1 QQAD

1.1 Summary

To assess the diversity of a rhizobium nodule population *recA*, *rpoB*, *nodA,* and *nodD* were amplified individually with *R. leguminosarum* bv. *trifolii* specific primers. The PCR samples were pooled, purified with AMPure XP beads, tagged with Nextera tags, PCR products were cleaned by gel purification, and the samples were sequenced internally at the University of York (Illumina Myseq, 2x250 bp paired end reads). The reads were assembled using PEAR. Primers were removed prior to the analysis and amplicons with no recognizable primers and truncated reads were discarded.

1.2 Sampling

Roots from white clover from three different DLF trial sites, UK, DK, and F and 50 Danish organic fields (DKO) were collected. Representative soil samples were collected from clover free sites. Root and soil samples were stored at 2°C until processing.

The roots were washed thoroughly with tap water and>100 pink nodules were collected. The samples were placed on ice immediately after collection and subsequently stored at -18°C.

1.3 DNA extraction

Pooled nodule samples (n>100) were thawed at ambient temperature. Nodule samples were crushed and homogenised using a sterile 0.5mL plastic tube in a sterile petri dish. 750μl of bead solution from the MO BIO PowerLyzer® PowerSoil® DNA isolation kit (QIAGEN, USA) was added and thoroughly mixed with the crushed nodules. DNA was isolated following the manufacturer's instructions. DNA concentration was checked using NanoDrop 3300 (Thermo Fisher Scientific Inc., USA). DNA from soil samples was extracted using the MO BIO PowerLyzer® PowerSoil® DNA isolation kit (QIAGEN, USA) following the manufacturer's instructions.

1.4 Amplification and primers

Primers were designed for two *R. leguminosarum* core genes, recombinase A (*recA*) and RNA polymerase B (*rpoB*), and two accessory genes necessary for establishing the symbiosis with *T. repens*, *nodA* and *nodD*. The primer sequences are listed in Table 1. The QQAD PCR programme is summarised in Table 3. The PCR was carried out individually for each primer pair using Platinum® *Taq* DNA polymerase (Thermo Fisher Scientific Inc., USA). Subsequently the four PCR reactions for each sample were pooled and purified using the Agencourt AMPure XP beads (Beckman Coulter, USA).

Target gene	Primer direction	Sequence
rpoB	Forward	5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG ANN NHN NNW NNN HGY TCG CAG TGG TGG ATG TT-3'
rpoB	Reverse	5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTC CGT CTT CRA GGA ACG GCA T-3'
recA	Forward	5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG ANN NHN NNW NNN HCG AGA ATG TTG TCG AGA TYG AGA CGA-3'
recA	Reverse	5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTT ATC GGT GAT TTC RAG CGC CTG-3'
nodA	Forward	5'-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG ANN NHN NNW NNN HCC GGA TCT SGA GGG GCT-3'
nodA	Reverse	5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA CTG CAN CCG TTT CGT TCG ATC AAT GA-3'
nodD	Forward	5'-TCG TCG GCG TCA GAT GTG TAT AAG AGA CAG ANN NHN NNW NNN HAT GCG TTT TAA GGG MYT GGA TCT-3'
nodD	Reverse	5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTG CRC CGG TCA GAT TCC GC-3'

Table 1: Primer sequences used for the QQAD analysis.

Component	Vol. in 20μl	Vol. in mastermix (x100)
Water	14.52	1452
Buffer	2	200
$MgCl_2$	1.2	120
dNTP	0.4	40
F primer	0.4	
R primer	0.4	
Platinum HS Taq	0.08	8
DNA	1	

Table 2: Mastermix for 96 well PCR plate.

Temperature (°C)	Time (seconds)	Cycles
95	180	
94	30	
70	300	2
72	120	
94	30	
70	60	30
72	60	
72	600	
4	hold	

Table 3: QQAD PCR programme.

