Biologie als data science

Zen en de kunst van de bioinformatica

Prof. dr. ir. Dick de Ridder

Inaugurele rede bij de aanvaarding van het ambt van
hoogleraar in de bioinformatica
aan Wageningen University op 30 april 2015

WAGENINGEN UNIVERSITY
WAGENINGEN UR

ISBN 978-94-6257-367-3

Biologie als data science

Zen en de kunst van de bioinformatica

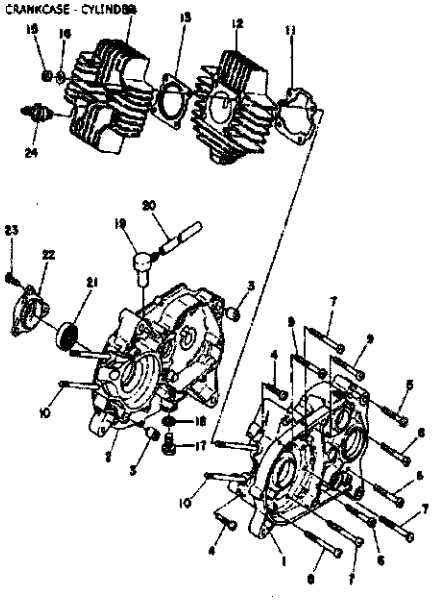
1. Introductie

Mijnheer de rector, collega's, studenten, familie, vrienden en alle overige aanwezigen: welkom! Ik stel het bijzonder op prijs dat u van heinde en verre bent gekomen om mijn openbare lezing bij te wonen. Het is een voorrecht om voor een breed publiek als dit te praten over mijn vak. Dat vak is de bioinformatica, ofwel de toepassing van computertechnologie – hardware en vooral software – in de verwerking, analyse en interpretatie van biologische data. Ik hoop vanmiddag duidelijk te maken wat de rol van de bioinformatica in de moderne biologie is en wat de toekomst brengt aan mogelijkheden en uitdagingen. Daarbij wil ik niet alleen een idee geven van het onderzoek en onderwijs dat mijn groep verricht, maar ook iets vertellen over het doel waar wij als onderzoekers hier in Wageningen en elders in de wereld naar toe werken. Kortgezegd is dat het begrijpen van hoe het leven werkt, op het niveau van moleculen tot cellen.

In dat opzicht is de ondertitel van mijn lezing misschien wat cryptisch. Het is een parafrase van de titel van een boek dat mij sterk heeft beïnvloed en dat ik nog steeds regelmatig cadeau geef aan promovendi en afstudeerders. Dat boek heet "Zen en de kunst van het motoronderhoud" en is geschreven door Robert Pirsig[1]. Het boek is al vrij oud, dat wil zeggen, ongeveer twee jaar jonger dan ik zelf. Mocht U onverhoopt bang zijn dat ik deze lezing gebruik om U bepaalde religieuze overtuigingen op te dringen, dan kan ik U geruststellen: mijn lezing gaat, net als het boek, vrij weinig over motorfietsen en eigenlijk helemaal niet over Zen. In essentie is het boek een filosofische verhandeling over de relatie tussen ratio, emotie en vooral kwaliteit. De schrijver heeft dat gegoten in de vorm van een "road novel". Het beschrijft een motortocht door de Verenigde Staten van een vader met zijn zoon en een stel vrienden. Het bevat ook een paar ideeën die volgens mij helpen om na te denken over wetenschappelijk onderzoek en onderwijs, en de organisatie daarvan. Daarvan wil ik er een aantal gedurende de lezing met u delen.

Een belangrijk thema in het boek is het onderscheid tussen de klassieke blik op de wereld – reductionistisch, objectief – en de romantische, ofwel holistische, subjectieve blik. Het boek illustreert dat aan de hand van de motorfietsen waarop de hoofdpersonen door de Verenigde Staten rijden. De verteller ziet zijn motor als een systeem dat volledig begrepen kan en zelfs *moet* worden om het te waarderen: de rol en plaats van alle onderdelen liggen vast, en als er iets misgaat kan een goede mecanicien door zorgvuldig onderzoek en redeneren het probleem vinden en oplossen. De vrienden van de verteller daarentegen hebben een romantischer wereldbeeld. Ze rijden op een mooie, nieuwe motor en waarderen die vooral om zijn uiterlijk, zijn nut en om de vrijheid die het ze geeft om te gaan en te staan waar ze willen.

Toen ik dit boek tijdens mijn studietijd las was voor mij het klassieke beeld heel herkenbaar. Misschien kwam dat omdat ik als tiener graag sleutelde aan mijn brommer, maar het "motoronderhoud" dat het boek beschrijft, heeft ook veel weg van het programmeren van computers. Ook daarin is het belangrijk om een volledig mentaal beeld op te bouwen van een probleem, en de datastructuren en de algoritmen die dat representeren in de computer. Vervolgens kan door zorgvuldig programmeren en debuggen een goed werkend stuk software worden gemaakt. Maar gaandeweg besefte ik dat de klassieke aanpak ook vernauwend kan werken. Zonder de romantische blik kun je misschien nog wel begrijpen *hoe* een systeem in elkaar zit (figuur (a)),

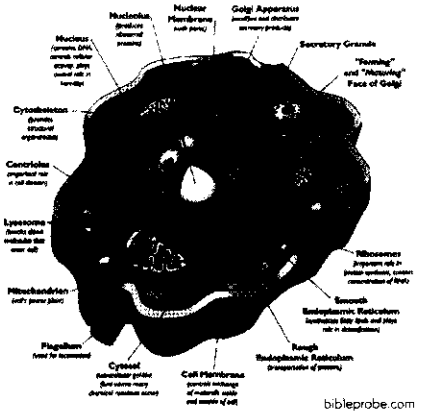


(a) www.cmsnl.com

(b) www.fs1eoc.co.uk

maar niet per se *waarom* dat systeem zo werkt, waartoe het dient of hoe je het zou kunnen verbeteren. Zo laat figuur (b) hier zien dat het rijden op een brommer heel andere doelen kan hebben dan alleen vervoer.

2. Biologie



Als brommersleutelende, programmerende scholier was ik geen groot liefhebber van het vak biologie. Het kwam op mij over als het leren van een groot aantal feiten: welke botten en aders bevinden zich waar, welke onderdelen heeft een cel, etc. Het volgde meer de romantische dan de klassieke aanpak. Ik kreeg daarmee niet de indruk dat biologie een wetenschap was als de natuurkunde, dat wil zeggen een poging te verklaren hoe de wereld in elkaar zit. Daarmee zat ik er flink naast. Dat werd me duidelijk toen ik ruim 10 jaar geleden de overstap maakte naar de bioinformatica, een onderzoeksgebied dat toen sterk in opkomst kwam. Ik leerde al snel dat veel biologen proberen het leven op het kleinste niveau te begrijpen, en dat dat fascinerend onderzoek oplevert. Dit figuur laat dat zien. Het is een cel, het basiselement van ieder organisme. We weten inmiddels dat zich in de celkern DNA-moleculen bevinden, die samen het genoom vormen. Dat DNA bevat bepaalde gebieden, genen, die afgelezen worden in de vorm van RNA-moleculen en vervolgens vertaald in eiwitten. Die eiwitten op hun beurt zorgen voor de structuur van de cel en voor allerlei functies. Ze zetten voedsel om in energie, zorgen voor de productie, vervoer, opslag en afbraak van andere eiwitten, geven signalen binnen en tussen cellen door enzovoorts. De cel is dus eigenlijk een complex, van de buitenwereld afgesloten systeem van verschillende soorten moleculen, die samen zorgen voor groei, activiteit en aanpassing aan de omgeving.

