Background

Disease resistance in plants, is frequently conferred by nucleotide-binding and leucine-rich repeat domain (NLR) proteins. In fact, the majority of cloned dominant resistance (*R*) genes broadly used by breeders code for NLR proteins. NLR proteins act as immune receptors that recognize pathogen effectors in the cytosol. In plant genomes, NLR-coding genes constitute huge and highly diverse gene families. The rice reference genome contains for example 276 genes that code for full length NLRs containing all canonical NLR domains (coiled-coil, NB-ARC and LRR domains), 209 proteins that lack one or the other of these domains and 138 pseudogenes with similarities to NLRs (Luo et al. 2012)

Knowledge on full NLR repertoires in elite varieties and highly resistant crop accessions would be of outstanding importance for resistance breeding efforts and the identification of novel, interesting resistance genes. However, this has until recently, been proven difficult to obtain due to the large size and frequently complex structure of the corresponding genes and their frequent location in genomic regions that are rich in repetitive sequences as well as their occurance in large gene clusters in plant genomes. These particular features of NLR-coding genes has complicated the identification of full NLR repertories in high quality reference genome sequences and renders their identification in re-sequencing data extremely difficult. *De-novo* assembly of short reads from repetetive and highly paralogous sequences pose tremendous problems.

Recently, RenSeq, a highly innovative novel technique based on the specific capture of size-calibrated NLR-coding genomic fragments has been devopped and solves the major issues in establishing NLR repertories in crop varieties (Jupe et al. 2013). RenSeq was successfully applied to the sequenced potato *Solanum tuberosum* clone DM and the sequenced tomato *Solanum lycopersicum* clone Heinz 1706 and identified respectively 217 and 105 novel NLRs (Andolfo et al. 2014; Jupe et al. 2013).

Objective

The main objective of the STSM is to better evaluate the diversity of NLR-coding genes in important rice varieties around the world and to identify with high efficiency novel rice *R* genes. For this, I applied for the first time RenSeq to rice and performed the main experimental steps in the laboratory of Prof J. Jones at the Sainsbury Laboratory (Norwich, UK).

