This thesis is presented for the degree of Master of Science of National Genomics Laboratory for Biodiversity

The Role of Teosinte *mexicana* Gene Flow in Sheath Pubescence in Mexican Highland Maize

Gabriela Carolina Cíntora Martínez
B.Sc. in Agricultural Parasitology ; M.Sc in Plant Biotechnology
August 2017

CINVESTAV UNIDAD IRAPUATO
LANGEBIO
Supervisors: Dr. Ruairidh James Hay Sawers (Coordinating)
Prof. Dr. Charles Stewart Gilmor III
Prof. Dr. Nayelli Marsch-Martínez

Thesis Declaration

I, G. Carolina Cíntora Martínez , certify that:

This thesis has been substantially accomplished during enrollment in the degree.

This thesis does not contain material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution.

No part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of CINVESTAV and where applicable, any partner institution responsible for the joint-award of this degree.

This thesis does not contain any material previously published or written by another person, except where due reference has been made in the text.

The work(s) are not in any way a violation or infringement of any copyright, trademark, patent, or other rights whatsoever of any person.

Third party editorial assistance was provided in preparation of the thesis by Rubén Rellán Álvarez

The work described in this thesis was funded by CONCACyT.

Technical assistance was kindly provided by lab 9 members for optimization of genotyping protocols, development of the mapping populations used and linkage map construction, all protocols which are described in the materials and methods chapter .

This thesis contains published work and/or work prepared for publication, some of which has been co-authored.

Gabriela Carolina Cíntora Martínez

February 2018

Abstract

Maize was domesticated 9000 years ago from the wild grass teosinte *parviglumis* (*Zea mays* subsp. *parviglumis*) in the basin of the balsas river in Mexico, after which it dispersed throughout North America, South America and the Caribbean, and later Europe, Africa and Asia. Early in this dispersion, maize colonized the central highlands of Mexico (elevations greater than approximately 2000 *masl*), which represented a drastic change in environment. In comparison to the area of domestication at approximately 1700 *masl*, the Mexican highlands are characterized by a reduction in both temperature and precipitation, as well as an increase in exposure to UV radiation. In this new niche maize came into contact with a different subspecies of *Zea mays*, *mexicana* with whom there is documented gene flow. Teosinte *mexicana* had already colonized and adapted to the highlands, and is characterized by intense stalk pigmentation as well as pubescence, traits both shared with highland maize. Here, I address the hypothesis that sheath pubescence in Mexican highland maize is the result of introgression from *mexicana*. Sheath pubescence was mapped genetically in a biparental mapping population derived from the Mexican highland landrace Palomero Toluqueño and the Midwest American inbred line B73, where two QTL were linked to pubescence on chromosomes 7 and 9. The location of the QTL was compared to previously identified regions of introgression from *mexicana* and suggests that only the effect in chromosome 9 is derived from *mexicana* introgression.

Dedication

I dedicate this work to my parents, who have not only made sacrifices in order for me to get this far, but have also motivated and supported me unconditionally. Mom, Dad, this is as much yours as it is mine. I love you.

To my brothers and sister, Rada, tío Chuby and tía Champi, for always providing random but interesting facts, perspectives, passionate discussions and lots of laughs. For keeping your deadbeat sister all these years. :)

For Michelle, maybe when you're old enough to read this you'll understand how much I love you and how important you are! I love you tiny monkey for never letting things be boring. Even when you're far away.

To all my friends. The band of renegades who never get things right the first time, or the 20th ... Misfits, thank you for the kart-wheels, midnight sprints through the street, cheap thrills on the terrace, sentimental bonfires and for being my home far-away-from home. Because we don't judge.

To those black curls. Meow. Thank you for all the dorilocos, adventures, sing your lungs out car rides, midnight constellation beers at sea level, and selfies. For obliging me to face my fears, whether climbing big rocks, flying to unknown places or letting go, but mostly for staying when things weren't okay. You'll never know how much you did.

To my lab 9 family. Prosthetic, yet equally sentimental and dysfunctional.

To Ruairidh, not only for being my boss, but for being a dear friend. Thank you for the patience, dedication and understanding even when I had none. Science will not be the same without you.

To the plot twist late in 2017. You are exactly what I needed.

To all the pandas.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Maize domestication and dispersal

Many important agricultural plant and animal species, such as rice, common bean, cotton, cattle, goats and sheep are the product of multiple independent domestication events (Second 1982, Sonnante *et al.* 1994, Wendel *et al.* 1995, MacHugh *et al.* 2001). In contrast, maize was domesticated from balsas teosinte (*Zea mays* subsp. *parviglumis* hereafter *parviglumis*) in a single well-defined event (Matsuoka *et al.* 2002), which occured in the basin of the Balsas River in Mexico, close to 9000 years ago (Piperno *et al.* 2009).

Once maize was domesticated it spread to the tropical lowlands, and Mexican Highlands, north to the US and south to South America (Vigoroux *et al.* 2008). A phylogenetic analysis of global maize diversity identified seven clusters of maize differentiation (Mir *et al.* 2013). However, these findings were not consistent with the initial results of Matsuoka *et al.* in 2002, that suggested that maize from the Mexican highlands was the most ancestral. It was later established that in fact, Mexican highland maize was placed closest to teosinte *parviglumis* due to the high levels of introgression from the related Chalco teosinte (*Zea mays* ssp. *mexicana*, hereafter *mexicana*) to the genomes of Mexican highland maize (van Heerwaarden *et al.* 2011).

The capacity for hybridization between *mexicana* and maize has been demonstrated experimentally by planting the two subspecies in close proximity (Ellstrand *et al.* 2007). An evaluation of reciprocal intermating (*i.e.* (*mexicana* x maize) and (maize x *mexicana*, the first mentioned subspecies being the female in the cross), revealed

that *mexicana* x maize hybridization occurs at less than a 1% rate, conversely maize x *mexicana* hybridization surpasses a 50%. Molecular markers have revealed substantial geneflow from *mexicana* to extant Mexican highland maize landraces (van Heerwaarden *et al.*) **Figure 1.3**.

1.2 Wild Relatives

1.2.1 Teosinte *parviglumis*

Teosinte *parviglumis* is a subspecies of *Zea mays* from which maize was domesticated (Matsuoka *et al* 2002). It is a green and glabrous annual grass that is widespread in the middle and low elevations (500-2000m) of southern and western Mexico (Buckler *et al.* 2006), in the states of Oaxaca, Michoacán, Guerrero, Jalisco and Nayarit [1]. **Figure 1.4** shows geographical distribution of *parviglumis* along with other subspecies [2].

[1] http://www.conabio.gob.mx/malezasdemexico/poaceae/zea-mays-parviglumis/fichas/ficha.htm#2.

[2] http://www.biodiversidad.gob.mx/usos/maices/grupos/Mapas_teocintles/teocintle_razas_Alta.jpg

Figure 1.1: Landrace Diversity in Mexico. Depicts the natural variation in ear shape, color, size. Image taken from CIMMyT.

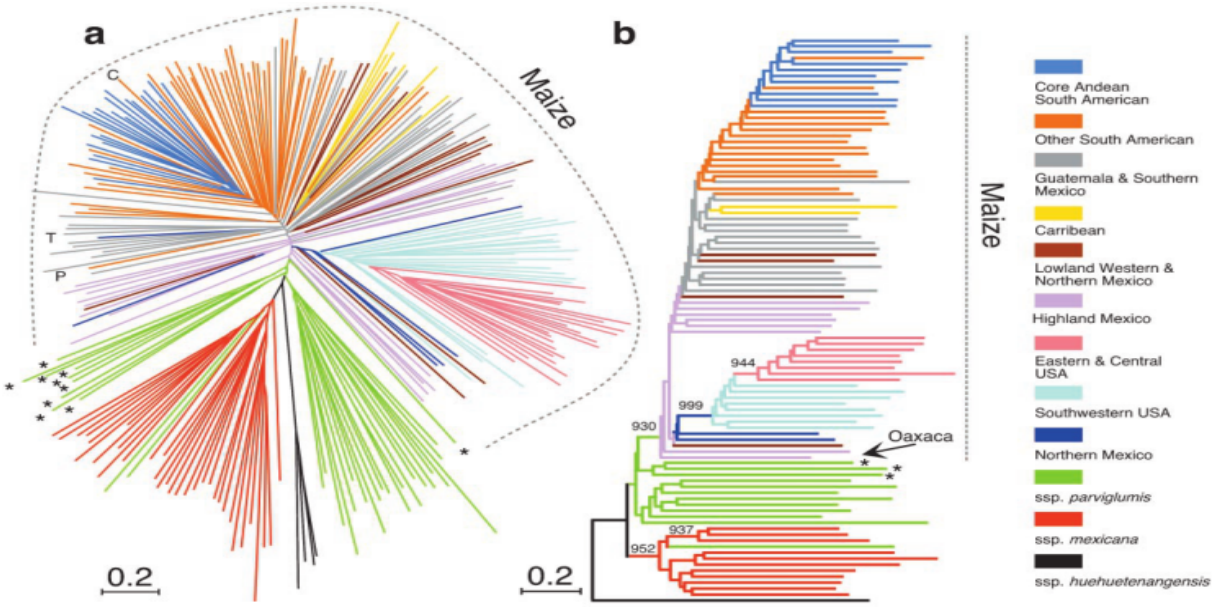

Figure 1.2: Maize and *parviglumis*Phylogeny. Tree based on 193 maize and 71 teosintes. b) Tree based on ecogeographically defined groups. The arrow points to highland maize from Oaxaca. Image taken from Matsuoka *et al.* 2002.

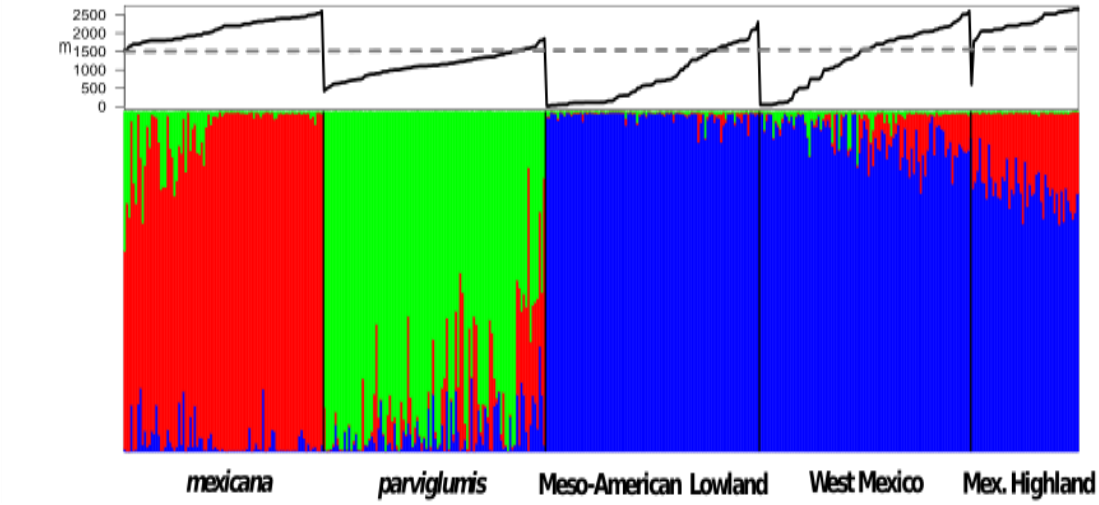

Figure 1.3: Introgression in Highland Populations. Depicts introgression from *mexicana*, *parviglumis* and *mays* in red, green and blue respectively. Mexican highland maize shows the highest levels of introgression from *mexicana*. The black line traces altitudes at which they are found inhabiting, the dotted line marks the altitude limit from which *mexicana* is present. Figure taken from van Heerwaarden *et al.* 2011.

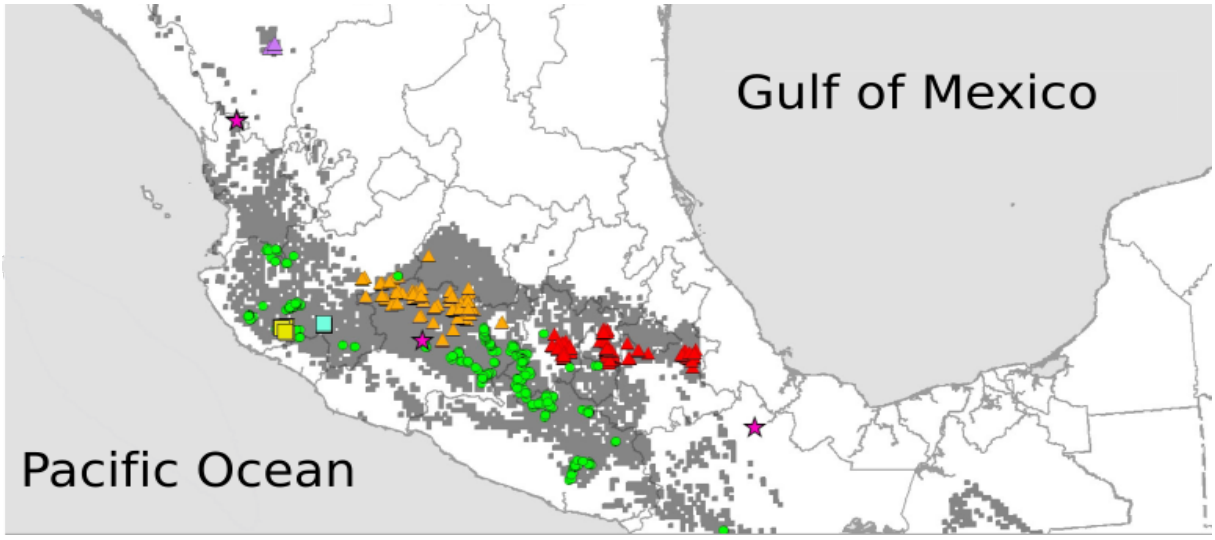

Figure 1.4: Distribution of *mexicana* and *parviglumis* subspecies in Mexico. Green dots represent *parvilgumis*, and red, orange and lilac triangles represent *mexicana*. Stars are species found in the *Zea* taxon.

