Molecular characterization of tin-1 locus of F2 NIL population in winter bread wheat

This work is submitted as a dissertation in partial fulfilment for the award of the degree of

Master of philosophy

In

Plant Sciences

DISS
BIO
5798

By

Arif Ali
Reg. No. 02041613019

Department of Plant Sciences
Quaid-I-Azam University
Islamabad, Pakistan
2018

List of Tables

Contents

List of abbreviations

AFLP	Amplified Fragment Length Polymorphism
ANOVA	Analysis of variance
BAC	Bacterial Artificial Chromosome
BC	Back Cross
cM	CentiMorgan
CI	Confidence Interval
DH	Double Haploid
DA	Day of Anthesis
Dart	Diversity array Technology
GXE	Genotypic x Environment
GDP	Gross Domestic Production
GC	Genomic control
G no	Grain number
GY	Grain Yield
Kb	KiloBase
Mya	Million Year Ago
NGS	Next Generation Sequencing
NIL	Near Isogenic Line
PCR	Polymerase Chain Reaction
PH	Plant Height
RAPDs	Random Amplified Polymorphism
RIL	Recombinant Inbred Line
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
SD	StandardDeviation
T/h	Ton Per Hectare
TGW	Thousand Grain Yield
TKW	Thousand Kernel Weight
USDA	United State Department of Agricultural

Abstract

The tillering inhibition gene (tin) that reduces tillering in wheat (*Triticum aestivum*) is also interlinked other yield components. The present research was conducted to study the effect on tin1 locus other yield components in bread wheat. Using BC1F2 line Cap_20_1_1 fixed line with 87% Caphorn (recurrent parent), reciprocal backcrosses were made with Caphorn at Quaid-i-Azam University Islamabad. Phenotypic screening of 800 BC2F2 populations revealed segregation of Tiller number, Plant height, Spike length, Grain per spike, grains per spikelet, TKW and GY. Significant variation was found in tiller number among the population ranging from 1-17 tillers. In BC2F2 population we identified 190 (24%) individuals with less than 4 tillers, whereas, 610 individuals (76%) with more than 5 tillers. Tiller number had significant positive relationship with PH(r=0.32) spikelet per spike(r=0.24) and GY(r= 0.34) .Whereas tiller number had a significant negative correlation with Spike length (r= -0.005) Grain per spikelet (r=-0.30) and thousand kernel weights (r= -0.28).Like tiller number, other agronomic traits are also segregating according to Mendel law of segregation. In order to determine the molecular basis of tin, we constructed fine map QTL from BC1F1 (heterozygous). This fine mapped QTL corresponded to 3.5Mbp region on the physical map and had 12 genes on it. Whereas, a Micro RNA was identified to be overexpressed in the zone by our collaborators in France. To validate F2 mapping population was constructed to create recombination between the two genes.

Keywords: QTLS Mapping, BCIF2, *Tin Gene*, Wheat, yield, *Tin 1* locus.

1 Introduction

The world population is growing at an exponential rate (figure1) from 3 billionin1960 to 6 billion in 1999 to 7.6 billion today. It is expected to reach at an alarming of 8.6 billion by 2030 and 9.8 billion in 2050 mainly in developing countries in Asia and Africa (US Census, 2018). This rapid flux has increased the ratio of the population affected by insufficient food, i.e. 816 Million people. In order to overcome the increased food demand due to the growing population, the yield of cereals is currently a major global priority (FAO 2018). Among cereal, wheat is considered the ultimate custodian of global security due to its wide range of adaptability. According to FAO 2017, the wheat output estimated globally was 754.8 million tons, which is a lower percentage than the previous year. Dimensioning land and water resources due to rapid urbanization have made the improvement of wheat yield to meet the rapidly growing population a daunting task. (Stratonovitch and Semenov, 2015; Allen *et al.*, 2017).

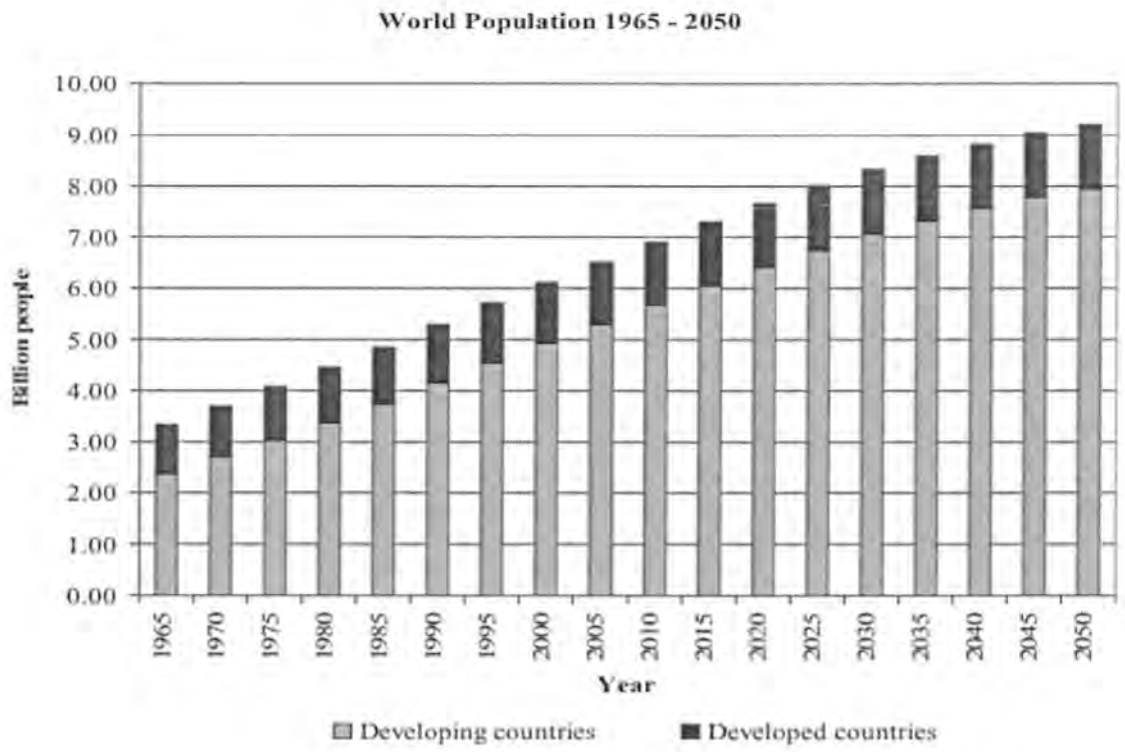

Figure 1 Graphical representation of the global population in developing countries in relation to the developed ones.

1.1 Wheat ultimate custodian of food security:

Agricultural crops, especially cereal grains, continuously play a central role to satisfy food demand of growing population as the significant proportion of their nutrient demand relies on cereal crops. In this respect, wheat is particularly important for several reasons (Shiferaw *et al.*, 2011). It comprises primarily of two types; Bread wheat (*Triticum aestivum*) shares 95% wheat production, while durum wheat (*Triticum durum*) shares 5% another account (Peng *et al.*, 2011). Wheat is the staple food of many countries; rich in basic nutrients such as protein, carbohydrates, and minerals for 30% of human population. (Curtis *et al.*, 2002).Photochemical analysis of wheat suggested that it is the significant source of mineral contributing 44% daily intake of which iron (25%) and zinc (11%) (Henderson *et al.*, 2007). The unique feature of wheat is its adaptation across a wide range of climatic conditions, nearly all-region from the equator to temperate land (varies latitude of 30°N to 60°N and 27°S to 40°S up to 3,000 meters above sea level), and with a temperature between 3° and 32° Celsius. Wheat is the main source of income for millions of small-scale farmers living in developing countries across the globe (Singh *et al.*, 2008; Gøtke *et al.*, 2015).

1.1 Evolution of Wheat

Among cereal crops, wheat is one of the complex genomic organism, global food crop, and a wide range of climatic adaptability. It is stapled cereal crops which covered 40% of the world population, especially Europe, North America, and the western part of Asia. (http://www.faostat.fao.org; http://www.croptrust.org).

The modern wheat *Tritium avestium as* result of two independent polyploidization event. The tetraploid A^uA^uBB and A^uA^uGG wheat originated through independent allopolyploidization events between two wild ploid grasses (Leonard and Martin, 1963; Kilian *et al.*, 2006).Strong evidence point to the wild out crossing *Aegilops speltoides* (SS) (or a genotype similar to it) as the female parent of tetraploid wheat and to wild *T. urartu* (A^uA^u) as the male parent (Dvorak and Zhang 1990);(Huangetal.2002);(Zhangetal.2002);(Kilianet2007a).as shown in figure 2.

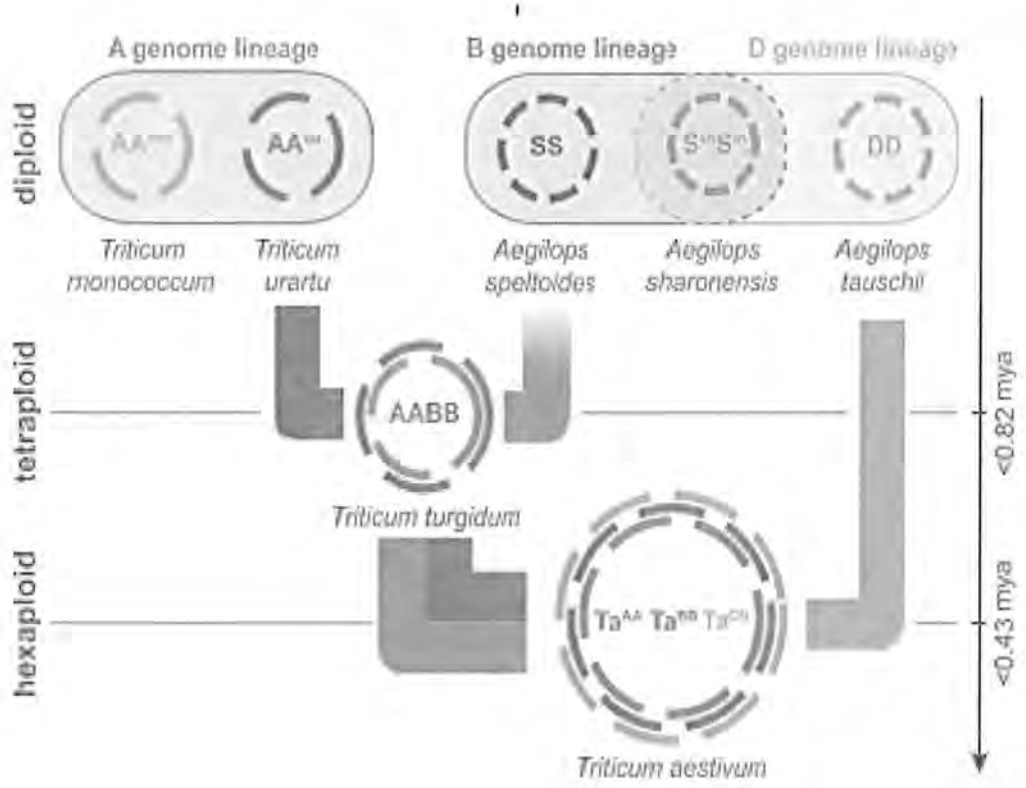

Figure 2 The schematics representation of the relationship between wheat genome and the polyploidization history and genealogy

Names and nomenclature for the genomes are indicated within circles that provide a schematic representation of the chromosomal complement for each species. Time estimates are from (Marcussen *et al.*, 2014), million years ago.

1.2 Wheat Production across Globe:

Globally, major wheat producing countries include China, India, and Russia which produced 129.0, 97.0 and 77.0 million tons wheat respectively (FAOSTAT, 2017). Among them, Pakistan ranked at 8th position in wheat production as shown table 1.

1.3 Wheat production in Pakistan:

Wheat is one of the important agricultural crops in Pakistan, as 80% farmers growing it on an area of around 9.0 million hectares during the winter or in "Rabi" season. Marketing year (MY) 2018/19 wheat production is forecasted to 26.3 million metric tons, one percent lower than the record production from a year ago. Wheat production area province wise is shown in figure 3.