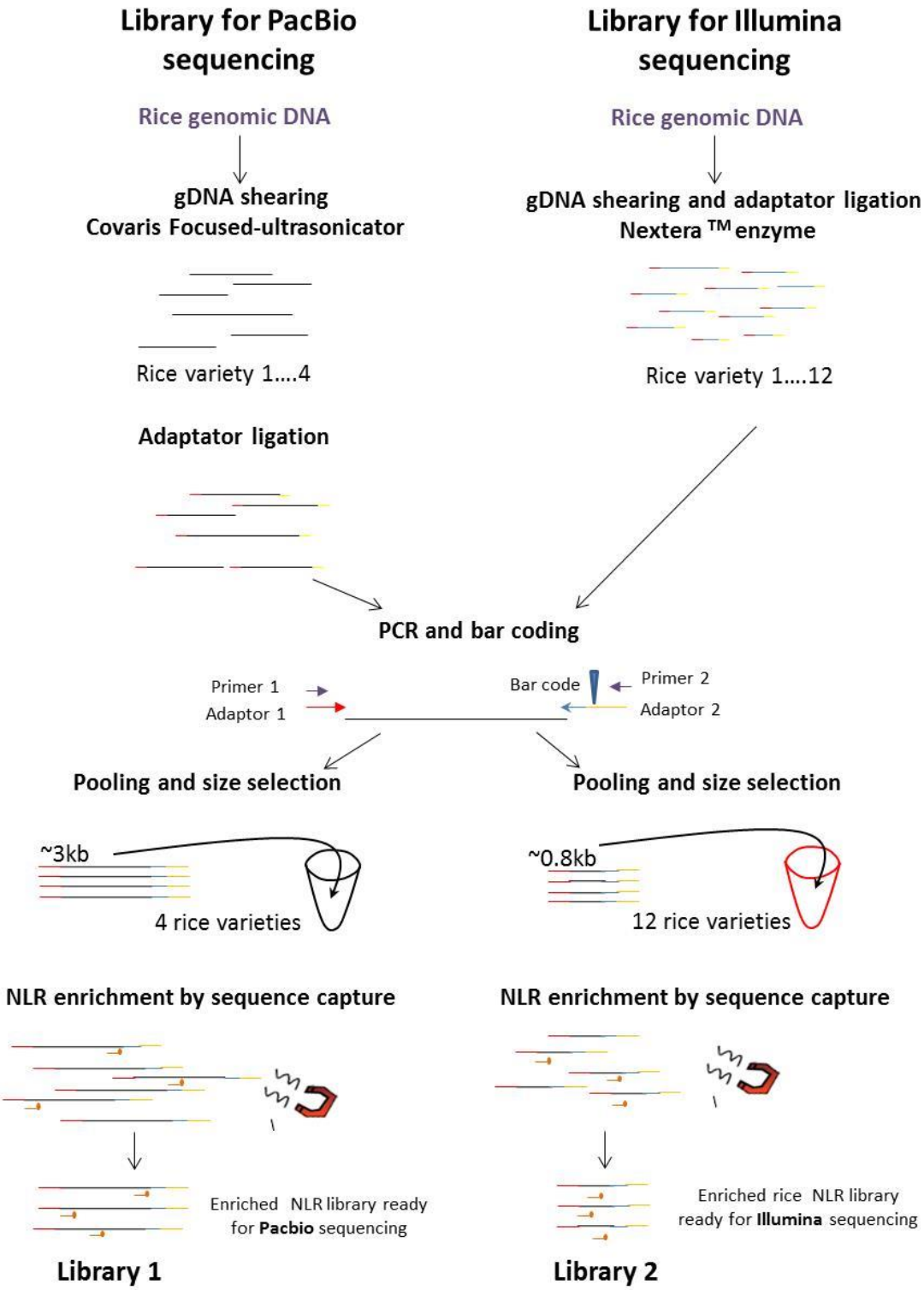
Description of the work

Two different multiplex RenSeq libraries were constructed with DNA from *Japonica* and *Indica* subespecies rice varieties with different degrees of resistance to the rice blast pathogen *M. oryzae* (table 1). Library 1 was for PacBio sequencing and was generated with DNA from 4 different rice varieties. Library 2 was for Illumina sequencing and contains DNA from 12 different rice varieties. A main diference between two libraries is the metodology that was used for DNA shearing. For the PacBio library, a method based on ultra-sonication that is well established in the host lab was used. For the Illumina library, a novel, not optimized enzyme/transposase-based method was used that promises to allow much more cost efficient DNA fragmentation. In addition size selection of fragments was different according to the sequencing metodology; for library 1, fragments of about 3 kb were selected for PacBio sequencing whereas for library 2 fragments of about 0.8 kb were selected for illumina sequencing (see flowchart in Figure 1)

Table 1. Rice varieties used for RenSeq analyses

N°	Rice variety	Blast resistance	Sub-species/group	Library 1 PacBio	Library 2 Illumina
1	Padi Boenor	Low resistance	Indica	X	X
2	IRAT13	Elevated resistance	Tropical Japonica		X
3	Tetep	Elevated resistance	Indica		X
4	Oryzica Llanos 5	Elevated resistance	Indica		X
5	Azucena	Elevated resistance	Tropical Japonica		X
6	Moroberekan	Elevated resistance	Tropical Japonica		X
7	IR64	Elevated resistance	Indica	X	X
8	Sariceltik	Very low resistance	Temperate japonica	X	X
9	Nipponbare	Elevated resistance	Temperate japonica	X	X
10	Maratelli	Low resistance	Temperate japonica		X
11	Kasalath	Intermediate resistance	Aus		X
12	Kitaake	Intermediate resistance	Temperate japonica		X

