

# Statusrapport: Detektion og kvantificering af kartoffelcystnematode fra jordprøver direkte ved DNA sekventering

KAF projekt

Elsa Sverrisdóttir, Aalborg Universitet, 31-03-2020

## Projektformål og strategi

Formålet med projektet er at undersøge muligheden for at detektere, kvantificere og pathotypebestemme kartoffelcystenematoder (KCN) direkte fra jordprøver ved hjælp af sekventering.

Projektet består af fire dele:

- 1) Primerdesign
  - Primersæt for ca. 40 genetiske loci, der koder for de pathotype relevante effektorgener vil blive designet ud fra litteraturstudie
- 2) PCR test og optimering
  - De 40 primersæt vil blive testet og analyseret for eksperimentel robusthed ved at analysere jordprøverne samt positive og negative kontroller. Her vil også detektionsgrænse blive evalueret.
  - Den del af primersættene (ca. 20) som er mest robust vil blive brugt i de efterfølgende målinger.
- 3) Amplifikation og sekventering
  - Amplifikationsprodukterne vil blive en unik DNA sekvens, så prøven er identificerbar efter sekventering. Herefter pooles alle prøverne og sekventeres samtidig på Illumina MiSeq.
  - Sekvenssættene vil blive analyseret for polymorfier, samt ”teoretiske sekvenser” fra litteraturen.
  - Den del af primersættene som er mest informativ i forhold til at adskille pathotyper vil blive udvalgt til den endelige analyse.
- 4) Udvikling af en matematisk prædiktionsalgoritme, som er i stand til at bestemme pathotypen på baggrund af sekvenserne
  - En prædiktionsalgoritme af typen beslutningstræ der er i stand til at forudsige pathotyper på baggrund af de opnåede sekvenser vil blive udviklet. Beslutningstræet vil blive evalueret mod resultater fra de ”klassiske” analyser (Lars Bødker projekt).

## Status 31-03-2020

Projektet har været teknisk mere udfordrende end forventet. Derfor er denne rapport en statusrapport, for det er ikke lykkedes at færdiggøre hele projektet. AAU vil dog færdiggøre projektet for egen regning og en endelig slutrapport vil blive publiceret i løbet af sidste halvdel af 2020. Dog med det forbehold, at vi pga. Corona krisen ikke i øjeblikket arbejder på det projekt.

Status er at de første to delmål er opnået, dvs. primerdesign og PCR test. Næste trin er at fortsætte PCR på alle prøver med 11 udvalgte primersæt, efterfulgt af sekventering. Der er opnået resultater fra PCR delen af 3), men validering ved sekventering er endnu ikke opnået.

## Materialer og metoder

### Udvælgelse af effektorgener og primerdesign

Proteinsekvenserne af effektorgener listet i (Akker et al., 2016) og (Cotton et al., 2014) for hhv. *G. Pallida* og *G. Rostochiensis* blev brugt til at lave en alignment (default indstillinger: gap open cost 10, gap extension cost 1). Derefter blev der bygget træer over effektorgener (algorithm: neighbor joining, distance measure jukes-cantor, bootstrap: 100 replicates). Ud fra træerne blev der udvalgt et effektorgen for hver familie. 20 effektorgener blev udvalgt for *G. rostochiensis* mens 31 effektorgener blev udvalgt for *G. pallida*.

Sekvensreads for helgenomssekventering af *G. pallida* og *G. rostochiensis* blev downloaded fra <https://ewels.github.io/sra-explorer/> og importeret til CLC Genomics Workbench. Her blev de trimmet og mappet til referencegenomet for hhv. *G. pallida* og *G. rostochiensis*, hvorefter varianter blev kaldt (basic variant detection (minimum coverage: 10, minimum count: 2, minimum frequency: 35%, maximum coverage: 100,000)). Variant tracks blev samlet i en track liste sammen med referencegenom, gtrack og repeat-regions-track. For *G. rostochiensis* blev der samlet 11 variant tracks, for *G. pallida* blev der samlet 17 variant tracks. Derefter blev der kigget på hvert enkelt af de udvalgte effektorgener, og primere blev designet så de flankerer relevante områder. Relevante områder blev defineret som områder inden for de valgte effektorgener hvor der optræder variationer på tværs af de forskellige variant tracks – dvs. varianter, der evt. kan bruges til at adskille de forskellige pathotyper fra hinanden og identificere dem. Det var også vigtigt, at de relevante variationer lå udenfor repeat regions. Primerne blev lagt i områder uden varianter og udenfor repeat regions. Primerne blev designet, således at de resulterende amplicons har en størrelse på maksimum 600 bp.

