

VOIKO EVOLUUTIOTA ENNUSTAA?

VILLE MUSTONEN

Evoluution tutkimus on perinteisesti keskittynyt jo tapahtuneiden asioiden ymmärtämiseen.

Genomidatan räjähdysenomainen kasvu, kokeellinen evoluutio ja evoluutioteorian lisääntynyt ymmärrys ovat avaamassa uuden tutkimussuunnan, jossa nopeasti muuttuvien lajien populaatioiden evoluutiota pyritään ennustamaan. Tässä kirjoituksessa kuvaan, millaiset seikat vaikuttavat evoluution ennustettavuuteen.

Evoluutio toimii erittäin pitkien aikaskaalojen yli. Kaikki tämänhetkinen elämä maanpäällä juontuu miljardeja vuosia sitten eläneestä viimeisestä universaalista yhteisestä esi-isästä. Tästä sekä sen jälkeisistä esi-isistä saadaan tietoa mm. vertailemalla evoluutioteorian avulla nykyisten organismien genomeja toisiinsa. Evoluutiotutkimus on pääsääntöisesti keskittynyt jo tapahtuneiden asioiden selvittämiseen ja ymmärtämiseen. Tämä fokus on ollut hyvä lähtökohta, sillä eri organismien perimien vertailu on opettanut meille valtavasti sekä biologiasta että evoluutiosta.

Evoluutiotutkimusta voidaan myös tehdä vertailemalle eri yksilöiden genomeja saman lajin sisällä. Hyvänä esimerkkinä tällaisesta tutkimuksesta käy ihmislajin evoluutiohistorian mallit, jotka kertovat meille esi-isiemme liikkeistä ja keskinäisistä vuorovaikutuksista. Meneillään oleva bioda-

tan, erityisesti genomidatan, räjähdysenomainen kasvu mahdollistaa yhä tarkemman mallintamisen kiinnostavista evoluutiohistorian kysymyksistä yli aikaskaalojen ja systeemien.

Yllämainittujen vakiintuneiden tutkimussuuntien lisäksi datavirtojen kasvu, kokeellinen evoluutio ja evoluutioteoriaan pohjautuvien mallien kehitys ovat avaamassa uuden suunnan, jossa nopeasti muuttuvien lajien populaatioiden evoluutiota pyritään ennustamaan. Avainkysymyksiä ovat: mitkä seikat vaikuttavat evoluution ennustettavuuteen, mitkä ovat ennustavan teorian ja mallintamisen rakennuselementit ja missä kulkevat ennustettavuuden rajat eri systeemeille? Mutta ennen kuin pääsen käsittelemään näitä kysymyksiä, on syytä pysähtyä miettimään, miksi ylipäätään haluaisimme ennustaa evoluutiota.

Ovatko evoluutiota hyvin ennustuvat mallit hyödyllisiä?

Joka vuosi ympäri maailmaa suuri joukko ihmisiä sairastuu kausittaisen influenssaviruksen vuoksi. Seuraukset ovat sekä inhimillisesti että kansantaloudellisesti erittäin suuret. Rokotus virusta vastaan auttaa merkittävästi, jos se on tehokas viruksen kyseisen kauden pääkantaan vastaan. Ongelma on, että rokotteen valmistamiseen liittyvien prosessien vuoksi päätös rokotteen kohdeviruskannasta pitää tehdä jo alkuvuodesta seuraavaa talvea varten. Jos pääkanta tulevana kautena on eri kuin se, mitä vastaan rokote oli suunniteltu, on rokotteen toimivuus mahdollisesti heikko. Osittain tästä syystä rokotteen tehokkuus vaihtelee influenssakausien välillä. Jos influenssaviruksen evoluutiota voitaisiin ennustaa siten, että rokote saataisiin tulevan kauden pääkantaan vastaavaksi mahdollisimman usein, ennustusten vaikuttavuus olisi huomattava.

Syöpien kohdalla genomidata sekä muut kvantitatiiviset datatyypit avaavat mahdollisuuden potilaan syöpää vastaan räätälöityyn terapiaan. Usein syöpä koostuu kuitenkin monista kilpailevista kloonista (tai syöpäsolulinjastoista), jotka ovat somaattisen evoluution seurauksena perimältään hieman erilaisia. Lääkkeiden vaste voi vaihdella eri kloonien välillä, ja tehokas lääke pääkloonin vastaan saattaa olla ilman vaikutusta jotain alakloonin vastaan. Tällöin lääke tappaa pääkloonin mutta jättää tilaa alakloonille, joka valtaa alaa nopeasti – syöpäpopulaatio tulee resistentiksi käytetylle terapialle. Kloonien terapiavasteen ja sitä kautta syöpäsolupopulaation evoluution ennustaminen voi auttaa terapian suunnittelussa. Syövän lisäksi monet bakteerit ja virukset voivat kehittävää resistenssin lääkkeitä vastaan. Resistenssievoluution dynamiikka voi olla hyvinkin nopeaa ja lääkeresistentit taudinaiheuttajat ovatkin nousseet maailmanlaajuisiksi terveysongelmaksi. Ennustava evoluution mallintaminen voi osaltaan auttaa resistenssi-ongelman ratkaisussa, esimerkiksi kehitetään terapiaprotokollia, jotka pyrkivät pienentämään resistenssin todennäköisyyttä.

