CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

The role of introgression from teosinte (*Zea mays* ssp. *mexicana*) in the adaptation of maize to the highlands of Mexico

A thesis submitted by:
María Rocío Aguilar Rangel

in fulfillment of the requirements
for the degree of Doctor of Philosophy

in

Plant Biotechnology

Thesis directors:
June K. Simpson, Ph.D. & Ruairidh J. Hay Sawers, Ph.D.

October 16, 2018

This work was performed from September 1st 2013 to November 30th 2017 at the Molecular Genetics laboratory of the Genetic Engineering Department and the Maize Genetics and Genomics laboratory of the Advanced Genomics Unit at Centro de Investigaciones y de Estudios Avanzados del Instituto Politécnico Nacional, Irapuato Unit and co-directed by June K. Simpson Williamson, PhD and Ruairidh J. Hay Sawers, PhD.

Committee members

- Luis Herrera Estrella, PhD
- Charles Stewart Gillmor, PhD
- Octavio Martínez de la Vega, PhD
- Víctor Olalde Portugal, PhD

Signed:

Exam date:

Abstract

Maize was domesticated ∼9,000 years ago in a mid-elevation region of southwest Mexico, from the lowland teosinte *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*). Following domestication, maize dispersed rapidly to a wide range of novel environments, including the highlands in central Mexico, where it encountered dramatically different environmental conditions and came into sympatry with the highland teosinte *Zea mays* ssp. *mexicana* (hereafter *mexicana*). Modern highland maize and *mexicana* share multiple phenotypic traits, prompting the hypothesis that maize adaptation to the highland niche was facilitated by hybridization with *mexicana* and subsequent acquisition of adaptive genetic variation. The use of high density molecular markers has revealed the extent of gene flow from *mexicana* to Mexican highland maize to be from 2 to 10%, and opened the door to the mapping of regions of introgression within the genome. To assess the importance of regions of *mexicana* introgression to fitness and agronomic traits in highland maize, QTL mapping was performed in a recombinant inbred line (RIL) population derived from the cross of the highland landrace Palomero Toluqueño, an ancient highland landrace with one of the highest levels of *mexicana* introgression, and B73, an inbred line adapted to low elevations in the US mid-west. The RIL population was evaluated in lowland (Valle de Banderas, Nayarit) and highland (Metepec, Estado de Mexico) field sites. Twenty one QTLs were identified for adaptive traits and flowering architecture and co-localized with *mexicana* introgression, however there was an overlap only on chromosomes 4 and 9 for ear diameter and tassel length. Development of heterologous inbred families (HIF) for the introgression region *inv4m* on chromosome 4 identified an effect of this region on male and female floral architecture closely related with the genetic background indicating epistatic interactions with other loci in the genome.

Resumen

El maíz fue domesticado hace alrededor de 9000 años, en una estrecha región geográfica a elevaciones medias en el suroeste de México a partir de teosintle endémico de esa región *Zea mays* ssp. *parviglumis* (en adelante *parviglumis*), y se extendió rápidamente a un amplio rango de nuevos ambientes, incluyendo los valles altos del Centro de México, donde se encontró con condiciones ambientales muy diferentes (climas más fríos y secos) y entró en simpatría con el teosinte de valles altos *Zea mays* ssp. *mexicana* (en adelante *mexicana*), adaptado a ese ambiente miles de años antes de la domesticación del maíz. El maíz moderno de valles altos y *mexicana* comparten múltiples características fenotípicas, por lo que se ha hipotetizado que la adaptación de maíz a ese nicho fue facilitada por la hibridación y subsecuente adquisición de genes de *mexicana*. El uso de una alta densidad de marcadores moleculares ha revelado flujo genético de *mexicana* al maíz de valles altos del 2 al 10%, y ha permitido el mapeo de regiones de introgresión dentro del genoma. Para probar si las regiones genómicas introgresadas de *mexicana* están ligadas a la adecuación y características de interés agronómico en el maíz de valles altos, se realizó el mapeo de QTLs en una población de líneas recombinates (RIL) derivadas de la cruza del maíz de valles altos Palomero Toluqueño, una raza ancestral con uno de los más altos niveles de introgresión de *mexicana*, y B73, la línea mejorada adaptada a elevaciones del medio oeste de Estados Unidos. Las RILs se evaluaron en dos sitios de campo a elevaciones bajas en Valle de Banderas, Nayarit y en elevaciones altas en Metepec, Estado de México. Se identificaron 21 QTLs para características adaptativas y arquitectura floral y fueron co-localizados con las regiones de introgresión de *mexicana* superponiendose solo en dos cromosomas, 4 y 9 para diámetro de la mazorca y longitud de la espiga. El desarrollo de líneas endogamicas heterólogas (HIF) para la región de cromosoma 4 denominada *inv4m* identificó efectos asociados a la arquitectura floral masculina y femenina cuya expresión cambia deacuerdo al fondo genético, indicativo de interacciones epistáticas de la *inv4m* con otros loci.

Acknowledgements

To Consejo Nacional de Ciencia y Tecnologia (CONACyT) for the financial aid to develop this project under CVU number 420966.

To Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato for providing the resources and installations necessary to carry out this investigation.

To my advisors June Simpson and Ruairidh Sawers for giving me the opportunity to join to their labs and for their support, dedication and patience throughout my studies and research.

To the members of my committee Luis Herrera, Stewart Gillmor, Octavio Martínez, and Víctor Olalde for their support and remarkable guidance throughout this process.

To my lab partners from both groups

To my folks from CINVESTAV

Especially to my family

Contents

List of Figures

List of Tables

List of Abbreviations

ASE	**A**llele **S**pecific **E**xpression
BC1S3	**B**ack**C**ross **1** **S**elf **3**
BC1S5	**B**ack**C**ross **1** **S**elf **5**
CAPS	**C**leaved **A**mplified **P**olymorphisms
DM	**E**ar **D**iameter
DNA	**D**eoxyribo**N**ucleic **A**cid
DO	**C**ob **D**iameter
DTA	**D**ays **T**o **A**nthesis
DTS	**D**ays **T**o **S**ilking
EH	**E**ar **H**eight
FKW	**F**ifty **K**ernell **W**eight
GxE	**G**enotype by **E**nvironment interaction
GBS	**G**enotyping **B**y **S**equencing
GWAS	**G**enome **W**ide **A**ssociation **S**tudies
HIF	**H**eterologous **I**nbred **F**amily
LM	**E**ar **L**ength
NIL	**N**ear **I**sogenic **L**ine
PCR	**P**olimerase **C**hain **R**eaction
PH	**P**lant **H**eight
PM	**E**ar **W**eight
PSR	**P**rior **S**tress **R**esponse
PT	**P**alomero **T**oluqueño
PTG	**T**otal **G**rain **W**eight
QEI	**Q**TL by **E**nvironment **I**nteraction
QTL	**Q**uantitative **T**rait **L**oci
RIL	**R**ecombinant **I**nbred **L**ine
RNA	**R**ibo**N**ucleic **A**cid
SNP	**S**ingle **N**ucleotide **P**olymorphism
STD	**ST**an**D** count
SSR	**S**ingle **S**equence **R**epeat
TBN	**T**assel **B**ranch **N**umber
TL	**T**assel **L**ength

Dedicated to my parents...

Chapter 1

Introduction

1.1 Genus Zea: teosinte and maize domestication

The species *Zea mays* belongs to the genus *Zea* and comprises four subspecies: (i) ssp. *mays*, cultivated maize; (ii) ssp. *parviglumis* (hereafter *parviglumis*), the direct progenitor of maize; (iii) ssp. *mexicana* (hereafter *mexicana*); and (iv) ssp. *Huehuetenangensis* [23]. The lineages giving rise to *parviglumis* and *mexicana* are thought to have diverged ca. 60,000 years ago and the two subspecies currently occupy distinct ecological niches. *Mexicana* is adapted to the drier and cooler elevations of northern and central Mexico (1600–2700 m), while *parviglumis* is adapted to the warmer, mesic middle elevations of southwestern Mexico (< 1800 m). Perhaps as a result of adaptation to distinct niches, the wild subspecies exhibit morphological differences: *mexicana* produces larger spikelets and seeds and fewer tassel branches compared to *parviglumis*; *mexicana* has red, hairy leaf sheaths in contrast to the green and glabrous leaf sheaths of *parviglumis*, traits thought to be important for adaptation to the cool temperatures and changing ultraviolet light conditions of the Mexican highlands (Figure **1.1**) [22, 52].

FIGURE 1.1: **Morphological differences between teosintes *parviglumis* and *mexicana*.** A, B, C and D represent the plant architecture, size of seeds and tassel, and leaf sheath respectively of teosinte *parviglumis*; E, F, G and H represent the plant architecture, size of seeds and tassel, and leaf sheath respectively of teosinte *mexicana*.

Cultivated maize was derived from ssp. *parviglumis* via a single domestication event [42]. After this initial event, teosinte populations have persisted in both allopatry and sympatry with maize [45], hence it is believed that introgression from other teosinte types may have contributed to the maize gene pool and thereby explain the remarkable phenotypic and genetic diversity in maize.

1.2 Domesticated corn has been spread across environments

Genetically, maize (*Zea mays* ssp. *mays* L.) is an extremely variable crop, adapted to a wide range of habitats through gradual evolution of hundreds of landraces or

farmer's varieties [45]. During its diffusion, maize adapted to diverse habitats resulting in a current distribution much wider than its closest relatives with respect to altitude (from 0 to 3,400 m) and latitude (from the central valley of Chile (40° South) to Canada (52° North)). Within Mexico alone maize landraces encounter extremely diverse environments in terms of mean annual temperature (from 12 °C to 29.1 °C) and precipitation (from 426 - 4245 mm), adapted to environments from 0 to 2900 masl [23, 66].

Maize landraces have proven to be adapted to specific, often marginal environments and several unique maize landraces are prevalent in different regions of the world [25, 48, 69], having distinct germplasm groupings which show specific adaptation to four broad classes of environments: lowland tropics (< 1200 masl), mid-elevation tropics (1200-1800 masl), the highlands (1800-2800 masl) and temperate environments (maximum photoperiod during growth > 14 h).

Highland maize from central Mexico is a germplasm resource of ancient origin that has desirable features important in adaptation including better seedling emergence from cool soil, better photosynthetically based growth at low temperatures and better frost tolerance than most other types of maize, however very little is known about the molecular and genetic mechanisms that underlie this variation.

1.3 Ssp. *mexicana*, potential source of beneficial variation during adaptation of cultivated maize to the highland niche

It is thought that gene flow from *mexicana* to maize has contributed to maize diversity [42]. Several genomic regions in maize landraces show signs of introgression from *mexicana*, suggesting an important role for gene flow during post-domestication maize dispersion [23]. The genomes of Mexican maize from elevations at which ssp. *mexicana* grows are composed of about 0.2 to 20% of *mexicana* alleles whereas Mexican maize from lower elevations contains only from 0.2 to 2% [42, 24].

A study from van Heerwarden [21] using several hundred markers revealed that admixture with *mexicana* may approach 20% in highland Mexican maize and these results were verified by Hufford [24] when evaluating sympatric and allopatric maize-*mexicana* populations based on genome-wide Single Nucleotide Polymorphisms (SNPs) data, demonstrating that highland maize genotypes may reach a content of 19% of from the *mexicana* genome versus 12% of the maize genome found in *mexicana* germplasm, corroborating asymmetric gene flow and suggesting that maize colonization of highland environments may have been facilitated by adaptive introgression from *mexicana* populations.

Five races are cultivated in the central Mexican highlands known as the "Mexican Pyramidal complex" or "Conico group": Chalqueño, Cónico, Palomero Toluqueño, Arrocillo Amarillo and Cacahuacintle [25, 48]. Palomero Toluqueño (PT) exhibits the highest level observed of *mexicana* genotype introgression (12%) [42]. PT is a popcorn cultivated in the highland valleys of central Mexico (2140-2887 masl) and has been classified as one of the most ancient races, probably representing an early adaptation during the dispersion of maize from its center of origin [22]. The PT genome is about 22% (140 Mb) smaller than B73, and shows a large number of hitherto unreported sequences [50], implying a large pool of unexplored alleles. Adaptive traits such as strong pubescence and accumulation of anthocyanins in the stem as well as

sparsely branched tassels and conical shaped ears are shown in PT [74, 53, 57]. Additionally, PT shows good *"per se"* performance when cultivated in high elevation-environments and was found to have resistance to the maize weevil *Sitophilus zeamais* [50].

1.4 Local adaptation through hybridization and introgression

Currently, there is much interest in understanding the processes contributing to the evolutionary transition from one environment to another. It has been reported that hybridization enhances the probability of adaptation to a challenging new environment [24] and that local adaptation mechanisms can also come into play when two species hybridize [29]. Hybridization between populations having different genetic systems of adaptation may lead to fertile individuals segregating blocks of genetic material belonging to entirely different adaptive systems [3].

Adaptation to local environments has been observed in many organisms. Exotic gene flow and natural introgression can increase the number of adaptive habitats occupied by the cultivated form(s) and this has been demonstrated in crops such as tomato, maize, and soybean [70]. In *Arabidopsis thaliana*, alleles associated with higher fitness have a tendency to be locally predominant and related to environmental conditions [18]. In *Medicago truncatula*, a study identified a candidate locus for local adaptation, and found a relationship to the growth rate under temperature and the soil moisture treatments [80].

In the maize context, it is suggested that *mexicana* introgression has been important in the development of highland maize races, based on the quantification of introgression throughout the genome, the Hufford study [24] found 9 regions with a high frequency of introgression (chromosomal regions with at least 50% of the opposite taxon in a given population). They compared these regions with putatively adaptive traits (presence of anthocyanins and macrohairs) under highland conditions identified by Lauter [32] from a cross between *parviglumis* and *mexicana* and six of them overlapped with QTLs detected for these traits. One of these introgression regions, on chromosome 4 overlaps with a QTL for both pigment intensity and macrohairs, and maps to the same position as a recently identified putative inversion polymorphism (***Inv4m***; [52]) showing significant differentiation between *parviglumis* and *mexicana*.

FIGURE 1.2: **Morphological differences between highland and lowland maize and teosintes for pigments and hairs** A, teosinte *mexicana*; B, teosinte *parviglumis* (modified from [32]); C, Palomero Toluqueño; D, lowland maize

1.5 Chromosomal inversions and local adaptation

Structural variations are widespread in maize genomes and are important for local adaptation and genetic improvement [78]. Inversions are rearrangements in the genome that change the order of genes along a chromosome and often appear linked to loci involved in climatic or habitat adaptation. A major effect of chromosomal inversions is to suppress recombination between alternative chromosomal arrangements in hybridizing populations [30, 29]. Inversions are predicted to be beneficial when they capture locally adapted alleles or when the breakpoints of the inversions themselves are beneficial [30]. Inversion breakpoints can directly disrupt coding sequences or alter gene expression of adjacent genes by separating regulatory elements from the corresponding coding sequences, providing new regulatory sequences, or moving the genes to different regulatory domains [51].

Chromosomal inversions have been involved in local adaptation of both animals and plants [4]. Numerous studies have found associations between putatively adaptive phenotypes and inversion polymorphisms for size or developmental time in Drosophila, adaptation to freshwater in sticklebacks and flowering time in plants [73, 51, 39]. This last example was found by Lowry and Willis [39] where an inversion polymorphism in the yellow monkeyflower *Mimulus guttatus* that occurs in two distinct ecotypes contributes to adaptation and reproductive isolation mainly due to modifications in flowering time.

Cytological studies of both *mexicana* and *parviglumis* have identified a number of inversion polymorphisms. Together they comprise approximately 5% of the maize reference genome and their role in patterning both molecular and phenotypic variation among populations is significant [52]. Hufford [23] found several newly identified inversion polymorphisms that were related to differentiation between populations and subspecies and associated with adaptation to environmental variables such as temperature and altitude, however a definitive demonstration of the influence of an inversion in adaptation and ecological reproductive isolation requires that the inversion polymorphism must be shown to be partitioned between reproductively isolated groups, links must be made between the inversion and given phenotypes and additionally field experiments must be conducted to show that the inversion contributes to adaptation, since focus on discrete environments may help to identify functionally important genes but, however this strategy may miss selectively important genes mediating adaptation to other environmental conditions and may not provide insights into how selection shapes genetic diversity across environmental clines [39, 80].

1.6 Regulatory differences involved in local adaptation

Cis-regulatory elements constitute an important part of the genetic basis for adaptation given that *cis*-regulatory regions underlie a variety of phenotypic differences with functional significance in morphology, physiology and behavior [75].

Stress plays an important role in facilitating local adaptation by directing and organizing the adaptive adjustment of an organism to an ever changing environment. Investigation of the transcriptome (link between genotype and phenotype) can help us to identify genes that have undergone regulatory evolution to confer greater stress tolerance.

1.7 Identification of genomic regions involved in local adaptation

To test different genomic regions for allelic differences in a locus of interest and confirm the effect of predicted differences, a single small segment may be transferred from a donor parent into the genetic background of a recipient parent to generate introgression lines. The composite of these lines is called mapping population, or because of their prevailing mode of construction, Recombinant Inbred Lines (RILs) or Near-Isogenic Lines (NILs) [79, 28].

The success of this approach has been facilitated by the development and use of molecular markers that enable the selection of individuals carrying a genomic donor segment at a precise location. Since the genomic composition of the starting material is known, only a few markers targeted at the donor introgressions are sufficient to successfully monitor subsequent generations.

Simple Sequence Repeats (SSRs) and Cleaved Amplified Polymorphic Sequence (CAPS) have been molecular markers widely used for genetic mapping [60]. SSRs are repetitions of a very short nucleotide motif (usually one to five base pairs) that vary in the number of repeat units among alleles [7], furthermore, they are co-dominant, allowing identification of heterozygotes. CAPS markers [2] have been useful to report polymorphisms (SNPs: single nucleotide polimorphisms or INDELs: insertions or deletions) in locus-specific PCR amplicons that result in differences in restriction fragment lengths.

1.8 Development of mapping populations

The starting point in mapping population construction is a cross between two genetically distinct parental lines which results in heterozygous offspring, used to generate either segregating (e.g., F2: F1 individuals derived by crossing the two parents) or fixed (e.g., RIL: single-seed selections from individual plants from an F2) populations.

Recombinant inbred lines (RILs) can be used to map and study quantitative trait loci (QTL) underlying phenotypic traits. Meiotic crossover events create a mosaic of parental genomes in each RIL. [49]

In the development of NILs by backcrossing the heterozygote to the recipient parent, the proportion of the donor parental genome is reduced by 50% in each backcross; in recurrent backcrosses, heterozygosity is further reduced. Once a single introgression at the desired position is obtained, this can be fixed by selfing, (losing 50% of heterozygosity in each round) or sibling mating (homozygosity increased in terms of selected traits), after which the homozygous NILs can be phenotyped.

NILs differing for a single QTL have been used to study the phenotype associated with a specific locus, however, their development has been limited by the cost, time and effort associated with developing the appropriate genetic materials. In addition, the presence of a single introgressed segment does not allow for expression of QTL effects that require specific genetic interactions (epistasis). Thus, the phenotype of the introgression can be masked by lack of penetrance, or confounded by novel epistatic interactions arising in the recipient background [67].

An alternative procedure utilizes multiple introgressions that allow detection of genetic interactions between loci. With this approach, NILs that are not entirely homozygous are selected from the introgression line (generated between the two parents of interest). Progeny of this line will segregate for those loci not yet fixed and

will represent a Heterogeneous Inbred Family (HIF; [67]) of nearly-isogenic individuals. HIFs are inbred to create almost completely homozygous genotypes except for a few small regions (<5% of the genome) [79].

1.9 Are B73 and PT regulatory differences potentially involved in local adaptation?

F1 hybrid transcription analysis provides tools to investigate genetic causes underlying shared and different phenotypes in different populations and help to identify targets of adaptive evolution through identification of stress response components

1.10 B73xPT mapping population

To identify chromosomal regions that contribute to highland adaptation, a BC5S1 cross between PT and B73 was generated.

1.11 What is the role of introgressed regions from *mexicana* into highland maize?

A QTL analysis allows us to assess chromosomal regions that contribute to highland adaptation from a mapping population derived from highland and lowland haplotypes. Whether *mexicana* introgression has been the driver will be tested by overlapping these regions with the adaptive QTLs.

QTLs for quantitative traits are influenced by genetic background (GB) and environment. Identification of QTLs with GB independency and environmental stability is prerequisite for effective marker-assisted selection (MAS). [72]

1.12 What is the effect of the *inv4m* in highland adaptation?

A collection of HIFs can be used as a mapping population to identify QTLs or determine the effect of the *inv4m*. Individual lines containing a heterozygous region coinciding with the region of interest but otherwise homozygous can be selected using the genotypic information of the population. Progeny of these families will segregate only for the heterozygous region, creating homozygous lines with different genotypes at the region of interest in a single generation. These near-isogenic lines (NILs) can then be tested to compare the effect of the segregating region. When more than one HIF can be selected, the same locus can be evaluated in different genetic backgrounds and epistatic interactions between the region of interest and other genomic regions can be tested [79].

HIFs provide a range of possibilities as described above and present the opportunity to evaluate dominance effects since the effect of the combination of alleles in heterozygous populations with respect to the homozygous populations developed at the region of interest can be determined. Additionally, these populations can be evaluated over contrasting environments allowing the significance of introgressed regions in adaptation to be determined.

Chapter 2

Objectives

The overall objective of this research was to evaluate the functional importance of previously identified *mexicana* introgression to the adaptation of maize to the highlands of central Mexico.

The specific aims were:

1. To identify *cis*-regulatory variation linked to stress associated genes in the highland landrace Palomero Toluqueño, and assess co-localization with previously mapped regions of *mexicana* introgression.

2. To map genomic regions linked to morphological and agronomic traits characteristic of Mexican highland maize, and assess co-localization with previously mapped regions of *mexicana* introgression.

3. Estimate the relative genotypic value of B73 and Palomero Toluqueño haplotypes of *inv4m*, a large scale inversion polymorphism found at high frequency in Mexican highland maize, and characterized previously to be of *mexicana* origin.

Chapter 3

Experimental strategy

3.1 Genetic material and genotyping

3.1.1 Plant material and isolation of a single PT haplotype

The Palomero Toluqueño (PT) accession Mexi 5 was provided by the International Maize and Wheat Improvement Center (CIMMYT; stock GID244857). The B73 stock used for population generation was provided by Thomas Brutnell, Boyce Thompson Institute/ Donald Danforth Center, USA.

A B73 x PT F1 hybrid stock was generated in an experimental field located in Valle de Banderas, Nayarit. Pollen pooled from multiple PT plants was used to pollinate B73 ears. Resulting F1 plants were grown under standard greenhouse conditions (27°C day, 24°C nights; 15h days, 9h nights; 30% humidity) at Ames, Iowa. Leaf tissue from a single B73 x PT F1 individual was harvested for RNA sequencing (see below, Allele Specific Expression), and the same individual crossed as a male to multiple B73 ears to generate a large bi-allelic BC1 population, used in the generation of further stocks (see below, Development of mapping populations). The PT haplotype carried by the selected F1 individual and captured in the resulting BC1 forms the basis for the characterization in this study.

3.1.2 Development of mapping populations

To study various aspects of functional variation between B73 and PT a number of mapping resources were derived from the original B73xPT BC1 population (Figure 3.1).

Generation of a BC1S5 Recombinant Inbred Line (RIL) population

Approximately 100 BC1S1 families were advanced through five generations of self pollination to generate a set of BC1S5 RILs, each carrying, on average, 25% PT and 75% B73 genome (Figure 3.1).