As the ancestor of one of the most important agronomic crops worldwide, it is a subspecies that has made an interesting case of study, particularly with regard to the genetic basis of morphological change during and after domestication.

Cultivated maize and *parviglumis* differ in plant architecture. Teosinte plants have main stalks that branch out at each node with leaves attached, while maize typically has one main stalk and no lateral shoots. Ears in teosinte are formed in secondary branches in the axils of the leaves along the primary branches, typically clustering in groups of 1-5, each of these ears are surrounded by a single husk. Male inflorescences are present in the distal part of the main stem and lateral shoots.[3] **Figure 1.5** displays a *parviglumis* plant in vegetative phase.

The most striking difference between maize and teosinte is the ear morphology. Teosinte has approximately 5 to 10 invaginations within the raquis where seeds lie in a highly indurated outer glume, which disarticulate completely at maturity for dispersal, in contrast, maize has a shallow raquis and the glume is not hardened nor does it completely wrap around the kernel and the number of kernels produced normally exceeds 100 .[3]

[3]https://teosinte.wisc.edu/morphology.html

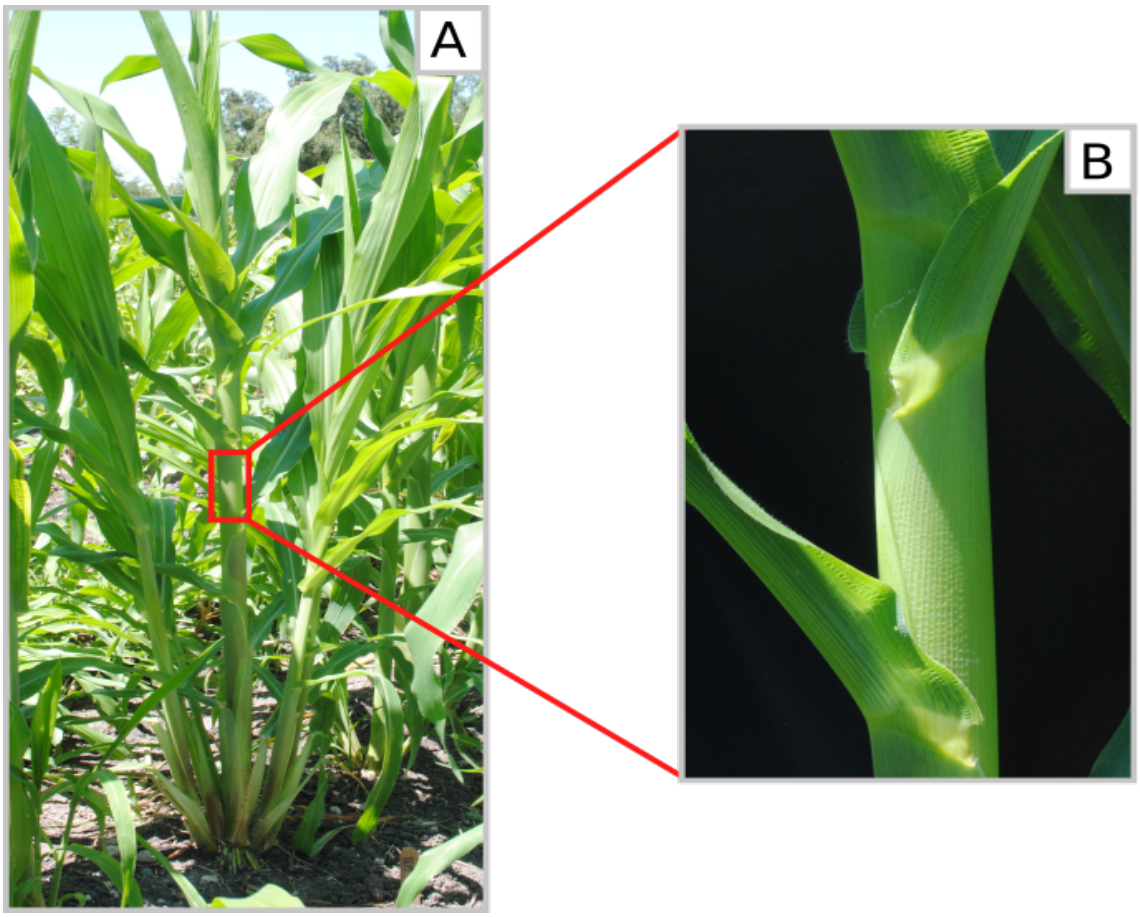

Figure 1.5: Teosinte *parviglumis* in vegetative phase. A) The main stalk along with tillers branching out laterally. B) Amplification of sheath. Image from laboratory field.

1.2.2 Teosinte *mexicana*

Teosinte *mexicana* largely resembles *parviglumis* with respect to overall morphology. *mexicana*, however, does display a number of potentially adaptive phenotypic traits absent in parviglumis, including intense pigmentation of the stem and a high density of macrohairs on the leaf sheath, which have been hypothesized to protect the plant from low temperatures and damage associated with excess UV radiation (Wilkes 1976, Lauter *et al.* 2004). **Figure 1.6** shows evident pubescence and pigmentation on sheaths.

Zea mays ssp. *mexicana* is restricted to highland populations on the Central Plateau of Mexico, *i.e.* altitudes greater than 1800 *masl* in Distrito Federal, Durango,

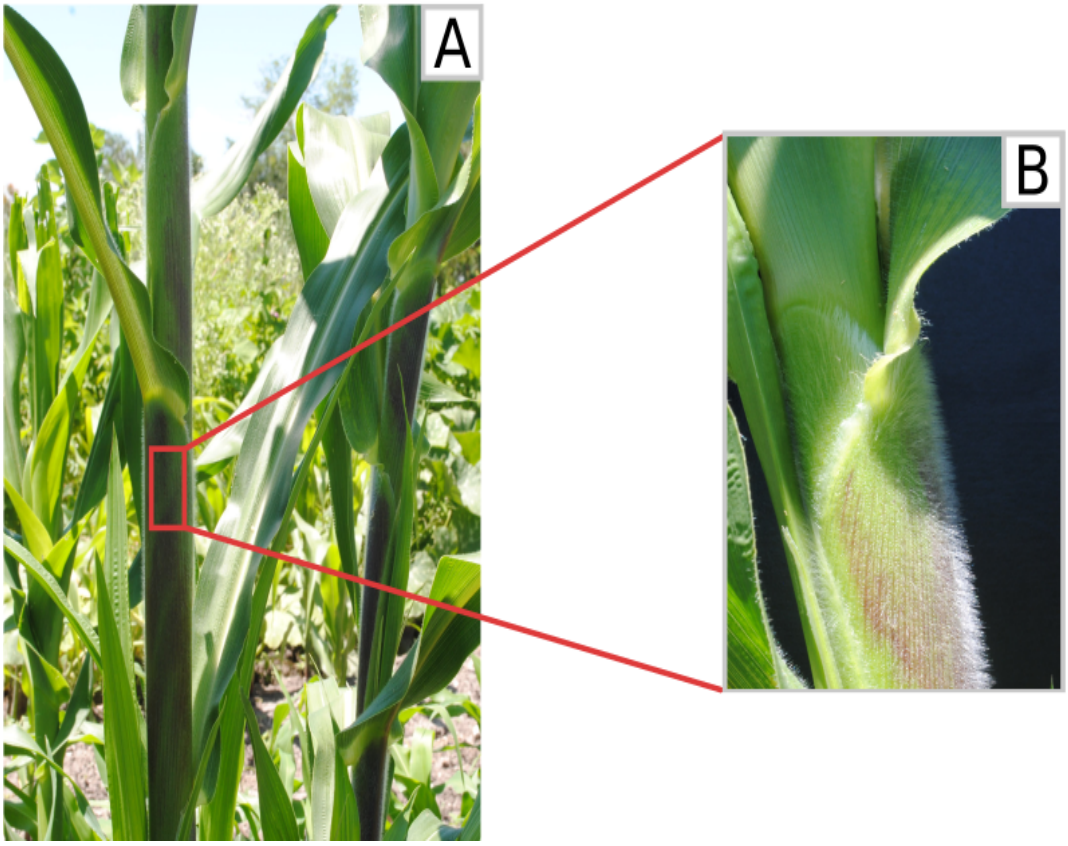

Figure 1.6: A) Stalk of teosinte *mexicana*. Unlike *parviglumis*, *mexicana* mimics a maize plant showing few tillers. B) Amplification of pubescence and pigmentation in sheath. Image from laboratory field.

Guerrero, Jalisco, Estado de México, Michoacán, Morelos, Nayarit, Oaxaca, Puebla, Quintana Roo, San Luis Potosi and Sinaloa. [4]

[4]http://www.conabio.gob.mx

1.3 Adaptation to the Highland Environment

Once maize was domesticated its dispersal occurred, leading it to colonize different environments throughout the American continent. One of the environments maize adapted to was the central highlands of Mexico. For maize to thrive in this new niche different strategies to overcome the climatic challenges that characterize the highlands were developed. The main changes in environmental factors include the reduction of atmospheric pressure and temperature as well as an increase in UV-B radiation (Körner 2007).

Studies have shown that atmospheric pressure declines by 11% per kilometer increase in altitude (Allen *et al.* 1998). This suggests that changes in atmospheric pressure declined from approximately 900 hPascals at the site of maize domestication, to 700-800 hPascals in the highlands which maize colonized. The biological implications these changes entail are respiration in animals and gas exchange in plants. A clear example in bird species is the change observed in eggshells which become more porous as altitude increases in order to promote oxygen diffusion to the embryo (Rahn 1983), as well as hyperventilation in warm blooded species along with an increase in red blood cell production (Beal 2002).

Temperature is another environmental factor which changes gradually with altitude and has been known to alter plant physiology. It is estimated that temperature decreases by 5.5 ^{o}K per kilometer ascended (Körner 2007). For the case of plants not adapted to cold or temperate environments, such as tropical or subtropical species, individuals may show severe symptoms of chlorosis or in extreme cases necrosis or death. Studies have demonstrated the impact of low temperature in unadapted maize, which when grown at low temperatures and exposed to high light intensity (like those in the highlands), suffers metabolic lesions in chlorophyll synthesis leading to increases in photodestruction of chlorophyll at rates higher than those it was synthesized (McWilliam and Naylor 1967). Needless to say, these cold-induced symptoms do not appear in highland maize which is already locally adapted. Studies of thermal acclimation and the dynamic response of plant respiration to temperature suggest that although enzyme capacity in respiration decreases with lower temperatures, it may be compensated by an increase in respiration rates (Atkin and Tjoelker 2003). Although these findings are not thoroughly conclusive, further

exploration of metabolic activity in plants subjected to cold stress coupled with fitness indicators may provide evidence that suggests this is one of the mechanisms by which plants compensate low temperature stresses.

UV-B radiation and its effects on living organisms have been subject to attention due to the significant reduction in the ozone layer that has taken place in the past decades. Studies have demonstrated the negative effects of this form of radiation in amino acids and proteins, as well as other cell components (Hollósy 2002). Perhaps one of the most detrimental effects UV-B conveys is on nucleic acids, and DNA in particular. It was previously elucidated that high mutagenesis rates were the result of exposure to UV-B however, although it was only recently established that UV-B is able to induce damage to DNA by inducing massive sequence specific-error prone-bypass repair in mosses exposed to radiation (Holá *et al.* 2015). The repair of UV radiation-induced lesions is of particular significance for replication and transcription, which may be affected resulting in cell death and/or mutagenesis (Stapleton 1992). On a morphological level UV-B stress in plants is usually reflected by a reduction in plant growth and development. However, research focused on biological adaptations to this stress suggests that anthocyanin biosynthesis in certain plant organs has the ability to mitigate DNA damage, this was established in tissue cultures of *Centaurea cyanus*. Conversely, a low concentration of anthocyanins is strongly correlated to a high incidence of DNA mutations in sorghum whereas maize and *Arabidopsis* anthocyanin-lacking mutants are hypersensitive to UV-B stress (Chalker-Scott 1999).

1.4 Maize Leaf Morphology and Development

The maize leaf is divided along the proximodistal axis into the sheath, which provides support by wrapping around the stem, and the blade, the site of the majority of photosynthetic activity. The boundary between blade and sheath is delimited by the ligule and auricule **(Figure 1.7)**. In general, the sheath is thicker in comparison to the blade, and inwardly curved so as to reinforce the whole leaf structure. The outer or abaxial part of the sheath is lined with photosynthetic cells while the adaxial is not. Both inner and outer epidermis of the sheath do not form specialized cells such as bulliform cells bicellular hairs and microhairs, with a few exceptions (Sylvester and Smith 2009) .

