

# BARCODE VAN DNA

## DEMOCRATISERING VAN DE TAXONOMIE DOOR DIGITAAL IDENTIFICATIESYSTEEM

FREK T. BAKKER

Het herkennen van biologische soorten aan de hand van een gestandaardiseerde DNA-barcode heeft de laatste tijd een enorme vlucht genomen. Gedreven door aan de ene kant de biodiversiteitscrises en de mogelijke *global change*, en aan de andere kant zowel razendsnelle technologische vooruitgang als ook het vooruitzicht dat niet genoeg klassieke taxonomen worden opgeleid voor de nabije toekomst, lijkt DNA-barcoding zich een strategische plek te veroveren op huidige, al dan niet toegepaste, biodiversiteitsonderzoeksaandachten. Maar er is meer en er blijven bovendien nog genoeg vragen over.

⌘ Hoe zou het zijn om in een bioliterate wereld te leven – een wereld waar je binnen een paar minuten de naam van elke plant, dier of schimmel kan weten, *any time, anywhere*? En niet alleen de naam maar gelijk ook alle andere relevante informatie: wat zijn z'n eigenschappen, is de soort bedreigd, is het gevaarlijk, hoort het daar eigenlijk wel thuis of is het invasief van elders? Aldus iBOL.org, de site voor het International Barcode of Life project dat in september 2010 officieel van start ging met het vieren van de oplevering van de éénmiljoenste DNA-barcode (zie foto). iBOL is het grootste biodiversiteit-genomicsproject dat ooit is ondernomen en heeft de ambitie om een digitaal identificatiesysteem voor leven tot stand te brengen, met in eerste instantie vijf miljoen barcodes in vijf jaar. Omdat per soort meerdere DNA-barcodes worden gegenereerd om mogelijke geografische variatie te kunnen verdisconteren, representeren die vijf miljoen barcodes dus minimaal een half miljoen soorten. Ter vergelijking: schattingen van het aantal eukaryote soorten op aarde lopen uiteen tot meer dan twintig miljoen, en tot op heden zijn 1,75 miljoen soorten officieel beschreven.

Zoals voor wel meer evolutiebiologische doorbraken (Darwin, natuurlijke selectie, seksuele selectie; Ernst Mayr, biologisch soortconcept), begon het ook met DNA-barcoding bij vogels: piloten en verkeersleiders vroegen zich af of effectieve identificatie van (resten) van vogels in en op hun straalmotoren en cockpits mogelijk was. Via een aantal wetenschappelijke workshops en meetings kwam het in 2003 tot de inmiddels beroemde publicatie door Paul Hebert's onderzoeksgroep aan de Universiteit van Guelph, Canada, getiteld *Biological identifications through DNA barcodes*. In deze publicatie werd de wetenschappelijke gemeenschap voorgelegd hoe volgens hen een nieuw systeem van soortidentificatie (maar ook -ontdekking) zou kunnen werken, gebruikmakend van een korte en gestandaardiseerde DNA-sequentie binnen het genoom, een regio dus die in ieder eukaryoot organisme te vinden zal zijn. De keuze voor dit 'barcodegen' viel op een regio van 648 baseparen in het cytochrome c oxidase 1 gene (CO1), dat gecodeerd is op het mitochondriaal genoom alwaar het een cruciale schakel in de oxidatieve fosforylatie ('ademhalingsketen') verzorgt. Hiermee werden gelijk twee vliegen in één klap gevangen omdat ten eerste het hier een essentiële genfunctie betreft die dus in alle, ook nog niet beschreven, leven kan worden verwacht, en ten tweede mitochondriaal DNA in meerdere kopieën per cel voorkomt en daarmee relatief eenvoudig uit zelfs historisch materiaal geïsoleerd kan worden. Een universele barcode dus, en CO1 is inderdaad uitermate geschikt gebleken voor het kunnen onderscheiden van vogels, insecten, vissen en zoogdieren.

Voor het identificeren van (land)planten en schimmels echter is CO1 niet geschikt: in planten vanwege te weinig CO1-sequentievariatie, in schimmels onder meer vanwege de aanwezigheid van zogenaamde introns in het CO1-gen. Op basis van resultaten van een aantal grote, datagedreven projecten is afgelopen jaar als plant DNA-barcoderegio gekozen voor een tweetal chloroplast-DNA-gecodeerde regio's (*rbcL* en *matK*). Voor de schimmels lijkt ribosomaal DNA geschikt, zo ook voor een andere bijzonder soortenrijke clade: de Nematoda.

