



DNA SEQUENCING

Ontrafelen van de levenscode

DNA is het molecuul van het leven. De code waarin alle instructies voor onze cellen wordt geschreven bestaat uit de letters A, C, G en T: de vier basen die de bekende wenteltrapstructuur van DNA vormen. Het lezen van die code komt neer op het bepalen van de basenvolgorde, ook wel *sequencing* genoemd. Dit kunstje beheersen biochemici al dertig jaar. Maar dan met relatief kleine genomen of met delen van grotere genomen. Ontrafeling van het menselijke DNA met drie miljard basenparen bleek heel wat moeilijker. In 1990 startte het eerste project om de sequentie van het menselijk genoom te bepalen. Sindsdien is de DNA-volgorde van liefst 95 procent van het menselijk genoom ontcijferd.

Het geheim van die ontcijfering is de polymerase kettingreactie (PCR), waarbij de DNA-bouwstenen zich vanzelf aaneen rijgen. In 1977 ontwikkelde Frederick Sanger een methode waarmee de volgorde waarin dit gebeurt kan wor-

den afgelezen met een kleurcode. De oorspronkelijke aanpak was langzaam en arbeidsintensief. Tegenwoordig moet alles sneller en goedkoper. Slimme trucs en krachtige computers maken dat de nieuwste sequencers een miljard basenparen per run analyseren in plaats van een paar miljoen. Kosten: bijna een miljoen euro voor een compleet humaan genoom. Het ultieme doel is die prijs omlaag te brengen naar duizend euro. Maar dat gaat nog wel even duren.

In deze Chemische Feitelijkheid

- De Context: Wat zijn de bouwstenen van DNA en waarom willen we hun volgorde weten?
- De Basis: Hoe kom je aan DNA en hoe bepaal je vervolgens de volgorde van de basen?
- De Diepte: Moderne analyse- en computertechnieken maken DNA-sequencing steeds sneller en goedkoper. Hoe werkt dat?

Van 95 procent van de basenparen in het menselijk DNA kennen we inmiddels de **exacte volgorde**. Dat hebben we vooral te danken aan het Human Genome Project en het Amerikaanse bedrijf Celera.

Strijd om vier letters

DNA wordt al in 1869 ontdekt, maar het zal nog tot 1952 duren voordat duidelijk is dat het molecuul de drager is van alle erfelijke eigenschappen. Een jaar later ontrafelen James Watson en Francis Crick de chemische structuur van DNA. Onder meer op basis van röntgenanalyses uitgevoerd door Wilkins en Franklin postuleerden zij in 1953 het wenteltrapmodel, gebaseerd op de dubbele helixstructuur.

Sindsdien staat DNA bekend als het instructieboekje voor het leven. De link met tekst of taal is makkelijk gelegd, omdat de code waarin de instructies zijn geschreven uit vier letters bestaat: A, C, G en T. Deze letters staan voor de vier basen (nucleotiden) waaruit DNA-moleculen zijn opgebouwd. Woorden in de DNA-taal bestaan uit combinaties van drie letters. En om de analogie verder



DNA-helix met de broncode van het leven.

door te voeren: deze drieletterige woorden vormen zinnen, de genen, die vertellen welke eiwitten een bepaalde cel moet produceren. Een gen codeert voor een eiwit. Alles wat cellen doen, staat beschreven in het DNA: welke cellen moeten groeien, en wanneer? Of welke cellen vormen een oog, en welke kleur heeft dat?

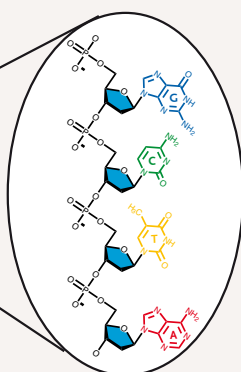
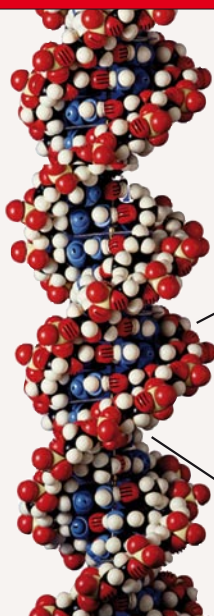
Het lezen van de codewoorden in DNA bestaat allereerst uit het bepalen van de basenvolgorde, ook wel *sequencing* genoemd. Dit kunstje beheerst men al dertig jaar. Frederick Sanger laat het in 1977 voor het eerst zien bij het DNA van een bacteriofaag. De Britse biochemicus weet dan met de hand – dus zonder hulp van machines of computers – de volgorde te achterhalen van de 5375 basenparen van het micro-organisme. Hij ontwikkelt daartoe een methode om de ketengroei van DNA selectief te beëindigen, waarna hij via gel-elektroforese de basenvolgorde van de DNA-fragmenten bepaalt. Deze vondst werd in 1980 bekroond met een Nobelprijs. Sangers tweede overigens, want in 1958 won hij de prijs al voor onderzoek naar de eiwitstructuur van insuline.