Et yderligere primersæt blev designet for udvalgte effektorgener for at kunne lave nested PCR. Disse primere blev designet så de ligger omkring 50-400 bp før (forward primer) eller efter (reverse primer) de oprindelige primersæt. For nemheds skyld bliver disse primersæt fremover omtalt som de yderste primere eller de yderste primersæt, mens de inderste (nestede) primersæt bliver omtalt som de primære primere eller de primære primersæt.

### Jordprøver, kontroller og ekstraktion af DNA

17 prøver fra Lars Bødkers projekt blev indsamlet. Prøverne stammer fra kartoffelmarker rundt omkring i Danmark. DNA blev ekstraheret med DNeasy PowerSoil Pro kit fra Qiagen. Derudover blev 11 prøver af ekstraheret DNA fra nematoder indsamlet fra James Hutton Institute, UK.

### PCR test og optimering

Ekstraheret DNA fra jordprøver blev blandet til en pool for at teste primerne. I første omgang blev alle 17 prøver poollet 1:1, men senere blev der lavet en *pallida* pool, kun bestående af prøver testet positive for *G. pallida*, og i særlig grad dem med kraftig eller meget kraftig infektion. Da der kun var én jordprøve der testede positiv for *G. rostochiensis* blev denne også brugt alene i primertest. DNA fra nematodepopulationer blev også poollet på tilsvarende måde; én *pallida*-pool og én *rostochiensis*-pool. Hvert primersæt blev testet på jordprøve-DNA med PCR, hvor mastermixet bestod af 1x Taq buffer (Ampliqon), 2 mM MgCl<sub>2</sub>, 200 μM af hver dNTP, 0,5 μM hver primer og 3 U Taq DNA polymerase (Ampliqon). Som udgangspunkt blev der brugt 1 ng DNA (fra jordprøver) til hver reaktion, men det blev opjusteret til 4 ng DNA. Temperaturgradienter blev kørt på hvert primært primersæt med annealing temperatur varierende mellem 47°C til 57°C over 12 trin. PCR programmet så således ud: 2 min ved 95°C, 30 cykler á 30 sek ved 95°C, 30 sek ved 47-57°C og 30 sek ved 72°C. Alle PCR reaktioner blev testet på 1% agarosegeler. Primære primersæt som ikke gav et resultat efter temperaturgradienten blev droppet, mens primersæt med positive resultater blev testet på nematode-DNA (positive kontroller) med samme betingelser som før, men mellem 20 og 80 pg nematode-DNA blev brugt til hver reaktion. De yderste primersæt blev testet en enkelt gang på både jordprøve-DNA og nematode-DNA. PCR reaktioner med positivt resultat for nematode-DNA blev poollet og oprenset med AMPure beads (ratio 1:0,7 prøve til beads) og elueret i 40 μL TE buffer (pH 8). Kvalitetskontrol blev foretaget med Agilent High Sensitivity D1000 ScreenTape på Agilent 2200 TapeStation og kvantificering blev udført med et Qubit flourometer (Invitrogen).

## Sekventering og databehandling af nematode-pool

Poolen af nematode-prøver blev sekventeret på MiSeq (Illumina, San Diego, USA) (2x300 bp). Reads blev importeret ind i CLC Genomics Workbench v. 12, hvor reads blev trimmet på kvalitet (quality limit: 0,05) og adaptore. Reads blev derefter merged hvorefter de blev mappet til genomet (hhv. *G. rostochiensis* og *G. pallida*) (mismatch cost 2, insertion og deletion cost 3, similarity fraction 0,8, length fraction 0,5). Til sidst blev mapping af merged og not-merged reads samlet til en mapping.