Table 1 Wheat production: leading producers (million tons)

Countries	Average	2016	2017 estimates	2018 estimates	Change: 2018 over 2017 (%)
European union	152.3	144.5	152.0	145.3	-4.4
China	129.6	128.8	129.8	129.0	-0.6
India	92.4	92.3	98..4	97.0	-1.4
Russian	73.6	73.3	85.8	77.0	-10.3
United state of America	55.4	62.8	47.4	47.0	-0.8
Canada	29.9	32.1	30.0	30.0	0.1
Ukraine	26.2	26.1	26.5	1.5	1.5
Australia	26.0	34.4	21.2	26.0	22.4
Pakistan	**25.4**	**25.5**	**25.8**	**26.0**	**1.0**
Turkey	21.6	20.6	21.5	21.0	-2.3
Argentina	16.1	18.4	18.5	16.5	-10.8
Kazakistan	14.2	15.0	13.9	13.8	-0.4
Egypt	9.1	9.0	8.8	9.0	2.3
Uzbekistan	6.9	6.9	6.9	6.9	0.0
Other countries	59.4	58.9	57.5	61.0	6.1
World	750.3	759.8	757.0	744.0	-1

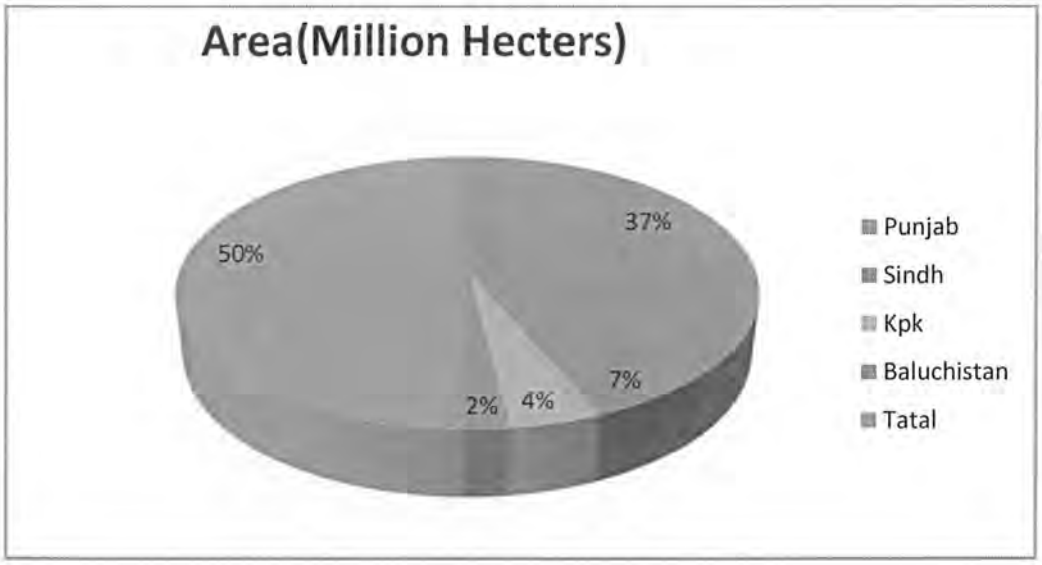

Figure 3The graphical representation of wheat production per province (percentage). Sources: Pakistan Grain and Feed Annual Report 2018

1.4 Winter wheat growth and developmental stage:

Wheat is the most important annual cereal monocotyledonous plant. On the basis of growth and development winter wheat is divided into following stages; Growth stage 1 (vegetative growth); Growth stage 2 (Heading to flowering); Growth Stage 3 (Grain filling to maturity), (Hanft and Wych, 1982).

Growth stage 1:

The vegetative growth is divided into 7 states from germination: germination, seedling development, tillering, stem elongation, boot, head emergence, flowering. Once the embryo is fully imbibed, plant development is resumed. With the resumption of growth, the radical and coleoptiles emerge from the seed. The first three seminal roots are produced and then the coleoptiles elongate pushing the growing point toward the soil surface. The seedling stage begins with the appearance of the first leaf and ends with the emergence of the first tiller. The emergence of the fifth leaf is normally accompanied by the appearance of the second crown tiller which originates from an auxiliary bud located in the node at the basis of the second leaf. Each tiller which is produced represents the potential for a wheat plant to develop an additional stem with its own leaves, roots, and head. A major change in the development of the wheat plant occurs at the end of the tillering stage. At this time, the growing points of the main shoot and tiller stop initiating new leaves and start producing reproductive structures. Conversion of the growing point signals the end of the vegetative stage and the start of the grain development.

Growth stage 2:

Simultaneously, the internode region elongates, moving the nodes and the growing point upward from the crown to produce a long stiff stem that will carry the head. Then the developing head within the sheath of the flag leaf becomes visibly enlarged, a stage known as booting stage. The booting stage ends when the first awn emerges from the flag leaf sheath and the head starts to force the sheath to open. Afterwords from the time of emergence of the tip of the head from the flag leaf sheath to when the head has completely emerged but has not yet started to flower is known as "heading stage", which is followed by "flowering or anthesis stage," which lasts from the beginning to the end of the flowering period. Pollination

and fertilization occur during this period. All heads of a properly synchronized wheat plant flower within a few days and the embryo and endosperm begin to form immediately after fertilization.

Growth stage 3:

The wheat grain development is generally divided into three phases spanning about four to six weeks under normal conditions. In the first phase, cell division phase (commonly called as "watery ripe" or "milk" stages), the number of cells in the endosperm (the major starch and protein storage portion of the kernel) is established. Not much weight is accumulated during this phase. Then 1 to 2 weeks after pollination, during the grain filling phase (commonly known as dough developmental stage), kernel accumulates most of its dry weight. The transport of nutrients from the leaves stems, and spike to the developing seed is completed during this stage. The developing kernel reaches the physiological maturity at the end of the grain filling phase even though it still contains approximately 30 percent of water, which hardens the kernel dough. This is the time when the final kernel weight is obtained. Finally, in the desiccation phase (commonly known as ripening phase), the growth of the kernel declines about three weeks into grain filling and reaches its maximum weight. As the kernel approaches maturity, it ripens i.e. it consistently losses its moisture to becomes a "hard kernel".

1.5 Grain yield and its component:

The main objective of wheat breeding is to achieve high yield (Jiang et al., 2014). In the 20th century, grain crops were improved through breeding by selection of high yielding lines. However, the current yield is not increasing to meet our food demand, therefore further knowledge is required to identify and incorporate alleles for high yield to secure global food insecurity. (Pedró *et al.*, 2012; Simmonds *et al.*, 2016).

Yield is defined as the annual outcome of a crop. It is a complex trait affected by different physiological aspects and strongly controlled by complex genetics. In order to meet desired grain yield, we need to better understand morphological aspects of wheat. (Gupta et al., 2008). The physiological trait associated with grain yield was first determined by (Slafer

and Rawson, 1994). Grain yield mainly depend on tiller number, grain weight and grains per spike.

$$GY = T.No^* GW^* G/S$$

Where GY = grain yield; T.N=Tiller number, GW= Grain weight, G/S= Grains per spike

1.6 Tiller inhibition (Tin) gene as an important component of Grain Yield

Grain yield of the wheat is extremely hampered in spring wheat due to low tiller number (Kebrom et al. 2012). There are several studies conducted on tiller number in common bread wheat. Tiller number is inhibited by a locus called *tin locus*. These studies on tiller inhibition have detected the presence of *tin* QTL on the short arm of chromosome 1A as well as long arm of 3A. Tin-1 on chromosome 1AS is linked with hairy glumes character denoted as Hg, positioned at distal region of chromosome. Similarly, (Kuraparty et al., 2006) target *Triticum monocum* mutant called *tin3*. *Tin3* mutant as recessive has the characteristic lack of tillering linked to spur 1205 marker.

The tiller no varies in spring and winter bread wheat. It is a complex trait dependent on genetic background and environmental factors. The degree of reduced tillering pattern mainly depends on genetic background (Duggan *et al.*, 2005) According to (Duggan *et al.*, 2002), effect of tin gene on spring season is different from winter season as number of tillers had reduced 30- 40% in summer and 50% reduction in winter season. When planted tin line in winter across all locations, they produce 2.3 tillers per plants while non-tin population produced 4 tillers per plant. In contrast, summer planted in line develops 2 tillering per plant while non-tin line developed 6 tillers per plants.

After tillering second major characteristic of the tin gene, that directly affects yield component is larger spike length (Gaju *et al.*, 2009). The Suengaga et al., 2005 selected oligoculum a source of reduced tillering line and observed physiological parameters under growth and field experiment, he reported that under controlled conditions in a glass house, reduced tillering line, spike length was 4.5 cm longer than free tillering line and 6.7cm longer when sown in the field. Similar findings were reported by (Heidari *et al.*, 2011) that oilgoculum had longer spike compared to free-tillering lines.

Beside longer spike length, tin line with reduced tiller number has more spikelet per spike. (Gaju *et al.*, 2009) reported that wheat genotype with tin locus has large spikes population with more spikelet per spike than free-tillering genotypes with a large spike (20.1) in growth chamber and field experiment.

Table 2 Summary of previous studies comparing traits containing the *tin* gene vs free-tillering lines.

Trait	*tin* lines	Free-tillering lines	Citation
Number of spikes/plant	6.2	9.3	Gaju et al. (2009)
as winter planted	2.3	4	Duggan et al. (2002)
as spring planted	2.6	6.4	Duggan et al. (2002)
Spikes/m^2	197-383	305-521	Mitchel et al. (2012)
Spike length (cm)	12.3	8.7	Gaju et al. (2009)
in greenhouse	10.4	5.9	Suenaga et al. (2005)
in field	16.8-17.2	10.3-10.6	Suenaga et al. (2005)
Number of spikelet's/spike	26.6	20.1	Gaju et al. (2009)
Number of seeds/spike	53.5	43.8	Gaju et al. (2009)
Plant height (cm)	65-67	76-81	Richards (1988)
Seed weight (mg)	38-42	35	Gaju et al. (2009)
	28.2	25.7	Mitchell et al. (2012)

As a first step towards isolating a candidate gene for tin, the gene was mapped in relation to molecular markers. Tightly linked flanking markers were identified that will assist in future development of a high-resolution map.

1.7 Genetic markers:

Genetic markers are a sequence of DNA that can be tagged on the chromosome, such sequence may be single base pair single nucleotide (SNP) or short fragment of nucleotide-like bi, tri, tetra, like simple sequence repeat (SSR) that exhibit polymorphism in two genotypes or lines. Genetic markers used various purpose like genetic variation, the pattern of heredity, or evolutionary studies and widely used in plant science toward genetic and physical mapping of the genome. Beside genetic markers are used for identification of gene controlling various parameter and phenotypes (traits association), genetic diversity in marker- assistant breeding for crops improvement. A genetic marker is classified into various on based on their function.

- Restriction length polymorphism (RFLP)

- Amplified fragment length polymorphism (AFLP)
- Single nucleotide polymorphism (SNP)
- Simple sequence repeat (SSR)
- Diversity arrays technology (DArT).

1.7.1 Application of simple sequence repeat (SSR) in plant breeding:

Simple sequence repeat also term as microsatellites, short tandem repeat sequence 1-6 nucleotides frequently present in the genome of wide taxa (Beckmann and Weber, 1992). Microsatellites are classified in various categories on the base of repeats nucleotide base pairs, mononucleotide (A) nucleotides (GT), trinucleotide (ATT),tetranucleotide (ATCG up to hexanucleotide TGTGCA)(Weber, 1990). Besides nucleus microsatellites are also distributed in chloroplast and mitochondria (Provan *et al.*, 2001)(Rajendrakumar *et al.*, 2006).

Various studies suggest that SSR markers have less repetition per locus and high level of polymorphism, (Zane *et al.*, 2002) therefore easily detectable by PCR (Kalia *et al.*, 2011). Microsatellite markers are often derived from the non coding sequence of the genomic region like bacterial artificial chromosome (BACs), development of SSR markers is expensive and laborious (Mir *et al.*, 2013) they are co-dominant and inheritable, therefore mostly abundant in the genome. (Tautz, 1989; Kalia *et al.*, 2011). Microsatellite detections or implication are PCR based and easily run on an agarose gel, acrylamide or capillary sequencer. SSR markers are widely used in plant breeding programs. Figure 7.

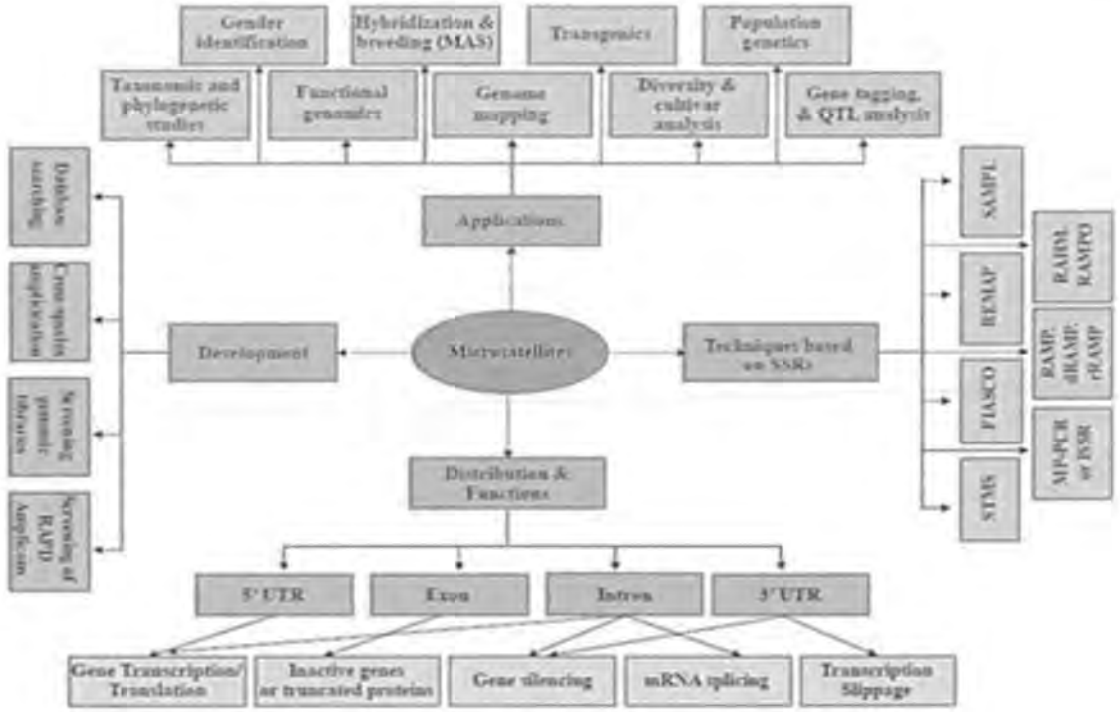

Figure 4 the Microsatellite: A summary of development, distribution and function and application.