WEDLOOP

Het duurt vervolgens nog tien jaar voordat het sequencen ook kennis oplevert over ons eigen genoom, want pas in 1990 start het eerste project om de basenvolgorde van het menselijke DNA te bepalen: het *Human Genome Project* (HUGO). De klassieke uitvoering van de Sanger-methode bleek geschikt voor bacteriële genomen tot zes miljoen basenparen, maar is tot dan toe nog nooit gebruikt voor zoiets groots als het menselijke genoom. Officiële doel van het drie miljard dollar kostende Amerikaanse HUGO-project is het begrijpen van het complete humane genoom. Bij de start denkt men dat het project vijftien jaar in beslag zal nemen. Het bedrijf Celera (afgeleid van het Latijnse woord voor snel) van de Amerikaanse onderzoeker Craig Venter begint in 1998 een project met eenzelfde doel als HUGO. Daarmee start een wedloop. Celera wil veel sneller en


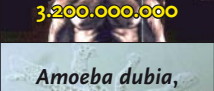
DNA IN DETAIL

In de kern van elke lichaamscel zit twee meter opgevouwen DNA, verdeeld over 23 paar chromosomen. Het totale DNA bestaat uit drie miljard basenparen (nucleotiden), waarin zich ongeveer 25.000 genen bevinden. Genen zijn stukjes DNA die coderen voor een bepaald eiwit. Tussen de genen zit 'junk'- ofwel 'non-coding' DNA.



Nucleotiden (A, C, G en T) bestaan uit een deoxyribose (een suiker), een fosfaatgroep en een base. DNA vormt bij kamertemperatuur een dubbele helix: de bekende wenteltrapstructuur. Hierin liggen de complementaire nucleotiden als paar tegenover elkaar (A - T en C - G).

OPGEHELDERDE GENOMEN

organisme aantal basenparen	jaar	bijzonderheden
 <i>bacteriøfaag phi-X174</i> 5375	1977	virus, eerste genoom
 <i>Pelagibacter ubique</i> 1.300.000	2005	kleinste bekende DNA van een zelfstandig organisme
 <i>Saccharomyces cerevisiae</i> 12.100.000	1996	bakkersgist, eerste eukaryoot
 <i>Caenorhabditis elegans</i> 97.000.000	1998	nematode worm, eerste meercellige
 <i>Arabidopsis thaliana</i> 120.000.000	2000	zandraket, eerste plant
 <i>Populus trichocarpa</i> 480.000.000	2006	populier, eerste boom
 <i>Gallus gallus</i> 1.000.000.000	2004	kip
 <i>Mus musculus</i> 2.500.000.000	2002	muis
 <i>Pan troglodytes</i> 3.100.000.000	2005	chimpansee
 <i>Homo sapiens</i> 3.200.000.000	2006	mens
 <i>Amoeba dubia</i> 670.000.000.000	2005	grootste bekende genoom

goedkoper – voor slechts 300 miljoen dollar – het hele humane genoom sequencen. Venter claimt bij aanvang dat dit hem slechts twee jaar zal kosten.

Celera kan sneller en goedkoper sequencen doordat men de klassieke methode verbetert en automatiseert. Het bedrijf beschikt over nieuwe en moderne DNA-analysers. Minstens zo belangrijk is de computerprogrammatuur, waarmee het bedrijf een grote snelheidswinst weet te boeken door dertig tot vijftig miljoen sequenties tegelijk te analyseren.

RUZIE

Er ontstaat discussie wanneer Celera de openbare resultaten van het HUGO-project gebruikt, maar de eigen resultaten niet toegankelijk wil maken. Terwijl dat eerder wél beloofd is. Celera wil de genen patenteren die het bedrijf tijdens het sequencen vindt om zo de investering in het dure gen-onderzoek terug te

verdiene. In maart 2000 stelt president Clinton dat een genoomsequentie niet gepatenteerd kan worden en dat de informatie vrij beschikbaar moet komen voor iedereen. De aandelen Celera kelderen en de Amerikaanse biotechsector verliest binnen twee dagen vijftig miljard (!) dollar.

De DNA-analyses gaan ondanks de ruzie gestaag door. Al in 2000 is een kladversie van het humane genoom gereed. In februari 2001 publiceren de wetenschappelijke tijdschriften *Nature* en *Science* tegelijkertijd een *draft sequence* waarin negentig procent van het genoom ontleed is. *Nature* publiceert de resultaten van het HUGO-project, terwijl *Science* kiest voor het uitbrengen van Celera's artikel. Uiteindelijk lijkt de competitie beide projecten goed te doen. Binnen het publieke HUGO-project schaft het Amerikaanse onderzoeksinstituut MIT bijvoorbeeld nieuwe sequencing-apparaten aan waarmee de analysesnelheid met een factor vijftien wordt verhoogd. Het complete genoom is in april 2003 klaar en in mei 2006 wordt de sequentie van het laatste chromosoom in *Nature* gepubliceerd. De onenigheid over de data duurt nog steeds voort.

ECHT COMPLEET?

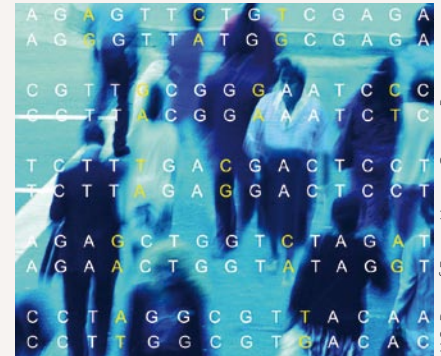
Overigens betekent compleet niet letterlijk dat alle ruim drie miljard basenparen van ons genoom nu bekend zijn. Schattingen lopen uiteen van 92 tot 95 procent. Verschillende chromosoomregio's blijken namelijk moeilijk te sequencen, bijvoorbeeld stukken DNA met veel herhalingen. Het baarde veel opzien dat menselijke eigenschappen, die zijn opgeslagen in de genen, maar een heel klein deel van het totale humane genoom uitmaken. Bij de start van HUGO schatte men zeker twee miljoen genen aan te treffen. De schattingen zijn meerdere malen bijgesteld en inmiddels gaan



Het ACGT van de mens. iStockphoto, Vasily Yakobchuk

GENETISCHE PATRONEN

En van de vervolgstudies van HUGO is het internationale HapMap-project, dat in 2002 van start ging. HapMap is een catalogus met veelvoorkomende genetische variaties in mensen. Het gaat om SNPs ofwel 'snips' (*Single Nucleotide Polymorphisms*), variaties van één basenpaar in het DNA van verschillende individuen. Iemand kan een A hebben op een bepaalde plek in het genoom terwijl een ander individu op die plaats een G heeft.



