**Calculation details for Forward breeding tool**

**Step 1: Input parameter file, provided by the user**



Input rules:

**Marker\_group\_name:** User must provide marker group name if missing the marker will be removed from analysis (missing is blank cell)

**Marker\_name:** user must provide marker name (missing or blank cell will lead removal of the marker)

**Fav\_allele:**

1. user must provide the fav\_allele value, if fav allele is missing the marker will be removed from the analysis
2. the favorable allele must match with the allele in the genotype file, if it does not match yield an error

**AlleleName**

If there is no allele name, substitute with fav\_allele value

**PriorityMarker**

1. If marker group has more than one marker and there is no YES, choose first markers in the file as priority marker
2. If marker group has more than one marker and there is multiple YES, choose first markers with YES in the file as priority marker
3. Not required if there is single marker in the marker group

**BreedingValue**

1. If no YES, means NO
2. If all values are NO, no molecular breeding value or index is calculated but allele translation part need to be completed
3. When multiple values for each marker group in mixed (say YES, NO, NO), any one of the values is YES, means that marker group is used for molecular breeding value calculation

**Model**

1. If no dominant, means additive (default is additive) – we may have to add recessive model in future
2. For each marker group if there is mixed values in the model (say dominant, additive, dominant, additive), choose the priority marker (first marker in absence) as correct reference

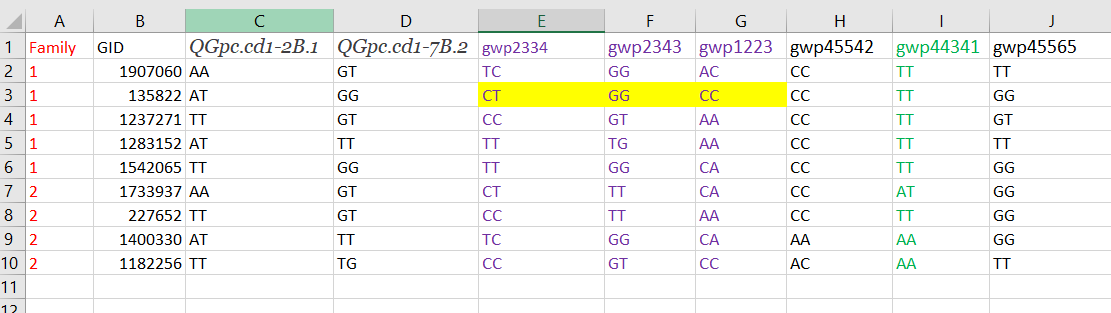
**SubstitutionEffect**

1. If breeding value is YES, there must be substitution effect, otherwise error

**RelativeWeight**

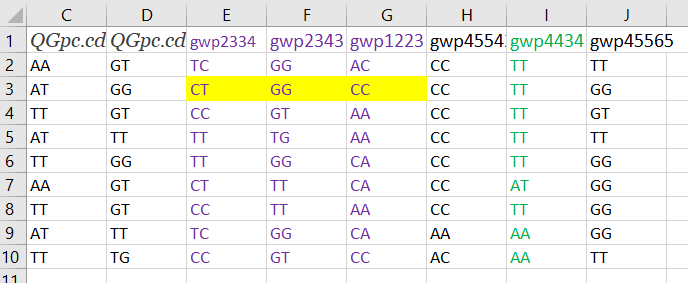
1. If breeding value is YES, there must be relative weight, otherwise error

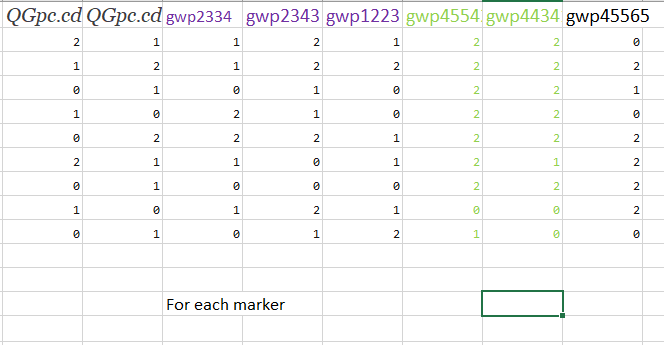
**Step 2: Input marker file**



It is important to note that some of the QTL are presented by more than one marker (maker group or haplotype). Fav\_allele is the one whose substation effects are noted and is allele of interest to breeder / geneticist. Allele name is the one breeder want to see in the output.