## Resultater

### Jordprøver og positive kontroller

Tabel 1 er en oversigt over de 17 jordprøver modtaget fra Lars Bødgers projekt. Én prøve er testet positiv for *G. rostochiensis*, 10 prøver er testet positive for *G. pallida*, mens seks prøver er testet negative for begge nematoder.

Tabel 1: Resultater fra udtagning af jordprøver.

Anonymt ID (AAU reference)	Dato rapport	Antal cyster	Levedygtige larver og æg	Infektionsgrad	% <i>G. ros</i>	% <i>G. pal</i>
1	29-05-2019	377	1.250	middel infektion	0%	100%
2	29-05-2019	531	5.850	meget kraftig infektion	100%	0%
3	29-05-2019	154	99	lav infektion	0%	100%
4	29-05-2019	188	2.200	kraftig infektion	0%	100%
5	29-05-2019	40	1.327	middel infektion	0%	100%
6	29-05-2019	63	1.500	middel infektion	0%	100%
7	29-05-2019	0	0			
8	29-05-2019	136	5.300	meget kraftig infektion	0%	100%
9	29-05-2019	130	5.350	meget kraftig infektion	0%	100%
10	29-05-2019	135	3.925	kraftig infektion	0%	100%
11	29-05-2019	0	0			
12	29-05-2019	1	29	lav infektion	0%	100%
13	29-05-2019	96	61	lav infektion	0%	100%
14	29-05-2019	0	0			
15	13-06-2019	0	0			
16	13-06-2019	1	0			
17	13-06-2019	4	0			

Tabel 2 viser de 11 nematodepopulationer modtaget fra James Hutton Institute (Skotland). Seks af populationerne er *G. pallida* populationer, mens fem er *G. rostochiensis* populationer.

Tabel 2: Nematodepopulationer modtaget fra James Hutton Institute. DNA er ekstraheret fra ~30 cyster per population.

Population	Ophav
Pa2/3	Lindley, Storbritannien
Pa3	Luffness, Storbritannien
Pa3 (EU standard)	Chavronney, Schweiz
Pa1	Skotland
Ro1	Storbritannien

R02	Europa
Ro3	Europa
Ro4	Europa
Ro5	Europa
Pa4	Sydamerika
P5A	Sydamerika

## PCR test og optimering

Alle 51 primersæt blev testet med temperaturgradient med jordprøve-DNA. Ud af dem var der positive resultater for 30 primersæt, se Tabel 3. 17 primersæt viste også et positivt resultat ved test på nematode-DNA. Nogle af resultaterne var mere eller mindre tvivlsomme pga. flere bånd eller bånd hvor der var usikkerhed om de matchede den forventede ampliconstørrelse. 15 af de yderste primersæt blev testet på nematode-DNA, hvoraf 11 viste sig at give positive bånd, og disse 11 blev udvalgt til sekventering. Deraf var der 9 der gav et forventeligt sekventeringsresultat (reads i de udvalgte områder), ét gav positive sekventeringsresultater men blot 43 reads, mens endnu et ikke viste resultater i det forventede område.

Tabel 3: Resultater fra primertests. Amp (bp): amplicon størrelse med primere. Annealing temp: gennemsnitlig teoretisk annealingtemperatur beregnet på baggrund af de designede primere. Temp fra temp grad: optimal temperatur bestemt ved temperaturgradient. Pos bånd jordprøver/nematode: angiver om primertest af primære primersæt med hhv. jordprøve-DNA eller nematode-DNA resulterede i et positivt bånd på agarosegel. Pos bånd nested PCR jordprøver/nematode: angiver om primertest af yderste primersæt med hhv. jordprøve-DNA eller nematode-DNA resulterede i et positivt bånd på agarosegel. Pos sekvens: angiver om sekventering af amplicons med nematode-DNA resulterede i reads i det pågældende område. Fortsættes med: angiver om resultaterne giver anledning til at det pågældende primersæt udvælges til videre forsøg.

