatic integrat bazat pe tehnologia NGS Ion Torrent de la INCDS Simeria va avea rolul de a contribui la impulsivitatea dezvoltării resurselor genomice și transcriptomice, dezvoltarea de markeri noi, identificarea genelor din transcriptom, detectarea regiunilor din genom implicate în procesele de adaptare și selecție.

Pentru prima dată la nivel național au fost dezvoltate librării genomice la specii de stejari, prin secvențierea integrală a ADN-ului cloroplastic, în scopul studierii evoluției filogenetice a cvercineelor. În acest scop, au fost dezvoltate 10 librării genomice pentru haplotipuri rare identificate în populații de stejar brumăriu (*Q. pedunculiflora*) și câte două librării genomice pentru stejar pufos (*Q. pubescens*), stejar pedunculat (*Q. robur*), gorun (*Q. petraea*), stejarul italian (*Q. virgiliana*) și gărniță (*Q. frainetto*).

De asemenea au fost dezvoltate 4 librării genomice pentru exemplare de stejar brumăriu și stejar pedunculat varietatea "tardifolia".

Estimarea potențialului genetic adaptativ, în cadrul ecosistemelor forestiere de cvercinee din stepă și silvostepă din România, va reprezenta o prioritate în contextul actual al schimbărilor globale de mediu, iar tehnologiile NGS vor permite identificarea și selectarea resurselor forestiere care dețin un fond de gene adaptat la secete severe. Utilizarea în practica silvică a materialelor forestiere provenite din aceste resurse va contribui la refacerea cu succes a ecosistemelor forestiere degradate și creșterea gradului de stabilitate a noilor arborete la schimbări climatice.

## Finanțare

Cercetările sunt realizate în cadrul programului Nucleu BIOSERV, proiectul (PN 19070304) "Estimarea potențialului adaptativ și a capacității de reziliență a ecosistemelor forestiere de cvercinee vulnerabile la schimbările climatice prin analize multidisciplinare și dezvoltarea de resurse genomice".

## Bibliografie

- Apostol EN, Curtu AL, Daia LM, Apostol B, Dinu CG, Șofletea N, 2017.** Leaf morphological variability and intraspecific taxonomic units for pedunculate oak and grayish oak (genus *Quercus* L., series *Pedunculatae* Schw.) in Southern Carpathian Region (Romania). *Science of the Total Environment* **609**, 497-505.
- Barbu I, Barbu C, Curcă M, Ichim V, 2016.** Adaptarea pădurilor României la schimbările climatice. 378.
- Brousseau L, Postolache D, Lascoux M, Drouzas AD, Källman T, Leonarduzzi C, Liepelt S, Piotti A, Popescu F, Roschanski AM, 2016.** Local Adaptation in European Firs Assessed through Extensive Sampling across Altitudinal Gradients in Southern Europe. *PLoS one* **11**, e0158216.
- Crăciunesc II, 2013.** Evaluarea nivelului de hibridare naturală la specii autohtone de cvercinee: analiză de caz în rezervația naturală Bejan-Deva Evaluation of natural hybridization in native species of oaks: a case study in Bejan-Deva Natural Reserve, Teză doctorat, Universitatea Transilvania Brașov.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, Gonzalez-Martinez SC, Neale DB, 2010.** Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**, 969–82.
- Giurgiu V, 2010.** Pădurile și schimbările climatice. *Revista pădurilor* **3**, 15.
- Kavaliuskas D, Fussi B, Westergren M, Aravanopoulos F, Finzgar D, Baier R, Alizoti P, Bozic G, Avramidou E, Konnert M, 2018.** The interplay between forest management practices, genetic monitoring, and other long-term monitoring systems. *Forests* **9**, 133.
- Kleinschmit J, 1993.** Intraspecific variation of growth and adaptive traits in European oak species. In "Annales des sciences forestières", Vol. 50, pp. 166s-185s. EDP Sciences.
- Konnert M, Fady B, An Gömöry D, A'hara S, Wolter F, Ducci F, Koskela J, Bozzano M, Maaten T, Kowalczyk J, 2015.** Use and transfer of forest reproductive material in Europe in the context of climate change. European Forest Genetic Resources Programme (EUFORGEN), Bioversity International, Rome, Italy. xvi and 75.
- Lesur I, Le Provost G, Bento P, Da Silva C, Leplé J-C, Murat F, Ueno S, Bartholomé J, Lalanne C, Ehrenmann F, 2015.** The oak gene expression atlas: insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics* **16**, 112.
- Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, 2014.** The genome of *Eucalyptus grandis*. *Nature* **510**, 356.
- Neale DB, Kremer A, 2011.** Forest tree genomics: growing resources and applications. *Nature reviews. Genetics* **12**, 111–22.
- Nețoiu C, Tomescu R, Buzatu A, 2016.** The basic environment and *Lymantria dispar* L. (Lepidoptera, Erebidae) infestations. *Muzeul Olteniei Craiova. Oltenia. Studii și comunicări. Științele Naturii* **32**, 8.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, 2013.** The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584.
- Olesen JE, 2017.** Climate change impacts on society. In "Climate Change, Impacts and Vulnerability in Europe 2016". European Environment Agency.
- Pham KK, Hipp AL, Manos PS, Cronn RC, 2017.** A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny. *Genome* **60**, 720-732.
- Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, Faye S, Francillon N, Labadie K, Le Provost G, 2018.** Oak genome reveals facets of long lifespan. *Nature plants*, 1.
- Plomion C, Bastien C, Bogeat-Triboulot M-B, Bouffier L, Déjardin A, Duplessis S, Fady B, Heurtz M, Le Gac A-L, Le Provost G, 2016.** Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* **73**, 77–103.
- Popa I, Leca Ș, Craciunescu A, Sidor C, Badea O, 2013.** Dendroclimatic response variability of *Quercus* species in the Romanian intensive forest monitoring network. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **41**, 326.
- Roschanski AM, Csilléry K, Liepelt S, Oddou-Muratorio S, Ziegenhagen B, Huard F, Ullrich KK, Postolache D, Vendramin G G, Fady B, 2016.** Evidence of divergent selection for drought and cold tolerance at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps. *Molecular ecology* **25**, 776–794.
- Sanger F, Nicklen S, Coulson AR, 1977.** DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences* **74**, 5463-5467.
- Sollars ES, Harper AL, Kelly LJ, Sambles CM, Ramirez-Gonzalez RH, Swarbreck D, Kaithakottil G, Cooper ED, Uauy C, Havlickova L, 2017.** Genome sequence and genetic diversity of European ash trees. *Nature* **541**, 212-216.
- Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, 2006.** The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604.
- Unver T, Wu Z, Sterck L, Turktas M, Lohaus R, Li Z, Yang M, He L, Deng T, Escalante FJ, 2017.** Genome of wild olive and the evolution of oil biosynthesis. *Proceedings of the National Academy of Sciences* **114**, E9413-E9422.