U.S. Dept of Energy Human Genome Program

De catalogus beschrijft niet alleen de variaties zelf, maar ook waar deze in het DNA zitten en hoe ze voorkomen bij bevolkingsgroepen in verschillende delen van de wereld. Deze database moet een hulpmiddel worden bij het vinden van genen die een rol spelen bij gezondheid, ziekte, gevoeligheid voor medicijnen, enzovoort.

onderzoekers uit van 20.000 à 25.000 genen. De rest van het DNA (omgerekend zo'n 96 procent) heeft een onbekende functie, populair aangeduid als 'junk-DNA'. Een betere term lijkt 'non-coding DNA', wat aangeeft dat het niet codeert voor de aanmaak van een eiwit. Niet iedereen gelooft namelijk dat junk-DNA helemaal geen functie heeft. Verspreid over het genoom liggen nog enkele tientallen 'gaten', waarvan de DNA-volgorde de komende jaren nog achterhaald moet worden. Onderzoekers verwachten niet dat die onbekende stukken ook genen bevatten, maar dat is niet zeker.

Het ophelderen van de basenvolgorde in het menselijk DNA is een hele prestatie. Desondanks weten we eigenlijk nog niet veel meer dan vijftien jaar geleden. Zoals een onderzoeker van MIT het stelt: 'We hebben nu een lijst met onderdelen. Als je alle onderdelen van een Boeing 747 zou bestellen weet je al een heleboel over het vliegtuig. Maar je kan het waarschijnlijk niet in elkaar zetten, laat staan dat je begrijpt waarom het ooit zal kunnen vliegen.'

De vier basen waaruit DNA bestaat kunnen zichzelf als kralen aaneen rijgen. Tenminste, onder de juiste omstandigheden. De sequentietechniek waarmee de **basenvolgorde** wordt bepaald maakt heel handig gebruik van deze eigenschap.

De code ontcijferen

DNA zit in alle cellen van *eukaryoten*, organismen waarvan de cel een kern bevat. Het eerste genoom dat in de jaren zeventig werd opgehelderd was van een bacteriofaag (een klein virus in bacteriën) en bestond uit iets meer dan vijfduizend basenparen. Heel andere koek dan het menselijke DNA, dat ruim drie miljard basenparen telt. Bij de start van het *Human Genome Project* (HUGO) in 1990 was nooit eerder zo'n groot genoom opgehelderd.

Uiteraard was voor het HUGO-project menselijk DNA nodig. Wiens DNA werd eigenlijk gebruikt? Om privacy-redenen is dat nooit onthuld. Als iemands persoonlijke genenkaart bekend is, valt immers ook veel af te leiden over lichamelijke eigenschappen en zelfs ziekten. Tijdens de genoomwedloop werd DNA verzameld van enkele tientallen individuen die anoniem bleven. Al is het algemeen bekend dat ook Craig Venter, de baas van Celera, DNA afstond voor de sequentieanalyses binnen zijn eigen bedrijf.

Het is het eenvoudigst om DNA te winnen uit zaadcellen – maar die heeft alleen een man – of uit rode bloedcellen. Met moderne technieken kan men tegen-

woordig ook DNA halen uit de meeste andere lichaamsmaterialen, zoals haar of huid. Voor een DNA-analyse volstaat een heel kleine hoeveelheid van één picogram ofwel 10^{-12} gram DNA.

KRALEN RIJGEN

Een DNA-sequentieanalyse is gebaseerd op de polymerase kettingreactie, vaak afgekort als PCR (polymerase chain reaction). Het is een manier om een kleine hoeveelheid DNA te vermenigvuldigen naar een grote hoeveelheid identiek DNA. Dit proces heet amplificatie. Onder invloed van het enzym polymerase worden de losse basen (A,C,G en T) aan elkaar geregen in precies dezelfde volgorde als in het oorspronkelijke DNA. De amplificatie vindt plaats in drie stappen. In de eerste stap wordt door verhitting (94 °C) het voorbeeld-DNA gedensatureerd waarbij de dubbele helix uiteenvalt in twee enkele ketens. Dat zijn de zogeheten templates. In de tweede stap vindt, na afkoeling tot 37 °C, hybridisatie plaats waarbij een klein stukje hulp-DNA, de primer, aan de template wordt gekoppeld. De primer is nodig om een beginnetje te maken voor de derde stap, het rijgen van de nucleo-

tiden. Deze ketenverlenging (extensie) vindt plaats onder invloed van het enzym polymerase en start als het mengsel weer is verhit tot 72 °C. De andere helft van de template wordt base voor base aaneengeregen: dit kan maar op één manier, omdat adenine (A) alléén tegenover thymine (T) kan binden en guanine (G) alléén tegenover cytosine (C). Hierbij ontstaat weer dubbelstrengs DNA en dat is een kopie van het oorspronkelijke DNA. De oplossing bevat dan twee maal zoveel van het gewenste DNA. Door de drie stappen een aantal malen (bijvoorbeeld dertig) te herhalen, groeit de hoeveelheid gewenst DNA exponentieel.

