Chapter 10 – Mapping Markers and Quantitative Trait Loci.

Block 1

The rf should not theoretically exceed 0.5, so values greater than 0.5 are indicative of variation around the value of 0.5. Thus we would probably make the cut-off on a point where it is very unlikely that a lod score is achieved by chance. In this case a lod of around 3, which is roughly where rf=0.25, i.e we draw the cut-off parallel to the x-axis at a lod score corresponding to a point below all but a few outliers on the right-hand side.

Block 2

Stricter threshold results in more smaller linkage groups. Looser thresholds have the opposite effect and in the extreme results in no linkage groups.

Block 3:

Not bad, given the number of markers, but chromosomes 1 and 5 have some very long gaps (>50cM) which suggests a problem.

Block 4:

Yes, unrippled, the Xover number was 543. The best orders with the 3 ripple approaches had 453 and 436 Xovers for ripple4, ripple 7 respectively With ripple 4 with likelihood, the chromosome length was 4609cM cf 5780cM

Block 5:

As error rate is increased, chromosome length decreases greatly and markers become more evenly distributed, up to point where chromosomes have perfectly even distribution. At this point, the signal in your data has been lost. There is clearly an optimum, which should be based on your estimated genotype error rate from examining your data. If your genotyping error rate is very high, your markers are not suitable for making a map.

Block 6:

One set of markers clearly align to two other groups of markers that do not align to each other. This will cause a problem for mapping. In this case, the cause of the problem is that the Avalon x Cadenza cross is segregating for a whole-arm translocation: Cadenza has the reference 5B and 7B chromosomes but Avalon has 5BL/7BL and 5BS/7BS. Chromosomes 1 here is a mix of 5B and 7B, with the odd patterns resulting from some individuals having e.g. the 5BS markers closest to 5BL, and others having the 5BS markers closest to 7BS. Anomalous patterns such as this are useful diagnostics of chromosomal re-arrangements such as this.

Block 7:

As expected, having a more conservative threshold results in more, shorter chromosomes. A more relaxed threshold results in the opposite. Bread wheat has 21 chromosomes, but it would be unwise to target this number for many reasons. Firstly, genome coverage is uneven, especially in wheat where the D genome has very limited diversity and hence marker coverage. If chromosomes have coverage gaps (esp >50cM apart) then they will be expected to map in separate pieces. Furthermore, poor genotypes, or chromosomal re-arrangements, as we have seen, can potentially

link together separate chromosomes erroneously, especially with smaller sample sizes of populations. It is best to use some of the diagnostic tools we have already seen in R/QTl2 to determine what thresholds to set, and in general, better to err on the side of caution: generate a map with lots of small chromosomes and use external information (such as genome sequences) to try and infer which fragments probably are on the same physical chromosome.

Block 8:

QTL on chromosome 9 now has higher LOD score but no more QTLs found.

Block 9 (R/QTL)

More degrees of freedom = lower power for detection but hopefully the other advantages outweigh this – e.g. better resolution, high diversity, but it does mean we need large sample sizes.

Block 10

They are really quite consistent between years, although some differences are apparent., e.g. chromosome 20 (2012 borderline significant but absent 2013, chromosome 7 significant in 2013, not 2012). Heritability is likely to be high if there is this level of consistency between environments.

Block 11

The significance of all QTLs is greatly reduced in the GRID approach. Maybe we are overestimating the significance of QTLs using LOCO, without further skimming. We could also look into this a bit further by doing a strong skim on the data and seeing what we find.

Block 12

The messy peaks on chromosomes 10 and 11 have been called as 2 and 3 separate QTLs respectively. This may be the case, but especially calling 3 QTLs on chr 11 looks a bit of a stretch from looking at the overall patterns. For MAGIC, we have extra information we can use for delimiting QTLs: are the allele patterns the same? If so, it suggests they may not be separate. We will look at these allele effects next, where we will see that on chromosome 10, it may be a stretch to call 2 QTLs. It is worth noting that the QTL separation analysis is dependent on having an accurate map.

Block 13

Chromosome 6: Soissons allele makes RILs flower an average of 2.5 days earlier. Soisson is known to be the only founder carrying the PpdD1 insensitive allele, which brings about early flowering.

Chromosome 10: It looks like Xi19 is carrying an early allele (approx. 2 days earlier) but the data is less clear cut. Partly this is because of an artefact. The 8 MAGIC founders are not equally related (or unrelated) to each other. In fact, Alchemy (ALC) is the daughter of Claire (CLA). Since these founders are closely related, it is often the case that the RILs cannot be accurately called for one or the other. The overall affect is that often, in cases where neither of these founder alleles has a big effect on phenotype, they are “neutral” when added together but they appear to have opposite large affects, such as at the end of chromosome 10. This pattern of opposite effect directions is frequent across chromosomes and makes interpretation of effects difficult. There are many challenges in working with multi-founder populations!

Chapter 11 – Data visualisation.

Flapjack

Map errors would cause an overly long or short interval to be defined, depending on the approach taken. There probably are some map errors here. Genotype errors would most likely lead to over-confidence in a narrow interval, as the error is more likely to disrupt than enhance a pattern. Genotype errors would also lead to map errors. In general, large QTL also have more defined cleaner peaks than more minor QTL. So in general the expectation would be that the process would be easier with larger QTLs.

Helium

There is a little bit of an art to this. I would suggest changing the colour scheme to two colours that are very different. Ensure you are showing the categorical data that you want to inspect. Then search for the line of interest that carries the allele of interest. Work back through the pedigree, if a parent shows the alternative allele, you have got a dead-end.

Trace back the origin in the UK pedigrees of the awn presence allele (in variety Soissons):

The last visible donor (line we have data for) looks to be "Courtot".

RhtD dwarfing allele is in Claire:

Maris-Hobbit seems to be the last donor we have data for. If you follow the parentage of Maris-Hobbit you can trace the potential original donors.

RhtD dwarfing allele is in Alchemy:

Alchemy has a parent which is Claire so we could assume the allele comes from Claire, and thus the Maris-Hobbit line. However, Alchemy could have received the allele from the paternal side by Consort – if you trace this lineage it seems you could go back to a Maris-Hobbit donor anyway.

RhtB dwarfing allele Soissons:

Soissons is a bit of a dead-end, the Iena parent has the alternative allele, so we must conclude the paternal parent HN-35 was the donor.

RhtB dwarfing allele is in Robigus:

Very limited pedigree information for Robigus! We can only see the direct parents, for which we do not have marker data for.

1B_1R introgression (present in Brompton):

In this case the maternal parent does not share the same allele as Brompton, but the paternal parent is an unknown. We must conclude that Brompton gets its 1B_1R allele from the unknown paternal line (CWW-92-1).

1B_1R introgression (present in Rialto):

Some dead ends present here, which means Riebesel-57-41 looks to be the donor, as this is the only plausible route. There could be other donors, that we don't have genotypes for. Riebesel-57-41 looks to of contributed the introgression to a lot of wheat cultivars in the past.

How could use a pedigree like this in your own work:

1. You can visualise the pedigree of a genotype of interest.
2. You can overlay information onto the pedigree such as breeding company, origin, and year of release. If you then overlay marker data (linked to a gene of interest) you can see how widespread a certain allele is based on these factors of interest (has a certain breeding company favoured a certain allele? etc.).
3. You can trace back to the original donor of a specific allele.
4. If you have genotype data for a specific line and both its parents. You can see which allele of a marker was taken forward the next generation. If you compare many of these examples, you can get an idea of past selection on that particular allele.