Generation of Near Isogenic Lines (NILs) segregating for *inv4m*

Starting from the B73xPT BC1 stock, marker-assisted backcrossing to B73 was used to generate a BC5 family carrying the PT allele of *inv4m*. BC5 individuals heterozygous at *inv4m* were self-pollinated to generate segregating stocks (Figure 3.1). The *inv4m* was genotyped using a novel Cleaved Amplified Polymorphic Sequence (CAPS) marker and three Single Sequence Repeat (SSR) markers (Tables *A*.1-*A*.2)

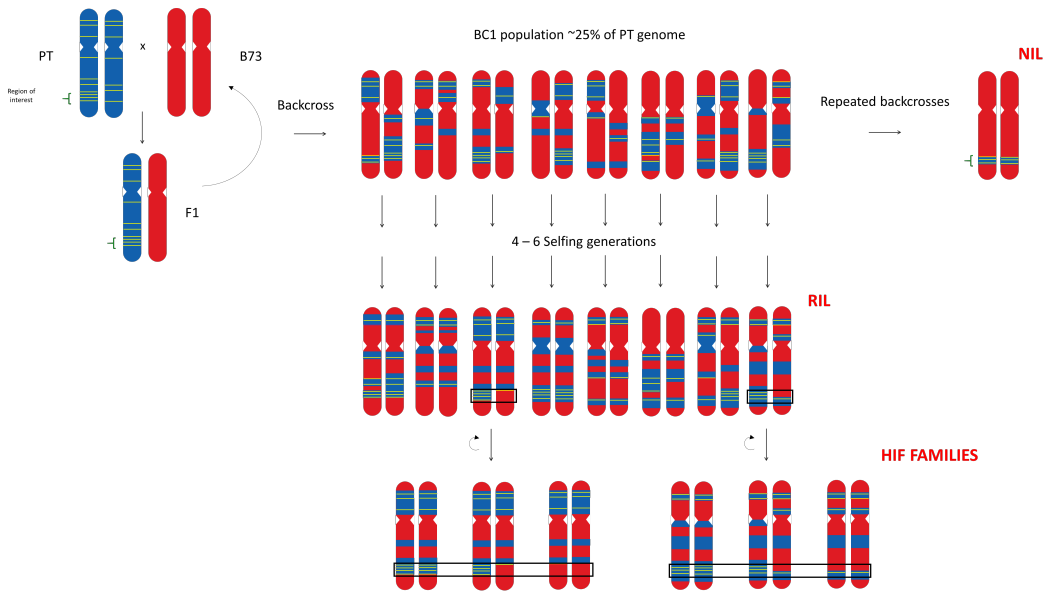

FIGURE 3.1: **Crossing scheme for genetic analysis of B73 and PT.** A BC1 population, generated by crossing a single B73xPT F1 individual as male to multiple B73 ears, self-pollinated for five generations to generate a BC1S5 recombinant inbred line (RIL) population. Approximately 100 independent pedigrees were advanced from the BC1S1 generation. A small number of BC1 individuals were also used to make additional backcrosses to B73 to generate Near Isogenic Lines (NILs) carrying the *inv4m* inversion polymorphism from PT. Heterogeneous Inbred Families (HIFs) segregating (*inv4m*) in advanced generations were selected and self-pollinated to generate pairs of near-isogenic lines homozygous for either the B73 or PT allele at *inv4m*.

Generation of NIL pairs from Heterologous Inbred Families (HIFs)

To evaluate further the effect of *inv4m*, Heterogeneous Inbred Families (HIFs) were identified from partially inbred generations of the B73xPT RILs [67]. HIFs provide materials for immediate characterization, homozygous segregants provide near-isogenic stocks fixed for either B73 or PT alleles. The strategy provides theoretical benefits: near isogenic comparisons of B73 and PT alleles will be made in a genetic background that contains a sizeable component of fixed PT content (25% on average), significant if the genotypic value of a given candidate region is strongly influenced by epistatic effects. With the selection of more than one HIF, the same locus can be evaluated in different genetic background and epistatic interactions between the region of interest and other genomic regions can be tested [67] Families segregating for *inv4m* were genotyped (as described below), and individuals identified as homozygous for B73 or PT alleles within a given family were self-pollinated to generate a pair of near-isogenic lines, differing at the *inv4m* (Figure 3.1).

3.1.3 Genotyping of *inv4m*

For analysis of the *inv4m* genotype, DNA was extracted from individual plants using the DNeasy plant mini kit (QIAGEN) according to the manufacturer ′s instructions. Microsatellite / simple sequence repeat (SSR) loci were amplified by PCR using fluorescently-labeled forward primers and unlabeled reverse primers (Table A.1). The PCR amplicons were separated by size using capillary electrophoresis

(Applied Biosystems 3730 XL DNA Analyzer) and the dye labeled products identified by fluorescence detection. The size of specific alleles or bins was scored using the software package Geneious (version 6.0.5.) [27].

A CAPS marker for the *inv4m* region was developed based on a previously reported diagnostic SNP: PZE104103103 [22] that falls within a site cleaved by the enzyme *Hinf I* (GANTC, where N could be any nucleotide; the diagnostic SNP C>T falls in the last position). Specific oligonucleotide primers (Table A.2) were designed to span the SNP PZE104103103 (Figure 3.2, Table A.3). The uncleaved PCR product amplified from B73 is 502 bp. In the B73 state (C) at PZE104103103, the product is cleaved in three places by *Hinf I*. In the PT state (T) at PZE104103103, the product is cleaved twice (Figure 3.2). Amplified products were digested with *Hinf I* and visualized by gel electrophoresis resolved on an agarose gel to differentiate B73 and PT alleles.

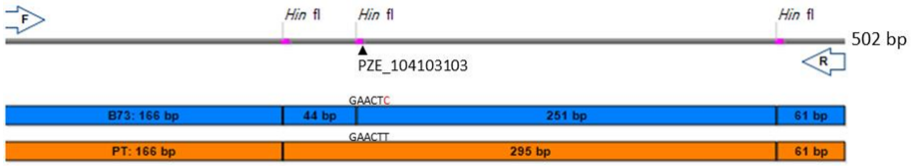

FIGURE 3.2: **The CAPS marker distinguishes B73 from PT alleles at the *inv4m*.** Specific primers were designed to amplify a ∼502 bp fragment from the B73 reference genome at the *inv4m* region spanning the SNP PZE104103103. In silico analysis of amplified fragments from B73 (blue) and PT (orange) detected three and two recognition sites, respectively, for the restriction endonuclease *Hinf I*.

3.1.4 High throughput genotyping

RILs and NILs were genetically analyzed using Genotyping By Sequence (GBS; [17]). 60 samples were sent to Cornell University, Ithaca, NY, and 30 samples were sent to Novogene Corporation sample collection site, Sacramento, CA.

The GBS approach uses restriction enzymes to reduce genome complexity avoiding repetitive regions and targeting lower copy regions with two to three fold higher efficiency prior to sequencing to identify polymorphisms [17, 15, 63].

High molecular weight DNAs were digested using restriction enzymes (REs). For Cornell samples, DNAs were digested using *ApeKI*, a suitable restriction enzyme for maize and teosinte partially sensitive to methylation (will not cut if the 3′ base of the recognition sequence on both strands is 5-methylcytosine), with relatively few recognition sites in the major classes of maize retrotransposons [17]. Novogene samples were digested by *Mse I* and *Hae III* enzymes. Barcode adapters were ligated to the ends of the resulting fragments. Products were PCR amplified and pooled for sequencing. Raw sequence data were filtered and aligned to the B73 reference genome (AGPv3.22) to identify polymorphisms [9] associated with PT alleles.

3.2 Allele specific expression analysis

3.2.1 RNA sequencing

Total RNA was extracted from leaf tissue of 14 day-old-seedlings harvested from the B73xPT F1 individual used to generate the RIL and NIL genetic resources described

above, using the Qiagen RNeasy Plant Mini Kit (cat ID 74904) according to the manufacturer's protocol. RNA purity/integrity was assessed by spectrophotometry and agarose gel electrophoresis.

Libraries for Illumina sequencing were prepared according to the TruSeq RNA Sample Preparation Guide (15008136 A, November 2010) and paired-end sequencing carried out on the Illumina HiSeq 2000 platform. Raw data is available in the NCBI *http://www.ncbi.nlm.nih.gov* Sequence Read Archive (SRA) under accession *SRP*011579.

3.2.2 Quantification of allele specific expression (ASE)

ASE analysis was based on the method proposed by Lemmon [33]. A total of 151,168,196 paired-end reads (101 bp length each) from a B73xPT F1 hybrid, were trimmed using Trimmomatic 0.32 [6] and aligned using BWA mem [35] against a set of 39469 annotated transcripts from the B73 reference genome AGPv3.22, selecting the longest predicted transcript per gene, and removing transcripts with base sequences consisting mainly of Ns. Aligned reads were filtered by mapping quality (view q 30) using samtools (version 1.3.1), bcftools (version 1.3.1) and vcfutils [37, 36, 34] to identify PT polymorphisms. These polymorphisms were incorporated into the B73 reference transcripts to generate PT pseudo-transcripts using vcfutils. A single fasta file containing the two sequences per locus (one from the B73 reference and one from the PT pseudo-reference) was created using Unix commands in the terminal and the B73xPT F1 reads were re-aligned against this F1 pseudo-reference using bowtie2 (version 2.2.9) [31].

ASE was quantified by counting read depth for each parental transcript with eXpress (version 1.5.1) [54] and subsequently a chi-squared test was used to identify any bias between B73 and PT alleles, adjusting for multiple comparisons using the Bonferroni correction. Pipeline details are provided in Appendix B

3.2.3 Gene Ontology annotation and enrichment analysis

ASE candidate genes were assigned to Gene Ontology categories (release 52 available at *ftp://ftp.gramene.org/pub/gramene*). Obsolete annotations were replaced by the corresponding "consider" or "replaced_by" category(ies) in the ontology file (go.obo) available at http://www.geneontology.org/ (dated 2016-09-19). Categories associated with at least 10 genes were considered in further analyses.

Enrichment analyses were performed comparing ASE candidates against the polymorphic gene set, using the Bingo [40] Cytoscape [61] plugin, under a hypergeometric test and controlling for multiple tests using Benjamini and Hochberg false discovery rate at 1%. Categories were called PT up or PT down using a threshold of abs(median log2 FC) ≥ 1.

3.2.4 Detection of Prior Stress Response (PSR) genes in the highland landrace Palomero Toluqueño

Candidate ASE genes were cross-referenced to a published study by Makarevitch in 2015 describing transcriptional responses in maize seedlings exposed to cold, heat, salt and UV stresses [41]. Only the B73 data from this study were used in the comparison with the B73xPT transcriptome. Genes (1) identified as ASE; (2) responding significantly to stress in the Makarevitch study (absolute log2 FC > 1; called as significant in the Makarevitch study; calls "up" or "on" in the published study were

considered here as "up", similarly, "down" or "off" were considered as "down") (3). Those for which the sign of ASE was concordant with the sign of stress response were considered to show Prior Stress Response (PSR) with respect to a given stress.

Published Fst values for population level differentiation between Mesoamerican and South American highland and lowland maize populations [65] were obtained from *https://github.com/rossibarra/hilo_paper/tree/master/fst*. where multiple SNPs were associated with a single gene, the values used correspond to the SNP showing the highest Fst in Mesoamerica.

3.3 Detection of *Cis*-regulatory variation associated with regions of crop-wild introgression in Mexican highland maize

3.3.1 ASE Co-localization with *mexicana* introgressed regions

Reported *mexicana* regions introgressed to highland maize [24] were evaluated for ASE enrichment in the PTxB73 F1. Assessment of levels of gene expression were done at the region of the inversion *inv4m* using the NILs and HIFs to confirm differences due to chromosomal re-arrangements.

3.3.2 Expression analysis of candidate genes

To asses the levels of gene expression for the genes showing ASE (log2 PT/B73 > 2 and $p < 0.05$) at the *inv4m*, the parentals, PT, B73, and derived lines NILs and HIFs were grown under green house conditions. Total RNA was extracted from seedlings 14 days after germination according to the protocols described in Appendix C.

Specific primers to quantify the relative levels of gene expression were designed using the Primer Express sofware 3.0 (Applied Biosystems) with the following parameters: fragment length 75 - 150 pb, primer length 18 - 24 nucleotides and Tm 60 - 65°C (Table A.4). Primers were validated as indicated in Appendix D to confirm the expression pattern by RT-qPCR and statistical trials were performed to validate their expression level differences.

3.4 Evaluation and QTL mapping in lowland and highland environments

3.4.1 Field evaluation

Two HIF families and a total of 100 BC1S5 RILS were evaluated during 2016 in one lowland (PV: Valle de Banderas, Nayarit) and one highland (MT: Metepec, Mexico State) location. The major environmental conditions are described in Table 3.1. Parental lines B73 and PT were grown for two years (2015 and 2016) and randomly inserted in each block of RILs as controls during 2016.

Each evaluated line was grown in a three-row plot in MT and two-row plot in PV, with 15 plants per row, 25 cm between plants within each row and 75 cm between rows. Weeds and insects were controlled by chemical methods as needed, and had no effect on crop performance.

TABLE 3.1: Environmental characteristics of the two experimental sites in Mexico where the BC1S5 lines were phenotyped.

	Lowlands	**Highlands**
Field site	Valle de Banderas	Metepec
State	Nayarit	Mexico
Latitude	20.8	20.2
Longitude	-105.2	-98.3
Elevation (masl)	54	2610
Min\Mean\Max (°C)	22.5\25.8\28.4	12.3\12.4\17.1
Precipitation (mm)	1173	809

3.4.2 Phenotyping

Plants were evaluated for a number of morphological and agronomic traits (Table 3.2). Flowering time was recorded for male and female inflorescence: time from sowing until 50% of plants in the row exhibited anthers was recorded as days to anthesis (DTA) and days to silking the time from sowing until silks emerged for at least 50% of the plants in the row (DTS). Morphological traits were measured after flowering on five plants per plot. Plant height (PH) was taken as the distance from the ground to the internode between stem and tassel main spikelet base. Ear height (EH) was taken as the internode base of the highest ear respectively. The number of primary branches on the tassel (TBN) and the main spike length from the junction to the tip of the tassel (TL) were recorded before harvest on five plants per plot. Ear length (LM), ear diameter (DM) and cob diameter (DO) were measured for five dry ears per row after harvest. Ear weight (PM), Fifty kernel weight (FKW) and total ear weight (PTM) were scored on five ears harvested per plot avoiding data from the border plants.

TABLE 3.2: Morphological and agronomic traits measured.

Major trait	**Component trait**	**Abbreviation**
Adaptation and survivor rate.	Stand count	STD
Phenology	Days to silking	DTS
	Days to anthesis	DTA
Morphology	Plant Height	PH
	Ear Height	EH
	Tassel branch number	TBN
	Tassel lenght	TL
	Ear lenght	LM
	Ear diameter	DM
	Cob dimeter	DO
Yield and yield components.	Ear weight	PM
	Fifty kernell weight	FKW
	Total kernel mass	PTG

Analysis of phenotypic data

Phenotypic data were fitted with a mixed linear model, for analysis of variation among lines within and between sites, including effects of experimental block and spatial coordinates as random effects nested to the field location. Genotype and location (highlands or lowlands) were considered fixed effects. For each line, the least square mean predicted from the PROC MIXED procedure SAS version 9.4 for windows (SAS Institute Inc) was used as phenotypic input for QTL analysis.

3.4.3 Library preparation and sequencing

Genomic DNA from plant material was extracted from 100 individual plants using the DNeasy 96 Blood & Tissue Kit (QIAGEN) according to the manufacurer ′s instructions. DNA concentration for each sample was measured by nanodrop ND-8000. Integrity was analyzed by electrophoresis on 1% agarose gels. Samples were arranged in 96-well plates and sent to the Genomic diversity facility at Cornell University, US for GBS analysis according to [17]. Samples from the following season, failed samples and samples from generations delayed in the crossing scheme were sent to Novogene Corporation, US for GBS services.

3.4.4 Detection of SNPs from raw sequencing data

SNPs were called in 94 lines using the TASSEL 5.2.19 production pipeline [19]. Lines showing an excess of recombination or sequencing errors were eliminated and data from the remaining lines were transformed to ABH format for the following steps.

Imputation and error correction

Imputation was performed using the TASSEL plugin "callGenosToABH" to remove sites with missing, ambiguous or heterozygous genotypes [19]. The output was a .csv file with the data arrangement specific to the Rqtl package.

3.4.5 Linkage map construction

A total of 60 individuals were genotyped from the Cornell analysis [17], and 30 individuals through Novogene GBS services. The two .csv files obtained were sorted independently by physical position according to the B73 reference genome AGPv3.22, and processed for linkage mapping by: (1) Removal of duplicate markers; (2) elimination of markers deviating strongly from the expected proportion 0.75 B73: 0.25 (according to the mapping population scheme BC1S5); (3) elimination of similar markers; and (4) elimination of singletons [68].

After processing independent files were combined and re-sorted by physical position. Remaining problematic markers were removed using the R package qtlTools to construct the linkage map. Pipeline details are provided in **??**.

3.5 QTL mapping

QTL analysis was performed using R/qtl sofware [10]. The analysis was performed on a subset of 81 individuals for which both molecular and phenotypic data were available. Data for lowland and highland sites was analyzed separately and together in a joint analysis. QTLs identified in either the separate or joint analysis were considered further. For each trait, the logarithm of odds (LOD) threshold was estimated at $p < 0.05$ using 1000 permutations of the dataset. QTL effects and associated variance explained were estimated by ANOVA using the *fitqtl* function. The confidence interval for each QTL was defined by a drop in LOD support of 1.5 from the QTL peak.

3.5.1 QTL co-localization with *mexicana* introgressed regions

Regions of *mexicana* introgression taken from physical positions reported previously [24] were compared with the same position in QTLs detected in this study.

3.5.2 Analysis of HIF derived NIL pairs

NIL pairs developed from a BC1S4 HIF heterozygous at the *inv4m* region were evaluated in conjunction with the RIL population. Phenotypic differences between NILs were assessed by *t-test*.

Chapter 4

Results

4.1 Identification of *Cis*-regulatory variation associated with stress-related genes in the highland landrace Palomero Toluqueño

4.1.1 Allele specific expression analysis

To identify *Cis*-regulatory differences between B73 and PT, allele specific expression analysis (ASE) was performed. The B73xPT F1 transcriptome was aligned against the B73 transcriptome. The alignment against the 39,469 B73 reference gene models identified 9,256 genes containing at least one sequence variant that could be used to distinguish the products of B73 and PT alleles. The number of reads per transcript were determined by eXpress.

For 2,386 (26%) of these 9,256 polymorphic transcripts, the number of reads corresponding to the B73 allele differed significantly ($p < 0.05$; Bonferroni correction for multiple tests) from the number of reads corresponding to the PT allele with an absolute $\log_2$ FC > 1, and these genes were considered to exhibit allele specific expression (ASE).

For 1,412 (59%) of the ASE candidate genes, accumulation of the PT transcript was lower than that of the B73 transcript ($\log_2$ PT/B73 < 1; hereafter, "PT-down"), while for the remaining 974 (41%) of the ASE candidates, the PT transcript was accumulated at higher levels ($\log_2$ PT/B73 > 1; hereafter, "PT-up").

4.1.2 Gene Ontology annotation and enrichment analysis

To obtain an overview of the ASE candidates, a Gene Ontology (GO) analysis was performed. The set of 2,386 ASE candidates were not enriched for any specific GO category with respect to the 9,256 polymorphic gene set, but, nonetheless, many individual genes belonged to biological process categories related to stress responses, including responses to heat (GO: 0009408), cold (GO: 0009409) and salt (GO: 0009651) (Figure 4.1).

Overall, 52 biological process categories were represented by at least 10 genes. Of these, 38 (73%) were PT-down (based on the median $\log_2$ PT/B73 of the associated genes), 11 (21%) were PT-up, and the remaining three categories had a median $\log_2$ PT/B73 close to 0. A similar pattern was observed for molecular function categories: 57 categories were associated with at least 10 ASE genes, 42 PT-down, 12 PT-up and three showing no trend.

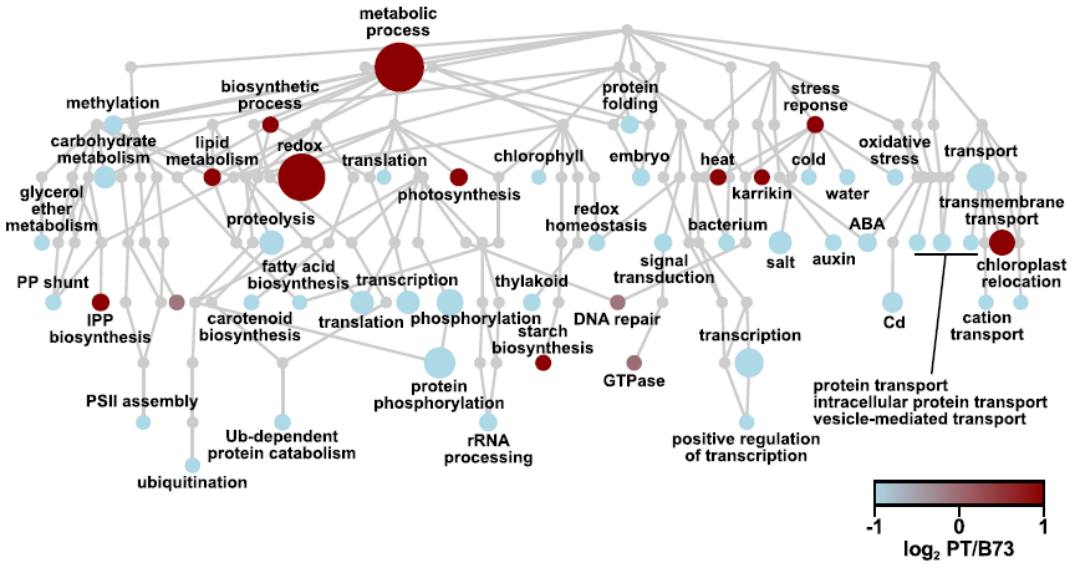

FIGURE 4.1: **ASE candidate genes are assigned to a range of biological process Gene Ontology categories.** Hierarchical tree of Gene Ontology biological process categories represented in ASE loci. Nodes represent categories, with the root GO:0008150 biological process as the uppermost node. Edges represent the parent-child (i.e., "is_a") relationship between categories. Node color indicates the median ASE ($\log_2$ PT/B73) for the genes in the category, with light blue indicating negative values and dark red indicating positive values. Node size is proportional to the number of loci assigned to corresponding category. Some category names were abbreviated for clarity

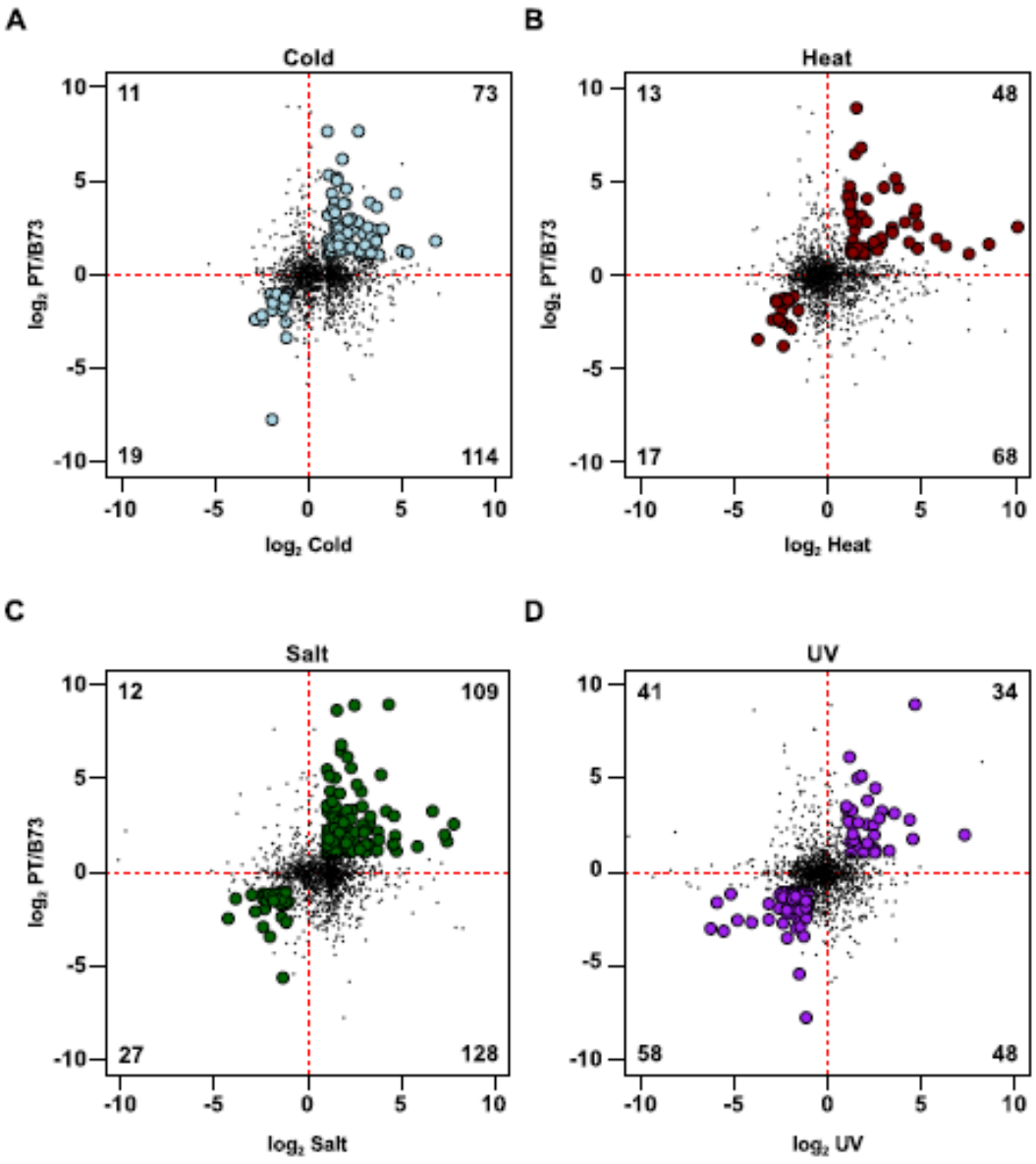

FIGURE 4.2: **ASE identifies prior stress response (PSR) in PT with respect to B73.** ASE ($\log_2$ PT/B73) in control F1 leaves for the 1,407 sequence variant, stress-responsive gene set against B73 stress response ($\log_2$ stress/control) for (A) cold, (B) heat, (C) salt and (D) UV treatments as reported in the Makarevitch dataset. Numbers in each quadrant represent the count of genes called as significant in ASE and stress comparisons. In each plot, the quadrants represent (clockwise from upper left) genes: up ASE/down stress, up ASE/up stress, down ASE/up stress, down ASE/down stress. Genes called as up ASE/up stress or down ASE/down stress are considered to show PSR and are shown as filled circles. Other genes are shown as points. Axes through the origin are shown as red dashed lines. A small number of genes outside the axis range are not shown, but are considered in the gene count.

