

Algemene informatie

Het aardappelgenoom

Elk organisme heeft een genoom, een chemisch 'instructieboek' of 'blauwdruk', dat beschrijft hoe alle genen bij elkaar moeten worden gezet. Dit wordt uitgeschreven als een DNA-sequentie, een lange zin die bestaat uit de chemische letters A, C, T en G. Zo'n sequentie bevat vele tienduizenden genen, die kunnen worden beschouwd als 'woorden' in de zin. Elk gen stuurt verschillende aspecten aan van de manier waarop het organisme groeit en zich ontwikkelt. Kleine veranderingen in deze instructie resulteren in verschillende variëteiten – elk met een iets andere versie van de DNA-sequentie voor de soort.

Inzicht in de volledige genoomsequentie, de precieze spelling van de DNA-letters, van de aardappel zal wetenschappers beter inzicht verschaffen in de manier waarop de aardappel groeit en zich ontwikkelt. En dit zal wereldwijd leiden tot verbeterde gewassen.

Elk exemplaar van het aardappelgenoom bestaat uit 12 chromosomen en heeft een lengte van ongeveer 840 miljoen basenparen, waarmee het een plantengenoom is van gemiddelde grootte.

Een hoogwaardige, goed geannoteerde genoomsequentie van de aardappel is een waardevolle basis die kan worden gecombineerd met bestaande aardappelgenetische kennis en de voortdurende vooruitgang in het onderzoek naar de genen die worden aan- of uitgeschakeld en welke chemische stoffen wanneer worden geproduceerd. Waarneming van de manier waarop deze veranderingen worden beïnvloed door veranderingen in het genoom zullen onderzoekers in staat stellen verschillende genenvarianten te identificeren die belangrijke kwantitatieve eigenschappen van de aardappel bepalen.

Het Potato Genome Sequencing Consortium (PGSC, een wereldwijd consortium voor de sequentiebepaling van aardappel gestart en gecoördineerd door een team wetenschappers van de Plant Breeding groep van Wageningen UR) wil in de nabije toekomst de aardappelonderzoeks- en veredelingswereld de beschikking geven over een dergelijk instrument. Daarmee wordt die in staat wordt gesteld het volledige potentieel van de op biotechnologie gebaseerde verbetering van dit belangrijke gewas te benutten.

Het PGSC streeft er naar de volledige genoomsequentie van aardappel in kaart te brengen in 2010 'gelezen' te hebben.

Over de aardappel

De aardappel maakt deel uit van de *Solanaceae*, een plantenfamilie die verschillende andere economisch belangrijke soorten omvat zoals tomaat, aubergine, petunia, tabak en verschillende pepersoorten. De aardappel is wereldwijd een belangrijke voedselbron. De aardappel is het op twee na belangrijkste voedselgewas, na tarwe en rijst, met een wereldwijde productie van 309 miljoen ton in 2007. Geschat wordt dat wereldwijd in 2020 meer dan twee miljard mensen afhankelijk zullen zijn van de aardappel voor voeding, voor diervoeding of voor inkomen. Verbetering van aardappelrassen zodat ze beter bestand zijn tegen omgevingsfactoren, zoals droogte, ziekten en plagen, zijn kerndoelstellingen van wereldwijde aardappelveredelingsprogramma's.

De genetische diversiteit van de aardappel behoort tot de breedste van alle cultuurplanten. Wilde aardappelsoorten zijn zeer wijd verbreid op het Amerikaanse continent, van het zuidwesten van de Verenigde Staten tot Zuid Chili en Argentinië en van zeeniveau tot de hooglanden van het Andes-gebergte. Veel wilde soorten kunnen direct worden gekruist met de gewone aardappel en bezitten een brede reeks waardevolle eigenschappen, zoals resistentie tegen ziekten en plagen en vorst- en droogtetolerantie. Daarmee vormen ze een nuttige bron bij het kweken van nieuwe rassen.

Wereldwijd wordt het financiële verlies dat door ziekten zoals aardappelziekte in aardappelgewassen wordt veroorzaakt geschat op zo'n € 3 miljard per jaar. Deze ziekten worden nu nog voornamelijk bestreden door frequente toepassing van fungiciden. *Verwacht wordt dat een van de eerste positieve resultaten van de kennis van de aardappelgenoomsequentie een belangrijke doorbraak zal zijn in ons vermogen de genen te karakteriseren en om die genen te selecteren die betrokken zijn bij de ziekteresistentie.*

