



Hvorfor DNA-sekventering og kartofler er vigtig

- for at brødføde verdens befolkning i fremtiden

Genomik, big data og statistisk prædiktion i planteforædling.

Af Kåre Lehmann Nielsen og Elsa Sverrisdóttir,
Institut for kemi og biovidenskab, Aalborg Universitet

Udvikling af nye og bedre afgrøder er essentiel for at brødføde de ni mia. mennesker, der vil være på planeten i 2050. Men planteforædling er omkostningsfyldt, besværligt og langsomt, fordi sandsynligheden for at kombinere de genvarianter, der betyder noget for vigtige agronomiske træk, er ufatteligt lille. Derfor planter, gror og evaluerer man ofte flere hundrede tusinde planter for blot at finde en enkelt, som er marginalt bedre end eksisterende afgrøder. Kortlægning af genvarianter ved DNA-sekventering og statistiske modeller som kan forudsige ydeevnen af planter ud fra generne, kan radikalt øge hastigheden, hvormed nye og bedre afgrøder kan udvikles. Fundamentet for disse muligheder er den ekstreme udvikling i DNA-sekventeringsteknologi, som er sket de sidste 10 år.

Udfordringen

De eksisterende sorter, som dyrkes på marken i dag, er hoved-

sageligt udviklet ved traditionel planteforædling. Det består let forsimplet i, at man tager to forældre, som man via observationer (fænotypebestemmelse) ved, har gode egenskaber, krydser dem, samler og planter frø herfra, og lader afkomspopulationen gro. Disse planter udsættes for omfattende fænotypebestemmelse typisk over flere år for forskellige agronomiske træk, hvorefter man udvælger de få afkom, som har bedre egenskaber end forældrene.

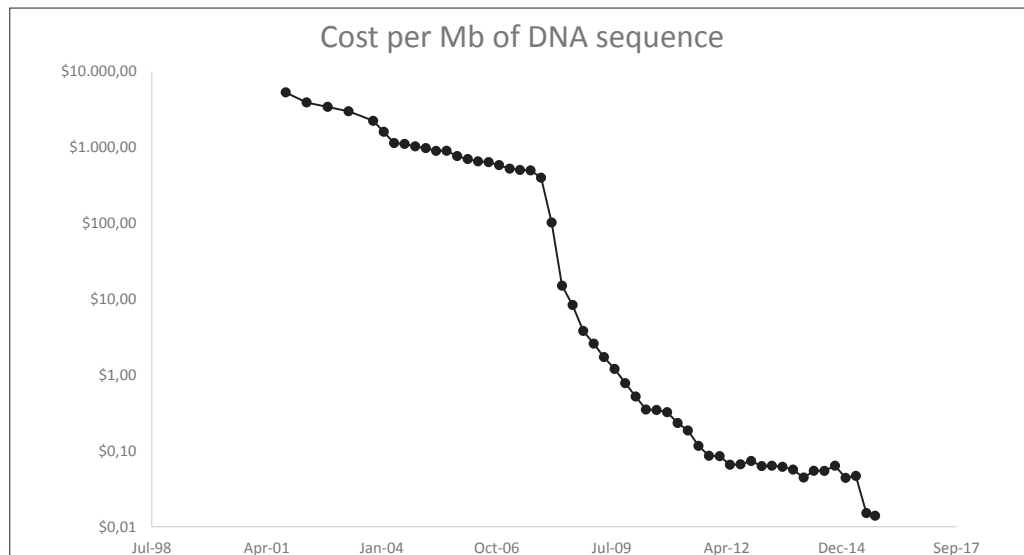
Princippet er jo ikke specielt udviklet, men processen er ufatteligt ineffektiv. Lad os i et tænkeeksperiment antage, at for en kompleks egenskab som udbytte (kompleks betyder, at flere gener påvirker egenskaben), er det ikke urealistisk, at 50 gener har en rimelig betydning for udbyttet. Hvis vi endvidere antager, at der i de to forældre tilsammen er for hvert af disse gener en genvariant, som påvirker trækket positivt og tre genvarianter, som ikke påvirker det nævneværdigt, så vil chancen for at få en enkelt optimal plante, som indeholder den positive genvariant for alle 50 gener være $0,5^{50}$. En ufatteligt lille chance. Det betyder, at man skal gro 10^{15} planter for at have en rimelig chance for at

få det optimale afkom af de to forældre. For kartofler, hvor en plante fylder ca. 0,25 m² ville et sådant forsøg fylde halvdelen af planeten Jorden (inklusive hav, land, is og søer). I realiteten, når man er et forædlingsfirma, der har kapacitet til at gro måske 250.000 planter om året, er der altså virkelig tale om et lottospil.

