

GENETISCHE informatie OP het INTERNET

Door Ellen Stolte (ir HH Stolte)

ellen.stolte@wur.nl, Celbiologie en Immunologie Wageningen Universiteit, Organismale Dierfysiologie Radboud Universiteit Nijmegen

Is DNA van ons allemaal ?!

Sinds de ontdekking van het DNA door Watson en Crick en het daaropvolgende besef dat dit DNA codeert voor het gehele bouwplan van ons lichaam en voor alle regelsystemen om het goed te laten functioneren heeft het onderzoek naar het DNA een grote vlucht genomen.

Het bepalen van de "code van het leven" wordt gebruikt in fundamenteel (biologisch) onderzoek om te begrijpen wat leven is en wat verschillende organismen definieert. Ook in toegepast (medisch) onderzoek wordt DNA gebruikt om ziektes met een genetische oorzaak te achterhalen en eventueel zelfs te behandelen.

De genetische informatie van vele diersoorten vormt een enorme bron van gemeenschappelijke kennis die op het World Wide Web vrij verkrijgbaar is. Onderzoekers maken gebruik van informatie die voor hun onderzoek van toepassing is en plaatsen op hun beurt informatie op het net die zij gevonden hebben, zodat de hoeveelheid informatie alleen maar groeit.

Op veel grotere schaal worden er projecten gestart om alle genetische informatie van een bepaalde diersoort (genoom) volledig beschikbaar te maken. Inmiddels zijn van de mens, muis, kip, zebravis en vele andere soorten alle genen gecodeerd en in kaart gebracht. Dit soort grootschalige projecten wordt vaak gefinancierd door overheden en onderzoeksinstituten, omdat met het vrijkomen van de genetische informatie de weg openligt voor nieuw onderzoek en nieuwe medicijnen. Vanzelfsprekend is deze informatie ook van belang voor grote bedrijven in de farmaceutische of voedingsindustrie en daarom ondersteunen zij financieel of materieel (sequence faciliteiten) dit soort grootschalige projecten. Echter, als compensatie voor hun financiële steun verwachten zij eerste inzage in de resultaten. In een enkel geval wordt zelfs een patent aangevraagd op een gen of een stuk genetische informatie. Hierdoor kan de volledige beschikbaarheid van genetische informatie in het geding komen.

Bepaling van de DNA sequentie

De genetische code is vastgelegd in het DNA (DeoxyriboNucleic Acid). DNA is een lange streng bestaande uit simpele eenheden die nucleotiden heten en bij elkaar gehouden worden door suiker- en fosfaat-groepen. Er zijn vier verschillende nucleotiden (A, G, C en T) en de volgorde hiervan bepaalt de genetische informatie die in het DNA is vastgelegd. In een cel zit het DNA in een gepaarde helix; een matrijsstreng en een kopie streng. In

deze helix zit de A altijd tegenover de T en de G tegenover de C. Om de genetische code af te lezen maken cellen een kopie van een stuk van het DNA dat RNA genoemd wordt. Dit RNA vormt een directe aanwijzing voor de volgorde van de aminozuren waaruit elk eiwit in een organisme is opgebouwd.

Om de genetische informatie van een organisme te ontfangen gebruikt men DNA sequencing. Tijdens DNA amplificatie wordt eerst de DNA helix een stuk ontwonden en vervolgens een starter ("primer") aan het begin van het te coderen stuk DNA op een van de twee strengen van de helix gezet. Vervolgens verlengt het enzym DNA polymerase dit startstukje, door elke keer als het op de matrijs streng een A ziet een T in te bouwen en als het een C ziet een G in te bouwen. Zo wordt uiteindelijk de hele streng gekopieerd. De starter zal vervolgens opnieuw aan het begin van het DNA binden en het proces begint van voren af aan. Als er in dit proces af en toe een speciale nucleotide wordt ingebouwd, die het proces van verlenging voortijdig beëindigt, en waaraan een fluorescent label is gehangen kun je de precieze volgorde van de nucleotiden achterhalen. Dit gebeurt door de fragmenten van verschillende lengte te scheiden in een glascapillair met een visceuze polymeer; de verschillende lengtes corresponderen met een verschillende kleur fluorescent label en daarmee met een verschillende nucleotide.