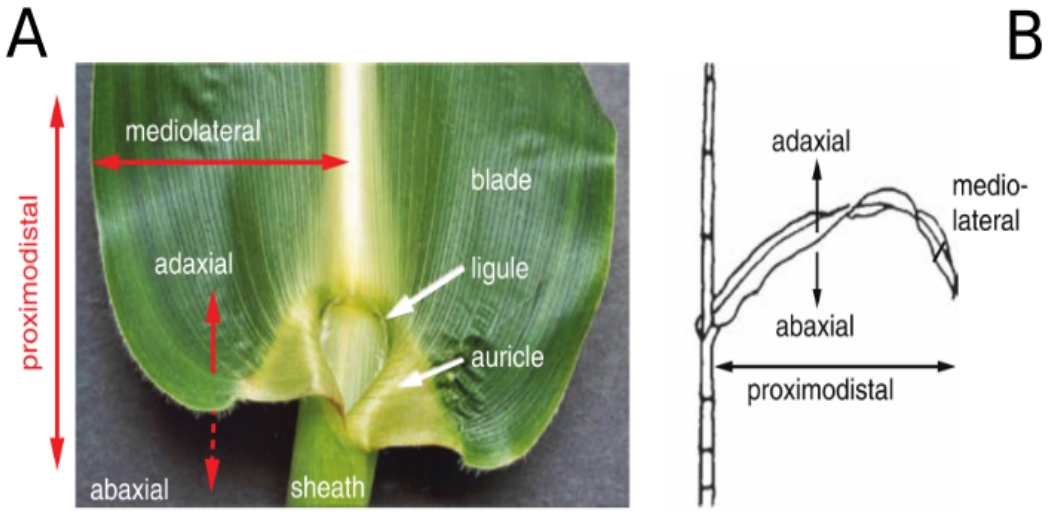

Figure 1.7: Maize Leaf Morphology. Image of the adaxial surface of the maize leaf showing distal blade and proximal sheath, separated by the ligule/auricle region. Image from Foster and Timmermans 2009.

Maize leaf development has three distinct phases: recruitment of founder cells, primordial development and post primordial cell division and differentiation. The first occurs when a ring of approximately 200 founder cells aligned in two epidermal layers are recruited and wrap around the shoot apical meristem (Poethig and Szymkowiak 1995, Schneeberger 1998). Once this has occurred, each founder cell divides consecutively in order to form a primordium which contains a complete set of lateral veins and the beginnings of the ligule. The last phase is growth and differentiation and initiates at the tip of the leaf and works its way down towards the basal cells. Adult leaf nervature arises from branches of lateral veins, which differentiate basipetally and anastomose close to the ligule. Typically one of these branched veins will cross into the sheath and join to the basipetal vascular system of the stem (Freeling 1992). All leaf domains *i.e.* blade, ligule, auricle, sheath as well as the subtending node and internode are derived from founder cells (Poethig & Szymkowiak 1995).

Leaf epidermal cells in maize are specialized or non-specialized, and cell fate is usually position dependent within each tissue. Specialized cell types in maize include three types of hairs: macrohairs, microhairs and bicellular hairs. Macrohairs emerge from multiple basal cells and are typically restricted to the adaxial surface of the leaf blade. The formation of these is likely the result of outer cell signaling pathways, this is supported by both stomata and bulliform cell formation in maize and trichome

Table 1.1: Arabidopsis genes regulating trichome initiation

Gene	Protein Family
transparent testa glabra1	WD40
glabra1	R2R3 MYB
myb23	R2R3 MYB
myb5	R2R3 MYB
glabra3	bHLH
enhancer of glabra3	bHLH
transparent testa8	bHLH
myc-1	bHLH

patterning in model plant *Arabidopsis* which have all been proven to be non lineage dependent (Larkin *et al.* 1996 , Hernández 1999).

1.5 Biology of Pubescence

1.5.1 Trichome initiation

Trichomes have been extensively studied and have been used to address basic biological questions concerning cell fate specification and differentiation. They have been morphologically classified as single-celled or multicellular, branched or unbranched, and glandular or non-glandular. In *Arabidopsis* several genes have been linked to this form of epidermal patterning and fall into two major groups, positive or negative regulators. Mutants of positive regulators fail to establish trichome cell fate, resulting in glabrous leaves. Mutants of negative regulators do induce trichomes, however these mutants develop more trichomes and/or clusters of trichomes (Reviewed in Hauser 2014, Pattanaik *et al.* 2014). Positive regulators include transcription factors from the families: bHLH (helix–loop–helix), R2R3 MYB-related, and a WD40 domain protein, which collectively form an activator complex (Balkunde *et al.* 2010). Genes acting as positive regulators are summarized in **Table 1.1** (Galway *et al.* 1994; Walker *et al.* 1999, Oppenheimer *et al.* 1991; Kirik *et al.* 2005; Payne *et al.* 2000; Zhang *et al.* 2003 and Zhao *et al.* 2012).

The activator complex formed by WD40, BHLH and R2R3 MYB proteins binds to the promoter of *GLABRA2* which codes for a homeodomain protein required downstream for production of trichomes (Pattanaik *et al.* 2014). In adjacent

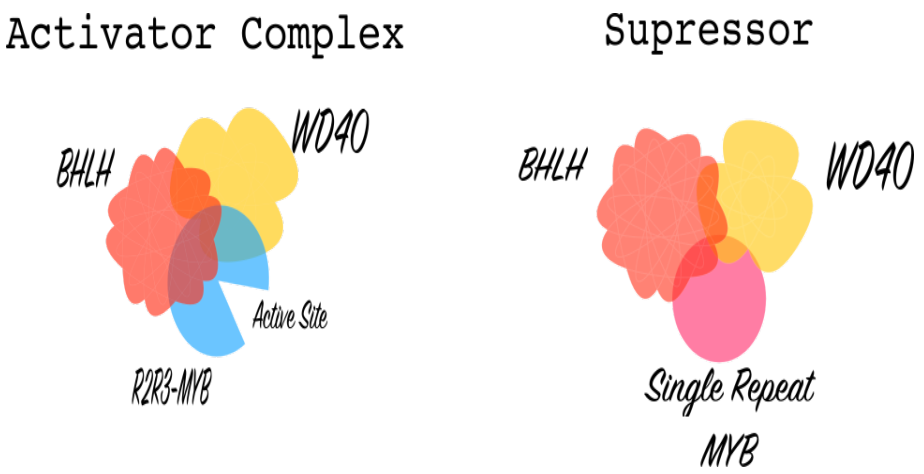

Figure 1.8: Cartoon ilustration of the activator complex required for trichome initiation. Suppressor complex substitutes the R2R3 MYB protein for a single repeat MYB that lacks the active site, therefore DNA binding is not completed.

cells a suppressor complex is formed, in which the R2R3-MYB is substituted for a single repeat MYB lacking the active site required for binding to the promoter of *glabra2*, thus regulating trichome density (Hauser 2014). **Figure 1.8** displays a cartoon version of the activator and suppressor complexes formed. Some of these transcription factors act specifically in trichome patterning while others are able to influence root hair development, anthocyanin biosynthesis and seed coat mucilage production (Nemie-Fayissa *et al.* 2014; Zhang *et al.* 2003; Song *et al.* 2009).

In maize, the *R* gene which codes for a Myc like BHLH was mapped genetically to 10L, and is responsible for anthocyanin biosynthesis in the aleurone. It has also been found to restore hairless mutant phenotypes in *ttg1* plants. However, previous attempts to map a glabrous leaf blade phenotype first observed in the K55 "pride of saline" inbred line, identified a region on chromosome 9, subsequently designated *Macrohairless1* (Moose *et al.* 2004). Mapping sheath pubescence in a cross of the subspecies, *mexicana x parviglumis* identified the same region on chromosome 9 and a second QTL on chromosome 4 (Lauter *et al.* 2004). Both of these loci remain to be fine mapped.

1.5.2 Regulation of Trichomes by Phytohormones and micro RNAs

Other factors contributing to regulation of trichomes are phytohormones, particularly giberellin (GA), Jasmonic acid (JA) and cytokinins (CK).

GA is known to have effects on seed germination, hypocotyl elongation, flowering, and trichome development. Studies with the GA biosynthetic mutant *ga1-3* have demonstrated GA to play a key role in trichome formation. Mutant *ga1-3* plants lack trichomes on leaves, however the phenotype can be rescued by exogenous application of GA (Perraza *et al.* 1998). Expression of *GL1*, *MYB23*,*GL3* and *EGL3* are induced with GA treatment, and as a result stimulates trichome initiation (Maes *et al.* 2008).

Studies in *Arabidopsis* have reported the induction of JA as a response to wounding, which in turn results in higher trichome formation (Traw and Bergelson 2003). It was also seen that JA mutants produce less trichomes in comparison to wild type plants and application of JA restores the hairs (Yoshida *et al.* 2009).

Cytokinins have also been known to act as positive regulators of trichome development. Application of 6-benzylaminopurine promotes expression of *GL1*, *MYB23* and *GL3* (Maes *et al.* 2008) resulting in more trichome production.

micro RNAS are small endogenous non-coding RNAs of approximately 20–22 nucleotides that have been proven to play an important part in trichome regulation as well as other regulatory cell processes. These micro RNAs modulate expression levels of their targets at a post-transcriptional level and in doing so balance many cellular functions (Fabian *et al.* 2010; Reviewed in Pattanaik *et al.* 2014). *SQUAMOSA BINDING PROTEIN LIKE GENES* are transcription factors that share a conserved DNA binding domain that has been proven to modulate plant growth, phase transition and trichome development (Chen *et al.* 2010). micro RNA156 has been reported as a regulator of *SQUAMOSA BINDING PROTEIN LIKE GENES*, in doing so it temporally regulates trichome development during the flowering stage (Yu *et al.* 2010).

1.5.3 Functional Perspective

Trichome morphology and functionality is diverse throughout plant species. Some trichomes act as specialized tissues producing glandular secretions, whereas other non-glandular trichomes act as physical barriers from herbivore attacks (Wagner 1991).

Several groups have investigated the functional importance of trichomes. A study by Lai *et al.* (2000) evaluated the effect of glandular trichomes on infection by late blight (*Phytophthora infestans*). This experiment evaluated glandular secretions from trichomes and identified the presence of polyphenolic compounds. Disease incidence correlated negatively with trichome density and presence of polyphenols, suggesting that these secretions conferred resistance to late blight. Previous reports regarding these secretions had identified resistance to insect predation (potato beetle and tuber moth).

Experiments in *Phillyrea* found that exposing leaves to high solar radiation induced an increase in flavonoid glycoside secretions along with an increase in trichome densities. Conversely, flavonoid glycoside concentrations and trichome densities were reduced in leaves acclimated to shade (Tattini *et al.* 2000). Similar studies in olive trees (*Olea europaea*) reported higher concentrations of flavonoids in the trichome layer in comparison to the lamina, as well as showing a significant UV absorbing capacity (Liakopoulos *et al.* 2006). Other reports in two different *Quercus* species, one with a higher trichome density than the other, found that higher trichome density was linked to higher efficiency in photosystem II in comparison to species with lower trichome density (Morales *et al.* 2002).

In *Arabidopsis*, responses to enhaced exposure to UV-B have been documented. WT plants and trichomeless mutants were subjected to intense UV radiation, after which significant decreases in height, rosette diameter, leaf size and delays in flowering were observed. Moreover, overexpression of the trichome phenotype showed a reduced sensitivity of mutant plants to UV-B radiation (Yan *et al.* 2012).

Other functions commonly attributed to trichomes are those of protection from herbivores, particularly from insect predation. In tomato *(Solanum lycopersicum)* it has been demonstrated that the mere rupture of glandular trichomes caused by

contact from insects heightens expression of plant defenses (Peiffer *et al.* 2009). In *Astralagus bisulcatus* hyperaccumulation of selenium has been observed in trichomes in young leaves, which is also speculated as a potential plant defense (Freeman 2006).

Lastly, another important role played by trichomes in birch (wood species) was elucidated. Prozherina *et al.* in 2003 found that rapid changes in epidermal differentiation towards glandular trichomes occurred as a result of exposure to freezing, that is to say, frost induced a rapid increase in glandular trichome density. It is possible that this form of epidermal patterning acts as a response to acclimation to low temperatures since trichomes serve as barriers from moisture on the leaf surface and the stomata. This may act by preventing the leaves from freezing and maintaining leaf gas exchange (Gutschick 1999).

1.6 QTL Mapping

Quantitative trait loci (QTL) mapping quantifies the association between specific regions of the genome and phenotypic traits using statistical and genetic resources. Quantitative traits are those that present a range of phenotypic variation, typically governed by polygenic effects. For example, in maize examples of traits regulated by mutiple loci include ear height (Li *et al.*2014), flowering time (Buckler *et al.* 2009), grain yield (Yang *et al.* 2016) and overall disease resistance in crops (Young, 1996). Other quantitative traits studied in different species are milk production in cattle (Zhang *et al.* 1998), blood pressure in mice (Pravenec *et al.* 1995). The variation in these traits is often due to the effects of multiple genes (loci) as well as environmental factors. Determining the location, number and identity of these loci can aid in crop breeding programs. In the case of biomedical sciences it may lead to the development of new therapeutic drugs, and in basic science it can shed light on biological and evolutionary processes.

In general, the main objectives of QTL mapping studies are to detect QTL, to obtain a confidence interval for their location, and to estimate their effect sizes (Broman 2009). There are two main issues in mapping which limit the outcome: missing data and model selection. If genotypic markers are available at each position along a genome for all the individuals in a mapping population, then mapping solely consists in identifying the regions of the genome related to the phenotype and

identifying potential interactions among themselves or with the environment (*model selection*). However, genotypic information at all positions in every individual is typically not available. Therefore, genotypes may be inferred for each individual at uniform spacing across the genome (*grid positions*) using the observed set of marker information (Broman, 2009).

There are different approaches that can be taken in QTL mapping, these vary according to the species in which the trait is to be evaluated as well as the time and resources available for the study. However, every mapping experiment starts with the development of a mapping population which needs to be genetically characterized. A typical manner of constructing these populations is using a biparental scheme and generating an F1 that can either be backcrossed to one of the parents, or selfed in order to generate an F2. The aim is to generate recombinant progeny in order to associate the desired trait to a particular chromosomal region or marker. The main limitation in the BC and F2 cross schemes is that replicates within these individuals are not possible and phenotypes can only be collected once. Overcoming this limitation is possible by generation of recombinant inbred lines (RILs). The main advantage of these lines is that they are perpetual and only need to be genotyped once. Other advantages are the higher number of breakpoints in comparison to those observed in an F2 or BC, the reduction in variation of measurements by phenotyping multiple individuals that possess the same genotype, as well as the collection of multiple destructive or invasive phenotypes (Broman, 2009). Lastly, the evaluation of traits across multiple environments is facilitated, allowing estimation of trait plasticity and $G \times E$ interactions (Broman 2009, Li *et al.* 2015). The limiting factor however, is the construction and upkeep which can result costly and time consuming.

