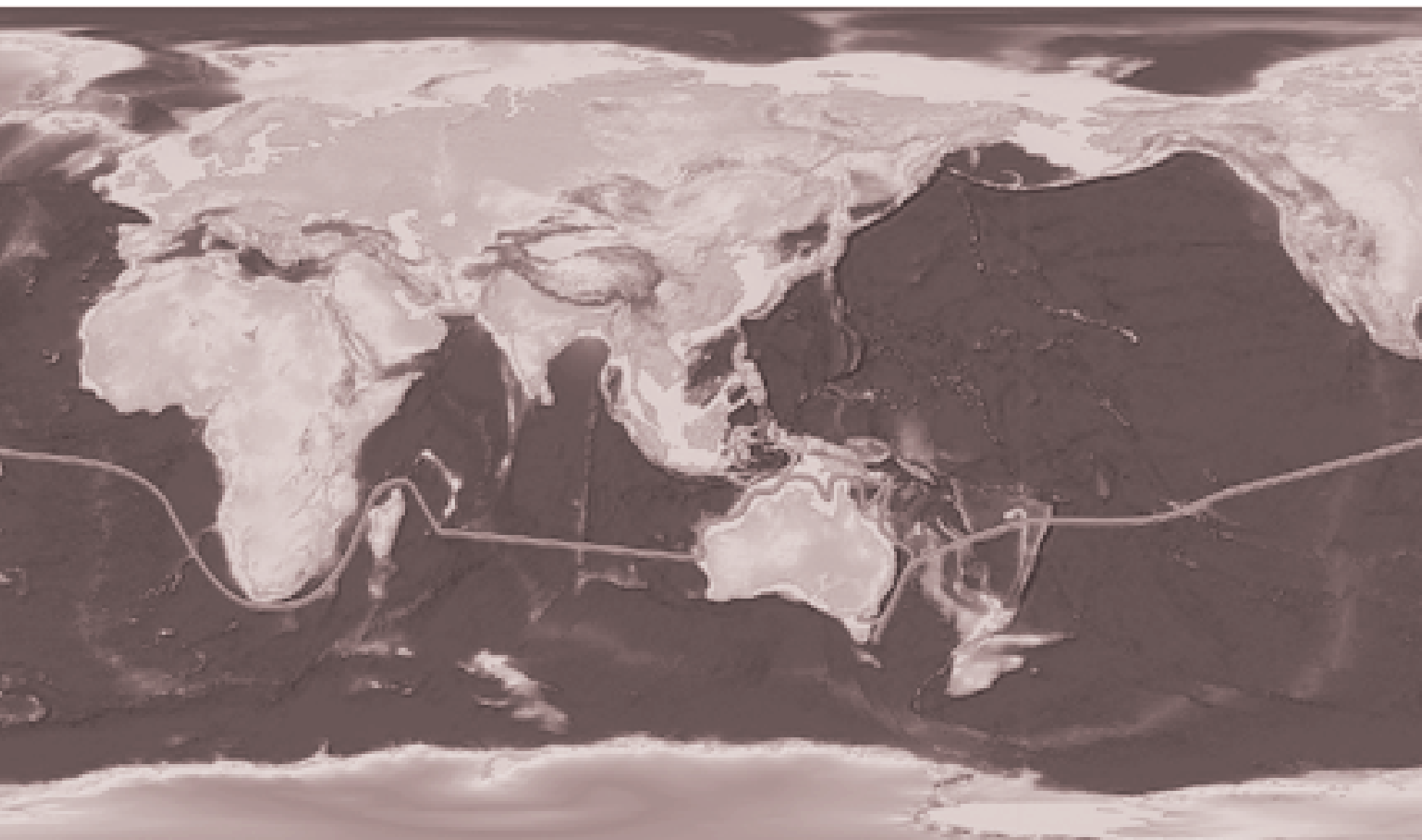


DE ONTRAFELING VAN EEN GENOOM IS NOG ZELDEN WETENSCHAPPELIJK NIEUWS. WAT WEL OPZIEBART, ZIJN DE RESULTATEN VAN EEN NIEUWE TAK BINNEN HET GENETISCH ONDERZOEK, DE META-OF ECOGENOMICS. HIER WORDT HET DNA VAN EEN HELE BIOTOOP ONTRAFELD OM INZICHT TE KRIJGEN IN HET FUNCTIONEREN VAN EEN VOLLEDIG ECOSYSTEEM. DE ONTDEKKING VAN NIEUWE BIOACTIEVE MOLECULEN BIEDT GROTE PERSPECTIEVEN VOOR TOEPASSINGEN IN DE INDUSTRIE EN DE GENEESKUNDE.

