

# Arvutite abil looduse saladusi muukimas

Kaasaegne teadus toodab bioloogia suurtele küsimustele vastuseid otsides meeletutes kogustes andmeid. Nende süstematiseerimine ja uuringutulemustest kätuvate mustrite leidmine on bioinformaatik Jaak Vilo igapäevatöö.

TEKST: KRISTJAN KALJUND

Tallinna Reaalkooli lõpetanud Vilo tee teaduse juurde algas Tartu Ülikoolis arvutiteadust õppides. Vahetusüliõpilasena 1989–1990 Helsingis õppinud Vilo astus põhjanaabrite juures 1991 ka doktorantuuri. Selle käigus hakkas ta kirjutama tarkvara ja programme tekstide analüüsiks. Teadmistele rakendusvaldkonda otsides oli valida loomuliku keele ning DNAsse kodeeritud teksti vahel. Vilo asus tööle viimase kallal ning otsis reegleid, mille alusel valk mingi kindla kuju võtab ning millised signaalid seda DNAs määravad.

Aeg oli põnev, äsja oli edukalt sekveneeritud esimese eukarüoodi genoom (pagaripärm), ning avaldatud esimesed mikrokiipide andmed, mille põhjal võis juba leida ühtemoodi käituvaid geenide alamgrupe ning neile omaseid erinevaid DNAs peidus olevaid geenide regulatsiooni mõjutavaid mustreid. «Süües kasvab isu ja siis tulid juba teised küsimused: kuidas geenide regulatsioon laiemalt toimib, milline on see molekulide võrgustik?» meenutab Vilo. «Kõik see kokku oli üks suur küsimus, mille ümber proovisin pisikestele jupikestele vastata.»

## Unikaalne võimalus

Seejärel siirdus Vilo Inglismaale, kus töötas Euroopa Bioinformaatika Instituudis, arendades bioloogiliste andmete «kaevandamise» tarkvara. Ilmselt oleks ta saareriiki jäänudki, kui Eestis poleks uue aastatuhande alguses hakanud idanema geenivaramu idee.

Eestlastest bioinformaatika eksperte võis tollal loetleda ühe käe sõrmedel ja nii kutsuti ka tosin aastat võõrsil töötanud Vilo kodumaale tagasi. «Inglismaal oli eluaegne töökoht ootamas, nii et loomulikult oli Eestisse erafirmasse tööle



## CV

### Jaak Vilo

- Sündinud 14.11.1966
- Lõpetanud Tallinna 11 Keskkooli
- 1991 Tartu Ülikooli bakalaureusekraad rakendusmatemaatikas
- 2002 Helsingi Ülikooli doktorikraad arvutiteaduses
- 2002–2006 EGeeni (Geenivaramu) informaatikadirektor
- 2007. aastast Tartu Ülikooli bioinformaatika professor
- Töövaldkonnad: andmekaevandus, klasteranalüüs, tekstialgoritmid, süsteemibioloogia, suuremahuliste arvutusülesannete lahendamise meetodid jm
- Paljude erialaste programmikomiteede ning nõukogude liige

tulek riski võtmine,» tunnistab mees. «Aga šanss oli unikaalne ning paremat pakkumist Eestisse naasta poleks ilmselt tulnud. Siin korda saadetud asjade mõju on kokkuvõttes suurem kui mõju, mida Inglismaal oleksin saavutanud, nii et otus oli õige.»

## Ühe ööga tulemusi ei tule

Töö Eestis algas täiesti nullist: tuli paika panna, kuidas ja mis tüüpi andmed geenivaramusse koonduvad ja mida nendega seal tehakse. Teiselt poolt tuli jätkata oma alustatud teadussuundade edasiarendamist. Kuna asjatundjaid nappis, hakkas Vilo ka ülikoolis tudengeid juhendama ja loenguid andma ning ta valiti 2007. aastal bioinformaatika professoriks.

Kuigi geenivaramu projekt vaevles aastaid rahahädas ja oli sunnitud esialgseid plaane mõnevõrra koomale tõmbama,

**Ilmselt oleks ta saareriiki jäänudki, kui Eestis poleks hakanud idanema geenivaramu idee.**

on Eesti nii tänu geenivaramu projektile kui ka bioinformaatika rühmade arengule ülikoolides oma rahvaarvu kohta väga tublisti maailmakaardil.

«Kui sa tahad, et inimese geen sulle mingit lugu jutustaks, siis ei saagi ühe ööga asju lahendatud,» selgitab Vilo, miks geeniuuringud nii kaua aega võtavad. «Äsja toodi Eestisse täismahus sekvenaator. Kui geenivaramut projekteeriti, ei osanud keegi sellest isegi unistada, aga peagi on reaalsus see, et tehniliselt saaks kõik Eesti elanikud sekveneerida. Keskklassi inimene suudaks seda lausa ise finantseerida.»