### HET DNA-BARCODINGPROCES

De DNA-barcode *production pipeline* bestaat uit ten eerste het verzamelen van taxonomische specimens die afkomstig kunnen zijn van expedities, botanische tuinen of bodemonsters, maar ook uit museum- of herbariumcollecties, omdat DNA de tand des tijds, vooral in planten, meestal goed doorstaat; met name herbariumcollecties representeren daardoor een belangrijke, moleculaire, meerwaarde; ten tweede een taxonoom die een specimen op naam brengt, meestal in nauwe samenwerking met de taxonomische wereldexpertise voor de betreffende groep; ten derde wordt vervolgens een klein gedeelte van het specimen opgestuurd naar een centrale DNA-barcode *high-throughput* faciliteit, waar het genereren van de sequenties voor een groot deel met behulp van PCR-amplificatie door robots wordt uitgevoerd (zulke faciliteiten zijn er momenteel in Canada, de Verenigde Staten, China, Duitsland, Frankrijk en binnenkort ook in Nederland); ten vierde wordt de barcode, compleet met afbeelding en collectiereferentie van het specimen, GPS-coördinaten van de vindplaats en rauwe DNA-data, toegevoegd aan een globale referentie bibliotheek ([www.boldsystems.org](http://www.boldsystems.org)). BOLD fungeert tevens als een virtuele werkbank in die zin dat per project, bijvoorbeeld 'alle mieren van de wereld', 'Orchideeën van Costa Rica' enzovoorts potentieel geïnteresseerde taxonomen kenbaar kunnen maken bij te willen dragen, en vervolgens, in de context van elkaars data (de DNA-barcodes) tot mogelijk nieuwe inzichten of soortconcepten kunnen komen; ten vijfde fungeert met het groeiende aantal DNA-barcode records in BOLD deze database uiteindelijk als referentiebibliotheek waartegen toekomstige DNA-barcodes die zijn gegenereerd van ongeïdentificeerd materiaal (bijvoorbeeld van planten die nog niet in bloei zijn, fragmenten van historisch, forensisch of medisch materiaal) kunnen worden vergeleken. In het geval zo'n *query* barcode nog niet in de bibliotheek voorkomt, en dus geen *match* geeft, kan dat dus wijzen op een nog niet beschreven soort, maar mogelijk ook op tot dan toe onvermoede genetische variatie binnen de betreffende soort.



DE CN TOWER IN TORONTO WAAR DE OPLEVERING VAN DE ÉÉNMIJLJOENSTE DNA-BARCODE EN DE OFFICIËLE START VAN IBOL WERDEN GEVIERD. OP DE FOTO DE PROJECTIE VAN DE DNA-BARCODE VAN CASTOR CANADENSIS (DE BEVER) ZOALS DIE DIE AVOND TE ZIEN WAS.

Wat dat betreft is de term barcode in dit verband wat ongelukkig omdat vaak analogie met productbarcodes in bijvoorbeeld supermarkten wordt vermoed. Echter bestaan biologische soorten bij de gratie van verschillen in binnensoortelijke genetische variatie, hetgeen in DNA-barcodevariatie tot uiting komt, terwijl producten in winkels natuurlijk standaard worden afgeleverd met een uniforme, en van te voren toegewezen barcode. Aangaande de analytische aspecten van het matchen van DNA-barcodes is momenteel veel aandacht, onder meer om naast puur kwalitatieve matching ook een kwantitatieve benadering te ontwikkelen ('negentig procent kans dat het soort X betreft'), wat vanuit juridisch of forensisch oogpunt van belang kan worden.

## TOEPASSINGEN

**N**aast fundamenteel wetenschappelijke toepassingen van DNA-barcodes (ecologische survey's en het op naam brengen van niet-bloeiend materiaal, het aan elkaar koppelen van verschillende levensstadia binnen een soort, het barcoderen van taxonomisch typemateriaal, *community analyses* enzovoorts) zijn er meer toepassingen in met name het bijtijds detecteren van plaagorganismen die meekomen bij im- en export van gewassen in onze havens en luchthavens, het monitoren van zoetwaterkwaliteit, of het snel kunnen detecteren van humane pathogenen of medisch relevante schimmels. Maar ook voor het vaststellen van voedsel fraude, met andere woorden het aanbieden of verkopen van voedsel onder incorrecte naam (zie verder [www.barcodeoflife.org](http://www.barcodeoflife.org)). Zoetwaterkwaliteit wordt afgeleid uit de samenstelling en het relatieve voorkomen van de waterinsektenfauna, vaak slechts in larvaal stadium. Voor sommige van deze groepen waren in de Verenigde Staten onlangs nog slechts twee professionele taxonomen beschikbaar, zij het gepensioneerd. Een CO1-bibliotheek is dan ook een goede investering voor toekomstig waterkwaliteitsbeheer, omdat anders niemand deze diertjes nog op naam zou kunnen brengen. Een ander project is bijvoorbeeld het Mosquito Barcoding Initiative, waarin de meest relevante van alle tweeduizend muskietensoorten wereldwijd van een CO1-barcode werden voorzien. Hier is het in mindere mate de verdwijnende taxonomische expertise maar veeleer het snel en effectief kunnen herkennen van mogelijke humane pathogenen. Naast deze zogenaamde cladegebaseerde toepassingen bestaan ook regio-gebaseerde DNA-barcoding projecten, zoals bijvoorbeeld het Moorea Biocode Project ([mooreabiocode.org](http://mooreabiocode.org)) waarin van ál het non-microbiële leven aanwezig op het tropisch eiland Moorea een DNA-barcode wordt gegenereerd. Op dit moment staat de specimenteller op 31.821 maar bij een complete inventaris zou in principe elk van het eiland afkomstige specimen, wellicht automatisch, op naam kunnen worden gebracht.