1.7.2 Random Amplified Polymorphic DNA's (RAPDs)

This technique was initially developed by William et al (Williams *et al.*, 1990) and Welsh and McClelland (Welsh and McClelland, 1990). RAPDs are PCR based implication and random amplify DNA sequence, they are inexpensive. RAPD require high quantity and quality of DNA. (Wolff *et al.*, 1993) Implications take place when two hybridization sites are similar or in opposite direction. RAPD primer should have minimum 40% GC.(Williams *et al.*, 1990). RAPD are highly polymorphic and band presence or absence can be easily visualized on gel electrophoresis (Jones *et al.*, 1997)(Jiang, 2013).RAPD marker has been widely used for genetic diversity in population and phylogenetic relationship among species and subspecies. (Wolff *et al.*, 1993). Some disadvantage this marker are not used for genome mapping and do not exhibit dependable amplification patterns (Dieffenbach *et al.*, 1993).

1.7.3 Amplified Fragment Length Polymorphism's (AFLPs)

AFLP is PCR based markers, require low quality and quantity of DNA, that should not contain any restriction enzymes or PCR inhibitors. (Blears *et al.*, 1998; Ridout and Donini, 1999). AFLP is multi-loci markers that occupy different positions of DNA. AFLP used two restriction enzymes for cutting of DNA. Each fragment of oligonucleotide sequence of nucleotides is used for ligation (Madhumati, 2014). This technique is completed in three steps.

Initially, oligonucleotide adopters are attached at both ends of the resulting restriction fragment and genomic DNA is digested. These fragments are selectively amplified using adopter and restriction site sequences as primer binding sites for PCR reaction. At 3 end of the primer extend into restriction fragment by oligonucleotide (1 to 4b) and only those fragment are amplified whose ends are complementary to 3 ends of selective primers. Finally, the amplified fragment is visualized by electrophoresis by either silver staining autography or fluorescence using polyacrylamide gel (Tautz, 1989; Schlotteröer *et al.*, 1991). However the limitation required greater technical skills, Moreover AFLP is dominant rather than co. dominant, and use Polyacrylamide gels for detection (Wolff *et al.*, 1993).

1.7.4 Single-nucleotide polymorphism (SNP)

The SNPs are the single base pair changes in the genomic sequence of the individual and may be (C/T or G/A) which is a transition or (C/G, A/T, C/A or T/G) is transverse of the basis in nucleotide substitution. The mRNA normally represents the change in a single base pair, including the SNPs which are the deletion/insertion in the single base. The SNPs cover the maximum and simplest numbers of markers and considered as the smallest unit for the inheritance. The frequency of SNPs in the plants ranges from 1 SNP in the every 100 to 300 base pairs (Xu, 2010). The possibility of finding the polymorphism in targeted gene increased because of the high density of SNPs markers, ultimately providing the advantage over preceding markers that at best closely linked to interested locus and not within (Lateef, 2015).

The coding and non-coding portion of genes or in between the genes (intergenic region) both have a fairly wide distribution of SNPs with different frequencies (Xu, 2010). A number of methods for detection platform and the allelic discrimination have been developed

based on the SNP genotyping. Various types of SNPs genotyping are developed depending on the molecular mechanisms among them oligonucleotide ligation, primer extension, allelic-specific hybridization and primer extensions are the most important (Sobrino *et al.*, 2005).

The recent throughputs in the genotyping method like chip-based NGS, allelic-specific PCR that makes SNPs the most attractive genotyping markers (Agarwal *et al.*, 2008). In near future, the SNPs markers are the markers of choice for the breeding especially in full sequence of plant genomes and will be available as an advantage of upcoming generations sequencing technologies(Liu *et al.*). More recently, it has become very cost effective and easier to quickly identify a large number of SNPs in a short time in any plant species.

1.7.5 Diversity array technology (DArT Seq)

It is a technique that provides a great opportunity for the genotyping of polymorphic loci (in several hundred to several thousand), which are distributed over the genome. It is highly reproducible microarray hybridization technology. There is no need for previous sequence information for the detection of loci for a trait of interest (Jaccoud *et al.*, 2001; Wenzl *et al.*, 2004). The most important benefit of this technique is that it is high throughput and very economical. To discover polymorphic markers by this technology, a single reaction assay can genotype several thousand genomic loci. As little as 50–100 ng genomic DNA is sufficient for the genotyping purpose. For the scoring and discovery of markers, an identical platform is utilized.

After the discovery of a marker, there is no need of specific assays for genotyping, except starting polymorphic markers assembly into an array of a single genotype. These polymorphic markers within the genotyping arrays are commonly used for genotyping (Huttner *et al.*, 2005).

Table 3 Comparison of important characteristics of the most commonly used molecular markers.

Characteristics	RFLP	RAPD	AFLP	SSR	SNP	DArT
Co-dominant/Dominant	Co-dominant	Dominant	Dominant	Co-dominant	Co-dominant	Dominant
Reproducibility	High	High	Intermediate	High	High	High
Polymorphism level	Medium	Very High	High	High	High	High
Required DNA quality	High	High	High	Low	High	High

Required DNA quantity	High	Medium	Low	Low	Low	Low
Marker Index	Low	High	Medium	Medium	High	High
Genome abundance	High	Very High	Very High	Medium	Very High	Very High
Cost	High	Less	High	High	Variable	Cheapest
Sequencing	Yes	No	No	Yes	Yes	Yes
Status	Past	Past	Past	Present	Present	Present
PCR requirement	No	Yes	Yes	Yes	Yes	No
Visualization	Radioactive	Agrose gel	Agrose gel	Agrose gel	SNP-VISTA	Microarray
Required DNA (µg)	10000	20	500-1000	50	50	50-100

1.8 Quantitative trait locus (QTL)

A Quantities trait locus is the region on a chromosome that is controlled by multiple factors such as genetic and environmental factor. QTL is software used identify marker (SSR, SNP, AFLPs) to identify the region of the genome that correlate to polygenic phenotypic traits (yield). By the help of QTL identify the specific sequence of the actual gene that controls multiple traits.

The reduction of tillering is one of complex trait associated with 'gigas' wheat demonstrated by the increase of spike length, number of spikelet per spike, seed per spikelet, seed weight per spikelet and additional a reduced plant height. QTL analysis of these traits revealed that they have significant impact on tin locus (Heidari, Sayed-Tabatabaei et al. 2011); (Suenaga *et al.*, 2005); (Kumar *et al.*, 2007);(Marza *et al.*, 2006);(Börner *et al.*, 2002; Duggan *et al.*, 2005; Cuthbert *et al.*, 2008; Cui *et al.*, 2012).

1.9 QTL Mapping:

With the advancement of plant breeding, breeders have used linkage analysis for identification of quantitative trait variation, i.e. quantitative trait loci (QTL) mapping for introgression and dissect quantitative traits. QTL mapping is the identification of both genetic positioning as well as a confidence interval, the probable gene decrypting percentage of trait variation. QTL mapping is very often available in the literature for the same trait in several different and independent experiments. The history of mapping quantitative traits can be traced back to the 1920s, when Sax (1923) used the morphological markers to demonstrate an

association between seed weight and seed coat color in beans. Today in 1961 used multiple genetic markers to systematically map the individual polygene, which controls quantitative traits in Drosophila. He notices that "the main practical limitation of the technique seems to be the availability of suitable markers". It is obvious that the numbers, as well as polymorphism rate of morphological or protein-based markers used at that time, were very limiting factors. There are numerous methods available for QTL detection i.e. single marker analysis (using ANOVA / regression / t-test), interval mapping, composite interval mapping, multiple interval mapping implemented since 1980's (Weller 1986, Lander and Botstein 1989, Haley and Knott 1992, Jansen 1992, Jansen 1993, Zeng 1993, Zeng 1994, Zhu 1998, Kao 1999 etc.). As this study does not focus on the methodology of QTL detection, only a brief description of the methods is given below. The single markers analysis is based on the simple idea that if there is an association between markers and trait values, it is likely that a QTL locus is close to that marker locus. The approach has been applied in many QTL studies for various organisms such as initially in Drosophila (Thoday 1961), maize (Edwards et al. 1987) and tomato (Weller et al. 1988), etc.

This method is mostly carried out on populations with two classes of genotypes (i.e. Recombinant Inbred Lines (RIL), Back Cross (BC) or Double Haploid (DH)). The major limitations of this technique are linked to the fact that: (i) the like hood of QTL detection significantly decreases as the distance between markers and QTL increases, (ii) the method cannot determine whether a marker is associated with one or more QTL, (iii) the effect of QTL is likely to be under estimated because they are confounded by recombination frequency. In order to overcome the drawbacks of the single markers analysis approach, Lander and Botstein (1989) proposed the interval mapping approach as a systematical way to scan the whole genome for statistical evidence of QTL. The interval mapping method is an extension of the single markers analysis by using two flanking markers to construct an interval for searching a putative QTL. 39 basic idea of the interval mapping approach is simple. LOD (Likely hood also named as logarithm (base 10) of odds score) is calculated between each interval increment from the first interval to the last one (walking step). This gives LOD score profile for the entire genome. When LOD exceeds a threshold value, we declare that a QTL have been found at that location. Interval mapping approach is widely used so that different software's are available for this method such as QTL cartographer, Map QTL, multi QTL etc. However, when multiple QTL are segregating in a single cross, alike for

the single markers analysis, interval mapping analysis also fails to account genetic variance originating from the different initial QTL. In 1994, Zeng et al. proposed the multiple QTL methods or marker-QTL-marker analysis, which was developed by Jansen and Stam (1994) by combining interval mapping with multiple regression analysis called composite interval mapping (CIM). The statistics consider a marker interval plus a few other well-chosen single markers in each analysis (to account for QTL effects outside the interval) so that n-1 tests for QTL-interval association are performed on a chromosome with n markers. The advantages of the CIM approach are: (i) Multiple QTL mapping can be performed in one single location (locus), (ii) Linked markers are used as cofactors so that the test is not affected by QTL present outside the confidence interval, thereby increasing the precision of QTL, (iii) By accounting genetic variance by other QTL, the residual variance is reduced, thereby increasing the power of detection of QTL. Finally, Multiple Interval Mapping (MIM) approaches is a multiple QTL oriented method which generalized simple QTL interval mapping analysis but allows searching for a number, positions, effects and interaction of significant QTL simultaneously (Korol et al. 1995). The MIM method consists in four components: (i) an evaluation procedure designed to analyze the likelihood of the data given a genetic model (number, positions and epitasis of QTL) (Kao and Zeng 1997), (ii) a search strategy optimized to select the best genetic model in the parameter space, (iii) an estimation procedure for all parameters of the genetic architecture of the quantitative traits simultaneously given the selected genetic model, (iv) the prediction procedure to estimate or predict the genotypic values of individuals based on the selected genetic model and estimated genetic parameter values for marker-assisted selection.

There are various terms that are associated with QTL i.e. heritability, maximum likely hood, of the odd score, confidence interval. A detailed description of these terms is given hereafter. Heritability (R^2 or H^2) is the proportion of phenotypic variation in a population that is attributable to genetic variation among individuals.H^2 or R^2 = Genotypic Variance = Phenotypic Variance= Genotypic Variance + Environmental Variance. So, R^2/H^2 value explains the inheritance of a trait at a particular locus in a segregating population. Maximum likely hood ratio (Lr) is a measure of the presence of a QTL in the considered interval. 40 Likely hood of odd (LOD) is a function of Maximum likely hood, defined as a log of the hypothesis that likely hood of the presence of QTL over the hypothesis that likely hood of the absence of QTL. Maximum likely hood and LOD score are related terms and are explained as.

Lr = -2 ln (likelihood of presence / Likely hood of absence) LOD = Log 10 (likelihood of presence / Likely hood of absence) LOD = 0.217 Lr the LOD score between the intervals is plotted on the chromosome to draw the QTL curve. The Confidence Interval (CI) corresponds then to a chromosomal region in which the true location of a QTL has a α=95% of chance to be found or also formulated as the significant threshold of the QTL location associated with x% (normally 95%) of probability of existence compared to a null hypothesis of the absence of a QTL on the considered region. Lander and Botstein (1989) proposed the LOD drop-off method to calculate the Confidence Interval (CI) of the QTL. Using this method, the CI is calculated by finding the location at either side of the estimated QTL location that corresponds to a decrease in the LOD score value of 1 or 2 units. Compared to QTL detection described previously, the meta-QTL approach defines consensus QTL that is statistically identified in a pool of independent linkage analyses (Etzel et al. 2002). As a consequence, the consensus confidence interval is reduced compared to the initial ones considered independently. Defining a minimum confidence interval is a crucial step toward QTL cloning either through map-based cloning approach (based on the construction of a BAC-clone based physical map spanning the QTL interval) as well as through a candidate gene approach (based on known gene function located under the QTL interval). As a consequence, the meta-analysis concept, summarized as the integration in a single analysis of different data obtained independently, was transposed by Goffinet and Gerber (2000) in the field of genetics and especially into the concept of the calculation of a concise QTL from independent studies.

The method provides modified Akaike criteria that can be used to decide how many QTL are actually represented by the QTL detected in single experiments. This statistical criterion is computed to choose between models with one, two, three, or more QTL models on the same linkage group based on different independent studies. Simulations were carried out to investigate the quality of the model obtained with this method in various situations. The method allows the length of the confidence interval of QTL location to be consistently reduced when there is colocalization of initial QTL loci. This meta-QTL approach was carried out on dairy cattle QTL (Khatkar et al. 2004, MacLeod et al. 2003, Lean et al. 2006, Charbonneau et al. 2006), in human science (Rice et al. 2006, Heijmans et al. 2005, Lawlor et al. 2006), as well as in the field of plant breeding and more precisely for the genetic architecture of flowering time of maize (Chardon et al. 2004).

