Resistance gene cloning in wheat
23 September 2020

— — —

Q&A session

Presenter: Burkhard Steuernagel, Postdoctoral Scientist, John Innes Centre, UK

The webinar recording is available on the IWGSC YouTube channel at https://youtu.be/QfX4RSVLjlU

Q: Is it possible to clone the gene of interest other than the resistant gene using these MutRenSeq techniques?

In theory, you could design a bait library that targets your genes of interest. You could just use an exome capture. MutChromSeq can be used for any gene.

Q: should the order of motif is the same.. as shown

The order of motifs follows a set of rules, which are the core of NLR Parser and NLR annotator. You find the rules for the NB-ARC motif on Figure 1 of the NLR-Annotator publication

Q: Can you please comment on the cost and how easy/difficult to do a chromosome flow sorting

I know only one lab in the world who can do that. I don't know about the current prices. It is likely more expensive to sequence entire chromosomes than NLR complements.

Q: Is gene cloning different from fine mapping the gene?

Fine mapping is to narrow down to an interval where you expect the gene. Cloning a gene is the next step: Determine the coding sequence and validate it.

Q: What about mutations downstream of the R gene that will not seen in Ren seq?

They will not be seen. Be lucky and have enough of your mutants having the mutation in the gene. My gut feeling is that downstream processes have redundancy to protect from single mutations. This is why we got so many genes with the method. However, I cannot exclude that downstream mutations are one of the reasons for our inconclusive experiments.

Q: Can we use the pipelines with transcriptome data, instead of whole genome ?

I have tried that once and it did not work. Later when the gene was clone with MutChromSeq I checked the data and found poor read support for that gene. In theory it should work but it will suffer when genes of interest are lowly expressed.

Q: Is the method being tried for other gene families and traits?

Yes. cer-q, rht18 and lys3 for mutchromseq. The mutrenseq has been tested for another gene family with another dedicated bait library but that is unpublished.

Q: What about Sr13 which is one of important Sr genes in durum? I think it is also found in Fielder. Sr13 has cloned in the map-based cloning technique by Wang et al. 2017. The sequencing of the BAC revealed two NLR genes CNL2 and CNL4 if i am not wrong

I am not sure what the question is. Your comments are correct, it has been cloned by Dubcovsky lab. Not with any of our methods, I think.

Q: when you denovo assemble the MutRenSeq/RenSeq sequencing data don't you get a lot false/chimeric contigs? R gene is a huge family and I suspect many genes will be very similar, hence it may be difficult to accurately assemble new R gene transcripts. What sequencing technology and assembly tools are you using?

You are correct, this is one of the main problems. Still not fully solved. We use Illumina sequencing and assemble with CLC. The assemblies are not perfect. We did pull out Sr33 though, which is from the large Mla cluster.

Q: I wonder what is the approximate cost of cloning a gene using a MutChromSeq pipeline?

Not sure, this is constantly changing. Consider the size of a wheat chromosome. You want at least 20 fold coverage. Sometimes, the purity of the chromosome is at 60%. That would mean 40% is "off-target". Then you know how many Gb you need to sequence with Illumina PE.

Q: If any transporter gene is responsible for the rust resistant so can we clone that gene by this method ?

MutChromSeq works for any single gene with a clear phenotype.

Q: Is it a good idea to use iwgsc refseq v1. 0 as a reference to identify R genes in progenitor species such as T. boeticum, if we have mapping information available...

RefSeq is always a good start. If that does not help, move to other genomes that have been sequenced.

Q: if the coverage of the sequencing is only low (such as three using the illumina sequencing), it's diffcult to get long contigs or it's difficult for denovel assembly. do you think it's poosible to run the pipeline?

I would think this is too low. With 10 fold you can start thinkin of (still very poor) denovo assemblies. 3 fold is also too low for SNP mining. The coverage is not even, so you have uncovered regions. If your one mutation you are after is in there your experiment is inconclusive

Q: How does the DArTseq genotyping working for AgRenSeq?

