

Construction of a consensus map of a nested association mapping (NAM) population of Asian common wheat and its application in revealing recombination landscape and genetic analysis of flowering duration
(アジアのパンコムギの NAM 集団のコンセンサスマップの構築と、その遺伝的組換えおよび開花期間の遺伝学的解析への応用)

Nie Jilu

Laboratory of Plant Breeding
Graduate School of Agriculture

General introduction

Common wheat (*Triticum aestivum* L.) serves as a primary food source for humans. Gene mapping in various crop species has identified numerous genes for yield and stress resistance, but its application in common wheat (genome constitution AABBDD, $2n = 6x = 42$) is still challenging due to its vast genome (~16 Gbp) and highly homoeologous subgenomes, which hampers genetic analyses. The concept of nested association mapping (NAM) populations was developed to combine the advantages of bi-parental and association mapping populations, which was initially established in maize (Yu et al., 2008), offering enhanced allelic variation and abundant recombination events while also minimizing the confounding effects of population structure. However, establishment and utilization of wheat NAM populations are still limited (Bajgain et al., 2016; Ren et al., 2017; Jordan et al., 2018; Ahmad et al., 2020; Priyanka et al., 2020; Sandhu et al., 2021; Donaire et al., 2023), highlighting the need for further analysis of wheat NAM populations. Now, an Asian common wheat NAM population consisting of 1,060 RILs generated from crosses between the common paternal accession ‘Norin 61’ (N61) and maternal 13 Asian wheat accessions was established (Nitta et al., in preparation). Utilizing the NAM population, I performed three analyses to advance genetic studies of agronomically important traits in wheat.

1. Construction of a consensus map using GRAS-Di markers

As a platform of high-resolution QTL analysis, a consensus map is necessary for the whole NAM population. Therefore, I constructed a consensus map of whole genome genotyping data with GRAS-Di (genotyping by random amplicon sequencing-direct) technology (Hosoya et al., 2019). By tidying the genotype data obtained from SNP calling, I got a total of 18,339 high-quality SNP markers. Of these markers, 6,467 SNP markers with the least missing rate in each haplotype were selected to represent the haplotypes in each NAM family and integrated into the consensus map. This consensus map spans an average of 99.2% of the physical length of each chromosome of the N61 reference genome, with a length of 3,699 cM and 3,844 unique map positions. Collinearity of the marker orders of the consensus map against physical positions on the N61 reference genome (Shimizu et al., 2021) was high, with Spearman’s rank correlation coefficients ranging from 0.911 on chromosome 4D to 0.999 on chromosome 3B. QTL analysis for heading time and flowering time demonstrated the power of the consensus map in detecting responsible loci so far reported. I detected five QTLs on chromosomes 2B, 2D, 5B, and 7B, which are close to the known gene position of *Ppd-B1*, *Ppd-D1*, *Vrn-B1*, and *FT-B1*, respectively (Welsh et al. 1973, Scarth and Law 1983; Worland 1996; Guo et al. 2010). The confidence intervals of QTLs detected using the consensus map were narrower and more accurately covered the reported functional genes than those detected in a single map.

2. Revealing the landscape and genetic architecture of genome-wide recombination

In breeding, researchers strive to introduce a favorable allele while minimizing linkage drag of neighboring undesirable alleles. Understanding the recombination landscape can help breeders to estimate the appropriate population size to introduce new allele and to break up the linkage of favorable and undesirable loci. Based on the SNP genotype data, I investigated the genome-wide recombination landscape. To count the crossovers (CO) events, collapsing SNPs based on haplotypes into blocks is a common practice (Gardiner et al., 2019; Sehgal et al., 2020; Brinton et al., 2020). Based on the average wheat haplotype lengths that analyzed the diverse CIMMYT spring bread wheat population (Sehgal et al., 2020) and the median haplotype length for the accessions of the 10+ Wheat Genome Project (Brinton et al., 2020), I adopted 20 Mb slide-window with a step of 10 Mb to estimate the CO numbers and positions.

A region-by-region analysis was performed for each chromosome, with respect to the five chromosomal regions originally defined by the IWGSC (2018), which were labeled as “R1”, “R2a”, “C”, “R2b” and “R3”. Consistent with previous reported recombination hotspot regions (IWGSC 2018; Gardiner et al., 2019; Brinton et al., 2020; Brinton et al., 2020; Sehgal et al., 2020), I found that the COs significantly favor the telomeric R1 and R3 regions in the NAM population. Among all the chromosomes, the recombination events were most frequent on chromosomes 5A, 5B, and 5D with 3.0, 2.6, and 2.5 COs, respectively, and were most rare on chromosome 4D with 1.0 COs. To explore the factors affecting CO numbers, linear regression analysis was performed on the CO numbers with the physical lengths of different chromosomal regions on different chromosomes, and with the pairwise genetic distances (PD) of

parental accessions. As a result, the CO frequencies were found to be positively correlated with the physical length of the corresponding chromosomal regions R1 and R3 ($R^2 > 0.8$, $p < 0.01$), while no significant correlation ($R^2 < 0.8$, $p > 0.05$) was found between the PD and the numbers of COs in all chromosomal regions.