The characterization of mapping populations has evolved quickly in the past few decades, which in turn has allowed the missing data problem to be addressed. Initially different types of biochemical and morphological markers were used, however variation caused by environmental factors and developmental stages as well as the limited number of available markers, has lead to their replacement with genetic markers (Collard *et al.* 2005, Winter and Kahl, 1995). Current markers exploit DNA polymorphisms through an ever-increasing number of molecular marker technologies including SSR, RFLP, AFLP, RAPD and SNP markers, all of which have advantages and disadvantages. SNP markers are preferred due to the reduction in costs of

next-generation-sequencing (NGS) technologies, allowing interrogation of millions of sites throughout an entire genome (Collard *et al.* 2005). High marker density is potentially the most appealing feature in NGS technologies. Previous detection of DNA polymorphisms required intense labor and resulted very time consuming since each polymorphic site had to be processed for each individual which in the end limited total marker yield.

Once genotypic markers have been obtained linkage between them must be analyzed. This refers to the arrangement of markers according to their genetic distance based on recombination frequencies observed in the genotyped individuals. Genetic map estimates can be obtained using the *rqtl* package as described by Broman (2009).

The ideal scenario in mapping optimizes both statistical power and resolution. Power refers to the ability to detect a QTL of real effect (Stange *et al.* 2013), while resolution is the precision with which QTL positions can be estimated (Vision *et al.* 2000). Interval Mapping (IM) is a simple approach which consists in testing one grid position at a time using a likelihood ratio to express statistical significance, under the assumption of the presence of a single effect (Jansen, 1993). This approach may have low power and/or a high false discovery rate as a result of not controlling for background effects throughout the genome (other QTL interacting) (Zeng, 1994).

Composite Interval Mapping (CIM) was implemented in order to overcome the limitations of IM. CIM implements the use of significantly associated markers as covariates and combines interval mapping with multiple regression. It's intended to control for genetic variation derived from other regions of the genome (Zeng, 1994). The markers used as covariates are typically identified by forward or backward stepwise regression. This not only allows the detection of multiple QTL but evaluates the significance of each multiple QTL model until it fits the observed data set best.

The standard approach employed in *rqtl* is a version of CIM. It estimates genotype probabilities given the observed markers at adjustable grid positions (typically 1 cM). A single scan tests each grid position for a QTL using the same maximum likelihood estimator as statistical significance (LOD Score) as IM. The LOD score or Logarithm of Odds, is the base 10 logarithm of the ratio between the likelihood of the presence of a QTL versus the likelihood of the absence. Therefore a LOD score of 3 implies

that the presence of a QTL in a certain grid position is a thousand times more likely than the absence. *rqtl* yields LOD scores for grid positions throughout the whole genome, however the decision to treat these as true QTL is based on a threshold which is calculated by permutation tests. Permutation tests randomize phenotypes and individual genotypes in order to estimate the highest observed LOD score after typically one to ten thousand permutations. A QTL is considered significant when it crosses the established threshold. Once an effect is detected, it is fixed into the model and a multiple QTL scan searches for additional QTL throughout the whole genome. This last step takes into account effects that may have gone undetected in the single scan.

Key points to keep in mind before constructing mapping populations are population size and resolution since these will undoubtedly affect the outcome of the analyses. A large population size (more than 100 individuals) reduces the false discovery rate or presence of "ghost QTL" since more individuals increase the robustness of a peak. In order to improve resolution several recombination events in the cross scheme are required. This is often confused with an increase in marker density. Albeit, high density markers contribute somewhat to resolution, adding more markers to few recombinants results in redundancy, and eventually these markers are discarded.

1.7 Introgression

1.7.1 Concept

Introgression is a relative term best described as the incorporation of alleles from one species into the gene pool of another (Anderson and Hubricht, 1938) through several cycles of hybridization and backcrossing. It is considered a relative term since alleles at a given locus introgress with respect to others leaving certain portions of the gene pool constant and unchanging (Harrison and Larson, 2014). It is through the constance in these static regions that the two different gene pools can be recognized.

Differential introgression is the observation that certain alleles at a locus will introgress more than others, and should be looked as a semipermeable boundary between species since it is the result of a selective process, where alleles at some loci are able to cross the boundary and alleles at other loci cannot (Harrison and

Larson, 2014), *i.e.* advantageous alleles which tend to introgress more easily than neutral ones. Under this premise, "adaptive" alleles will introgress at higher rates than neutral ones, thus allowing one to identify genomic regions that are important for local adaptation by assessing these patterns of differential introgression (Nachman and Payseur 2012).

1.7.2 Identification of Introgression

One of the long standing problems of population genetics is the ability to identify gene flow between species. Identifying introgression requires great computational power due to the large nature of genomic data sets. However this issue has been addressed through several methods which not only resolve the requirements for computational power, but also allow identification, quantification and dating of introgression events.

A simple and computationally efficient approach that is currently in use is the four-taxon test **(Figure 1.9)**. This test evaluates the excess of shared derived variants, which considers ancestral ("A") and derived ("B") alleles and is based on the the premise that SNP patterns, termed "ABBA" and "BABA", should be equally frequent under the scenario of no introgression (Durand *et al.* 2011). An excess of ABBA over BABA patterns is indicative of gene flow between two of the taxa and can be detected using Patterson's D statistic. This has been proven to be suitable for a whole genome assessment of introgression. In order to target smaller windows within the genome, the related f_d statistic is recommended (Martin *et al.* 2015).

The targeted scanning approach consists in analyzing sequence data, for a particular region using the four taxon test, originally implemented to test for admixture between modern humans and neanderthal populations, and requires the following: 1) an outgroup, 2) A population from which introgression is suspected to be derived from and 3) two populations (P1 and P2) which are going to be scanned for introgression. The scanning tests SNPs in P1 and P2 at a given region and determines which population it is shared with, whether ancestral or derived. An excess of derived alleles is indicative of introgression.

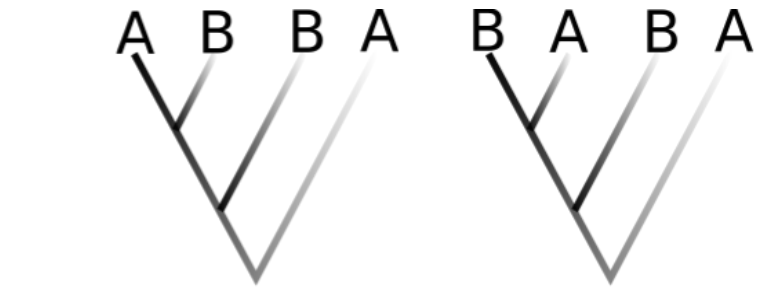

Figure 1.9: a) Patterson's D statistic for identification of introgression at whole genome levels. b) Modified Patterson's statistic f_d for targeted region scanning. The ABBA BABA trees illustrate the four-taxon test which counts the frequency for each ABBA and BABA scenario at each polymorphic site.

Chapter 2

Materials and Methods

ABSTRACT

Highland landraces Michoacan 21 and Palomero Toluqueño (PT) were crossed to the US inbred line B73 to generate a BC5S1 NIL and a BC1S5 RIL mapping population, respectively. The B73xPT BC5S1 RIL population was evaluated in two different field sites in contrasting environments, the warm lowlands of Valle de Banderas situated at 36 masl and the central highlands of Mexico, Metepec (2550 *masl*). Hair phenotypes were collected using a semi-quantitative scale that focuses on density and pattern of macrohairs along the sheath. Genotypic data was obtained using Genotyping By Sequencing (GBS). SNPs were called using the TASSEL production pipeline. Marker data was arranged into a linkage map using R software and rqtl package. QTL analyses were performed using *rqtl*.

2.1 Seed Stock and Mapping Populations

2.1.1 Recombinant Inbred Lines

The Mexican highland landrace Palomero Toluqueño (PT) accession mexi-5 and the US inbred line B73 were used to generate a BC1S5 (1 backcross, 5 generations of selfing; **Figure 2.1**) Recombinant Inbred Line (RIL) mapping population. One individual was selected at the F1 generation and was backcrossed as a male to several B73 recurrent parent females to generate a large BC1 stock, capturing a single PT haplotype from the the original open-pollinated accession. Backcross progeny carried on average 25% PT genome. Individual BC1 families were self pollinated for 5

generations to obtain an expected degree of homozygosity of 97%. The total number of 96 families were generated for evaluation.

2.1.2 Nearly Isogenic Lines

Mexican landrace Michoacán 21 was crossed to american inbred line B73 to generate an F1. The F1 was then phenotypically scored for the presence of macrohairs on leaf sheaths and progeny with said trait were backcrossed again into recurrent parent B73. This was conducted for a total of 6 generations, after which an individual with presence of macrohairs was then selfed in order for its progeny to segregate for this trait. At this stage individuals contained a total of 0.7% of landrace genome spread throughout all their chromosomes. **Figure 2.1** shows the cross scheme for the NIL.

2.2 Collection of Phenotypic Data

2.2.1 Macrohair Evaluation

PT x B73 mapping populations were planted in two separate locations. The central highlands of Mexico, Metepec, and the lowlands close to the pacific Valle de Banderas, Nayarit. Both locations were evaluated at flowering, when macrohair density along leaf sheaths was confidently scored.

Macrohairs were evaluated using a semi-quantitative scale which consists on scoring density in an ordinal manner with values ranging from 0-4 (zero being glabrous and four being the highest density), and pattern scored using a nominal scale which denotes macrohair distribution on the sheath. **Figure 2.2** represents the Density and Pattern scales.

40 BC5S1 Michoacán 21 NILs were evaluated under greenhouse conditions using the same semi-quantitative scale. In these individuals tissue was harvested for Scanning Electron Microscopy (SEM) along with tissue for RNA Seq analyses. Tissue was harvested from sheaths corresponding to leaves 7 through 10. Tissue was also collected for light microscopy, from leaf blades (proximal and distal) and sheaths, where individual macrohairs were counted with the help of Ruthenium red staining.

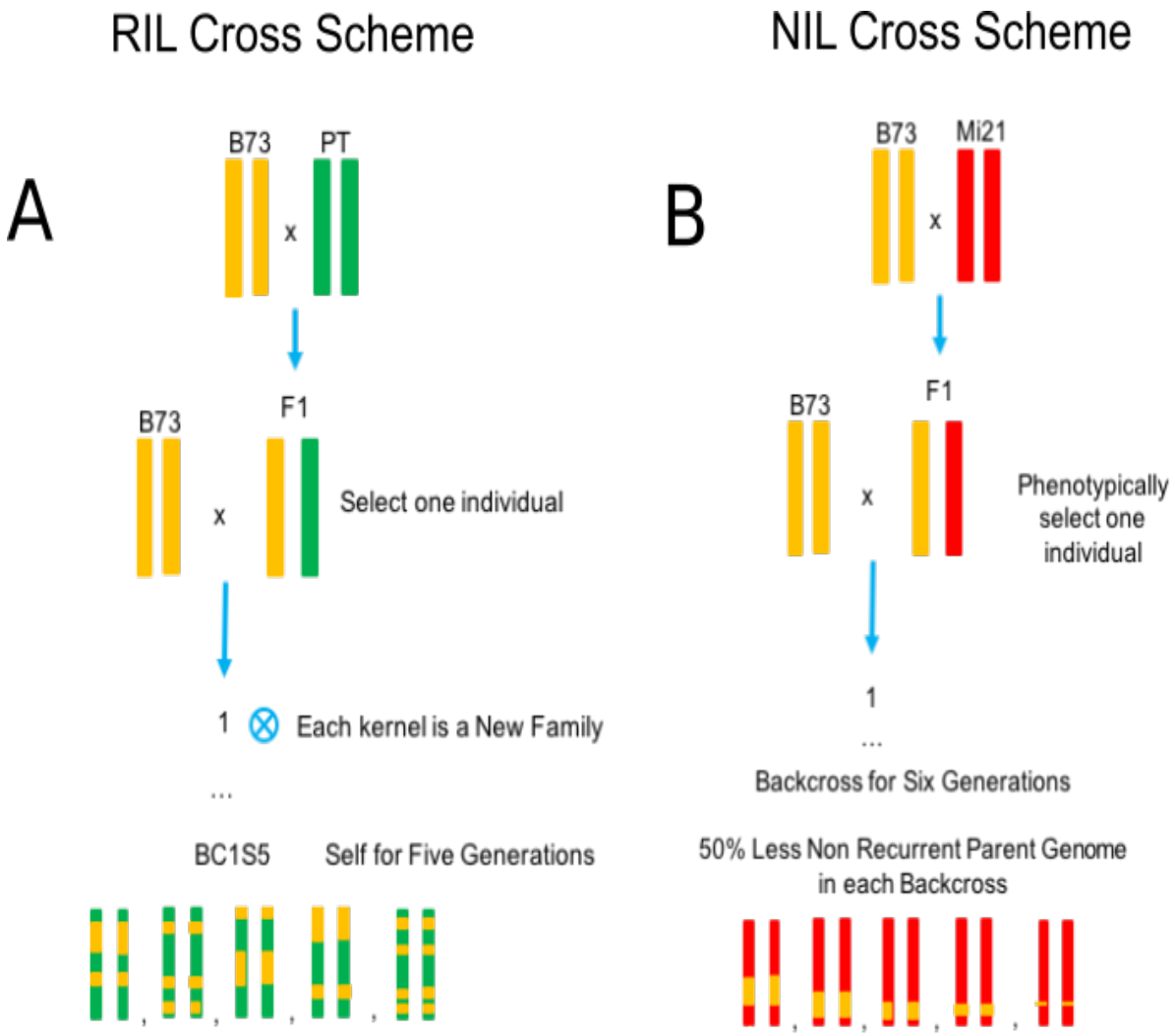

Figure 2.1: A) Recombinant Inbred Line mapping populations. Two individuals were crossed (Palomero Toluqueño PT and US Inbred line B73). One individual was selected and then backcrossed into the recurrent parent B73 leaving approximately 25% of landrace genome. Afterwards each of the progeny resulted in a new family which was then selfed down for five generations, retaining an approximate of 3% heterozygosity. B) Nearly Isogenic Lines. American inbred B73 was crossed to Mexican landrace Michoacán 21 the F1 was scored for presence of macrohairs and a hairy individual was then backcrossed to B73. Each backcross eliminates 50% of landrace genome, this was conducted for 6 generations, each one phenotypically selecting a hairy plant. Afterwards a hairy BC6 plant was selfed in order for its progeny to segregate for the trait. By this stage the BC6 individuals contained less than 1% of landrace genome and still maintained the hairs phenotype.