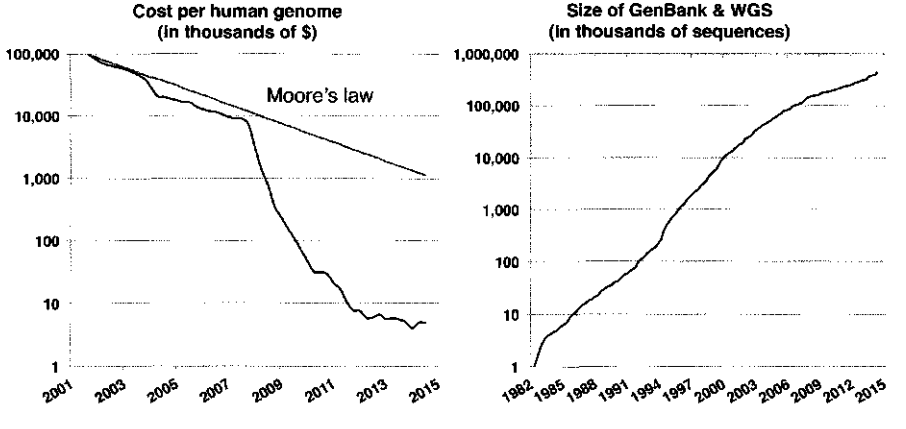
Dit figuur geeft misschien de indruk dat de handleiding voor het bouwen van een levende cel nu wel bijna compleet is. Die indruk is misleidend, om drie redenen. Ten eerste, omdat het voorbijgaat aan het feit dat we de meeste lessen pas de afgelopen 60 tot 70 jaar hebben geleerd, en we nog lang niet uitgeleerd zijn. Nog steeds worden nieuwe soorten moleculen gevonden, en nieuwe interacties tussen die moleculen, die van essentieel belang zijn om te begrijpen hoe het leven op dit niveau werkt. Ten tweede is het figuur statisch, terwijl vooral de dynamische interacties tussen alle moleculen zorgen voor de complexiteit van en de variatie tussen cellen. Het genoom

wordt dan wel het "boek van het leven" genoemd, maar dit boek wordt niet simpelweg van kaft tot kaft gelezen: elke pagina bevat instructies over hoe paragrafen van andere pagina's gelezen, vertaald en geïnterpreteerd moeten worden. En ten derde is het misleidend omdat de reductionistische aanpak weliswaar een beeld geeft *welke* processen plaatsvinden in de cel, maar we nog ver af staan van het volledig begrijpen *hoe* die processen werken en vooral *waarom* ze zo werken.

Desalniettemin stelt de kennis die we nu hebben ons al in staat om een scala van praktische problemen aan te pakken. Wageningen kent veel voorbeelden van zulk onderzoek. Zo kunnen we voorspellen welke planten we wanneer moeten kruisen om de juiste combinatie van genen te krijgen voor voedzame gewassen die tegen droogte en ziekte kunnen[2]. We kunnen micro-organismen genetisch aanpassen om afval op een duurzame manier om te zetten in brandstof, voedsel of geneesmiddelen[3]. Daarnaast leren we hoe we met de juiste combinatie van gewassen en bodembacteriën groenten en fruit kunnen verbouwen met minder mest en bestrijdingsmiddelen[4].

3. Data

De belangrijkste factor achter ons snel toegenomen begrip van de werking van de cel is de snelle ontwikkeling van de moleculaire biologie. Dit gebied houdt zich bezig met methoden om processen in de cel te beïnvloeden, en de concentraties van en interacties tussen moleculen te meten. Het meest bekende voorbeeld is misschien wel het werk van Watson & Crick, die in 1953 het helix-model van DNA-moleculen voorstelden op basis van röntgendiffractiemetingen[5]. Sindsdien is er een stormachtige technologische vernieuwing op gang gekomen, die ons in staat stelt om van veel soorten moleculen te meten of ze aanwezig zijn, in welke concentraties, en met welke



(a) www.genome.gov/sequencingcosts/ (b) www.ncbi.nlm.nih.gov/genbank/statistics

andere moleculen ze een interactie hebben. Er zijn veel grafieken die dat illustreren, maar deze twee zijn het meest bekend en veelzeggend. Links ziet u een grafiek die weergeeft hoe de kosten van het aflezen van de 3,3 miljard basen in een menselijk genoom zijn gedaald over de afgelopen jaren. Het Human Genome Project, dat liep van pakweg 1990 tot 2003, kostte naar schatting 2,7 miljard dollar[6]. In 2014 waren de kosten gedaald tot iets meer dan 4000 dollar. Er zijn weinig gebieden waar de ontwikkeling zo snel gaat. Zelfs de microelektronica, waar de wet van Moore zegt dat de kosten voor een bepaalde hoeveelheid rekenkracht elke twee jaar halveren (de grijze lijn in de figuur), blijft daar ver bij achter.

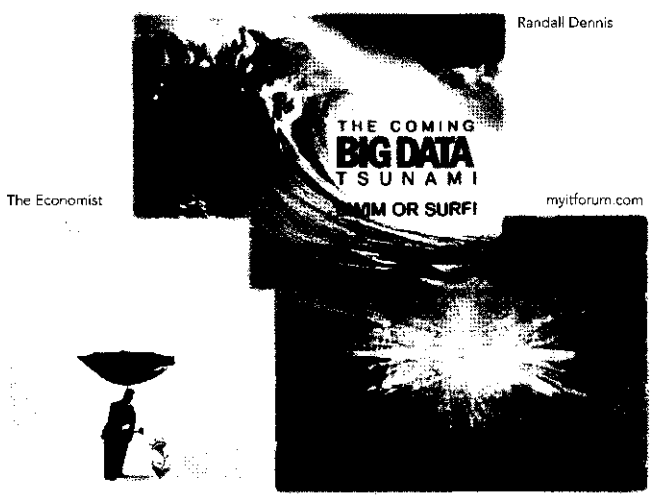
Het gevolg van die dalende kosten van metingen is een stijging van de hoeveelheid meetdata. De grafiek rechts laat zien hoe een van de belangrijkste databases, GenBank, is gegroeid sinds de start in 1982. Het aantal DNA sequenties in die database is exponentieel toegenomen van een paar 100 tot bijna een half miljard in 2014. Het aantal organismen waarvan we het genoom kennen is in die tijd gestegen van twee tot vele tienduizenden. Ook in Wageningen hebben collega's van mij de laatste jaren hieraan bijgedragen. Een aantal voorbeelden – de genomen van aardappel[7], tomaat[8], varken[9], kip[10] en aardappelrot[11] – hebben de voorkant van *Nature* gehaald, maar de lijst is nog veel langer. En ook hier neemt het gemak waarmee zulke genomen kunnen worden gesequenced snel toe. Waar het aflezen van het tomatengenoom nog een groot, internationaal project van een jaar of 5 was, zijn recent hier in Wageningen de genomen van 150 tomaten in pakweg 1 jaar afgelezen[12]. Ter illustratie: de verwachting is dat we in 2015 25 *petabasen* – 25 met 15 nullen, een miljoen miljard – aan DNA kunnen aflezen, het equivalent van 250.000 menselijke genomen[13]. Als de groei zich exponentieel door zou blijven zetten kunnen we binnen 10 jaar *elk jaar de hele mensheid* sequencen. Dat lijkt me weinig zinvol meer, en daarom zal die groei binnen afzienbare tijd wel afvlakken, maar het illustreert de enorme technologische stappen die we zetten.

De technologie om DNA te lezen is misschien het meest indrukwekkend, maar er is ook apparatuur ontwikkeld om allerlei andere moleculen en interacties te meten: massa-spectrometrie voor metabolieten en eiwitten, microarrays en sequencing voor mRNA moleculen, precipitatietechnieken voor eiwit-DNA interacties, hybridisatietechnieken voor eiwit-eiwit interacties, kristallografie voor 3D eiwit-structuren, imaging voor localisatie, etc. Naast het verzamelen van dit soort ruwe data is er veel werk gestoken in het aanleggen van databases, waarin opgedane kennis over de eigenschappen, structuren en functies van moleculen worden opgeslagen. Op dit moment zijn er naar schatting tussen de 1000 en 2000 van dit soort databases. Het gevolg is dat er een enorme hoeveelheid data is waarop we onze biologische experimenten kunnen baseren, waarmee we de uitkomsten moeten vergelijken en die we kunnen doorzoeken naar nieuwe inzichten.

