THE SAINSBURY LABORATORY

Target enrichment and next generation sequencing as tools to facilitate cloning of *R* genes from *Solanum* species

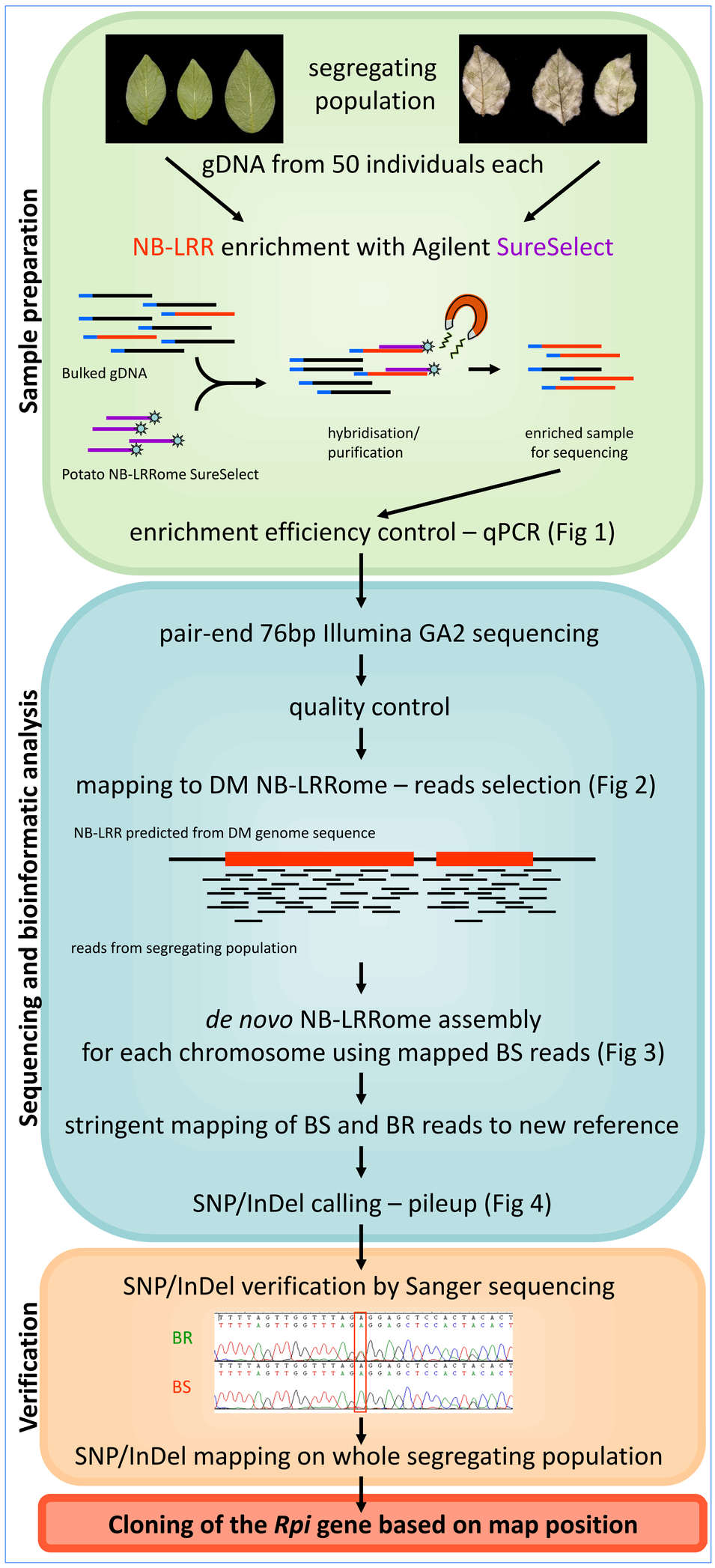
Kamil Witek[1], Jadwiga Śliwka[2], Walter Verweij[1], Florian Jupe[3], Henryka Jakuczun[2], Ingo Hein[3], Ewa Zimnoch-Guzowska[2] and Jonathan D. G. Jones[1]

[1] The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK; kamil.witek@tsl.ac.uk,

