Channeling targeted DNA double strand breaks into alternative repair pathways

Project Summary

We have expressed TAL DNA binding domains fused to the FokI nuclease under meiotic promoters (e.g. DMC1, SPO11) in Arabidopsis. The aim of this work is to target DNA double strand breaks to specific sites in the genome, in order to bias initiation of meiotic recombination. Our preliminary data show that while these nucleases are expressed in meiotic-stage floral buds they do not support wild type levels of crossover recombination when the endogenous nuclease (SPO11-1) is mutated. Additionally these transgenic lines show occurrence of developmental phenotypes, leading us to the hypothesis that the resulting DSBs enter a mutagenic pathway. To investigate this in this project we are performing whole genome DNA sequencing and mutation discovery. This has been performed using support from the OpenPlant project and bioinformatic mutation discovery is ongoing. In parallel we have crossed these nuclease lines to mutants in canonical and alternative end joining pathways to test the hypothesis that we can shunt DSBs into crossover recombination via removing competing repair pathways. These lines will be grown and DNA sequencing repeated, in addition to phenotypic analysis in the next part of this project.

Report and Outcomes

Prior to the start of this project a large number of TALEN constructs had been transformed into Arabidopsis. The constructs used arrays of TAL DNA binding domains fused to the FokI nuclease, under the control of meiotic promoters – specifically DMC1 and SPO11-1. As FokI functions as a dimer our constructs contained paired arrays designed to bind adjacent sites, positioning the FokI proteins together to allow binding and DNA double strand break formation. As a functional test of these constructs they were initially transformed into spo11-1 mutants, which are sterile due to an absence of crossovers and subsequent aneuploidy. The expectation, if the TALEN constructs were able to generate DSBs which successfully entered the crossover repair pathway, would be restoration of fertility. Although some T1 lines showed increased fertility of spo11 homozygotes, in no case was fertility restored to wild type levels.

We confirmed that TALEN proteins were detectable via western blotting. This led to the hypothesis that DSBs generated by the TALEN constructs may be processed by recombination pathways other than that which leads to inter homolog crossover repair. Consistent with this we observed developmental phenotypes in many of our transformed lines, which led to the further hypothesis that mutagenic repair was occurring potentially via the non-homologous end joining pathways (NHEJ).

First DNA was extracted from multiple of the meatless transformed lines, which were selected according to the range of phenotypes they displayed and varying target site specificities (i.e. different TAL arrays). Genomic DNA was extracted from replicate individuals per line and used to generate barcoded DNA sequencing libraries using the Illumina Truseq kit. These libraries were then subjected to sequencing using a Hiseq instrument. These data were analysed by aligning reads to the Arabidopsis TAIR10 reference sequence using a standard polymorphism discovery pipeline. Briefly, reads were aligned using bowtie2 with default parameters, followed by processing using samtools (BAM
compression, sorting and indexing). The samtools mpileup and bcftools call functions were then used to call polymorphisms. These calls were filtered using vcftools to keep sites with qualities greater than 100 and SNPs and indel polymorphisms separated. The sites were then used as an input for the GATK pipeline. Here the original .fastq files were realigned using BWA, followed by read deduplication using picard. GATK was then used for local realignment of reads around indel sites and base quality scores were recalibrated. The GATK haplotype called was then used for a further round of variant discovery, recalibration and filtering. This cycle was repeated a further time and limited change in the final variant sites observed. The resulting SNP and indel variants from each library are currently being used for further bioinformatics analysis. Specifically, we are overlapping these variants with predicted TAL DNA binding sites.

As our TAL domains contain degenerately binding repeats this task is more complicated that initially anticipated and we are in the process of testing various methods for overlapping. We have also performed a limited amount of Sanger DNA sequencing to validate predicted variants and have been successful in confirming these mutations.

Therefore, in summary we have used support from OpenPlant to progress our understanding of genetic changes occurring as a consequence of meiTALEN expression in the *Arabidopsis* genome. We are in the process of testing the role of NHEJ pathways in causing the variants and phenotypes we observe, which will be completed during the remainder of the project.

---

**Project Expenditure**

The funds have been used to support genomic DNA library preparation from meatless transformed lines, and to pay for subsequent sequencing. Replicate libraries were generated and barcoded using the Illumina Tru-seq kit. Sequenced was performed using a HiSeq instrument at the Beijing Genomics Institute. The remaining funds will be used to complete sequencing of meiTALEN spo11 xrcc1 ku70 plants, once they are identified in the F2 or F3 generation (ongoing).

---

**Follow-On Plans**

The second major objective of this research is to cross meatless transformed lines with mutations in canonical and non-canonical NHEJ pathways. Therefore, a subset of lines have been crossed to xrcc1 ku70 double mutants – the rationale being that disruption of these pathways may then channel FokI-derived DSBs into the crossover repair pathway. The genetics of this part of the project is ongoing and the F2 population segregating for the meiTALEN, spo11-1, xrc11 and ku70 will be obtained – we will attempt to identify meiTALEN spo11 xrcc1 ku70 lines and measure fertility. These lines will also be used for DNA sequencing using the remainder of the OpenPlant funds.