# Het genoom van een biotoop

**Bruno Cammue**



Als basiscomponenten van elke levende cel vormen eiwitten (of proteïnen) enerzijds de directe bouwstenen van verschillende cellulaire structuren, anderzijds worden zij ingezet als werktuigen in de aanmaak van de diverse belangrijke celcomponenten (vetten, suikers ...) of andere eiwitten. Zodoende liggen zij aan de basis van alle eigenschappen van een organisme. Zelf zijn ze opgebouwd als ketens van aminozuren. De informatie die nodig is om eiwitten aan te maken ligt algemeen vervat in het DNA dat eveneens een strengvormige structuur heeft die is opgebouwd uit slechts vier bouwstenen, de zogenaamde nucleotiden die worden getypeerd als A, C, G, en T.

Elk triplet van nucleotiden codeert voor één aminozuur, op basis van een genetische lettercode

die universeel geldig is voor (nagenoeg) alle levende wezens. Hierbij wordt één *gen* meestal gezien als de sequentie van nucleotiden die nodig is om alle aminozuren van één eiwit aan te maken. Zo bezit elke menselijke cel ongeveer 30 000 genen, of de informatie die nodig is voor de aanmaak van (minstens) evenveel verschillende eiwitten. Het geheel van deze genen wordt het (cel)genoom genoemd. Aangezien organismen kunnen bestaan uit één enkele cel (bijvoorbeeld bij bacteriën) of uit miljarden ondubbelde en dus genetisch identieke cellen (bijvoorbeeld bij de mens) komt het genoom van de cel overeen met dat van het organisme.

Tot het begin van de jaren 1990 was de aanpak van het biologisch-georiënteerd wetenschappelijk onderzoek gericht op het bestuderen van één of en-

kele genen tegelijk, of op hun overeenkomstige eiwitten en de specifieke eigenschap(en) die ze aan het organisme verlenen. Sindsdien heeft de enorme evolutie in de analysetechnieken en de bio-informatica het mogelijk gemaakt steeds meer genen en zelfs volledige genomen te betrekken in het onderzoek om zo te komen tot een meer geïntegreerde studie van het bestudeerde organisme. Dergelijke 'genoomwijde' benaderingen werden geduid onder de algemene term '*genomics*'. Een eerste echte mijlpaal voor dergelijke *genomics*-benaderingen kwam er met de publicatie van de eerste volledige genoomsequentie, namelijk deze van de bacterie *Haemophilus influenzae* in 1995.

De term *genomics* is echter vooral doorgebroken als gevolg van de enorme explosie in de ontwikke-

ling van nieuwe moleculaire technieken en aanverwante informatica-ondersteuning tijdens de periode van het spraakmakende Humane Genome Project. Dit project werd opgestart in 1990 op initiatief van het Amerikaanse National Institute of Health en het Department of Energy en betrof een wereldwijde samenwerking tussen de academische wereld en een aantal industriële en niet-commerciële organisaties. Het doel was duidelijk: het menselijke genoom ontrafelen binnen een periode van vijftien jaar. Dit onderzoek geraakte echter in een stroomversnelling door de start in 1998 van een concurrerend en eerder commercieel getint gelijkaardig project onder leiding van de microbioloog Craig Venter. Bijna gelijktijdig publiceerden beide consortia in de respectievelijke toptijdschriften *Nature* en *Science* de eerste ruwe beschrijving van de sequentie van het menselijk genoom, waarvan de grootte ongeveer 1700 maal deze van de eerder vermelde *Haemophilus*-bacterie bedroeg. Sindsdien is het genoom van tientallen bijkomende organismen in kaart gebracht, zoals diverse types van bacte-

riën, de bakkersgist, de fruitvlieg, de rijstplant, de kip enzovoort.

