

IMPACT del 2 - Udvikling af ny genotyping teknologi for genetisk diverse afgrøder som kartoffel - Slutrapport.

Ansvarlige:

Kåre Lehmann Nielsen, kln@bio.aau.dk, Aalborg Universitet, Sektion for bioteknologi
Katrine Amalie Hamborg Nielsen, Aalborg Universitet, Sektion for bioteknologi

Formål: Vedligeholdelse af MASPot populationen som genetisk ressource for fremtidige forskningsprojekter og forædling af kartofler.

Resume

Planteforædling er i gang med et betydeligt teknologiskifte, hvor man tidligere alene har baseret selektion på fænotype, er der stigende brug af molekylære (genetiske) markører, som kan anvendes til selektion på et meget tidligere stadie og dermed spare tid og penge. Det er blevet muligt at bruge mange markører på samme tid og derfor selektere på mange træk på én gang (Genome Assisted Selection). Centralt for dette teknologiskifte er molekylær biologiske genotypeteknologier som er nødvendige for at måle markørstatus af de individer som analyseres. I modsætning til andre store afgrøder er disse teknologier en udfordring for kartoffelforædlingen, fordi kartofflens komplekse genetik ikke tillader effektiv brug af de eksisterende geno-typeteknologier. Aktiviteter: Udvikling af en ny genotypeteknologi består af flg. faser og strækker sig over 2 år. Denne rapport omhandler begge år.

Projektet har været opdelt i 4 faser: 1) Udvikling af arrays og design af primere. 2) Proof of technology. 3) Benchmarking. 4) Anvendelsesfase.

Status: Fase 1-3 er næsten afsluttet (se nedenfor) og fase 4 i gang, men dog for nuværende på pause pga. to uforudsete begivenheder: i November 2019 havde Aalborg Universitet en fejlfunktion af et ildslukningsanlæg i det serverrum, hvor vores computerservere og forskningsdata lagres. Anlægget nærmest eksploderede og forårsagede altødelæggende fysisk skade på vores servere. Vigtige data er selvfølgelig redundante og kan genskabes. Dog pga. størrelsen af data > 300 Terrabyte og omfanget af ødelæggelsen af både array og servere og leverings- og samletider på komponenter var vores servere ude af drift indtil slutningen af januar. I den tid har vi ikke kunnet arbejde med generede data fra dette projekt. Da universiteterne blev lukket ned den 11/3 2020 pga. Corona situationen har vi færdiggjort analyserne af de første data, men laboratoriarbejdet vedr. en del af 3) og det meste af 4) er nu på pause indtil laboratorierne igen er tilgængelige. Vi vil derfor indsende en endelig slutrapport på projektet i sidste halvdel af 2020.

Aktiviteter:

Fase 1: Udvikling af arrays og design af primere. Et genome capture array med 2377 80-mer oligonukleotider er blevet designet ud fra flg. kriterier: a) På baggrund af Ea Høgh Riis Niensens (erhvervs PhD KMC, Danespo og AAU) arbejde med mapping af markører fra litteraturen til genomsekvensmodellen, samt DNA områder med relevans udpeget fra projekterne MASPOt og GenSAP blev 318 områder der er tæt på QTLs for agronomisk vigtige træk udvalgt. b) De 824 Mbp genomsekvens fra modelkartofflen som blev sekventeret i 2014 blev opdelt i 80 nt lange fragmenter som blev sorteret, således at de fragmenter som er mest unikke (dvs. ikke ligner andre dele af sekvensen) blev prioriteret.

Herefter blev der fra denne pool udvalgt tilfældige sekvenser således af de følger gendensiteten på genomet. Dvs. vi har tættest markørdækning der hvor generne hyppigst forekommer i enden af kromosomerne, hvor også chancen for overkrydsning er størst. c) Herefter blev sekvenserne filtreret for ekstremt højt eller lavt GC indhold og hvis sekundær struktur af oligonukleotider kunne forudsiges. De resulterende 2377 sekvenser blev sendt til MyBaits, Arbor Biosciences, USA, og et capture array blev fremstillet.

Primerne til barcoding af genomiske fragmenter fra individuelle prøver blev indkøbt således at op til 96 prøver kan analyseres på samme sekventering og bagefter kædes tilbage til de individuelle prøver.