1.10 QTLs For tiller number:

QTL that controls tiller number was associated with vernalization (vrnA) gene on chromosome 5A in wheat. Kato et al. (2000). According to (Li et al. 2002), a QTL that had significant effect number of tiller per plant is located on short arm of chromosome 6A and 1D. Moreover also a numbers of QTLs were identified on different chromosome such as 1A,1B,2A,2D,3A,3B,5A,6D and 7A that had significant effect on tiller number (Spielmeyer and Richards, 2004),(Kuraparthy et al., 2007); (Kato *et al.*, 2000; Xie *et al.*, 2006); Li et al. 2002;(Huang *et al.*, 2003; Cui *et al.*, 2011);(Yan et al., 2011); (An et al., 2006).

Spielmeyer et al., 2004 has identified and mapped the tin locus which was connected with the microsatellite marker gwm136. Beside this, the study also demonstrated synteny relationships between this region of rice on chromosome 5, and some ESTs of chromosome 1A were non-orthologous to rice regions (figure 6).

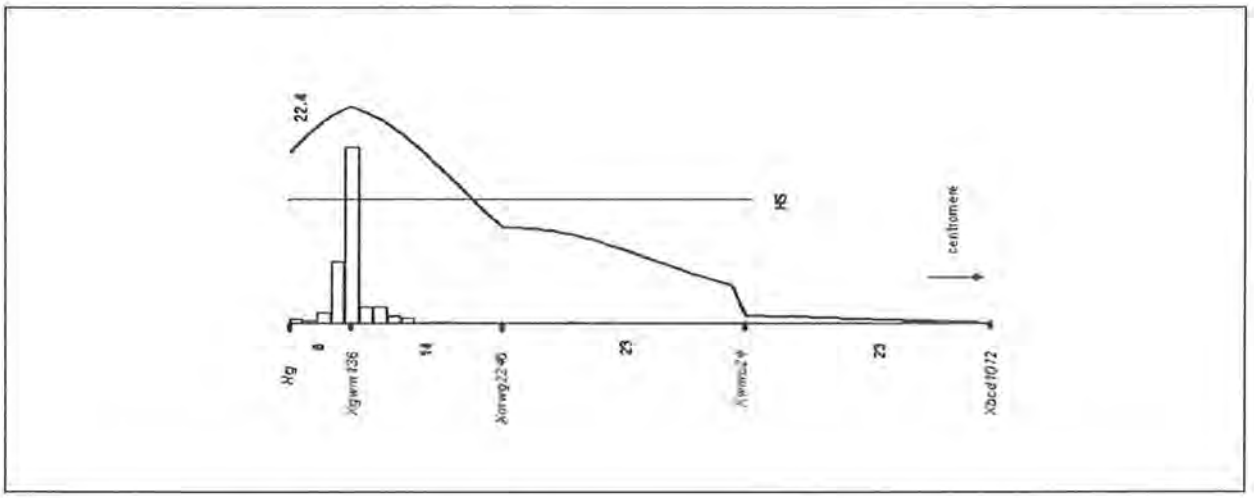

Figure 5 The tin locus. This locus for tillering in wheat is located on chromosome 1A. The LOD score obtained is 22.4. The marker at the peak of the QTL is the marker gwm136 (Spielmeyer et al., 2004).

1.11 Development of population:

In order to dissect *Tin-1* locus, single seed dissent recombinant inbred lines (RILs) population was constructed "Oligoculum" the tiller number mutant from Spielmeyer et al., (2004)was reciprocally crossed with "OC37" in INRA France. OC37 line had 30-40 tillers as

compare to 1-4 tillers in Oligoculmn. Thus, an unbiased population with the SSD (single seed descent is explained in Figure7.

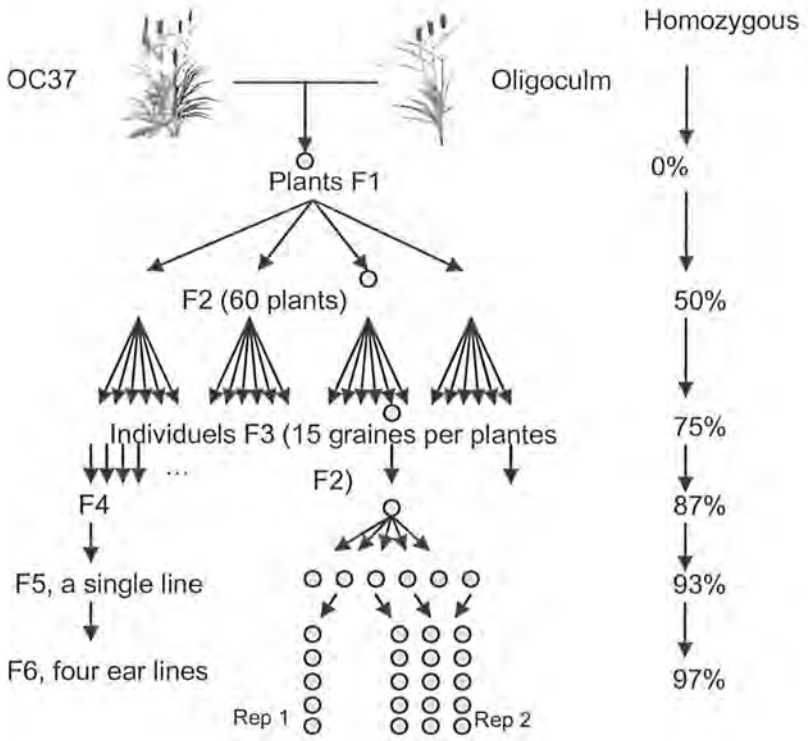

Figure 7: Population Development Strategy

In order to achieve this population faster, generations F3 and F4 were produced within the same year (accelerated generation) in the greenhouse. The F6 population was therefore developed in 5 years. In the F4 population of about 1000 plants, aberrant phenotype plants were removed by specific selection criteria. The population F6 was planted in fields for the year 2006-2007, it is composed of 423 families (plants descending from the same plant F4, or spur line). The DNAs of these plants were taken from the F5 population for genotyping.

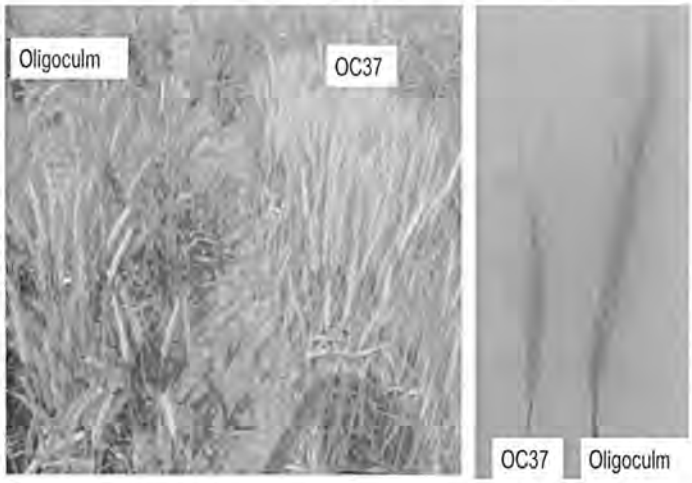

Figure 6 the comparison of oligoculum (low tiller line) and OC37 (high tiller line)

Construction of a genetic map

For the analysis of yield components in soft wheat, we developed the genetic map from SSR markers (microsatellites) of the F5 population of the Oligoculm X OC37 cross. In parallel, phenotyping of the F6 population in the field was performed. From this map and phenotypic data, we have completed the search for QTLs Phenotypic data on the F6 population comprised on information about early flowering, ear morphology (size, number of spikelet's, number of flowers, compactness) of number of ears (tillering) and size of the plant.

This tillering QTL co localizes with the tin locus on the same chromosome which was already reported. The associated markers gpw7072 and psp2999. The psp2999 marker is reported to be associated with said QTL in Suenaga's 2004 publication for tiller number.

1.12 Fine Mapping of QTL

Initially, the tillers QTL on chromosome 1A had a LOD score of 22.4 and accounted for 49% of tillering character. It has been performed since the addition of numerous SSR and COS markers that reduced the initial confidence interval from 4.57cM to 0.96cM. The physical map of oligoculmn Loci was constructed in France, carrying 25 genes on 150Mbp region. There was no further polymorphic marker on the region. So, to delimit the region further crosses were made to dissect the locus. The recombinant (line 147) from the mapping population was cross with caphorn (21-32 tillering). The F1hetrozygotes were backcross to

the parent (pi) caphorn to create BC1F1 population which selfed to produce BC1F2 population.

Aims and objectives of research:

The main purpose of my study was to dissect Tin1 locus. In order to achieve my goal, we had to do QTL mapping of BC1F2 population. If there were no recombinants in the region, we wanted to do develop a new dense mapping population (BC2F2) to dissect the region. So the objectives of my study can be listed as

- Phenotypic screening to BC1F2 population.
- Molecular mapping of the markers within tin1 locus.
- QTL mapping in BC1F2 population.
- Identification of the gene responsible for tiller inhibition on the tin1 locus.
- Development of BC2F2 population by the reciprocal cross of recombinant lines with Caphorn.
- Molecular mapping of tin1 Locus on BC2F2 population.
- Identify the impact of Tin1 locus on grain yield.

2 Materials and Methods

2.1 Plant Material:

Seeds of 300 BC1F2 mapping populations were taken from INRA Clermont Ferrand France. Plant material along with parent's oligoculmn, OC37 and Caphorn were planted in the field at Department of Plant Sciences, Quaid-i-Azam University Islamabad in the growing season of 2016-2017. Each line was space planted according to environmental conditions, normal cultural and agronomic practices. Genotyping of the material was done at INRA Clermont Ferrand France.

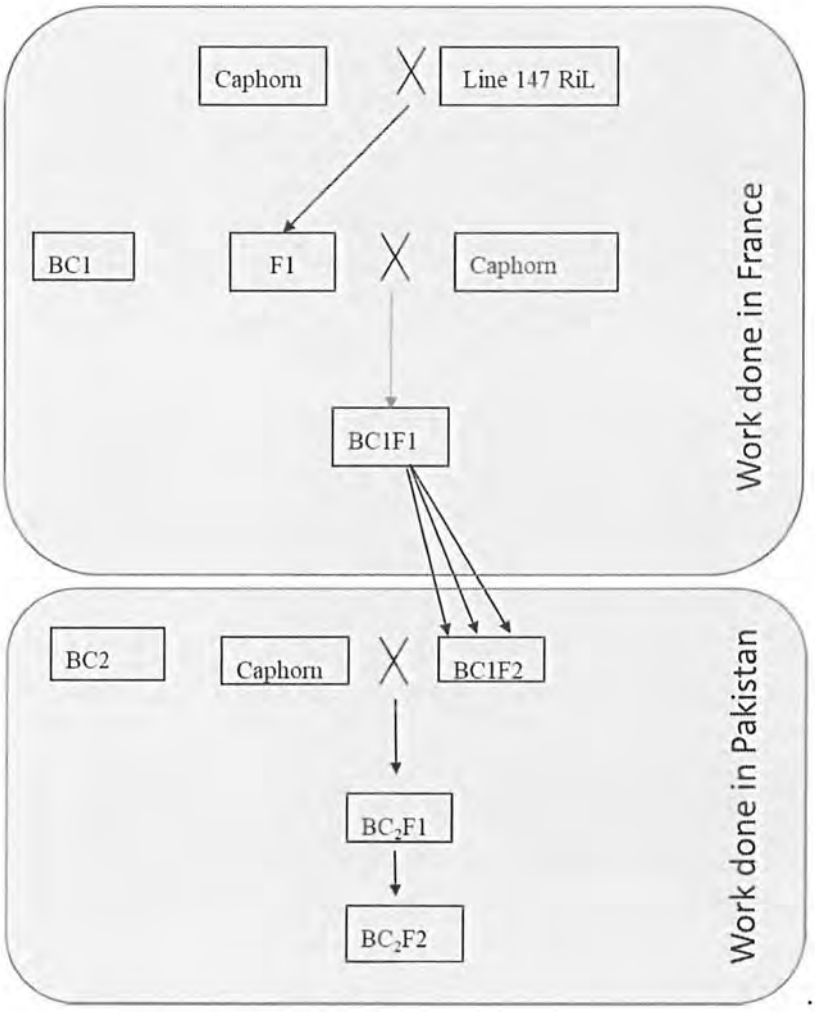

Figure 7Schematic representation of work done.

2.2 Construction of BC2F2 mapping population

The recombinant line from BC1F2 line (C20-1-1) was back reciprocally crossed with caphorn at Quaid-i-Azam University Islamabad. The F1 seeds were planted in the glass house of Department of Plant Sciences, Quaid-i-Azam University Islamabad during off season of 2017. F2 seeds were collected and thrashed for further analysis.

2.3 Analysis of BC1F2 Population tin1 Winter Bread Wheat:

Experiments were carried out at the Plant physiology lab Department of plant science Quaid-i-Azam University Islamabad in 2017 to 2018, to evaluate tillering inhibition line in F2 Population of winter bread wheat. A total of 800 seeds from F1 heterozygous (C20-1-1 X Caphorn) Hexaploid winter wheat cultivars (*Triticum avestivum* L.) were planted for evaluation.

2.4 Growth chamber

Seeds were sown in seed trays containing wells with Peat moss was used to ensure nutrient supply to young seedlings. The seeds were sown on 12th November 2017and germinated under open conditions photoperiod and temperature, (16: 8 h, 20: 16°C, light: dark). Seedlings at the two-leaf stage on 24th November 2017 were transferred to 4°C to vernalize for 30 d. while average photoperiod and the temperature were maintained at 8:16h, 15: 3.8c^{o}, light: dark.

The vernalized seedlings at three leaf stage were transferred to a hardening stage (photoperiod and temperature, 12: 12 h, 15: 15°C, light: dark) for 6 d to gradually acclimatize. Finally, all the plants were transplanted into the field.