SANGER-METHODE

Voor het sequencen wordt hetzelfde DNA-kopieerprincipe gebruikt, maar Sanger bedacht een methode om het oorspronkelijke DNA te vermenigvuldigen in de vorm van heel veel kleine deelfragmenten. Dat gebeurt met de zogeheten ketenterminatietechniek. Door naast de 'gewone basen' (deoxynucleotiden) een klein percentage chemisch afwijkende nucleotiden toe te voegen, kan de ketengroei worden gestopt. Deze chemische variant, de dideoxynucleotide, lijkt precies op een gewone nucleotide, maar heeft geen 3'-hydroxygroep, waardoor aan deze bouwsteen geen volgend nucleotide kan worden gekoppeld en de ketengroei stopt.

Bij het sequencen wordt aan het reactiemengsel behalve de gewone DNA-bouwsteen ongeveer vijf procent van de speciale variant toegevoegd. Deze dideoxynucleotiden fungeren als 'ketenstoppers' waardoor allemaal kleine stukjes van het oorspronkelijke DNA ontstaan. De clou is dat de meeste nucleotiden van het gewone type zijn en slechts een kleine minderheid dideoxynucleotiden. Ter verduidelijking

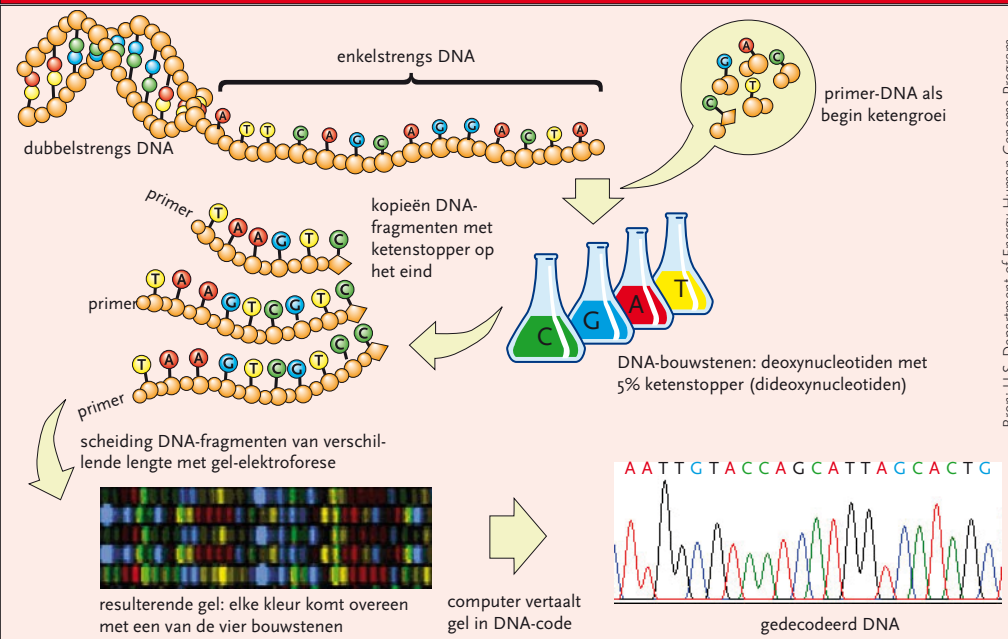
ZELF DNA ISOLEREN

Thuis experimenteren met DNA? Volg dan de onderstaande stappen.

- Doe 100 gram spliterwt (het kan ook met spinazie, kippenlevertjes, aardbeien of broccoli) in een blender
- Voeg een mespuntje zout en 200 ml (1 glas) water toe en mix 15 seconden (hoogste stand). Je hebt nu erwtenzellen-soep
- Zeef de soep in een kopje en voeg 2 eetlepels vloeibare zeep toe
- Meng door even met het kopje te zwenken en laat 5-10 minuten staan
- Schenk de mix in reageerbuisjes of andere kleine glazen potjes, elk ongeveer 1/3 vol
- Voeg als enzym een middel toe dat vlees mals maakt, dat kan ook ananassap of contactlensvloeistof zijn
- Voeg voorzichtig een laagje alcohol (70-95% isopropylalcohol of ethanol) toe zodat het bovenop de erwtenmix blijft staan
- DNA zal als een witte draderige smurrie naar de alcoholaag gaan. Gebruik een houten stokje om het uit het buisje te halen



BASENVOLGORDE AFLEZEN



nemen we het dideoxynucleotide van thymine als voorbeeld. Tijdens de ketengroei zal wanneer een T nodig is meestal een gewone deoxy-T worden gebruikt. Onder invloed van polymerase gaat de ketengroei dan gewoon verder. Echter in vijf procent van de gevallen krijgt het enzym te maken met een dideoxy-T, waarna de ketengroei stopt. Vroeger of later zullen alle kopieën afgesloten worden met zo'n T. Wat resulteert is een mengsel gekopieerde DNA-fragmenten met allerlei verschillende lengtes, die allemaal een T op het eind hebben. Door die fragmenten via elektroforese op lengte te scheiden, kan worden afgeleid op welke plekken een T zat in de oorspronkelijke DNA-keten. Als bij de scheiding bijvoorbeeld blijkt dat er ketens bestaan van 3, 9, 15 en 22 nucleotiden lang, dan zat in het oorspronkelijke DNA op al die plaatsen een T.

De werkelijkheid is nog wat ingewikkelder. Dan zitten er immers vier verschillende ketenstoppers in het mengsel. Om ze van elkaar te kunnen onderscheiden worden ze voorzien van een kleurlabel. Wanneer de DNA-fragmenten op lengte zijn gescheiden zijn de vier verschillende nucleotiden met behulp van een laser te achterhalen. Aan de hand van de kleur kan worden afgeleid welke nucleotide op welke plek zit.