4. Data science in de biologie

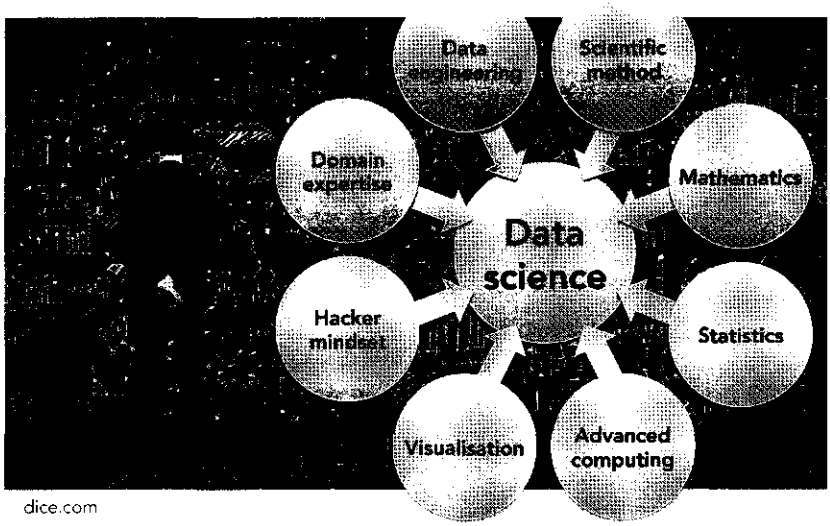
Nu we hebben vastgesteld dat er gigantische hoeveelheden data geproduceerd worden in de biologie, en dat er nog veel meer aankomt, dringt zich de vraag op: *wat moeten we met al die data?* Deze vraag, en de problemen die de data met zich meebrengen, worden de laatste tijd kortweg aangeduid met *"big data"*. Nu is die aandacht voor "big data" niet uniek voor de biologie. In veel gebieden treedt een soortgelijk effect op: namelijk dat de snelheid waarmee data gegenereerd wordt, en het gemak waarmee het opgeslagen wordt, veel hoger ligt dan de snelheid waarmee die geïnterpreteerd kan worden. In de hoge energie fysica bijvoorbeeld genereert de Large Hadron Collider bij het CERN in Genève in één jaar ook 25 petabyte aan data. In de astronomie en de aardobservatie zijn tientallen petabytes aan beelddata publiekelijk beschikbaar voor analyse. Retailketens, zoals Walmart en Albert Heijn, leggen enorme databases aan met transacties. En internetbedrijven als Facebook en Google verzamelen honderden petabytes aan data over gebruikers, hun voorkeuren en netwerken. Van Google wordt zelfs geschat dat ze enkele exabytes (duizenden petabytes) aan totale opslagcapaciteit hebben.

Als het gaat over "big data" wordt er vaak gesproken van een data "zondvloed", "tsunami" of "explosie", waarvoor we ons moeten hoeden. Door terminologie te gebruiken die we normaal voor natuurrampen of terroristische aanslagen bewaren, klinkt het alsof we een enorm probleem hebben. In mijn optiek valt dat allemaal reuze mee. Ja, er is een praktisch probleem in hoe we die data opslaan, veilig stellen en doorzoekbaar maken. Maar Google en andere bedrijven laten zien dat we dat



prima kunnen oplossen, als we er maar genoeg geld voor over hebben. En ja, er is een probleem dat we die data niet makkelijk in zijn geheel kunnen analyseren, interpreteren en integreren. Maar in feite is dat een luxeprobleem; we vinden ook niet dat bibliotheken een bedreiging vormen omdat niemand alle boeken makkelijk kan doorlezen, of dat er meer muziek gedownload kan worden via Spotify dan een mens in zijn leven kan beluisteren. Op basis van specifieke, persoonlijke voorkeuren voor literatuur of muziek weet iedereen hier een keuze in te maken en verbanden te vinden.

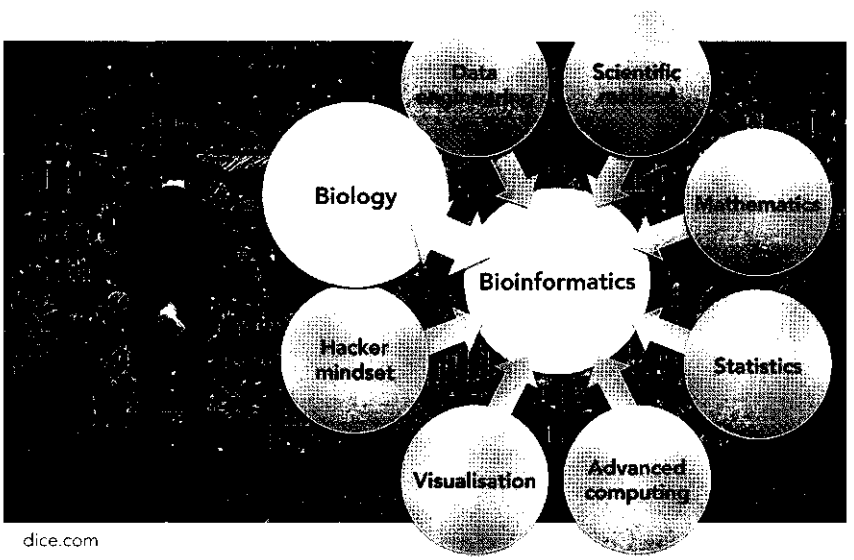
De meest interessante vraag met betrekking tot "big data", zeker in de biologie, is wat we eruit kunnen leren. Jim Gray van Microsoft noemde dit in 2007 het vierde wetenschappelijke paradigma[14]. De meeste wetenschap, zeker de bètawetenschap, heeft als doel te generaliseren, dat wil zeggen regels en wetten te vinden die het onderwerp van onderzoek beschrijven. De traditionele manier van wetenschap, het "eerste paradigma", is dat een onderzoeker een hypothese opstelt, daarmee een voorspelling doet, die empirisch test door een experiment uit te voeren, en dan op basis van de resultaten de hypothese al dan niet verwerpt. In de tweede en derde paradigma's generaliseren we op basis van modellen en simulaties. Maar sinds kort kunnen we ook proberen verbanden af te leiden uit de grote hoeveelheden data waar we de beschikking over hebben, vaak zelfs zonder een duidelijke hypothese te stellen. Jim Gray noemde dit data-intensieve wetenschap, en de onderzoekers die zich bezighouden met het bedenken van computeralgoritmen en modellen om de data te bestuderen *data scientists*.



dice.com

Daarmee ontstond het idee dat het stellen van een hypothese niet langer nodig was, de zogenaamde hypothese-vrije aanpak. Dat bleek niet helemaal juist; zeker wanneer je grote hoeveelheden data doorzoekt is het essentieel om voorkennis en vooronderstellingen te gebruiken om de zoektocht te laten slagen. Het is wel zo dat de data zelf tot nieuwe hypothesen kan leiden. Zo kunnen we met bioinformatica tools voorspellingen doen over de effecten van genetische varianten, of over de mechanismen die ten grondslag liggen aan bepaalde ziekten. Maar die hypothesen moeten over het algemeen weer ouderwets in het laboratorium worden getest.

Mijn stelling is dat biologie de laatste jaren snel het karakter van zo'n data science krijgt, en dat wij bioinformatici de data scientists van de biologie zijn. Niet alleen wordt het verwerken en analyseren van meetgegevens steeds belangrijker voor vrijwel elk soort onderzoek, maar ook de interpretatie van de resultaten en het stellen van nieuwe hypothesen gebeurt steeds vaker op basis van al aanwezige data. Bioinformatici zorgen voor de gereedschappen en modellen die dit mogelijk maken, op basis van kennis van data analyse, wiskunde, statistiek en informatica. Dat combineren we met wat in dit figuur de "hacker mindset" wordt genoemd, iets dat ik (misschien wat te positief) zou willen vertalen met "doelgericht en niet bang om de handen vuil te maken". Daarbij wordt kennis van en interactie met het toepassingsdomein – de biologie – steeds belangrijker. En waar nu data analyse vaak wordt uitbesteed door biologen aan bioinformatici, verwacht ik dat er binnen afzienbare tijd steeds meer onderzoekers komen die vrijwel al hun werk met de computer verrichten, en juist de experimentele validatie van hun voorspellingen uitbesteden.