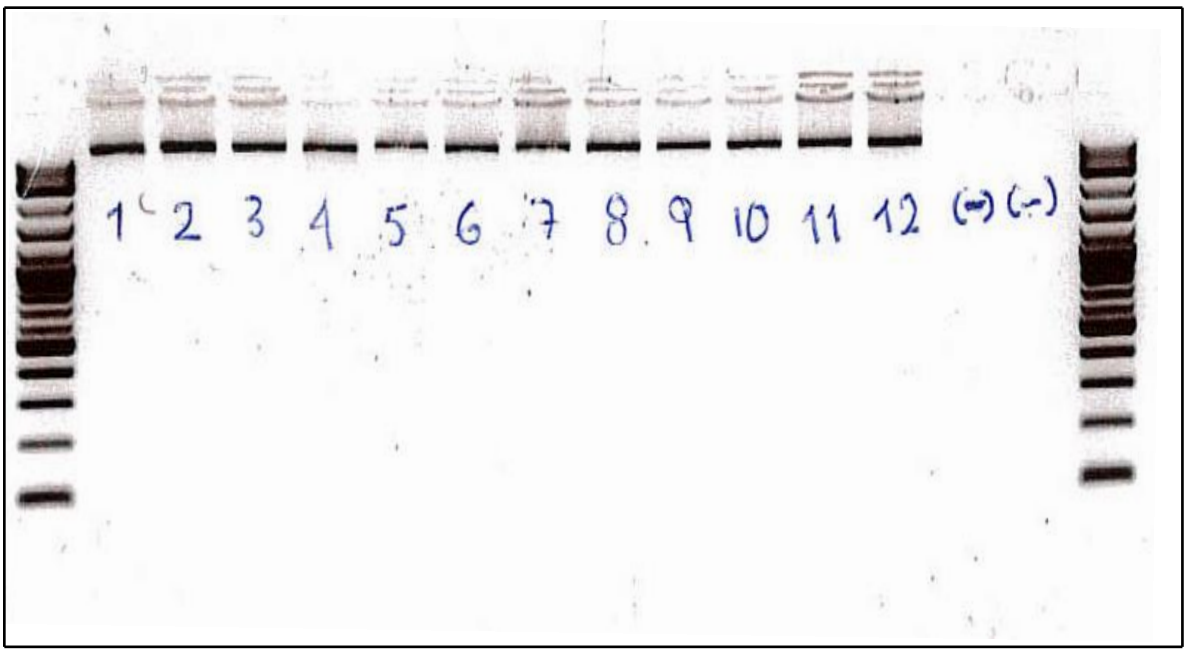
Flowchart 1. Overwiew of the 2 rice *NLR* enrichement (RenSeq) experiments



Main results

1) DNA quantification and quality control

The quality of the DNA from the 12 rice varieties was verified in 1.5% agarose gels and the concentration was measured using PicoGreen ^{R} reagent. The DNA concentration varied between 172 – 243 ng/ul and all samples exhibited a good DNA quality (figure 1).



Sample No	1	2	3	4	5	6	7	8	9	10	11	12
DNA ng/ul	214	218	207	183	172	224	231	247	176	237	229	243

2) DNA shearing

For PacBio-sequencing 4 micrograms of DNA of 4 rice varieties (table 1) were sheared seperately using Covaris AFA technology (Adaptive Focused Acoustics™) that employs ultrasonic acoustic energy in combination with specifically engineered tubes to generate DNA fragments of desired size, in this case, long DNA fragments (figure 2). For Illumina-sequencing, 20 ng of DNA per rice variety (table 1) was fragmented and adaptators were ligated with Nextera Enzyme Mix^{TM}. For this library we standarized the condidions (data not shown) to generate DNA fragments of mid size (figure 3).

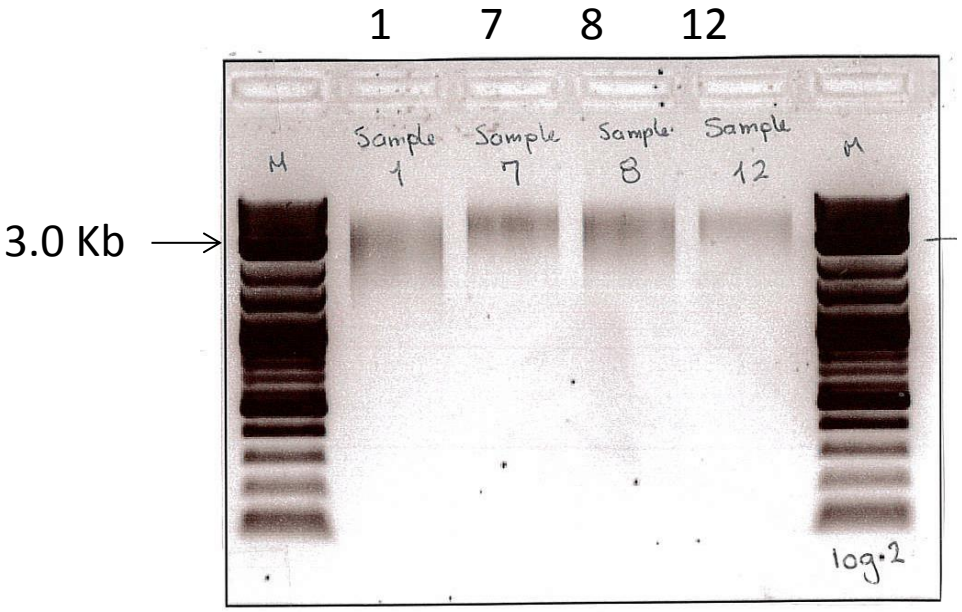


Figure. 2 DNA fragments bigger than 3kb were obtained after DNA shearing with Covaris focused-ultrasonicator. The size of DNA fragments obtained after shearing was homogenous for all samples