2.5 Experimental Design:

Before transplanted seedling, field trial was irrigated by ploughing to ensure uniform emergence. At three leaves stage, all F1 seedlings C20-1-1 x Caphorn were transferred to the field. The field was prepared by a mixture of loam and sandy soil. All seedlings were space planted with10 cm plant to plant and 15cm rows to rows distance. After two weeks of planting NPK fertilizer was added at the rate of 100:80:80 kg/hac to ensure healthy growth.

2.6 Evaluation of Phenotypic Date:

Since the main aim of this study was to identify high tillers genotypes containing high grain yield plant characteristics, for this purpose we determined all phenotypic yield-related traits, such as number of tillers, plant height, spike length, spikelet's per spike, seeds per spikelet, grains per spike, Thousand Kernel Weight, and Grain Yield was evaluated at different stages of wheat.

2.6.1 *Tiller number:*

Tiller number of each individual of F2 population was counted manually at two different stages i-e initially tiller emergence and number and harvesting time.

2.6.2 Plant Height:

Plant height was measured when plants reached physiological maturity. It was measured using ordinary meter rod in cm. The plant height was measured from the base of the plant to tips of spike excluding the awn length of each individual in F2 population at physiological maturity.

2.6.3 Spike Length:

Spike length is one most important agronomic trait of Oligoculum winter wheat, which ultimately depends on yield (Richard 1998) . Spike length, was measured from termination of the rachis of the spike to the tip. Spike length was measured by a 30 cm scale of each individual of F2 population at physiological maturity.

2.6.4 Spikelet's per spike:

The smaller seed bearing units on a spike has grown on rachis are termed as spikelets. Per spike of each spike of the individual plant was counted manually after harvesting F2 population.

2.6.5 Seeds per spikelet:

The number of seeds produced in each spikelet of the spike is known as seeds per spikelet.Seeds per spikelet were manually counted as a number of seeds in each spikelet of the spike.

2.6.6 Thousand Kernel Weight:

Thousand kernel weights or TKW is calculated by counting and weighing of 1000 seeds in grams using the electric balance.

2.6.7 Grain yield (GY)

The total grain yield is obtained in grams by weighing the number of seeds obtained per unit i.e. the number of seeds obtained per plant. It is measured by using an electric balance

2.7 Statistical Analysis:

The collected data was further evaluated using a number of statistical software's, such as XL- STAT and R. The statistical analyses applied on the data were correlation test and ANOVA, where the mean of all variables were used to generate results.

2.7.1 Correlation Test

Correlation test is an analytical tool that allows the study of two variables and the strength of their association. Pearson correlation was used. It explores the extent of linear association exhibited by two different variables, with the strength and trend of association between the said variables being represented with a single value and sign. The values of correlation co efficient range between +1 and -1 being negatively correlated and no correlation being exhibited between two variables at 0. For the computational purpose, the statistic tools used were XL- STAT 2014 and R.

2.7.2 Analysis of variance (ANOVA)

ANOVA is another statistical tool used to understand and signify the difference between three or more independent groups with the help of their simultaneous comparison.

XL- STAT2014 was used for application of ANOVA on data. The p-value obtained once the test applied determines the level of significance of the data. If the p-value is less than 0.05, then the difference between the obtained data is significant. The closer the value reaches 0.01., the higher the 'p-value becomes significant.

2.8 Genomic plant DNA extraction:

For genomic analysis 3-4inch young leaves at four-leaf stage were harvested and washed with 70% ethanol followed by distilled water. Leaf sample was immediately cryopreserved in liquid nitrogen. Leaf sample was grind by pestle mortar. Approximately 0.5 g fine powder was transferred into a 1.5 ml Eppendorf tube. Add 600ul of preheated extraction buffer at 65^0 in water bath [(5MNacl, 1M Tris pH (8.0), 0.5 M EDTA pH (8.0), 20% SDS)]. Add 0.38g sodium bisulfate/100ml in extraction buffer add to powder in Eppendorf tube. Eppendorf tube containing samples were incubated for 30 mins in the water bath at 65^0C.

The sample was incubation added 600 ul is amyl alcohols: chloroform (1:24), and shacked vigorously by vortex. Samples were centrifuged for 10mins at 4500rpm. Take supernatant into new autoclaved Eppendorf tube, and add 70% chilled ethanol.Samples were placed in a freezer at -20c° for 1 hr. The samples were centrifuged for 10 mins at 10000 rpm. Discard supernatant, and the pellet was washed with 70% ethanol, and briefly, air-dried overnight. At last, step add 50 ul TBE buffer or PCR water to each sample and stored at -20c°.

2.9 Molecular marker:

PCR markers (table 4) were used to fine map the BC1F2 population. PCR reactions contained 25 ng genomic DNA, 1× buffer, 2mM MgCl2, 0.2mM of each dNTPs, 2 units of Taq polymerase and 10 pmol of reverse and forward primers. PCR reactions were optimized according to protocols by Röder et al. (1998). PCR fragments were separated by capillary electrophoresis using the ABI genetic analyzer Prism3130 (Applied Biosystems, Foster City, CA, USA).

Table 4 Marker used for B2C2.

Primer	Forward	Tm	Reverse	Tm
gpw1106	GCAAGAAGAGAAGCACCACC	60°c	ATCCACGTATCCTCCACTGC	60°c
Gpw7080	ATGCCAACCAGACATCACAG	60°c	CAAAACCTACAGCTCCCTCG	60°c
Gwm369	CTGCAGGCCATGATGATG	60°c	ACCGTGGGTGTTGTGAGC	60°c
F24R23	CATTGCCAGCATACATTCTC	57.1	GCTGACACGGGTTTTAT	60.1
Gwm136	GACAGCACCTTGCCCTTTG	60°c	CATCGGCAACATGCTCATC	60°c
Cfa2153	TTGTGCATGATGGCTTCAAT	60°c	CCAATCCTAATGATCCGCTG	60°c

2.10 Molecular analysis of BC2F2 population:

DNA was extracted as described earlier of the mapping population. The DNA was lyophilized at -80°C for 1 day. The DNA was sent to genotyping Plate at INRA Clermont Ferrand for further analysis.

3 RESULTS

3.1 Phenotypic measurements

The quantitative characteristics of a population of many individuals, we often observe a distribution of characters following a Gauss curve *. In most cases, for a segregated population, the average values of the population are framed by the values of both parents. When most values of the population are higher than parental values, we speak of the positive transgression of the character. Similarly, there are negative transgressions of a character. These may be due to a complementarily of the parental genes. The graphs of phenotypic measurements of the F2 population are mentioned below.

3.1.1 Height of Plants

Plant height was recorded at maturity. The F2 population had mean 74cm height. Plant height of Caphorn was 75cm while C20-1-1 was 65cm high. The population was distributed normally showing the segregation of genes in the population. The detail regarding the analysis of the frequency distribution of the F2 population is shown in supplementary table 4 and histogram.

3.1.2 Tillers per Plant

Tillers per plant were recorded at maturity. The F2 population had mean 9 tillers per plant. Caphorn was 17 while C20-1-1 had 2 tillers. The population was distributed normally showing segregation of genes in the population. The details regarding the analysis of the frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.1.3 Morphology of Spike

For each measurement taken (spike length, number of spikelet's per spike, and number of seeds per spikelet) from F2 population.

3.1.4 Length of spike

Length of the spike was measured at maturity. The F2 population had mean 12.1cm length. Spike length of Caphorn was 10.5cm while of C20-1-1 was 15.5cm. The population was normally distributed showing the segregation of genes in the population. The details regarding the analysis of the frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.1.5 Number of spikelet's per spike

Spikelets per spike were counted at maturity. The F2 population had mean spikelets per spike 22.76. Spikelets per spike of Caphorn were 25 while C20-1-1 was 21. The population was normally distributed showing segregation of genes in the population. The details regarding the analysis of the frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.1.6 Number of seeds per spikelet

Seeds per spikelet were counted at maturity. The F2 population had mean 5. A seed per spikelet of Caphorn was 4, while C20-1-1 was 6. The population was normally distributed showing segregation of genes in the population. The details regarding the analysis of the frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.1.7 Thousand Kernel weight

Thousand kernel weights were measured after threshing. The F2 population had mean3.45g thousand kernel weights, of Caphorn was 3g while of C20-1-1 was 5g. The population was normally distributed showing segregation of genes in the population. The details regarding the analysis of Frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.1.8 Grain yield

Grains yield was measured by electric balance after threshing. The F2 population had mean 18g. Grain yield of Caphorn was 15g while C20-1-1 was 38g. The population was

normally distributed showing segregation of genes in the population. The details regarding the analysis of the frequency distribution of the F2 population are shown in supplementary table 4 and histogram.

3.2 Analysis of Variance (ANOVA).

. All agronomic parameters related to grain yield were recorded. ANOVA was applied on the means of all observed parameters and rate of significance between the variables was observed.

3.2.1 Evaluation of Yield related parameters

A number of grain yield-related parameters were recorded for elevation during growing season 2017-2018 and the results obtained through ANOVA are presented as follows;

3.2.1.1 Number of Tillers

The mean values for a number of tillers were calculated to be 8.12. The maximum value observed for a number of tillers was 17.0, while minimum value observed in a number of tillers recorded was 1.0. The Standard deviation (S.D) was estimated at 2.9. The difference for this variable was significant and was found in overall F2 population with- value (>0.0001). The details regarding the analysis of variance of this relationship are shown in supplementary table 4 and Mean histogram for one way ANOVA.

3.2.1.2 Plant Height:

The mean value for plant height was calculated to be 73.9cm. The maximum value for plant height observed was 90cm, while the minimum value for plant height was observed as 35cm. The Standard deviation (S.D) among all population was noticed as 7.05. The difference for this variable was highly significant having the p-value (>0.0001). The detail regarding the analysis of variance of this relationship is shown in supplementary table 4 and Mean histogram for one way ANOVA with one way ANOVA.

3.2.1.3 Spike Length:

For spike length, the calculated mean was observed to be 12.1 cm. The maximum value for spike length was estimated 16.5 cm, while the minimum value for spike length was observed 12.1cm. The Standard deviation (S.D) was estimated at 1.1, while the significant difference p-value was a highly significant value (>0.0001). The details regarding the analysis of variance of this relationship are shown in supplementary table 4 and mean histogram for one way ANOVA.

3.2.1.4 Spikelet Per spike:

. The mean value for spikelet per spike was recorded as 22.76. The maximum value for spikelet per spike was observed 28, while the minimum value for spikelet per spike was observed 15. The Standard deviation (S.D) was noted at 1.79, While the significant difference- value was a highly significant value (>0.0001). The details regarding the analysis of variance of this relationship are shown in supplementary table 4 and Mean histogram for one way ANOVA.

3.2.1.5 Seed per spikelet:

The mean value for seed per spikelet was observed at 5.0. The maximum value for seed per spikelet was calculated 8, while the minimum value for seed per spikelet was calculated 3. The Standard deviation (S.D) was estimated at 0.74, While the significant difference for this variable was highly significant value with (p>0.0001). The detail regarding the analysis of variance of this relationship is shown in supplementary table 4 and Mean histogram for one way ANOVA.

3.2.1.6 Thousand Kernel Weight (TKW):

The mean value for thousand kernel weight was recorded as 3.4g. The maximum value for thousand kernel weight was 5.6g, while the minimum value for thousand kernel weight was observed 1.8g. The Standard deviation (S.D)was noticed at 0.4, While the significant difference of this variable was highly significant p-value (>0.0001). The detail regarding the analysis of variance of this relationship is shown in supplementary table 4 and Mean histogram for one way ANOVA.

3.2.1.7 Grain yield (GY):

The mean value for Grain yield was calculated as 17.9g. The maximum value for grain yield was observed 47.6g, while the minimum value for grain yield was observed 2.9g. The Standard deviation (S.D) was noticed at 11.88, While the significant difference for this variable was a highly significant overall whole population with p-value (>0.0001). The details regarding the analysis of variance of this relationship are shown in supplementary table 4 and Mean histogram for one way ANOVA.

Table 5 Descriptive statistics of grain yield parameters in winter breed wheat.

Variable	Minimum	Maximum	Mean	Std. deviation	P- Value
Tiller NO	1.0000	20.0000	8.1258	2.9155	>0.0001
Plant height	35.0000	90.0000	73.9417	7.0569	>0.0001
spike length	9.0000	16.5100	12.1017	1.1208	>0.0001
spikelet per spike	15.0000	28.0000	22.7695	1.7944	>0.0001
seed per spikelet	3.0000	8.0000	5.0093	0.7400	>0.0001
Tkw	1.8700	5.6700	3.4522	0.4364	>0.0001
Grian yield	2.9100	47.60000	17.9128	11.8871	>0.0001

3.3 Correlation Matrix:

Correlation provides information regarding the signs of any traits using the mean value of concerned observation; the correlation between traits was calculated and analyzed.

3.3.1 Inter-correlation between physiological parameter

A correlation test was carried out between the physiological parameter under study i-e Parameter concerned grain yield-related parameter showed the significant correlation among yield parameter. Both values positive and negative.

3.3.2 Number of Tillers:

Correlation test showed that tiller number had a significant positive correlation with grain yield 0.56, with an r-value of 0.34. It also showed a positive correlation with plant height and spikelet per spike. The r value recorded here was 0.32 and 0.24 while the number of tillers has significantly negative correlated with spike length, seed per spikelet, and thousand kernel weights. The r value was recorded here with -0.005,-0.30, and -0.28

respectively. The detail regarding co-relationship is shown in supplementary table 5 and figure 11.

