Short-term Scientific Mission (STSM) Report

COST Action FA0807

Integrated Management of Phytoplasma Epidemics in Different Crop Systems

STSM TITLE: Napier Glass Stunt Phytoplasma Genome assembly and annotation

NAME: Mr Praphat Kawicha

SUPERVISOR AND HOST: Dr Michael Kube, Humboldt-University of Berlin

START DATE: 20th August 2012 **END DATE:** 29th October 2012 (Note, the STSM comprised two periods of stay in Berlin - the first was for 1 week starting on the 20th August. A 6-week period was spent back at Nottingham conducting finishing experiments, followed by a second visit to Berlin for 3 weeks to complete the training in genome assembly and annotation skills.)

Introduction

Our research team recently submitted the napier grass stunt phytoplasma (Group 16SrXI) for genome sequencing through the phytoplasma genome sequencing initiative (PGSI) annotation school that is being operated through the COST Action FA0807, and the quality of the sequencing has proven to be extremely good and acceptable for further follow-up studies. In particular, it was felt that completion of the genome sequence is possible through a relatively small amount of additional work. This STSM involves improving the genome assembly in order to reduce the risk of misassembles and the number of gaps, and we set up a genome database to design and conduct final experiments for PCR and sequencing.

Description of the work

WEEK 1 : Introduction to the handling of database

Location: Humboldt-University of Berlin

Napier grass stunt phytoplasma (NGS) genome was sequenced by using the Illumina method through the PGSI annotation school. Having assembled the genome by using CLC genomic workbench software version 4.5, NGS sequences were ordered into 44 contigs. There were 19 contigs which were larger than 5 kb and an estimated genome size was around 495.4 kb.

To fill the gaps between the 44 contigs, PCR technique were then applied. Forward and reverse primers for a PCR experiment including 89 primers, were designed by GAP 5 software. Appropriate primer sites for primer design are shown in Figure 1. The designed primers were synthesized by Eurofins MWG Operon Company.

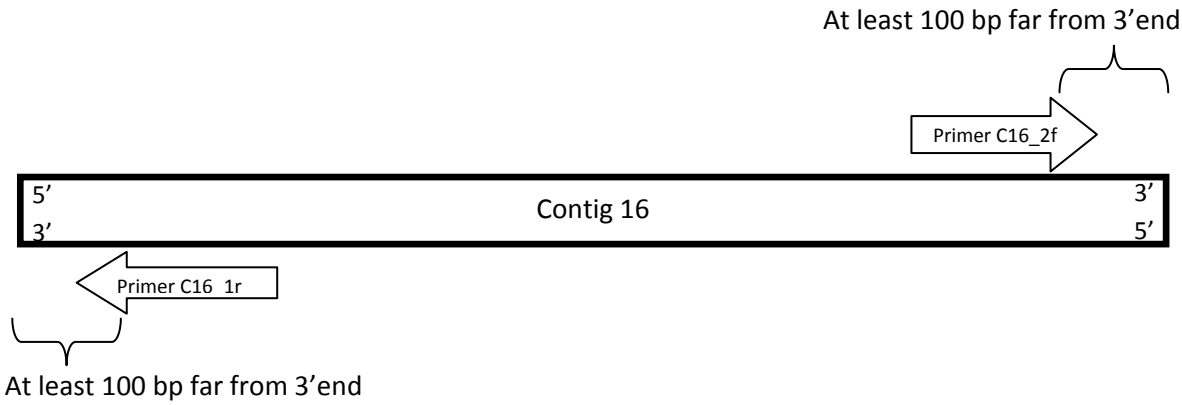

Figure 1. Location of primer site on the contig.

WEEK 2-7 : Conducting the finishing experiments (PCR, sequencing)

Location: University of Nottingham

Gaps were filled by PCR for contigs which had the possibility to be assembled based on a gene at the start and the end of a contig and a BLAST result against the relaxed NGS genome assembly database.

PCR was performed with 1X PCR buffer (GoTaq® Flexi DNA polymerase, Promega cooperation); 12.5 ng of template DNA; 0.5 mM each primer; 0.2 mM (each) dATP, dCTP, dGTP, and dTTP; 2 mM of MgCl2 and 1.25 U of GoTaq® Flexi DNA polymerase in a final reaction volume of 12.5 µl. Amplification consisted of 30 cycles of the following steps: denaturation for 30 sec at 94^{o}C; annealing for 30 sec at 48-51^{o}C (depended on Tm of primer); and extension for 1.5 min at 72^{o}C. PCR solution was electrophoresed in a 1% agarose gel, and the PCR products were visualized by UV transillumination after staining with ethidium bromide.

1,600 PCR reactions which had different pairs of primer in each reaction were done. Only 116 PCR reactions provided single band products. The products were purified by GenElute^{TM} PCR clean up kit, Sigma Aldrich$^{®}$, and sequenced by Eurofins MWG Operon Company. In total, 49 PCR products were successfully sequenced but the rest failed. Having low melting temperatures (Tm), some primers were probably non-specific resulting in non-specific products or mixed and impure products, and this probably resulted in the sequencing failure.

WEEK 8-10 : analysing of the data for genome assembly

Location : Humboldt-University of Berlin

The sequences of PCR products were aligned with and used to join the contigs (Figure 2) by using GAP 5 software. Twenty sequences of PCR product were assembled into the contigs but only a half of them were able to join the contigs. To this point, we have been able to assembly some contigs and decrease the

total number to 31 contigs (Table 3). Sixteen of them are between 5 kb to 128.8 kb length and the rest are small contigs of 200 bp to 3.8 kb. The estimated genome size is around 496.5 kb.