4.1.3 Detection of prior stress response (PSR) genes in Palomero Toluqueño

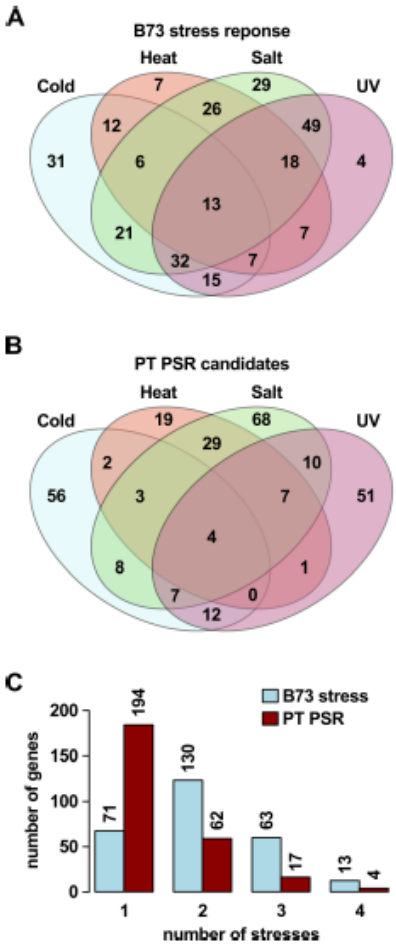

FIGURE 4.3: **PSR candidates may respond to multiple stresses in B73.** (A) Number of genes from the 277 PSR gene set that responded to cold, heat, salt, UV or a combination of stresses in the Makarevitch B73 study. (B) Number of genes called as PSR in PT with respect to each stress from the same 277 gene set. (C) Counts with respect to number of stresses of genes in A and B. Numbers above bars give counts.

To identify evidence of prior stress response (PSR) in PT, the ASE gene set was compared with a previous study reporting changes in the transcriptome of B73 seedlings exposed to cold, heat, salt or UV treatments reported by Makarevitch *et al.*, 2015 [41]. A total of 1,407 stress responsive genes identified in the Makarevitch study were also present in the 9,256 polymorphic gene set for which ASE was evaluated. Of these 1,407 genes, 432 (31%) showed ASE, a slight enrichment compared with the 2,386 (26%) ASE genes in the 9,256 polymorphic gene set as a whole (ASE, Makarevitch: 432; ASE, non-Makarevitch: 1,963; non-ASE, Makarevitch: 984; non-ASE, non-Makarevitch: 5,886; $\chi^2 = 15.7$, $d.f. = 1$, $p < 0.001$).

From this 432 gene set, a gene was considered to exhibit PSR in PT if the sign of ASE was concordant with the sign of B73 stress response: i.e., PT-up and induced by stress in B73, or PT-down and repressed by stress in B73. On this basis, a set of 277 PSR candidates was identified (Figs. 4.2A-4.2D). The majority of these 277 genes respond to two or more stress treatments (Figs. 4.3A-4.3C), but often in different directions such that they present stress-specific PSR (Figs. 4.3B-4.3C): 194 were identified as showing PSR with respect to one treatment, 62 with respect to two, 17 with respect to three, and 4 with respect to all four (Figure 4.3C). Of the 277 genes, 92 showed PSR with respect to cold, 65 with respect to heat, 136 with respect to salt, and 92 with respect to UV (Figure 4.3B). The number of PSR genes with respect to any given stress was proportional to the number of genes responding to that stress in the 1,407 polymorphic gene set (cold, PSR: 92, non-PSR: 631; heat, PSR: 65, non-PSR: 374; salt, PSR: 136, non-PSR: 736; UV, PSR: 92, non-PSR: 444; $\chi^2 = 5.2$, $d.f. = 3$, $p = 0.16$), and there was no indication of an enrichment for PSR with respect to any one of the four treatments. In contrast to the complete ASE gene set, the majority of the 277 PSR genes were PT-up (181 PT-up, 96 PT-down), although this general trend was not observed when the UV treatment was considered alone, where the majority of PSR genes were PT-down (34 PT-up, 58 PT-down; Figure 4.2D).

4.1.4 Hormone related genes and transcription factors showed prior stress responses in PT

A primary aim of the analysis was the definition of a small number of candidate genes for future functional analysis. For this purpose, the PSR candidate genes were cross-referenced with the classical maize gene list, a curated set of 4,908 well-annotated genes, many linked with existing functional data (the "combined set" gene list was obtained from *www.maizegdb.org* and filtered for unique gene identifiers). Of the 277 PSR candidate genes, 48 were present in the classical gene list (Figure 4.4), including 9 genes associated with hormone homeostasis (Table 4.1) and 12 transcription factors (TFs; Table 4.2; [26]) that were considered of special interest.

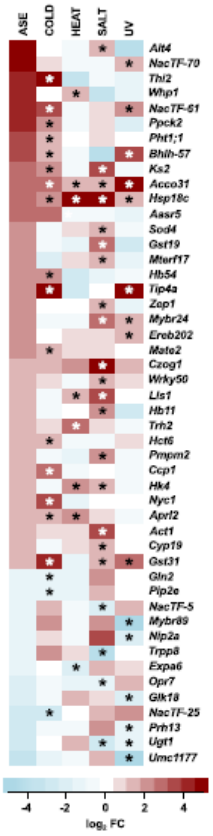

FIGURE 4.4: **Classical PSR candidate genes.** Heatmap representation of ASE ($\log_2$ PT/B73) and B73 response to cold, heat, salt and UV stress ($\log_2$ stress/control) as reported in the Makarevitch dataset for PSR candidates in the maize classical gene list. Asterisks (*) in the stress columns indicate a given gene was called as PSR with respect to that stress.

The 277 PSR candidates were cross referenced with a published study of population level differentiation between Mesoamerican and South American highland and lowland maize by Takuno *et al.*, in 2015 [65]. *Fst* estimates and significance were reported for 183 of the PSR candidates, 22 which showed significant differentiation ($p < 0.1$) between highland and lowland Mesoamerican populations, including the hormone associated gene *Czog1* (GRMZM2G168474). The number of PSR candidates showing significant *Fst* was as expected based on the overlap with the 1,407 polymorphic gene set (*Fst* reported for 1,032 of 1,407 genes; PSR, *Fst* $p < 0.1$: 22; PSR, *Fst* $p \geq 0.1$: 161; non-PSR, *Fst* $p < 0.1$: 100; non-PSR, *Fst* $p \geq 0.1$: 749; $\chi^2 < 1$, *d.f.* = 1, $p = 1$).

TABLE 4.1: **ASE and stress-responsive hormone-related genes.** List of genes involved in hormone biosynthesis, transport or catabolism present in the 277 PSR gene set. ASE call indicates biased expression of the PT allele (1) or B73 allele (-1). Response to stress indicates the name of the stress for which the gene was called as differentially expressed in the Makarevitch dataset. Prior stress response indicates the stress condition for which the sign of the ASE call and the stress response coincide.

Gene id	Symbol	Molecular Function	hormone	ASE call	Response to stress	Prior stress response
GRMZM2G070563	-	auxin efflux carrier	auxin transport	1	heat, salt, uv	heat, salt
GRMZM2G072632	-	auxin efflux carrier	auxin transport	1	heat, salt, uv	heat, salt
GRMZM2G112598	-	auxin efflux carrier	auxin transport	1	heat, salt, uv	heat, salt
GRMZM2G475148	-	auxin efflux carrier	auxin transport	1	heat, salt	heat, salt
GRMZM2G072529	*Acco31*	1-aminocyclopropane-1-carboxylate oxidase	ethylene biosynthesis	1	cold, heat, salt, uv	cold, heat, salt, uv
GRMZM2G020761	-	putative cytochrome P450 (castasterone C-26 hydroxylase)	brassinosteroid catabolism	-1	cold, salt, uv	cold, uv
GRMZM2G148281	*Opr7*	12-oxo-phytodienoic acid reductase	jasmonate biosynthesis	-1	salt, uv	salt
GRMZM2G168474	*Czog1*	*cis*-zeatin O-glucosyl transferase	cytokinin homeostasis	1	salt	salt

TABLE 4.2: **ASE and stress-responsive TFs.** List of TFs present in the 277 PSR gene set. PlantTFDB family indicates the TF family according to the PlantTFDB [26]. ASE call indicates biased expression of the PT allele (1) or B73 allele (-1). Response to stress indicates the name of the stress for which the gene was called as differentially expressed in the Makarevitch dataset. PSR indicates the stress condition for which the sign of the ASE call and the stress response coincide.

Gene id	Symbol	PlantTFDB family	ASE call	Response to stress	Prior stress response
GRMZM2G159937	*Bhlh57*	bHLH	1	cold, salt, uv	cold, uv
GRMZM2G148333	*Ereb202*	ERF	1	uv	uv
GRMZM2G010920	*Glk18*	G2-like	-1	heat, uv	uv
GRMZM2G127537	*Hb11*	HD-ZIP	1	salt, uv	salt
GRMZM2G041127	*Hb54/ZmHdz10*	HD-ZIP	1	cold, heat, salt	cold
GRMZM2G049695	*Mybr24*	MYB-related	1	salt, uv	salt, uv
GRMZM2G121753	*Mybr89*	MYB-related	-1	cold, salt, uv	uv
GRMZM2G127379	*NacTF25/ZmNAC111*	NAC	-1	cold, salt, uv	cold
GRMZM2G162739	*NacTF5*	NAC	-1	cold, salt, uv	salt
GRMZM2G003715	*NacTF61*	NAC	1	cold, uv	cold, uv
GRMZM2G312201	*NacTF70*	NAC	1	uv	uv
GRMZM2G071907	*Wrky50*	WRKY	1	salt	salt

4.2 Identification of *Cis*-regulatory variation associated with crop-wild introgressed regions in Palomero Toluqueño.

Single nucleotide polymorphisms (SNPs) identified from the B73xPT F1 transcriptome data were used to assess ASE, in order to identify co-localization and enrichment of *cis*-regulatory variation for the nine previously reported regions introgressed from *mexicana* [24].

The presence of ASE in four introgressed regions from *mexicana* located on chromosomes 4, 5, 6 and 10 was identified (Table 4.3), but no evidence of any particular enrichment was found (Figure 4.5).

TABLE 4.3: **Regions of introgression from teosinte *mexicana* to Palomero Toluqueño captured in the analyzed B73xPT F1 individual.**

Region	Chromosome	Start (Mb)	Stop (Mb)	Captured
1	1	120	145	no
2	2	73	78	no
4	4	169	180	yes
5	5	102	135	yes
6	6	46	56	yes
7	7	30	31	no
9a	9	107	125	no
9b	9	43	44	no
10	1	120	145	yes

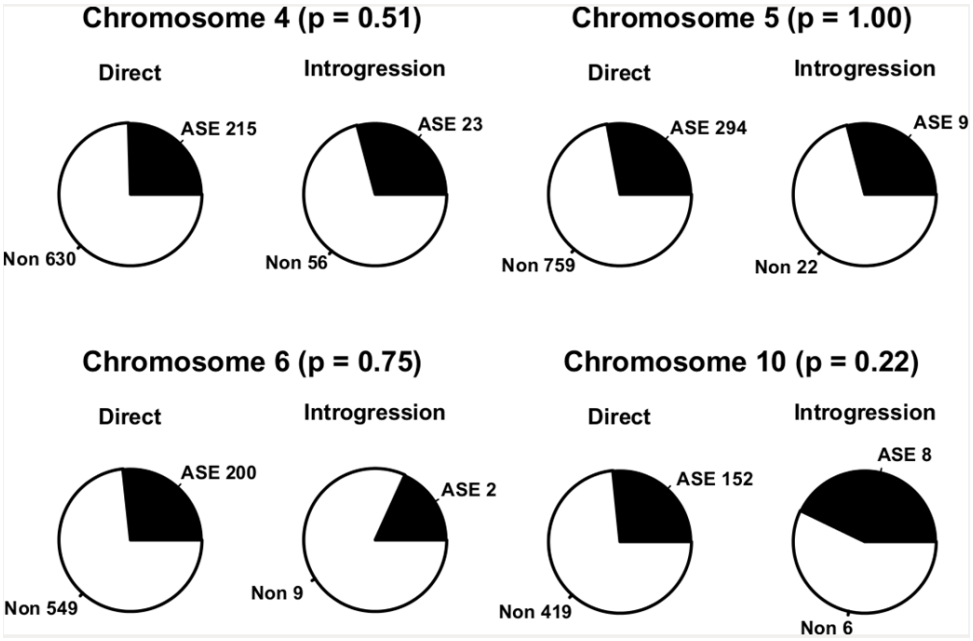

FIGURE 4.5: **Candidate *cis*-regulatory variants are equally distributed between regions of introgression.** For each chromosome, the left circle indicates genomic regions with no evidence of introgression ; the right circle represents genomic regions with a signature of introgression. Black sections of the pie chart indicate the proportion of ASE genes detected.

4.2.1 *Cis*-regulation at the *inv4m*.

Given the predicted importance of inversions in local adaptation [30, 29], distribution of ASE at *inv4m* was analyzed. *Inv4m* harbors 358 genes and at least one variant was identified for 294 of them. Twenty-nine differentially expressed genes (9.8%) between PT and B73 alleles were identified, their distribution shows a bias towards PT-down alleles (7 PT-up and 21 PT-down) reflecting the observed bias to the PT-down ASE genes called, but contrary to the PSR tendency (181 of 277 PSR genes identified presented higher expression of the PT allele), however they were distributed along the region of the inversion (Figure 4.6). The number of effective counts for each allele are shown in Table 4.4.

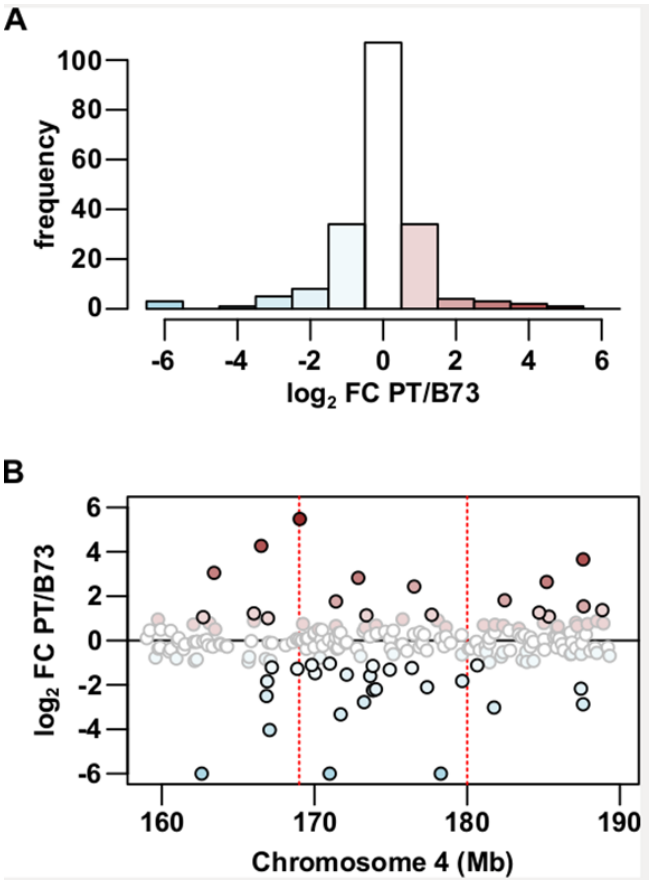

FIGURE 4.6: **Allele specific expression in the region of *inv4m*.** A. Shows the distribution PT/B73 alleles. B. Shows the physical position of ASE genes surrounding the inversion. PT induced genes (red) and PT repressed or B73 induced genes (blue).

TABLE 4.4: **ASE genes located in the *inv4m*.** Selected candidates to analyze by qRT-PCR are in bold

Locus	Annotation	chr	start	stop	B73	PT	ratio	log2_fc	chi_sq	p_value	fdr	bon	holm	Reference
GRMZM2G344967	Cell morphogenesis N-terminal	4	168867433	168885815	2318.970934	957.873365	0.413059668	-1.275577894	565.3569176	5.75E-125	5.40E-124	5.22E-121	4.66E-121	Phytozome12
GRMZM2G019177	**Metallo-beta-lactamase family protein**	**4**	**169036645**	**169041098**	**69.674222**	**3104.30263**	**44.55453597**	**5.477500411**	**2901.397837**	**0**	**0**	**0**	**0**	**Phytozome12**
GRMZM2G417682		4	169825706	169831187	1462.486157	680.889724	0.46557003	-1.102929904	285.0143969	6.07E-064	3.05E-063	5.51E-060	4.41E-060	
GRMZM2G096343	**Signal peptide peptidase**	**4**	**170036780**	**170046677**	**1042.902775**	**376.258941**	**0.360780458**	**-1.470806898**	**313.1524733**	**4.49E-070**	**2.43E-069**	**4.08E-066**	**3.33E-066**	**Phytozome12**
GRMZM2G042380	**Protein Kinase family protein-related**	**4**	**170996275**	**171000497**	**599.621621**	**0.016542**	**2.76E-005**	**-15.14563111**	**599.5719968**	**2.07E-132**	**2.02E-131**	**1.88E-128**	**1.69E-128**	**Phytozome12**
GRMZM2G123537	Pumilio homolog 4	4	171007080	171012916	5453.896547	2655.129932	0.486831737	-1.038504872	965.9722515	4.48E-212	7.27E-211	4.06E-208	3.81E-208	Phytozome12
GRMZM2G142661	**Voltage-gated shaker-like K+ channel, subunit beta/KCNAB**	**4**	**171405281**	**171409735**	**153.74098**	**521.018974**	**3.388940112**	**1.760834142**	**199.9127602**	**2.18E-045**	**8.68E-045**	**1.98E-041**	**1.48E-041**	**Phytozome12**
GRMZM2G102230	Large subunit ribosomal protein L23e (RP-L23e, RPL23)	4	171718132	171721643	163.292186	16.255564	0.09954894	-3.328450238	120.4123594	5.14E-028	1.53E-027	4.67E-024	3.10E-024	Phytozome12
GRMZM2G075336	ABC transporter I family member	4	172109718	172113857	3184.089715	1097.622323	0.344720916	-1.536499258	1016.730256	4.15E-223	7.20E-222	3.77E-219	3.55E-219	Phytozome12
GRMZM2G162007	2-haloacrylate reductase	4	172864338	172866679	37.847898	267.881632	7.077847018	2.823310579	173.079515	1.57E-039	5.70E-039	1.43E-035	1.03E-035	Phytozome12
GRMZM2G054210	**AP-1 complex subunit gamma-1**	**4**	**173239464**	**173253984**	**444.657169**	**64.85905**	**0.145863048**	**-2.777313648**	**283.1050432**	**1.58E-063**	**7.91E-063**	**1.44E-059**	**1.15E-059**	**Phytozome12**
GRMZM2G125838	**Calcinerum B Subunit**	**4**	**173401199**	**173403388**	**186.432225**	**409.465794**	**2.196325201**	**1.135091684**	**83.47732551**	**6.45E-020**	**1.62E-019**	**5.85E-016**	**3.52E-016**	**Phytozome12**
GRMZM5G855924		4	173648501	173648980	180.095299	59.393242	0.329787853	-1.600389833	60.83375223	6.21E-015	1.37E-014	5.64E-011	3.09E-011	
GRMZM2G411764	Large subunit ribosomal protein L39e	4	173821634	173822427	378.229797	171.035427	0.452199769	-1.144967838	78.15806478	9.51E-019	2.33E-018	8.64E-015	5.10E-015	Phytozome12
GRMZM2G111491	Protein kinase domain (Pkinase) // WD domain, G-beta repeat (WD40) // HEAT repeat (HEAT)	4	173835335	173846164	1244.839072	262.36151	0.210759379	-2.24633126	640.4337759	2.69E-141	2.77E-140	2.44E-137	2.21E-137	Phytozome12
GRMZM2G010491	Thioredoxin H2	4	174019150	174020537	6220.424518	1356.283882	0.21803719	-2.197353863	3122.710137	0	0	0	0	Phytozome12
GRMZM2G039263	T-complex protein 1 subunit alpha (CCT1, TCP1)	4	174931116	174936588	1328.602934	534.470031	0.402279731	-1.313729045	338.4983194	1.35E-075	7.83E-075	1.23E-071	1.02E-071	Phytozome12
GRMZM2G064537	Chaperone protein DNAJ 1, mitochondrial	4	176377675	176381799	1876.636992	796.659474	0.424514425	-1.236114517	436.297079	6.92E-097	5.10E-096	6.29E-093	5.43E-093	Phytozome12
GRMZM2G091189	**Chaperonin, choroplastic**	**4**	**176529663**	**176531723**	**64.700134**	**349.465677**	**5.401313033**	**2.433310162**	**195.7945642**	**1.73E-044**	**6.76E-044**	**1.57E-040**	**1.17E-040**	**Phytozome12**
GRMZM2G441144	Thioredoxin-like 3-1, chloroplastic	4	177376776	177378440	196.682917	45.70518	0.232380019	-2.105442069	94.04041433	3.09E-022	8.20E-022	2.81E-018	1.75E-018	Phytozome12
GRMZM2G126732	Oxidoreductase, 2OG-FE II Oxygenase family protein	4	177657802	177660210	276.420617	0	0	-Inf	276.420617	4.53E-062	2.23E-061	4.11E-058	3.27E-058	Phytozome12
GRMZM2G126795	E3 Ubiquitin-protein ligase RFWD2	4	177676332	177683190	519.999838	1161.838015	2.234304571	1.159825861	244.9440918	3.29E-055	1.49E-054	2.98E-051	2.33E-051	Phytozome12
GRMZM2G475683	SAUR-like auxin responsive protein	4	178275310	178276595	146.504454	1.148248	0.007837632	-6.995366445	143.0954282	5.60E-033	1.82E-032	5.09E-029	3.52E-029	Phytozome12
GRMZM2G478709	**Nuclear protein Skip related**	**4**	**179677461**	**179679867**	**2894.931541**	**817.725372**	**0.282467948**	**-1.823840923**	**1162.182655**	**1.01E-254**	**2.06E-253**	**9.17E-251**	**8.72E-251**	**Phytozome12**
GRMZM2G101518	Nucleolar protein 7 / Estrogen receptor coactivator-related	4	180661015	180664967	639.061994	294.162764	0.460303956	-1.119341253	127.4671271	1.47E-029	4.49E-029	1.33E-025	8.97E-026	Phytozome12
GRMZM2G161566	Shikimate kinase	4	181757980	181760735	460.53041	56.742135	0.123210398	-3.020804084	315.2012853	1.61E-070	8.75E-070	1.46E-066	1.19E-066	Phytozome12
GRMZM2G082184	Aquaporin transporter	4	182093907	182096555	89.848112	0	0	-Inf	89.848112	2.57E-021	6.70E-021	2.34E-017	1.44E-017	Phytozome12
GRMZM2G503738	Uncharacterized conserved protein	4	182452172	182461267	128.360042	451.351822	3.516295375	1.814056264	179.9578315	4.95E-041	1.84E-040	4.49E-037	3.28E-037	Phytozome12

4.2.2 Generation of inbred lines segregating for *inv4m*

In order to establish an approach to identify allelic differences at *inv4m*, the CAPS marker was used to assay PT and B73, along with teosintes *mexicana* and *parviglumis* confirming their predicted restriction pattern (Figure 3.2); PT shares the restriction pattern with *mexicana* and B73 with *parviglumis* (Figure 4.7).