DNA-sekventering

Frederick Sanger og Walter Gilbert fik i 1980 Nobelprisen i kemi for metoder til at bestemme rækkefølgen af baserne i DNA, som de havde udviklet få år før. Dermed fik videnskaben mulighed for at afkode den genetiske, som er afgørende for egenskaberne (fænotyperne) for alt liv. Indtil ca. år 2005 blev

DNA-sekventering udført hovedsageligt på samme måde som Frederick Sanger gjorde i 1977. Men siden er det gået meget stærkt, figur 1. Hvor det i 2005 kostede ca. 1.000\$ at sekventere 1 mio. baser, koster det i dag 10 cents, altså 10.000 gange mindre. Der er virkelig tale om en revolution: Hvis samfundets



Figur 1. Prisen for sekventering af 1 Mbp fra 2001 til 2015. Bemærk Logaritmisk skala. Data fra: Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Tilgængelig her: www.genome.gov/sequencingcostsdata.

produktivitet havde forbedret sig med samme hastighed, kunne vi have reduceret vores arbejdsuge fra 37 timer i 2005 til 13 sekunder i 2015!

Det, der virkelig har betydet noget for effektiviteten af DNA-sekventering, er miniaturisering af reaktionen og auto-

Titration!
www.metrohm.dk



Metrohm
Nordic

BEDRE!

OMNIS

VI HÆVER STANDARDEN TIL ET NYT NIVEAU

MEGET HURTIGERE

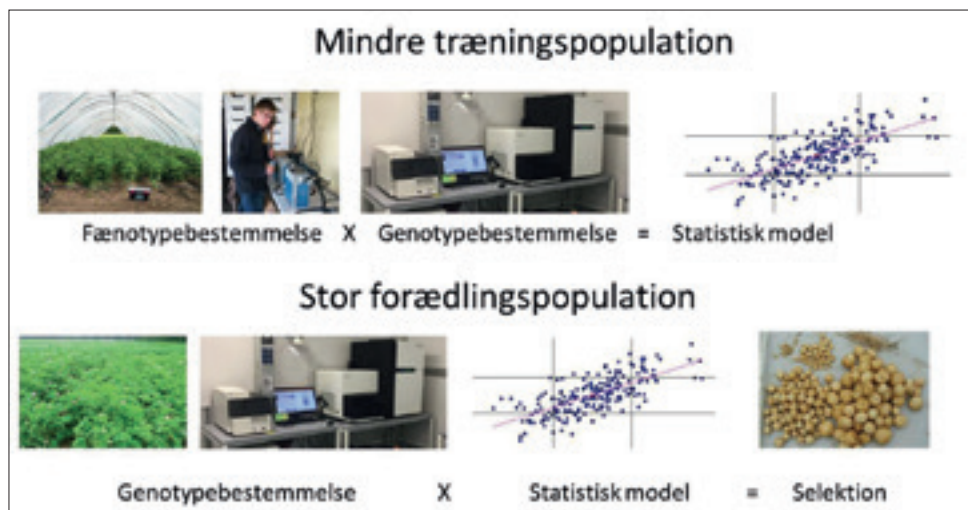
Udfører op til fire automatiske titreringer med det samme system på samme tid.

MERE SIKKERT

Ingen udsættelse for kemikalier længere som følge af kontaktfri udskiftning af reagens.

MEGET NEMMERE

En grafisk brugerflade & letforståelig brugervejledning gør OMNIS simpelt og sjovt at styre.



Figur 2. Forædlingsprogram ved brug af genomisk prædiktation. En mindre repræsentativ population af planter bliver både fænotype- og genotypebestemt. Disse data bruges til at bygge den statistiske model. Herefter anvendes den statistiske model på genotypedata på den meget større forædlingspopulation til at foretage selektion af de bedst egnede planter.