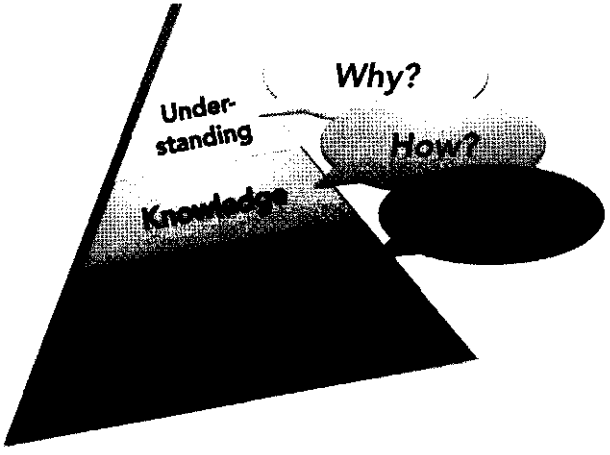


dice.com

5. Bioinformatica: van data naar begrip

Om wat concrete voorbeelden te laten zien van het werk dat we doen, is het handig de data-informatie-kennis-begrip piramide er bij te halen[15]. Traditioneel wordt de bovenste trede aangeduid met "wijsheid", maar dat lijkt me wat te hoog gegrepen, dus ik houd het bij "begrip". Deze manier van denken over data-analyse is al vrij oud, maar nog steeds actueel. Stel dat we aan een biologisch probleem werken waarvoor we een grote set meetgegevens verzamelen, bijvoorbeeld DNA sequentiedata van een bepaalde plant. De eerste stap is dan die van de ruwe data naar informatie, waarin we proberen te leren *wat* er precies in de data aanwezig is – welke genen zien we, en wat voor functies hebben die genen? Vervolgens proberen we uit die informatie kennis te destilleren, waarmee we vragen kunnen beantwoorden over *hoe* bepaalde processen werken. Bijvoorbeeld: als deze plant een bepaalde set van genen heeft, kan het zich goed verdedigen tegen bepaalde ziekten of maakt het bepaalde smaakstoffen aan. En tenslotte de laatste stap, in de hoop op basis van die kennis begrip op te bouwen: *waarom* heeft deze plant die mechanismen? Kunnen we dat verklaren op basis van evolutie of uit de interactie met de omgeving?

De piramide vorm geeft aan dat in elke stap de hoeveelheid gegevens sterk vermindert: het is niet ongebruikelijk om uit vele terabytes aan originele meetdata uiteindelijk één of twee vormen van begrip af te leiden. Voor mij geeft het ook aan hoe we van het terrein van de informatica bewegen in de richting van de biologie. In zekere zin gaat dit terug naar het onderscheid tussen de reductionistische en de holistische aanpak waar ik het eerder over had. Op het onderste niveau werken we met de "nuts en bolts", specialistische algoritmen en hardware voor dataverwerking. Vervolgens



kunnen we modellen opstellen voor biologische processen, door die data met al aanwezige kennis te combineren. Die modellen leiden tot nieuwe hypotheses, die na toetsing tenslotte biologisch begrip kunnen opleveren. Ik denk dat dit ook illustreert hoe het ontwikkelen van bioinformatica tools op het onderste niveau niet veel zin heeft zonder een idee te hebben van het uiteindelijke doel: het verkrijgen van biologisch inzicht. Maar omgekeerd kan een bioloog het zich niet veel langer permitteren om uitkomsten van tools te accepteren zonder enig begrip te hebben van de onderliggende stappen in de data analyse, en vooral de beperkingen daarin.

Voor mij persoonlijk is het beklimmen van deze piramide een uitdaging. Hoe kan ik mijn kennis over algoritmen en modellen zo goed mogelijk gebruiken om van data naar begrip over de biologie te komen? Daarvoor is stevige interactie nodig met wetenschappers in de biologie, wiskunde en natuurkunde, met onderzoekers in het bedrijfsleven en vooral ook tussen "scientists" en "engineers". Voor mij was dit de belangrijkste reden om naar Wageningen te komen. Wageningen University & Research Centre heeft een in Nederland unieke focus in onderzoeksgebied, en combineert daarbij topwetenschap met een ingenieursaanpak en een interesse in de toepassing van de opgedane kennis. We hebben het laatste jaar hard gewerkt aan het uitbouwen van de Bioinformatics Group tot een groep die alle kennis en vaardigheden in zich heeft om fundamenteel bioinformatica-onderzoek (voor zover dat geen *contradictio in terminis* is) te combineren met uitdagende toepassingen.

In de komende secties wil ik graag wat specifieke voorbeelden geven van het werk dat we in de Bioinformatics Group verrichten in de drie stappen van data naar informatie, van informatie naar kennis en van kennis naar begrip.

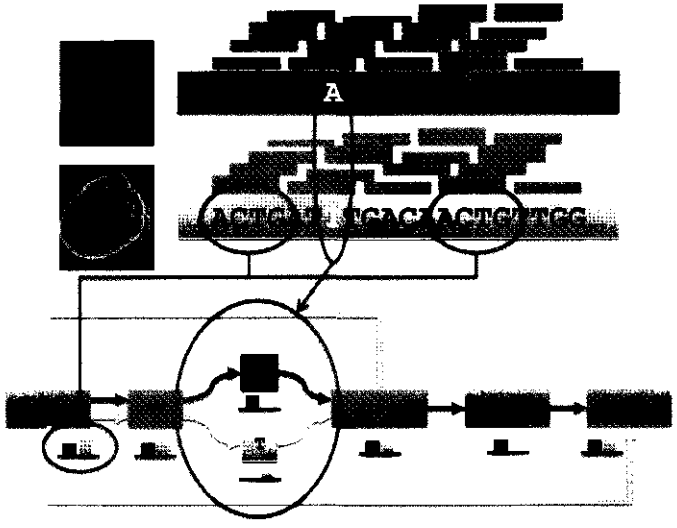
6. Van data naar informatie: *wat*?

Allereerst de dataverwerking, waarin we onszelf de volgende vraag stellen: hoe krijgen we zoveel mogelijk betrouwbare informatie uit de ruwe meetdata? Dat is vaak geen eenvoudige opgave, en het vereist inzicht in de manier waarop de data gemeten is. Verschillende soorten technologie brengen verschillende beperkingen en fouten met zich mee. De data moet daarom eerst worden opgeschoond, genormaliseerd en verwerkt tot een niveau waarop verdere analyse mogelijk is. Daar zijn gespecialiseerde computeralgoritmen voor nodig.

Een goed voorbeeld is het aflezen van DNA sequenties, waar ik eerder over sprak. Een heel genoom in één keer aflezen lukt technisch nog niet. Daarom moeten we het reconstrueren op basis van overlappende, kortere stukken, een proces dat assembleren heet. Een goede analogie voor dat assembleren is het oplossen van een legpuzzel. De technologie die het sinds een jaar of tien mogelijk maakt om genomen snel en

goedkoop af te lezen levert maar heel korte stukjes DNA op, ook wel "reads" genoemd. En waar wij mensen van elk chromosoom twee kopieën hebben, hebben veel planten meer kopieën – van vier tot soms zelfs twaalf – met heel veel herhalingen in het genoom. In de legpuzzel-analogie betekent dat dus het tegelijkertijd oplossen van een paar puzzels, met miljoenen stukjes die erg op elkaar lijken.