Once confirmed that the PT accession used to generate the mapping population captures the characteristic *mexicana inv4m* haplotype (Figure 4.7) analysis to identify segregating stocks were performed using markers at the inversion.

SSRs inside the *inv4m* (Table A.1) and the CAPS marker to identify the characteristic SNP PZE104103103 were used to confirm the genotype for development of NILs and HIFs (B73 introgressed lines with the PT region at the *inv4m*). The lack of recombination between highland and lowland haplotypes at *inv4m* [52] facilitated the generation of test materials and assessment of the region as a block.

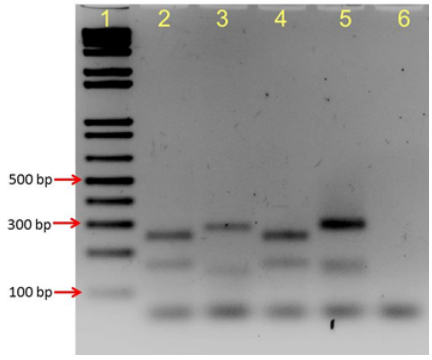

FIGURE 4.7: **Locus PZE104103103 in *parviglumis*, *mexicana*, B73 and PT from the mapping population.** DNA from the different genotypes was amplified using primers RS427 and RS428. PCR products were digested with *HinfI*, electrophoresed on a 2% agarose gel, and stained with ethidium bromide. Lane 1-DNA size markers, lanes 2 and 3 correspond to *parviglumis* and *mexicana* respectively. Lane 4 shows the B73 haplotype and lane 5 the PT haplotype carried by the mapping population. Lane six corresponds to the negative control.

NILs segregating for PT and B73 haplotypes at *inv4m*.

To analyze functional differences between B73 (standard) and PT (inverted) alleles at *inv4m*, NILs were developed by backcrossing the PT allele into the B73 background. SSR and the CAPS markers were used to select for the PT allele during backcrossing (Figure 4.8). After five generations of backcrossing (BC5), heterozygous plants at the *inv4m* (*inv4m*-B73/*inv4m*-PT) were self-pollinated to generate BC5S1 stocks, segregating for the inverted PT allele producing three genotypic classes: *inv4m*-B73/*inv4m*-B73, *inv4m*-B73/*inv4m*-PT and *inv4m*-PT/*inv4m*-PT , at the *inv4m* region (Figure 4.10). Individuals Confirmed for each genotypic class were used for analysis of transcript accumulation by quantitative PCR.

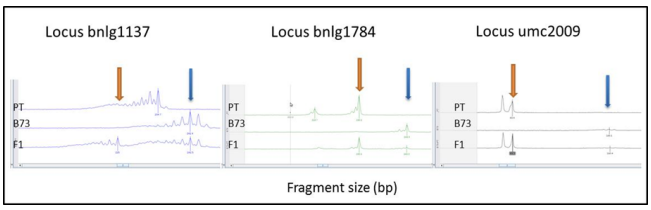

FIGURE 4.8: **PCR amplification products obtained by SSR markers using genomic DNA from PT, B73 and the B73xPT F1 as template *bnlg1137, bnlg1784* and *umc 2009*.** Products were resolved on an Applied Biosystems 3730 XL DNA Analyzer. Orange arrows indicate products amplified from PT alleles and blue arrows indicate the products of B73 alleles. The size used was GeneScanTM-500 LIZ®, indicated in Table A.1

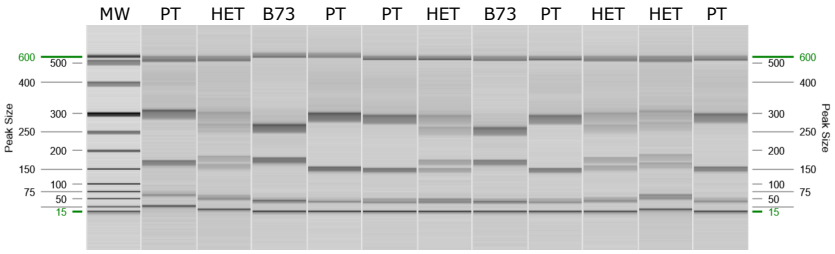

FIGURE 4.9: **Genotype pattern with the CAPS marker in a BC1S5 family segregating for *inv4m*.** Fragments amplified with the primers RS425 and RS426 were digested using the enzyme *Hinf I*. Digested products were analyzed on a gel matrix by automatic high-resolution capillary electrophoresis (QIAGEN).

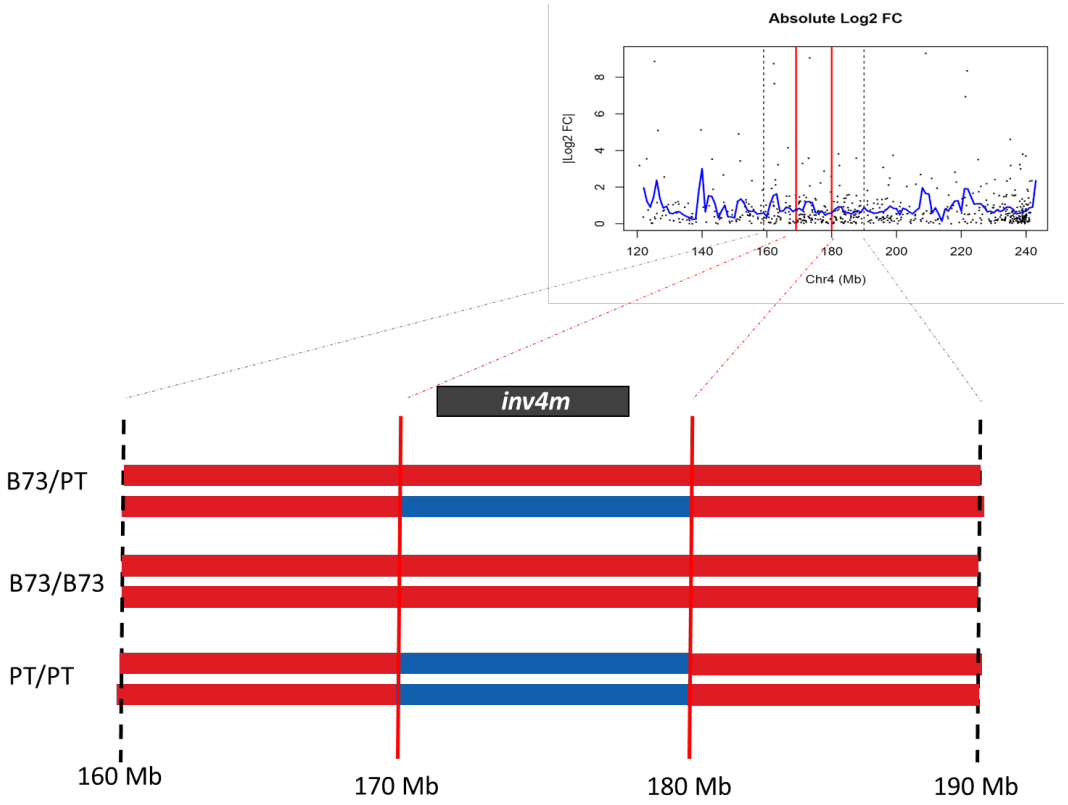

FIGURE 4.10: **Chromosomal context representation for the BC1S5 segregants at *inv4m*.** Upper right panel represents the portion of chromosome 4 (from 120 to 240 Mb) that spans *inv4m* showing ASE gene distribution. Lower panel shows and enlarged image of the *inv4m* region. Each pair of horizontal lines represent the chromosome section spanning and segregating for *inv4m*. Possible genotypes are depicted by red or blue. Blue fragments indicate the PT alleles in the B73 background (red). Red vertical lines represent the limits of the inversion [52], dotted vertical lines show 10 Mb regions flanking *inv4m*

HIFs segregating for PT and B73 haplotypes at *inv4m*

A collection of ∼150 PTxB73 BC1S3 families generated from B73 (recurrent parent) and PT (donor parent) was screened to identify heterogeneous inbred families (HIFs; [67]) segregating for B73 or PT alleles at *inv4m*. Multiple individuals were pooled per family and screened by SSR genotyping to identify families where both B73 and PT alleles were present. Individual plants from HIF families were genotyped using the CAPS marker to confirm segregation at *inv4m*. Heterozygous individuals at the inversion were self-pollinated to generate pairs of near-isogenic sibling families from the segregation with the region fixed for either the B73 or PT allele.

Three BC1S3 HIF families (heterozygous for *inv4m*) were identified (HIF 56, HIF 122 and HIF 134), and three near-isogenic pairs generated by selfing these individuals. Homozygous plants *inv4m*-B73/*inv4m*-B73 and *inv4m*-PT/*inv4m*-PT were genotyped by GBS at Cornell to confirm the genotype at *inv4m*, characterize the flanking regions and to identify unlinked PT gene content throughout the genome (Figure 4.11).

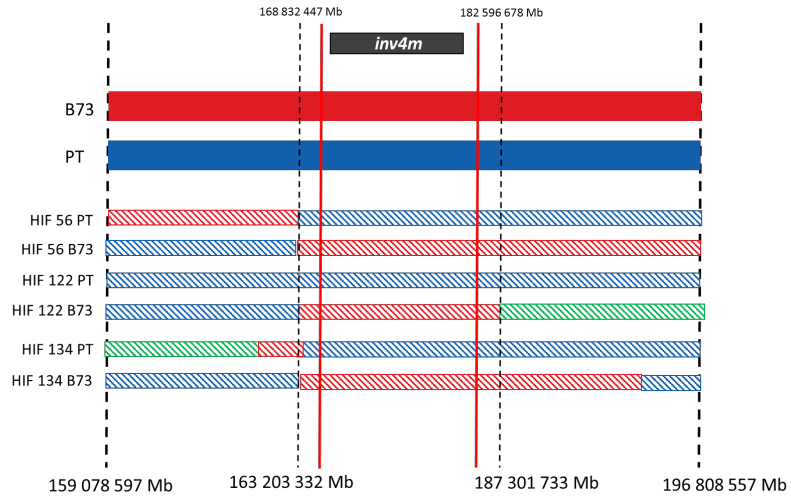

FIGURE 4.11: **HIF genotypes at *Inv4m* for the families 56, 122 and 134.** Scheme of *inv4m* plus 10 Mb of flanking region on each side, in B73 and PT parents and three HIF families. Homozygous regions for B73 are shown in red and for PT in blue, heterozygous regions are shown in green. Vertical red lines indicate the limits of the inversion reported by Pyhajarvi *et al*., in 2013 [52]. Vertical central dotted lines indicate the break points observed in GBS data.

4.2.3 Effect of the *inv4m* inversion on gene expression

To validate *cis*-acting regulatory variation within *inv4m*, eight candidate genes were selected on the basis of the B73xPT F1 ASE expression experiment and analyzed by real-time PCR in the *inv4m* segregants of the B73xPT BC5S1 population for B73 or PT alleles (*inv4m*-NIL). The candidate gene panel consisted of four genes identified as PT-up and four as PT-down in the previous F1 ASE experiment (Tables 4.4 and A.4). The candidate panel included the PT-up gene GRMZM2G019177, ortholog of the *Arabidopsis thaliana* gene *ASI4* (AT4G33540), previously reported to be

AS-inducible with putative role in arsenic detoxification [13]; the PT-up gene GRMZM2G091189 encoding a chaperonin10 *cpn10*, key cellular component in numerous folding pathways leading to biologically active proteins (Viitanen 1995, Levy-Rimler 2002); and the PT-down gene GRMZM2G478709, ortholog of the *Arabidopsis thaliana* gene AT1G77180 that encodes a putative transcription factor involved in response to abscisic acid, salt and osmotic stress. Ten plants of each parent PT and B73, and fifty BC5S1 segregants were grown under green house conditions, as described previously for the B73xPT F1 hybrid. 14 days after germination, leaf tissue was harvested and frozen in liquid nitrogen prior to nucleic acid extraction. Individual plants were genotyped for *inv4m* using the CAPS marker (Figure 4.9) and six individuals per genotypic class (*inv4m*-B73/*inv4m*-B73; *inv4m*-B73/*inv4m*-PT; *inv4m*-PT/*inv4m*-PT) were assayed for analysis of transcript accumulation by qRT-PCR.

For six out of eight candidate genes (three PT-up, three PT-down), the median level of transcript accumulation, as determined by qRT-PCR, in individuals homozygous for *inv4m*-B73 or *inv4m*-PT reflected allele specific expression as estimated in the previous B73xPT F1 transcriptome analysis (Figure 4.12; Table 4.5), although the difference was significant in only three cases (the PT-up gene GRMZM2G091189, and the two PT-down genes GRMZM2G042380 and GRMZM2G054210). For half of the eight genes, the median level of transcript accumulation in heterozygous plants (*inv4m*-B73/*inv4m*-PT) was intermediate between that of the two parents, consistent with the expected additive effect, linked to *inv4m*. For all four PT-up candidates, the median level of transcript accumulation in *per se* PT plants was higher than that in B73, and the range, in all cases, was greater than that observed between homozygous for *inv4m*-B73 and *inv4m*-PT segregating individuals, indicating the action of *trans* acting effects (Figure 4.12). Contrary to expectations, in three of four PT-down candidates (GRMZM2G054210, GRMZM2G096343 and GRMZM2G47870), median transcript accumulation in the *per se* PT parent was actually higher than in B73, although for two of these cases (GRMZM2G054210, GRMZM2G096343) the pattern in the *inv4m* segregants fulfilled the expectation of lower accumulation in the *inv4m*-PT homozygote, suggesting strong *trans* acting effects.

TABLE 4.5: **Allele expression differences between B73 and PT.** Results from the B73xPT F1 RNAseq data are the counts for B73 and PT alleles. Data for each genotype is the median of the relative levels of gene expression from six independent biological replicates grown 14 days under greenhouse conditions.

gene_id	B73 counts	PT counts	Relative level of gene expression per genotype					Dominance effect		Cis effect	Trans effect
			B73	PT	HOM B73	HET	HOM PT	Expected	Observed		
GRMZM2G019177	69.674222	3104.30263	1.0014189	2.9790779	1.02123878	0.98704782	1.90763266	1.46443572	0.98704782	0.44820359	0.55179641
GRMZM2G042380	599.621621	0.016542	0.99015443	0.53959899	0.97565629	0.59953886	0.30330569	0.63948099	0.59953886	1.49227053	-0.49227053
GRMZM2G054210	444.657169	64.85905	1.00697059	1.8297114	1.53490914	0.44650113	0.95499785	1.24495349	0.44650113	-0.70485295	1.70485295
GRMZM2G091189	64.700134	349.465677	0.9908031	2.17306993	0.98579393	1.54912998	1.75158937	1.36869165	1.54912998	0.64773486	0.35226514
GRMZM2G096343	1042.90278	376.258941	1.00007516	3.10384626	1.56594509	1.08182055	1.20049869	1.38322189	1.08182055	-0.17371015	1.17371015
GRMZM2G125838	186.432225	409.465794	1.00851622	1.7655288	0.8338053	1.20041728	1.50882611	1.17131571	1.20041728	0.89169034	0.10830966
GRMZM2G142661	153.74098	521.018974	1.00078762	1.59274192	1.10508957	1.23785818	1.04894605	1.07701781	1.23785818	-0.09484435	1.09484435
GRMZM2G478709	2894.93154	817.725372	1.00902987	2.57020841	1.38419365	1.60046047	1.63031836	1.50725601	1.60046047	0.15765314	0.84234686

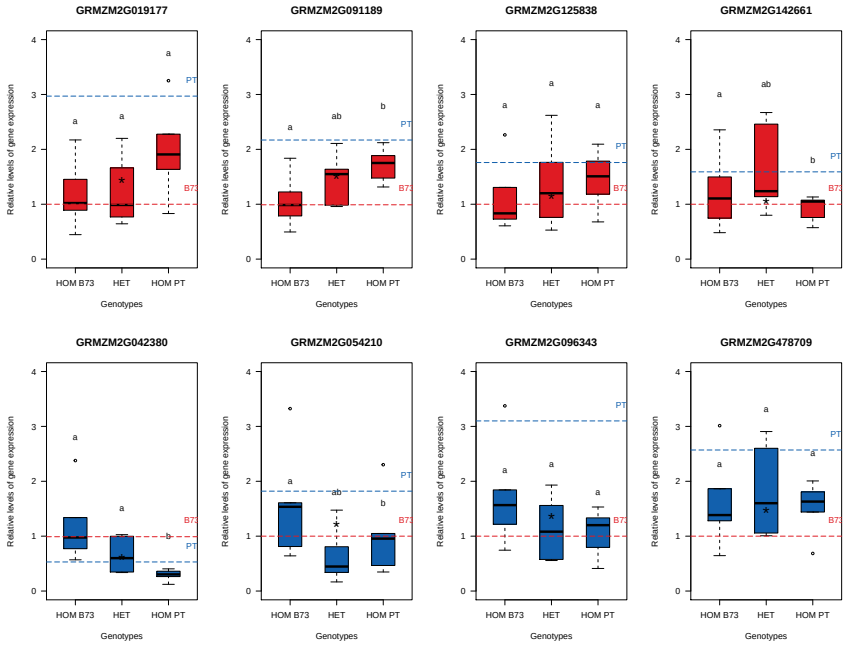

FIGURE 4.12: **qRT-PCR for eight genes at *inv4m* in the BC5S1 segregants.** Accumulation of transcripts identified as PT-up (red) or PT-down (blue) in the prior ASE analysis, analyzed by qRT-PCR in a BC5S1 stock, segregating *inv4m*-B73 and *inv4m*-PT alleles in the B73 background. Boxes show 1st quartile, median and 3rd quartile of transcript accumulation in relative levels of gene expression for six individuals per genotypic class (HOM B73: *inv4m*-B73 and *inv4m*-B73; HET: *inv4m*-B73 and *inv4m*-PT; HOM PT: *inv4m*-PT and *inv4m*-PT).Whiskers extend to the most extreme points within 1.5x box length; outlying values beyond this range are not shown. Dotted lines indicate the median transcript accumulation in B73 (red) and PT (blue) parents *per se*. Letters above each box indicate means groups assigned on the basis of Tukey LSD. Black asterisks indicate the expected level of transcript accumulation in *inv4m*-B73 and *inv4m*-PT under additivity.

4.3 Co-localization of introgression regions with QTL linked to morphological and agronomic traits in highland maize

4.3.1 Field trials

To evaluate the relationship between *mexicana* introgression and phenotypic variation, the B73xPT BC1S5 RILs and two HIF families (HIF 56 and HIF 134) were evaluated in lowland (Valle de Banderas, Nayarit, Mexico, hereafter *lowlands*) and highland (Metepec, Mexico State, Mexico, hereafter *highlands*) field sites, during 2016. Parental lines B73 and PT were evaluated in the same field sites, over two years (2015 and 2016). Two complete blocks (each block consisting of one 15 plant single row plot per genotype) were grown in the lowland site, and three complete blocks were grown in the highland site. The position of each line within the block was randomized. The parental lines B73 and PT were included at random positions within each block, at a frequency of 5%. A number of phenotypes described in Table 3.2 were evaluated as indicators of adaptation, phenology [58], and plant morphology.

B73 and PT parents were compared both within and between sites. On the basis of individual ear and grain weight, B73 was found to perform better than PT in the lowlands, and PT to perform better than B73 in the highlands - a textbook example of local adaptation [58] (Figure 4.13).

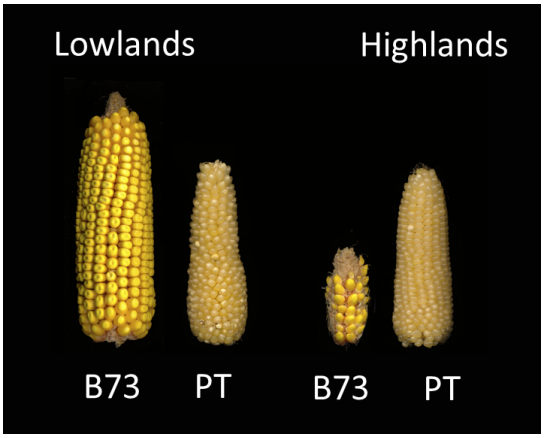

FIGURE 4.13: **Ear size variation in B73 and PT in plants grown in Valle de Banderas (lowlands) and Metepec (highlands).**

Phenotypic variance was analyzed using the linear model:

$$P = G + E + GEI + error \tag{4.1}$$

where P is the phenotype, G is the genetic contribution, E is the effect of the environment (*plasticity*), and GEI is the genotype-specific effect of the environment (*genotype x environment interaction*) [62]. For all traits, the GEI term was significant, indicating G and E main effects and their interaction to differ between B73 and PT parents, *i.e* all traits showed plasticity, differed between B73 and PT parents, the impact of the environment was genotype specific (Table 4.6).