Näiden hyvinkin käytännöllisten sovellusten lisäksi ennustamisen motiiviksi voi mainita, että yleisesti ottaen tieteellisten teorioiden ja mallien kehittäminen tapahtuu testaamalla niiden tekemiä

ennustuksia, jolloin esimerkiksi niiden puutteellisuus voidaan havaita.

Ennustavan mallintamisen osatekijät

Ennustusten kehittäminen on mielekkäintä systeemeille, joissa evoluutio tapahtuu nopeasti, aika-kaaloilla tunneista muutamiin vuosiin. Tällaisiin systeemeihin lukeutuvat mm. virukset, bakteerit, parasiitit ja syövätkin. Ennustusten tarkkuutta näille systeemeille voidaan jälkikäteen arvioida vertaamalla ennustetta tapahtuneeseen evoluutioon. Evoluutio on stokastinen prosessi, jossa sattumalla on, olosuhteista ja systeemistä riippuen, mahdollisesti suurikin merkitys. Tästä syystä ennusteet ilmaistaan luonnollisesti todennäköisyyksinä ja eri ennustusmenetelmien suorituskykyä arvioidaan käyttämällä informaatioteorian konsepteja.

Kun mutaatio tapahtuu yksilössä, sen kohtalo on joko hävitä populaatiosta tai levitä populaatioon mutatoituneen yksilön jälkeläisten kautta. Todennäköisyydet näille eri kohtaloille riippuvat monista seikoista, kuten populaation koosta, sekä itse mutaation vaikutuksesta yksilön kelpoisuuteen (*fitness*) kyseisessä ympäristössä. Esimerkkinä kelpoisuutta lisäävästä mutaatiosta käy bakteeripopulaation yksilössä tapahtuva sattumanvarainen mutaatio, joka tekee yksilöstä resistentin käytetylle antibiootille. Tällaisen yksilön jälkeläiset kasvattavat osuuttaan populaatiosta nopeasti, joten muutaman sukupolven jälkeen mutaatio voi löytyä suuresta osasta populaation bakteereja.

Juuri kelpoisuuden arviointi on ennustavan mallintamisen kannalta avainasemassa ja tehdään evoluutioteorian avulla tai kokeellisesti. Edellä mainitun influenssaviruksen evoluution menestyksenkäs ennustaminen perustuu kelpoisuusmallin oppimiseen jo tapahtuneesta evoluutiosta kerätyn genomidata-aikasarjan avulla (Luksza, Lässig 2014). Kyseinen malli arvottaa viruskantojen mutaatiot niin, että joka kannalle voidaan rakentaa arvio kelpoisuudesta. Kyseisessä kelpoisuusmallissa uudet mutaatiot voivat joko parantaa kelpoisuutta, esimerkiksi muuttamalla viruksen epitooppi-sekvenssiä, jonka ihmisen immuunisysteemi pyrkii tunnistamaan, tai heikentää kelpoisuutta, jos ne esimerkiksi destabiloivat proteiineja, joita sekvenssi koodaa. Näihin kelpoisuusarvioihin perus-

tuen voidaan sitten ekstrapoloida lähitulevaisuuteen, tässä tapauksessa seuraavaan kauteen, miten eri kantojen suhteelliset määrät kehittyvät ja siten varautua paremmin.

Kelpoisuuden mallintamien voi olla hyvinkin vaikeaa, varsinkin jos mutaation kelpoisuus riippuu vahvasti yksilön perimästä (niin sanottu epistaattinen vuorovaikutus). Äärimmäisessä tapauksessa kelpoisuusmaasto (*fitness landscape*) on sellainen, että vain yksi tai muutama polku (uusi mutaatio) johtaa ylöspäin ja kaikki muut heikentävät kelpoisuutta. Tämä voidaan havaita kokeissa esimerkiksi siten, että replikaattipopulaatioiden perimä vaihtuu hyvin samalla tavalla. Tällainen kelpoisuusmaasto johtaa suureen toistettavuuteen mutta tekee ennustamisesta erittäin vaikeaa, jos ennuste pitää tehdä luonnolliselle populaatiolla ilman replikaatteja, koska kelpoisuusmaastoa ei voida helposti arvioida. Joten suurikaan toistettavuus ei takaa ennustettavuutta yksittäiselle populaatiolle käytännössä.

