Decoding the polyploid wheat genome using gene networks
19 November 2020

— — —

Q&A session

Presenter: Philippa Borrill, Lecturer in Plant Biology, University of Birmingham, UK

The webinar recording is available on the IWGSC YouTube channel at https://youtu.be/36W7JUPGpRo

---

Q: could you explain how to develop double mutant population to study candidate gene?

The first step is to select mutant lines - one with a mutation in the A homoeolog, one with a mutation in the B homoeolog. These mutant lines are then crossed together, the F1 selfed and double homozygous mutants can be compared to double homozygous wild type siblings in the F2 population. Usually we would backcross a few times to clean up the background which has many EMS mutations. I recommend reading Adamski et al., 2020 eLife for more detail (DOI: 10.7554/eLife.55646) and also look at http://www.wheat-training.com/functional-studies/ for step by step guides

Q: Why do you have a so big variance in the double homozygous for the NAM-2 gene? DO you think it is possible that in same case there is a sort of redudancy with another gene?

There are certainly other genes involved in senescence, but I think the variation we see in the homozygous mutant in NAM-2 is due to inherent variation in the senescence process which can be affected by slight environmental differences (e.g. between the edge or middle of the bench in a greenhouse)

Q: Is redundancy also observed at pathogen resistant loci?

Yes- this is seen at the MLO locus, but for other pathogen resistant loci (especially those introgressed from wild relatives) the resistance gene may only be present on one homoeolog, therefore redundancy wouldn't be observed in those cases.

Q: Which percentage of the wheat genes are expressed from the three subgenomes?

~50 % of genes in wheat have 3 homoeologs, of these ~70 % have a similar expression level of the 3 homoeologs, while ~30 % have expression mainly from 1 or 2 homoeologs

Q: As Stay green is one of the target trait for heat stress tolerance. If senescence is related with GPC QTLs/Gene. Can we target GPC gene/QTLs for heat/drought stress tolerance?

Interesting idea, I think it may be possible and senescence loci can affect heat/drought stress tolerance (although not all do). There are some limitations to using GPC to look for heat drought tolerance- firstly not all variation in GPC is caused by changes to senescence and some heat/drought tolerance loci will not affect GPC so would be missed by this approach.

Q: Is it possible to apply this method for detection of R genes or their networks...

In principle yes, as long as you can have a time course of gene expression (e.g. during infection) this approach could be applied to many different biological questions. I think Katherine Denby at York has done some work on pathogen resistance using similar approaches but I'm not sure this was focussed specifically on R genes.

Q: What is dosage compensation?

Dosage compensation could refer to when 1 homoeolog can substitue for another – for example if 1 homoeolog is knocked out the other 2 homoeologs in hexaploid wheat may fill in the role of the missing homoeolog. Dosage compensation could occur at the gene expression level or at the protein level.

Q: Knocking out homologous and then going for screening sounds a lengthy process. A brief Guide to the process?

The first step is to select mutant lines - one with a mutation in the A homoeolog, one with a mutation in the B homoeolog. These mutant lines are then crossed together, the F1 selfed and double homozygous mutants can be compared to double homozygous wild type siblings in the F2 population. Usually we would backcross a few times to clean up the background which has many EMS mutations. I recommend reading Adamski et al., 2020 eLife for more detail (DOI: 10.7554/eLife.55646) and also look at http://www.wheat-training.com/functional-studies/ for step by step guides

Q: Which RNAseq data was used for creating the gene networks? Did you use the the TILLING population?

For the genie3 network we used 850 RNA-seq samples from a diverse range of tissues, developmental stages and environmental conditions (see Ramirez-Gonzalez et al., Science DOI: 10.1126/science.aar6089 for more details). We did not use the TILLING population.

Q: Is there any other criteria you know about for filtering the genes prior to building gene networks rather than selecting only the transcription factors? For example, is it advisable to look at suites of genes that have been implicated in the trait of study?

It is possible to focus on genes which are implicated in your trait of interest, however we wanted to take a less biased approach, so we could find new gene families that might be involved.

Q: How much time does it cost to build this gene network ?