PUZZELN

Deze Sanger-methode is uitstekend geschikt om in DNA-fragmenten van maximaal zo'n 700 basenparen lang de sequentie te bepalen. Maar hoe moet

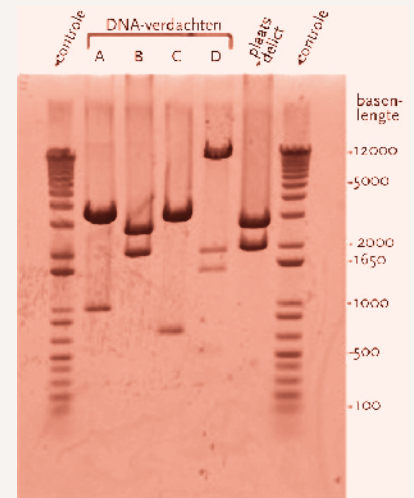
dat nu met het humane genoom dat maar liefst drie miljard basenparen lang is? Voor deze geweldige uitdaging stonden de onderzoekers toen ze zich stortten op het humane genoom project. In feite zijn er twee benaderingen toegepast, de ene is gevolgd door het publieke HUGO-project en de andere door het commerciële Celera-project. In het eerste geval hebben de onderzoekers het hele humane genoom in stukken verdeeld van zo'n 100.000 tot 300.000 basenparen. Van elk deelfragment zijn in bacteriën duizenden klonen gemaakt: zogeheten BAC-klonen (Bacterial Artificial Chromosome). Op deze wijze is een grote BAC-bibliotheek opgezet en deze klonen zijn afzonderlijk gesequencet. BAC's zijn echter nog veel te groot om in één keer te sequencen. De zogeheten *shotgun* methode bracht de oplossing. Hierbij wordt BAC-DNA vooraf gefragmenteerd tot kleinere delen, die simultaan worden gesequencet. In een volgende stap moet dan de basenvolgorde in de afzonderlijke fragmenten weer als een puzzel in elkaar gelegd worden om de volgorde in de oorspronkelijke BAC op te helderen. Voor het leggen van deze puzzel zijn computerprogramma's ontwikkeld.

De Celera-benadering is nog veel complexer, omdat de onderzoekers niet uitgaan van BAC-klonen, maar in één keer het hele humane genoom sequencen met behulp van de zogeheten *whole-genome shotgun* methode. Deze aanpak bespaart veel werk en tijd, maar levert natuurlijk een veel ingewikkelder puzzel op.

DNA-FINGERPRINT

De volgorde waarin de letters en woorden uit het DNA-alfabet gerangschikt zijn in menselijk DNA is grotendeels gelijk. Van de drie miljard basenparen zijn er slechts drie miljoen voor ieder mens uniek. Je kunt daarmee dus onderscheid maken tussen individuen – net als bij een vingerafdruk. De techniek wordt dan ook *DNA-fingerprinting* genoemd. Centraal bij deze techniek staat het zogenaamde junk-DNA of *non-coding* DNA. Junk-DNA heeft veel herhalende patronen, zogenaamde Short Tandem Repeats (STR's). Dit zijn stukjes DNA van twee tot tien bouwstenen met dezelfde volgorde die herhaald voorkomen, bijvoorbeeld: AATCAATCAATCAATC. Het aantal herhalingen binnen zo'n STR is voor ieder individu verschillend.

De basenvolgorde wordt op een gel vastgesteld en hiervan wordt een foto gemaakt. De verschillende gescheiden bandjes zijn te zien als donkere streepjes. De analyse geeft zodoende een 'streepjescode' voor een persoon. Deze code wordt vergeleken met die van een familielid, of bij forensisch onderzoek met het DNA van een misdadiger. Zo kan de identiteit van de onbekende persoon worden vastgesteld.



De kans dat er twee mensen (die niet direct familie van elkaar zijn) hetzelfde DNA-profiel hebben, is ongeveer één op de miljard. Omdat de streepjescode komt uit een stuk DNA zonder genen zegt het overigens helemaal niets over de eigenschappen van die persoon.

Vergeleken met een gewone legpuzzel komt het erop neer dat als onderzoekers in het HUGO-project 1000 puzzels van elk 500 stukjes in elkaar leggen, de Celera-onderzoekers in een keer een puzzel van 500.000 stukjes leggen. Dankzij zeer krachtige computers en software zijn ze daar toch in geslaagd.

Al in 1990 voorspelden insiders dat de DNA-technologie een hoge vlucht zou nemen, waardoor het sequencen van DNA steeds sneller en goedkoper zou worden. Die voorspelling kwam uit. Toch lijkt het **duizend dollar genoom** nog ver weg.

Nieuwe generaties sequencers

In 1998 was het HUGO-project al acht jaar onderweg, maar waren er wereldwijd nog maar 200 miljoen basenparen geanalyseerd. Oorzaak: de klassieke sequentiemethode was erg tijdrovend en arbeidsintensief. Met de techniek konden in het begin slechts zes tot acht monsters tegelijk geanalyseerd worden in 32 laantjes (acht monsters keer vier basenparen). De monsters zelf waren ieder zo'n 100 tot 200 basenparen groot. Nadat de methode was geautomatiseerd konden zestien monsters van ieder vierhonderd basenparen in één nacht gemeten worden.

Op zich een verbetering, maar met die snelheid zou het sequencen van het menselijk genoom nog steeds vele tientallen

jaren in beslag nemen. Nieuwe technieken waren dus hard nodig om de doelstellingen van het HUGO-project te halen: in 2003 drie miljard basenparen, meer dan honderd keer zoveel als er in alle jaren daarvoor waren gesequencet. Gelukkig kwamen die technieken er.