In een samenwerking met de groep van Jack Pronk in Delft hebben we gewerkt aan de genomen van giststammen, die gebruikt worden om onder andere bio-ethanol en bier te maken. Daarvoor hebben we algoritmes ontworpen die DNA reads gebruiken om niet alleen iets te leren over één genoom, maar over meerdere genomen tegelijk. Een van de ideeën die we onderzocht hebben is dat, als twee genomen veel op elkaar lijken, we ze in de computer efficiënt kunnen opslaan in een zogenaamde co-assembly graaf. In zo'n graaf is elke knoop een uniek stukje sequentie en komt elk pad tussen die knopen overeen met een van de genomen; in dit voorbeeld kan het donkere genoom worden teruggevonden door het donkergrijze pad te volgen en het lichte genoom door het lichtgrijze pad te volgen. Deze graaf kunnen we op een paar manieren gebruiken om meer te leren over de verschillen tussen de genomen. Zo verwachten we bijvoorbeeld van een gebied dat in één genoom vaker voorkomt dan in het andere meer reads aan te treffen. Door dit wiskundig te modelleren hebben we een efficiënt algoritme kunnen ontwerpen om zulke gebieden te vinden, zonder dat daarvoor een referentiegenoom nodig is[16]. In een andere toepassing hebben we gekeken of we automatisch alle kleine variaties tussen genomen kunnen vinden. Daar is een mooie methode uitgekomen die snel alle zogenaamde "bubbles" in co-assembly grafen zoekt[17].



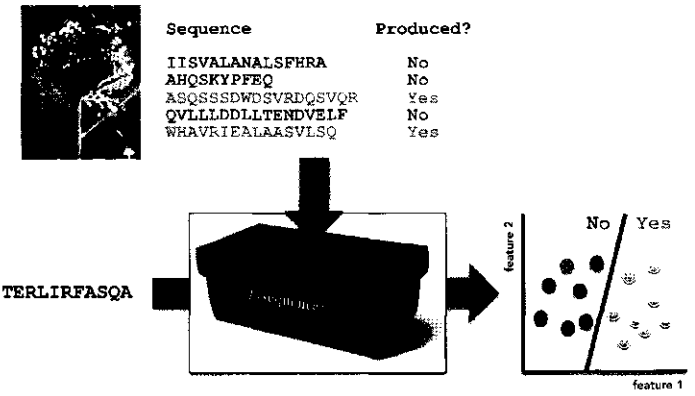
In de toekomst bouwen we verder op deze ideeën. Als we twee genomen kunnen co-assembleren, kunnen we dat ook proberen met honderden genomen. De resulterende graaf is dan een representatie van de "rode draad" door die genomen en alle variaties daarop, een zogenaamd *pangenoom*. Maar waar het nog relatief eenvoudig is om een paar giststammen te co-assembleren, omdat ze een klein genoom hebben, is het co-assembleren van een paar honderd plantgenomen andere koek. En juist daar is behoefte aan in de plantenveredeling. Net als het 150 tomaten project waar ik het eerder over had, zijn er nu ook projecten onderweg om honderden genomen van andere gewassen af te lezen. De uitdaging is dus om door de bomen het bos te kunnen blijven zien. Daarvoor doen we onderzoek naar hoe we zo'n pangenoom graaf efficiënt kunnen opslaan en doorzoeken, en hoe we het kunnen visualiseren zodat biologen makkelijk de meest relevante genetische variatie kunnen bekijken.

Ook in andere projecten proberen we de laatste technologie in te zetten om de samenstelling van genomen te ontrafelen. Ik geef twee voorbeelden. Het eerste is dat we via onze collega's bij Plant Research International sinds kort de beschikking hebben over nieuwe technologie die lange stukken DNA met camera's opneemt. Deze zogenaamde "optical mapping" gaat ons helpen om snel en tegen lage kosten veel te weten te komen over de grootschalige structuur van genomen. Een tweede voorbeeld is een samenwerking met Chris Maliepaard, waarin we proberen uit te zoeken hoe we betrouwbaar genetische varianten kunnen vinden en toekennen aan individuele chromosomen in gewassen waarin van elk chromosoom meerdere kopieën zijn, zoals in aardappels. In beide gevallen staat de software om de data te verwerken nog in de kinderschoenen, wat voor de komende jaren genoeg uitdagingen biedt.

7. Van informatie naar kennis: *hoe*?

Als we de ruwe data hebben verwerkt tot informatie, zijn we meestal nog ver verwijderd van begrip van de biologie. Als we een genoom geassembleerd hebben kunnen we iets zeggen over welke genen we hebben gevonden. Maar dat levert ons niet direct inzicht in hoe die genen samenwerken in specifieke biologische processen. Daarvoor is een volgende stap nodig, het maken van *modellen* om te generaliseren op basis van de observaties. Een belangrijke vraag hierbij is hoe we alle kennis die we al hebben over een proces mee kunnen nemen in zo'n model. Om terug te vallen op de motorfiets-analogie: we proberen hier te achterhalen hoe alle onderdelen samenwerken om een werkende motor te krijgen.

Een voorbeeld van dit soort werk is het voorspellen van gedrag of functie van moleculen. Dit gebeurt vaak met behulp van zogenaamde *black-box* modellen: generieke methoden uit de patroonherkenning en machine learning die een bepaalde uitvoer voorspellen op basis van een aantal kenmerken. Zo hebben we de afgelopen jaren



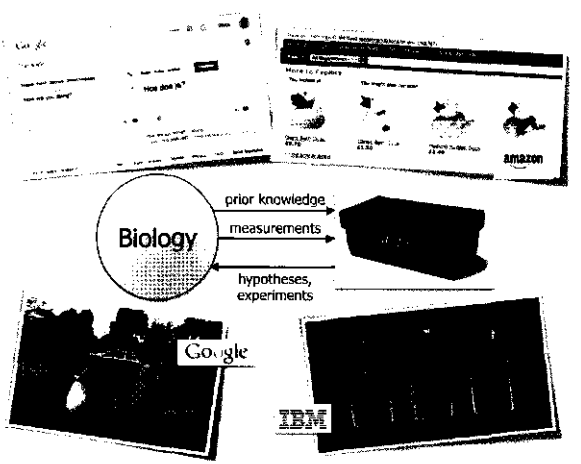
gewerkt aan de voorspelling van de produktie van eiwitten door een bepaalde schimmel. Die schimmel wordt gebruikt om eiwitten op industriële schaal te maken, met toepassingen in onder andere voeding, wasmiddelen etc. In een project met DSM hebben we een algoritme gemaakt dat, op basis van de sequentie van aminozuren in een eiwit, kan voorspellen of dat eiwit goed geproduceerd kan worden of niet[18]. Die voorspeller werd vervolgens de basis van een methode om eiwitten te herontwerpen zodat ze beter geproduceerd kunnen worden; hier zal ik later op terugkomen. Het werk leverde ons ook een aantal interessante aanknopingspunten op om de onderliggende biologische processen mee te gaan bestuderen. Een voorbeeld van een minder generieke aanpak is een model dat we hebben ontwikkeld van translatie, het proces waarmee eiwitten worden geproduceerd op basis van mRNA door ribosomen. Er is de laatste jaren een nieuw soort meetdata beschikbaar gekomen die het mogelijk maakt om de beweging van de ribosomen over het mRNA molecuul te modelleren. Dit geeft ons inzicht in eventuele opstoppingen en vormen van regulatie.

Ook in deze richting zetten we ons werk voort, onder andere in samenwerkingen met de collega's van de Plant Sciences Group. Ten eerste bestuderen we, in een samenwerking met Ben Scheres, de manier waarop een bepaalde familie van eiwitten de groei van de wortels van planten aanstuurt. Het doel is om een model te bouwen dat, gegeven de DNA-sequentie rond een gen en de concentraties van die eiwitten, kan voorspellen welke biologische processen aan- of uitgeschakeld worden in bepaalde groeistadia. In een tweede project, met Richard Immink, willen we de interacties tussen leden van een belangrijke familie van regulator-eiwitten modelleren, de evolutie daarvan, en de effecten op ontwikkelingsprocessen in planten. Ten derde werken we samen met Francine Govers om de moleculaire interacties te modelleren tussen tomaten en een schimmel, *Phytophtora infestans* ofwel aardappelrot, op basis van metingen aan metabolieten, mRNA en eiwitten. Het doel is te leren welke processen verant-

woordelijk zijn of geraakt worden met de golven van aanvals- en verdedigingsmoleculen waarmee de schimmel en de plant elkaar bestoken.