I never used DArTseq

Q: What is the optimum concentration of EMS used for mutagenesis?

I am not sure. Here is a protocol for it. https://link.springer.com/protocol/10.1007/978-1-4939-7249-4_17
Maybe that helps?

Q: Though rice being a largely grown cereal, rust infection is not reported yet. have any ideas related to NLR family or any other gene family specific to rice (resistance) or wheat(suseptible) ?

No.

Q: Have breeder friendly markers been developed for each of these genes you have cloned? Where are they being used in resistance breeding in wheat?

We deliver the gene sequence. That can be used to develop the perfect marker.

Q: How to select probable motif number during motif analysis in MAST? Is it based on nature of gene or published data?

My work is based on the motifs defined by Jupe et al 2012. So I have not done this myself yet.

Q: Regarding AgRenSeq and assuming you have more than one resistant individuals within your association panel, how would you recommend to choose that resistant accession for which you do the denovo assembly? How does this deal with the quantitative nature of some resistances (having more than one gene within the panel explaining resistance)?

You start with one resistant accession and move through each of them until you find a good candidate. For our genes, we eventually tested them all. That can easily be scripted with our pipeline so it is not much effort. Quantitative nature…good question. Would you expect these to be conferred by NLRs?

Q: How you control miss-assemblies during deno assembly or didi you performed any quality control steps after that?

Misassemblies happen. Mainly paralogues collaps. That will result in funny allele frequencies. Sometimes it can be resolved by lowering allele fequency thresholds in the pipeline. But it remains a problem.

Q: How do you suggest to proceed if there is more that one candidate gene? If the trait is not explained by a single gene?

Then forward genetics becomes difficult. I guess you need to do some clever crosses to figure out what you are dealing with. Then, if several genes are required you would need more mutants to still have enough in each gene.

Q:What is the best strategy to apply when the candidate gene is located on an alien translocated fragment?

That's where mutational genomics is very good. We create the reference of the wildtype as part of the pipeline.

Q: How many strong functional motifs identified which distinguish TIR, NBS and LRR region of R genes along with other signature sequence like WRKY, CC, NLS, TrD, KINASE etc.?

Check figure 1 of the NLR annotator paper for motifs. There are 3 for a TIR (not reported in the paper). The motifs are defined in Jupe et al 2012. WRKY will not be reported by current version.
There are 2 motifs for CC but they don't find every coil-coil.

Q: have you used transcriptome data

did not work for me as complexity reduction method but I only tried once. It should work unless your gene of interest is too lowly expressed.

Q: How do you take care of slippage in SNP identification?

Do you mean SNPs that are missed by the SNP caller? The pipeline uses an internal over-sensitive SNP caller. The many false positives are got rid of in downstream process.

Q: Is it possible to identify multiple major NLR genes in a panel using AgRenSeq if the panel has been phenotyped using several pathogen races

Yes.

Q: For the EMS mutant, what's the percentage of EMS-type variation? Should it be greater than 90%? If only small part (30%) is EMS-type and most are other types, could we consider this is a contamination but not a real mutant?

If there are that few EMS-type variations only, something is going on. Problems with the assembly, contamination, residual heterogeneity.

Q: Why the number NLRs drop in recent assemblies (slide 5)?

because all these little partial contigs were only fractions of genes. In newer assemblies most of them were joined into the actual genes. There is more going on but that is the major reason.

Q: different methods gives different numbers of NBS LRR gene in genome, how to we finally select true NBS-LRR

it will always change. Assemblies constantly get better. Different methods are often complementary. A final call is only possible with manual annotation. Considering that manual annotation of all NLRs in a genome is a hell of a lot of work (I did one chromosome then gave up) and that that genome is only one accession, who wants to do this? How relevant is the exact number in context of crop improvement?

Q: How to tackle the problem of gaps during chromosome walking in map-based cloning of diseases resistance genes in wheat?

don't do map-based cloning; use our methods ;-)