TABLE 4.6: **Two way ANOVA for effects of genotype and environment in the parental lines B73 and PT.** Asterisks indicate significance at $p < 0.05$

Trait	ID	Source of variation	*df*	F	*p*
Stand	STD	Genotype	1	0.416	0.52
		Environment	1	5.43	0.02*
		Genotype x Environment	1	8.64	0.004*
Days to silking	DTS	Genotype	1	519.7	< 2e-16 *
		Environment	1	2485.8	< 2e-16 *
		Genotype x Environment	1	42.6	5.21e-09 *
Days to anthesis	DTA	Genotype	1	282.91	< 2e-16 *
		Environment	1	1831.39	< 2e-16 *
		Genotype x Environment	1	70.81	1.35e-12 *
Plant height	PH	Genotype	1	20.18	9.83e-06 *
		Environment	1	113.37	2e-16 *
		Genotype x Environment	1	26.33	5.00e-07 *
Ear height	EH	Genotype	1	0.001	0.975369
		Environment	1	171.266	< 2e-16 *
		Genotype x Environment	1	11.811	0.000667 *
Tassel branch number	TBN	Genotype	1	295.31	< 2e-16 *
		Environment	1	38.79	1.48e-09 *
		Genotype x Environment	1	13.98	0.000219 *
Tassel length	TL	Genotype	1	120.64	<2e-16 *
		Environment	1	383.42	<2e-16 *
		Genotype x Environment	1	92.78	<2e-16 *
Ear length	LM	Genotype	1	153.341	<2e-16 *
		Environment	1	383.42	<2e-16 *
		Genotype x Environment	1	92.78	<2e-16 *
Ear diameter	DM	Genotype	1	198.798	< 2e-16 *
		Environment	1	4.594	0.0336 *
		Genotype x Environment	1	48.973	7.09e-11 *
Ear weight	PM	Genotype	1	78.185	1.45e-14 *
		Environment	1	2.955	0.0883
		Genotype x Environment	1	75.184	3.57e-14 *
Fifty kernel weight	FKW	Genotype	1	78.185	1.45e-14 *
		Environment	1	2.955	0.0883
		Genotype x Environment	1	75.184	3.57e-14 *
Total grain weight	PTG	Genotype	1	78.185	1.45e-14 *
		Environment	1	2.955	0.0883
		Genotype x Environment	1	75.184	3.57e-14 *

To further investigate trait distribution across the B73xPT RIL population, pairwise correlations across environments were examined for all traits in the BC1S5 population. The flowering traits days to silking (DTS) and days to anthesis (DTA) were highly correlated to each other with a correlation close to 1. Ear weight (PM) and total grain weight (PTG) were also highly correlated (0.97), as were plant height (PH) and ear height (EH) by 0.71. Ear length (LM) and ear diameter (DM) were moderately correlated 0.41, while tassel branch number (TBN) and tassel length (TL) were the most weakly correlated traits in this study (0.31) (Figure 4.14).

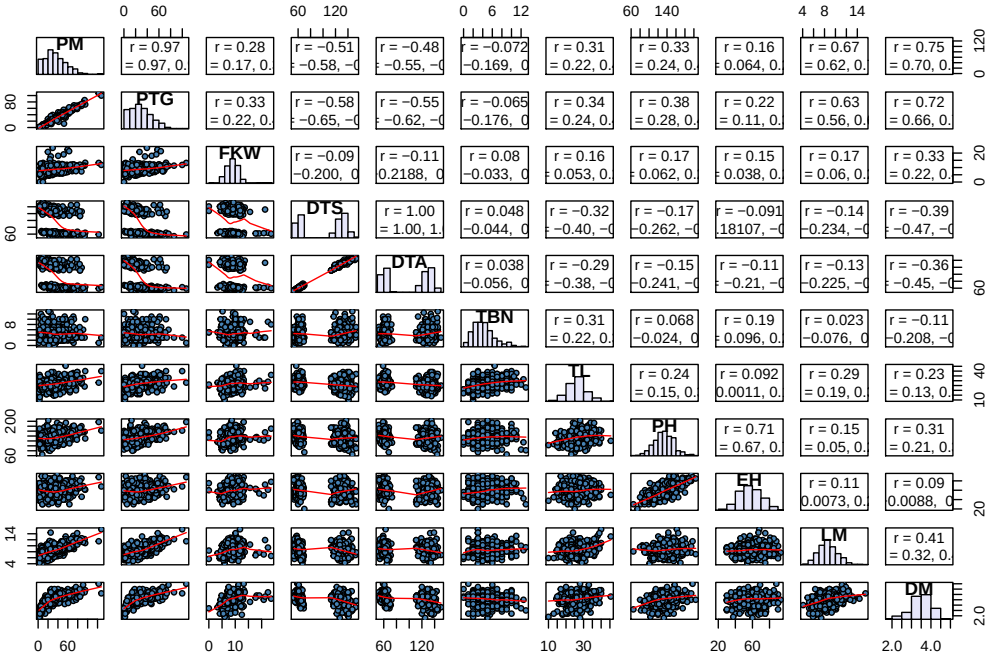

FIGURE 4.14: **Correlation analysis of the B73xPT BC1S5 recombinant inbred line (RIL) population.** Histograms in the diagonal show the phenotypic distribution of each trait. The values above the diagonal are pairwise correlation coefficients between traits, and the plots below the diagonal are scatter plots of compared traits. Ear weight (PM); total grain weight (PTG); fifty kernels weight (FKW); days to silking (DTS); days to anthesis (DTA); tassel branch number (TBN); tassel length (TL); plant height (PH); ear height (EH); ear length (LM); and ear diameter (DM).

Analyzed traits were grouped in two major categories: 1) plant fitness and phenology [58]; and 2) plant morphology. The first category contained germination (GER), stand count (STD), days to silking (DTS), days to anthesis (DTA) and total grain weight (PTG); the second category contained plant height (PH), ear height (EH), tassel branch number (TBN), tassel length (TL), ear diameter (DM), ear length (LM), ear row number (NHG), and the ratio ear length/ear diameter (LM/DM). A summary of all traits is shown in Table 4.7.

To further analyze the *mexicana* introgression effect. The signature region of introgression on chromosome 4 (*inv4m*) was evaluated through the HIF families 56 and 134, phenotyped for the same number of traits in both locations. Analysis of variance was performed for the linear model: trait $\sim$ *inv4m* * Family * Location, significance was recorded in Table 4.8.

TABLE 4.7: **Traits measured in the B73xPT BC1S5 population, means and standard deviations by location.**

Trait	Description	Mean (±SD)	
		Valle de Banderas	Metepec
Yield and yield components			
PM	Ear weight (g)	46.85±18.66	18.25±19.00
PTG	Total grain weight (g)	38.70±15.72	21.42±16.31
FKW	50 kernel weight (g)	8.88±1.80	8.78±2.69
Phenology			
DTS	Sowing to 50% silking (days)	63.89±30.79	131.00±7.40
DTA	Sowing to 50% anthesis (days)	61.07±3.80	129.86±7.44
Morphology (per plant basis)			
PH	Plant height (cm)	141.52±16.97	132.47±23.14
EH	Ear height (cm)	61.05±10.73	56.94±12.34
TL	Tassel length (cm)	29.00±4.44	25.584±4.74
TBN	Tassel primary branch no.	4.81±2.68	4.62±2.39
Morphology (per ear basis)			
LM	Ear lenght (cm)	8.99±1.81	8.46±1.78
DM	Ear diameter (cm)	3.73±0.38	3.38±0.63
DO	Cob diameter (cm)	2.39±0.38	2.52±0.42

Plant fitness and phenology

Stand count (STD), representing primarily germination and seedling establishment, was strongly plastic and showed a rank-changing GEI effect, PT being superior in the highlands, but B73 in the lowlands (Figure 4.15**A**). Flowering time (chronological days) was very sensitive to environmental conditions, PT flowered earlier in both environments, but more clearly so in the highlands, reflected by significant GEI although this did not produce rank-changes between the two parental genotypes (Figures 4.15**B-C**). On the basis of total grain weight (PTG), B73 performance was better than PT in the lowlands, and PT was shown to perform better than B73 in the highlands (Figure 4.15**D**). Broadly, the RIL population captured the phenotypic range of the parents in both environments, with evidence for transgressive segregation (i.e. values more extreme than either parent) in most of the traits across sites, with the exception of flowering time: no RIL flowered as early as the PT parent, possibly indicating unintentional selection during the generation of the RIL population. The HIF families 56 and 134 showed plasticity for STD, DTA, DTS and PTG across environments, broadly following the behavior of the B73 parent. A family effect between 56 and 134 was significant for all these traits (STD, DTS, DTA, PTG) and for the interaction with field location, except for PTG (Table 4.8). The *inv4m* effect was not significant for any trait in this group at $p < 0.05$ (Table 4.8), however a strong *inv4m* stable (non-plastic) effect was observed on PTG that showed an interaction with family: for family 56, the *inv4m*-B73 haplotype was associated with higher PTG in both environments; for family 134 the *inv4m*-PT haplotype was associated with higher PTG (Figure 4.16**D**).

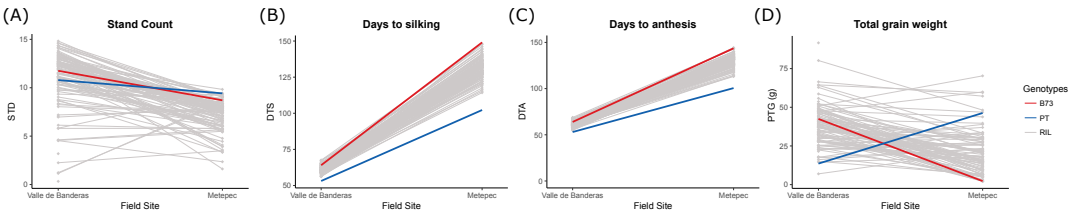

FIGURE 4.15: **Reaction norms for fitness and phenology traits in B73 (red), PT (blue) and the RIL population (gray) grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Stand count (STD), represents germination and establishment of plants from 0 to 15 days; (B) Days to silking (DTS); (C) Days to anthesis (DTA); and (D) Total grain weight (PTG).

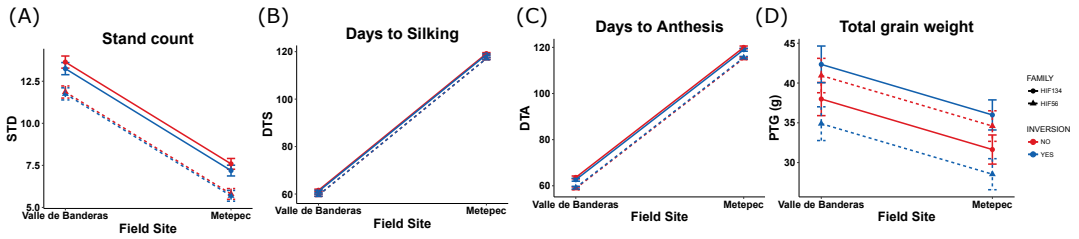

FIGURE 4.16: **Reaction norms for fitness and phenology traits in the HIF families 56 and 134 grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Stand count (STD), (B) Days to silking (DTS), (C) Days to anthesis (DTA), and (D) Total grain weight (PTG). Dashed lines correspond to the family HIF 56 and solid lines to HIF 134. *inv4m*-PT genotypes are represented in blue and *inv4m*-B73 are represented in red. Triangles and circles represent the mean $\pm$ SE for HIF families 56 and 134 respectively.

Plant morphology

Plant height (PH) and ear height (EH) differed between B73 and PT parents and were plastic between lowland and highland sites where scored (Figure 4.17**A-B**). A mild rank-changing GEI effect was seen for EH. Extensive transgressive segregation was seen in the RIL population for both PH and EH. A stable increase in PH and EH was associated with *inv4m-PT* in the HIFs, where the *inv4m* was significantly different, as well as the *inv4m**Location term at $p < 0.05$, most clearly in family 134 (Table 4.8, Figures 4.18**A-B**).

Male inflorescence traits were more stable across environment, and strong GEI was observed in the parents (Figure 4.17**C-D**). Rank changing GEI was observed within the RIL population for both male and female inflorescence architecture traits (Figures 4.17**C-D** and 4.19**A-D**). Clear differences in TL and TBN were observed linked to the *inv4m* within the HIFs, although the sign of the effect was different between families 56 and 134, there were shorter and more branched tassels associated with the *inv4m*-PT in family 56 and larger and less branched tassels in family 134 for the same allele (Figures 4.18**C-D**), although the location term did not shown significance for TBN.

In the B73 and PT parents, female inflorescence architecture traits were strongly plastic, with rank changing GEI. B73 has bigger ears than PT in lowlands but smaller in highlands. The proportion ear length over ear diameter shown to be very stable for PT, as well as the kernel row number across environments, but B73 decreased this two parameters in the highlands. Ear traits were plastic in the RILs and the HIFs,

where the *inv4m* effect was significant for all traits. In both HIF families 56 and 134, (*inv4m*-B73) was associated with a greater number of kernel rows and greater ear diameter under both environments (Figures 4.20**D** and 4.20**B**). Ear length and the ratio LM/DM were greater for *inv4m*-PT genotypes in both families 56 and 134, however, LM/DM was the only trait where the family term did not shown significance (Figures 4.20**A** and 4.20**C**; Table 4.8).

The results described above show that plant morphology traits are clearly affected by the *inv4m*, interestingly, fitness and phenolgy traits did not show an effect of *inv4m*, however PTG is highly related with the shape of the ear, that was clearly affected by the *inv4m* tending to produce long and thin ears with fewer rows in both HIF families. The question is why PTG is higher in family 134, big seeds or more seeds?

In FKW seed weight of family 134 is high, indicating large seeds in this family Therefore PTG is not related to ear shape

menos hileras en las dos familias por la inversion

In family 56, *inv4m*-PT for PTG is related with less weight, short tassel and more branches whilst records for ears belonging to family 134 show more weight, longer tassels and fewer branches. With respect to ear shape the question arises as to why PTG was higher in one family in comparison to the other.

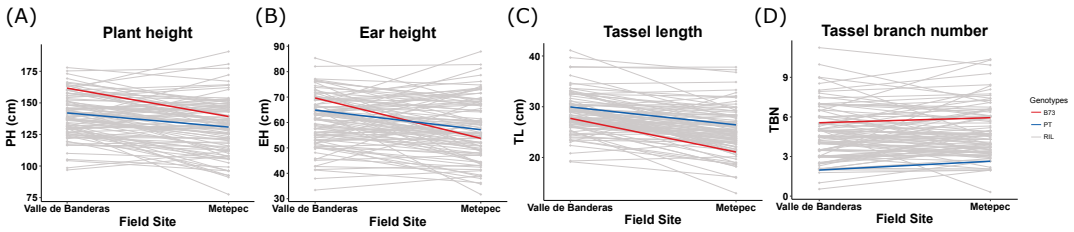

FIGURE 4.17: **Reaction norms for plant morphology traits in B73 (red), PT (blue) and the RIL population (gray) grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Plant height (PH); (B) Ear height (EH); (C) tassel length; and (D) tassel branch number (TBN). Gray lines represent each genotype of the RIL population.

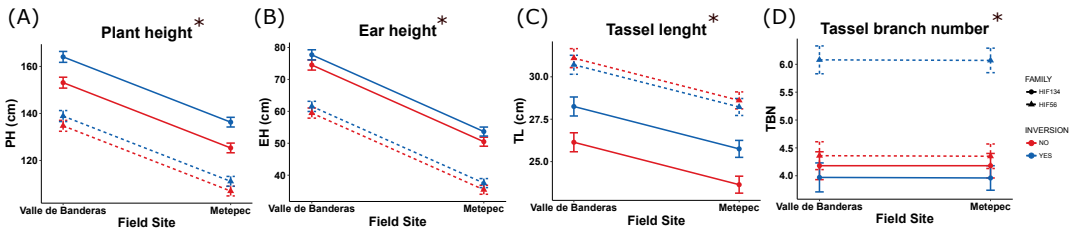

FIGURE 4.18: **Reaction norms for plant morphology traits in the HIF families 56 and 134 grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Plant height (PH); (B) Ear height (EH); (C) tassel length; and (D) tassel branch number (TBN). Dashed lines correspond to the family HIF 56 and solid lines to HIF 134. *inv4m*-PT genotypes are represented in blue and *inv4m*-B73 are represented in red. Triangles and circles represent the mean $\pm$ SE for HIF families 56 and 134 respectively.

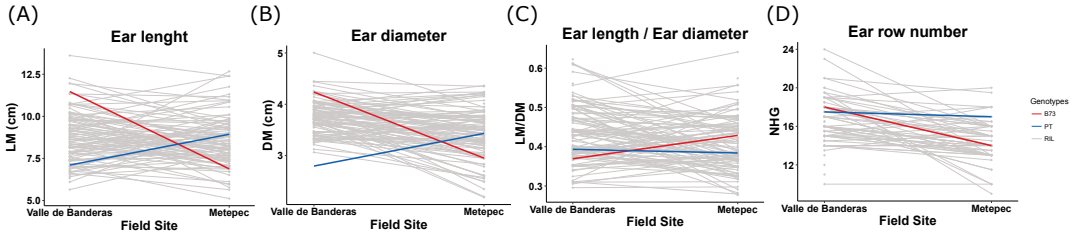

FIGURE 4.19: **Reaction norms for inflorescence traits in B73 (red), PT (blue) and the RIL population (gray) grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Ear length (LM); (B) ear diameter (DM); (C) the ratio ear length over ear diameter (LM/DM); and (D) kernel row number (NHG).

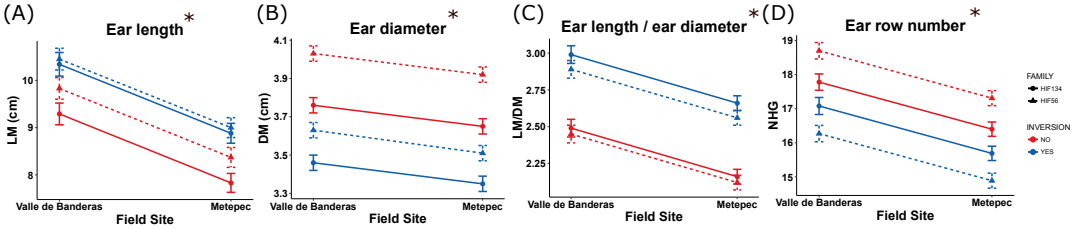

FIGURE 4.20: **Reaction norms for inflorescence architecture traits for the HIF families 56 and 134 grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** (A) Ear length (LM); (B) ear diameter (DM); (C) the ratio ear length over ear diameter (LM/DM); and (D) kernel row number (NHG). Dashed lines corresponds to the family HIF 56 and solid lines to HIF 134. *inv4m*-PT genotypes are represented in blue and *inv4m*-B73 are represented in red. Triangles and circles represent the mean $\pm$ SE for HIF families 56 and 134 respectively.

TABLE 4.8: **Three way ANOVA for effects of *inv4m*, the HIF families (56 and 134) and the field location. Analysis of significance was developed under the linear model: trait $\sim$ *inv4m* + Family + Location + *inv4m* * Family + *inv4m* * Location + Family * Location.** Asterisks indicate significance at $p < 0.05$

Trait	Family	inv4m	Location	Family * inv4m	Family * Location	inv4m * Location
STD	***		***		***	
DTS	***		***		***	
DTA	***		***		***	
PTG	***		***			
PH	***	***	***			***
EH	***	***	***			***
TL	***	***	***		***	
TBN	***	***			***	
LM	***	***	***		***	
DM	***	***	***		***	***
LMDM		***	***		***	***
NHG	***	***	***		***	

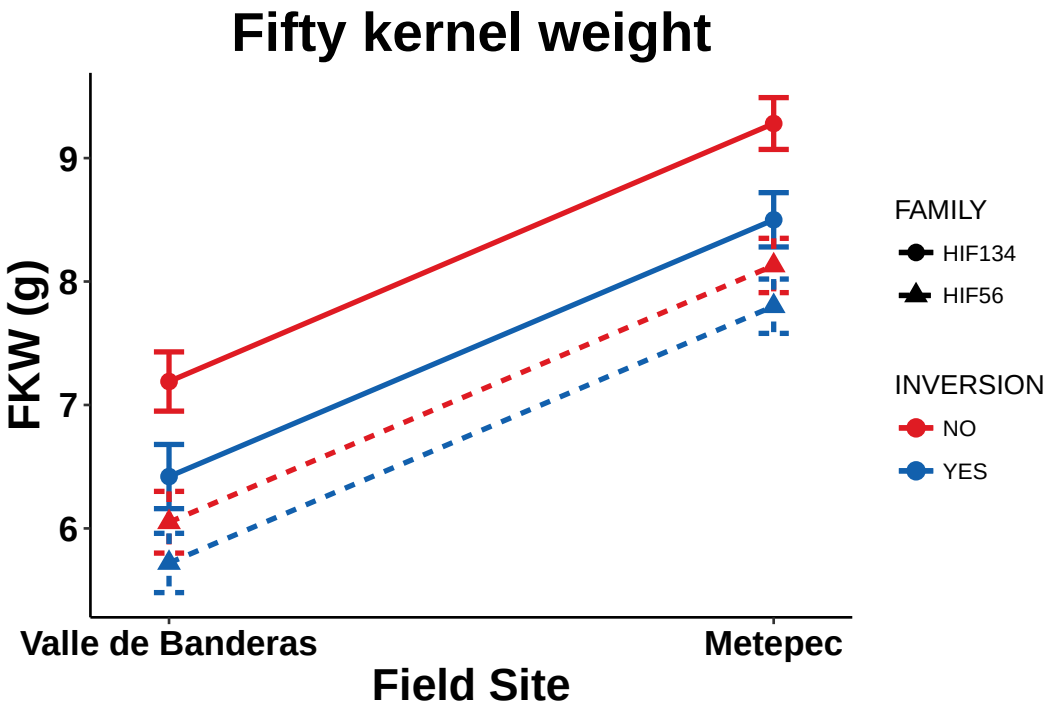

FIGURE 4.21: **Reaction norms for FKW in HIF families 56 and 134 grown at Valle de Banderas (lowlands) and Metepec (highlands) field sites.** Dashed lines corresponds to the family HIF 56 and solid lines to HIF 134. *inv4m*-PT genotypes are represented in blue and *inv4m*-B73 are represented in red. Triangles and circles represent the mean $\pm$ SE for HIF families 56 and 134 respectively.

4.3.2 Genotypic analysis and construction of a linkage-map

A total of 94 individuals of the B73xPT BC1S5 population, two samples per genotype of the HIF families and the parental lines B73 and PT were analyzed by genotyping-by-sequencing (GBS) using the Cornell platform [17]. Genotypes were called with respect to the B73 reference genome AGPv3.22 using the TASSEL 5.2.19 production pipeline. Missing data were imputed using the ABHGenotypeR plugin for TASSEL (Figure 4.22; Table 4.9). After SNP calling, individual sites were filtered by 1) frequency (minimum of 0.1 and maximum of 0.9), and 2) call rate (at least 70 individuals). The data were then filtered further to remove individuals showing high levels of missing data or an apparent excess of recombination or heterozygosity. The filtered set consisted of 74 individuals and 60,906 sites. A selection of 11,630 non-redundant sites was then made using the ThinSitesByPositionPlugin from TASSEL for generation of the linkage map. Genotyping of the HIF families confirmed the limits of introgression at *inv4-m* and identified non-linked segregating sites (probablemente aqui necesito una figura de los HIFs en la region de la inversión). As was expected, no recombination was observed across *inv4m* in either the RILs or HIFs.

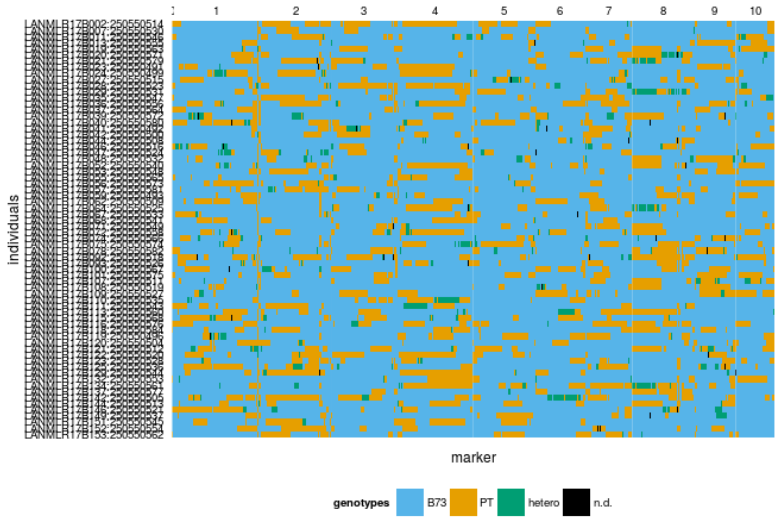

FIGURE 4.22: **RIL genotypes after imputation and filtering by site and individual** Each line is represented horizontally and its genotype across the ten chromosomes (numbered) colored in blue for B73, yellow for PT and green for the heterozygous regions.