Inmiddels is het "sequencen" zoals bovenstaand proces genoemd wordt voor een groot deel geautomatiseerd.

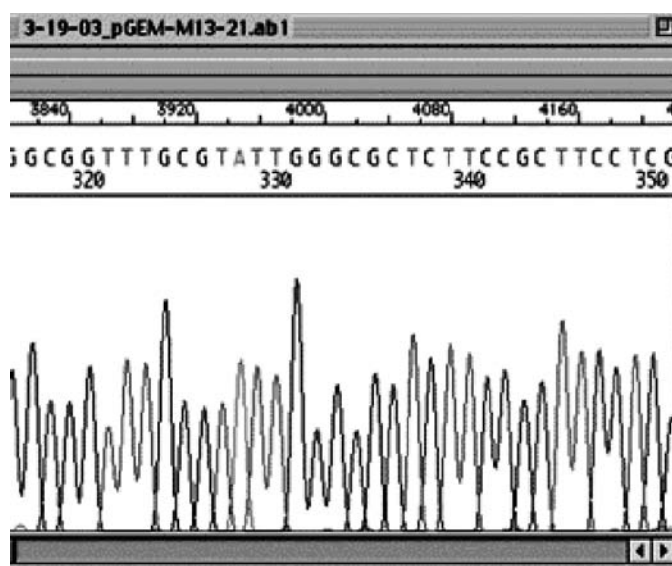


Fig 1. Chromatogram na DNA sequencing, de verschillende fluorescente signalen worden afgelezen en vertaald in de nucleotiden volgorde.

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG }	UGU } Cys UGC } UGA } Stop UGG } Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

Fig 2: Vertaling van de genetische code; in dit schema is af te lezen hoe de combinatie van de 3 verschillende nucleotiden die een codon vormen samen coderen voor een aminozuur. Bijvoorbeeld GGG rechtsonderaan codeert voor het aminozuur Glycine, met de afkorting G.

NB; Bij dit figuur moet rekening worden gehouden dat dit de RNA vertaling is van het DNA die in de cel plaatsvindt en niet de DNA vertaling van het sequencing proces. Het enige verschil hierin is dat de T van het DNA vervangen is door een U in het RNA.

Robots worden ingezet voor de toevoeging van chemicaliën evenals voor de waarneming van de fluorescente signalen. Uiteindelijk krijgt de onderzoeker naar keuze een chromatogram van deze signalen te zien, of een directe vertaling hiervan in de bijbehorende nucleotiden.

De volgorde van de verschillende nucleotiden kan vervolgens worden vertaald in de aminozuurvolgorde van een eiwit. Drie nucleotiden vormen samen een codon, dat correspondeert met een aminozuur.

atgaagttcttctgtgtctttccaccattgggttctgctggg
M K F F L L L F T I G F C W
acacaacaaggacggacatctattgttcattgttgaatggc
T Q Q G R T S I V H L F E W
cttgaatgtgagcgatatttagctccgaaggattggagggg
L E C E R Y L A P K G F G G
aatgaaatgttgcaattacaacccttcagaccttggtggg
N E N V A I Y N P F R P W W
agctataaattatgcacaagatctggaaatgaagatgaattta
S Y K L C T R S G N E D E F
tgtaacaatgttgggttcgtattatgttgatgctgaatta
C N N V G V R I Y V D A V I

Fig 3; De nucleotiden volgorde (kleine letters) is vertaald in de aminozuur volgorde (hoofdletters).