This will depend on the cost of RNA-seq and the number of RNA-seq samples. I wouldn't recommend trying this with fewer than 10 timepoints and 3 replicates. There is also the cost of the analysis by someone with bioinformatics skills.

Q: Rust is commonly found on many of the cereals but there are not any report of rust on rice .... why?

Sorry I don't know but I imagine this is due to host specificity.

Q: Why after knockout the all three homeologous genes, we call it a wild type? Should it be the mutants?

After knocking out the three homoeologs we would indeed call it a mutant (usually I refer to them as triple mutants because all three homoeologs are mutated).

Q: Agronomically, delayed senescence in wheat is getting more time to get nutrition and as a result, the grains will get much time to fully develop. But, in a specific environment, e.g. terminal drought, is there any advantages for the delayed senescence?

Yes, the environment in which a plant grows has a strong effect on whether delayed senescence is advantageous or not. In terminal drought delayed senescence can sometimes help improve yield

because it can contribute to extending grain filling, but there is a limit when delaying senescence for too long can have a negative effect on nutrient content (e.g. protein content).

Q: Is there any online tool to study gene regulatory network to select candidates?

Yes- have a look at https://knetminer.com/Triticum_aestivum/ and their recent BiorXiv paper (Hassani-Pak et al., 2020, https://doi.org/10.1101/2020.04.02.017004). This incorporates the genie3 transcription factor-target network which I mentioned (also see Harrington et al., 2020, G3 https://doi.org/10.1534/g3.120.401436) and a co-expression networks that we developed using 850 RNA-seq samples (IWGSC et al., 2018, Science, DOI: 10.1126/science.aar7191)

Q: Can you let us know the link and what data is used as an input in this online tool?

CSI is available through cyverse.org and the input is a matrix of gene expression values (e.g. in tpm)

Q: The RNA-Seq data are from which wheat variety? In eFP browser and wheat-expression.com.

The data in the eFP browser comes from Azhurnaya (spring wheat) and in wheat-expression.com data comes from a range of varieties- you can add a column to indicate which variety the data is from, and filter for your favourite variety. After Azhurnaya, the most samples come from Chinese Spring the reference sequence variety.

Q: Is there isoform divergence between close homeologs in expression?

Interesting question, I think it is possible but this hasn't been studied extensively. The current annotation was not developed specifically to look at isoforms so we think it underestimates isoform diversity so we didn't investigate gene expression data at the transcript level yet.

Q: Do you have any thoughts on how deletion lines (ie. whole chromosome arms) might be useful for studying gene networks in light of some of the new resources/tools you mentioned.

I think deletion lines could be interesting especially if they have small deletions because that could be used to test the effect on the network structure. However, I would be cautious about using whole chromosome arm deletions because so many genes will be missing it will be hard to interpret changes in the network.

Q: Can you speak on how CSI was chosen for network analysis, whether DEGs were included with TFs for the CSI algorithm, and how you view TF versus target genes (ie by GENIE3 prediction) as candidates for causal genes?

We chose CSI because many algorithms cannot deal with large numbers of genes, since we are working in wheat we have lots of genes! CSI seems to be able to handle up to ~350 genes, whereas other comparable algorithms often cease working after ~50 genes. We selected TFs which were differentially expressed during our timecourse of senescence to include as input for the CSI network. At the moment we are focussing our efforts on characterising TFs at the functional level (e.g. making mutants) because we hypothesise these will have strong effects on senescence. However, downstream target genes will certainly be interesting to explore in the future - e.g. using targets identified in the GENIE3 network or through other approaches.

Q: Are genes exhibiting the dosage function contributed by its promoter or protein function or both?

We are not sure but I think several different mechanisms could contribute- promoter sequence, epigenetic marks and also protein turnover.

Q: Do you see a bias towards any of the sub-genomes for regulatory genes?

Not in the networks we have looked at.

Q: Has any group used promiscuous guide RNA for knocking out 2 or 3 copies in the same attempt?

Yes there have been several publications on this - e.g. Zhang et al., 2017, Plant J https://doi.org/10.1111/tpj.13599 . Also, there is a tool to help design promiscuous gRNAs: Cram et al., 2019, BMC Plant Biol https://doi.org/10.1186/s12870-019-2097-z