1.4.1 Detailed PCR protocol

The PCR should be performed individually for the four primer sets to ensure an equal amount of product is produced for each gene. Make the mastermix for approx. 100 samples containing everything except primers and DNA (Table 2). Always make more mastermix than the actual amount of samples you need to run. Store everything on ice, and add the polymerase (Platinum HS Taq) last. Split the mastermix into four tubes and add a primer set to each (forward and reverse primer). Pipette the mastermix, now containing primers, into the PCR tubes on ice. Add 1μl DNA to each tube. Make a no-template-control (1 μl water instead of DNA) and a positive control (DNA extract from DLF nodule samples stored in York) for each primer pair.

Run the PCR as specified in Table 3.

1.4.2 PCR product cleanup

Check a few μl of sample on a gel to ensure amplification has been carried out. Expected fragment size is around 300bp. Pool the remaining PCR product for each sample (four 20μl PCR reactions has been done for each sample, yielding a total of 80μl for each sample). Purify the product using the AMPure XP beads.

1. Vortex beads at room temperature until they are completely resuspended
2. For 20μl PCR product use 16μl beads (For 80μl PCR product use 64μl of beads)
3. Pipette up and down 10 times
4. Incubate at room temperature for 5 minutes to allow DNA to bind to beads
5. The following steps should be carried out on a magnetic stand
   (a) Place on a magnetic stand for 2 minutes or until supernatant has cleared
   (b) Remove supernatant
   (c) Add 200μl freshly made 80% ethanol
   (d) Incubate for 30 seconds
   (e) Remove supernatant

(f) Repeat step c to e

(g) Air dry for 10 minutes

6. Remove tubes from magnetic stand
7. Add 525μl 10 mM Tris pH 8.5 to elute DNA from beads
8. Pipette up and down 10 times (make sure beads are fully resuspended)
9. Incubate for 2 minutes
10. Return to magnetic stand for 2 minutes
11. Carefully transfer 50μl supernatant to a new sterile tube without disturbing the pellet

1.5 Sequencing

Sequencing libraries were prepared using Phusion High-Fidelity DNA polymerase. The PCR product was purified on a dense agarose gel (1.5%) prior to sequencing. The amplicon bands (approx. 300 bp) were excised and DNA was extracted using the QIAquick® Gel Extraction Kit (QIAGEN, USA). To normalise the amount of PCR product from all samples used for sequencing, visual approximation of the quantity of each band was calculated using GelAnalyzer 2010a [Lazar, 2010]. The samples were sequenced internally at the University of York by the Technological facility using Illumina MySeq.

1.5.1 Nextera Index PCR

The amplification was tested with 5, 10, and 15μl purified PCR product. 5μl was chosen since the DNA concentration was very high and a higher starting volume resulted in less clear bands, when the amplification was assessed on a gel. Instead of Platinum HS Taq polymerase, a proofreading polymerase, Phusion High Fidelity (Thermo Fisher Scientific Inc., USA), was used. The components for 50μl reaction volume is specified in Table 4. Due to the price of the Nextera tags no controls are run. The tags should be added in unique combinations as shown in Figure 1.

Component	**Vol. in 50μl**	**Vol. in mastermix (x25)**
Water	23.5	587.7
Buffer	10	250
dNTP	1	25
XT index S	5	
XT index N	5	
Phusion	0.5	12.5
DNA	5	

Table 4: Nextera index PCR for 50μl reaction volume.

Temperature (°C)	**Time (seconds)**	**Cycles**
95	180	
95	30	
55	30	10
72	30	
72	300	
4	hold	

Table 5: Nextera index PCR programme.