Het aantal gesequeneerde genomen blijft versneld groeien, in die mate zelfs dat een nieuw gepubliceerde genomesequentie niet meer *per se* tot de top van de wetenschappelijke publicaties behoort. Bovendien maken enerzijds de dalende kost per gesequeneerd nucleotide en anderzijds de drastisch toegenomen analysesnelheid het zelfs aanvaardbaar om de genomesequentie te gaan bepalen van nieuwe stammen of rassen van species waarvan reeds een genomversie is gekend. Tijd dus om de 'genomics-benadering' op een hoger niveau toe te passen? De stap die hiervoor wordt gezet, is vergelijkbaar met de eerder vermelde evolutie van de één-genbenadering naar de genomwijde aanpak in de beginjaren 1990. Meer specifiek gaat men zich nu, sinds ongeveer het begin van dit millennium, niet enkel meer richten op het genoom van één bepaald organisme maar op de totale genetische informatie die aanwezig is in één bepaalde leefomgeving, in één bepaalde biotoop. Hierbij wordt de

specificiteit van een dergelijke biotoop bepaald door de diversiteit van organismen die er in kunnen (over-)leven in interactie met elkaar en hun omgeving. De 'omics'-term die hiervoor meestal wordt gebruikt is *metagenomics*, omdat het onderzoek gericht is op het (meta)genoom van een biotoop, waarbij niet meer het organisme zelf maar de gehele biotoop als eenheid van organisatie wordt aanzien. Alternatieve benamingen of nauwverwante termen voor dit ontluikend onderzoeksdomein zijn onder andere *ecogenomics* of *environmental genomics*.

**In de metagenomics wordt  
niet meer het organisme zelf  
maar de gehele biotoop  
als eenheid van organisatie  
aanzien**

Momenteel maakt men bij het inventariseren van een dergelijk metagenoom meestal gebruik van verdere ontwikkelingen van de zogeheten *shotgun sequencing*-methode, die eerder door Craig Venters groep werd toegepast voor de ontrafeling van het menselijk genoom. Hierbij wordt totaal-DNA uit een bepaald biotoopstaal geïsoleerd, gefragmenteerd en via hogedoorvoersequencers geanalyseerd. De miljoenen willekeurig bekomen fragmentsequenties worden nadien via krachtige algoritmen in een groter geheel geschikt, als stukjes van een puzzel, op basis van overlappende delen 'genetische lettercode' aan de uiteinden van de fragmenten. Hierdoor kunnen meerdere verschillende genomen uit de biotoop geheel of gedeeltelijk worden gereconstrueerd. Twee pioniersstudies uit het voorbije jaar hebben het potentieel van deze *shotgun-sequencing*-methode sterk benadrukt en het *metagenomics*-landschap opvallend bijgekleurd.

Een eerste was gericht op een zeer specifieke biotoop, namelijk een oude ijzermijn in Californië, en in het bijzonder op organismen die voorkomen in het extreem zure drainagewater van deze mijn. Uit de in totaal 76 miljoen nucleotiden lange sequenties konden de genomen van twee dominante micro-organismen worden ontcijferd. Een dergelijke succesvolle reconstructie van twee nagenoeg volledige genomen uit één enkele studie werd in dit geval natuurlijk geholpen door de lage microbiële diversiteit ten gevolge van de extreme condities van deze biotoop.

Een tweede studie betreft monumentaal onderzoek van, alweer, Craig Venters groep, waardoor deze man stilaan blijkt uit te groeien tot de Bill Gates van het genomisch onderzoek. Craig onder-



zocht het genoom in tweehonderd liter waterstalen van de Sargasso Zee, één van de oceanografisch best bestudeerde mariene regio's in de buurt van de Bermuda-eilanden. *Shotgun sequencing* van deze stalen resulteerde in nagenoeg twee miljoen verschillende DNA-sequenties met een totale lengte van 1.6 miljard nucleotiden. Hieruit konden reeds 1.1 miljoen verschillende genen worden gereconstrueerd, ongeveer veertigmaal zoveel genen als de mens er bezit. De afgeleide eiwitsequenties uit deze ene studie vertienvoudigde het aantal eiwitgegevens dat aanwezig is in de publieke databank SWISSPRO.