CAPILLAIRE SCHEIDING

De introductie van capillaire sequencers betekende al een grote stap voorwaarts. De gel-elektroforese stap, waarbij de DNA-fragmenten op lengte worden gescheiden, maakte plaats voor een capillaire elektroforese. Hiermee verloopt het scheidingsproces veel sneller en met een veel hogere resolutie. Bovendien

is capillaire elektroforese minder arbeidsintensief omdat het gemakkelijk geautomatiseerd kan worden. Dat betekent snelheidswinst: alleen al in januari 2003 analyseerde men op die manier binnen het HUGO-project anderhalf miljoen basenparen.

De eerste capillaire sequencer had één capillair dat, net als bij gel-sequencing, in drie uur tijd een sequentie van vijfhonderd basenparen kon verwerken. Tegenwoordig bestaan er apparaten met 96 capillairen die ongeveer twee uur nodig hebben om 96 monsters van elk achthonderd basenparen te sequencen. Er zijn ook langere capillairen op de markt, die in drie uur tijd DNA-fragmenten van duizend basenparen kunnen verwerken. De snelste machines lezen binnen een etmaal ongeveer 2,1 miljoen basen af. Om fouten uit te sluiten moet een analyse wel meerdere keren worden uitgevoerd; soms zelfs in zesvoud. Met deze snelheid duurt het sequencen van het menselijk genoom 24 jaar.

PARALLELE SYNTHESE

Sequencers gebaseerd op de polymerase kettingreactie met ketenterminatie gevolgd door DNA-fragment analyse (de Sanger-methode) zijn van groot praktisch nut, maar lijken inmiddels uitontwikkeld. Voor een nieuwe stap voorwaarts is een nieuwe generatie sequencers nodig. Daartoe behoort de zogeheten 'massief parallelle pyrosequencing' techniek: 'massief' omdat het heel veel DNA-fragmenten kan analyseren (200.000 tot 400.000), 'parallel' omdat dat allemaal tegelijk gaat, en 'pyro' omdat de ingebouwde nucleotide een lichtflits genereert. Het Amerikaanse bedrijf 454 Life Sciences ontwikkelde in 2000 de zogenaamde *Sequencing-by-synthesis* technologie die gebaseerd is op pyrosequencing. Zowel de monstervoor-

TOPTOMATEN

Bij het veredelen van planten is het essentieel om snel te weten welk plantje bijvoorbeeld veel tomaten gaat geven en welke niet. Genetici zoeken daarom naar markers die aangeven of een plant veel of weinig oogst gaat opleveren. Vaak is een verschil van één base al een aanwijzing voor een andere eigenschap. Zo'n afwijking in een base heet een SNP ofwel *Single Nucleotide Polymorphism*. Een manier voor het vinden van SNP's is het ophelderen van het hele genoom van de tomaat, en dan voor elk plantje de hele sequentie bestuderen en kijken waar de SNP zit. Onhaalbaar voor plantenveredelingsbedrijven, want sequencing is een dure techniek.

Het Wageningse biotechbedrijf Keygene bedacht een slimme techniek om markers in een genoom op te zoeken zónder dat de hele DNA-sequentie bekend is: *Amplified Fragment Length Polymorphism*, kortweg AFLP. Met deze techniek heeft Keygene in komkommers drie genetische markers gevonden die verraden of een plant resistent is tegen een bepaald virus dat in Spanje veel problemen veroorzaakt bij de komkommerteelt.



foto: Bastienne Wentzel

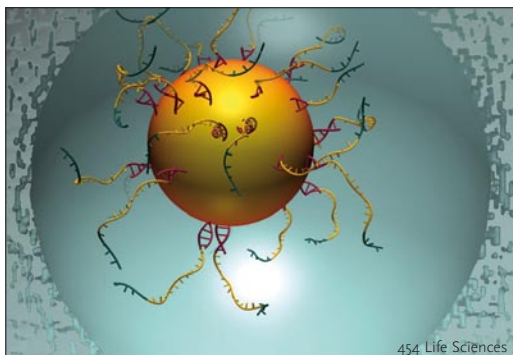
GENETISCHE VOETAFDruk

Dat de eerste mensen uit Afrika stammen is geen nieuws. Maar hoe zijn ze vervolgens uitgezwermd over de aardbol, nadat ze zestigduizend jaar geleden het Afrikaanse continent voor het eerst verlieten? De Amerikaanse National Geographic Society probeert dit samen met IBM te achterhalen door genetische voetafdrukken in kaart te brengen. Binnen het zogeheten *Genographic Project* speuren zij naar *markers* in het DNA van Y-chromosomen en het mitochondriale DNA. Mutaties in dit DNA zijn niet ontstaan door 'vermenging', omdat dit DNA alléén van vader-op-zoon of moeder-op-dochter wordt doorgegeven. Het is de bedoeling om het DNA van zeker 100.000 mensen te verzamelen en analyseren. Het project startte in april 2005 en zal vijf jaar duren.



bereiding als de daadwerkelijke sequenc-ing gaan bij pyrosequencing sneller dan bij gebruikelijke sequentiemethoden.

Het DNA wordt aan ultrakleine kraaltjes (*beads*) gebonden; aan elk kraaltje één fragment enkelstrengs DNA. Vervolgens wordt met behulp van de polymerase kettingreactie dit DNA-fragment (van ongeveer 100 basenparen lang) vermenigvuldigd. Het kraaltje raakt vol met stukjes identiek DNA en wordt overgebracht naar een picotiterplaat met 1,6 miljoen gaatjes ofwel *wells*. Ieder kraaltje krijgt zijn eigen gaatje. Vervolgens laat men over de plaat een reactiemengsel stromen met één van de vier basen, zodat een koppelingsreactie kan plaatsvinden. Telkens als de base de complementaire base detecteert wordt die ingebouwd in de DNA-streng. Hierbij komt een serie chemische reacties op gang, die eindigt met de omzetting van luciferine in oxyluciferine. Dat



454 Life Sciences

Massief parallelle pyrosequencing van het Amerikaanse 454 Life Sciences is een snelle techniek waarbij de DNA-strengen aan een kraal worden gebonden.

gaat gepaard met een lichtflits, die wordt geregistreerd door een sensor. Daarna leidt men een mengsel met de volgende base over de plaat en begint het proces opnieuw. Van alle 1,6 miljoen DNA-fragmenten wordt zo parallel en automatisch de volgorde bepaald. Per run van vijf uur kan deze techniek twintig miljoen basenparen analyseren – vijftig keer zo snel als de snelste capillaire sequencers. De allernieuwste versie van deze machine kan nog eens vijf keer zoveel DNA tegelijk analyseren.