[2] Plant Breeding and Acclimatization Institute, Research Centre Młochów, Platanowa 19, 05-831 Młochów, Poland,

[3] The James Hutton Institute, Invergowrie, Dundee DD2 5DA, Scotland UK

Late blight caused by oomycete pathogen *P. infestans* is the most destructive disease in cultivated potato. Since *P. infestans* is known to quickly overcome resistance genes used in breeding programs, there is a constant necessity to identify and clone novel *Rpi* genes (*Rpi- Resistance to P. infestans*). Classical map-based cloning is a laborious and time-consuming effort; therefore we are developing a technique which combines target enrichment and next generation sequencing to accelerate cloning of new *Rpi* genes.
This technique allows to avoid classical polymorphism discovery for fine-mapping of *Rpi* genes. Ideally, our approach will allow to 'land' on the gene (cluster of genes) conferring resistance. Here we present the developed pipeline, discuss current troubleshooting and communicate potential of the approach. This newly developed technique should be applicable to facilitate cloning not only *Rpi*, but also other *R* genes from *Solanaceae* which are of NB-LRR type.
Briefly, samples consisting of combined 50 susceptible (BS) and 50 resistant (BR) individuals are first enriched for NB-LRR genes using Agilent SureSelect with probes designed against 470 NB-LRR genes predicted from sequenced doubled monoploid potato genome (DM). Such enriched sample is sequenced using Illumina GA2 platform. Next, obtained data are analysed using various bioinformatic tools. Predicted SNP/InDels are confirmed by Sanger sequencing and fine-mapped using segregating populations.



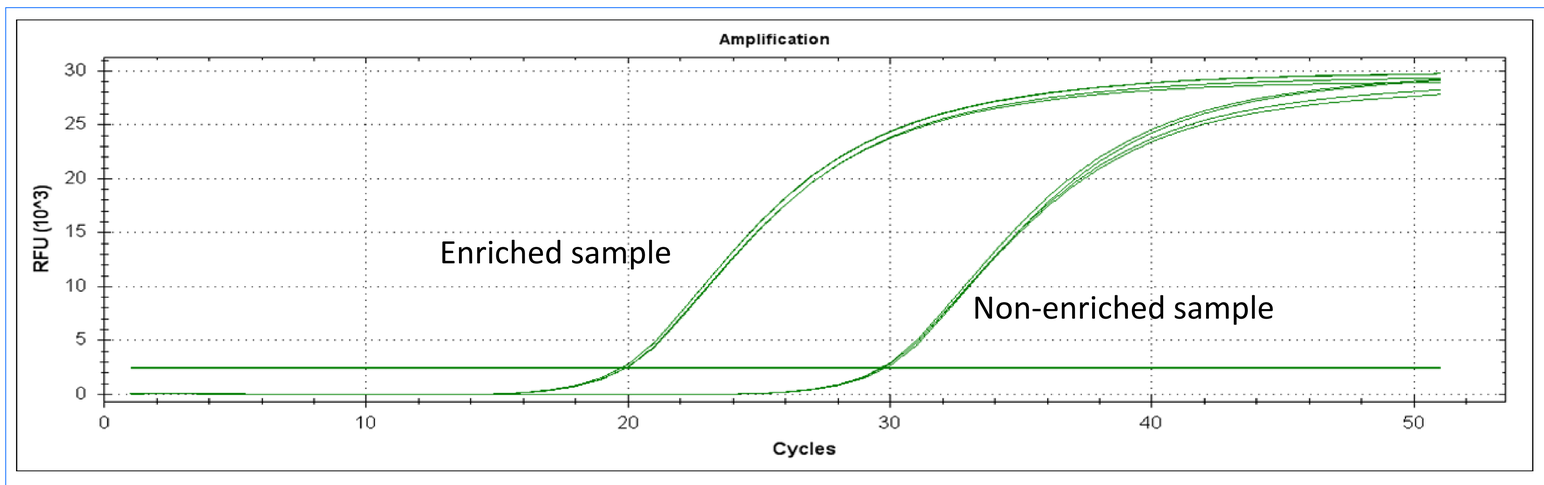


Fig 1. Enrichment efficiency control.
Exemplary qPCR with oligos for *R3a* gene on enriched and non-enriched samples. For different samples and various *R* genes, the difference is between 8-11 cycles, what indicates 250-2000-fold enrichment.