Toisaalta evoluutiota voidaan myös ennustaa tekemällä kokeita ilman, että yritetään eksplisiitisti mallintaa tai mitata kelpoisuusmaastoa. Tämä vaatii, että on olemassa riittävän hyvä malli kohdepopulaatiosta, jonka evoluutiota halutaan ennustaa. Potilaan syövästä eristetty syöpäsolulinja, jonka lääkeväste pyritään kartoittamaan kokeilla, voidaan siten myös nähdä evoluution ennustamisen valossa.

Ennustaminen on helpompaa, jos riittää arvioida, miten jo havaitut kannat kehittyvät toisiinsa nähden ilman uusia mutaatiota. Uusia mutaatiota kuitenkin tapahtuu, ja ne voivat muuttaa systeemin suunnan radikaalistikin ja siten tuhoata nykyhetken kelpoisuuteen perustuvan ennustettavuuden. Esimerkiksi alkuvuodesta arvotetut viruskannat saattavat mutatoitua ja kantojen suhteelliset voimasuhteet muuttuvat ennen seuraavaa kautta. Influenssaviruksen esimerkissä tämä olisi erityisen harmillista, koska rokotteen valmistus olisi jo käynnissä, joten ennustuksen muuttaminen uuden datan perusteella ei tässä tapauksessa auttaisi. Eri sovelluskohteissa on erilaiset tarpeet ennustusaikahorisontille. Jos on tarve tehdä ennustuksia pitkälle aikavälille, myös uusien mutaatioiden vaikutus pitäisi pystyä ottamaan huomioon. Tämä on tosin usein mahdotonta, koska useat

eri mutaatiot voivat johtaa samankaltaiseen fenotyyppiin ja siten kelpoisuuteen. Yleisesti ottaen ennustettavuus lisääntyy ja tulee mahdolliseksi, kun siirrytään yksittäisten mutaatioiden ennustamisen sijaan kohti niiden vaikutusten ennakoimista fenotyyppeihin ja kelpoisuuksiin.

Jos olosuhteet ovat kuitenkin äärimmäiset, kuten lääketerapian aikana, saattaa yksittäisten mutaatioiden ennustaminen tulevaisuudessa onnistua. Esimerkkinä tästä mahdollisuudesta käyvät evoluutiokokeissa havaitut biologisten replikaattien identtiset resistenssimutaatiot, joiden kelpoisuudet eivät ainakaan ratkaisevasti riipu yksilöiden perimästä (ei epistaattista vuorovaikutusta; Vázquez-García ym. 2017).

Joitakin ennustamisen mallintamisen pullonkauloja

Niin kuin useilla muillakin tieteen aloilla evoluution ennustamisen kehityksen veturina toimii kokeiden, datan sekä teorian tuottoisa yhdistely. Kuitenkin kaikilla näillä osa-alueilla on vielä suuria avoimia ongelmia.

Kokeellinen evoluutio, jossa populaatiota propagoidaan laboratoriossa eri olosuhteissa, on osoittautunut oivaksi tavaksi kerätä tietoa evoluutiosta ja sen ennustettavuudesta. Näistä kokeista voidaan kerätä geneettisen datan aikasarjaa, josta voidaan suoraan nähdä, miten evoluutio muokkaa populaation perimää ajan funktiona. Tällaiset kokeet ovat muun muassa näyttäneet, että aseksuaalisesti lisääntyvät organismit muuttuvat ajoittain niin kutsutun kloonien keskinäisen häirintädynamiikan mukaan (*clonal interference*). Tässä dynamiikassa ensimmäistä kelpoisuutta lisäävää mutaatiota kantavat yksilöt aluksi kasvattavat osuuttaan populaatiossa. Jos toinen vielä parempi mutaatio tapahtuu yksilössä, jolla ei tuota ensimmäistä mutaatiota ole, syntyy kilpailu tilanne. Koska mutaatioita ei voi yhdistää aseksuaalisessa lisääntymisessä, johtaa tämä ensimmäisen mutaation häviöön ja poistumiseen populaatiosta siitäkin huolimatta, että mutaatio itsessään oli myös kelpoisuutta lisäävä. Ilmiö ymmärretään nyt varsin hyvin sekä kokeellisesti että teoreettisesti ja on havaittu luonnollisissa populaatioissa, kuten influenssaviruksen evoluutiossa. Tämä ja muut vastaavat perustutkimuksen tulokset ovat avainasemassa ennustavan mallintamisen

kehittämisessä.