TABLE 4.9: **Proportion of SNPs per genotype in the PTxB73 BC1S5 population.** The expected proportion of genotypes per line is 75% B73 and 25% PT due to the crossing scheme used for this population

B73	PT	Heterozygous	n.d.
76.388	21.452	1.958	0.202

In the RIL population, the Cornell SNP dataset (61 individuals total) was combined with data for an additional 30 lines genotyped using the Novogene GBS platform. Additional filtering resulted in a combined dataset comprised of 2067 SNP markers for the Cornell dataset and 1505 for the Novogene dataset. Heterozygous data were replaced by "NA" in both data sets. Additional steps to remove non informative markers are described below:

- From 2067 Markers in the Cornell dataset:
    1. Removal of duplicate/redundant markers reduced the number of SNPs to 950 markers.
    2. Elimination of markers showing strong segregation distortion with respect to the proportion B73/PT (75/25) reduced the dataset to 788 markers.
    3. Further thinning of linked markers reduced the dataset to 351 markers.
    4. Finally, markers with error LOD > 0.01 in 5 or more individuals were removed to finish with 340 markers from Cornell (Figure 4.23).

- From 1505 Markers in the Novogene dataset:
    1. Removal of duplicate/redundant markers reduced the number of SNPs to 1281 markers,
    2. Elimination of markers showing strong segregation distortion with respect to the proportion B73/PT (75/25) reduced the dataset to 861 markers.
    3. Further thinning of linked markers reduced the dataset to 768 markers.
    4. Finally, markers with error LOD > 0.01 in 5 or more individuals were removed to finish with 741 markers from Novogene (Figure 4.24).

Little overlap was observed between the sites called in the Novogene and Cornell platforms, because genomic DNA was digested with different restriction enzymes. The two datasets (Figure 4.26), were ordered by physical position according to the B73 reference genome AGPv3.22, allowing for missing data at sites not typed in both platforms. A final filtering step resulted in a dataset of 744 markers (Figure 4.25). Missing values were imputed and the linkage map was constructed using Rqtl.

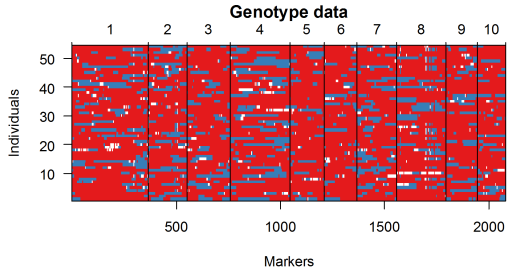

FIGURE 4.23: **Genotypes of 61 individuals analyzed by Cornell GBS.** Each row represents a different RIL typed across the ten chromosomes (numbered); red fragments depict B73 genotype, blue PT genotype, and white represents missing data.

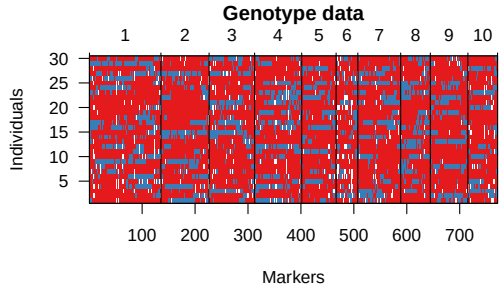

FIGURE 4.24: **Genotypes of 30 individuals analyzed by Novogene GBS.** Each row represents a different RIL typed across the ten chromosomes (numbered); red fragments depict B73 genotype, blue PT genotype, and white represents missing data.

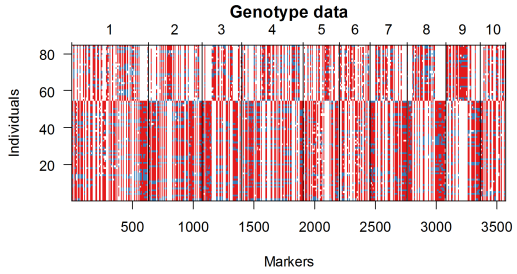

FIGURE 4.25: **Genotypes of 91 individuals combining Cornell and Novogene data** RILs analyzed using Novogene (upper part of the figure) or Cornell (bottom part of the figure) platforms. Each row represents a different RIL typed across the ten chromosomes (numbered); red fragments depict B73 genotype, blue PT genotype, and white represents missing data.

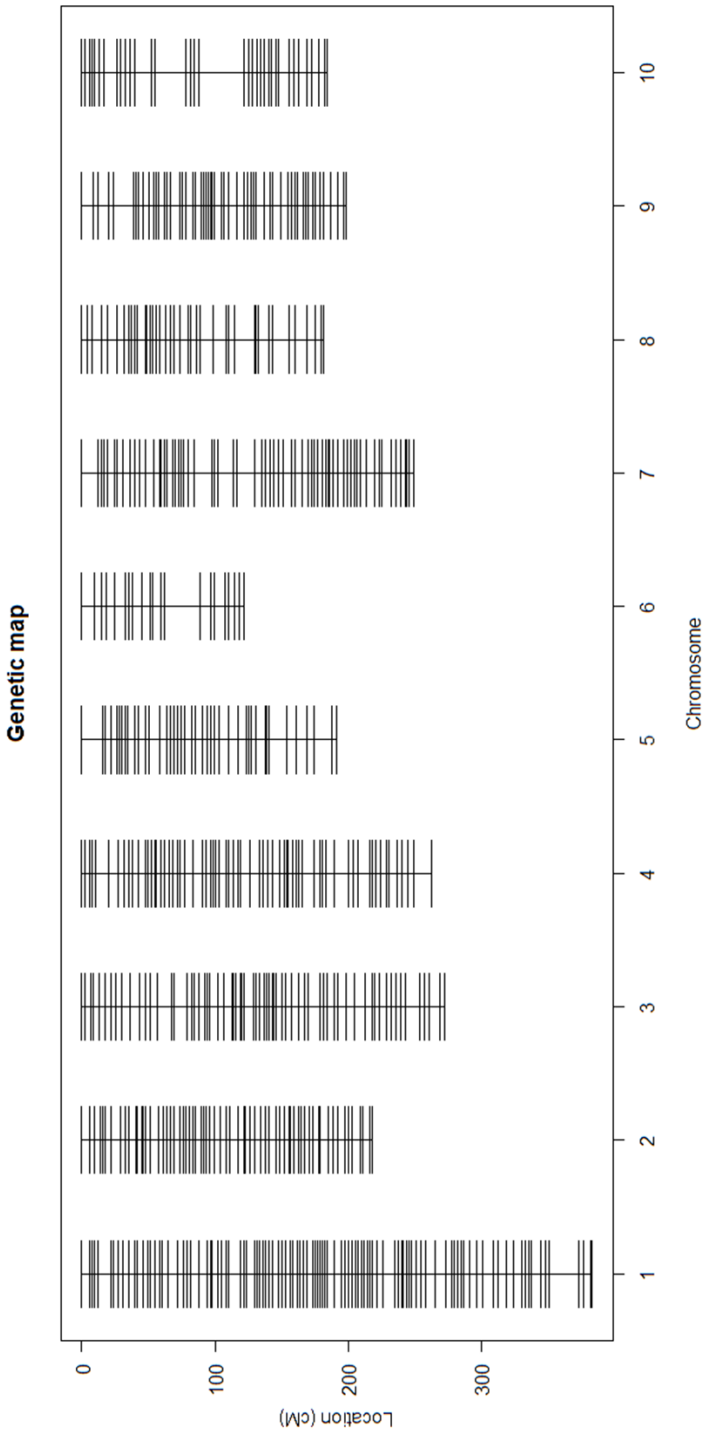

FIGURE 4.26: **PTxB73 BC1S5 Linkage Map.** Linkage map based on analysis of 744 sites in 81 individuals from the B73xPT BC1S5 population. Vertical lines represents each chromosome; marker position (cM) indicated by horizontal lines

4.3.3 QTL mapping

QTL mapping was performed using genetic and phenotypic information from 81 individuals evaluated in highland and lowland field sites. Analysis was performed on data from each location separately, on the mean of the two field sites, and on the difference between the two field sites. QTL identified to show a similar effect in both locations and/or on the basis the mean were considered stable (constitutive). QTL identified in a single site and/or on the basis of the difference between sites were considered to show QTL x environment interaction (conditional). QTL identified in both environments, but showing opposing effects in the different sites were considered antagonistic.

A total of 21 QTL were found associated with 12 of the traits evaluated (Table 4.11). The phenotypic variation explained by each QTL ranged from 0.04% to 24.19%. QTLs with the largest effects (>20%) were detected *qDO1-1* on chromosome 1 for DO, and *qTBN7-1* on chromosome 7 for TBN. The number of QTLs detected per trait ranged from 1-3, jointly explaining 16.35-43.27% of total phenotypic variation (Table 4.10). The largest number of QTLs detected were for inflorescence traits.

TABLE 4.10: Summary of the quantitative trait loci (QTL) identified in the B73xPT BC1S5 population.

Traits	No of QTL	Variation explained by each QTL (%)	Variation explained by all QTL (%)
STD	1	19.76	19.76
PH	1	17.8	17.8
EH	1	19.19	19.19
TL	3	9.95 - 17.48	38.65
TBN	2	19.08 - 24.19	43.27
LM	2	17.24 - 20.59	37.83
DM	2	18.34 - 18.35	36.69
DO	2	13.89 - 24.92	38.81
DTS	1	16.35	16.35
DTA	2	18.37 -22.04	40.41
PM	3	8.58 - 12.12	30.24
PTG	2	12.64 - 13.62	26.26

TABLE 4.11: **QTL linked to STD, PH, EH, TL, TBN, LM, DM, DO, DTS, DTA, PM, PTG in the BC1S5 RIL population.** (∗) Effects calculated as PT - B73. (⋆) Genetic position range for drop one LOD support interval.

Trait	Chr	QTL	Peak position (cM)	Peak psotion (bp)	LOD				Var (%)				Effect size ∗				Genetic interval⋆ (cM)	Physical interval (bp)
					MT	PV	MEAN	DIFFERENCE	MT	PV	MEAN	DIFFERENCE	MT	PV	MEAN	DIFFERENCE		
STD	6	qSTD6-1	114.42	166,674,035	0.55	2.64	0.88	**3.98**	3.1	13.94	4.89	20.25	0.66	-2.48	-0.91	3.15	111.00 - 118.00	166,461,089 - 167,436,381
PH	1	qPH1-1	225	182,355,315	2.81	2.99	**3.54**	0.09	14.77	15.65	18.24	0.53	-20.87	-17.83	-19.35	-3.04	56.00 - 234.00	22,439,341 - 194,993,115
EH	1	qEH1-1	206	161,179,291	2.92	2.94	**3.75**	0.04	15.32	15.41	19.2	0.24	-11.58	-10.34	-10.96	-1.25	200.07 - 226.00	159,922,182 - 182,355,315
TL	1	qTL1-1	258.44	227,682,365	2.37	0.06	0.66	3.11	9.34	0.28	2.63	15.64	-2.91	0.41	-1.24	-3.33	253.00 - 276.00	215,520,063 - 248,308,743
	4	qTL4-2	257	239,423,939	2.23	1.17	2.24	0.27	8.76	5.67	9.38	1.24	-3.35	-2.24	-2.79	-1.11	236.00 - 262.42	237,458,023 - 241,177,024
	9	qTL9-3	132	126,095,248	1.7	1.63	2.22	0.01	6.59	7.96	9.28	0.06	2.81	2.57	2.69	0.25	113.00 - 139.00	86,165,548 - 135,440,886
TBN	7	qTBN7-1	116.67	104,569,849	2.45	6.54	4.89	0.88	12.56	30.01	24.31	3.97	-1.58	-2.31	-1.95	0.72	114.03 - 125.00	87,515,189 - 125,915,539
	7	qTBN7-2	220.22	166,745,705	0.9	0.49	0.05	3.59	4.4	1.88	0.21	17.44	0.99	-0.61	0.19	1.6	195.00 - 225.00	158,082,626 - 168,745,410
LM	8	qLM8-1	87	123,597,178	0.01	4.01	1.09	2.76	0.04	20.6	6.01	14.86	0.07	-1.54	-0.71	1.62	79 - 97	120,248,364 - 149,259,762
DM	1	qDM1-1	125	64,446,634	3.28	0.3	2.77	2.24	16.89	1.42	13.3	12.29	-0.52	-0.09	-0.31	-0.42	121.71 - 156.00	64,129,797 - 82,587,417
	4	qDM4-2	161	179,842,509	0.33	3.32	1.7	0.34	1.55	16.89	7.93	1.77	-0.14	-0.27	-0.21	0.14	158.11 - 172.00	162,825,835 - 191,538,464
DO	1	qDO1-1	144	77,384,773	4.21	6.19	7.71	0.02	21.46	25.19	32.74	0.1	-0.43	-0.43	-0.45	0.03	136.00 - 149.00	72,104,649 - 79,013,374
	6	qDO6-2	108	165,554,795	0.23	3.61	1.51	0.92	1.05	13.47	5.3	5.76	-0.09	-0.31	-0.17	0.2	104.00 - 110.00	162,833,157 - 165,811,056
DTS	8	qDTS8-1	88	127,739,327	3.23	0.47	3.11	2.56	16.78	2.64	16.19	13.53	-7.28	-1.05	-4.17	-6.23	45.00 - 168.00	91,483,033 - 167,192,651
DTA	7	qDTA7-1	189	157,886,703	1.64	0	0.11	1.57	7.39	0	0.64	7.51	-5.16	0.05	-1.83	-4.89	186.00 - 195.00	157,455,779 - 159,904,263
	8	qDTA8-2	91	127,739,327	2.02	0.63	0.6	1.35	9.2	3.52	3.27	6.41	-6.44	-1.56	-4.61	-5.05	84.00 - 166.00	123,404,566 - 167,192,651
PM	1	qPM1-1	289	256,395,222	0.01	2.83	0.65	1.16	0.03	12.12	3.38	5.02	-0.72	-11.54	-5.49	11.03	283.00 - 295.00	250,895,974 - 257,956,805
	8	qPM8-2	87	125,181,626	0.17	2.27	1.1	0.39	0.92	9.54	5.79	1.63	-4.57	-11.4	-8	6.83	83.00 - 96.00	123,404,566 - 123,404,566
	8	qPM8-3	160	164,070,712	1.57	0.02	0.69	1.79	8.58	0.09	3.59	7.92	13.58	-1.09	6.11	14.64	154.00 - 167.00	159,541,552 - 167,192,651
PTG	7	qPTG7-1	150	138,140,415	1.66	0.25	0.1	2.78	8.59	1.4	0.56	12.64	9.4	-3.39	1.86	13.51	141.02 - 159.00	133,886,872 - 139,199,029
	8	qPTG8-2	159	165,779,093	1.25	0.81	0.02	2.96	6.37	4.48	0.11	13.62	8.47	-6.25	0.86	14.46	86.00 - 167.00	123,597,178 - 167,192,651

QTLs linked to plant fitness and phenology

Flowering time regulation is a complex trait, and a key determinant of the adaptation of a plant to its environment. Linkage analyses and Genome Wide Association Studies (GWAS) have shown that flowering time variation in maize involves numerous chromosomal regions. Here, the major QTL for days-to-silking (DTS) and days-to-anthesis (DTA) co-localized to the same region of chromosome 8 (*qDTS8-1; qDTA8-2*) and were detected in both lowland and highland sites. The corresponding physical interval included the well defined flowering candidates *Zea Centroradialis8* (*Zcn8*; 123Mb), a maize *FLOWERING LOCUS T*-like floral activator [44], and *Vegetative to generative transition 1* (*vgt1*; 132Mb), a QTL identified as a *cis*-regulator of the *Rap2* gene, a negative regulator of flowering time [11, 16, 1]. *Vgt1* has a key role in maize alti-latitudinal adaptation, and significantly, has been suggested to play a key role not just in maize adaptation to temperate latitudes but is also related with the differentiation of maize varieties according to elevation in tropical Central America [16]. Flowering time was clearly plastic, plants taking longer to flower in the highlands (Figures 4.15**B** and 4.15**C**). A second QTL for DTA on chromosome 7, was conditional, being detected only in the highlands, the PT allele effect of this QTL was associated with decreasing the number of days to anthesis by 5.16 days and did not co-localize with any previously reported major flowering time QTL.

Yield is the result of many trait interactions during plant development and growth, and has been the main focus of cereal breeding programs [8]. Yield is typically determined on the basis of productivity per unit area. Here, ear weight and total grain weight (PM and PTG, respectively) were used as an indicator of yield, and a measure of overall plant adaptation. Three QTLs were identified for PM, one on chromosome 1 and two on chromosome 8; and two QTLs for PTG on chromosomes 7 and 8. All QTLs detected for yield components showed QEI. For the lowland-conditional PM QTL *qPM1-1* and *qPM8-1*, the PT allele was negative. The third PM QTL *qPM8-2* showed antagonistic pleiotropy, the PT allele increasing PM in the highlands, and decreasing PM in the lowlands. The QTL *qPM8-2* co-localizes with the previously described flowering time QTL *qDTA8-2* and *qDTS8-1*, and the conditional QTL *qPTG8-2*, was detected based on the difference between highlands and lowlands. Both QTL detected for PTG (*qPTG7-1* and *qPTG8-2*) were identified based on the difference between environments. The PT allele had a negative effect in lowlands but positive in highlands. The QTL *qPTG7-1* was located close to the flowering time QTL *qDTA7-1*.

QTL linked to plant morphology

Five QTLs were identified which are associated with the tassel (male inflorescence) architecture (TL: tassel length and TBN: tassel branch number). QTLs for TL were located on chromosomes 1, 4 and 9, and did not overlap with TBN QTLs identified on chromosome 7, suggesting independent regulation of branching and length. Both TL and TBN were plastic with respect to elevation and QEI was observed for the TL QTL *qTL1-1* and the TBN QTL *qTBN7-2*.

In the parents, PT tassels were larger and much less branched than those of B73, a trait characteristic of Mexican highland maize landraces. In both parents, plants in the highlands tended to have smaller tassels than in lowlands. Interestingly, the PT allele at QTL *qTL4-2* was associated with a decrease in tassel length by 3.35 cm in highlands or 2.24 cm in lowlands (Table 4.11). The PT allele at *qTL9-3*, however, showed a stable positive effect, increasing tassel length by 2.69 cm (Table 4.11). The

QTL *qTL1-1* showed antagonistic pleiotropy, with the PT allele increasing TL in the lowlands but reducing it in the highlands. QTL *qTBN7-1* was the major-effect QTL for TBN, and was stable across the environments, explaining an mean of 24.19% of the variance across the two sites, the PT allele decreased the number of tassel branches by 1.95. The second TBN QTL *qTBN7-2*, showed antagonistic pleiotropy, the PT allele decreasing tassel branching in the lowlands, but increasing it in the highlands.

Tassel architecture QTLs co-localized with a number of genes previously reported to control male inflorescence development [76, 77]. The *qTL1-1* interval contains the genes *Tasselless1* (*Tls1*, GRMZM2G171822) and *Indeterminate growth1* (*Id1*, GRMZM2G011357). Loss-of-function mutations of *Tls1* give rise to plants with reduced tassels, while loss-of-function mutations of *Id1* extend tassel growth in association with delayed flowering. The QTL *qTL4-2* co-localizes with *fea2*, a regulator of shoot meristem proliferation in maize [64]. The QTL interval *qTBN7-1* contains the gene *ramosa1* (*ra1*, GRMZM2G003927) a zinc finger transcription factor that affects maize inflorescence branching, and has been associated with reduced TBN and increased TL[77]. The confidence interval for the QTL *qTBN7-2* contains the gene *ramosa3* (*ra3*, GRMZM2G014729) also involved in tassel development. Loss of function *ra3* mutants produce more tassel branches [77].

Five QTLs were identified linked to variation in ear (female inflorescence) architecture, for ear length (LM), ear diameter (DM) and cob diameter (DO). A single QTL *qLM8-1* was associated with LM, located in chromosome 8, showing antagonistic pleiotropy, the PT allele decreasing ear length by 1.54 cm in the lowlands, but increasing ear length in the highlands (Table 4.11). Two stable QTLs were identified for DM, one on chromosome 1 (*qDM1-1*) and the other on chromosome 4 (*qDM4-2*); in both QTLs, the B73 allele showed a positive effect. Two stable QTLs were also identified for cob diameter, one overlapping with the ear diameter QTL *qDM1-1* on chromosome 1, and the second on chromosome 6, *qDO6-2*.

The QTL effect for ear traits correlate with the size of the parental ears, where B73 is bigger than PT, but also the QTLs *qLM8-1* could explain how PT is able to develop a bigger ear under highland conditions (Figure 4.13) since the QEI gives a positive effect due to the PT allele.

It is remarkable that the QTL identified for DM *qDM4-2* overlaps with the previously described inversion polymorphism *inv-4m*, however, the closest gene to this QTL peak is *tunicate1* (*tu1*), associated with glume morphology.

Height adaptations are essential to plant fitness and agricultural performance [47]. Plants with similar PH but lower EH are expected to be more resistant to lodging. EH is generally correlated with PH in maize [12]. In this work, stable PH and EH QTL were located in chromosome 1.

Although more than 280 QTLs controlling plant height have been identified in maize [12] genetic regulation related to the highland/lowland environment has not been studied. The QTL for PH on chromosome 1 covers a broad genetic interval that co-localizes with previous QTLs [20, 12]. Elongation of plant parts is a complex phenomenon mediated by many plant hormones, including auxins, brassinosteroids, and gibberellins (GAs). However, our understanding of how these hormones regulate cell elongation remains limited [46]. Over 40 maize genes involved in hormone synthesis, transport, and signaling have been determined to have large effects on plant height [47]. The closest well defined gene to the peak of the PH QTL *qPH1-1* is *brachytic2 (Br2)* locus (GRMZM2G315375), its mutation results in a reduction in plant height [46]. PH was clearly plastic with respect to the elevation of the field site,

plants in the highland site being generally shorter. However, no QTL x environment interaction (QEI) effects were detected for PH.