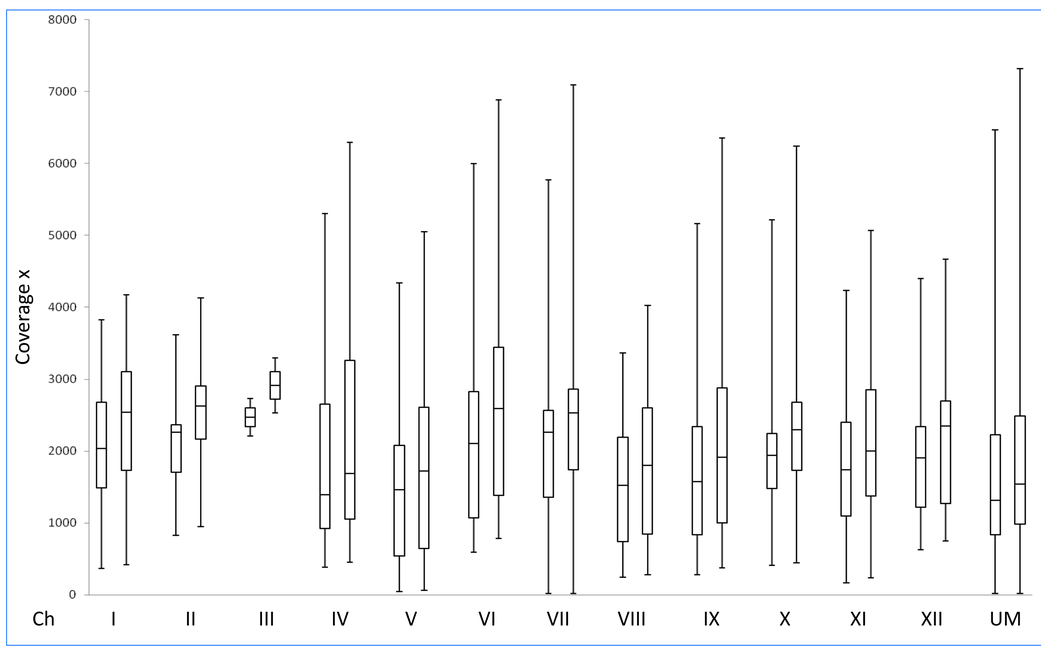


Fig 2. Average read depth per gene per chromosome.
Reads are mapped to DM NB-LRRome using BWA allowing 10% miss-match. Results are normalised to the gene length.
Box plot shows similarly high coverage for most NB-LRRs across all chromosomes (and unmapped genes). Multigenic families are usually higher covered and very few genes have low coverage.
Chromosome, Ch; left bars, BS; right bars, BR; unmapped genes, UM.