Hoopis iseküsimus on aga see, mida säärase andmehulgaga peale hakata. «Tänu täismahus sekveneerimisele on kogu maailmas aru saadud, kui unikaalne ikkagi iga inimene on. Mis siis, et mul on selline DNA? Kui palju riske ma tänu sellele reaalselt vähendada saan?»

Personaalsete ravimite loomine pole kuskil läinud nii libedalt kui loodeti. Keerulisemate haiguste mõjutajad on mööda genoomi laiali ja nende leidmine äärmiselt keeruline. Problemaatiline on ka uute

ravimite katsetamine, sest ei loomkatsed ega ka inimese tüvirakkudest toodetud rakuliinide kasutamine ei anna tulemusi, mida saaks üks ühele kõigile inimestele üle kanda.

«Meie algteadmised on ikka üsna na- did, isegi kui kõik toorandmed on käes,» tunnistab Vilo. «Selles osas loodetakse palju süsteemibioloogiale, mis püüab lahendusi leida arvutimudelite põhjal. Tuumapommi suudame arvuti peal uu- rida, miks siis mitte inimest? Samas on tuumapommi puhul arvutamist küll pal- ju, aga reeglid füüsikalisele lihtsad, ent füüsikareeglite bioloogiliste reeglite tu- letamine on teema, mida võib uurida veel aastasadu.»

Muide, ka üks maailma kuulsamaid arvutiteadlasi, algoritmide uurimisele aluse pannud Donald Knuth, on öelnud, et arvutiteadusele ta lõputut kasvu prognoosi- da ei julge, aga bioloogias jagub relevant-

### «Mõnes valdkonnas kahekordistub and- mete hulk üheksa kuuga.»

seid probleeme veel 500 aastaks.

Ent kaasaja bioloogia ei oleks enam võimalik ilma informaatikata. «Andme- mahud on sellised, mida saabki üksnes arvutitega tõlgendada,» selgitab Vilo. «Andmebaasid on suured ja ükskõik mida sa mõõdad, pead tulemusi võrdlema eel- miste andmetega ja kümnetest erineva- test katsetest ühe tervikpildi moodusta- ma. See ongi bioinformaatika suurimaid väljakutseid.»

Uuringutes on ammu juba võetud suund sellele, et andmed on avalikud, sest see võimaldab teistel neid töödelda ja mitte tööd duplitseerida. «Inimese DNA sekveneerimise võidujooksus oli ka kõige suurem intriig see, kas genoomi primaar- järjestus muutub avalikuks või mitte, kõik muu oli tehniliste detailide küsimus. See on viimase 20 aasta jooksul toimunud muutus. Nüüd on suund sinna, et teadusartiklid oleksid samamoodi avalikud,» selgitab teadlane.

Hetkel küsitakse mõne teadusartikli *online*-lugemise eest keskmiselt 30–50 dollarit. «Seda raha ei oleks vaja võtta, sest see on juba kinnimakstud töö,» ütleb Vilo. «Avatud kirjastamise korral tekiks rohkem võimalusi ka sinna, kus on väik- semad ressursid, info leviks kiiremini ja vabamalt. Tänu IT-le saavad sellised as- jad võimalikuks.»

Seda et teadusajakirju on palju, teab vast igaüks, kes raamatukogus mõne vas- tava andmebaasiga tutvunud, aga fakt, et iga päev avaldatakse maailmas 5000–6000 teadusartiklit ainuüksi biomeditsii-



OVE MAIDLA/TARTU ÜLIKOOI

nis, ületab ilmselt julgemaidki pakkumisi. Säärast tekstimassi ei suuda käsitsi keegi isegi läbi sirvida, saati siis lugeda ja end kõigega kursis hoida. «Üks bioinformaa- tika valdkond tegelebki teadusartiklite analüüsimisega arvutite abil,» selgitab Vilo.

Ta ise on tsiteeritavuse poolest üks Eesti mõjukamaid teadlasi, kelle viimase kümne aasta jooksul ilmunud artiklitele on ISI andmetel viidatud üle 2700 kor- ra ja Google Scholar andmetel ligi 5000 korda.

#### Ülikool ei ole moesõu

Ülikoolis õpetab Vilo peamiselt algoritmide koostamist ning andmestruktuuri- de ehitust, sest see annab baasi reaalse- te ülesannete lahendamiseks. «Ülikool peab õpetama fundamentaalsed asjad väga hästi selgeks, see ei ole mingi moesõu, et täna see ja homme too trend,» arvab ta. «Peamine kasu, mis mina sain pikast õppimisest, oligi korralik baasharidus. Selle peale saab ehitada kõikvõimalikke kitsamaid uurimissuundi ja vajadusel ka eriala vahetada – bioinformaatika seosed kehtivad ka keeletehnoloogias. Kui ühen- dad endale huvitava ja kasuliku, siis pik- kamööda jääb midagi külge ka.»