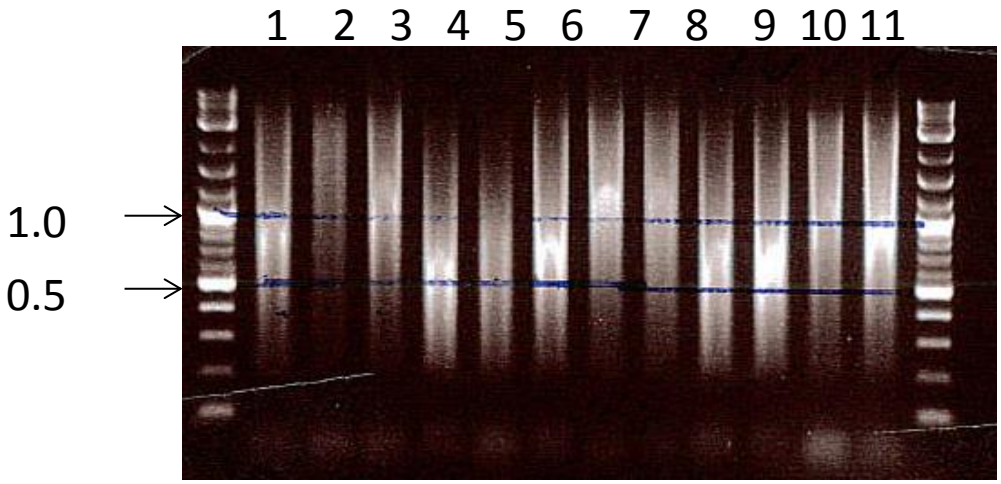


Figure. 3 DNA fragments between 0.3 to 1.2 kb were obtained after DNA shearing with Nextera Enzyme. The size of DNA fragments was highly variable between samples probably due to small differences in the DNA concentration of each rice genotype.

3) Library construction

NEBNext Ultra DNA Library Prep Kit and Nextera™ DNA sample prep kit was used to construct the DNA libraries 1 and 2 respectively. One bar code and two adaptors were ligated to DNA fragments for further identification and amplification respectively. After PCR amplification, samples were pooled in library 1 and 2. DNA fragments were separated by SageELF technology based on agarose gel electrophoresis to recover fragments of around 3 kb for library 1 and 0.8 kb for library 2 (figure 4).

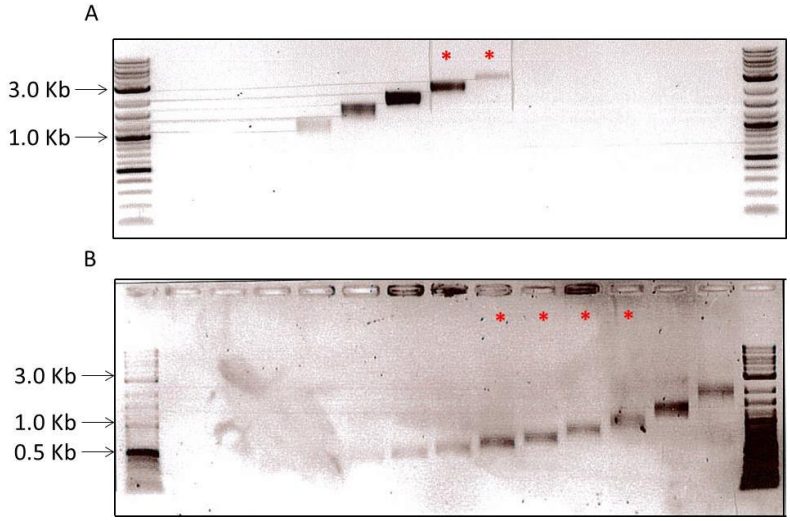


Figure 4. Separation of DNA fragments by size using SageElf cassettes. Fractions marked by starts were recovered: (A) fragments between 2 and 3kb for PacBio sequencing and (B) Fragments between 0.5 and 1.0 kb for illumina sequencing.