Primersæt	Amp (bp)	Annealing temp	Temp fra temp grad	Pos bånd jordprøver	Pos bånd nematode	Pos bånd nested PCR jordprøver	Pos bånd nested PCR nematoder	Pos sekvens	Fortsættes med
GPLIN_11	533	50	47	Ja	Ja	Nej	Ja	Ja	Ja
GPLIN_26	573	54	49	Ja	Ja	Smear	Ja	Ja	Ja
GPLIN_28	565	52	48	Ja	Ja	Ja	Ja	Ja	Ja
GROS_1	545	51	48	Ja	Ja	Ja	Ja	Ja	Ja
GROS_2	442	51	49	Ja	Ja	Ja	Ja	Ja	Ja
GROS_6	536	50	48	Ja	Ja	Ja	Ja	Ja	Ja
GROS_8	505	51	47	Ja	Ja	Ja	Ja	Ja	Ja
GROS_10	533	52	48	Ja	Ja	Ja	Ja	Ja	Ja
GROS_17	491	51	48	Ja	Ja	Ja	Ja	Ja	Ja
GPLIN_30	445	51	47,5	Ja	Ja	(Ja)	Ja (svagt)	(43 reads)	Ja
GROS_21	522	52	54	Ja	Ja	Ja	Ja	Nej	Ja
GPLIN_6	498	51	48	Ja	Ja	Ja	Nej		Nej
GPLIN_17	356	49	47	Ja	Ja	Ja	Nej		Nej
GPLIN_21	626	51	54	Ja	Ja	Nej	Nej		Nej
GROS_19	477	51	47,5	Ja	Ja	(Ja)	Nej		Nej
GPLIN_4	523	54	48	Ja	Ja				Nej
GROS_3	533	49	48	(Ja)	Nej				Nej
GPLIN_1	490	52	47,5	Ja	Ja/Nej				
GPLIN_3	546	52	48	Ja	Nej				
GPLIN_5	510	53	48	Ja	Nej				
GPLIN_9	275	51	47,5	Ja	Nej				

GPLIN_10	521	52	47,5	Ja	Nej
GPLIN_13	516	52	47,5	Ja	Nej
GPLIN_14	579	52	47,5	Ja	Nej
GPLIN_18	456	50	48	Ja	Nej
GPLIN_22	534	52	47,5	Ja	Nej
GPLIN_23	504	50	47,5	Ja	Nej
GROS_11	456	50	49	Ja	Nej
GROS_15	539	50	48	Ja	Nej
GROS_16	386	49	47,5	Ja	Nej
GPLIN_2	384	52		Nej	
GPLIN_7	451	50		Nej	
GPLIN_8	442	50		Nej	
GPLIN_12	401	51		Nej	
GPLIN_15	611	50		Nej	
GPLIN_16	445	49		Nej	
GPLIN_19	500	50		Nej	
GPLIN_20	430	50		Nej	
GPLIN_24	397	51		Nej	
GPLIN_25	460	52		Nej	
GPLIN_27	608	53		Nej	
GPLIN_29	491	53		Nej	
GROS_4	440	49		Nej	
GROS_5	534	51		Nej	
GROS_7	449	50		Nej	
GROS_9	446	53		Nej	
GROS_12	464	53		Nej	
GROS_13	425	51		Nej	
GROS_14	650	49		Nej	
GROS_18	488	49		Nej	
GROS_20	486	51		Nej	

## Fremtidigt arbejde

### Amplifikation og sekventering

De 11 primersæt med positive resultater fra primertests samt sekventering bliver udvalgt til videre behandling. Der bliver udført PCR på alle jord- og nematodeprøver enkeltvis med både de yderste primersæt og de primære primersæt. Derefter bliver PCR produkterne barcoded med barcode primere, således at det hele kan samles i en enkelt pool, som til sidst bliver sekventeret.

Sekventeringsresultaterne bliver mappet til genomet (hhv. *G. pallida* og *G. rostochiensis*) og analyseret for polymorfier.

### Udvikling af en matematisk prædiktionsalgoritme

En prædiktionsalgoritme af typen beslutningstræ der er i stand til at forudsige pathotyper på baggrund af de opnåede sekvenser vil blive udviklet. Beslutningstræet vil blive evalueret mod resultater fra de ”klassiske” analyser (Lars Bødker projekt).