Maar de aardappel heeft, net als de mens, twee licht verschillende, exemplaren van het genoom (het is polyploid). Het ene exemplaar van de genoomsequentie is afkomstig van de moederplant en het andere, licht afwijkende, van de vader. Dit maakt bestudering van de aardappelgenetica gecompliceerd en over veel belangrijke eigenschappen bestaat veel onduidelijkheid. Toch is inzicht in de genetische samenstelling een eerste vereiste bij de ontwikkeling van efficiëntere veredelingsmethoden. De aardappelgenoomsequentie zal een belangrijke stap voorwaarts zijn bij het verkrijgen van beter inzicht in de manier waarop aardappeleigenschappen gekoppeld zijn aan genen, ter ondersteuning van toekomstige veredelingsprogramma's. Momenteel duurt het ontwikkelen van een nieuw ras zonder gebruik te maken van genomische aardappelveredeling zo'n 10-

12 jaar. Aangenomen wordt dat de mogelijkheid om gebruik te maken van genomische informatie niet alleen de tijd die nodig is om nieuwe rassen te ontwikkelen drastisch zal bekorten maar ook de kosten zal verlagen.

Potato Genome Sequencing Consortium

Het international Potato Genome Sequencing Consortium (PGSC, internationaal consortium voor bepaling van de aardappelgenoomsequentie) is een samenwerkingsverband van 16 onderzoeksgroepen in 14 landen: Argentinië, Brazilië, China, Chili, Indië, Ierland, Italië Nederland, Nieuw Zeeland, Peru, Polen, Rusland, het Verenigd Koninkrijk en de Verenigde Staten. Het PGSC vindt z'n basis in lange-termijn onderzoek naar de moleculaire genetica van aardappelen door de partner-organisaties, waaronder partners die wereldwijd een leidende positie innemen op het gebied van de sequentiebepaling van genomen en rekenkundige analyses.

Elke partner zorgt zelf voor de financiering die nodig is om onafhankelijk aan het project deel te kunnen nemen, meestal door subsidies van overheden en industriële organisaties.

Technische details

Het PGSC bepaalt de sequentie van twee genotypes:

RH89-039-16 (RH), een diploïde, heterozygote aardappel

DM1-3 516R44 (DM), een dubbele monoploïd.

RH89-039-16

Het PGSC is oorspronkelijk begonnen met de sequentiebepaling van het RH-genotype. Dit onderdeel van het project is gebaseerd op een diploïde aardappelgenomische BAC-klonenbibliotheek (bacterieel kunstmatig chromosoom) van 78 000 klonen, die zijn 'gefingerprint' en gerangschikt in ~7000 fysische 'map contigs'. Verder is de sequentie van de BAC-uiteinden bepaald; deze zijn openbaar. Er zijn ongeveer 30 000 BACs verankerd in de Ultra High Density genetische kaart van de aardappel, bestaande uit 10 000 unieke AFLPTM merkers. De BAC-bibliotheek en de fysische map is gemaakt en ter beschikking gesteld aan het consortium door de Plant Breeding groep van Wageningen UR.

Uit deze geïntegreerde genetisch-fysische map zijn momenteel 50 tot 150 zaad-BACs geïdentificeerd voor elk chromosoom. Fluorescerende in situ hybridiseringstests met geselecteerde BAC-klonen bevestigen deze ankerpunten. De zaadklonen vormen het uitgangspunt voor een 'BAC-by-BAC' sequentiestrategie. Deze strategie wordt aangevuld met 'whole genome shotgun' sequentiebenaderingen met gebruikmaking van zowel 454 GS FLX als Illumina GA2 instrumentaria. Voor de samenvoeging en annotatie van de sequentiegegevens zal gebruik worden gemaakt van vrij beschikbare en van speciaal ontwikkelde instrumentaria. De beschikbaarheid van de geannoteerde gegevens zal bijdragen aan de karakterisering van germplasma-collecties op basis van allelische variantie en zal aardappelveredelaars helpen bij het volledig benutten van het genetische potentieel van de aardappel.

DM1-3 516R44

Met de sequentiebepaling van DM werd begonnen om het proces van 'whole genome shotgun sequencing' te versnellen. Verwacht wordt dat 'whole genome shotgun' sequentiebepaling van DM1-3 516R44 (CIP801092), een dubbel monoploïde aardappelkloon, de complexiteit bij de bepaling zal verminderen.

Sequentietechnologieën

Het PGSC heeft drie verschillende technologieën ingezet om de genomsequentie te bepalen.

Sanger sequentiebepaling – de traditionele 'one read at a time' technologie, die werd gebruikt voor bepaling van de sequentie van het menselijk genoom

Solexa sequentiebepaling- een 'Next Generation Sequencing' platform waarbij tegelijkertijd miljoenen stukjes van het genoom kunnen worden gelezen; hun plaats in het genoom is echter niet bekend.

454 sequentiebepaling – een ander 'Next Generation Sequencing' platform waarbij tegelijkertijd honderdduizenden sequenties kunnen worden gelezen.

Sequentiebepaling in lekentermen

Wat is een genomsequentie?