4) Capture of DNA target

The enrichment for rice *NLRs* was achieved by hybridization of the PCR-amplified, size-selected genomic libraries with the bait library that consists of 20 000 different 100 bp RNA sequence tags complementary to known rice NLRs and designed by the home laboratory. Captured DNA was amplified by PCR (figure 5). Based on the features of each each library, a visible band was expected to appear after 26 and 14 cycles for PacBio and Illumina libraries respectively. A correct amplification was obtained with the PacBio library but a weak amplification was obtained for the illumina library. Four additional cycles were required to visualize a weak band suggesting that capture for library 2 failed.

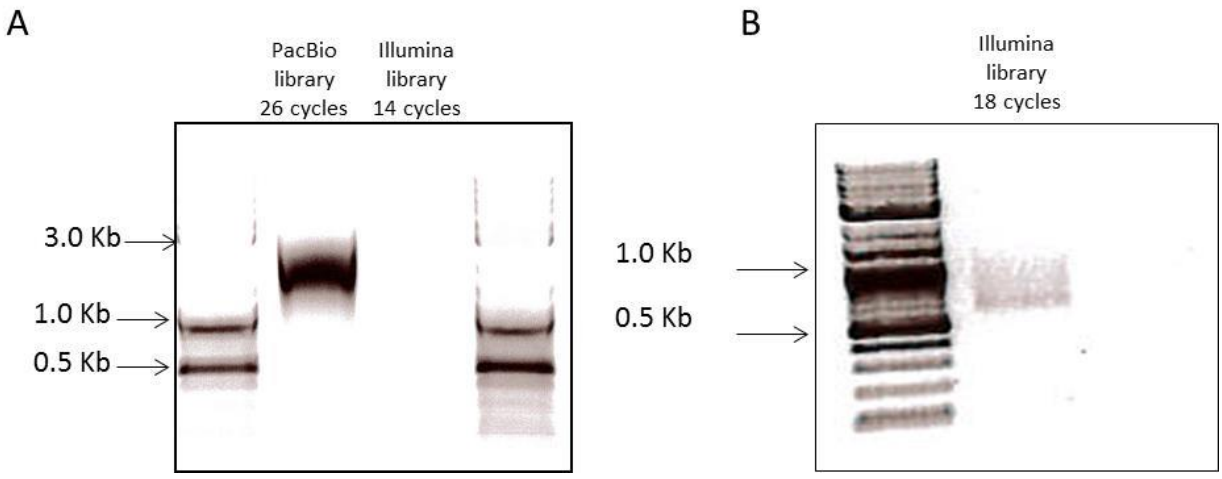


Figure 5. PCR after DNA capture library 1 and 2. (A) PCR product for PacBio and Illumina libraries after 26 and 14 cycles. (B) PCR product for illumina library after 18 cycles (4 cycles more than recomended)Monitoring of *NLR* enrichement by qPCR

To check whether the captured DNA was indeed enriched for rice *NLRs*, qPCR for three known rice NLRs *RGA4*, *RGA5* and *Pi-ta* was performed with DNA libraries before and after capture. By this, enrichment of target rice *NLRs* was confirmed for the PacBio library, but not for the the Illumina library. Indded for the Illumina library, no differences in *RGA4*, *RGA5* and *Pi-ta* amplification was detected before and after capture indicating that the *NLR* enrichement for this library did not properly work and needs to be repeated.

5) Quality control before sequencing

Before sequencing, the quality and size of the PacBio library was confirmed by electrophoresis using Agilent Bioanalyzer (figure 6).

Figure 6. Three replicates of the *NLR*-enriched rice library with DNA fragments ~3 kb

Conclusion

A RenSeq libray for four rice reference varieties, Padi Boenor**,** IR64, Sariceltik and Nipponbare was suscesfully generated. The sequencing of these library will show if the rice RenSeq bait library allows indeed efficient *NLR* enrichment and determination of whole genome NLRomes in diverse rice varieties. In addition, this analysis will provide a much more precise image of the differences in the NLR repertoires of rice varieties with highly different leves of resistance to *M. oryzae* and other pathogens.

Perspectives

- A novel Ilumina library will be constructed. For this, DNA shearing for samples with low DNA concentration or very different fragments sizes will be repeated. With these new DNA samples, a novel Illumina library will be constructed, sized and the captured with the RNA bait library.

-We will continue the collaboration to analyze the NLR repertoires of a large panel of different rice varieties .

Bibliography

Andolfo, G. et al., 2014. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. , pp.1–12.

Jupe, F. et al., 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. , pp.530–544.

Luo, S. et al., 2012. Dynamic Nucleotide-Binding Site and Leucine-Rich Repeat-Encoding Genes in the Grass Family. *Plant Physiology*, 159(1), pp.197–210.