Op het internet zijn vele (gratis) programma's voor het vertalen van de nucleotiden volgorde. Hier hoeft slechts de nucleotiden volgorde te worden ingevoerd om vervolgens de aminozuur volgorde te kunnen bekijken.

Hoe decodeer je een dier ?

Veruit de meeste genetische informatie die verkrijgbaar is, heeft betrekking op de medische wetenschap. Op het web is daarom ook van de mens de meeste informatie beschikbaar, gevolgd door de muis en rat, omdat deze dieren veel in medisch onderzoek worden gebruikt als proefdieren. Voor meer fundamenteel onderzoek worden andere diersoorten gebruikt en elke onderzoeksrichting heeft een favoriete diersoort. De ontwikkelingsbiologie maakt bijvoorbeeld veel gebruik van zebrafish, de genetica van het fruitvliegje enzovoorts. Zodra er genoeg mensen geïnteresseerd zijn in dezelfde diersoort kan er gezamenlijk worden besloten een deel of het gehele genoom van dit dier te sequencen en op het web te zetten. Verschillende laboratoria over de hele wereld nemen ieder een stuk van het genoom voor hun rekening. Belangrijk hierbij is natuurlijk dat niet hetzelfde deel van het genoom twee keer gesequenced wordt en een ander deel niet. De stukken genetische informatie worden in een database gezet. Vervolgens voorspellen computer algoritmes op basis van vertalingen van de genen en vergelijkingen van deze genen met informatie die in andere databases beschikbaar is, welk gen het is. Ook de locaties van de genen op de chromosomen worden in kaart gebracht.

Zoeken naar de bekende weg...

Hoe maakt men nu gebruik van de genetische informatie die op het web bekend is? Een van de voornaamste methoden is het vergelijken van de genetische informatie van verschillende diersoorten. Diersoorten zijn in meer of mindere mate evolutionair aan elkaar verwant. Hiervan wordt gebruikt gemaakt in medisch onderzoek.

• • • Popular genomes • • •



Homo sapiens
NCBI 36 | Vega



Mus musculus
NCBI m36 | Vega



Danio rerio
Zv6 | Vega



New Chicken assembly and genebuild (Gallus gallus)

Fig 4. Een deel van de Ensemble Genome Browser homepage. Hier kan heel gemakkelijk op verschillende manieren worden gezocht naar genetische informatie van diersoorten.



Als een bepaalde stof voor een mens giftig is en voor een muis niet, dan er kan bijvoorbeeld een verschil zijn in het afbreekmechanisme voor deze stof. Als het gen dat codeert voor het enzym dat verantwoordelijk is voor de afbraak bekend is dan is onderling vergelijk mogelijk. Een enkel aminozuur verschil kan de katalytische eigenschappen behoorlijk verhogen of verlagen en een oorzaak zijn voor de verschillende respons op het gif. Dezelfde methode kan ook worden gebruikt om mensen onderling te vergelijken en kan soms de genetische basis van een ziekte aantonen. Hiermee kan dan een oorzaak worden gevonden voor een ziekte door de mutatie te vinden die het verschil verklaart. Het is echter nog niet of slechts gedeeltelijk mogelijk om een gevonden mutatie weer ongedaan te maken en daarmee een ziekte te genezen. Wel zijn andere behandelmethoden vaak mogelijk, door bijvoorbeeld een medicijn toe te dienen dat de omzetting van een bepaalde stof mogelijk maakt als het gemuteerde enzym dat niet meer kan.