Wat deze projecten met elkaar gemeen hebben is dat ze vaak black-box modellen gebruiken, iets dat soms nog op weerstand stuit. Hoe kunnen we ooit specifieke biologische processen begrijpen met behulp van generieke algoritmen? Mijn eerste antwoord zou zijn dat het voor veel toepassingen niet nodig is om de onderliggende biologie volledig te begrijpen. Neem, ter vergelijking, de natuurkunde, waarin de gravitatiewet van Newton eeuwenlang heeft geholpen om nuttige voorspellingen te doen, apparatuur te bouwen etc. – zonder dat we, tot de dag van vandaag, precies begrijpen *hoe* zwaartekracht dan wel precies werkt. Black-box modellen worden tegenwoordig op een soortgelijke manier succesvol toegepast. Google, het bekendste voorbeeld, gebruikt ze in hun vertaal-apps, routeplanners en zelfrijdende auto's; bedrijven als Amazon en bol.com genereren er aanbevelingen mee voor hun klanten; en IBM zet hun Watson systeem, waarmee ze een paar jaar geleden de beste menselijke kandidaten versloegen in de quiz show Jeopardy, nu in voor medische diagnostiek.

Maar dat is niet een echt antwoord op de vraag; geen bioloog zal uiteindelijk tevreden zijn met een model waarmee we een cel perfect kunnen simuleren, maar waarvan we niet weten hoe het intern werkt. Daarom denk ik dat we op de lange termijn naar steeds gedetailleerdere modellen zullen gaan voor specifieke processen. Maar veel van de data die nodig is om zulke modellen op te stellen kan nog niet eens gemeten worden. Daarom zal het voor de afzienbare toekomst nodig blijven om black-box modellen te bouwen die processen op een hoger niveau beschrijven. En die modellen kunnen, net zo goed als andere modellen, bruikbare hypotheses opleveren.

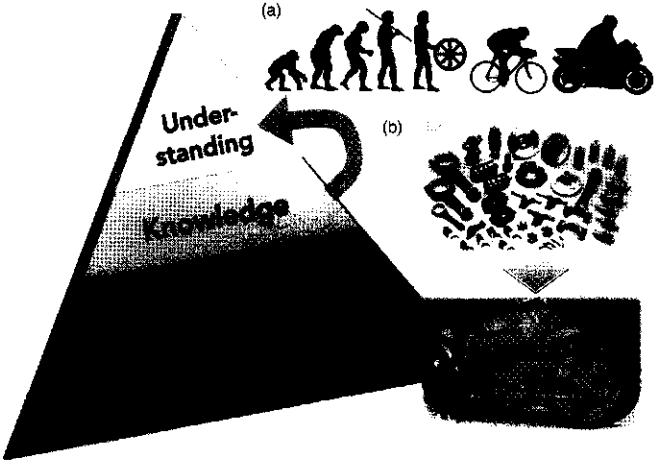


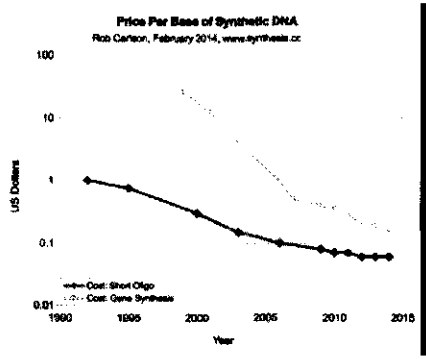
8. Van kennis naar begrip: *waarom*?

De laatste stap naar de top van de piramide is die van kennis naar begrip: als we ideeën hebben opgedaan over hoe bepaalde processen werken en hoe ze met elkaar interacteren, kunnen we dan ook begrijpen *waarom*? Kunnen we verklaringen geven die verder gaan dan de voorspellingen die uit onze modellen komen? In de biologie wordt begrip vaak gezocht in de evolutie. Daarmee kunnen we de losse kennis die we verzamelen van een "verhaal" voorzien, in termen van mutaties, selectiedruk en aanpassing. De meest kernachtige samenvatting van dat idee is de beroemde uitspraak van Dobzhansky, "Nothing in biology makes sense, except in the light of evolution" – ofwel, in de biologie is niets zinvol behalve in het licht van de evolutie[9].

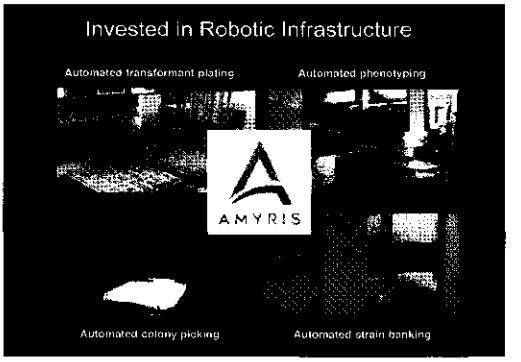
Maar voor het begrijpen van een biologisch systeem geeft evolutie ons een smalle basis om op te generaliseren. We moeten het doen met de variatie die we nu observeren, in feite een momentopname van een stochastisch proces (met randvoorwaarden) van miljarden jaren, dat maar een klein deel van alle mogelijkheden heeft verkend. Vertaald in termen van motoronderhoud, zegt het dat we motorfietsen kunnen begrijpen door te leren dat ze ontstaan zijn toen fietsen werden voorzien van verbrandingsmotoren; dat fietsen ontstaan zijn doordat de industriële revolutie de technologie bracht om een voertuig licht en stevig genoeg te maken om door de passagier te worden aangedreven; etc., wellicht tot de uitvinding van het wiel (zie figuur (a)).

Ik denk dat huidige technologische ontwikkelingen in de synthetische biologie ons gereedschappen geven om op een nieuwe manier tot begrip te komen. Om een andere beroemde wetenschapper aan te halen: van de natuurkundige Richard Feynman is de stelling "What I cannot create, I do not understand". Kortom, we





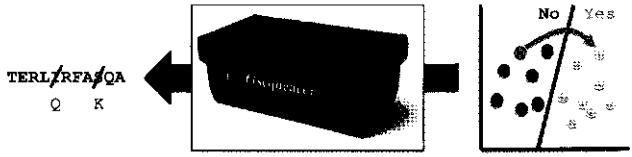
(a) www.synthesis.cc

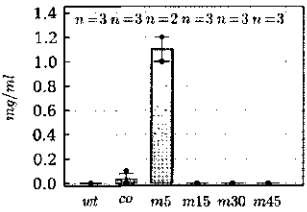


(b)

kunnen pas zeggen dat we een motorfiets begrijpen als we er een zelf kunnen bouwen (figuur (b)). Als ingenieur spreekt me dit erg aan, en ik heb hier grote verwachtingen van. Het synthetiseren, ofwel schrijven, van DNA is nog relatief duur ten opzichte van het lezen van DNA, op dit moment in de orde van 10 cent per base, en er kunnen maar relatief korte stukken geschreven worden. Maar er is enorme vooruitgang in methoden om korte stukken DNA te combineren tot langere stukken, en in automatisering van het laboratoriumwerk dat nodig is om DNA in cellen tot expressie te brengen. Een paar jaar geleden heeft Craig Venter deze technologie gebruikt om een bacterie te maken met een volledig synthetisch genoom (waar de quote van Feynman, overigens incorrect verwoord, in verstopt was)[20]. Maar naast dit soort huzarenstukjes wordt synthetische biologie ook gebruikt om nieuwe genetische richtingen te verkennen. Een biotechnologiebedrijf dat hier het voortouw in heeft genomen, Amyris uit de Verenigde Staten, kan op een bijna volledig geautomatiseerde manier op dit moment meer dan 100,000 verschillende gistgenomen *per maand* bouwen en testen op functionaliteit. Die schaal geeft ze de mogelijkheid om een enorm aantal varianten van bepaalde processen te proberen en de meest veelbelovende DNA sequenties te selecteren voor verder onderzoek.