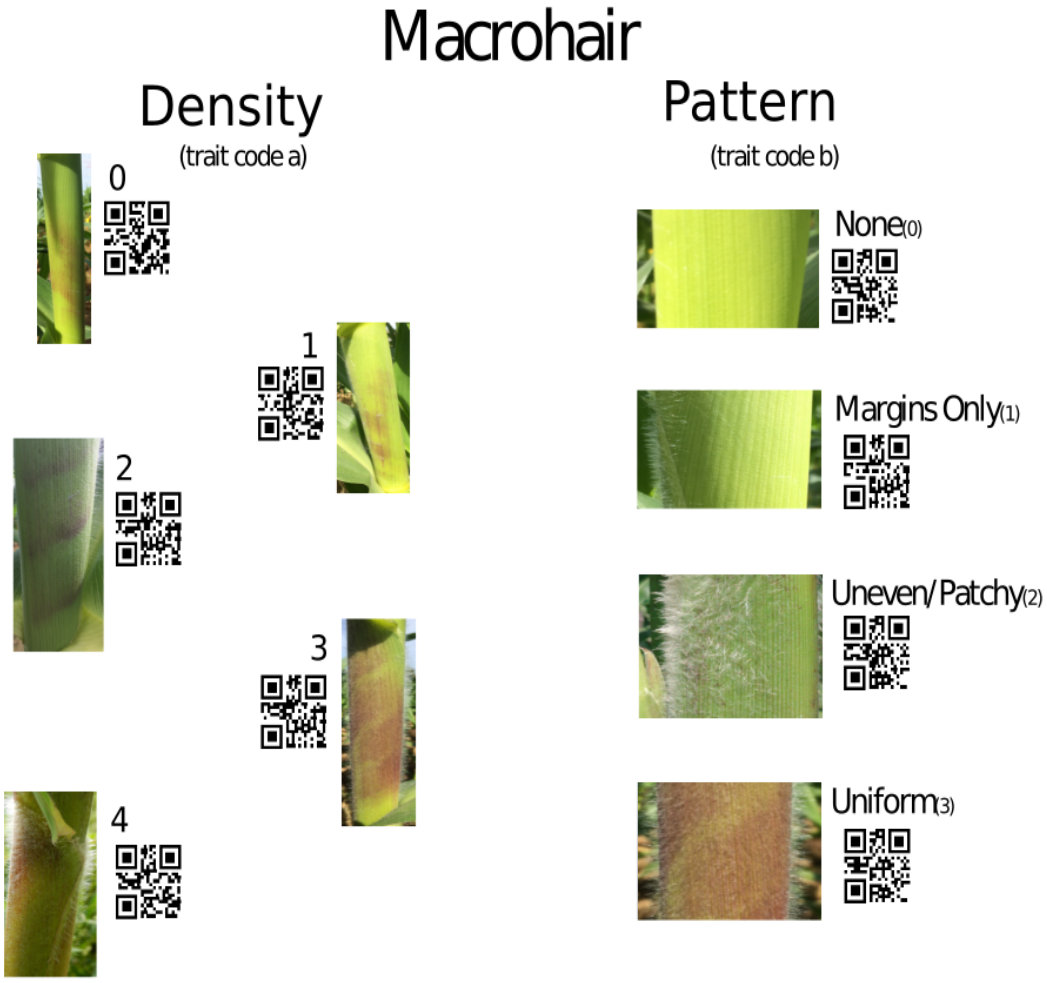

Figure 2.2: Semi-quantitative macrohair scale. Barcode scans automatically assign the value for density and pattern traits.

2.3 Genotypic Data

RIL populations were genotyped using the Genotyping By Sequencing (GBS) approach (Elshire *et al* 2011). This method consists in a reduction of genome complexity using methylation sensitive restriction enzymes. The advantage of this is that one is able to sequence regions of the genome that are being expressed, as well as target those in lower copy regions with higher efficiency.

2.4 QTL Analysis

Once marker data was obtained, the linkage map was constructed using *rqtl* and guidelines in Genetic Map Construction with R/qtl (Bromman, 2010).

1. Genome probabilities were estimated using `calc.genoprob` using `step = 1` parameter. This estimates genotypes at 1cM intervals.

2. Single QTL scan was conducted using `scanone` function based on Haley-Knott regression.

3. LOD cutoff was estimated using ten thousand permutations in Haley-Knott regression with a 5% cutoff.

4. QTL interval was obtained using the `lodint` function and the parameter `expandtomarkers = TRUE` in order to find the closest flanking SNPs.

5. The closest marker to the QTL was identified using `find.marker`.

6. Multiple QTL scan was done using `addqtl` by fixing the effect on chromosome 7 into the function, using grid position 101. The peak with the next best LOD score was included into the analysis and evaluated using `fitqtl` to see if the additional QTL fit the model.

7. QTL positions were re-evaluated using `refineqtl` to see if there were significant changes in grid positions.

Chapter 3

Results

ABSTRACT

A total of 96 families from two RIL populations were scored for presence of macrohairs, in highland and lowland locations, over two years. Genotypic data was obtained for 54 RIL families using Genotyping by Sequencing approach, and 2081 SNP markers were used to build a genetic linkage map. Single QTL analyses identified a QTL on chromosome 7 associated with the presence of stem macrohairs. Multiple QTL analysis identified a second effect on chromosome 9. Together,the QTL on chromosomes 7 and 9 explained approximately 60 % of the phenotypic variation observed. The QTL on chromosome 9 co-localizes with a previously reported region of introgression from *Zea mays* ssp. *mexicana* containing the *macrohairless* locus. The larger effect QTL on chromosome 7 does not colocalize with any previously reported region of introgression. On the basis of the B73 reference genome, a number of candidate transcription factors were identified to be present within the QTL intervals.

3.1 Field Evaluation

A total of 2304 plants we individually scored for presence of macrohairs along leaf sheaths. Table **3.1** summarizes the total number of plants evaluated from different mapping populations in the two given sites.

Table 3.1: Summary of Field Evaluations

No. RILs Phenotyped	96
Total No. Individuals Phenotyped	≥ 2300
Locations	Puerto Vallarta, Metepec
Individuals/RIL Phenotyped	≥ 24

Table 3.2: Cross Summary from rqtl

No. of RILs	54
% Phenotyped	100
% Genotyped	97.6
Number of Chromosomes	10
% Palomero Toluqueño	21.9
% B73	78.1

3.2 QTL Mapping

3.2.1 Exploration of Genotypic Data

Interpretation of Genotyping by Sequencing (GBS) results shows that the observed percentage of each of the parental lines in all of the RIL families (78.1% B73 21.9% Palomero Toluqueño) meets the expected value (75% B73 25% PT) and is presented in **Table 3.2**. In a first instance half of the mapping populations were sent out for sequencing, data from the remaining is still pending.

Checking for marker order is a critical step. Incorrect placement of genetic markers can associate QTL to chromosomal regions where in fact the QTL are not located in, which may destroy the results of the QTL analysis (Bromman, 2009). Another equally important aspect of the genetic markers to keep in mind, is the presence of mislabeling errors. This refers to errors in marker labels so they do not match the true markers that were genotyped. Further exploration of marker data referring to order and genotyping errors, indicates the presence of a minimal amount of markers placed incorrectly **(Figure 3.1)**, which were either corrected or discarded, as well as few individuals with genotyping errors, in this case genotyping errors may appear as apparent tight double crossovers.

Figure 3.1 shows that the majority of markers are aligning to their respective chromosome, *i.e.* markers labeled as "chromosome 1" are in fact aligning there and

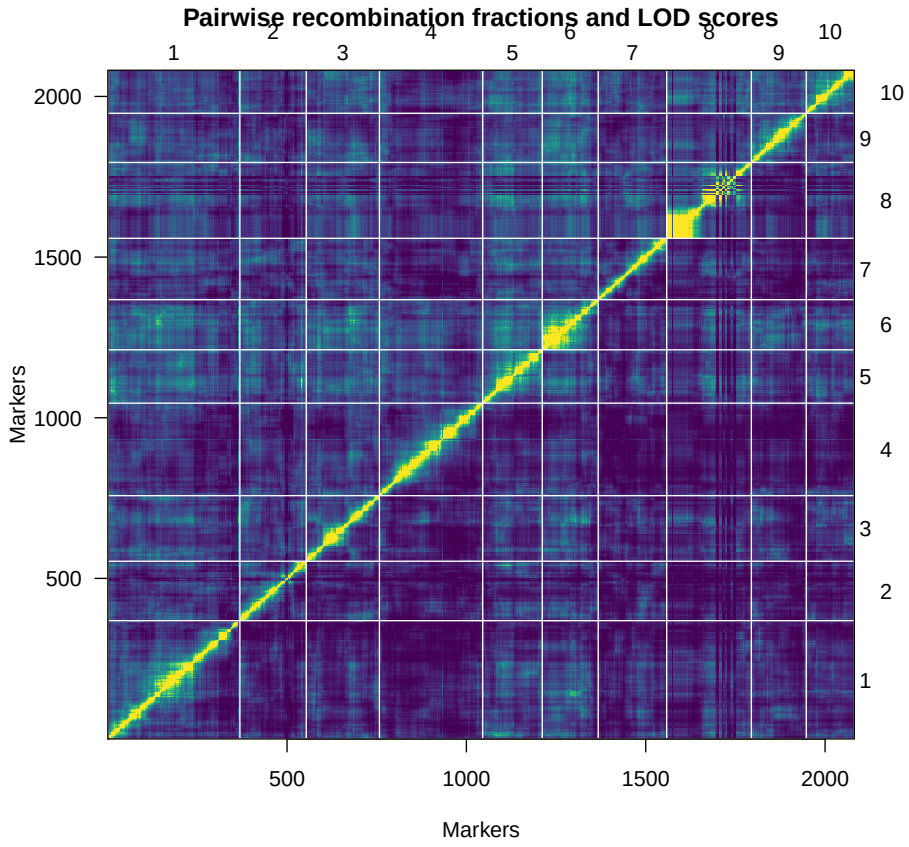

Figure 3.1: This figure shows on the top and right the 10 chromosomes in maize, the remaining borders refer to the markers obtained by genotyping. The tendency shows that each marker is aligning to its respective chromosome, and the color represents the intensity of the correlation. Chromosome 8 seems to have a few markers that align elsewhere, however the effect of this is minimal.

not elsewhere. This is the case for all the markers, except on chromosome 8 where there appear to be some markers not correlating exactly and exclusively. This could be the result of a genotyping error, in which case said markers were discarded.

Once the markers were imputed and arranged into proper order, the map estimates were obtained for the number of markers remaining, in this case 2081 SNPs were used to build the linkage map. The genetic map estimate is presented in **Figure 3.2**, which clearly shows the arrangement of markers. Density of markers per chromosome is visible, and for the case of chromosome 8, there appear to be gaps in certain sections. This is not necessarily the result of improper imputation or errors in genotyping, it

may be indicative that there were no recombination events in between these gaps, and therefore no markers to align there.

