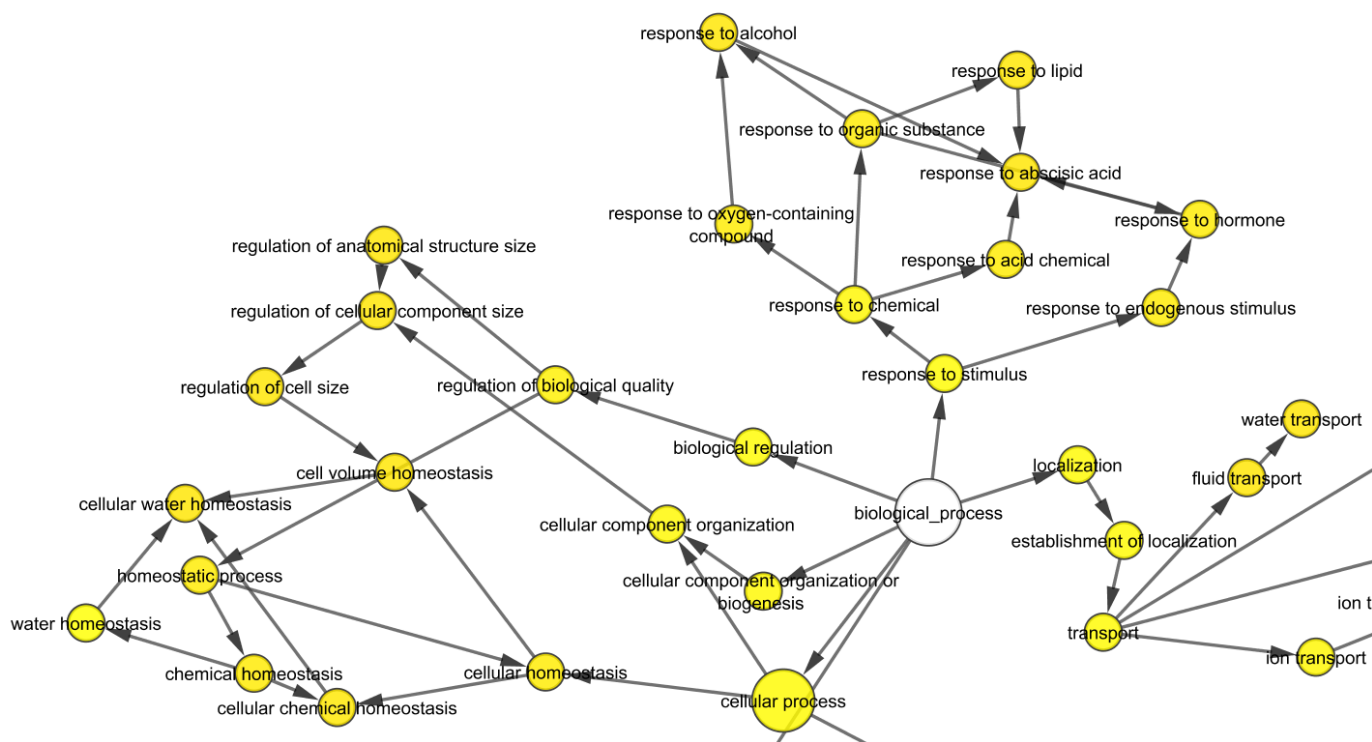


Viens no projekta pamata mērķiem ir veicināt pēcdoktoranta profesionālo attīstību, palielinot pētniecisko kapacitāti un kompetenci. Komandējuma laikā apguvu jaunas prasmes, tādās kā genoma sekvenču bioinformātikās analīzes metodes izmantojot sadarbības partnerim pieejamo *UPPMAX* sistēmu (*Uppsala Multidisciplinary Center for Advanced Computational Science*), kas ir augstas kapacitātes digitālā darba vide ar daudzām instalētām un darba kārtībā esošajām bioinformātikām programmām. *UPPMAX* digitālais centrs ļauj izanalizēt liela apjoma datus, tādus kā priedes visa genoma sekvenci (6,58 GB), kuras apstrāde nebija iespējama ar mums pieejamiem resursiem. Strādāju ar *UPPMAX* klāsteri, kas satur 6080 kodolus, kas savukārt tiek reprezentēti ar 334 mezgliem katrs ar diviem 10-kodola Intel Xeon V4 CPU. 32 lielākie mezgli satur 256 GB atmiņas, bet pārējie satur vēl 128 GB. Dažādu uzdevumu apstrāde jaudīgajā *UPPMAX* sistēmā aizņēma no dažām sekundēm līdz pat divām dienām.