In ons eigen werk hebben we al een paar voorzichtige stappen in deze richting gezet. Ik had het eerder over een model dat voorspelt of eiwitten wel of niet goed geproduceerd kunnen worden door een schimmel. Dat model is gebruikt als de basis voor een algoritme om eiwitten te herontwerpen voor betere productie, door een klein aantal veranderingen aan te brengen[21]. We hebben een aantal herontwerpen experimenteel getest, door het benodigde DNA te synthetiseren en in het genoom van de schimmel in te bouwen. We vonden dat teveel mutaties de productie van het eiwit





volledig verstoren, maar het bleek inderdaad mogelijk om met een klein aantal mutaties – in dit voorbeeld 5 – de productie wel 10x te verhogen. Een soortgelijke aanpak gebruiken we nu om de regulatie van de groei van wortels te begrijpen: door het DNA waarop de regulatoren binden te herontwerpen, leren we wat de belangrijkste bepalende elementen in dat DNA zijn.

Maar in feite gaat het hier nog steeds om experimentele validatie van voorspellingen. Recent onderzoek elders laat zien hoe we met de computer stukken DNA kunnen ontwerpen en in het lab slim alle combinaties kunnen testen[22]. Daarmee kunnen datasets worden aangelegd van tien- tot honderdduizenden sequenties van genen of regulatoire gebieden met het resulterende fenotype, bv. eiwitexpressie. Vervolgens kunnen we algoritmes en modellen maken om uit die sequenties regels af te leiden over de interacties tussen moleculen, om uiteindelijk tot begrip te komen van de werking van de belangrijkste processen in de cel. Op voorlopig bescheiden schaal gaan we hier in de Bioinformatics Group ook mee verder, specifiek in een aantal projecten op het gebied van eiwitinteracties en enzymatische reacties, in samenwerking met Plant Research International. Het doel is om een relatief groot aantal sequenties te genereren en te testen, en daaruit af te leiden wat de belangrijkste kenmerken van een eiwit zijn die de interacties of reacties beïnvloeden. Maar ik denk dat we hier nog maar aan het begin staan van een nieuwe doorbraak, zoals de sequencing ons die tien jaar geleden heeft gegeven.

9. Onderwijs

De ontwikkeling van het biologie onderzoek in de richting van data science die ik eerder schetste heeft ook gevolgen voor het onderwijs. Het feit dat er de afgelopen jaren diverse MSc programma's in de bioinformatica zijn opgezet laat zien dat hier al het nodige gebeurt. Een belangrijke vraag is hoe we in het onderwijs de interactie die nodig is tussen biologen, wiskundigen, fysici en informatici vorm kunnen geven, want zoals iedereen weet gaat dat niet vanzelf. Moeten we *multidisciplinaire* onderzoekers opleiden, die gespecialiseerd zijn in één van die takken van wetenschap maar goed kunnen samenwerken; *interdisciplinaire* onderzoekers, die wat minder diepe kennis van een breder gebied hebben; of *transdisciplinaire* onderzoekers, die een overzicht hebben van het hele gebied? Het lijkt me te ambitieus om studenten in een paar jaar op te leiden tot echt transdisciplinaire onderzoekers. Een multidisciplinaire aanpak daarentegen heeft het gevaar dat individuele studenten weliswaar goed problemen kunnen aanpakken op hun specifieke trede van de piramide, maar dat ze nog het overzicht missen om te bepalen welke problemen het beste aangepakt kunnen worden. Ik denk dat op dit moment de bachelors in de biologie wat meer multidisciplinair zouden moeten worden, door ze te laten zien wat de mogelijkheden en beperkingen van bioinformatica zijn – dat wil zeggen door ze de beginselen van programmeren en data analyse te leren en hoe om te gaan met de bestaande technologie. Vervolgens kunnen we in de MSc fase interdisciplinaire onderzoekers opleiden, die in staat zijn om bij te dragen aan vernieuwing van die technologie. Hiervoor is het belangrijk dat de MSc Bioinformatica studenten met verschillende achtergronden toelaat.

De uitdaging is hoe het onderwijs zo te organiseren dat we studenten afleveren die klaar zijn om te werken in een snel veranderende omgeving. "Zen en de kunst van het motoronderhoud" heeft het over twee belangrijke elementen in onderzoek. De eerste is "gumption", ofwel een optimistische blik op een probleem, de overtuiging dat het op te lossen is, die de nodige motivatie geeft om je maanden tot jaren vast te

bijten in een bepaald probleem. Het tweede is "stuckness", ofwel "vastzitten". Ik denk dat we ons allemaal situaties kunnen herinneren waarin we volledig vastzaten bij het oplossen van een probleem. Maar het moment dat je zelf doorkrijgt hoe je zo'n probleem op moet lossen levert een enorme leerervaring op. Vooral in de masterfase is het belangrijk om studenten deze twee concepten te laten ervaren: door ze te laten zien hoe mooi onderzoek kan zijn, in onderwijs en in afstudeerprojecten, maar ook door ze zich af en toe te laten vastbijten in een probleem zonder direct uit te leggen hoe dat opgelost moet worden. Daarom zetten we onze vakken vaak op in de vorm van projecten of opdrachten, waarin studenten zelfstandig zo'n probleem te lijf gaan.

In "Zen en de kunst van het motoronderhoud" staat overigens ook een interessante observatie van de hoofdpersoon over wat een universiteit is. De oorsprong van het woord is het Latijnse *universitas magistrorum et scholarium*, dat wil zeggen een samenkomen van docenten en studenten, waarbinnen een manier van denken wordt gevormd en overgedragen via onderwijs, discussies en debatten – zoiets als verbeeld in dit beroemde schilderij (figuur (a)). De fysieke implementaties daarvan die we nu bijna duizend jaar kennen – de campussen, gebouwen en studieprogramma's – zijn een tijdelijke en, op zijn best, zwakke afspiegelingen van dat Platonische ideaal (figuur (b)). Nu maakte de hoofdpersoon die opmerking nadat hij was ontslagen door zijn benoemingsadviescommissie, dus misschien is het niet geheel zonder vooroordelen. Maar het is wel iets dat we niet uit het oog moeten verliezen in de organisatie van ons onderwijs, waarin de relatie tussen student en docent gaandeweg meer trekken krijgt van die van een klant tot een leverancier. Dat heeft gevolgen voor het onderwijs: de "stuckness" wordt zoveel mogelijk uit de programma's verwijderd om ze studeerbaarder te maken.

Gelukkig komt hiervoor de laatste jaren meer aandacht. De recente protesten van studenten tegen het "rendementsdenken" wijzen op problemen in het onderwijs. Een soortgelijke beweging vindt plaats in het onderzoek, waarvan in de laatste decennia

(a)

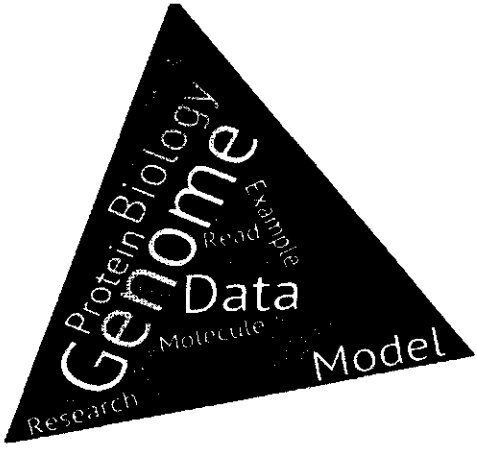
(b)

de kwaliteit ook steeds meer wordt gevat in kwantitatieve metingen aan productie, valorisatie en impact. Hierover is de laatste jaren een broodnodig debat aangewakkerd door de "Science in Transition" beweging. Ik denk dat we als studenten, docenten en onderzoekers een gedeelde taak hebben om de universiteit zo te helpen organiseren dat het ideaalbeeld van de *universitas* niet uit het oog verloren raakt: geen "fabriek" voor studenten, publicaties of patenten, maar bovenal een plek om samen te ontdekken hoe we de wereld beter kunnen begrijpen.