Er bestaat ook een 'niet-massieve' versie van de pyrosequencing techniek. Die analyseert 96 monsters tegelijkertijd met een snelheid van één basenpaar per minuut. Deze aanpak leent zich goed om kleine verschillen tussen twee DNA-fragmenten te achterhalen, terwijl de massieve methode geschikt is om snel een compleet genoom op te helderen.

BASE VOOR BASE

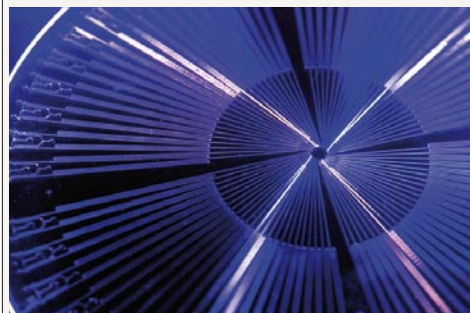
Snelheid is niet de enige belangrijke factor bij de keuze voor een bepaalde sequentiemethode. De pyrotechniek is zeer geschikt om snel een heel genoom van bijvoorbeeld een bacterie te analyseren, terwijl het met capillaire sequencing gemakkelijker is om specifieke genen te analyseren die slechts uit honderden tot duizenden basenparen bestaan. Pyrosequencing kan bovendien problemen geven met homopolymeren: DNA met repeterende kopieën van een bepaalde base, bijvoorbeeld tienmaal A naast elkaar. In dat geval worden bij de inbouwreactie meerdere basen tegelijk gekoppeld. Het lichtflitsje waarmee dit gepaard gaat is evenredig sterker – bij het inbouwen van twee basen is het twee keer zo intens – maar het verschil bij zeven of acht basen is niet groot. De analyse van zo'n DNA-fragment is daardoor onnauwkeurig.

Een oplossing voor dit probleem is een nieuwe techniek van het Amerikaanse bedrijf Solexa. De DNA-fragmenten worden hierbij niet aan kraaltjes gebonden, maar aan een oppervlak en vervolgens vermenigvuldigd. Dit levert clusters DNA op met dezelfde basenvolgorde. Vervolgens wordt de Sanger-ketenterminatiereactie uitgevoerd, maar dan met een nucleotidenreagens van 100 procent gelabelde dideoxynucleotiden. Na aflezen van het label wordt de ketenstopper verwijderd, waarna de volgende gelabelde base zich kan binden. Het DNA wordt op deze manier stap voor stap afgelezen.

GOEDKOOP GENOOM

Het sequencen van DNA is steeds goedkoper geworden. Bij de start van HUGO schatte men de kosten voor een volledig humaan genoom op drie miljard dollar, ofwel een dollar per basenpaar. Dat is inmiddels afgenomen tot enkele centen per base: de kosten liggen tussen de vier en acht euro voor één monster van 500 basenparen. Voor een compleet menselijk genoom betekent dit dus 25 tot 50 miljoen dollar. Veel goedkoper is de massief parallelle pyrosequencing: één run van 20 miljoen basenparen kost 5.000 euro; omgerekend 0,025 cent per base ofwel 780.000 euro voor het hele menselijke genoom.

Daarmee heeft men het zogenaamde 'duizend-dollar genoom', het sequencen van het menselijk genoom met één machine binnen een redelijke tijd voor duizend dollar, nog niet binnen handbereik. Volgens de Amerikaanse fabrikant Applied Biosystems is daarvoor *single molecule sequencing* nodig. Deze techniek zal één DNA-molecuul kunnen analyseren – dus zonder de noodzaak van amplificatie. Naar verwachting zal het nog jaren duren voor deze methode op de markt komt. Overigens heeft de Amerikaanse X-Prize Foundation (bekend van de toeristische ruimtevaart) een prijs van tien miljoen dollar uitgelooft voor het sequencen van honderd menselijke genomen binnen tien dagen. Voorlopig is dat nog onhaalbaar.



De nieuwe generatie sequencers decodeert DNA duizenden keren sneller en duizenden keren goedkoper dan de klassieke sequencers.

Ook homopolymeren kunnen hiermee nauwkeurig worden geanalyseerd. Naar verwachting kunnen met deze techniek tien tot veertig miljoen stukken DNA van 25 à 30 basenparen tegelijk afgelezen worden. Een andere nieuwe techniek is SOLiD (*Sequencing by Oligonucleotide Ligation and Detection*) waarmee per run veertig miljoen fragmenten van ieder 25 basenparen kunnen worden uitgelezen. In totaal dus een miljard basenparen. Een nauwkeurige analyse van het hele menselijke genoom kost dan nog slechts een paar weken.