Pieejamās bioinformātikās analīzes programmas izmanto komandrindas principu *Linux* vidē, katrai no izmantotām programmām ir sava lietotāja rokasgrāmata, tomēr iesācējam ir grūti saprast, kurš rīks var tikt izmantots konkrētā pētnieciskā uzdevuma realizēšanai. Iemācījos sekvenču analīzes pamatprincipus, *UPPMAX* sistēmas izmantošanas principu, dažādu programmu izmantošanas veidus. Analīzeju vairākus priežu references genomus: *Pinus taeda* v.1.01, v2.01, *Pinus lambertiana* v.1.01. un *Pinus sylvestris* nepabeigto genomu, kurš nav publiski pieejams ārpus Zviedrijas universitātes laboratorijas. Pirmās priežu genomu versijas ir sadalītas augstās un zemās kvalitātes gēnu kopās. Datu precīza strukturēšana darba gaitā bija ļoti svarīga un tam tika pievērsta papildus uzmanība. Darba gaitā tika risinātas dažādas pētnieciskās problēmas. Piemēram, veikta pieejamās transponējamo elementu datubāzes sakārtošana. Transponējamie elementi bieži vien ieņem vienu otrā, tādejādi veidojās klāsteru struktūras sastāvošās no vairākiem gariem termināliem atkārtojumiem. Literatūrā ir aprakstīti gadījumi, kad arī šāda himēriska struktūra pārvietojās miežu genomā, taču tas ir samērā rets gadījums. Šādas insercijas satur īpašas struktūras, kas pierāda to transpozīciju. Tomēr priežu retrotranspozonu datubāzē automātiskās anotēšanas rezultātā ir ievietoti vairāki simti elementi ar šādu struktūru un šo elementu atkārtojamība genomā vai specifiski signāli netika analizēti. Tapēc retrotranspozonu ģimeņu skaits pieejamajā datubāzē ir daudz lielāks (vairāk kā 19 tūkst.), kamēr praktiski vairākas sekvenču savstarpēji ir līdzīgas un reprezentē vienādu sekvenču kopu, ņemot vērā pieņemtos mobilo elementu klasificēšanas principus. Šī nepilnība tika novērsta, jo tā nopietni ietekmētu rezultātus- palielinātos atkārtoto datu apjoms; idenificētas gēnu kopas būtu līdzīgas dažādām sekvenču kopām, kas praktiski reprezentē vienu transponējamo elementu ģimeni; dažādi atkārtojumi vienas struktūras iekšienē uzrādītu līdzību ar tiem pašiem gēniem. Zināms, ka garie terminālie atkārtojumi satur svarīgus signālus, kas var regulēt gēnu darbību un, iespējams, ka evolucionāri izdevīgāk uzturēt tikai šo daļu kodējošo sekvenču

tuvumā. Lai pārbaudīt šo hipotēzi, nolēmu izolēt no datubāzes visus garos terminālos atkārtojumus un atlasīt to reprezentatīvas sekvences. Tas ir arī izdevies un visas analīzes tika atkārtotas arī ar šo datu kopu.

Rezultātā ieguvu 15 datu kopas, katra no tām satur simtiem sarakstu ar vairāku simtu gēnu nosaukumiem un sekvencēm, kas satur attiecīgo transponējamo elementu. Dati tika arī sakārtoti pēc vairākiem nozīmīgiem parametriem. Katrai identificētai gēnu kopai turpmāk tiks noteikta funkcionāla piederība un novērtēta iespējamība, ka kopas pārstāvji ir iesaistīti līdzīgajos procesos (attēls 1.). Attēlā redzama daļa no bioloģisko procesu tīkla, kur vairāki no identificētiem gēniem piedalās šūnas atbildes veidošanā, ūdens transporta nodrošināšanā, ūdens homeostāzē un šūnas lieluma regulācijā. Iespējams, ka šādu gēnu darbība varētu būt svarīga atbildot uz sausumu. Tā kā visi šie gēni satur vienādu sekvenci UTR rajonā, un šo gēnu grupu reprezentācija tīklā ir statistiski būtiska salīdzinot ar visu pieejamo priedei anotēto gēnu grupu frekvenci genomā, tā varētu būt svarīga operatīvās atbildes veidošanā, ietekmējot šo gēnu transkripcijas faktoru piesaisti, metilēšanas statusu. Lai turpmāk pierādīt vai noliegt šīs sekvenču nozīmi, tiks sīkāk analizēta struktūra, iesaistīto gēnu strukturālā variācija populācijā, un šo lokusu ekspresija atbilstošo apstākļu ietekmē. Pirms tam, tiks izanalizētas visas gēnu-transponējamo elementu kopas un atlasīti tikai nozīmīgi tīkli un to dalībnieki.



Attēls 1. Viens no identificēto gēnu saistības tīkliem, kas tika veidots pamatojoties uz transponējamā elementa *PtRLX1124* esamību šo gēnu UTR rajonā un bioloģiskajos procesos iesaistīto gēnu ontoloģijas klasifikāciju.