Maar wat als de genetische informatie van het gekozen onderzochtsdier nog niet bekend is? Ook dan is de informatie op het web zeer nuttig, alleen moet deze anders gebruikt worden. In dit geval zal eerst de sequentie van het onbekende gen bepaald moeten worden, alvorens het vergeleken kan worden met andere diersoorten. Echter, het vinden van dit gen kan vergemakkelijkt worden, door hetzelfde gen in andere (evolutionair verwante) soorten te vergelijken en op basis daarvan een voorspelling te maken van hoe het gen er ongeveer uit zal zien. Dit wordt gedaan met een “alignment”. Het principe laat zich vergelijken met het zoeken van een woord in een vreemde taal. Als je de vertaling van een woord niet weet in het Portugees, kun je een gerichte gok doen als je de vertalingen van het Latijn, Spaans en Italiaans vergelijkt. Met behulp van een alignment is gericht zoeken naar het gen mogelijk en als het gen gevonden is kun je het vervolgens vergelijken met de al bekende genen om een voorspelling te doen over de structuur en functie van dit gen.

Zoeken naar een speld in een hooiberg....

Al deze zoekstructuren en vergelijkingen zijn natuurlijk bijzonder handig, maar wat als de functie van het gen nog onbekend is? Hoe kan bepaald worden, welke genen betrokken zijn bij een bepaald proces of ziekte? Hier kan men bijvoorbeeld gebruik maken van een “micro-array”. Dit is een soort chip waarop tienduizenden stukjes gen worden geplakt die door toeval worden gekozen uit DNA van controle materiaal (gezond persoon) en test materiaal (ziek persoon). Vervolgens wordt deze chip gehybridiseerd met controle materiaal dat groen gelabeld is en test mate-

Zebravis LVPHTP-TVAPYLTSPVCSVLELIEPEVVFAGYDNTQPDTTDHLTSLNQLAGKQMIRVVVKWAKVLPGFRSLPIEDQI
Stekelbaars LVPHGGPGGVAPFLPPSICSVLELIEPEEVVSYGYDNTRPDTTDHLSSLNRLAGKQVMRMVKWAKVLPGFRGLPIEDQI
Muis LVPQLAS-ITRALTPSPSMILENIEPEIVYAGYDNSKPDTAESLLSTLNRLAGKQMIQVVVKWAKVLPGFKNLPLEDQI
Mens LVPQLST-ISRALTPSPVMVLENIEPEIVYAGYDSKPDTAENLLSTLNRLAGKQMIQVVVKWAKVLPGFKNLPLEDQI
Klauwpad LVPLISA-VSPAITLSAAVILENIEPEIVYAGYDNTQPDTAENLLSSLNQLAGKQMVQVVKWAKVIPGFRNLPLEDQI

Fig. 5 Een stuk van een alignment. Hier is hetzelfde gen van verschillende diersoorten met elkaar vergeleken. Rood betekent gelijke aminozuren in alle soorten (*). Uitwisseling voor een ander aminozuur wordt in verschillende kleuren weergegeven; groen betekent een vergelijkbaar aminozuur (:), blauw een minder vergelijkbaar aminozuur (.) en zwart een totaal ander soort aminozuur. Als er aminozuren missen in een bepaalde soort wordt dit aangegeven met een streepje (-).