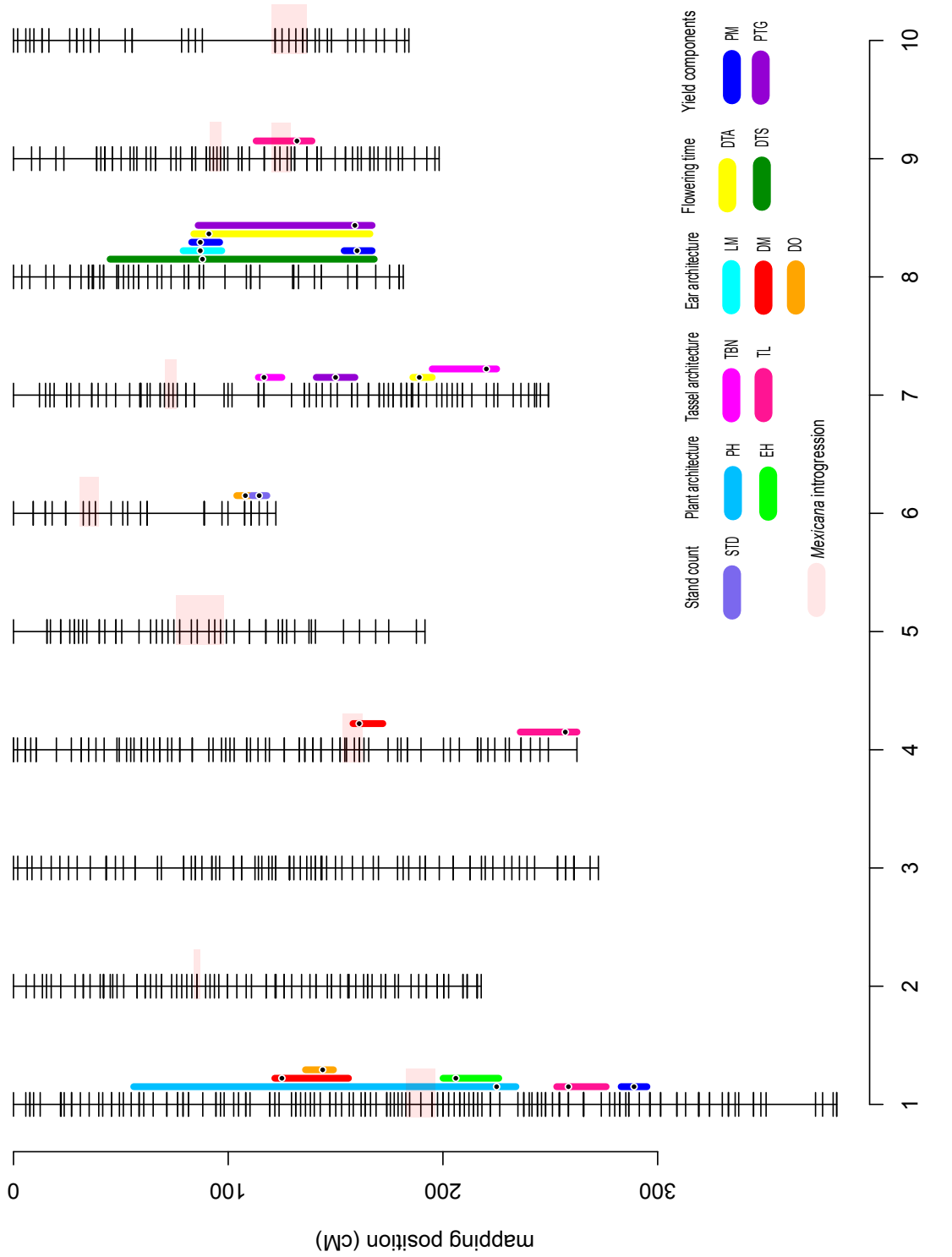

FIGURE 4.27: **QTLs detected in the B73xPT BC1S5 population and their co-localization with previously reported *mexicana* introgression [24].** QTL intervals are shown based on 1-LOD support interval (dropping 1 LOD score from the peak). Traits are indicated by colors; the black dot in the QTL represents the peak and the color is distributed along the confidence interval. *Mexicana* introgressed regions are marked in light red throughout the linkage map to indicate co-localization with the QTL detected

4.4 Co-localization of detected QTL with *mexicana* introgression regions in highland maize

Confidence intervals were defined for all QTLs on the basis of a 1-LOD support interval, meaning that the interval in which the LOD score is determined is within 1 unit from its maximum. Genetic intervals (cM) were then converted to physical intervals (bp) on the basis of the anchored position of the nearest flanking markers to the B73 reference genome. Physical QTL intervals were then compared to a set of previously defined candidate *mexicana* introgression regions, identified to be present at high frequency in Mexican highland maize [24] to determine possible co-localization.

Co-localization of QTL from adaptive traits at *inv4m* was expected due to its putative role in local adaptation [24, 55], however the QTL *qDM4-2* was the only QTL identified inside this region. An overlap on chromosomes 4 and 9 was observed for DM and TL QTL peaks respectively, and the confidence interval for PH QTL on chromosome 1 also overlapped with the *mexicana* introgression interval, however these regions did not co-localize with the QTL peak (Figure 4.27).

4.5 Genotypic value of B73 and Palomero Toluqueño haplotypes at *inv4m*

Further information on the *inv4m* effect was identified based on the differences between and within HIF families. Accordingly with previous reports, *inv4m* has been related to flowering time [55], however no statistical difference was observed between the *inv4m*-B73 or *inv4m*-PT region for flowering time in this study (Figures 4.16**B-C**, Table 4.8). Reaction norms for a number of traits such as PTG, PH, EH, TL TBN, LM, DM, LMDM, and NHG (Figures 4.16**D**, 4.18**A-D** and 4.20**A-D**), showed significant differences between inverted and non inverted haplotypes and interestingly all the mentioned features are related to male and female inflorescence architecture.

To investigate whether the differences between HIF families were related with differences between the genotype at the QTL position for each trait, PT introgression was interrogated at each QTL position in both HIF families: 56 and 134 (Table 4.12). QTL physical position was co-localized with GBS data in the two families, however there were regions where the genotype was not determined due to recombination detected between markers surrounding the same position and these were defined as ND (non determinate regions) in addition heterozygous regions for some QTL (SEG) were also identified.

Most of the analyzed traits carried the same allele at the QTL for both families. Differences at QTL regions were clear for STD and DO; in both cases HIF 56 carries the B73 allele at the QTL region and HIF 134 the PT allele. For STD, HIF 134 was more successful in both environments and the difference between families was significant (Table 4.8), however, QTL for the PT allele showed a negative effect on size in the lowlands, but the *inv4m* effect was not statistically different between both families (Figure 4.16, Table 4.8). With respect to DO, the PT allele at the QTL produced a negative effect that was also observed due to *inv4m* in the reaction norms at both locations. *inv4m* and families were statistically different (Table 4.8) and HIF 134 showed a smaller diameter confirming the presence of the PT allele at the QLT region. Another interesting case was PTG, where the B73 allele was present in both families, but HIF 134 was segregating within the inversion at the QTL region and contrary to family 56, showed a positive effect for the inv4m in family 134.

TABLE 4.12: **Genotype of the HIF families 56 and 134 at the QTL regions for STD, PH, EH, TL, TBN, LM, DM, DO, DTS, DTA, PM, and PTG.** ND: Not determined genotypes; SEG: heterozygous regions detected.

Trait	Chr	QTL	Peak position (bp)	HIF GENOTYPE	
				Fam 56 inv4m + / -	Fam 134 inv4m + / -
STD	6	*qSTD6-1*	166,674,035	B73/B73	PT/PT
PH	1	*qPH1-1*	182,355,315	ND	ND
EH	1	*qEH1-1*	161,179,291	ND	ND
TL	1	*qTL1-1*	227,682,365	B73/B73	B73/B73
	4	*qTL4-2*	239,423,939	SEG	SEG
	9	*qTL9-3*	126,095,248	B73/B73	B73/B73
TBN	7	*qTBN7-1*	104,569,849	ND	ND
	7	*qTBN7-2*	166,745,705	SEG/PT	PT/SEG
LM	8	*qLM8-1*	123,597,178	B73/B73	B73/B73
DM	1	*qDM1-1*	64,446,634	SEG	ND
	4	*qDM4-2*	179,842,509	PT/B73	PT/B73
DO	1	*qDO1-1*	77,384,773	B73/B73	PT/PT
	6	*qDO6-2*	165,554,795	B73/B73	PT/PT
DTS	8	*qDTS8-1*	127,739,327	B73/B73	B73/B73
DTA	7	*qDTA7-1*	157,586,703	ND	ND
	8	*qDTA8-2*	127,739,327	B73/B73	B73/B73
PM	1	*qPM1-1*	256,395,222	B73/B73	B73/B73
	8	*qPM8-2*	125,181,626	B73/B73	B73/B73
	8	*qPM8-3*	164,070,712	B73/B73	HET/B73
PTG	7	*qPTG7-1*	138,140,415	ND	ND
	8	*qPTG8-2*	165,779,093	B73/B73	HET/B73

To clarify the *inv4m* effect on the detected QTL, adjustment was made to an interaction model of QTL fitting *inv4m* as a QTL (covariate?). RIL lines were split based on their genotype at the trait QTL and at *inv4m*

The low number of rows per ear has been reported previously and fewer rows are related to higher branching, this effect could be confounded by demography and local adaptation. Additionally QTL detected in the HIF families and their genotype at the QTL was investigated in order to identify differences due to either the QTL or inv4m. Table

to see if the inv4m was pleiotropic with an specific QTL, it was fixed as QTL

RIL lines were splited

Chapter 5

Discussion

5.1 Summary

Low temperatures, higher solar radiation and reduced atmospheric pressure have conditioned maize dispersion through the Highlands in Central Mexico. In this scenario, gene flow from teosinte *mexicana* to cultivated maize has been proposed as a source of useful genetic variation, facilitating maize highland adaptation. Previous population genetic analyses have identified the large-scale inversion polymorphism *inv4m* to be derived from *mexicana*, and to be present at high frequency in Mexican highland maize.

In this work, an allele specific expression (ASE) study, comparing the Mexican highland maize landrace Palomero Toluqueño (PT) with the lowland inbred B73, identified 10% of the genes located in *inv4m* to be differentially expressed between the PT and B73 allele. Field evaluation of BC4S5 Near Isogenic Lines (NILs) carrying the PT allele of *inv4m* in a B73 background, however, did not identify clear morphological or performance differences with respect to the B73 recurrent parent. Nonetheless, quantitative trait loci (QTL) mapping in a BC1S5 B73 x PT recombinant inbred line (RIL) population linked mild effects on a number of morphological traits to *inv4m*. A second previously-reported region of *mexicana* introgression was also linked to variation in inflorescence morphology. Subsequent evaluation of NILs derived from two separate Heterologous Inbred Families (HIFs) that segregated for *inv4m* confirmed effects of *inv4m* on both female and male inflorescence architecture and total grain weight. Importantly, the *inv4m* effect was modified in the different genetic backgrounds of the two HIFs, indicative of epistatic interactions between *inv4m* and other loci, and consistent with the absence of effect in the BC4S5 NIL.

5.2 The genetic architecture of local adaptation in Mexican highland maize

The strict criterion for local adaptation is that a population must have higher fitness at its native site than any other population introduced to that site [58]. B73 was observed to perform better than PT in the lowlands and PT to perform better than B73 in the highlands (Figure 4.13). Most of the observed differences in plant performance across sites were related with seedling emergence, male and female architecture and time to flowering in the two field sites.

The putatively adaptive traits sheath pubescence and anthocyanin pigmentation are shared between Mexican highland maize and *mexicana*, and have been attributed to *mexicana* introgression into maize. Sheath pubescence results in the formation of

a boundary layer that influences how quickly gases and energy are exchanged between the stem and the surrounding air, potentially maintaining temperature under cold conditions and enhancing frost tolerance. Anthocyanin pigmentation can provide protection from the high levels of solar radiation experienced by plants in the highlands [25, 43]. Previous mapping of the genetic architecture of sheath pubescence and pigmentation in a cross between *parviglumis* and *mexicana* [32] identified a number of QTL, six of which overlap with regions of introgression from *mexicana* to Mexican highland maize as reported by [22]. In my work, I identified QTL for pigments on chromosome 2 and macrohairs on chromosomes 7 and 9. Loci identified on chromosomes 2 and 9 co-localize with previous reports [32], the region on 9 co-localizing with a reported region of introgression. The locus on chromosome 7 was not previously identified in the teosinte cross [32], nor does the region overlap with a a reported region of *mexicana* introgression [22]. The QTL for stem pigments on chromosome 2 was the best supported in my study, and co-localizes with the *b1* locus that has been well characterized to regulate stem pigmentation in maize [59]. There is no evidence for *mexicana* introgression at *b1*. Multiple *B1* alleles have been described to condition stem pigmentation [59]. Different *B1* alleles are based on different variants, typically transposon insertion into the regulatory region [59]. In the case of the alleles *B1-Peru* and *B1-Bolivia* the result is increased pigmentation not in the stem, but in the kernel aluerone, a tissue where anthocyanin pigmentation is normally regulated by the *b1* paralog *r1*. Gain-of-function alleles of *b1* appear to arise readily, and multiple distinct, presumably independently derived, alleles can be identified even within maize varieties of similar origin [59]. I would suggest that given selective pressure for increased stem pigmentation, the required genetic variation can arise *de novo* at sufficient frequency to not require introgression of exisiting *B1* alleles from *mexicana*.

Previous evaluation of highland and lowland maize cultivars, grown in highland, lowland and mid-altitude sites, has clearly shown a dramatic cultivar/environment interaction for flowering time and yield components [25]. Highland cultivars were generally characterized by rapid emergence from cold soils, earliness to flower, relatively fewer leaves, few tassel branches, a lower leaf area index, and shorter plant height relative to their lowland counterparts [25].

Flowering time was clearly environment sensitive, although PT flowers earlier than B73 in both environments. Mean duration from sowing to flowering was extended from 60 days in the warmest site (lowlands) to 130 days at the coolest location (highlands), reflecting the environmental effect imposed over both genotypes. An adaptation to high altitude environment is a reduction in plant size and height, which minimizes wind chill, abrasion and desiccation [38]. PT was observed to be shorter than B73 in the highlands, even in the lowlands, besides both genotypes were shorter in the highlands. Stand count and the yield component total grain weight showed stronger GEI effects in response to the different environments. PT had higher grain weight and higher plant density than B73 specifically in the highlands leading to a rank-change from lowlands as evidence of local adaptation to their respective niches.

Highland maize PT tassel is longer and less branched than B73, however the highlands environment propitiates more branching and shorter tassels, these phenotypes were clearly plastic although the slope was not significantly modified and no rank-change was observed between PT and B73. Ear traits were more severely affected due to environmental conditions, B73 ears were shorter and thinner than PT ears in the highlands. The opposite effect was observed in the lowlands indicating a clear GEI effect, although PT ears were shown to be more stable across the

environments.

Twenty-one genomic regions involved in 12 of the phenotypes measured were identified. QTL explained from 0.04 to 30.01 of the total phenotypic variation depending on the basis on which they were considered. Eleven QTL were constitutive for 8 of the traits: PH, EH, TL, TBN, DM, DO, DTS and PM, however although the target of breeding is constitutive QTL, environment-specific QTL appear to be very important in highland adaptation. Six QTL with significant QEI were detected for STD, TL, TBN, LM, DTA, PM and PTG. The most critical QEI that showed opposite phenotypic effects in different environments (antagonistic pleiotropy) was that for ear weight. QTL for PM (*qPM8-3*) showed a positive effect of 13.58 in highlands and minus 1.09 in lowlands, whilst PTG QTL (*qPTG8-2*) effect was 8.47 in highlands and -6.25 in lowlands for PT allele. Of the environment specific QTL, the PT alleles showed a positive effect under highland conditions for STD, TBN, LM, PM and PTG traits, whereas PT alleles for TL and DTA showed a negative effect under highland conditions. Only one QTL identified at the *inv4m* was constitutive for DM, the effect size for the PT allele of this QTL was -0.27 and -0.14 in lowlands and highlands respectively.

Rapid evolution in domesticated plants has been shaped by agricultural practices, accompanied by selection for agronomic performance and adaptation to novel environments [71]. Fine changes in regulatory regions controlled by *Cis* regulatory elements (CREs) are also thought to be important in evolution because it is assumed they show less pleiotropy and consequently a lower risk of effects caused by deleterious alleles [75, 33]. There are many examples of modified CREs that underly trait divergence between crops and their ancestors conferring positive or negative gene regulation, tissue specific expression patterns and/or shifts in the expression profile [33]. PT and B73 differ in 26 percent of CREs at the whole genome level, although *cis*-regulation was identified in four of the nine *mexicana* introgression regions reported, no enrichment was associated with these regions. At the chromosomal inversion *inv4m* 10% of the genes were identified as differentially *cis*-regulated for either the PT or B73 alleles. These genes showed a bias towards PT-down regulated alleles (21 PT-down and 7 PT-up) reflecting the bias observed towards the PT-down ASE genes identified, but contrary to the PSR tendency (181 of 277 PSR genes identified presented higher expression of the PT allele).

The ASE gene list at *inv4m* includes the PT-up gene chaperonin10 (*cpn10*, GRMZM2G091189), a key cellular component in numerous folding pathways leading to biologically active proteins (Viitanen 1995, Levy-Rimler 2002). Chaperonin genes have been also involved in stress response. The genes GRMXM2G019177 (PT-up) and GRMZM2G478709 (PT-down), orthologs of *Arabidopsis thaliana* genes *ASI4* and AT177180 respectively are relevant because the stress imposed by the highland environment relates to cold and UV, but also to low fertilizer input cropping systems [5]. *ASI4* is inducible by As and has been implicated in arsenic detoxification whilst the transcription factor AT177180 is involved in response to abscisic acid, salt and osmotic stress.

Limitations in the number of QTLs identified in this study were derived from resolution of detection, population size and genetic diversity in addition to the inaccuracy of the map positions since biparental populations have the disadvantage that they only allow the analysis of genetic variation present between the two parents. Larger population sizes and allelic diversity would provide a greater number of recombinants improving the resolution of detection. In terms of complementation, Genome Wide Association studies (GWAS) is a powerful approach developed to investigate allelic associations with a given trait in question, however a number of

factors limit their successful use. One is the population structure which may result in false positive associations. The advantage of QTL analysis is that we do not have population structure and effects from particular loci are due to the trait in question. QTL detection power can identify spurious GWAS associations, and their unique genomic structure is superior for investigating the role of low-frequency alleles. Future QTL studies will therefore benefit from a combined approach of GWAS and classical linkage analysis.

5.3 The mechanistic basis of allelic variation in highland adaptation

The conditional QTL identified highlight the importance of the environmental interaction. Yield components were one of the most susceptible traits depending heavily on the environment. Flowering time is a complex trait that determines adaptation, this trait is strongly dependent on many environmental factors, including photoperiod and temperature, however although *inv4m* has previously been strongly related to flowering time in maize landraces[55, 71], the flowering QTL identified in this study did not lie within this region. QTL identified for flowering overlapped with the major QTL previously reported for flowering time including the genes *Zea Centroradialis8*, *FLOWERING LOCUS T*-like and *Vegetative to generative transition 1* clearly associated with maize alti-latitudinal adaptation. The PT allele for these QTL decrease flowering time at least by six days. Moreover, the effect of the conditional QTL on chromosome 7 under highland conditions leads to a further reduction of at least 5 days.

A major QTL was detected for TBN, this QTL includes the *Tasselless1*, *ramosa1* and *ramosa3* genes, key regulators of branch number. Mutations with loss of function of *Tls1* give plants with reduced tassels whereas mutations in the *ra3* gene produce more tassel branches. This is consistent with the effect of the PT allele under highland conditions suggesting that allele variation clearly modifies the effect, in agreement with reports suggesting that *mexicana* introgresion following maize dispersion provided the allelic variants to allow adaptation to the completely different environment of the highlands [24, 55, 71]. Analysis of candidate genes corroborated expression differences between PT and B73 alleles, but *trans*-acting effects were also identified.

5.4 The role of *mexicana* introgression

Since its domestication in southwest Mexico from teosinte *parviglumis*, maize experienced a genetic bottleneck resulting in a loss of diversity. Following domestication gene flow into maize from multiple teosinte species has been documented in geographical regions outside this center of origin. Gene flow from *mexicana* into maize has potentially played a role in adaptation to this niche. Investigation across these adaptive regions points to *inv4m* as the main *mexicana* contribution to modern maize lines linked to flowering time in landraces [24, 71].

We dissected the genetic architecture for adaptive, morphological and phenological traits by QTL analysis and investigated regulatory variation between highland and lowland maize through allele specific expression analysis. Genomic regions underlying highland adaptation were identified in six of the ten chromosomes and

regulatory variation was found to be dispersed throughout the whole genome. Co-localization with four of the nine reported regions of *mexicana* introgression was determined [24] with strong confidence intervals for the QTLs detected.

We did not identify effects of *inv4m "per se"*, however allelic effects could remain undetectable, either because they have no effect at phenotypic level or because their effects are only visible under specific, genetic or environmental conditions, generally referred as 'cryptic' or 'hidden' genetic variation [56]. Analysis of HIF families lead us to determine that the *inv4m* effect depends on the genetic background. Ear and tassel morphology were the phenotypes that clearly responded in differentially depending on different backgrounds suggesting limitations based on physiological, developmental or fitness constraints.

Chapter 6

Conclusions

PT mexi 5 carries gene content from teosinte *mexicana* and harbors *inv4m*. It is locally adapted to the Mexican highlands Whereas B73 is adapted to mid-low elevations and their phenotypic differences provide tools to examine the genetic architecture of highland adaptation.

The differences in gene expression between PT and B73 under benign conditions provide insights into how PT can respond and cope with the challenges of the highland environment.

There are a range of environmental effects influencing the parental phenotypes. The behavior of the RIL population showing different reaction norms across the environments can provide insights to the genomic regions involved in local adaptation

There are QTLs showing parental phenotypes whose effect depends on the environment. On the other hand, We have also detected QTLs for adaptive traits (phenology and yield components) that show environmental effects.

PT alleles affect yield and flowering under highland environments. The yield QTLs detected are located mainly on chromosome 8, which has not previously been shown to have undergone introgression. New variation selected

Their is no colocalization of the QTLs with introgressed regions from *mexicana*, this is particularly the case with yield traits that we expect would be more related with adaptation, however phenotypes associated with macrohairs do colocalize with previous QTL reports and introgressed regions.

PTG and TBN were affected in different way by the *inv4m* presence, suggesting epistatic interactions with other genomic regions

Chapter 7

Perspectives

The identification genomic regions involved in local adaptation will serve as landmark to explore the mexican landrace diversity panel as source for breeding programs contributing to the efforts enhancing adaptation providing new strategies to marginal lands to set up the basis for further collaborations generating knowledge through scientific papers and providing strategies for a better management and exploitation of natural resources in crops.

Given the differences in response to environmental stress observed between PT and B73 alleles under benign conditions, apply the stress conditions can provide us insights in the deferentially regulation lend the adapted crops face with environmental challenges

The development of larger and more diverse mapping populations in order to improve the detection resolution and corroboration of key genes involved in adaptation to the highlands. Genome edition and/or mutant generation of ear and tassel morphology genes for adaptation and their analysis under different backgrounds and environments in order to validate their function in highland adaptation

Bibliography

[1] Philipp Alter et al. "Flowering Time Regulated Genes in Maize Include the transcription factor ZmMADS1." In: *Plant physiology* 172.1 (2016), pp. 389–404. ISSN: 1532-2548. DOI: 10.1104/pp.16.00285. URL: http://www.plantphysiol.org/content/172/1/389.abstract.

[2] Konieczny Andrzej and Frederick Ausubel. *Konieczny and Ausubel 1993.pdf*. 1993.

[3] Michael L Arnold. "Transfer and origin of adaptations through natural hybridization: were Anderson and Stebbins right?" In: *The Plant cell* 16.3 (2004), pp. 562–570. ISSN: 1040-4651. DOI: 10.1105/tpc.HistPersp.

[4] Diego Ayala, Anna Ullastres, and Josefa González. "Adaptation through chromosomal inversions in Anopheles". In: *Frontiers in Genetics* 5.MAY (2014), pp. 1–10. ISSN: 16648021. DOI: 10.3389/fgene.2014.00129.

[5] Jeannette S. Bayuelo-Jiménez and Iván Ochoa-Cadavid. "Phosphorus acquisition and internal utilization efficiency among maize landraces from the central Mexican highlands". In: *Field Crops Research* 156 (2014), pp. 123–134. ISSN: 03784290. DOI: 10.1016/j.fcr.2013.11.005. URL: http://dx.doi.org/10.1016/j.fcr.2013.11.005.

[6] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. "Trimmomatic: A flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15 (2014), pp. 2114–2120. ISSN: 14602059. DOI: 10.1093/bioinformatics/btu170.

[7] N.Manikanda Boopathi. *Genetic mapping and marker assisted selection*. 2013, p. 304. ISBN: 9788132209577. DOI: 10.1007/978-81-322-0958-4.

[8] S Bouchet et al. "Association mapping for phenology and plant architecture in maize shows higher power for developmental traits compared with growth influenced traits". In: *Heredity* 118.3 (2017), pp. 249–259. ISSN: 0018-067X. DOI: 10.1038/hdy.2016.88. URL: http://www.nature.com/doifinder/10.1038/hdy.2016.88.

[9] Peter J. Bradbury et al. "TASSEL: Software for association mapping of complex traits in diverse samples". In: *Bioinformatics* 23.19 (2007), pp. 2633–2635. ISSN: 13674803. DOI: 10.1093/bioinformatics/btm308.