**Step 3: Counting dosage for each favorable allele across all markers**



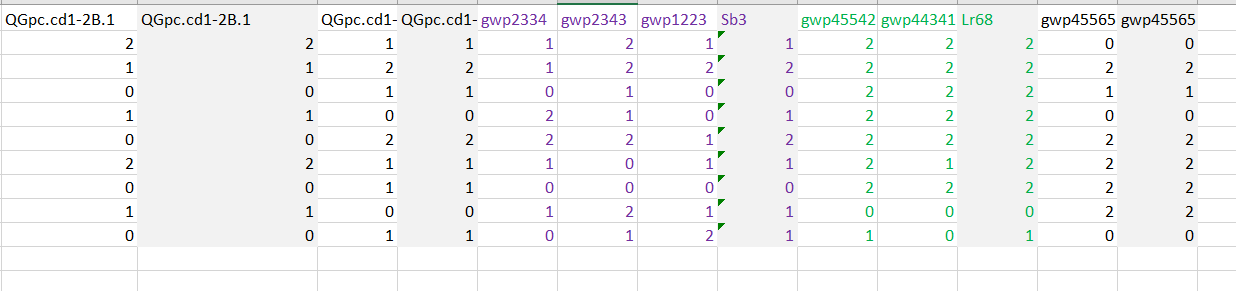


The idea here is just to count specific allele in the markers using excel formula like:

=SUM(LEN(C2)-LEN(SUBSTITUTE(C2,B$15,"")))

The reference here is favorable allele in cell B$15.

Creating consensus for those that are in haplotypes (marker groups)



Here we calculate mode value, if mode value yield NA because no one value is consensus, we replace the values with priority marker (means that marker that breeder thing most useful or highly linked with the QTL), if no marker is designated as most useful may be just yield the first marker. The excel formula here is:

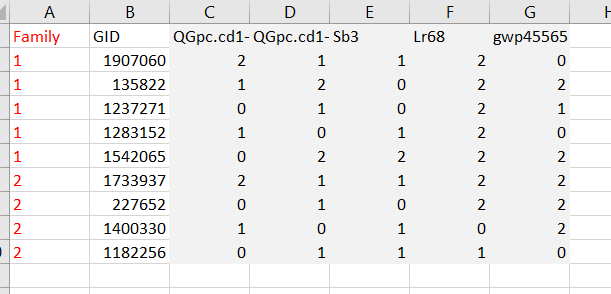
=IFNA(MODE(AA3:AC3),AB3)

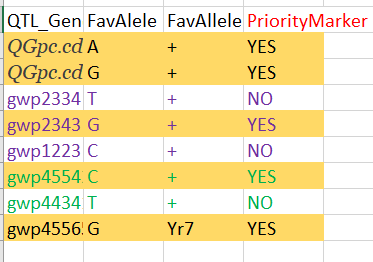
**Dealing with missing values**

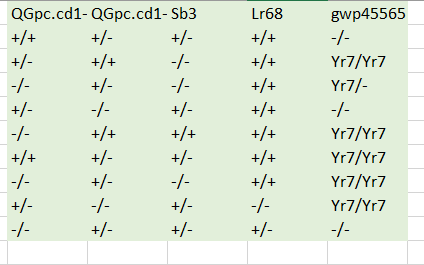
**Consensus building, missing values are removed for example we have values (1,2,NA, 1), the consensus will be 1, example (NA, NA, NA,1), the consensus would be 1, if there is no non-NA, produce is as NA, this will yield Not calculated (NA or NC) in the breeding values.**