Wereldwijd worden dergelijke databanken intensief gebruikt door wetenschappers voor het opsporen van sequentiehomologieën. Enerzijds omvat dit de invoer van steeds nieuwe sequentiegegevens uit hun lopend onderzoek, waardoor de databanken geleidelijk aangroeien. Anderzijds is er een continue consultatie van deze gegevens (via de zogenaamde BLAST-searches, vergelijkbaar met Google-zoektochten op het internet) waarbij onderzoekers hun laatste sequentieresultaten kunnen vergelijken met deze uit de databanken. Hierbij gaan ze er van uit dat homologe eiwitsequenties wijzen op gelijkaardige structuren en aldus gelijkaardige functies van deze eiwitten. Omwille van de plotse massale invoer van nieuwe genen en eiwitten uit Venters werk werden de databanken sterk verrijkt met eiwitten met ongekende functie. Daardoor werden de resultaten van BLAST-searches plots vaak gedomineerd door sequentiehomologieën met dergelijke eiwitten, wat het voorspellen van eiwitfuncties sterk bemoeilijkte.

Daardoor zag het National Center for Biotechnology Information (NCBI), dat het beheer van dergelijke databanken mee coördineert, zich genoodzaakt om een nieuwe BLAST-webpagina te creëren specifiek voor metagenoomsequenties. Het is dus duidelijk dat het verwerken van dergelijke resultaten waarschijnlijk meer tijd, inspanning en kosten zal vergen dan het genereren van de ruwe sequentiegegevens. De mega-*metagenomics*-studie van Venters groep is voorlopig de enige, maar waarschijnlijk niet de laatste van die omvang. Het blijft daarom aan te raden om het hoofd koel te houden binnen een stilaan (over)verhitte *genomics*-wereld en de basisdoelstellingen van dergelijke studies voor ogen te blijven houden.

Een eerste betrachting van het *metagenomics*-onderzoek blijft om vanuit fundamenteel wetenschappelijk standpunt een beter inzicht te krijgen in de samenstelling en het functioneren van diverse ecosystemen, gaande van simpele microbiële niches (zoals het mijnrainswater) tot complexere biotopen (zoals de Sargasso Zee). Micro-organismen worden verantwoordelijk geacht voor de meeste biogeochemische processen die vorm hebben gegeven aan de diverse biotopen op aarde.



Conventieel microbiëel onderzoek is gebaseerd op het analyseren van bepaalde microbiële species na opkweek op specifieke artificiële voedingsbodems. Volgens de huidige schattingen zou echter meer dan 99 procent van alle microbiëel leven uit onze omgeving niet op te kweken zijn via deze standaardmethoden. Daardoor zijn ze dus tot nog toe onzichtbaar gebleven voor wetenschappelijk onderzoek.

Doordat de klassieke opkweek van micro-organismen niet vereist is bij een *metagenomics*-aanpak, zou een dergelijke benadering in de eerste plaats kunnen bijdragen tot een meer representatieve inventarisatie van biotopen. Dergelijke taxonomische gegevens zijn nodig als basis voor verdere ecologische en evolutieve studies die respectievelijk gericht zijn op de interacties van de diverse organismen onderling en met hun biotoop, en op de druk van die omgeving op het ontstaan van nieuwe species. Ondanks de veel grotere output van sequentiegegevens in het Sargasso-experiment in vergelijking met de ijzermijnstudie konden hieruit ook slechts twee (quasi-)volledige genomen worden samengesteld. Dit heeft voornamelijk te maken met de veel hogere biodiversiteit van een dergelijke zeebiotoop. De onderzoekers konden inderdaad aantonen dat de reeds onderzochte zeestalen sequenties bevatten van tenminste 1 800 verschillende species, waaronder 148 nog ongekende soorten. Interessant hierbij was de vaststelling dat sommige van de opgepikte genen duidelijk homologie vertoonden met reeds gekende sequenties uit zoogdieren. Los van enige wetenschappelijke evidentie zouden speculanten onder ons kunnen hopen op het aldus ontdekken van de eerste sporen van verdwenen zeevaarders in de beruchte Bermudadriehoek.

Een tweede belangrijke betrachting van een *metagenomics*-benadering is eerder gericht op toegepast onderzoek. In de farmaceutische sector bijvoorbeeld zijn onderzoekers blijvend op zoek naar nieuwe microbiële bronnen van antibiotica sinds de ontdekking van het eerste antibioticum penicilline (uit de schimmel *Penicillium notatum*) in 1929 door Alexander Fleming. Bodembacteriën van de orde van de actinomycetales bleken hierbij een succesvolle bron te zijn. Veertig jaar na de eerste ontdekking van actinomycine in 1940 bleken actinomyce-

tales in 1980 al verantwoordelijk voor de productie van zeventig procent van alle natuurlijke antibiotica die worden gebruikt in de geneeskunde. Gedurende de laatste decennia is echter het succes in het ontdekken van structureel nieuwe types antimicrobiële componenten significant gedaald. Isolatie van reeds eerder gekende klassen van antibiotica (de zogenaamde 'herontdekking') via conventionele microbiële screening benadert stilaan een frequentie van 99.9 procent.