Een genoom is de volledige set DNA-letters die beschrijft hoe een organisme is opgebouwd. DNA is een lange keten van vier verschillende chemische stoffen: adenine, cytosine, guanine en thymine. Dit vier-letter alfabet is bepalend voor de instructies, of genen, die bepalen hoe een organisme, zoals een plant of een persoon, werkt. Deze instructies zijn vele duizenden letters (bekend als basen) lang en zijn ingebed in een bibliotheek die honderden miljoenen basen lang is. Geschat wordt dat de aardappel 840 miljoen basen bevat, ongeveer een zesde van de grootte van het menselijk genoom dat 5 miljard basen bevat. Bij genoomsequentiebepaling wordt de precieze volgorde van alle basen bepaald zodat we het hele genoom kunnen uitschrijven. Een complicerende factor is dat bij de aardappel, net als bij de mens, het ene exemplaar van het genoom afkomstig is van de ene ouder en het andere exemplaar van de andere ouder. Deze verschillen enigszins en verschillende combinaties van deze verschillen zijn verantwoordelijk voor de verschillen die we zien tussen aardappelrassen, net zoals mensen van elkaar verschillen en van hun ouders.

Hoe werd de sequentie van het genoom bepaald?

Bepaling van de precieze sequentie is een grote uitdaging. Zelfs met de best beschikbare technologieën kunnen we maar een paar honderd letters per keer lezen; dit betekent dat we de sequentie van kleine stukjes moeten vinden en die dan moeten samenvoegen, zoals we de complete werken van Shakespeare met een paar woorden per keer lezen. De oorspronkelijk gebruikte strategie bestond uit het willekeurig opbreken van de genoomsequentie in kleinere stukken van zo'n 100.000 basen. Deze kunnen worden gescheiden en opgekweekt in bacteria als BAC's (bacteriële kunstmatige chromosomen). Als we bij onze Shakespeare-analogie blijven is dit het lezen van een bladzijde per keer. We kunnen zorgvuldig de vingerafdrukken bepalen ('fingerprinting') om ze aan elkaar te koppelen en om uit te werken welke van deze stukken samengaan met andere, net zoals we bij een legpuzzel alle randstukjes of de stukjes met lucht bij elkaar zoeken en op een hoopje leggen. Of het zoeken naar kernwoorden, zoals de naam van de personen, om uit te vinden over welk stuk van Shakespeare de bladzijde gaat die we lezen. Op dit punt hebben we nog niet de volgorde, we weten alleen maar welke delen samen gaan en, soms maar niet altijd, min of meer de volgorde waarin ze zich in het genoom bevinden.

De sequentie van elk van deze kleinere stukjes kan dan per keer worden bepaald. Ze zijn nog te groot om in een keer te lezen zodat ze in vele kleine stukjes

worden opgebroken die individueel kunnen worden gelezen. Elk van deze kleine sequentielezingen kan dan worden vergeleken met elk van de andere en waar de 'woorden' in de ene identiek zijn aan de 'woorden' in een andere, kunnen ze worden samengevoegd om langere zinnen te vormen, en daarna tot alinea's, die 'contigs' worden genoemd. Deze techniek staat bekend als 'shotgun sequencing'. Met voldoende kleine sequentielezingen kan de volledige sequentie van elk van deze BAC's worden bepaald en dan worden samengevoegd om de uiteindelijke genomesequentie te vormen. De sequentie van het menselijk genoom werd op deze manier bepaald.

Deze benadering heeft nadelen. Elk minuscuul fragment van een sequentie moet een voor een worden gelezen; een tijdrovende en kostbare benadering. En je wilt ook niet de sequentie van een grote hoeveelheid BAC's bepalen die hetzelfde deel van het genoom dekken, want op die manier zouden dezelfde delen steeds maar weer opnieuw worden gelezen – verspilling van tijd en geld.

Enkele jaren geleden kwam een belangrijke nieuwe set sequentie-technologieën beschikbaar. Deze lasen niet de sequentie van een stukje DNA per keer maar ze konden tegelijkertijd de sequentie van miljoenen stukjes DNA bepalen. In plaats van het maken van tienduizenden aparte BACs, kon het hele genoom in één keer worden gelezen - een benadering die 'whole genome shotgun' werd genoemd. Dit gaat niet zonder problemen. Hoewel we veel individuele korte sequenties kunnen lezen, en in het geval van de aardappel zijn er meer dan een miljard korte sequenties gelezen, is het samenstellen van de legpuzzel erg lastig. En het wordt nog moeilijker als zich twee licht verschillende (ongeveer een letter van elke 60-100) versies in het mengsel bevinden, zodat een speciale onderzoeksstam van de aardappel werd geïdentificeerd die maar één versie van het genoom heeft. Dit is de variëteit waarvan nu de sequentie is bepaald.