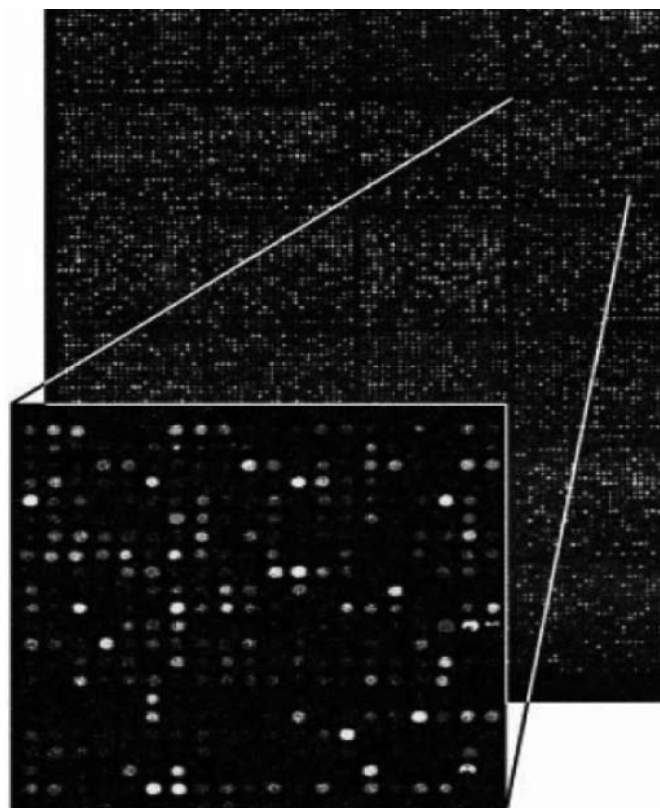
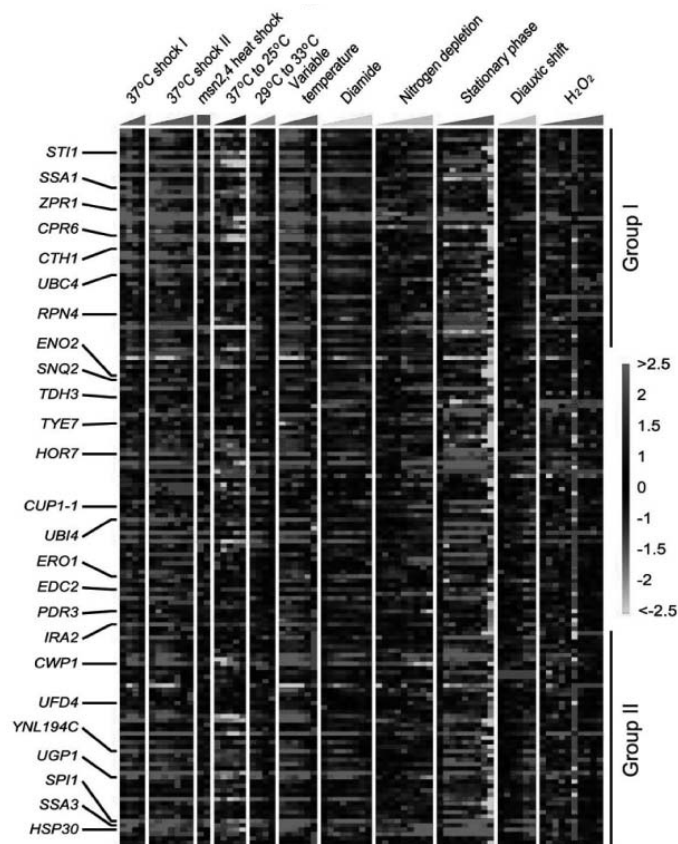


Fig 6 Micro-array slide met tienduizenden (stukjes van) genen die geprint zijn op een glazen plaatje.

riaal dat rood gelabeld is. Genen die aangeschakeld zijn in een controle persoon zullen binden aan de stukjes gen op de chip die corresponderen met datzelfde gen. Zo zullen de genen die alleen aanstaan in een gezond persoon groen worden, genen die alleen aanstaan in een ziek persoon rood, en genen die bij beide groepen aanstaan geel.

Met analyse software worden vervolgens de kleuren van de stipjes op de chip gemarkeerd. Deze stippen corresponderen met bepaalde stukjes gen en de sequentie van de stukjes wordt in een zoekmachine (bijv Ensembl) gestopt om te kijken of dit gen bekend is. Zelfs als het gen nog niet bekend is kan vaak wel worden bekeken wat de sequentie van het gehele gen is, of het op een gen in een andere diersoort lijkt en waar het gen op het chromosoom is gelokaliseerd.