Daartegenover staat de reeds eerder vermelde lage portie (één procent) van de totale biodiversiteit die via deze standaard microbiële technieken kan worden opgespoord. Alternatieve benaderingen dringen zich duidelijk op. Het meest voor de hand liggend is de ontwikkeling van nieuwe microbiële opkweekmethododes. Hierbij wordt momenteel onderzoek verricht naar onder andere het toepassen, tijdens de kweek, van ingewikkelde simulaties van natuurlijke omgevingsfactoren of het inkapselen van cellen in gel-microdruppels. Hoe dan ook blijft het ontwikkelen van nieuwe kweekmethododes sterk afhankelijk van het te isoleren micro-organisme, en zijn deze technieken dus moeilijker algemeen implementeerbaar in grootschalige screeningsprogramma's. Het alternatief is het toepassen van kweekonafhankelijke, moleculaire benaderingen, zoals onder meer toegepast in de metagenoomstudies van Craig Venter.

Dergelijke methoden omvatten vaak het aanleggen van een zogenaamde *DNA-bibliotheek*. Hierbij wordt DNA uit het staal (metagenoom) geïsoleerd, gefragmenteerd (bijvoorbeeld in stukken van 50–150 000 nucleotiden lang), geïnserieerd in een DNA-vector (zoals de zogenaamde cosmiden of bacterieel artificiële chromosomen [BACs]) en overgebracht in een gastcel zoals bijvoorbeeld de bacterie *Escherichia coli*. Het geheel van al deze fragmenten, apart verpakt in de verschillende gastcellen van de DNA-bibliotheek, zou dus het oorspronkelijk geïsoleerde totaal-DNA moeten omvatten. Om representatief te zijn voor het metagenoom van bijvoorbeeld een bodemstaal zou dergelijke bibliotheek theoretisch minstens tien miljoen BACs moeten bevatten, elk met een *insert* van gemiddeld 100 000 nucleotiden lang, tenminste als er wordt van uitgegaan dat alle organismen uit de onderzochte biotoop op gelijke wijze vertegenwoordigd zouden



zijn in het geïsoleerde DNA-staal en de overeenkomstige DNA-bibliotheek. In de praktijk is dit zeker niet het geval en zal de aangelegde bibliotheek vele malen groter moeten zijn om het volledige metagenoom te vertegenwoordigen. Eens dergelijke complexe bibliotheken zijn aangelegd, kunnen zij worden aangewend voor het opsporen van nieuwe bioactieve moleculen (zoals geneesmiddelen of industriële biokatalysatoren). Verschillende screeningsstrategieën worden hierbij aangewend die grotendeels op te delen zijn in activiteitsgebaseerde en sequentiegebaseerde benaderingen.

### **Het grote voordeel van deze aanpak is dat men totaal nieuwe genen en dus nieuwe types van biomoleculen kan opsporen**

In *activiteitsgebaseerde* strategieën wordt de expressie van elk van de verschillende DNA-fragmenten in de gastcel geëvalueerd, met andere woorden wordt de specifieke activiteit van de overeenkomstige nieuwe genproducten (eiwitten of via eiwitten geproduceerde metabolieten) nagegaan. Hierbij kan het bijvoorbeeld gaan om antimicrobiële werking voor het ontwikkelen van nieuwe antibiotica of om specifieke enzymatische activiteiten die toepasbaar zijn in industriële processen. Het grote voordeel van deze aanpak is dat men totaal nieuwe genen en dus nieuwe types van biomoleculen kan opsporen, en bovendien dat men enkel selecteert op volledige genen en functionele genproducten. Het grote nadeel blijft de vereiste expressie van de DNA-constructen in de gastheer. Naast informatie die vervat is in het DNA, de zogenaamde universele genetische code, zijn processen die optreden na afschrijving van een gen immers even belangrijk voor de correcte vorming van functionele eindproducten.