3.3.3 Plant Height (PH):

Correlation test showed Plant height has significant positive correlation with grain yield, tiller number, and spikelet per spike. The r value was recorded here to be r=0.56 r=0.32 and r=0.15) while plant height showed strongly negative correlated with seed per spikelet, and thousand kernel weight. The r value recorded here was r = -0.20, r= -0.064) and r =- 0.0037. The detail regarding co-relationship is shown in supplementary table 5 and figure 11.

3.3.4 Spike length (SPL):

Correlation test showed Spike length has significant positive correlation with the number of seed per spikelet, thousand kernel weights, spikelet per spike, and grain yield. The r value recorded here was r = 0.37, r = 0.21, r = 0.24 and r = 0.037 respectively. While it showed significantly negatively correlations with tillers number and plant height. The r value recorded was r = -0.095 and r = - 0.054) respectively. The detail regarding co relationships is shown in supplementary table 5 and figure 11.

3.3.5 Spikelet per spike (sp/s):

Correlation showed mostly correlation to a number with other physiological and agronomic parameters. Correlation test showed Spikelet per spike has positive correlation in with tiller number, grain yield, Plant height, spike length, and thousand kernel weights. The r-value of these agronomic was recorded are r = 0.24, 0.23, 0.15, 0.12), and 0.017.while spikelet per spike showed a significant negative correlation with seed per spikelet. With an r value r = -0.017.The detail regarding co relationships is shown in supplementary table 5 and figure 11.

3.3.6 Grain per spikelet:

Grains per spikelet significant positive correlation with thousand kernel weight, thousand. The value for r this positive correlation was calculated to be 0.37, 0.37) while seed per spikelet showed a significantly negatively affect a number of tillers, plant height, on grain

yield and spikelet per spike. The value for r for this negative correlation was calculated to be -0.30 - 0.20 -0.0019, and 0.12 respectively. The detail regarding co relationships' showed in supplementary table 5 and figure 11.

3.3.7 Thousand kernel weight (Tkw):

Correlation test showed thousand kernel weights have significant positive correlation with seed per spikelet, spikelet per spike, Seed per spike and plant height. The values of r for this positive correlation were a calculation to be 0.37, 0.21, 0.037 and 0.017. While thousand kernel weights showed a significant negative correlation with plant height and number of tillers and grain yields. The value of r for these negative correlation was calculated to be -0.006,-0.28 and -0.019. The detail regarding co-relationship are shown in supplementary table 5 and figure.11

3.3.8 Grain yield (GY):

Correlation test showed that Grain yield have significant positive correlation with a number of tillers, plant height, spikelet per spike. The value of r for these significant positive correlations was calculated to be 0.56, 0.34 and 0.23. Grain yield has also significantly positive correlation thousand kernel weight and spike length. The values of r for this correlation were calculated to 0.19 and 0.037. While Grains yield were showed significantly negative correlation, grain per spike. The value of r for this negative correlation was calculated to be -0.0019. The details regarding co relationships are shown in supplementary table 5 and figure 11.

Table 5 Correlation matrix of yield-related grain parameters.

Variables	Tiller NO	Plant height	spike length	spikelet per spike	seed per spikelet	Tkw	Gran yield
Tiller NO	**1**	**0.3163**	**-0.0735**	**0.2362**	**-0.3116**	**-0.2834**	**0.2388**
Plant height	**0.3163**	**1**	-0.0331	**0.1431**	**-0.2130**	-0.0060	**0.1656**
spike length	**-0.0735**	-0.0331	**1**	**0.1331**	**0.3515**	**0.1834**	0.0139
spiklet per spike	**0.2362**	**0.1431**	**0.1331**	**1**	-0.0284	0.0169	**0.0788**
seed per spiklet	**-0.3116**	**-0.2130**	**0.3515**	-0.0284	**1**	**0.3727**	-0.0063
Tkw	**-0.2834**	-0.0060	**0.1834**	0.0169	**0.3727**	**1**	**0.0891**
Gran yield	**0.2388**	**0.1656**	0.0139	**0.0788**	-0.0063	**0.0891**	**1**

Values in bold are different from 0 with a significance level alpha=0.05

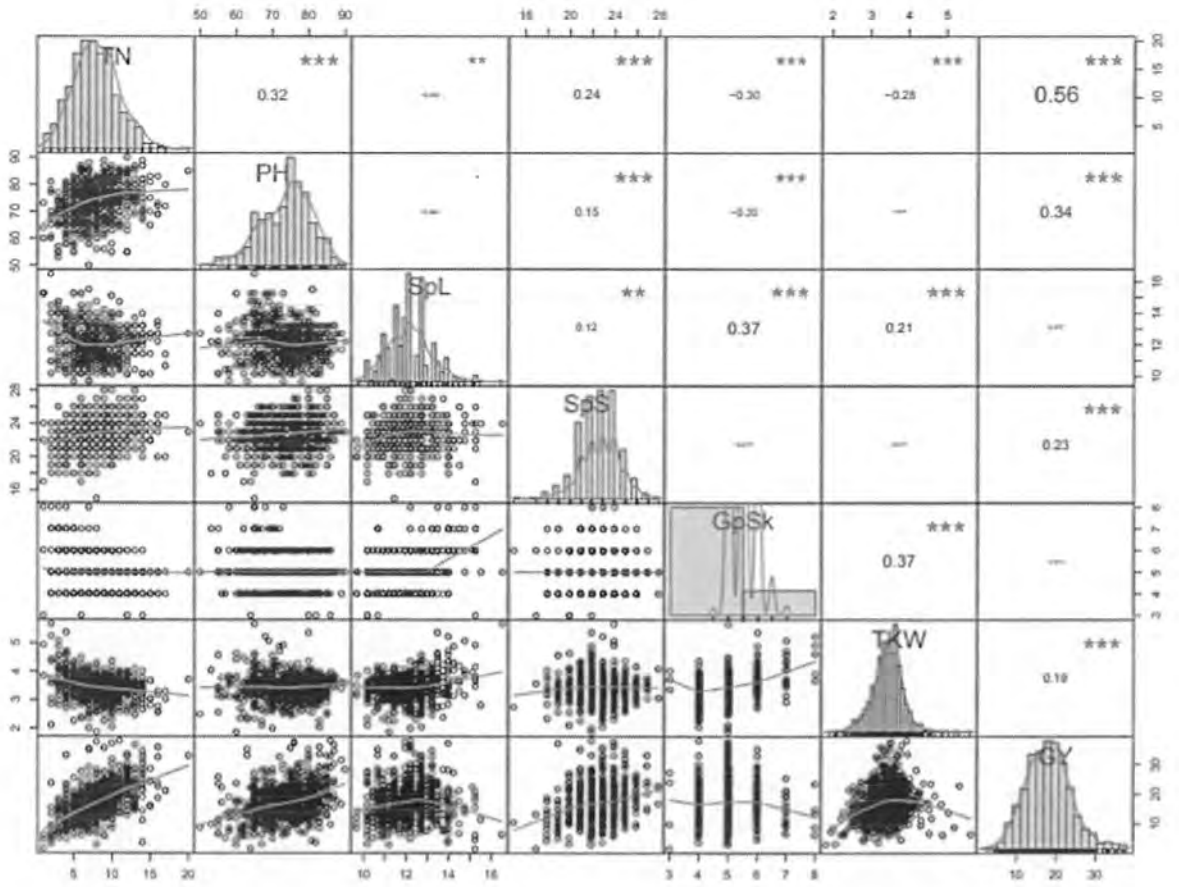

Figure 8 The correlation matrix graph of yield-related grain parameters.

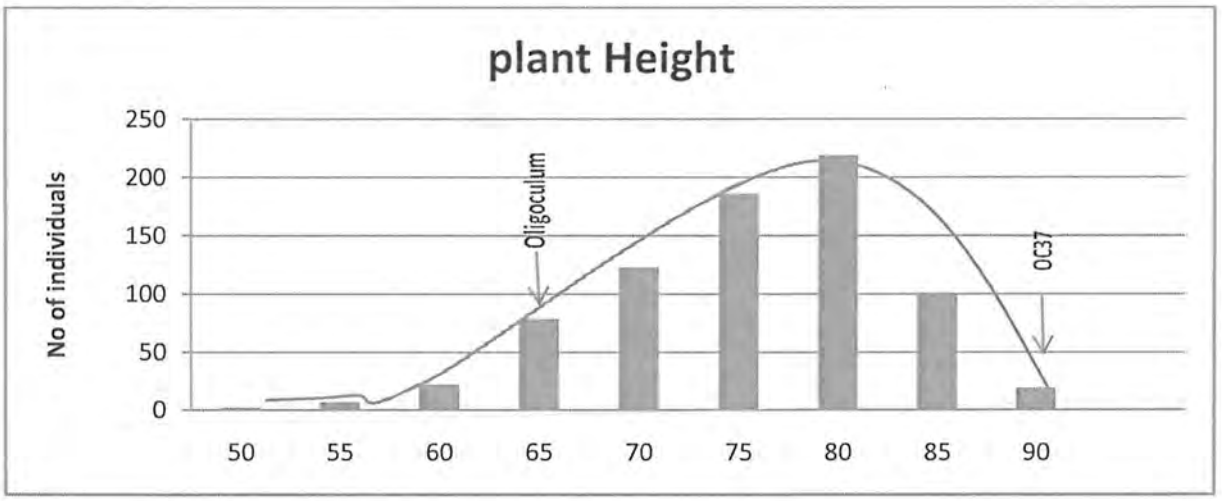

Figure 9Represented the phenotypic frequency distribution of plant height of F2 population of Nil in winter bread wheat.

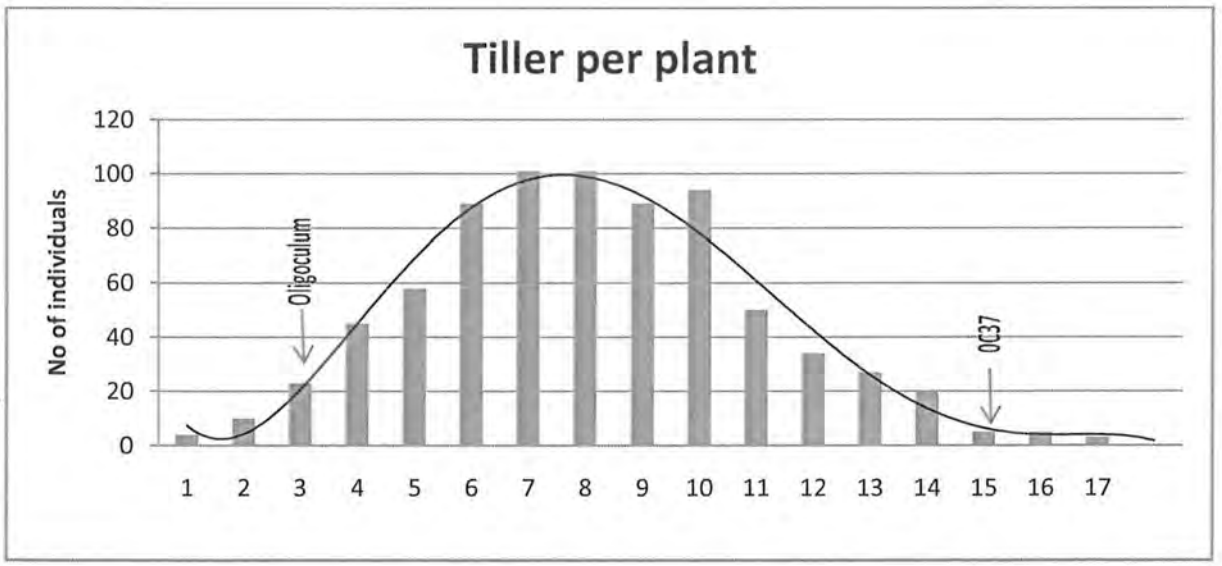

Figure 10 Represented the phenotypic frequency distribution of tiller per plant of F2 population of Nil in winter bread wheat.

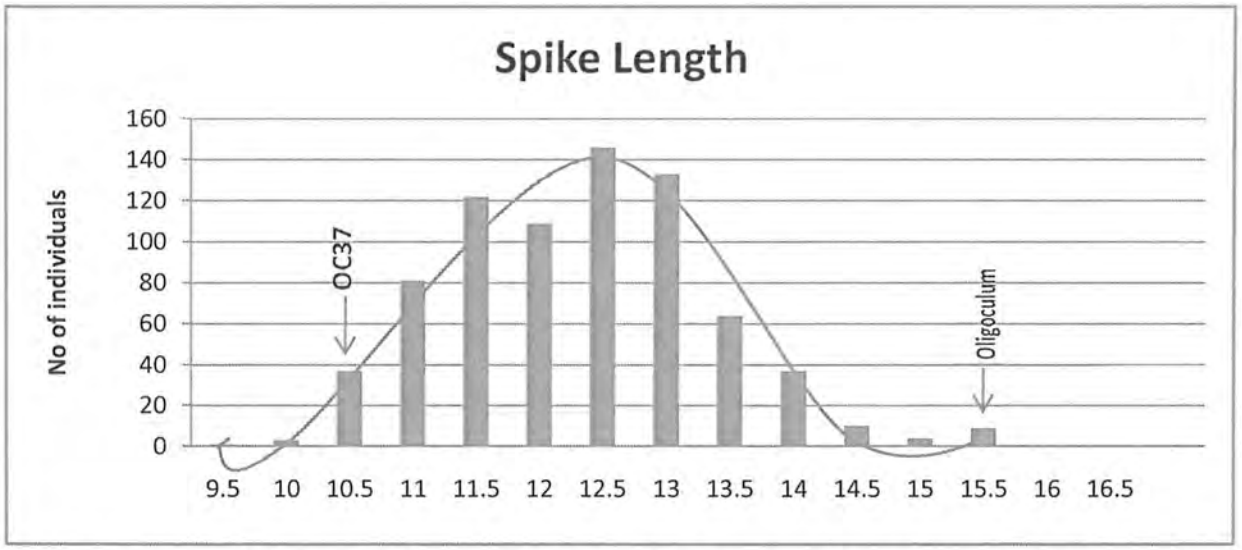

Figure 11 represented the phenotypic frequency distribution of spike length of F2 population of Nil in winter bread wheat.