		N701	N702	N703	N704	N705	N706	N707	N708	N709	N710	N711	N712
	PCR 1	1	2	3	4	5	6	7	8	9	10	11	12
S502	A	1 1JJ	2 2PB	3 3ORP	4 4HPK	5 5HPK	6 6LN	7 7MJ	8 8HB	9 9HB	10 10JSK	11 11JSK	12 12ISL
S503	B	13 13ISL	14 14JPV	15 15AK	16 15AK	17 17EB	18 18NG	19 19JJ	20 20JJ	21 21PM	22 22PM	23 23PR	24 24WB
S504	C	25 25MSH	26 26MSH	27 27HMH	28 28EP	29 29EP	30 30HB	31 31AJ	32 32AJ	33 33vP	34 34vP	35 35AMT26	36 36BR
S505	D	37 38RM	38 39BJS	39 40PG	40 41JM2	41 42KFN	42 43IL	43 44FG861	44 45UM	45 46AK	46 47AK	47 48CC	48 49CC
S506	E	49 50KGN	50 35AMT27	51 106-11F	52 105-19F	53 105-18F	54 105-36F	55 105-15F	56 105-14F	57 106-30F	58 105-31F	59 106-7F	60 105-2F
S507	F	61 106-9F	62 105-12F	63 105-27F	64 105-5F	65 105-17F	66 107-16F	67 105-26F	68 106-10F	69 105-6F	70 105-4F	71 105-25F	72 106-36F
S508	G	73 105-16F	74 107-15F	75 105-23F	76 105-13F	77 105-29F	78 105-21F	79 107-28F	80 106-24F	81 105-31F	82 107-11F	83 105-20F	84 106-6F
S517	H	85 107-20F	86 107-29F	87 106-32F	88 106-28F	89 107-27F	90 107-13F	91 105-34UK	92 106-7UK	93 105-12UK	94 106-9UK	95 106-32UK	96 105-13UK

Figure 1: Plate layout showing the added tags in each well.

1.5.2 QIAquick Gel extraction

PCR products from the Nextera PCR were loaded onto a dense gel to purify the product prior to sequencing.

- 1.5% agarose gel
- 10μl PCR product
- Run for 2 hours at
  - 80V for large gels
  - 60V for small gels
- Load wells with empty well between each sample to prevent transfer of product between wells
- Expected fragment size is approx 300bp

Follow the manufacturer's instructions when using the extraction kit except for the following alterations:

- Step 1: Excise the band from the gel using the wide end of an autoclaved 200μl pipette tip
  - Blow the excised gel piece into a sterile Eppendorf tube using a needleless syringe
  - For demonstration, contact David Sherlock, the inventor of the method
- Step 9: Add 30μl Buffer EB heated to 50°C and incubate for 5 minutes prior to centrifugation
- Gel analysis: Use 2μl loading dye and 2μl purified DNA on normal 1% agarose gel. This is used to normalise the DNA concentration of the samples prior to sequencing.

1.6 Data analysis

Sequencing reads were merged using the PEAR assembler [Zhang et al., 2014]. The reads were clustered into gene groups using CD-HIT [Li and Godzik, 2006, Fu et al., 2012]. All subsequent analysis were done using the R software and figures were produced using the 'ggplot2' package [R Core Team, 2017, Wickham, 2009]. Reads differing by only a single nucleotide were merged within each sample to exclude possible PCR errors. The five most abundant amplicons for each sample were not allowed to merge with each other. Filtering parameters were set to include amplicons with n>500 reads. Neighbor-joining trees were created online using the Clustal Omega online multiple sequence alignment tool and inherent phylogenetic analysis tool [McWilliam et al., 2013].

References

[Fu et al., 2012] Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). Cd-hit: accelerated for clustering the next generation sequencing data. *Bioinformatics*, 28(23):3150–3152.

[Lazar, 2010] Lazar, I. (2010). Gelanalyzer 2010a: Freeware 1d gel electrophoresis image analysis software. *Available at: http://www.gelanalyzer.com/*.

[Li and Godzik, 2006] Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22:1658–9.

[McWilliam et al., 2013] McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., and Lopez, R. (2013). Analysis tool web services from the embl-ebi. *Nucleic acids research*, 41(W1):W597–W600.

[R Core Team, 2017] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Wickham, 2009] Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[Zhang et al., 2014] Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). Pear: a fast and accurate illumina paired-end read merger. *Bioinformatics*, 30(5):614.