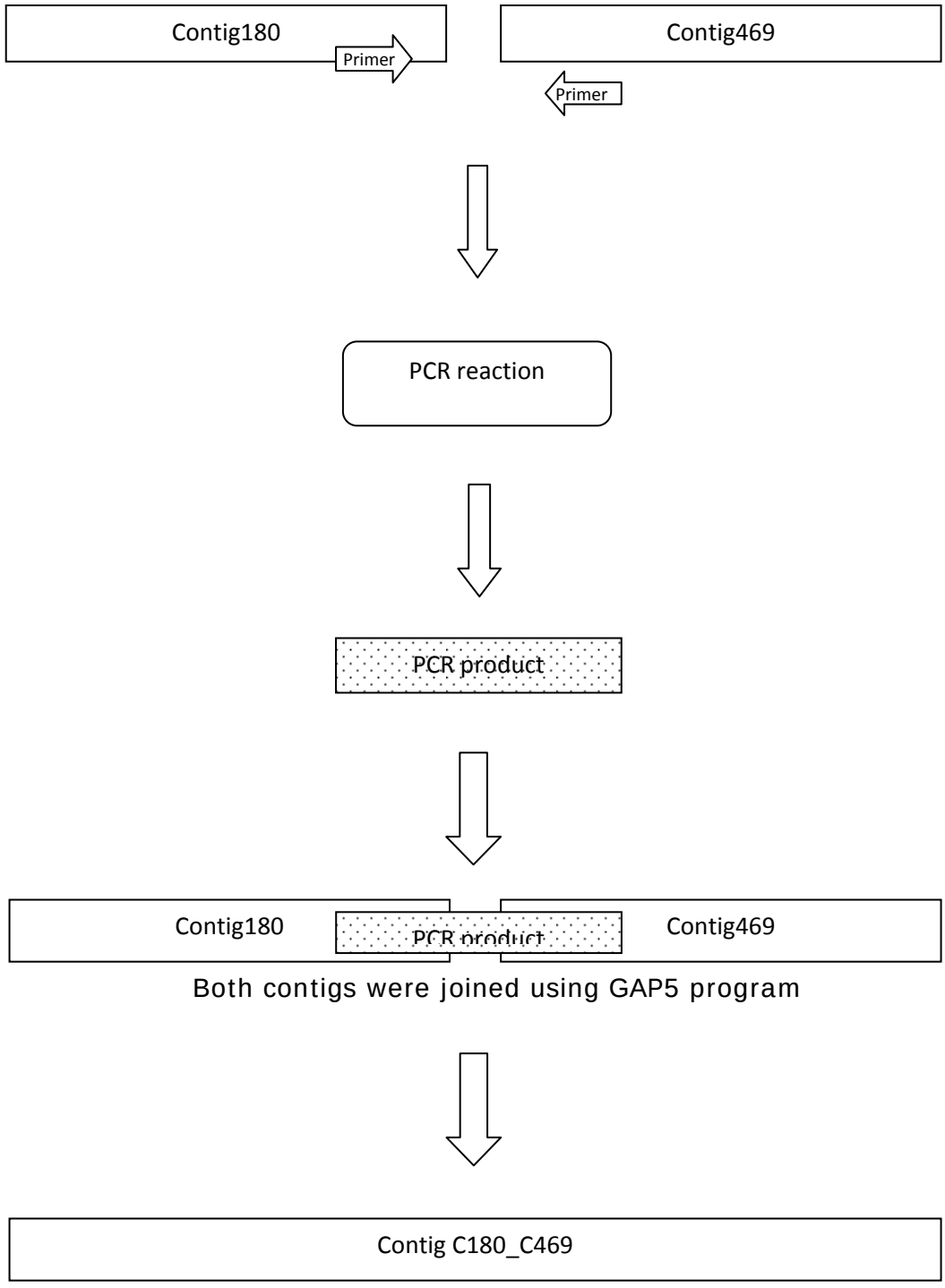

Figure 3. Scheme for filling the gap and joining the contigs.

Table 3. Updated sequnces contigs of NGS phytoplasma.

Contig name	Comprised contigs	Length (bp)
C180_C469	180, 469	128848
C9625_C457	9625, 457	65212
C63	63	60145
C275_C1323	275, 1323	33435
C86_C433_C3887_C111_C85_C2523	86, 433, 3887, 111, 85, 2523	32616
C408_C2093	408, 2093	25863
C175	175	25085
C20	20	17780
C21	21	14126
C286	286	13553
C1451	1451	13333
C492	492	11151
C1136	1136	9715
C368	368	8517
C2188_C4850	2188, 4850	6299
C16	16	5054
C980	980	3832
C258	258	3694
C22290	22290	3176
C232	232	3102
C709	709	3022
C6744	6744	2740
C25640	25640	1162
C245	245	951
C2677	2677	878
C775	775	851
C40167	40167	760
C54944	54944	658
C31837	31837	435
C1303	1303	388
C13323	13323	200
Total contig number : 31	Estimated genome size :	496581 bp

Conclusion and future collaboration

Ten gaps in the NGS genome sequences were filled. However, there were a number of PCR products which failed to be sequenced and we expected that some of them would be able to fill the gaps. For further experiments, they will be cloned and re-sequenced. After getting the sequences and filling the rest of gaps, the number of gaps and contigs should be reduced. Thereafter, the genome database will be updated and we will re-design and conduct other experiments for PCR and sequencing in order to completely fill the gaps of the rest. Finally, NGS genome annotation will be performed.