- Warren RL, Keeling CI, Yuen MMS, Raymond A, Taylor GA, Vandervalk BP, Mohamadi H, Paulino D, Chiu R, Jackman SD, 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *The Plant Journal* **83**, 189-212.
- Zimin AV, Stevens KA, Crepeau MW, Puiu D, Wegrzyn JL, Yorke JA, Langley CH, Neale DB, Salzberg SL, 2017. An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing. *Gigascience* **6**, 1-4.

## Abstract

### Setting priorities for development of genomic and transcriptomic resources in oak forest species from Romania

A summary of achievements in the development of genomic and transcriptomic resources in forest species are presented in an international context.

The first steps in the development of genomic resource in oak species of Romania are presented and national research priorities were discussed and determined.

The application of Next Generation Sequencing (NGS) technologies may foster at the national level to identify valuable genotypes of tree forest species with increased productivity and adaptive capacity.

In the framework of the Research Station of Simeria was implemented a fully equipped NGS laboratory for applications that are of interest to National Institute for Research and Development in Forestry "Marin Drăcea".

At the national level there is an increased interest in development of genomic and transcriptomic resources for forest species that would allow to produce valuable new knowledge about molecular mechanisms involved in adaptation processes of tree species to biotic and abiotic stress factors.

For the first time, at the national level have been developed genomic resources for oak species such as *Quercus pedunculiflora*, *Q. pubescens*, *Q. robur*, *Q. petraea*, *Q. virgiliana*, and *Q. frainetto*.

Additionally, genomic resources have been developed for late bud flushing variety (var. *tardifolia* Czern.) found in *Quercus pedunculiflora* and *Q. robur*. The late bud flushing variety (var. *tardifolia* Czern.) in oak species is of great importance in forestry because of harder and more durable wood (Kleinschmit, 1993), as well as this variety is less affected by defoliators (Nețoiu et al., 2016).

The employment of forest material from certified genetic resources can contribute towards successful restoration of degraded forest ecosystems and to foster resilience and stability of forests to global changes.

**Keywords:** oak species, NGS, genomic resources, transcriptomic resources.