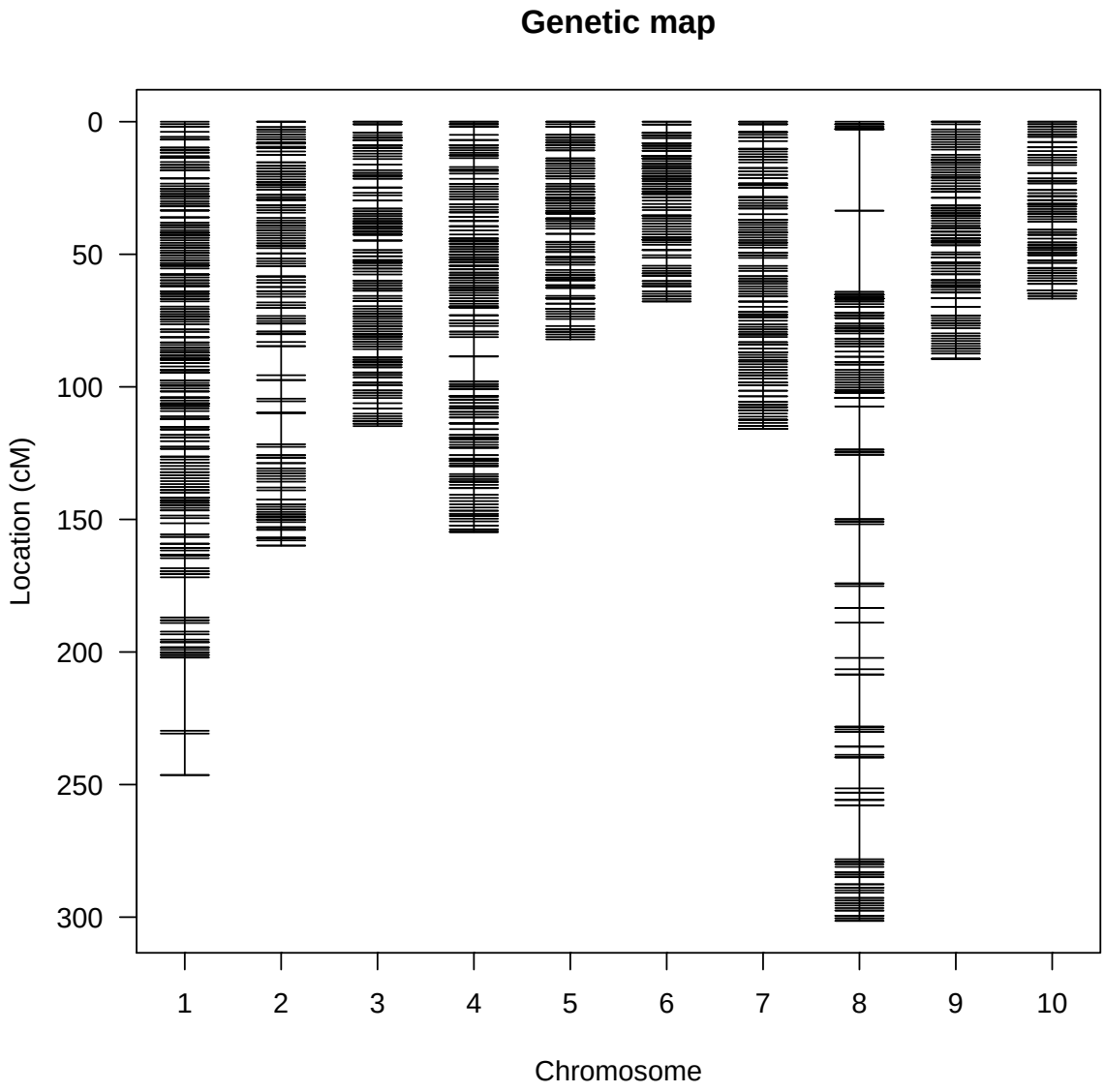

Figure 3.2: Genetic Map constructed using a total of 2081 SNP markers. The distance between markers is represented by cM.

Table 3.3: Summary of Single QTL Scan

SNP	Chromosome	Position	LOD
S1_179707307	1	120	2.308
c2.loc25	2	25	0.779
S3_194254890	3	145.03	1.235
S4_191045584	4	155.44	0.958
S5_67673484	5	79.84	0.955
S6_161010798	6	112.81	0.541
c7.loc101	7	101	4.697
S8_2902011	8	0	1.004
S9_92916094	9	69.42	2.382
S10_1905196	10	2.86	0.593

3.3 QTL Analysis

3.3.1 Single QTL Scan

Using the `scanone` function in rqtl package, LOD scores were estimated at 1cM intervals throughout the genome. **Table 3.3** summarizes the highest LOD scores for all chromosomes, along with the SNP marker and position number in the grid. The chromosomes with the highest LOD scores are chromosome 7 and 9 with 4.69 and 2.38 LOD values respectively.

LOD scores estimated at 1cM grids using Haley-Knott regression are plotted in **Figure 3.3**. After ten thousand permutations the threshold was set at 3.59 for a 5% cutoff. A strong signal on the long arm of chromosome 7 is the only peak to have crossed the threshold with a LOD value of 4.69

A LOD score represents the ratio of maximum likelihood estimators (MLE) between the odds of a QTL being present in a given grid position, versus the odds of no QTL being present at the same grid position, therefore a higher LOD score states that it is more likely for a QTL to be present rather than absent. This is the case for the QTL identified on chromosome 7, a LOD score of 4.69 implies that the odds of a QTL being present versus it the odds of it being absent in a given position, are four times more likely given the observed data.

3.3.2 Confidence Interval and Position Estimates

Table 3.4 shows the 99% probability estimate for the position of the QTL on chromosome 7. The interval is defined by position 95 in the grid which has a LOD score of 3.01 and position 121.21 with a LOD score of 2.76. Physical position estimates are in **Table 3.5** where the closest flanking markers are used to establish the base-pair interval which ranges from approximately 138 million bp to 158 million bp.

3.3.3 Multiple QTL Scan Model

After taking a closer look at the initial peaks derived from the `scanone` function **(Figure 3.3)**, it appears that there is an additional effect on chromosome 9 that could represent a potential QTL . In order to consider this new effect as a potential QTL the function `addqtl` was used and plotted **(Figure 3.4)**. This function calculates a new set of LOD scores **(Table 3.6)** taking into account the previously detected one in `scanone`. Recalculated LOD scores show a clear increase from 2.38 to 3.59 for the effect on chromosome 9. Position estimates were redefined using `refineqtl`, the effect on chromosome 7 was shifted from position 101 to position 102 in the grid.

The effect of each haplotype (B73 or PT) on each of the QTL was estimated using the effect plot provided by *rqtl* **(figure 3.5)**. Individuals are grouped by haplotype in both chromosomes 7 and 9, while plotting the hair index. By doing this it is evident that individuals with a B73 haplotype in both QTL are significantly less hairy than those with PT haplotypes in either chromosome 9 or 7, with the latter apparently having a larger effect on hairs than chromosome 9. Finally, individuals with with PT on both QTL are the "hairiest", however this is based on two observations.

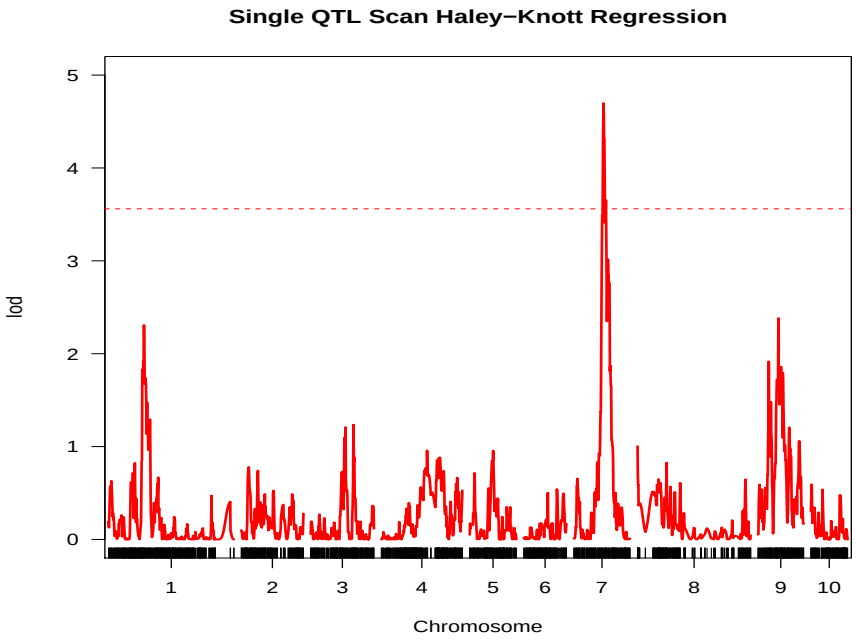

Figure 3.3: Single QTL Model Using Haley-Knott Regression

Table 3.4: 90% Probability QTL Position Estimates

Marker	Chromosome	Position	LOD
c7.loc95	7	95.00	3.01
c7.loc101	7	101.00	4.69
S7_161434966	7	121.21	2.76

Table 3.5: SNP Markers Flanking QTL on Chromosome 7

Physical Position	SNP	Chromosome	Position	LOD
138,606,362	S7_138606362	7	94.12	2.56
.	c7.loc101	7	101	4.69
157,455,779	S7_157455779	7	111.19	2.85

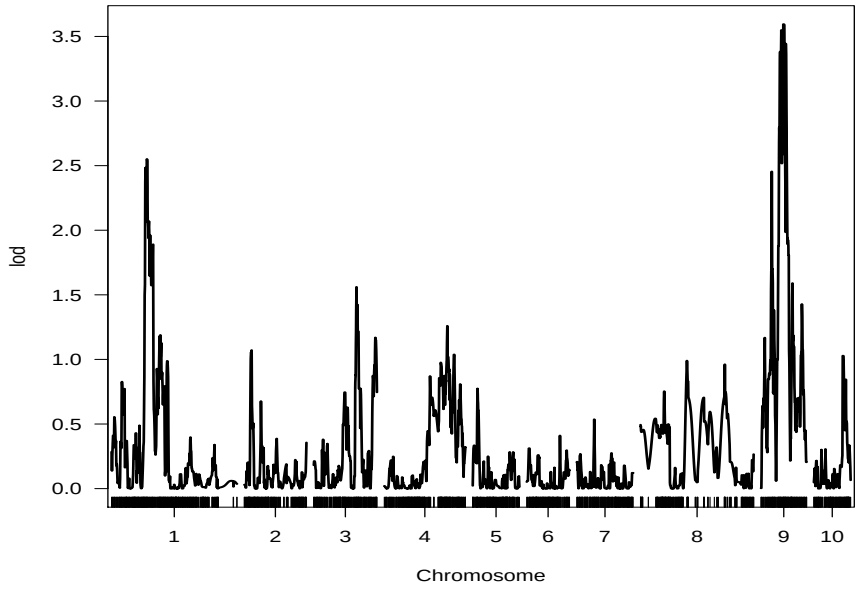

Figure 3.4: Additional QTL Scan using `addqtl`

Table 3.6: Summary of `addqtl` Output

SNP	Chromosome	Position	LOD
S1_79707307	1	120.2	2.551
S2_7950068	2	24.7	1.071
S3_194231426	3	145.0	1.561
S4_224623977	4	214.2	1.259
S5_4608846	5	16.2	0.774
S6_161010798	6	112.8	0.410
S7_100805625	7	59.2	0.536
S8_149888531	8	157.6	0.990
S9_108403546	9	76.8	3.594
10_140737865	10	98.0	1.028

After having fixed the effect of chromosome 7 and 9 the full model becomes $y \sim Q1 + Q2$ with a significance level of 0.001, this supports the model that states the presence of two effects that explain a total of 50% of the phenotypic variation observed **(Table 3.7)**. In this case there is a difference in effect size, where the QTL on chromosome 7 accounts for approximately 36% of variation and the QTL on chromosome 9 17.69% of variation **(Table 3.8)**

Table 3.7: Full QTL Model $y \sim Q1 + Q2$

.	df	SS	MS	LOD	%Var	Pvalue(Chi2)	PvalueF
Model	2	21.30	10.65	8.29	50.71	5.04e-09	1.45e-08
Error	51	20.69	0.40				
Total	53	41.99					

Table 3.8: % Variation Explained by Each effect

Marker	LOD	%var	Pvalue(F)
7@102.0	6.434	36.04	1.40e-07
9@76.8	3.594	17.69	8.32e-05

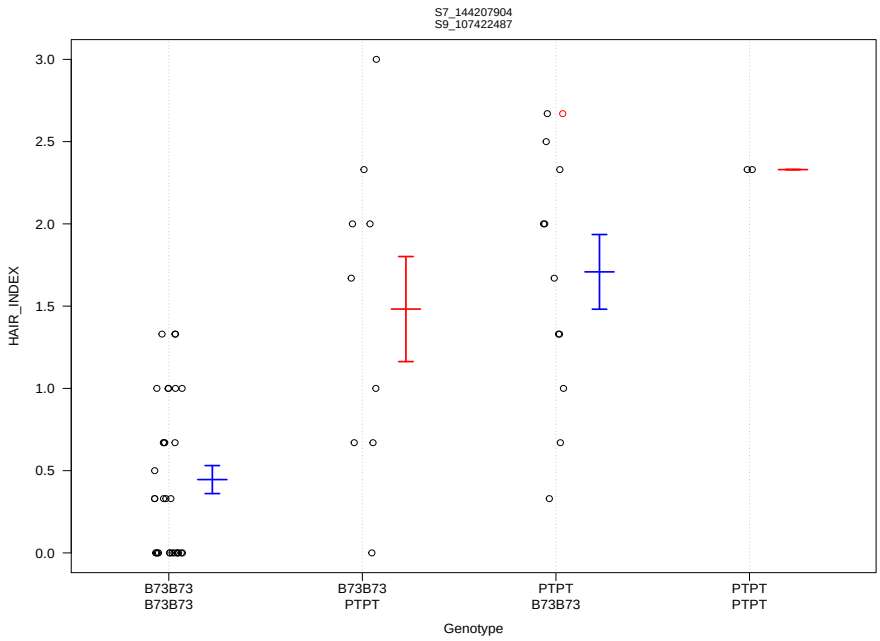

Figure 3.5: Effect Plot. This plot shows the effect of each haplotype on the hair index for each chromosome (7 and /or 9). The top SNP refers to chromosome 7 and the bottom to chromosome 9.

Chapter 4

Discussion

4.1 QTL on chromosomes 7 and 9 were linked to sheath pubescence in Palomero Toluqueño

Single QTL analysis identified a QTL of large effect on chromosome 7 which explains approximately 50% of the phenotypic variance observed in sheath pubescence. Based on the physical position of SNP markers in the B73 reference genome, the QTL interval was estimated (90% confidence; closest flanking markers) to be between 138 Mb - 157 Mb. The QTL on chromosome 7 was fixed into a multiple QTL model in order to search for additional QTL, identifying a QTL on chromosome 9. The QTL on chromosome 9 was detected using the `addqtl` function, which identified the "next best" peaks with the highest LOD scores throughout the genome taking into account the effect on chromosome 7. The closest flanking marker to the QTL on chromosome 9 is located at approximately 108 Mb, and the 90% interval is located from 10 Mb - 151 Mb. The broad interval for chromosome 9 is due to the relatively low number of individuals in the mapping population associated with a limited number of recombinants in the region of the QTL.