Sen sijaan useat muut luonnolliset tilanteet ovat vielä heikosti ymmärrettyjä. Näihin lukeutuvat muun muassa, miten vuorovaikutukset lajien välillä vaikuttavat systeemeihin, esimerkiksi bakteeripopulaation yhteisevoluutio niitä syövien faagien kanssa ja miten antibiootit mahdollisesti vaikuttavat tähän. Samoin myös muuttuvat olosuhteet, joihin muun muassa lääketerapiat voidaan aika yleisesti lukea, ovat vielä heikosti ymmärrettyjä. Myös useimmat evoluutiokokeet ovat toistaiseksi keskittyneet populaatioihin, jotka ovat alussa homogeenisiä, eli geneettisesti identtisiä, vaikka yhä paremmin on tiedossa, että populaation geneettinen heterogeenisuus vaikuttaa dynamiikkaan suuresti ja muun muassa huonontaa joidenkin syöpien kohdalla prognoosia. Näihin perustutkimusaiheisiin huolellisesti pureutumalla saadaan tarvittavaa lisätietoa evoluution ennustamisen käytännön sovelluksia varten.

Maailmassa meneillään oleva niin kutsutun ison datan kasvu tapahtuu myös biologian puolella. Kuitenkin biologia, biolääketiede ja evoluutio eroavat oleellisesti monista muista ison datan osaluista. Biologisen datan hinta on vieläkin korkea verrattuna esimerkiksi tilanteeseen, jossa jotkut suoraan internetistä saatavat aineistot, esim. videot tai keskustelupalstat, ovat tutkimuksen kohteena. Samoin joidenkin periaatteessa nopeisiin lukeutuvien systeemien evoluutio voi tuntua hitaanpuoleiselta tutkijan näkökulmasta, joten pitkän aikasarjan kerääminen vaatii kärsivällisyyttä, ellei systeemejä voi tavalla tai toisella tutkia rinnakkaisesti. Esimerkiksi syöpäsolulinjojen propaagointi eri lääkkeiden vaikutuksen alla voidaan tehdä rinnakkaisesti.

Lähitulevaisuus

Perinteinen fokus evoluutiotutkimuksessa jo tapahtuneiden asioiden tulkintaan on saamassa rinnalleen uuden suunnan, missä nopeasti muuttuvien organismien populaatioiden perimän evoluutiota pyritään ennustamaan. Satunnaisuus mutaatioissa, lisääntymisessä ja olosuhteissa tekevät evoluution ennustamisesta haastavaa. Systeemeillä on todennäköisesti jokin ennustettavuuden aikahorisontti, jonka taakse ei höydyllisiä ennustuksia voida tehdä. Influenssavirukselle tämä hori-

sontti näyttäisi olevan sellaisella etäisyydellä, että höydyllisiä ennustuksia voidaan tehdä jo tämänhetkiselällä tieteellisellä ymmärryksellä. Missä rajat kulkevat muilla systeemeille, on aktiivinen tutkimuskohde. Lähivuodet näyttävät, kuinka ennustettavaa erilaisten populaatioiden evoluutio on sekä miten ennustettavuutta voidaan parantaa yhdistämällä kokeita, dataa ja teoriaa uusin tavoin.

Kirjallisuutta

- Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. *Nature* 2017 Nov 2;551(7678):45–50. doi: 10.1038/nature24287. Epub 2017 Oct 18. PMID:29045390.
- Lipinski KA, Barber LJ, Davies MN, Ashenden M, Sottoriva A, Gerlinger M. Cancer Evolution and the Limits of Predictability in Precision Cancer Medicine. *Trends Cancer*. 2016 Jan;2(1):49–63. Review. PMID:26949746.
- Luksza M, Lässig M. A predictive fitness model for influenza. *Nature* 2014 Mar 6;507(7490):57–61. doi: 10.1038/nature13087. Epub 2014 Feb 26. PMID:24572367.
- Lässig M, Mustonen V, Walczak AM. Predicting evolution. *Nat Ecol Evol*. 2017 Feb 21;1(3):77. doi: 10.1038/s41559-017-0077. PMID:28812721.
- Morris DH, Gostic KM, Pompei S, Bedford T, Luksza M, Neher RA, Grenfell BT, Lässig M, McCauley JW. Predictive Modeling of Influenza Shows the Promise of Applied Evolutionary Biology. *Trends Microbiol*. 2018 Feb;26(2):102–118. doi: 10.1016/j.tim.2017.09.004. Epub 2017 Oct 30. Review. PMID:29097090.
- Vázquez-García I, Salinas F, Li J, Fischer A, Barré B, Hallin J, Bergström A, Alonso-Perez E, Warringer J, Mustonen V, Liti G. Clonal Heterogeneity Influences the Fate of New Adaptive Mutations. *Cell Rep*. 2017 Oct 17;21(3):732–744. doi: 10.1016/j.celrep.2017.09.046. PMID:29045840.

Kirjoittaja on Helsingin yliopiston bioinformatiikan professori.