## LINGUA FRANCA

**D**ankzij deze taxonomische lingua franca van DNA-barcodes is er sprake van een ongekende amplificatie van de taxonomie, een veld dat historisch vaak in een meer geïsoleerde lokale of regionale setting plaatsvond. Het is voor goede taxonomische soort-afbakening natuurlijk sowieso goed om een globaal perspectief te hanteren, omdat evolutie zich immers niet aan politieke of landsgrenzen houdt. Daarnaast is er dikwijls sprake van uiterst waardevolle reciproque illuminatie, dat wil zeggen dat op morfologie (maar ook op geografie) gebaseerde soortomschrijvingen fungeren als initiële hypothesen die vervolgens al dan niet door DNA-barcodes bevestigd worden. De taxonoom prikkelt dit om mogelijk over het hoofd geziene morfologische kenmerken nog eens in groter detail te bestuderen, wat weer kan leiden tot nieuwe soortrevisie. Bij planten zijn overigens de soortsgrenzen vaak minder duidelijk dan bij dieren, en kunnen nauw verwante soorten bovendien vaak

hybridiseren, hetgeen het toekennen van een effectieve DNA-barcode sterk compliceert.

Tot voor kort opereerden verschillende taxonomische communities (entomologen, ornithologen, botanici enzovoorts) vaak gescheiden, maar binnen DNA-barcodingverband ziet men steeds vaker cross-community meetings waar gemeenschappelijke uitdagingen, zij het wetenschappelijk, qua infrastructuur en financiering of in het trainen van een volgende generatie taxonomen, worden herkend en collectief aangegaan. Zo is bijvoorbeeld een typisch gedeeld probleem wíé alle, met behulp van DNA-barcodes, versneld ontdekte nieuwe soorten officieel gaat beschrijven, dat wil zeggen volgens de standaards die de internationale codes voor nomenclatuur vereisen? In sommige gevallen (insecten, crustaceaën) is kristalhelder dat niet genoeg taxonomen voor deze taak beschikbaar (zullen) zijn en wordt soms gesorteerd op nummers in plaats van namen als tijdelijke *identifiers*. Met behulp van de meest recente high-throughput DNA-sequentietechnologie is het mogelijk uit bodemmonsters onverwacht grote hoeveelheden tot nu toe onbekende, want niet kweekbare, biodiversiteit te ontdekken. Voor het officieel beschrijven dáárvan is niet genoeg taxonomische capaciteit beschikbaar, maar wordt ze soms al wel in bijvoorbeeld ecologische (bodem)modellen meegenomen.

## DE TOEKOMST

**A**l met al is met het ontwikkelen van DNA-barcoding een belangrijke stap gezet in het mobiliseren en mogelijk amplificeren van taxonomische wetenschap en haar meer toegepaste disciplines. In principe is het herkennen van soorten aan de hand van DNA-sequenties echter niet nieuw, moleculaire systematiek bestaat al ongeveer twintig jaar. Het is echter de globale schaal en standaardisatie in DNA-barcoding die volledig nieuw is. Om dit goed te laten verlopen heeft bijvoorbeeld het Consortium for the Barcode of Life (CBOL) goed werk verricht om tot die noodzakelijke datastandaards, als ook de noodzakelijke globale participatie door taxonomische communities te komen (dit laatste een waardevolle *community-exercise in itself*). Het democratiseren van de taxonomie door het beschikbaar komen van DNA-barcodebibliotheken kan het werk van taxonomen verlichten doordat het aantal datapunten enorm toeneemt, en meer aandacht aan het ontdekken van nieuwe soorten (in plaats van identificeren van bestaande) kan worden besteed. Voorts is het goed om te realiseren dat volgens de CBOL-standaards voor elke soort met een DNA-barcode ergens een gecureerd specimen in een museum of herbarium moet liggen. Financiering voor DNA-barcoding is dus ook voor biologische collecties en taxonomie. Tenslotte, sommigen verwachten een toename van netto biobeleezenheid wanneer DNA-barcodebibliotheken toenemend in gebruik zullen zijn. Misschien is dit te vergelijken met het gebruik van Google Earth: omdat de technologie beschikbaar is, wordt het ook gebruikt en resulteert het daardoor in toegenomen belangstelling voor (en dus kennis van) topografie en geografie. Zal de gebruiker van morgen, gewapend met een DNA-barcoder in z'n iPhone ook daadwerkelijk meer soortidentificaties uitvoeren, domweg omdat die functie nu eenmaal beschikbaar is? Het zou een geweldige kans zijn om onze biobeleezenheid te behouden.

*Dr. F.T. Bakker is universitair docent bij de leerstoelgroep Biosystematiek van Wageningen Universiteit. Hij was voorzitter van de science advisory board van Consortium for the Barcode of Life (CBOL) en is thans lid van het steering committee van International Barcode of Life (iBOL).*