4.2 The QTL on chromosome 9 co-localizes to with the locus *Macrohairless1*

In 2004 Moose *et al.* identified a locus linked to a "macrohairless" phenotype on leaf blades from a US inbred maize line. This locus was named *Macrohairless1* (*Mhl1*)

and was genetically mapped to chromosome 9, 6.7cM from *glossy15* and 14.7 cM from the SSR *umc1120* (physical positions at approximately 95 Mb and 122 Mb, respectively), spanning the peak of the QTL. Although the *mhl1* mutation is linked to loss of MH on leaf blade and the PT QTL is linked to sheath pubescence, the underlying variants may be alleles of the same gene. The production of MH in additional domains of the leaf (abaxial sheath, compared with the "typical" adaxial blade) may be explained by a potential gain of function or neomorphic allele of *Mhl1*. Such "gain of function" has been previously reported in maize for the *B* gene, linked to anthocyanin biosynthesis. The *B-I* allele pigments leaf blades, sheaths and tassels while the *B-PERU* allele pigments those tissues as well as the aleurone (Ludwig 1990). This may be the case of *Mhl1*, where a different allele is able to produce MH in a different site.

In 2004 Lauter *et al.* identified 5 QTL linked to sheath MH in a *mexicana* x *parviglumis* F2. Four of these QTL had minor effects and were situated in 5S, 2C, 10L and 4L. When combined these effects explained close to 20% of the phenotypic variation observed while another in 9L explained 46% of the variation. This major effect appears to be syntenic with both *Mhl1* and the QTL on 9L reported in this work. It is possible that the loci in teosinte and this work are the same given the co-localization of the QTL with regions of introgression.

4.3 The Origin of Allelic Diversity

Two large effect QTL linked to sheath MH were identified in the B73 x PT population, located on chromosomes 7 and 9, respectively. Interestingly enough, although the QTL on chromosome 9 co-localizes with signals of *mexicana* introgression (**Figure 4.1**; Hufford *et al.* 2013), the QTL on chromosome 7 does not, raising the question as to the origin of this allele. Cloning of genes underlying the major effect QTL would allow targeted sequencing and the use of gene-level population genetic approaches to compare hypotheses of *mexicana* origin, *de novo* mutation or presence in ancestral *parviglumis* populations. For example, if all highland accessions of both maize and *mexicana* share the same architecture and the result of ABBA/BABA scanning in these QTL denotes introgression, then it is likely that this trait was derived solely from *mexicana*. Conversely, if the architecture is not shared between highland maize and *mexicana*, it would suggest phenotypic convergence. This would

open the possibility for one of the following scenarios: i) the development of two independent mutations or ii) natural selection of standing variation from an ancestral population *i.e.* the existence of a rare MH allele in ancestor *parviglumis*.

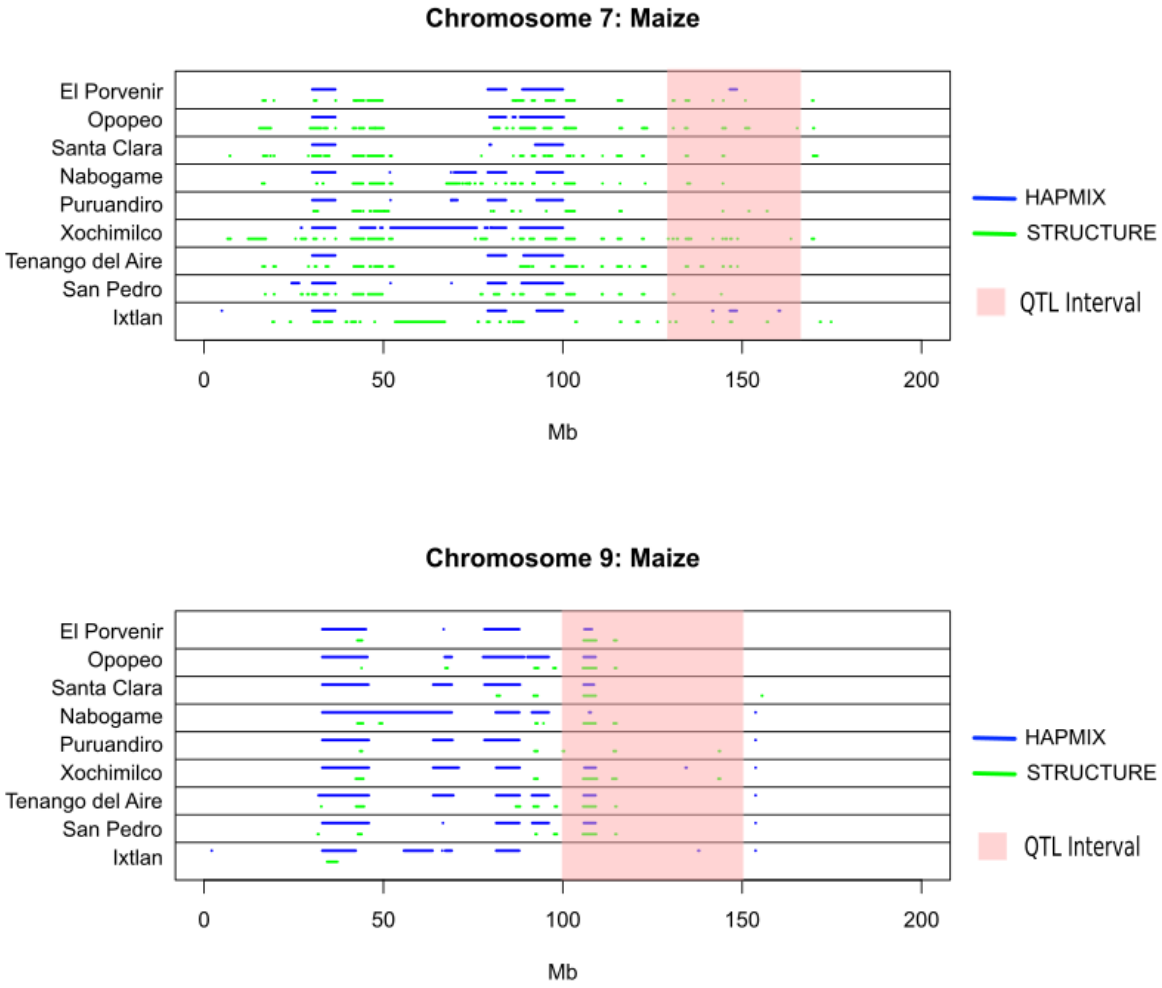

Figure 4.1: Introgression map modified from Hufford *et al* 2013. Introgression of teosinte *mexicana* was estimated in regions throughout chromosomes 7 and 9 in several maize populations using two different softwares (HAPMIX and STRUCTURE). Introgression is plotted using blue and green lines respectively. Intervals for the QTL detected are denoted in the pink shaded areas.

Table 4.1: Levels of Expression of Transcription Factors within QTL Intervals in Leaves 8, 11 and 13.

Transcription Factor	Gene ID V2	Chr.	L8	L11	L13
MYB_RELATED	GRMZM2G103783	7	117.18	100.02	90.38
BHLH	GRMZM2G435015	7	0	0	0
BHLH	GRMZM2G153454	7	99.14	64.12	74.60
BHLH	GRMZM2G089501	7	68.02	95.38	152.75
BHLH	GRMZM2G013688	7	52.99	35.58	0
BHLH	GRMZM2G144275	7	171.84	117.28	125.30
BHLH	GRMZM2G085467	7	59.77	38.54	34.72
BHLH	GRMZM2G058451	7	243.37	257.42	205.77
MYB	GRMZM2G104551	7	22.13	154.78	205.75
MYB	GRMZM2G172327	7	0	0	0
MYB	GRMZM2G031323	7	6.86	0	0
MYB_RELATED	GRMZM2G034110	9	360.63	72.9	75.32
MYB_RELATED	GRMZM2G170148	9	74.77	56.47	62.68
MYB	GRMZM2G098179	9	0	20.72	34.43

4.4 Identification of the Potential Loci Involved

In the model plant *Arabidopsis thaliana*, trichome initiation is regulated by the combined action of transcription factors of the WD40, MYB, MYB-related and BHLH families (Larkin *et al.* 1994). The genes transparent testa glabra *(TTG1)* and glabra1 *(GLAB1)* encode WD40 and MYB-related transcription factors, respectively (Walker *et al.* 1999 and Oppenheimer *et al.* 1991), whose mutant phenotypes are characterized by the absence of trichomes on leaves and stems. In maize, there are a number of genes encoding for WD40, MYB and MYB-Related transcription factors within the chromosome 7 and 9 QTL intervals **(Table 4.1)** [1].

There are 15 potential candidate genes that code for transcription factors in the MYB, MYB-related and BHLH families within the QTL intervals, therefore locating the underlying gene(s) is still premature. In order to further identify MH candidate genes, a BC6S1 NIL family was developed by recurrent crossing of a pubescent landrace donor to B73, selecting for sheath MH at each BC generation. The BC6S1 family segregated 3 pubescent : 1 glabrous individual, indicating a single, dominant factor to be regulating production of sheath MH. Pubescent and glabrous individuals

[1] http://grassius.org/grasstfdb.php

will be analysed using genotyping-by-sequencing to identify the region segregating. Leaf sheath tissue was harvested from both pubescent and glabrous individuals for future RNAseq analysis, that will both further indicate the segregating region (based on identification of sequence polymorphisms) and provide insight into the transcriptional events associated with sheath MH initiation.

4.5 Biological Implications of Pubescence in Maize

Sheath pubescence is a trait observed at high frequency in both Mexican highland maize and the highland populations of both maize and the highland maize wild-relative *mexicana*. The prevalence of this trait raises questions regarding it's biological function in plants. The highland environment subjects plants to reduced atmospheric pressure, increases in UV-B radiation and low temperatures.

Trichomes serve numerous functions in a wide range of plant species, primarily as protection against both biotic and abiotic stresses. For the case of highland maize, MHs may represent an adaptation to low temperatures, providing thermal insulation from the cold weather. In birch trees, trichome density on leaf surfaces increased temporally as cold season progressed (Prozherina *et al.* 2003).

Summary and Conclusions

The combined use of QTL mapping and introgression maps has proven to be an efficient manner to determine key aspects in the genetic architecture underlying local adaptation to the highlands. Using a biparental mapping population, two major effect QTL which combined accounted for close to 60% of the phenotypic variation observed in sheath pubescence, were mapped to 9L and 7L. There appear to be additional small effect QTL, however these were not well supported in the relatively small mapping population, and level of replication, used.

Conservative estimates of the physical position of the QTL defined intervals of approximately 30 Mbs. Mapping resolution was limited by the number of recombinants available in these regions, reflecting the population size. Generation of additional recombination in these regions will be required to fine map the QTL. The list of annotated genes within the physical interval of the QTLs includes transcription factors of the BHLH, MYB and MYB-Related families, which in *Arabidopsis* have been shown to assemble a multimeric complex regulating trichome initiation.

The QTL on chromosome 9 co-localized to a previously defined region of *mexicana* introgresion, the QTL on chromosome 7 did not. This result suggests that sheath pubescence in Mexican highland maize arose as the result of both introgression from *mexicana* and from direct descent, whether drawing on standing variation present in ancestral *parviglumis* populations of through *de novo* mutation. Identification of the genes underlying these QTL and determination of causative variation would facilitate definition of the origin of allelic variants. In general, mapping putatively adaptive traits in Mexican highland maize will ascertain the importance of gene flow from wild relatives in local adaptation.

Perspectives

Understanding of the mechanistic basis of sheath pubescence, and the origin of associated genetic variation, would be greatly facilitated by identification of the underlying genes. Attempts to fine map QTL will be based on 1) generation of further recombination events in the QTL intervals by additional back crossing; 2) mapping in a second B73 x highland maize population (currently available); 3) GWAS analysis of a maize landrace diversity panel. Generation and subsequent crossing of near isogenic lines will allow further exploration of possible epistasis between QTL. Such material will provide also the basis for functional studies, including evaluation of the effect of sheath pubescence on plant performance.

Bibliography

1. Allen, R.G.*et al.* Crop evapotranspiration. Guidelines for computing crop water requirements. *Irrigation and Drainage*, paper 56, FAO (1998)

2. Atkin, O. K. & Tjoelker, M. G. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends Plant Sci.* 8, 343–351 (2003).

3. Anderson, Edgar, and Leslie Hubricht. Hybridization in Tradescantia. III. The Evidence for Introgressive Hybridization. *American Journal of Botany* vol. 25, no. 6, 1938, pp. 396–402. JSTOR, JSTOR, www.jstor.org/stable/2436413.

4. Balkunde, R., Pesch, M. & Hülskamp, M. Trichome Patterning in Arabidopsis thaliana. *Curr. Top. Dev. Biol.* 91, 299–321 (2010).

5. Beall, C.M. Biodiversity of human populations in mountain environments. *Mountain Biodiversity.A Global Assessment*. pp. 199–210, Parthenon. (2002)

6. Becraft, P. W., Kang, S.-H. & Suh, S.-G. The Maize CRINKLY4 Receptor Kinase Controls a Cell-Autonomous Differentiation Response. *Plant Physiol.* 127, 486–496 (2001).

7. Bennetzen, J. & Hake, S. Handbook of Maize: Its biology. Handbook of Maize Its biology (2009). doi:10.1007/978-0-387-79418-1

8. Boyd, R. S. Plant defense using toxic inorganic ions: Conceptual models of the defensive enhancement and joint effects hypotheses. *Plant Sci.* 195, 88–95 (2012).