10. Woorden van dank

Daarmee ben ik gekomen bij het einde van mijn rede. Ik hoop dat ik wat heb kunnen overdragen van het belang van de bioinformatica voor de moderne biologie, en de bijdrage die de Bioinformatics Group daaraan wil leveren. De groep is in 2002 opgericht aan de Universiteit Wageningen, met de aanstelling van professor Jack Leunissen. Helaas is Jack veel te vroeg gestorven, aan een ernstige ziekte. Hoewel ik hem persoonlijk niet goed heb gekend, merk ik aan de manier waarop de leden van de groep en de collega's over hem praten dat hij een goed onderzoeker was, met een brede kennis van de bioinformatica, maar vooral een man met een warm hart. Ik hoop met de groep het pad dat hij ingeslagen is te vervolgen en, waar dat kan, nieuwe wegen in te slaan. Ik ben de Raad van Bestuur van Wageningen UR en met name de rector, Martin Kropff, dankbaar voor de kans die ze me daartoe hebben geboden. Tussen het wegvallen van Jack en mijn aantreden is de groep draaiend gehouden door Harm Nijveen en Sandra Smit, met hulp en steun van Gabino Sanchez, Ton Bisseling en Ernst van den Ende. Daarvoor, en voor de warme ontvangst in Wageningen, ben ik ze erg dankbaar.

Het mooiste van het werken bij een universiteit is dat je er voor wordt betaald om elke dag te blijven leren. Ik sta hier vandaag dan ook als hoogleraar omdat ik van veel mensen iets heb mogen leren. Daarbij is het ondoenlijk om uitputtend te zijn. Ik heb veel opgestoken van alle studenten die ik de afgelopen jaren heb begeleid, de onderzoekers waar ik mee heb samengewerkt en de vrienden waarmee ik oorlogsverhalen heb uitgewisseld. Toch wil ik er een paar mensen uitlichten. Ten eerste zijn dat de promovendi en postdocs die ik begeleid en heb begeleid, en over wier



werk ik vandaag wat heb verteld: Rogier, Yunlei, Bastiaan, Jurgen, Wynand, Alexey, Domenico, Marco en Emrah: bedankt! Ook van de nieuwe lading hier in Wageningen – Linke, Saulo, Sevgin, Siavash, Ehsan, Sven, Sander en Luca – heb ik hoge verwachtingen, ik zie er naar uit om de komende tijd met jullie samen te werken.

Daarnaast wil ik een paar mensen in het bijzonder bedanken. Frank Staal en Jack Pronk, bij wie ik heb gewerkt in Rotterdam en Delft, hebben me laten zien hoe je een groep zo kunt organiseren dat het loopt als een geoliede machine waarin iedereen zijn bijdrage kan leveren aan het grote geheel. Bob Duin, mijn mentor tijdens en na mijn promotie, heeft me laten zien hoe een onderzoeker in het leven zou moeten staan: enthousiast maar kritisch, gedreven door een kernvraag maar open voor nieuwe ideeën en denkrichtingen, en altijd op zoek naar samenwerking. Van Marcel Reinders, bij wie ik de laatste tien jaar heb gewerkt, heb ik geleerd hoe je veel kunt bereiken door jezelf continu uit te dagen en hoe je het beste uit studenten en medewerkers kunt halen door tegelijkertijd wetenschappelijk streng en persoonlijk zacht te zijn. En tenslotte mijn moeder, Sonja de Ridder, die uit ervaring weet hoe belangrijk een goede opleiding is en mij altijd heeft aangemoedigd en gesteund tijdens mijn studie, promotie en latere carrière.

Tenslotte wil ik hen bedanken die het belangrijkst voor mij zijn, maar dat buiten kantooruren helaas niet altijd even goed merken. Barbara, bedankt voor alle steun en hulp de afgelopen vijfentwintig jaar, waarbij ik soms op erg ongelegen momenten aan het werk was of in het buitenland zat. Dolf, Felix en Boris - ik hoop dat ik jullie oud en wijs mag zien worden. En, al wil ik de druk niet teveel opvoeren, ik ben benieuwd of mijn uittreede eerder komt dan jullie intreeredes.

Dames en heren, bedankt voor uw aandacht.

Ik heb gezegd.

[1] R.M. Pirsig, "Zen and the art of motorcycle maintenance". Vintage, 1999.
[2] J.-L. Jannink, A.J. Lorenz, H. Iwata, "Genomic selection in plant breeding: from theory to practice". *Briefings in Functional Genomics* 9(2):166-177, 2010.
[3] J.D. Keasling, "Manufacturing molecules through metabolic engineering". *Science* 330(6009):1355-1358, 2010.
[4] R. Mendes, P. Garbeva , J.M. Raaijmakers, "The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms". *FEMS Microbiology Reviews* 37(5):634–663, 2013.
[5] J.D. Watson, F.H.C. Crick, "A structure for Deoxyribose Nucleic Acid". *Nature* 171:737-739, 1953.
[6] International Human Genome Sequencing Consortium *et al.*, "Initial sequencing and analysis of the human genome". *Nature* 409:860-921, 2001.
[7] Potato Genome Sequencing Consortium *et al.*, "Genome sequence and analysis of the tuber crop potato". *Nature* 475(7355):189-195, 2011.
[8] The Tomato Genome Consortium, "The tomato genome sequence provides insights into fleshy fruit evolution". *Nature* 485:635–641, 2012.
[9] M.A.M. Groenen, A.L. Archibald, H. Uenishi *et al.*, "Analyses of pig genomes provide insight into porcine demography and evolution". *Nature* 491:393–398, 2012.
[10] L.W. Hillier, W. Miller, E. Birney *et al.*, "Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution". *Nature* 432:695-716, 2004.
[11] B.J. Haas, S. Kamoun, M.C. Zody *et al.*, "Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*". *Nature* 461:393-398, 2009.
[12] The 100 Tomato Genome Sequencing Consortium *et al.*, "Exploring genetic variation in the tomato (*Solanum* section *Lycopersicon*) clade by whole-genome sequencing". The Plant Journal 80(1):136-148, 2014.
[13] http://omicsmaps.com/
[14] T. Hey, S. Tansley, K. Tolle (eds), "The fourth paradigm: data-intensive scientific discovery". Microsoft Research, 2009.
[15] R. Ackoff, "From data to wisdom". *Journal of Applied Systems Analysis* 16:3–9, 2009.
[16] J.F. Nijkamp, M.A. van den Broek, J.M.A. Geertman *et al.*, "*De novo* detection of copy number variation by co-assembly". *Bioinformatics* 28(24):3195-3202, 2012.
[17] J.F. Nijkamp, M. Pop, M.J.T. Reinders, D. de Ridder, "Exploring variation-aware contig graphs for (comparative) metagenomics using MaryGold". *Bioinformatics* 29(22):2826-2834, 2013.
[18] B.A. van den Berg, M.J.T. Reinders, M. Hulsman *et al.*, "Exploring sequence characteristics related to high-level production of secreted proteins in *Aspergillus niger*". *PLoS ONE* 7(10):e45869, 2012.
[19] T. Dobzhansky, "Nothing in biology makes sense except in the light of evolution". *The American Biology Teacher* 35(3):125-129, 1973.
[20] D.G. Gibson, J.I. Glass, C. Lartigue *et al.*, "Creation of a bacterial cell controlled by a chemically synthesized genome." *Science* 329(5987):52-56, 2010.
[21] B.A. van den Berg, M.J.T. Reinders, J.-M. van der Laan *et al.*, "Protein redesign by learning from data". *Protein Engineering Design and Selection* 27(9):281-288, 2014.
[22] E. Sharon, Y. Kalma, A. Sharp et al., "Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters". *Nature Biotechnology* 30(6):521-530, 2012.

Prof. dr. ir. Dick de Ridder

'Ontwikkelingen in de levenswetenschappen worden sterk gedreven door de introductie van nieuwe technologie, die het mogelijk maakt concentraties, eigenschappen en interacties van vele duizenden moleculen in de cel te meten. De uitdaging is uit de ruwe meetgegevens informatie, kennis en uiteindelijk begrip van het biologisch systeem te destilleren. De bioinformatica houdt zich bezig met de ontwikkeling van computeralgoritmes en -modellen om dit mogelijk te maken en levert daarmee onmisbare gereedschappen voor de moderne levenswetenschappen.'

ISBN 978-94-6257-367-3