Fase 2: Proof of technology. Oprensat genomisk DNA fra 2 individer med kendt genotype (forældre fra MASPot) blev fragmenteret og tagget med to forskellige indeksskombinationer og poolet. Dette vil blive oprenset med det bestilte capture array. Efter et betydeligt optimeringsarbejde med hybriserings- og amplificeringsbetingelserne har vi nu opnået at ca. 35% af alle opnåede sekvensreads kommer fra et af de forventede områder. Det svarer til en berigelse på ca. 1500X, hvilket betragtes som yderst tilfredsstillende og indenfor det økonomisk viable. Det svarer til en omkostning på ca. 100 kr pr. prøve.

Et custom bash script er udarbejdet, så prøverne kan analyseres automatisk ved brug af de software komponenter som vi allerede har udviklet i MASPot.

Fase 3: Benchmarking. 76 genomiske DNA prøver fra det valideringspanel, som blev brugt i både MASPot og GenSAP for hvilke vi har mange fænotype data, samt udviklet prediktionsalgoritmer til på baggrund af andre metoder er blevet oprenset og fragmenteret. Prøverne afventer nu sekventering.

Fase 4: Anvendelsesfase. Vi har modtaget 114 prøver fra Holland gennem et samarbejdsprojekt mellem Danespo, AAU og flere europæiske partnere, hvor vi anvender denne metode. Prøverne er endnu ikke analyseret, da vi modtog dem dagen før Corona lock-down.

Afledte projekter:

Metoden bliver anvendt i EU- ERA-net projektet Diffugat : <https://www.suscrop.eu/projects-first-call/diffugat>. AAU og Danespo er danske partnere i projektet.

Publikationer:

Projektet har endnu ikke givet publikationer, men vi forventer publikationer i løbet af 2020/21, hvor vi kan publicere på projekter hvor metoden er blevet anvendt.

Bilag:

Protocol for Impact genotyping – tagging, capture and amplification.

Kåre Lehmann Nielsen rev 02/2020

Input requirements: 50 ng highly purified gDNA A260/A280=1.8-2.0. (best option for quantification is qubit; however nanodrop required for ratio determination).

Abbreviations:

Qubit -Qubit 4 quick reference guide. ThermoFisher Scientific 2017.
NextProt -Nextera DNA library Prep reference guide (Document #15027987v01 – jan 2016) (Illumina)
MinElute -MinElute Reaction Cleanup kit using a microcentrifuge (Qiagen MinElute Handbook 01/2020)
HybProt -MyBaits hybridization capture for targeted NGS v4.01 April 2018
Tapestation -Agilent 2200 Tapestation system user manual 09/2015. Agilent technologies.

Per sample:

1. Follow NextProt tagmentation page 9
2. Following thermocycling end purify Rxns using MinElute.
3. Amplify the purified rxns using **a separate combination of barcodes** for each reaction for **8 cycles** according to the “Amplify Tagmented DNA “ section in NextProt page 12-14.
Note: Remember to note down which barcodes identifies which samples!
4. Pool 50% all reactions to a single tube
Note: store the remaining 50% in -20 freezer for backup.

Only a single pooled sample:

5. Purify Rxns using MinElute.
6. Quantify rxn using Qubit dsDNA HS assay, according to Qubit.
Note: Reference is not needed.
7. Analyze rxn using Tapestation HSD1000 following Tapestation page 56.
Note: use ladder if available.
8. Follow appendix A2 in HybProt page 15-16 using Phusion polymerase and P5reamp and P7 reamp primers.
9. Quantify rxn using Qubit dsDNA HS assay, according to Qubit.
Note: Reference is not needed.
10. Analyze rxn using Tapestation HSD1000 following Tapestation page 56.
11. Follow HybProt page 7-13 using Phusion polymerase – at page 12 split the rxn in three portions (5, 10, 10 μ l) and run 0, 8 and 14 cycles of PCR, respectively (3 reactions going into section 3.2).
12. Purify using MinElute.
13. Quantify rxn using Qubit dsDNA HS assay, according to Qubit.
Note: Reference is not needed.
14. Analyze rxn using Tapestation HSD1000 following Tapestation page 56.
Note: use ladder if available.
15. Determine A260/A260 by Nanodrop analysis.

Submit for sequencing using a single lane of HiSeq3000/4000 or similar throughput – about 300 mio reads are needed for an average of 250 reads pr site pr sample.