Dergelijke posttranscriptionele processen kunnen variëren tussen verschillende organismen. Voor het screenen van een metagenomische DNA-bibliotheek maakt men best vooraf een rationele keuze van de gastheer. Een eerste optie is de darm-

bacterie *Escherichia coli*; als alomgebruikt werkpaard in de biotechnologie is hiervoor immers al heel wat know-how voorhanden. Het bijkomende voordeel is dat *Escherichia coli* ook algemeen wordt toegepast in industriële fermentatieprocessen. Geselecteerde genen uit de screening van de DNA-bibliotheek kunnen daarom veel sneller worden geïmplementeerd in dergelijke toepassingen. Anderzijds zou het aanwenden van een bodemorganisme, zoals de bacterie *Streptomyces lividans*, aangewezen kunnen zijn voor expressie van bodemmetagenomen. Ook hybride systemen zijn voorhanden waarbij de DNA-bibliotheek bijvoorbeeld wordt aangemaakt in *Escherichia coli* en nadien wordt overgebracht in *Streptomyces lividans* voor verdere expressietesten.

Een sequentiegebaseerde screening daarentegen omvat geen expressiestap en is dus onafhankelijk van een gastheer. Bij deze methode wordt eerst van alle DNA-constructen uit de metagenomische bibliotheek de sequentie bepaald (zoals in het Sargasso-experiment) en wordt hierin vervolgens gezocht naar geconserveerde regio's die belangrijk zijn voor het type biomoleculen waarop men zich richt. Dergelijke evolutief geconserveerde gelijkenis tussen sequenties kan immers wijzen op een evolutief behoud van de functie (activiteit) van de overeenkomstige eiwitten, zoals eerder vermeld. Op basis van deze methode kunnen dus nieuwe varianten van gekende functionele proteïnen worden opgespoord, maar voorkennis over de gezochte proteïnen is steeds vereist. Verwacht wordt echter dat door de enorme evolutie in de bio-informatica de kennis omtrent het verband tussen bepaalde genetische motieven en de eigenschappen van de overeenkomstige eiwitten drastisch zal toenemen, waardoor ook het succes van sequentiegebaseerde screeningsmethoden gaandeweg zal stijgen.

De lijst van bioactieve moleculen die werden ontdekt via screening van metagenomische DNA-banken (metagenomische bioproductie) is momenteel nog beperkt, maar groeit snel. Hiertoe behoren biokatalysatoren zoals amylases, lipases, proteases en agarases, die toepasbaar zijn in diverse industriële processen, alsook nieuwe geneesmiddelen zoals de turbomycine-klasse van antibiotica. Momenteel zijn er weliswaar nog geen therapeutica uit metagenomische screeningsprogramma's commercieel beschikbaar, maar dit is eerder te wijten aan het feit dat de tijd die nodig is

voor de verdere ontwikkeling (via klinische studies) en goedkeuring van dergelijke geneesmiddelen langer is dan de periode dat *metagenomics* als onderzoeksveld bestaat. De meeste nieuwe bioactieve moleculen die momenteel via metagenomische prospectie ontdekt werden, resulteerden voornamelijk uit screening van bodemstalen, waarbij 'herontdekking' van reeds gekende moleculen nog nergens werd gerapporteerd.

### **De enorme genetische diversiteit die waargenomen werd in een staal van de Sargasso Zee is nog maar het topje van de ijsberg aan genetisch potentieel**

Dit wijst erop dat het 'bodem-DNA' dat tot nu toe is gekend, slechts een kleine portie vormt van de beschikbare genetische diversiteit van het bodemmetagenoom. Bovendien wordt aangenomen dat een bodem een veel grotere genetische diversiteit bevat dan deze van watersystemen (zoals de Sargasso Zee). Door de dominante vaste fase in een bodem is deze immers veel minder onderhevig aan vermenging en vormt als het ware een aaneenschakeling van microbiotopen die op korte afstand van elkaar significant kunnen verschillen. Het is dus duidelijk dat de enorme genetische diversiteit die door het recente metagenomisch onderzoek van Venter werd waargenomen in een staal van de Sargasso Zee, nog maar het topje is van de ijsberg aan genetisch potentieel dat verder ontgonnen kan worden uit diverse biotopen voor zowel fundamenteel wetenschappelijke als meer toepassingsgerichte doeleinden. ●

J. Craig Venter et al., 'Environmental Genome Shotgun Sequencing of the Sargasso Sea', in: *Science*, 2004, nr. 304, 66-74.

R. Daniel, 'The soil metagenome – a rich source for the discovery of novel natural products', in: *Current Opinion in Biotechnology*, 2004, nr. 15, 199-204.