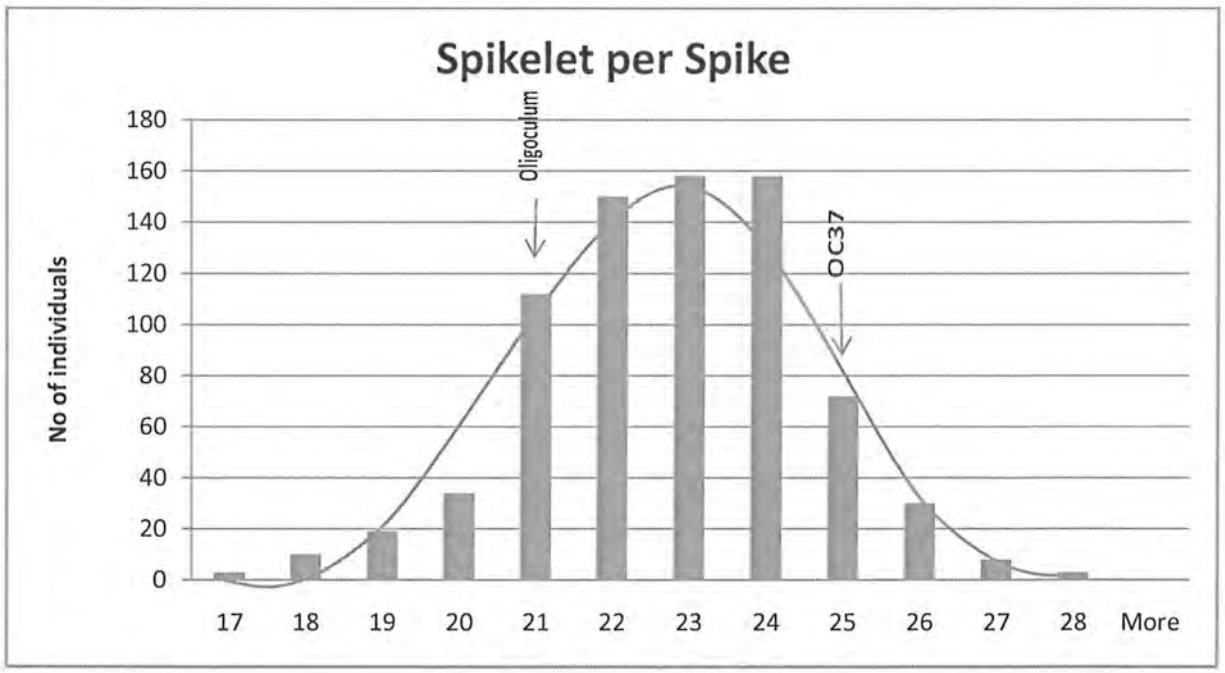

Figure 12Represented the phenotypic frequency distribution of spikelet per spike of F2 population of Nil in winter bread wheat.

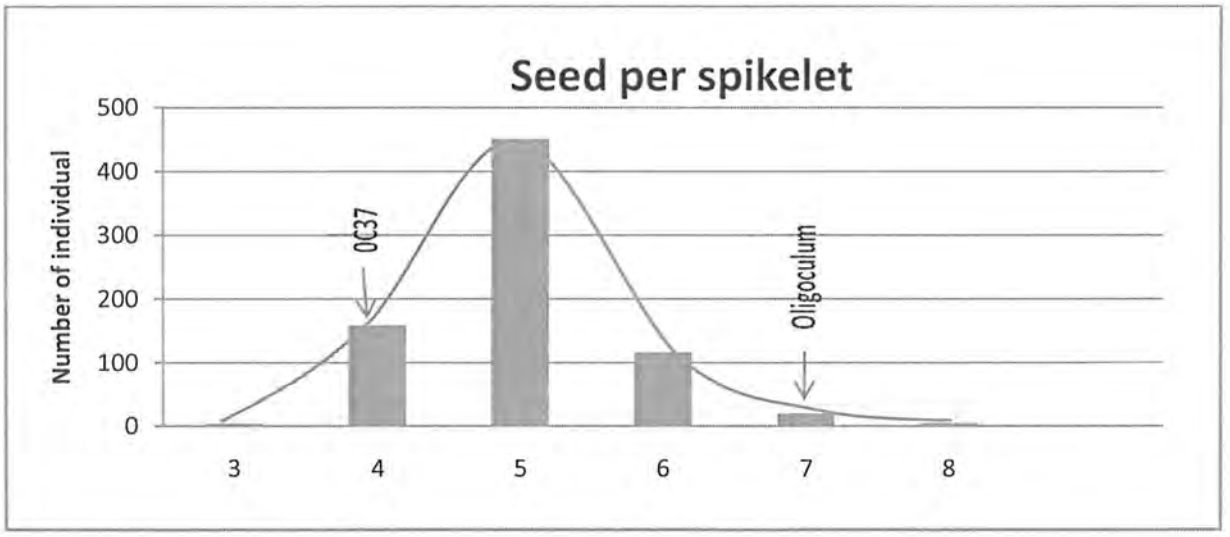

Figure 13Represented the phenotypic frequency distribution of seeds per spike of F2 population of Nil in winter bread wheat.

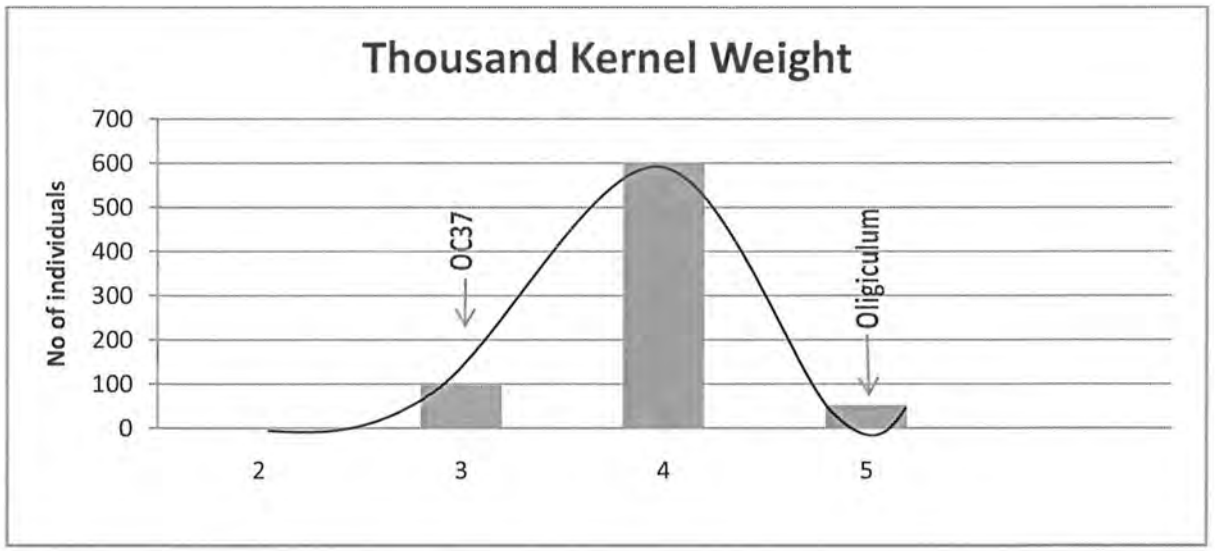

Figure 14Represented the phenotypic frequency distribution of thousands of F2 population of Nil in winter bread wheat.

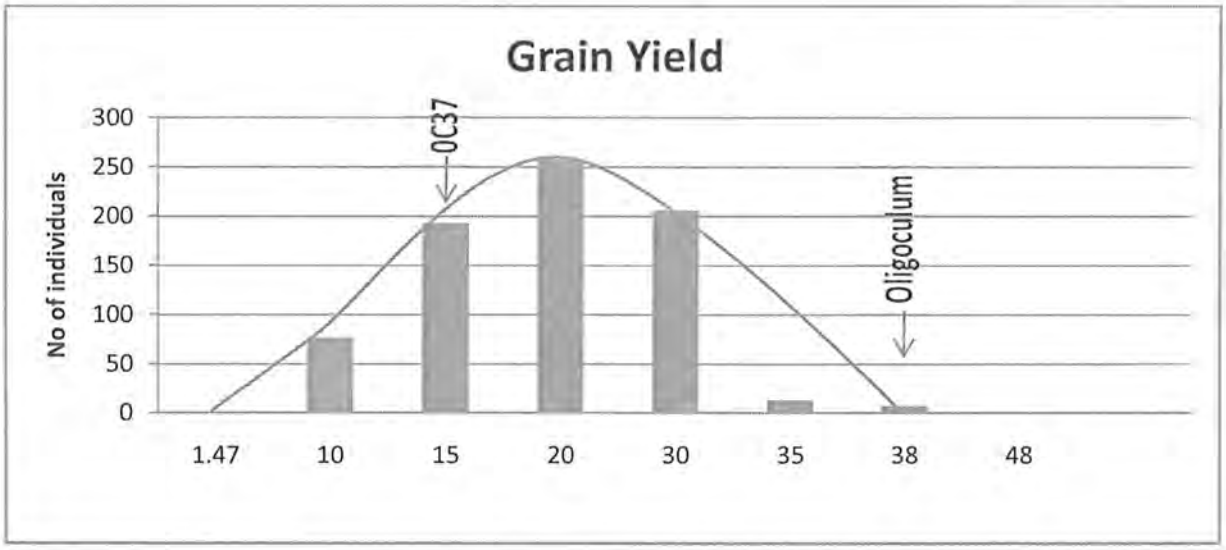

Figure 15represented the phenotypic frequency distribution of grains yield of F2 population of Nil in winter bread wheat.

Table 6 Analysis of variance (Variable Tiller NO):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	37299.6534	37299.6534	1473.7370	< 0.0001
Error	757	19159.3466	25.3096		
Corrected Total	758	56459.0000			

Computed against model Y=0

Table 7Analysis of variance (Variable Plant height):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	3093023.0708	3093023.0708	2156.5624	< 0.0001
Error	757	1085717.9292	1434.2377		
Corrected Total	758	4178741.0000			

Computed against model Y=0

Table 8Analysis of variance (Variable spike length):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	83576.2822	83576.2822	2232.2851	< 0.0001
Error	756	28304.4805	37.4398		
Corrected Total	757	111880.7628			

Computed against model Y=0

Table 9Analysis of variance (Variable spikelet per spike):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	285034.9356	285034.9356	1967.6777	< 0.0001
Error	56	109513.0644	144.8586		
Corrected Total	757	394548.0000			

Computed against model Y=0

Table 10Analysis of variance (Variable seed per spikelet):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	14154.1827	14154.1827	2050.0254	< 0.0001
Error	755	5212.8173	6.9044		
Corrected Total	756	19367.0000			

Computed against model Y=0

Table 11Analysis of variance (Variable Tkw):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	6588.2425	6588.2425	1945.6899	< 0.0001
Error	754	2553.0969	3.3861		
Corrected Total	755	9141.3394			

Computed against model Y=0

Table 12Analysis of variance (Variable Gran yield):

Source	DF	Sum of squares	Mean squares	F	Pr> F
Model	1	168767.4279	168767.4279	707.7545	< 0.0001
Error	755	180033.3399	238.4548		
Corrected Total	756	348800.7678			

Computed against model Y=0

3.4 Fine mapping of tin1 Locus.

The *Tin1* was fine mapped using the molecular markers (gpw1106, gpw7080, gwm136, F24R23, gwm369 and cfa 2153) presented below.