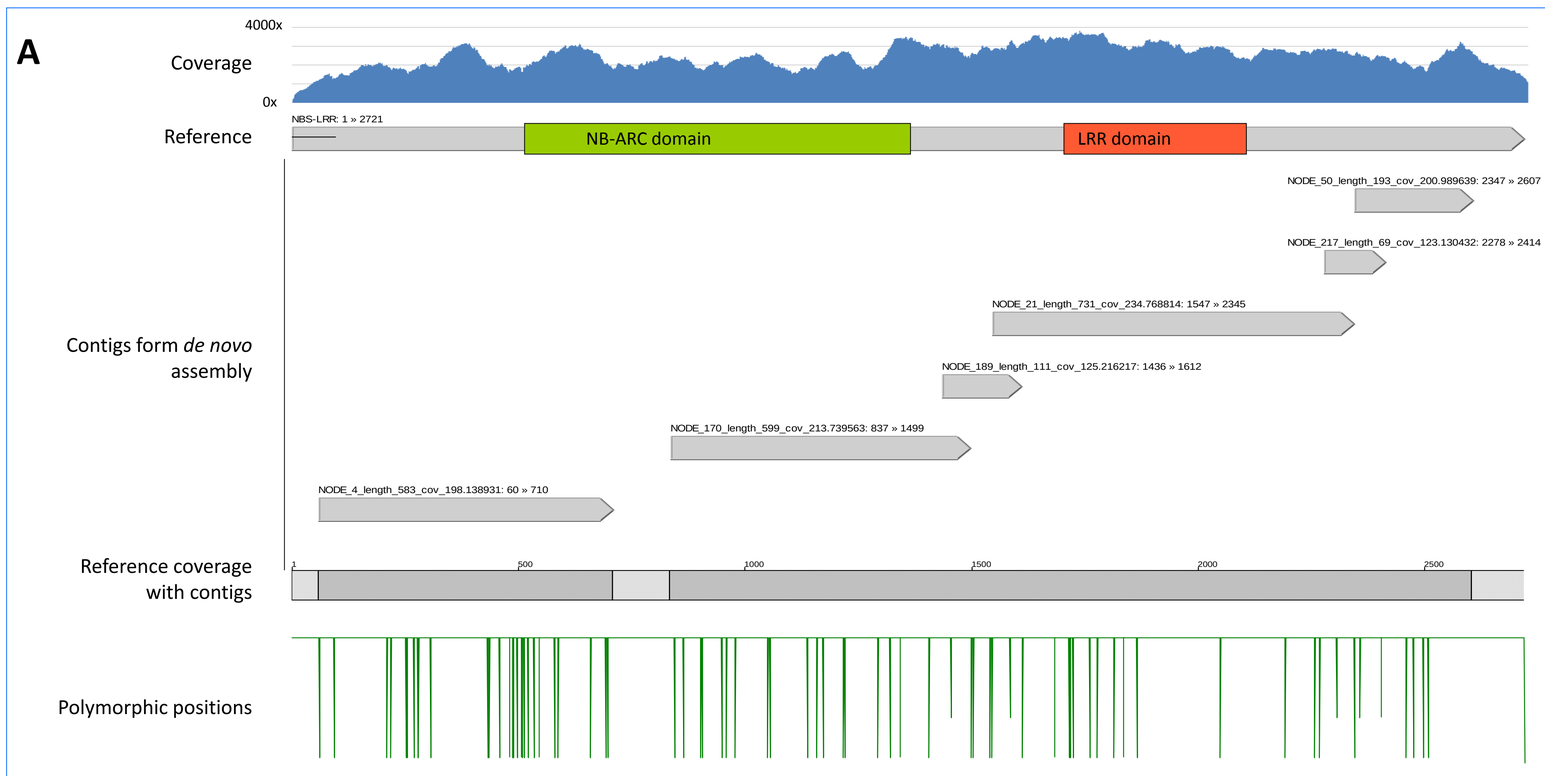


Fig 3. *De novo* assembly of NBS-LRR gene from chromosome XI using BS reads.
Preselected reads for each chromosome are assembled using Velvet with various settings to optimise assembly for N50 and/or number of contigs >1kb. As shown on panel A, even with the best assemblies (with high coverage over whole gene) usually it is not possible to reconstruct gene, probably due to high similarity of *R* genes and high complexity of their loci (an example here is a putative NBS-LRR gene localised on Ch XI). Panel B shows statistics for assembly of all *R* genes on chromosome XI using Velvet.

Fig 4. SNP prediction with new reference.
BS assembly is used as reference for SNP prediction with SAM tools (pileup). Both BR and BS reads are aligned and polymorphism exclusive for BR is detected and later confirmed by Sanger seq. SNP calling remains the biggest challenge, as many predicted SNPs are false positive.

Summary:

- target enrichment against potato NB-LRRome is very efficient
- bioinformatic analysis of obtained data is still a challenge
- although it is not possible to assembly whole *R* gene, shorter assemblies are mostly correct (around 85%, confirmed by Sanger sequencing)
- using presented approach we were able to find markers within *Rpi* genes linked to resistance less than 0.1 cM.

B

Min contig length:	137
Max contig length:	3.304
Mean contig length:	461
N50 contig length:	642
Number of contigs:	166
Number of contigs >=1kb:	16
Number of contigs in N50:	33
Number of predicted NB-LRRs on Ch XI:	56

Funding: BBSRC bioscience for the future; The Gatsby Charitable Foundation