9. Broman, K. W. Genetic map construction with Rqtl. 1–41 (2010).

10. Bromman, K. A Guide to QTL Mapping with rqtl. (2009).

11. Broman, K. W. A shorter tour of R / qtl. Rqtl 1–8 (2012).

12. Buckler IV, E. S., Goodman, M. M., Holtsford, T. P., Doebley, J. F. & Sanchez G, J. Phylogeography of the wild subspecies of *Zea mays*. *Maydica* 51, 123–134 (2006).

13. Buckler, E. S. & Stevens, N. M. Maize Origins , Domestication , and Selection. *Darwin's Harvest* 67–90 (2005).

14. 2. Buckler, E. S. et al. The Genetic Architecture of Maize Flowering Time. *Science* (80-.). 325, 714–718 (2009).

15. Chalker-Scott, L. Environmental significance of anthocyanins in plant stress responses. *tPhotochem. Photobiol.* 70, 1–9 (1999).

16. Chen, X. *et al. SQUAMOSA* promoter-binding protein-like transcription factors: Star players for plant growth and development. *J. Integr. Plant Biol.* 52, 946–951 (2010).

17. Cheruiyot, D. J., Boyd, R. S. & Moar, W. J. Exploring Lower Limits of Plant Elemental Defense by Cobalt, Copper, Nickel, and Zinc. *J. Chem. Ecol.* 39, 666–674 (2013).

18. Doebley, J., Stec, A., Wendel, J. & Edwards, M. Genetic and morphological analysis of a maize-teosinte F2 population: implications for the origin of maize. *Proc. Natl. Acad. Sci. U. S. A.* 87, 9888–9892 (1990).

19. Doebley, J. The genetics of maize evolution. *Annu. Rev. Genet.* 38, 37–59 (2004).

20. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28, 2239–2252 (2011).

21. Eagles, H. A. & Lothrop, J. E. Highland maize from central Mexico - Its origin, characteristics, and use in breeding programs. *Crop Sci.* 34, 11–19 (1994).

22. Ellstrand, N. C., Garner, L. C., Hegde, S., Guadagnuolo, R. & Blancas, L. Spontaneous hybridization between maize and teosinte. *J. Hered.* 98, 183–187 (2007).

23. Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6, 1–10 (2011).

24. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. *Annu. Rev. Biochem.* 79, 351–379 (2010).

25. Foster, T. M., Timmermans & P., M. C. in Handbook of Maize: Its Biology pp 161-178 (2009). doi:10.1007/978-0-387-79418-1

26. Freeman, J. L. Spatial Imaging, Speciation, and Quantification of Selenium in the Hyperaccumulator Plants *Astragalus bisulcatus* and *Stanleya pinnata*. *Plant Physiol.* 142, 124–134 (2006).

27. Galway, M. E. et al. The *TTG* Gene Is Required to Specify Epidermal Cell Fate and Cell Patterning in the *Arabidopsis Root. Dev. Biol.* 166, 740–754 (1994).

28. Gutschick, V. P. Biotic and abiotic consequences of differences in leaf structure. *New Phytol.* 143, 3–18 (1999).

29. Harrison, R. G. & Larson, E. L. Hybridization, introgression, and the nature of species boundaries. *J. Hered.* 105, 795–809 (2014).

30. Hauser, M.-T. Molecular basis of natural variation and environmental control of trichome patterning. *Front. Plant Sci.* 5, 320 (2014).

31. Hernandez, M. L., Passas, H. J. & Smith, L. G. Clonal analysis of epidermal patterning during maize leaf development. *Dev. Biol.* 216, 646–658 (1999).

32. Holá, M., Vágnerová, R. & Angelis, K. J. Mutagenesis during plant responses to UVB radiation. *Plant Physiol. Biochem.* 93, 29–33 (2015).

33. Hollósy, F. Effects of ultraviolet radiation on plant cells. *Micron* 33, 179–197 (2002).

34. Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics*.

35. Körner, C. The use of 'altitude' in ecological research. *Trends Ecol. Evol.* 22, 569–574 (2007).

36. Lai, A., Cianciolo, V., Chiavarini, S. & Sonnino, A. Effects of glandular trichomes on the development of *Phytophthora infestans* infection in potato (*S. tuberosum*). *Euphytica* 114, 165–174 (2000).

37. Larkin, J. C., Young, N., Prigge, M. & Marks, M. D. The control of trichome spacing and number in *Arabidopsis*. *Development* 122, 997–1005 (1996).

38. Larkin, J., Oppenheimer, D., Lloyd, A., Paparozzi, E. & Marks, M. Roles of the *GLABROUS1* and *TRANSPARENT TESTA GLABRA* Genes in *Arabidopsis* Trichome Development. *Plant Cell* 6, 1065–1076 (1994).

39. Lauter, N., Gustus, C., Westerbergh, A. & Doebley, J. The inheritance and evolution of leaf pigmentation and pubescence in teosinte. *Genetics* 167, 1949–1959 (2004).

40. Li, D. *et al.* The genetic architecture of leaf number and its genetic relationship to flowering time in maize. *New Phytol.* 210, 256–268 (2016).

41. Li, Z. Q., Zhang, H. M., Wu, X. P., Sun, Y. & Liu, X. H. Quantitative trait locus analysis for ear height in maize based on a recombinant inbred line population. *Genet. Mol. Res.* 13, 450–456 (2014).

42. Li, S., Wang, J. & Zhang, L. Inclusive composite interval mapping of QTL by environment interactions in biparental populations. *PLoS One* (2015).

43. Liakopoulos, G., Stavrianakou, S. & Karabourniotis, G. Trichome layers versus dehaired lamina of Olea europaea leaves: Differences in flavonoid distribution, UV-absorbing capacity, and wax yield. *Environ. Exp. Bot.* 55, 294–304 (2006).

44. Ludwig, S. R. & Wessler, S. R. Maize R gene family: Tissue-specific helix-loop-helix proteins. *Cell* 62, 849–851 (1990).

45. Maes, L., Inzé, D. & Goossens, A. Functional specialization of the *TRANSPARENT TESTA GLABRA1* network allows differential hormonal control of laminal and marginal trichome initiation in *Arabidopsis* rosette leaves. *Plant Physiol.* 148, 1453–64 (2008).

46. McWilliam, J. R. & Naylor, A. W. Temperature and Plant Adaptation. I. Interaction of Temperature and Light in the Synthesis of Chlorophyll in Corn. *Plant Physiol.* 42, 1711–1715 (1967).

47. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* 32, 244–257 (2015).

48. Matsuoka, Y. *et al.* A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci.* U. S. A. 99, 6080–6084 (2002).

49. Mir, C. *et al.* Out of America: Tracing the genetic footprints of the global diffusion of maize. *Theor. Appl. Genet.* 126, (2013).

50. Moose, S. P., Lauter, N. & Carlson, S. R. The Maize macrohairless1 Locus Specifically Promotes Leaf Blade Macrohair Initiation and Responds to Factors Regulating Leaf Identity. *Genetics* 166, 1451–1461 (2004).

51. Morales, F., Abadía, A., Abadía, J., Montserrat, G. & Gil-Pelegrín, E. Trichomes and photosynthetic pigment composition changes: Responses of Quercus ilex subsp. ballota (Desf.) Samp. and Quercus coccifera L. to Mediterranean stress conditions. *Trees - Struct. Funct.* 16, 504–510 (2002).

52. Nachman, M. W. & Payseur, B. A. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 409–421 (2012).

53. Oppenheimer, D. G., Herman, P. L., Sivakumaran, S., Esch, J. & Marks, M. D. A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. *Cell* 67, 483–493 (1991).

54. Pattanaik, S., Patra, B., Singh, S. K. & Yuan, L. An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*. *Front. Plant Sci.* 5, 259 (2014).

55. Peiffer, M., Tooker, J. F., Luthe, D. S. & Felton, G. W. Plants on early alert: Glandular trichomes as sensors for insect herbivores. *New Phytol.* 184, 644–656 (2009).

56. Perazza, D., Vachon, G. & Herzog, M. Gibberellins promote trichome formation by Up-regulating *GLABROUS1* in arabidopsis. *Plant Physiol* 117, 375–383 (1998).

57. Piperno, D. R. *et al.* Late Pleistocene and Holocene environmental history of the Iguala Valley, Central Balsas Watershed of Mexico. *Proc. Natl. Acad. Sci.* U. S. A. 104, 11874–11881 (2007).

58. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl. Acad. Sci.* U. S. A. 106, 5019–24 (2009).

59. Piperno, D. R. A few kernels short of a cob: on the Staller and Thompson late entry scenario for the introduction of maize into northern South America. *J. Archaeol. Sci.* 30, 831–836 (2003).

60. Pires, N. & Dolan, L. Origin and Diversification of Basic-Helix-Loop-Helix Proteins in Plants Research article. 27, 862–874 (2010).

61. Pravenec, M. *et al.* Mapping of quantitative trait loci for blood pressure and cardiac mass in the rat by genome scanning of recombinant inbred strains. *J. Clin. Invest.* 96, 1973–1978 (1995).

62. Prozherina, N., Freiwald, V., Rousi, M. & Oksanen, E. Interactive effect of springtime frost and elevated ozone on early growth, foliar injuries and leaf structure of birch *(Betula pendula). New Phytol.* 159, 623–636 (2003).

63. Rahn, H. Altitude adaptation: organisms without lungs. *In Hypoxia, exercise and altitude: Proceedings of the Third Banff International Hypoxia Symposium*, pp. 345–363, Liss (1983)

64. Runyon, J. B., Mescher, M. C. & De Moraes, C. M. Plant defenses against parasitic plants show similarities to those induced by herbivores and pathogens. *Plant Signal. Behav.* 5, 929–931 (2010).

65. Schilmiller, A. L., Last, R. L. & Pichersky, E. Harnessing plant trichome biochemistry for the production of useful compounds. *Plant J.* 54, 702–711 (2008).

66. Schneeberger, R., Tsiantis, M., Freeling, M. & Langdale, J. A. The rough sheath2 gene negatively regulated homeobox gene expression during maize leaf development. *Development* 125, 2857–2865 (1998).

67. Selinger, D. A., Lisch, D. & Chandler, V. L. The maize regulatory gene *B-Peru* contains a DNA rearrangement that specifies tissue-specific expression through both positive and negative promoter elements. *Genetics* 149, 1125–1138 (1998).

68. Stange, M., Utz, H. F., Schrag, T. A., Melchinger, A. E., & Würschum, T. (2013). High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. *Theoretical and Applied Genetics*, 1–12.

69. Stapleton, A. Ultraviolet Radiation and Plants: Burning Questions. *Plant Cell* 4, 1353–1358 (1992).

70. Sylvester, A. W. & Smith, L. G. in Handbook of Maize: Its Biology (eds. Bennetzen, J. L. & Hake, S. C.) 179–203 (Springer New York, 2009). doi:10.1007/978-0-387-79418-1_10

71. Sylvester, A. W. & Smith, L. G. in (2009). doi:10.1007/978-0-387-79418-1

72. Szymanski, D. B., Lloyd, a M. & Marks, M. D. Progress in the molecular genetic analysis of trichome initation and morphogenesis in *Arabidopsis*. Tibs 5, 214–219 (2000).

73. Tattini, M., Gravano, E., Pinelli, P., Mulinacci, N. & Romani, A. Flavonoids accumulate in leaves and glandular trichomes of *Phillyrea latifolia* exposed to excess solar radiation. *New Phytol.* 148, 69–77 (2000).

74. van Heerwaarden, J. *et al.* Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci.* U. S. A. 108, 1088–1092 (2011).

75. Vigouroux, Y. *et al.* Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am. J. Bot.* 95, 1240–1253 (2008).

76. Vision, T. J., Brown, D. G., Shmoys, D. B., Durrett, R. T., & Tanksley, S. D. (2000). Selective mapping: A strategy for optimizing the construction of high-density linkage maps. *Genetics*.

77. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting Natural Selection in Genomic Data. *Annu. Rev. Genet* 47, 97–120 (2013).

78. Wagner, G. J. Secreting glandular trichomes: more than just hairs. *Plant Physiol.* 96, 675–679 (1991).

79. Wagner, G. J., Wang, E. & Shepherd, R. W. New approaches for studying and exploiting an old protuberance, the plant trichome. *Ann. Bot.* 93, 3–11 (2004).

80. Walker, A. R. *et al.* The TRANSPARENT TESTA GLABRA1 locus, which regulates trichome differentiation and anthocyanin biosynthesis in Arabidopsis, encodes a WD40 repeat protein. *Plant Cell* 11, 1337–1349 (1999).

81. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63 (2009).

82. Wilkes, G. H. Maize and Its Wild Relatives. *Science* (80-.). 177, 1071–1077 (1972).

83. Winter, P. & Kahl, G. Molecular marker technologies for plant improvement. *World J. Microbiol. Biotechnol.* 11, 438–448 (1995).

84. Yan, A., Pan, J., An, L., Gan, Y. & Feng, H. The responses of trichome mutants to enhanced ultraviolet-B radiation in *Arabidopsis thaliana. J. Photochem. Photobiol. B Biol.* 113, 29–35 (2012).

85. Yang, C., Zhang, L., Jia, A. & Rong, T. Identification of QTL for maize grain yield and kernel-related traits. J. *Genet.* 95, 239–247 (2016).

86. Young, N. D. Qtl Mapping and Quantitative Disease Resistance in Plants. *Annu. Rev. Phytopathol.* 34, 479–501 (1996).

87. Yu, N. *et al.* Temporal control of trichome distribution by microRNA156-targeted SPL genes in *Arabidopsis thaliana*. *Plant Cell* 22, 2322–2335 (2010).

88. Zeng Z-B (1994). Precision mapping of quantitative trait loci. *Genetics* 136: 1457–1468.

89. Zhang, Q. *et al.* Mapping Quantitative Trait Loci for Milk Production and Health of Dairy Cattle in a Large Outbred Pedigree. *Genetics* 149, 1959–1973 (1998).