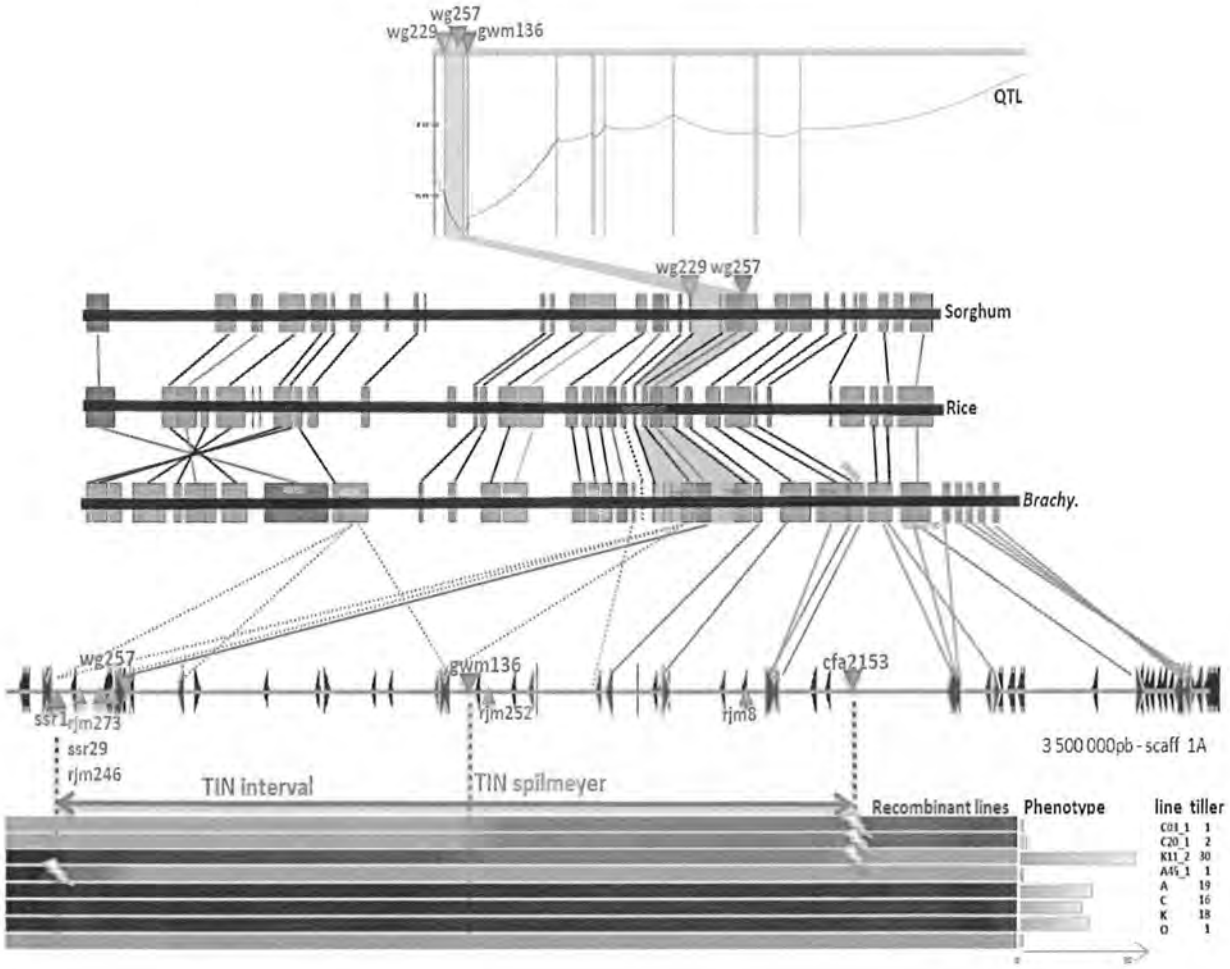

Figure 16The fine mapped QTL corresponded to 3.5Mbp region on the physical map and had 12 genes on it. Spilmier et al., 2017 reported Starch synthase to be responsible for tiller inhibition. Whereas, a Micro RNA was identified to be overexpressed in the zone by our collaborators in France. To validate F2 mapping population was constructed to create recombination between the two genes.

4 Discussions

Near Isogenic Lines (NILs) are set of a line that is genetically identical, except for one or a few loci. Such lines can be created by crossing a donor line (containing a gene or trait of interest) with a recurrent parent to produce a heterozygous F1, and then repeatedly back-crossing with a recurrent parent to create BC1, BC2, etc termed as NILs. NILs are dynamic tools which can be used to dissect and validate complex traits. Even though QTL on Chromosome 1A is not as complex as another quantitative trait due to its high LOD score (22.4) and High heredity (49%), but Creating NILs of the QTL will help us identify and validate the candidate gene responsible for tiller inhibition in bread wheat. Markers (gpw1106, gpw7080, gwm136, F24R23, gwm369 and cfa 2153) within QTL in an initial cross between Oligoculmn X OC37, recombinant inbred line (RILs) were selected. Line 147 of the mapping population was used for backcross with high tiller number recurrent parent (Caphorn) to develop BC1F2 population in INRA Clermont Ferrand, France. The BC1F2 population was screened in INRA using the markers of the initial cross. The phenotyping was done at Quaid-i-Azam university Islamabad. Phenotyping revealed that the trait was segregating in mendelian trait in BC1F2 population. Furthermore, the QTL region was reduced to 0.4cM or 100Mbp carrying 10 genes including, Starch synthase and Micro RNA. Both of which were not present in high tiller number lines and are expressed in Tiller inhibition mutant. Unfortunately, no recombinant was found between these two genes to validate the candidate gene. Using BC1F2 line Cap_20_1_1 fixed line with 87% Caphorn (recurrent parent), reciprocal backcrosses were made with Caphorn at Quaid-i-Azam University Islamabad. These BC2F1 were planted in the field during planting season 2017-2018. Phenotypic screening of 800 BC2F2 populations revealed segregation of Tiller number, Plant height, Spike length, Grain per spike, grains per spikelet, TKW and GY. Significant variation was found in tiller number among the population ranging from 1-17 tillers. Caphorn produces 30 tillers on average in the field in winter planting season. Even though, we gave vernalization treatment to plants maximum potential was not achieved due to short vernalization time. Tiller inhibition mutant in winter season gives maximum 4 tillers in the winter season. In BC2F2 population we identified 190 (24%) individuals with less than 4 tillers. Whereas, there were 610 individuals (76%) with more than 5 tillers. So, the tiller number is segregating according to Mendel law of segregation (https://en.wikipedia.org/wiki/Mendelian_inheritance) *i.e.*1:3 phenotypically in the

population. The previous study on Tin1 and Tin3 locus also found some variations in their mapping populations (Duggan *et al.*, 2005; Suenaga *et al.*, 2005; Heidari *et al.*, 2011; Motzo et al. 2004; Gaju *et al.*, 2009). Tiller number had a significant positive relationship with PH and GY. This signifies the genomic background of recurrent parent Caphorn with high tiller number, tall stature, and higher GY. Similar results were presented in earlier studies of Heidari *et al.*, (2011), Motzo*et al.*, (2004), Gaju *et al.*, (2009). Whereas tiller number had a significant negative correlation with Spike length, Spikelet per spike, Grain per spikelet and thousand kernel weight. It has been reported thantTin1 Mutant plant has the ability to produce long spike using the maximum vegetative production (Heidari *et al.*, 2011). Interestingly, TKW of tiller inhibition lines was also higher in previous studies (Motzo et al. 2004). The high TKW in lines carrying Tin1 mutations is thought to be linked with delayed maturity and high grain filling period (Hyles et al., 2017). Hyles et al., (2017) proposed cellulose synthase-like (Csl) protein recently annotated to Starch synthase in TGCV1 wheat genome release as the candidate gene (https://plants.ensembl.org/Triticum_aestivum/Info/Index). Csl is reported to increase the starch content of the seeds, hence increase TKW but no report of tiller number variation has been reported so far. Micro RNA present within the rejoin of initial QTL is also over expressed in Tin1 carrying lines, suggesting inhibition of tiller pathway genes.

Plant height of Caphorn and C20_1_1 was recorded to 90cm and 65cm respectively. 220 (27%) individuals had plant height less than 70cm, indicating that plant height is also segregating Mendelian in the BC2F2 population. A significant positive correlation was observed to grain yield and spike length. PH in wheat is significantly negatively correlated to yield in post-green revolution cultivars. The significant impact of tiller inhibition with plant height lead was also observed in Tin1 mutant by Hyles et al., (2017). Similarly, important agronomic traits (number of spikelet per spike, seed per spikelet, thousand kernel weight, and grain yield) were also segregated, according to medal law of segregation.

The spike length and spikelet per spike showed a significant negative correlation to a number of tillers in NILs. Showing the impact of Tin1 locus on Spike length. The spike length of C20-1-1 and Caphorn was 15.5cm and 10.5cm, respectively. Seeds per spikelet in C201-1 were 7 as compared to 5 in Caphorn. Similar studies were also suggested by Richards (1988), Motzo et al. (2004), Suenaga *et al.*, (2005) and Gaju et al. (2009).

In order to determine the molecular basis of tin, we constructed fine map QTL from BC1F1 (heterozygous). This fine mapped QTL corresponded to 3.5Mbp region on the physical map and had 12 genes on it, as mention previously in material and method. Spilmier et al., 2017 reported Starch syntheses to be responsible for tiller inhibition. Whereas, a Micro RNA was identified to be overexpressed in the zone by our collaborators in France. To validate F2 mapping population was constructed to create recombination between the two genes.

5 Conclusion and future perspective

Rapid increasing population, decreasing water and land resources in addition to climate change are reducing the potential yield of the field crops. Winter wheat has 2 times more yield potential than spring wheat. The main reducing factor of yield in spring wheat is a number of tillers. Identification of the gene responsible for tiller number in bread wheat could help us to increase wheat yield drastically. BC2F2 population will help us develop recombinant in the zone of *Tin1* locus. The genotypic results have identified the markers (F24-R27) linked to the trait. The marker can be used in marker-assisted selection (MAS) to select against the tiller inhibition locus. The material made in this study would be pivotal for

- Identification of the gene responsible for tiller inhibition in Oligoculm mutant.
- To study the impact of the gene on various pathways of yield in bread wheat.
- To study the QTL for seeds per spikelet.
- Development of high tiller number lines in spring wheat cultivars.
- To understand molecular mechanisms involved in tiller formation in bread wheat.

6 References

Agarwal, M., N. Shrivastava, and H. Padh, 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant cell reports, 27(4): 617-631.

Allen, A.M., M.O. Winfield, A.J. Burridge, R.C. Downie, H.R. Benbow, G.L. Barker, P.A. Wilkinson, J. Coghill, C. Waterfall, and A. Davassi, 2017. Characterization of a wheat breeders' array suitable for high throughput SNP genotyping of global accessions of hexaploid bread wheat (Triticum aestivum). Plant biotechnology journal, 15(3): 390-401.

An, D., J. Su, Q. Liu, Y. Zhu, Y. Tong, J. Li, R. Jing, B. Li, and Z. Li, 2006. Mapping QTLs for nitrogen uptake in relation to the early growth of wheat (Triticum aestivum l.). Plant and Soil, 284(1-2): 73-84.

Beckmann, J.S. and J.L. Weber, 1992. Survey of human and rat microsatellites. Genomics, 12(4): 627-631.

Blears, M., S. De Grandis, H. Lee, and J. Trevors, 1998. Amplified fragment length polymorphism (RFLP): A review of the procedure and its applications. Journal of Industrial Microbiology and Biotechnology, 21(3): 99-114.

Börner, A., E. Schumann, A. Fürste, H. Cöster, B. Leithold, M. Röder, and W. Weber, 2002. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (Triticum aestivum l.). Theoretical and Applied Genetics, 105(6-7): 921-936.

Cui, F., A. Ding, J. Li, C. Zhao, L. Wang, X. Wang, X. Qi, X. Li, G. Li and J. Gao, 2012. QTL detection of seven spike-related traits and their genetic correlations in wheat using two related ril populations. Euphytica, 186(1): 177-192.

Cui, F., J. Li, A. Ding, C. Zhao, L. Wang, X. Wang, S. Li, Y. Bao, X. Li, and D. Feng, 2011. Conditional QTL mapping for plant height with respect to the length of the spike and internode in two mapping populations of wheat. Theoretical and Applied Genetics, 122(8): 1517-1536.

Huang, X., H. Cöster, M. Ganal, and M. Röder, 2003. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (Triticum aestivum l.). Theoretical and Applied Genetics, 106(8): 1379-1389.

Huttner, E., P. Wenzl, M. Akbari, V. Caig, J. Carling, C. Cayla, M. Evers, D. Jaccoud, K. Peng and S. Patarapuwadol, 2005. Diversity arrays technology: A novel tool for harnessing the genetic potential of orphan crops. In: Discovery to Delivery: BioVision Alexandria 2004, Proceedings of the 2004 Conference of The World Biological Forum. CABI Publishing: UK. pp: 145-155.

Jaccoud, D., K. Peng, D. Feinstein, and A. Kilian, 2001. Diversity arrays: A solid-state technology for sequence information independent genotyping. Nucleic acids research, 29(4): e25-e25.

Jiang, G.-L., 2013. Molecular markers and marker-assisted breeding in plants. In: Plant breeding from laboratories to fields. InTech.

Jones, C., K. Edwards, S. Castiglione, M. Winfield, F. Sala, C. Van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes and A. Daly, 1997. Reproducibility testing of rapid, aflp and SSR markers in plants by a network of European laboratories. Molecular Breeding, 3(5): 381-390.

Kalia, R.K., M.K. Rai, S. Kalia, R. Singh, and A. Dhawan, 2011. Microsatellite markers: An overview of the recent progress in plants. Euphytica, 177(3): 309-334.

Kato, K., H. Miura, and S. Sawada, 2000. Mapping QTLs controlling grain yield and its components on chromosome 5a of wheat. Theoretical and Applied Genetics, 101(7): 1114-1121.

Kilian, B., H. Özkan, O. Deusch, S. Effgen, A. Brandolini, J. Kohl, W. Martin, and F. Salamini, 2006. Independent wheat b and g genome origins in outcrossing Aegilops progenitor haplotypes. Molecular Biology and Evolution, 24(1): 217-227.

Wolff, K., E. Schoen, and J. Peters-Van Rijn, 1993. Optimizing the generation of random amplified polymorphic DNAs in chrysanthemum. Theoretical and Applied Genetics, 86(8): 1033-1037.

Xie, X., M.-H. Song, F. Jin, S.-N. Ahn, J.-P. Suh, H.-G. Hwang and S. McCouch, 2006. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between Oryza Sativa and Oryza rufipogon. Theoretical and Applied Genetics, 113(5): 885-894.

Xu, Y., 2010. Molecular plant breeding. Cabi.

Yan, W.-H., P. Wang, H.-X. Chen, H.-J. Zhou, Q.-P. Li, C.-R. Wang, Z.-H. Ding, Y.-S. Zhang, S.-B. Yu and Y.-Z. Xing, 2011. A major QTL, ghd8, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. Molecular Plant, 4(2): 319-330.

Zane, L., L. Bargelloni and T. Patarnello, 2002. Strategies for microsatellite isolation: A review. Molecular Ecology, 11(1): 1-16.