**Step 4: First desired output matrix**

After calculating the numeric dosage matrix of favorable allele, we would like to recode to produce the QTL genotype as breeder would like to see:



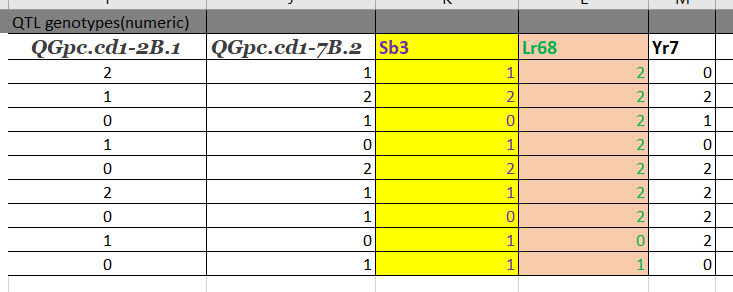


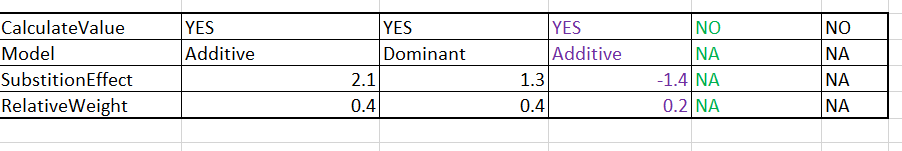


The example excel formula is:

=IF(C2=2,C$22&"/"&C$22, IF(C2=1, C$22&"/-","-/-"))

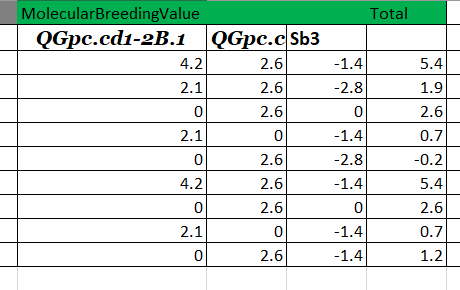
**Step 5: Calculation of Molecular breeding value**





With the formula like: =IF(I$14="Additive",I3\*I$15,IF(I3>0,2,0)\*I$15)

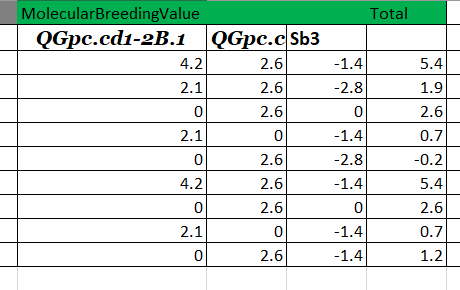
If model is additive it simply multiplies the dosage value with the substitution effect, if not ( alternative is dominant), it recodes the value greater than 0 as 2 before multiplying, thus heterozygote and homozygote will have same value 2. The dominant model assumes the favorable allele indicated is dominant.

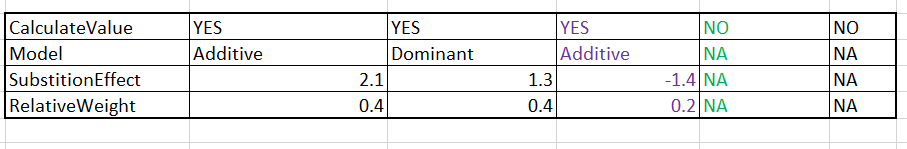


The sum is total value for all loci that are used to calculate molecular breeding value. At backend it can be all loci but set allele substation value 0 for those loci that do not go into the index.

**Step 6: Weighted Molecular Breeding value / Index**

Using the output table in step 4 and multiplying this with the relative weight, we can get this value.





The Multiply with relative weight with simple multiplication formula:

=(O3\*I$16)