mativering af detektion. Hvor man i 1995 lavede Sanger-sekventering i fire plastkrør med efterfølgende gelelektroforese og eksponering og fremkaldelse af røntgenfilm for en enkelt reaktion, er reaktionen nu virkelig *high throughput*. Reaktionen foregår i 1 µm clustre af ca. 1.000 identiske DNA-molekyler kovalent bundet til en glasoverflade og detektion sker via lys og fluorescerende grupper kovalent koblet til de nukleotider, som udgør byggestenene i DNA. Der kan være rigtig mange (ca. 1.000 pr. mm²) af sådanne clustre på overfladen af et mikroskopobjektglas. En illustrativ video af teknologien fra sekventeringsfirmaet Illumina kan ses her (<http://www.illumina.com/techniques/sequencing/dna-sequencing.html>). Det bør retfærdigvis siges, at der findes andre sekventeringsteknologier på markedet, men Illumina har langt den største markedsandel. Kernereaktionen i teknologien er den samme, som Frederick Sanger udnyttede: DNA-syntese sker via en kopiering af en template (den komplementære DNA-streng) og begynder fra en primer (en lille stump DNA eller RNA). Enzymet DNA-polymerase katalyserer dannelsen af en fosfodiesterbinding mellem primerens 3'-OH gruppe og nukleotidets 5'-PO₄²⁻ under fraspaltning af pyrofosfat og vand. Kun det nukleotid, som er komplementær til templatestrengen og kan danne de rigtige Watson-Crick baseparringer, bliver påsat. I Illumina-sekventering udnyttes det, at 3'-OH gruppen er blokeret med en fluorescerende gruppe, som har forskellig farve for hver af de fire baser. Det bevirker, at reaktionen efter inkorporering af et nukleotid går i stå, hvorefter en laser og detektor kan aflæse farven af de fluorescerende grupper, der sidder på hvert cluster på overfladen, og man får dermed direkte information om DNA-sekvensen af clustrene. Herefter kløves den fluorescerende gruppe af efterladende en fri 3'-OH gruppe, som kan udnyttes til at køre cyklus igen. I dag kan man køre 250-500 cyklusser og på den måde få 250-500 basepar DNA-sekvensinformation fra hvert molekylecluster. Processen er uhyre effektiv: en enkelt kørsel på vores nye sekvenator i Aalborg, som er blevet købt for penge fra Poul Due Jensens fond genererer op til 1 trillion basepar DNA-sekvens.

Muligheder

Denne teknologirevolution betyder, at selvom alle højere livsformer indeholder rigtigt meget DNA (0,5-10 mia. basepar), er det nu teknologisk muligt og økonomisk realistisk at sekventere genomer helt eller delvist af hele populationer af planter.

Dermed kan man identificere den genetiske variation ved DNA-sekventering, genotypen, af de enkelte individer. Ved at bestemme både genotypen og fænotypen af en træningspopulation kan man identificere hvilke genvarianter, der påvirker fænotypen. Med denne information kan man lave prædiktive statistiske modeller, som kan forudsige, hvilke egenskaber et individ har i en population, hvor man ikke kender fænotypen allerede på frø eller spirestadiet. Frøet indeholder jo de samme gener, som den voksne plante, og med denne forudsigelse kan man på dette tidligere stadie kassere langt størstedelen af afkommet, som man derfor ikke behøver at bruge ressourcer på at gro og evaluere. Processen, som kaldes genomisk assisteret forædling, er illustreret i figur 2. Når

chancen for at finde det optimale afkom er mikroskopisk (jf. tænkeeksperimentet i begyndelsen af artiklen), er det centralt at bruge den begrænsede vækst og evalueringskapacitet, der findes hos forædlerfirmaerne til at gro planter, som faktisk har en rimelig chance for at være bedre end forældrene. Det er altså nemmere at få syv rigtige i lotto, hvis man på forhånd kender seks rigtige tal.

Kartofler

På Aalborg Universitet har vi efterhånden i en del år arbejdet sammen med den danske kartoffelindustri om at udvikle molekyllære og statistiske metoder til at understøtte genomisk assisteret forædling i det praktiske forædlingsarbejde. Arbejdet er støttet af både offentlige og private fonde. Men hvorfor kartofler? Udover at kartofler smager godt, kan kartofler producere ca. det dobbelte antal kalorier pr. m² end korn, majs og ris, og derfor er kartofler også "den fattige mands" afgrøde. Både historisk i Irland i første halvdel af 1800-tallet, og i dag, hvor dyrkning af kartofler er under stærk vækst i lavindkomstområder i Afrika og Asien. Hvis målet er at brødføde verdens befolkning, også i en fremtid hvor vi har 10 mia. munde at mætte hver dag, er udbytte pr. hektar en vigtig egenskab. Der er jo ikke mere regnskov at konvertere til landbrugsjord. Herudover har kartofler en særdeles indviklet genetik; i stedet for to kopier af hvert kromosom har de fire, og der er en enorm genetisk diversitet tilstede. Den genetiske diversitetsforskel mellem to elitekartoffelsorter er ca. seks gange større end forskellen mellem en elitetomat og de vildarter af tomat, der findes i naturen. Det betyder, at de genetiske kombinationsmuligheder er endnu større end i vores tænkeeksperiment, og dermed er chancen for at få en vinder endnu mindre. Blandt andet dette har forårsaget, at kartoffelforædlingen de sidste 50 år ikke har kunnet følge med andre afgrøder: hvor udbyttet af korn, majs og ris er blevet 250% forøget siden 1960, er udbyttet på kartoffelmarker kun forøget med 150%. Men det betyder også, at effektive prædiktive algoritmer baseret på genotyper har potentiale til at have en endnu større effekt i kartoffelforædling end i forædlingen af andre afgrøder, simpelthen fordi kartofflen i mindre grad har optimeret sit genetiske potentiale i de sorter, vi har til rådighed i dag.

E-mail:

Kåre Lehmann Nielsen: kl@bio.aau.dk