Using the CO numbers of each NAM individual as a phenotype, I detected three QTLs on chromosomes 2A, 4A, and 7A controlling the CO numbers in the single NAM families of 'Nobeokabouzu' $\times$ N61, 'PK1' $\times$ N61 and 'Chinese Spring' (CS) $\times$ N61, respectively. As a verification of the QTL result, I observed the behavior of chromosome pairing during meiosis in three RIL lines in the CS $\times$ N61 subpopulation each possessing the N61 allele (CO promoting) or the CS (CO suppressing) allele. According to previous studies that suggested the 'double' chiasmata may occur in the same arm (Gennaro et al., 2012; Dreissig et al., 2017), I recorded the paired chromosomes configuration as single-chiasmata rod (X.rod, one crossover), double-chiasmata rod (XX.rod, two crossovers), one-one chiasmata ring (X.X.Ring, two crossovers), one-double chiasmata ring (X.XX.Ring, three crossovers) and double-double chiasmata ring (XX.XX.Ring, four crossovers) to indicate the putative COs per cell (Gennaro et al., 2012; Dreissig et al., 2017; Draeger et al., 2023). For each selected RIL, at least 30 pollen mother cells (PMCs) at metaphase I of meiosis were observed. The average CO numbers for the low recombination RILs were 47.2, 47.2, and 47.8, respectively. While, the numbers were 52.3, 52.6, and 53.7 for the high recombination RILs, respectively. The Kruskal-Wallis test showed no significant difference within low recombination RILs ($p = 0.716$) nor within high recombination RILs ($p = 0.260$), but a significant difference was found between low and high recombination RILs ($p < 0.01$), where the difference of average putative CO numbers was 5.4 between low and high recombination RILs. Furthermore, significant differences in the CO numbers for all types ($p < 0.01$) except for the XX-rod ($p = 0.250$) were indicated by Wilcoxon rank-sum test, with X-XX rings and XX-XX rings being more prevalent in the high recombination rate RILs. Combining genomic and cytological analyses, I showed that recombination frequencies vary among wheat accessions and detected so far unreported candidate QTL loci, providing a basis and tools for future genetic and breeding studies.

3. Genetic analysis of flowering duration

The flowering duration (FD) in wheat is defined as the period from the first tiller flowering to the end of flowering in a plant. Flowering progression was recorded by randomly selecting ten tillers and recording the percentage that flowered. This was repeated three times daily until the average percentage reached 90%, which was recorded as the end of flowering day. In the NAM population, 'Shirokagekomugi' $\times$ N61 RIL subpopulation have the shortest average flowering duration of 10.7 days, while the 'BT1' $\times$ N61 RIL population had longest of 14.5 days. Linear regression showed a weak correlation between flowering duration and time ($R = 0.42$, $p < 0.01$), indicating they are not totally under the same genetic control. QTL analysis revealed one QTL on chromosome 1A from the 'Shirokagekomugi' $\times$ N61 RIL population consistently in two-year observations. As the NAM founder lines are from a diversity panel of hexaploid wheat core collection consists of 185 globally collected accessions (Takenaka et al., 2018), I performed same analysis on the diverse hexaploid wheat collection to find undiscovered QTLs for flowering duration. In the core collection, flowering duration and time were not correlated ($R = -0.16$, $p = 0.04$). Welch's t-test was performed founding that cultivars had significantly shorter FD compared to landraces ($t = -3.29$, $df = 68.09$, $p = 0.002$). Further dividing cultivars into modern and old cultivars based on the Green Revolution, the average FD for modern cultivars was 10.8 days and 13.8 days for old cultivars. Interestingly, no significant difference was found between old cultivars and landraces ($t = -0.542$, $df = 18.72$, $p = 0.594$). However, modern cultivars had significantly shorter FD than both old cultivars ($t = -2.188$, $df = 42.28$, $p = 0.034$) and landraces ($t = -3.605$, $df = 41.55$, $p = 0.001$). These results suggest that shorter FD has been selected for in modern breeding. In the GWAS analysis conducted on hexaploid wheat core collection, I identified two significant SNPs associated with flowering duration on chromosomes 1B and 7B. These loci accounted for an average of 17.5% of the dominant effect. The integration of QTL and GWAS approaches provided comprehensive insights into the genetic factors influencing flowering duration, highlighting novel loci that could serve as valuable tools for future wheat breeding in the face of environmental challenges.

General discussion

In this dissertation, I successfully constructed a consensus map of the NAM population using GRAS-Di technology. The utility of this map is demonstrated through the well-characterized heading trait. Using the map and cytological observations, I revealed that the recombination rate in wheat is variable and to some extent controlled by genetic factors. The detected QTLs may serve as regulators for the recombination rate in wheat. Additionally, the genetic factors related to flowering duration identified using this map can be utilized as potential materials and tools for adjusting flowering duration by manipulating allele combinations. In summary, this dissertation highlights the significance of constructing and utilizing a consensus map of the NAM population.