Voor het samenstellen van deze enorme legpuzzel waren een grote computer en nieuwe computerprogramma's nodig. De aardappel is een van de grootste plantengenomen waarvan de sequentie op deze manier is bepaald. In de eerste stadia werden sequentieoverlappingsen gebruikt om korte 'contigs' te creëren. Deze kunnen worden samengevoegd door gebruik te maken van andere informatie, zoals kennis van het soort 'read' dat zich aan tegenovergestelde kanten van hetzelfde DNA-fragment bevindt (we kunnen tegelijkertijd slechts de sequentie van één eind van een stukje DNA lezen) en met gebruikmaking van heel veel informatie uit de grote collectie BAC's waarvan de vingerafdruk al is

bepaald samen met een deel van de sequentie. Deze vormen dan grotere platforms met gebieden waarvan we nu de sequentie kennen, maar ook met gaten waarvan we de sequentie niet kennen, maar waarvan we wel weten hoe groot ze zijn.

Als de delen een grootte hebben waarvan we weten dat deze een sequentie van een bruikbare grootte hebben, normaal gesproken als ze de volledige instructies voor een individueel gen bevatten, zijn ze bruikbaar voor de rest van de wetenschappelijke wereld en worden ze als ontwerp-sequentie vrijgegeven. Dat lijkt veel op een legpuzzel waarvan we de belangrijkste stukken hebben samengesteld en waarvan we het algemene plaatje kunnen herkennen maar waarvan er nog vele lastige stukjes moeten worden geplaatst om de details van de gaten te vullen. De aardappelsequentiebepaling bevindt zich in september 2009 in dit stadium, waarbij het grootste deel van het genoom er is in vele honderden grote stukken en we schatten dat 95% van de genen hun complete sequentie hebben in één sequentiestuk.

De volgende stap is verbetering van de concept-samenstelling. We zijn druk bezig met het samenvoegen van contigs en platforms en met het uitwerken van de volgorde waarin ze in het genoom zitten. Er zullen enkele delen zijn (hopelijk zeer weinig) waarbij we kleine foutjes hebben gemaakt bij het samenvoegen. De uitgebreide analyse waaraan we nu beginnen zal ons helpen deze te onderkennen en te corrigeren zodat, als de genoomsequentie rond eind 2009 netjes wordt gepubliceerd, de wetenschappelijke wereld daarin vol vertrouwen zal kunnen hebben. We maken het ontwerp vroeg beschikbaar omdat er al veel informatie is die door andere onderzoekers kan worden gebruikt.

Wat doe je als je de sequentie hebt?

Het verkrijgen van de aardappelsequentie is slechts het begin. We moeten nu bepalen waar alle genen, de sets instructies voor hoe de plant groeit, zich bevinden. We moeten uitzoeken wat ze doen, gebruik makend van geavanceerde computerprogramma's om hun functie te voorspellen en van individuele onderzoekers die deze genen in het lab testen. En dan moeten we kijken hoe deze tussen rassen verschillen door relevante gebieden te bekijken en na te gaan hoe die korte sequenties eruit zien in andere rassen. Belangrijke onderzoeksterreinen zijn de identificatie van genen die de voedingskwaliteit van

de aardappel beïnvloeden, de resistentie tegen ziekten en plagen, met name het aardappelcystenaaltje in het Verenigd Koninkrijk en de aardappelziekte *Phytophthora infestans*. Belangrijk voor de landbouw zijn genen die aardappelen in staat stellen droogte te verdragen en nog een groot aantal andere eigenschappen. Beschikbaarheid van het genoom is nog maar een begin om het complete genetische instructieboek voor de aardappel te kunnen lezen.

Welke voordelen brengt kennis van de genoomsequentie met zich mee?

Alle bovengenoemde eigenschappen kunnen worden ontwikkeld zonder de genoomsequentie te kennen, maar het zou lang duren en het zou zeer kostbaar zijn, met veel vallen en opstaan. Als we de genetische vingerafdruk van de gewenste eigenschappen kennen, dan kunnen deze al heel vroeg in de veredelingsprogramma's worden geselecteerd. Daarmee kunnen vele jaren worden bespaard op de huidige ontwikkelingscyclus van 10-12 jaar terwijl ook minder middelen nodig zijn voor de ontwikkeling van deze nieuwe stammen. Deze genetische selectiebenadering is zeer veelbelovend en de technologie om direct de genoomsequentie te benutten wordt al gereed gemaakt in het Verenigd Koninkrijk en in Nederland.

Inzicht in de genetische blauwdruk van de aardappel maakt het ook mogelijk gewassen aan te passen om zo specifieke eigenschappen te modificeren. De commerciële vooruitzichten voor de toekomstige ontwikkeling van genetisch gemodificeerde aardappelen zijn op dit moment nog onduidelijk.