De analyse wordt meer gecompliceerd als er niet alleen een test en controle situatie worden gebruikt, maar ook verschillende stadia en/of gradaties van een ziekte. Hiervoor is speciale software ontwikkeld die uit de dui-



zenden puntjes selecteert welke genen echt aan of uit staan (geel, lichtgroen, of echt groen), en vormt clusters van genen die een zelfde patroon hebben. Vervolgens kunnen de genen uitgezocht worden die de meest waarschijnlijke kandidaat zijn voor het ontstaan van een ziekte. microarray.icmb.utexas.edu/hsf/figure4.shtml

Patent op een gen ?

Als het gen is gevonden dat verantwoordelijk blijkt te zijn voor het ontstaan van een ziekte, kan het tegenwoordig gepatenteerd worden.

Dit is het geval bij het BCRA (BReast CAncer) gen. Dit gen vergroot de kans bij draagsters op het ontwikkelen van borstkanker tot 85% gedurende hun leven. Ook de kans op ovarium- en eileiderkanker is bij draagsters verhoogd. Het bedrijf Myriad heeft patenten gekregen voor twee genen, BRCA-1 en BRCA-2, die gebruikt kunnen worden voor kanker "screening testen". Myriad heeft in de VS een aantal laboratoria de rechten gegeven om deze tests uit te voeren, waarbij voor elke test royalty's moeten worden betaald aan het bedrijf. Onderzoeksinstituten die door de federale overheid zijn ingesteld krijgen korting, maar moeten ook nog steeds betalen. Op dit moment probeert Myriad ook Europese en Canadese patenten te krijgen voor de genen. Echter, hier is

grote controverse over ontstaan, omdat de prijs voor kankeronderzoek en diagnostische testen hierdoor aanzienlijk zou stijgen. Voorlopig heeft de Europese Patent Office beslist dat het principe van algemene beschikbaarheid van zorg sterker moet wegen dan commerciële belangen.

Servers, databases en software packages

In bovenstaand verhaal wordt een aantal maal gerefereerd naar de beschikbaarheid van databases en programma's op het web. Hieronder worden er een aantal meer uitgebreid belicht.

Browsers en servers:

Ensembl genome browser, Ensembl is een gezamenlijk project van EMBL, het European Bioinformatics Institute (EBI) en het Wellcome Trust Sanger Institute (WTSI) om een software systeem te ontwikkelen dat automatisch genen herkent en benoemt (annoteert) voor geselecteerde diersoorten. Via EMBL kunnen onderzoekers genen die gevonden zijn toevoegen aan de database. Het wordt voornamelijk gefinancierd door de Wellcome Trust.

<http://www.ensembl.org> <http://www.ebi.ac.uk/ensembl/>

ExPASy (Expert Protein Analysis System) proteomics server van het Swiss Institute of Bioinformatics (SIB). Deze server kan worden gebruikt voor de analyse van eiwit sequenties en structuren. <http://us.expasy.org/>

NCBI National Centre for Biotechnology Information. In 1988 opgezet als een nationale databank voor moleculair-biologische en biomedische informatie. Met als zoekterm een naam van een gen zijn zowel de genetische sequentie als ook alle wetenschappelijke artikelen over dat gen te vinden. <http://www.ncbi.nlm.nih.gov/>

Gratis programma's die op bovenstaande servers te vinden zijn:

Vertaalprogramma's; deze programma's zetten de DNA sequentie om in aminozuur volgorde.

<http://us.expasy.org/tools/dna.html>

ClustalW; Hiermee wordt bepaald hoeveel twee of meerdere sequenties op elkaar lijken en welke regio's het meest gelijkend zijn. Hiermee kan een evolutionaire (phylogenetische) stamboom worden berekend. <http://align.genome.jp/>

BLAST; Basic Local Alignment Search Tool Dit programma vergelijkt nucleotide of aminozuur sequenties met sequentie databases en berekent statistische significantie van de overeenkomst. BLAST kan worden gebruikt om evolutionaire verwantschappen te ontdekken en leden van bepaalde gen families te herkennen.

<http://www.ncbi.nlm.nih.gov/BLAST/>