Meer weten

AANBEVOLEN LITERATUUR

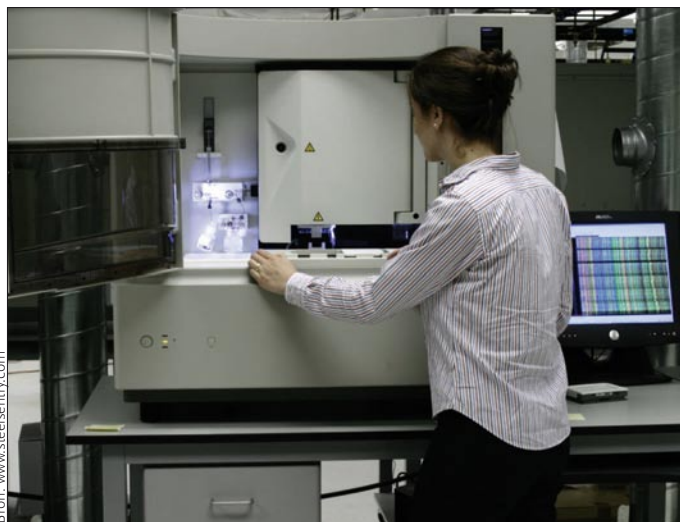
- J.D. McPherson et al, *Nature* vol 409 (15-02-2001), p. 934
- J.C. Venter et al., *Science* vol 291 (16-02-2001), p. 1304
- J. Shreeve, *The genome war: how Craig Venter tried to capture the code of life and save the world*, 1st edition, Knopf, 2004, 416 pp, ISBN 0375406298
- L. Alphey, *DNA Sequencing*, 1st edition, Garland Science, 1997, 224 pp, ISBN 1859960618

AANBEVOLEN WEBSITES

- www.ornl.gov/sci/techresources/Human_Genome/home.shtml: informatie over het Human Genome Project (HUGO)
- <http://seqcore.brcf.med.umich.edu/doc/educ/dnapr/sequencing.html>: hoe werkt DNA-sequencing?
- www.kennislink.nl/web/show?id=94541: dossier over DNA
- www.pbs.org/wgbh/nova/genome/program.html: online documentaires over DNA: "Cracking the code of life" (Engels)
- www.dnaftb.org/dnaftb/: lessen over DNA, genetica en sequencing (Engels)
- www.454lifesciences.com/enabling-technology/the-process.asp: uitleg van Massive Parallel Pyrosequencing
- <https://www3.nationalgeographic.com/genographic/>: het Genographic Project

VOOR OP SCHOOL

1. Waar staat de afkorting DNA voor? Wat betekent die naam? Waar staan de letters A, C, T en G voor?
2. A,C,T,G zijn basen. Geef aan waarop die basische eigenschap berust. Noteer een zuur-basereactie met een van de vier basen.
3. Waardoor kan een T alleen met een A binden en een G alleen tegenover een C staan in de dubbele helix van DNA?
4. Stel je start met één DNA-molecuul en je laat dat dertig maal de cyclus met PCR doorlopen. Hoeveel moleculen heb je dan verkregen?



Moderne sequencer die gebruikt wordt bij het ontrafelen van de volgorde van basenparen in DNA.

5. Voor de reactie waarmee DNA wordt vermeerderd of gesequencet is het enzym Taq polymerase nodig. Wat doet dit enzym? Waarom is juist dit enzym zo geschikt voor deze reactie?
6. Bepaling van de 'streepjescode' gaat met behulp van gelelektroforese. Wat is de basis van deze techniek?
7. Frederick Sanger kreeg ook de Nobelprijs voor onderzoek naar insuline. Uit hoeveel aminozuren bestaat dit eiwit? Geef de code (in A,C,T, G) voor een deel van de eiwitketen. Gebruik zonnig Binas tabel 70.
8. Geef een argument voor verlening van patent op de genomsequentie en een argument tegen.
9. Geef aan hoe de 'streepjescode' gebruikt wordt in de civiele rechtspraak (dus niet alleen bij misdadigers in de strafrechtpraak).
10. In DNA komt deoxyribose voor. In RNA ribose en in het 'stopnucleotide' vind je dideoxyribose. Wat is het verschil tussen deze suikers? Beredeneer de oorzaak van de verschillende activiteiten van deze moleculen.

COLOFON

Chemische Feitelikheden: actuele encyclopedie over moleculen, mensen, materialen en milieu. Losbladige uitgave van de KNCV, verschijnt drie maal per jaar met in totaal tien onderwerpen.

Redactie:

Alexander Duyndam (C2W)
Marian van Opstal (Bèta Communicaties)
Arthur van Zuylen (Bèta Communicaties)
Gerard Stout (Noordelijke Hogeschool Leeuwarden)

Basisontwerp: Menno Landstra

Redactie en realisatie:

Bèta Communicaties
tel. 070-306 07 26
betacom@planet.nl

Uitgever:

Roeland Dobbelaer
Bèta Publishers
Postbus 249, 2260 AE Leidschendam
tel. 070-444 06 00
fax 070-337 87 99
info@betapublishers.nl

Abonnementen opgeven:

Abonnementenland
De Trompet 1739, 1967 DB Heemskerk
tel. 0251-31 39 39
fax 0251-31 04 05
aboservice@aboland.nl

Abonnementen kunnen elk moment ingaan. Abonnementen worden automatisch verlengd tenzij vóór 1 november van het lopende jaar een schriftelijke opzegging is ontvangen.

Abonnementen:

- papieren editie en toegang tot digitaal archief op internet: (inclusief verzamelmap): € 75,-
KNCV- en KVCV-leden: € 65,-
- alleen toegang tot digitaal archief op internet: € 60,-
KNCV- en KVCV-leden: € 50,-

DNA-SEQUENCING

editie 52
nummer 233
maart 2007

Met dank aan:

- Dr. Nathalie van Orsouw
Keygene
e-mail: nathalie.vanorsouw@keygene.com
- Jon Wittendorp
Keygene
e-mail: jon.wittendorp@keygene.com