Eesti ühiskonna suurimaks traagikaks peabki Vilo noorte andekate haridustee katkemist ja liiga vara tööle asumist. «Firmadele on neid vaja, aga nad ei jõua kunagi oma tippvõimeni. Täpselt nagu spor- dis – kui saad külavõistlusel hakkama ega viitsi enam treenida, oled kaotatud talent. Selline inimene ei tee asju, mida mitte keegi pole veel maailmas teinud. Kui lase- me generatsioonide viisi sadadel tuden- gitel ennast mitte realiseerida, siis see on kriminaalne.»

Seni, kuni riik ei oska või ei suuda teha sedalaadi strateegilisi otsuseid ja näidata, et ka akadeemiline karjäär on konkurent- sivimeline, võiksid firmad ühe lahendu- sena initsiatiivi enda kätte võtta, arvab Vilo. «IT-sektor peaks võtma jalad kõhu alt välja ja ütlema, et meie oleme valmis haridusse panustama, kui riik oma osa juurde paneb.» Tegelikult ongi praegu käivitumas IKT riiklik programm, mis peaks veidi kaasa aitama IT-hariduse pa- randamisele.

#### Rahvuskangelane

«On väga ohtlik, kui meil tekib illusioon, et doktorant, kes teeb päeval kaheksa tun- di palgatööd ja õhtuti või nädalavahetus- tel nokitseb veidi oma doktoritöö kallal, võib saada maailmatasemele,» ütleb Vilo. «Ta ei ole kuidagi konkurentsivimeline inimesega, kes mõnes välisülikoolis pü- hendab 70–80 tundi nädalas üksnes oma doktoritööle. Ilma vaevata medaleid ei ja- gata, ole kui suur talent tahes.»

Vilot ennast on Soome tagasi kutsutud ja Inglismaalgi imestati, kui ta otsustas Eestisse naasta. «Üks Hiina päritolu USA teadlane, kes mind USAsse kutsus, arvas,

et peaksin selle otsuse tõttu rahvuskang- elane olema. Aga me peame oma tu- dengitele ka kodumaal võimaluse looma. Minu kui rühmajuhi ülesanne ongi kanda hoolt, et inimesed saaksid teha oma tööd, ja mina saan siis need hallid juuksed.»

Kuigi IT-alaste tulevikuennustustega on alt läinud ka väga tuntud eksperdid, ei ole valdkonna areng Vilo meelest siiski päris prognoosimatu. Tõde, usub ta, on kuskil vahepeal: ei ole tulevik nii sujuv ja probleemitu, kui ilusamad lubadused kinnitavad, ega ka nii trööstitu, kui küm- ned ulmefilmid arvavad.

«Oma lapsele siniseid silmi või vastu- pidavusgeeni valida pole ehk eetilise, aga teisalt oleks ebaeetiline keelata valiku- võimalus vanematel saada terve laps, kui neil neljast kolm võib-olla muidu üldse iseseisvalt elada ei suudaks. Tehnoloogia nende valikute jaoks on aga täpselt sama,» veeretab Vilo vastutuse masinatelt inimestele.

Kuigi arvutid arenevad kiiresti ning Moore'i seadus, mis ennustab protsessorivõimsuse kahekordistumist iga 18 kuu tagant, peab laias laastus paika juba aastakümneid, ütleb Vilo, et andmetega on

### «Füüsikareeglite bioloogiliste reeglite tuletamine on teema, mida võib uurida veel aastasadu.»

asi veel pöörasem: «Andmete pealetuleku kiirus ületab arvutite kiiremaks muutu- mise kiiruse, mõnes valdkonnas kahekor- distub andmete hulk üheksa kuuga, nii et tekib tohutu lõhe andmete ja arvutamise vahele.»

Riistvara võimsuse kasvatamine käib usinalt ka Eestis. Tartu Ülikooli teadus- arvutuste keskus, mille nõukogusse Vilo kuulub, plaanib oma arvutipargi ehk au- rumasina, nagu seda hellitavalt kutsutakse, võimsust järgmise aastaga vähemalt kolmekordistada. «Järgmise 4–5 aastaga oleme praeguse seisuga võrreldes võim- suse võib-olla 50kordseks kasvatanud,» usub Vilo. «Aga investeerida ei saa vaid paari arvutisse, vaid peab hoolitsema, et oleks inimesed, kes masinaparki käigus hoiavad, arendavad ja kasutama õpeta- vad.»

Jaak Vilo tegemiste nimekirja on nii pikk, et mahukad tööpäevad on talle hästi tuttavad. «Noorena võid teha mõned nädalavahetused järjest tööd, aga taastu- mise võimalus peab ka olema. Sel suvel sain lõugastada küll, müttasin füüsilist tööd teha – see võtab muu mõtlemisvõi- me ära ja saadki rahus maaharimisele keskenduda,» on mees rahul.