[10] Karl W. Broman et al. "R/qtl: QTL mapping in experimental crosses". In: *Bioinformatics* 19.7 (2003), pp. 889–890. ISSN: 13674803. DOI: 10.1093/bioinformatics/btg112.

[11] E. S. Buckler et al. "The Genetic Architecture of Maize Flowering Time". In: *Science* 325.5941 (2009), pp. 714–718. ISSN: 0036-8075. DOI: 10.1126/science.1174276. arXiv: ISSN:2278-8867. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1174276.

[12] Hongguang Cai et al. "Identification of QTLs for plant height, ear height and grain yield in maize (Zea mays L.) in response to nitrogen and phosphorus supply". In: *Plant Breeding* 131.4 (2012), pp. 502–510. ISSN: 01799541. DOI: 10.1111/j.1439-0523.2012.01963.x.

[13] P. Catarecha et al. "A Mutant of the Arabidopsis Phosphate Transporter PHT1;1 Displays Enhanced Arsenic Accumulation". In: *the Plant Cell Online* 19.3 (2007), pp. 1123–1133. ISSN: 1040-4651. DOI: 10.1105/tpc.106.041871. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.106.041871.

[14] Tomasz Czechowski et al. "Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: Unprecedented sensitivity reveals novel root-and shoot-specific genes". In: *Plant Journal* 38.2 (2004), pp. 366–379. ISSN: 09607412. DOI: 10.1111/j.1365-313X.2004.02051.x.

[15] Stéphane Deschamps, Victor Llaca, and Gregory D. May. "Genotyping-by-Sequencing in Plants". In: *Biology* 1.3 (2012), pp. 460–483. ISSN: 2079-7737. DOI: 10.3390/biology1030460. URL: http://www.mdpi.com/2079-7737/1/3/460/.

[16] Sébastien Ducrocq et al. "Key impact of Vgt1 on flowering time adaptation in maize: Evidence from association mapping and ecogeographical information". In: *Genetics* 178.4 (2008), pp. 2433–2437. ISSN: 00166731. DOI: 10.1534/genetics.107.084830.

[17] Robert J. Elshire et al. "A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species". In: *PLoS ONE* 6.5 (2011), pp. 1–10. ISSN: 19326203. DOI: 10.1371/journal.pone.0019379. arXiv: NIHMS150003.

[18] A. Fournier-Level et al. "A Map of Local Adaptation in Arabidopsis thaliana". In: *Science* 334.6052 (2011), pp. 86–89. ISSN: 0036-8075. DOI: 10.1126/science.1209271. arXiv: arXiv:1011.1669v3.

[19] Jeffrey C. Glaubitz et al. "TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline". In: *PLoS ONE* 9.2 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0090346. arXiv: NIHMS150003.

[20] P Gustafson et al. "The genetic relationship among plant-height traits found using multiple-trait QTL mapping of a dent corn and popcorn cross". In: *Genome* 50.4 (2007), pp. 357–364. ISSN: 0831-2796. DOI: 10.1139/G07-018.

[21] J. van Heerwaarden et al. "Genetic signals of origin, spread, and introgression in a large sample of maize landraces". In: *Proceedings of the National Academy of Sciences* 108.3 (2011), pp. 1088–1092. ISSN: 0027-8424. DOI: 10.1073/pnas.1013011108. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1013011108.

[22] Matthew B Hufford, Paul Bilinski, and Tanja Pyha. "Teosinte as a model system for population and ecological genomics". In: 28.12 (2012). DOI: 10.1016/j.tig.2012.08.004.

[23] Matthew B. Hufford et al. "Inferences from the Historical Distribution of Wild and Domesticated Maize Provide Ecological and Evolutionary Insight". In: *PLoS ONE* 7.11 (2012). ISSN: 19326203. DOI: 10.1371/journal.pone.0047659.

[24] Matthew B. Hufford et al. "The Genomic Signature of Crop-Wild Introgression in Maize". In: *PLoS Genetics* 9.5 (2013). ISSN: 15537390. DOI: 10.1371/journal.pgen.1003477. arXiv: 1208.3894 [q-bio.PE].

[25] J C Jiang et al. "Genetic analysis of adaptation differences between highland and lowland tropical maize using molecular markers." In: *Theoretical and Applied Genetics* 99.(7-8) (1999), pp. 1106–1119.

[26] Jinpu Jin et al. "PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants". In: *Nucleic Acids Research* 45.D1 (2017), pp. D1040–D1045. ISSN: 13624962. DOI: 10.1093/nar/gkw982.

[27] Matthew Kearse et al. "Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data". In: *Bioinformatics* 28.12 (2012), pp. 1647–1649. ISSN: 13674803. DOI: 10.1093/bioinformatics/bts199.

[28] Joost J B Keurentjes et al. "Development of a near-isogenic line population of Arabidopsis thaliana and comparison of mapping power with a recombinant inbred line population". In: *Genetics* 175.2 (2007), pp. 891–905. ISSN: 00166731. DOI: 10.1534/genetics.106.066423.

[29] Mark Kirkpatrick. "How and why chromosome inversions evolve". In: *PLoS Biology* 8.9 (2010). ISSN: 15449173. DOI: 10.1371/journal.pbio.1000501.

[30] Mark Kirkpatrick and Nick Barton. "Chromosome inversions, local adaptation and speciation". In: *Genetics* 173.1 (2006), pp. 419–434. ISSN: 00166731. DOI: 10.1534/genetics.105.047985.

[31] Ben Langmead and Steven L Salzberg. "Fast gapped-read alignment with Bowtie 2". In: 9.4 (2012), pp. 357–360. DOI: 10.1038/nmeth.1923.

[32] Nick Lauter et al. "The inheritance and evolution of leaf pigmentation and pubescence in teosinte". In: *Genetics* 167.4 (2004), pp. 1949–1959. ISSN: 00166731. DOI: 10.1534/genetics.104.026997.

[33] Zachary H. Lemmon et al. "The Role of cis Regulatory Evolution in Maize Domestication". In: *PLoS Genetics* 10.11 (2014). ISSN: 15537404. DOI: 10.1371/journal.pgen.1004745.

[34] Heng Li. "A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data". In: *Bioinformatics* 27.21 (2011), pp. 2987–2993. ISSN: 13674803. DOI: 10.1093/bioinformatics/btr509. arXiv: 1203.6372.

[35] Heng Li. "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM". In: *arXiv preprint arXiv* 00.00 (2013), p. 3. ISSN: 2169-8287. DOI: arXiv:1303.3997[q-bio.GN]. arXiv: 1303.3997. URL: http://arxiv.org/abs/1303.3997.

[36] Heng Li. "Improving SNP discovery by base alignment quality". In: *Bioinformatics* 27.8 (2011), pp. 1157–1158. ISSN: 13674803. DOI: 10.1093/bioinformatics/btr076.

[37] Heng Li et al. "The Sequence Alignment/Map format and SAMtools". In: *Bioinformatics* 25.16 (2009), pp. 2078–2079. ISSN: 13674803. DOI: 10.1093/bioinformatics/btp352. arXiv: 1006.1266v2.

[38] Lorna Little et al. "Leaf and floral heating in cold climates: Do sub-Antarctic megaherbs resemble tropical alpine giants?" In: *Polar Research* 35.2016 (2016). ISSN: 17518369. DOI: 10.3402/polar.v35.26030.

[39] David B. Lowry and John H. Willis. "A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation". In: *PLoS Biology* 8.9 (2010). ISSN: 15449173. DOI: 10.1371/journal.pbio.1000500.

[40] Steven Maere, Karel Heymans, and Martin Kuiper. "BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks". In: *Bioinformatics* 21.16 (2005), pp. 3448–3449. ISSN: 13674803. DOI: 10.1093/bioinformatics/bti551.

[41] Irina Makarevitch et al. "Transposable Elements Contribute to Activation of Maize Genes in Response to Abiotic Stress". In: *PLoS Genetics* 11.1 (2015). ISSN: 15537404. DOI: 10.1371/journal.pgen.1004915. arXiv: 1512.00567.

[42] Yoshihiro Matsuoka et al. "A single domestication for maize shown by multilocus microsatellite genotyping." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.9 (2002), pp. 6080–6084. ISSN: 00278424. DOI: 10.1073/pnas.052125199.

[43] Frederick Meinzer and Guillermo Goldstein. "Some consequences of leaf pubescence in the Andean giant rosette plant Espeletia timotensis". In: *Ecology* 66.2 (1985), pp. 512–520. ISSN: 00129658. DOI: 10.2307/1940399.

[44] X. Meng, M. G. Muszynski, and O. N. Danilevskaya. "The FT-Like ZCN8 Gene Functions as a Floral Activator and Is Involved in Photoperiod Sensitivity in Maize". In: *The Plant Cell* 23.3 (2011), pp. 942–960. ISSN: 1040-4651. DOI: 10.1105/tpc.110.081406. URL: http://www.plantcell.org/cgi/doi/10.1105/tpc.110.081406.

[45] C. Mir et al. "Out of America: Tracing the genetic footprints of the global diffusion of maize". In: *Theoretical and Applied Genetics* 126.11 (2013), pp. 2671–2682. ISSN: 00405752. DOI: 10.1007/s00122-013-2164-z.

[46] D. S. Multani. "Loss of an MDR Transporter in Compact Stalks of Maize br2 and Sorghum dw3 Mutants". In: *Science* 302.5642 (2003), pp. 81–84. ISSN: 0036-8075. DOI: 10.1126/science.1086072. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1086072.

[47] Jason A. Peiffer et al. "The genetic architecture of maize height". In: *Genetics* 196.4 (2014), pp. 1337–1356. ISSN: 19432631. DOI: 10.1534/genetics.113.159152. arXiv: ISSN:2278-8867.

[48] Hugo Perales R., S. B. Brush, and C. O. Qualset. "Landraces of Maize in Central Mexico: An Altitudinal Transect". In: *Economic Botany* 57 (2003), pp. 7–20. ISSN: 0013-0001. DOI: 10.1663/0013-0001(2003)057[0007:LOMICM]2.0.CO;2.

[49] Daniel A Pollard. "Quantitative Trait Loci (QTL)". In: 871 (2012), pp. 31–39. DOI: 10.1007/978-1-61779-785-9. URL: http://link.springer.com/10.1007/978-1-61779-785-9.

[50] B. M. Prasanna. "Diversity in global maize germplasm: Characterization and utilization". In: *Journal of Biosciences* 37.5 (2012), pp. 843–855. ISSN: 02505991. DOI: 10.1007/s12038-012-9227-1.

[51] Marta Puig et al. "Human inversions and their functional consequences". In: *Briefings in Functional Genomics* 14.5 (2015), pp. 369–379. ISSN: 20412657. DOI: 10.1093/bfgp/elv020.

[52] Tanja Pyhäjärvi et al. "Complex patterns of local adaptation in teosinte". In: *Genome Biology and Evolution* 5.9 (2013), pp. 1594–1609. ISSN: 17596653. DOI: 10.1093/gbe/evt109. arXiv: 1208.0634.

[53] J. C. Reif et al. "Grouping of accessions of Mexican races of maize revisited with SSR markers". In: *Theoretical and Applied Genetics* 113.2 (2006), pp. 177–185. ISSN: 00405752. DOI: 10.1007/s00122-006-0283-5.

[54] Adam Roberts and Lior Pachter. "Streaming fragment assignment for real-time analysis of sequencing experiments". In: 10.1 (2013). DOI: 10.1038/NMETH.2251.

[55] J Alberto Romero Navarro et al. "A study of allelic diversity underlying flowering-time adaptation in maize landraces". In: *Nature Genetics* 49.3 (2017), pp. 476–480. ISSN: 1061-4036. DOI: 10.1038/ng.3784. URL: http://www.nature.com/doifinder/10.1038/ng.3784.

[56] Arnaud Le Rouzic. "Evolutionary potential of hidden genetic variation". In: (2007), pp. 33–37. DOI: 10.1016/j.tree.2007.09.014.

[57] José Ariel Ruiz Corral et al. *Ecología, adaptación y distribución actual y potencial de las razas mexicanas de maíz*. 2013, p. 155. ISBN: 9786073701877.

[58] Outi Savolainen, Martin Lascoux, and Juha Merilä. "Ecological genomics of local adaptation". In: *Nature Reviews Genetics* 14.11 (2013), pp. 807–820. ISSN: 1471-0056. DOI: 10.1038/nrg3522. URL: http://www.nature.com/doifinder/10.1038/nrg3522.

[59] D a Selinger and V L Chandler. "B-Bolivia, an allele of the maize b1 gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream." In: *Plant physiology* 125.3 (2001), pp. 1363–1379. ISSN: 0032-0889. DOI: 10.1104/pp.125.3.1363.

[60] K Semagn, Å Bjørnstad, and M N Ndjiondjop. "Principles , requirements and prospects of genetic mapping in plants". In: *African Journal of Biotechnology* 5(25).December (2006), pp. 2569–2587. ISSN: 16845315. DOI: 10.5897/AJB2006.000-5112. arXiv: 1684-5315.

[61] Paul Shannon et al. "Cytoscape: a software environment for integrated models of biomolecular interaction networks". In: *Genome Research* 13 (2003), pp. 2498–2504. ISSN: 1088-9051. DOI: 10.1101/gr.1239303.metabolite.

[62] Mohamed El-Soda et al. "Genotype ?? environment interaction QTL mapping in plants: Lessons from Arabidopsis". In: *Trends in Plant Science* 19.6 (2014), pp. 390–398. ISSN: 13601385. DOI: 10.1016/j.tplants.2014.01.001. URL: http://dx.doi.org/10.1016/j.tplants.2014.01.001.

[63] Chengfu Su et al. "High Density Linkage Map Construction and Mapping of Yield Trait QTLs in Maize (Zea mays) Using the Genotyping-by-Sequencing (GBS) Technology". In: *Frontiers in Plant Science* 8.May (2017), pp. 1–14. ISSN: 1664-462X. DOI: 10.3389/fpls.2017.00706. URL: http://journal.frontiersin.org/article/10.3389/fpls.2017.00706/full.

[64] F. Taguchi-Shiobara et al. "The fasciated ear2 gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize". In: *Genes and Development* 15.20 (2001), pp. 2755–2766. ISSN: 08909369. DOI: 10.1101/gad.208501.

[65] Shohei Takuno. "Independent Molecular Basis of Convergent". In: 200.August (2015), pp. 1297–1312. DOI: 10.1534/genetics.115.178327.

[66] Roberto Tuberosa, Andreas Graner, and Emile Frison. *Genomics of plant genetic resources: Volume 1. Managing, sequencing and mining genetic resources*. 2014, pp. 1–710. ISBN: 9789400775725. DOI: 10.1007/978-94-007-7572-5.

[67] M. R. Tuinstra, G. Ejeta, and P. B. Goldsbrough. "Heterogeneous inbred family (HIF) analysis: A method for developing near-isogenic lines that differ at quantitative trait loci". In: *Theoretical and Applied Genetics* 95.5-6 (1997), pp. 1005–1011. ISSN: 00405752. DOI: 10.1007/s001220050654.

[68] Hans Van Os et al. "SMOOTH: A statistical method for successful removal of genotyping errors from high-density genetic linkage data". In: *Theoretical and Applied Genetics* 112.1 (2005), pp. 187–194. ISSN: 00405752. DOI: 10.1007/s00122-005-0124-y.

[69] Jean-Philippe Vielle-Calzada and Jaime Padilla. "The Mexican Landraces: Description, Classification and Diversity". In: 2009, pp. 544–560. ISBN: 978-0-387-79417-4. DOI: 10.1007/978-0-387-79418-1.

[70] Giannina Vizzotto and Rachele Falchi. "The Kiwifruit Genome". In: *Springer-Verlag* (2016), pp. 189–204. DOI: 10.1007/978-3-319-32274-2. URL: http://link.springer.com/10.1007/978-3-319-32274-2.

[71] Li Wang et al. "The interplay of demography and selection during maize domestication and expansion". In: (2017), pp. 1–13. DOI: 10.1186/s13059-017-1346-4.

[72] Xiaoqian Wang et al. "Genetic background effects on QTL and QTLÂÃ—Âenvironment interaction for yield and its component traits as revealed by reciprocal introgression lines in rice". In: *Crop Journal* 2.6 (2014), pp. 345–357. ISSN: 22145141. DOI: 10.1016/j.cj.2014.06.004.

[73] M. Wellenreuther et al. "Local adaptation along an environmental cline in a species with an inversion polymorphism". In: *Journal of Evolutionary Biology* 30.6 (2017), pp. 1068–1077. ISSN: 14209101. DOI: 10.1111/jeb.13064.

[74] H G Wilkes. "Hybridization of Maize and T e o s i n t e , in Mexico and Guatemala and the I m p r o v e m e n t of Maize 1". In: (1977), pp. 254–293.

[75] Gregory A. Wray. "The evolutionary significance of cis-regulatory mutations". In: *Nature Reviews Genetics* 8.3 (2007), pp. 206–216. ISSN: 14710056. DOI: 10.1038/nrg2063.

[76] Xun Wu et al. "Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize". In: *Plant biotechnology journal* 14.7 (2016), pp. 1551–1562. ISSN: 14677652. DOI: 10.1111/pbi.12519.

[77] Guanghui Xu et al. "Complex genetic architecture underlies maize tassel domestication". In: *New Phytologist* 214.2 (2017), pp. 852–864. ISSN: 14698137. DOI: 10.1111/nph.14400.

[78] Ning Yang et al. "Contributions of Zea mays subspecies mexicana haplotypes to modern maize". In: *Nature Communications* 8.1 (2017), p. 1874. ISSN: 2041-1723. DOI: 10.1038/s41467-017-02063-5. URL: http://www.nature.com/articles/s41467-017-02063-5.

[79] Nengjun Yi, Shizhong Xu, and David B Allison. "Quantitative Trait Loci". In: 883.October (2003), pp. 867–883. DOI: 10.1007/978-1-61779-785-9.

[80] Jeremy B. Yoder et al. "Genomic signature of adaptation to climate in Medicago truncatula". In: *Genetics* 196.4 (2014), pp. 1263–1275. ISSN: 19432631. DOI: 10.1534/genetics.113.159319.

Appendix A

Primers sequence

TABLE A.1: **SSR markers at the *inv4m***

Locus	Marker	Physical position (bp)	Primer Sequence		Allele size (bp)		Repeat
			Forward	Reverse	B73	PT	
Chr 4	bnlg1137	169770082	ATGAGCTCAGTCACACTGTAGTG	ACTGATGACTGGTCCATGCA	242	224	AG(34)
Chr 4	bnlg1784	170944265	GCAACGATCTGTCAGACGAA	TTGGCATTGGTAATGGGTCT	243	230	AG(13)
Chr 4	umc2009	180471579	CACTTCAAGGCTGTGTCATTGTG	CCTCTCCGGATCAGAACAAGAAG	100	64	

TABLE A.2: **CAPS marker at the *inv4m***

Oligo code	Locus	Marker	Physical position (bp)	Primer Sequence		SNP	
				Forward	Reverse	B73	PT
RS 427/428	Chr4	PZE104103103	179617762	CTGAGCAGGAGATGATGGCCACTC	GGAAAGGACATAAAAGAAAGGTGCA	G	A

TABLE A.3: **Expected polymorphisms between B73 and PT at *inv4m*.** Detected polimorphisms using the CAPS marker for the SNP PZE104103103

Locus	Marker	Context sequence)	Size of PCR product (bp)	Enzyme	Restiction fragments expected (bp) +		
					B73	PT	
Chr4	PZE104103103	aagaagatgcatgaatggttttgcacgtagcacttgctagggcgtacgca[A/G]attcttcttcttggccctgtgaattccgaattctcgacgattgattctg	502	*HinfI*	166; 44; 251; 61	166; 295; 61	You can simply put text here

TABLE A.4: **specific primers used for RT-qPCR at the *inv4m***

Oligo code	Gene	Primer Sequence		Amp Len	Description
		Forward	Reverse		
RA 1041/1042	GRMZM2G019177	ATACTCCAAAGCTGGCCAACA	TCAGCCCACTTCCTGTGATCT	100	DESCRIPTION
RA 1045/1046	GRMZM2G042380	ATGTCCTCTCCATGCCATCTCA	AGGGTCGTAGTTACATGGAATGGT	75	DESCRIPTION
RA 1047/1048	GRMZM2G054012	AAGGTACGTCACTTCGCATCAG	CCTGCCGTATCATCGAAATACA	75	DESCRIPTION
RA 1063/1064	GRMZM2G091189	AAGCCTCTGAACGACCGTGTT	ATGGCTTCTCCTTGGCAGTTT	100	DESCRIPTION
RA 1065/1066	GRMZM2G096343	GCGGCTGACACTGGATAAAAA	GCGGCAGAGATCATTCAAAAC	100	DESCRIPTION
RA 1069/1070	GRMZM2G125838	CGGCAACTACTGTTGGATCGT	CTGCACCTCCTCGATGTCGTA	100	DESCRIPTION
RA 1077/1078	GRMZM2G142661	GCAGCCCGAGTACAACCTCTT	AACCAATGCCATAGGTGCTGTAA	80	DESCRIPTION
RA 1083/1084	GRMZM2G478709	TTCTGCAGGTTTGGAATGCTT	TAACAGCACAGGAGCCATGCT	95	DESCRIPTION

TABLE A.5: **specific primers used for RT-qPCR close to the *inv4m***

Oligo code	Gene	Primer Sequence		Amp Len	Description
		Forward	Reverse		
RA 1085/1086	GRMZM2G020775	CTGCCAGTGTTCTCGTATGTGAA	GCAGCTAAGGCATCGAAACTTC	75	DESCRIPTION
RA 1087/1088	GRMZM2G077258	ACCGTATATTTTCCAGGCATCCA	CGAACCAGGCGTGAGAACAGA	100	DESCRIPTION
RA 1089/1090	GRMZM2G125455	CCCGCTGATTCCAAATGATG	TGCGACTGATGATCTTGCTACA	100	DESCRIPTION
RA 1091/1092	GRMZM2G444623	CTTCCAGCTGTCTCCACGAAA	CCGTCACCAGTTTAGCGAAAA	100	DESCRIPTION
RA 1093/1094	GRMZM2G003501	CTCAAGCACTGACCCGACTACAA	TACAACGCTCCGGACCACAA	127	DESCRIPTION
RA 1095/1096	GRMZM2G171060	AGCTGGCAAGGTGCATTGA	CCTTTGCCCTTTCCCTGATC	72	DESCRIPTION
RA 1099/1100	GRMZM2G063473	GGGCTATGCAGTGAGGGTTTG	CTCCACAGTTCCGTGCTTTGT	80	DESCRIPTION
RA 1101/1102	GRMZM2G063276	GGATGGTTTAGACGAAATGGAAA	TCGCGCTGCTGTTCTATTACTT	78	DESCRIPTION
RA 1103/1104	GRMZM2G136996	TGGTGCCCAGAAGCAAGTTAT	AAAGCCACACCCGTGCTAAG	75	DESCRIPTION
RA 1105/1106	GRMZM2G344967	CATCTACTGCCAGGTGCTTGAG	CATCACGTGGCATGCTTGAG	104	DESCRIPTION

Appendix B

ASE pipeline

Appendix C

Total RNA isolation and qRT-PCR

Total RNA isolation

Total RNA was isolated using the PureLink Micro-to-Midi Total RNA Purification System (AmbionR) with the modification of a previous extraction with Trizol (invitrogen) instead lysis buffer according to the manufacturer's instructions. RNA quality and integrity was assessed by spectrophotometry using nanodrop ND-8000 and agarose gel electrophoresis respectively. Only RNA samples with a 260/280 ratio between 1.9 and 2.1, and a 260/230 ratio greater than 2.0 were used for further analysis.

First-strand cDNA synthesis

First-strand cDNA synthesis was conducted using *Superscript II* RT reverse trasncriptase (invitrogen) according to manufacturer's instructions: 500 ng of total RNA were mixed with 0.5 μL of reverse primer mix (5 μM each) and 1 μL of dNTP Mix (10 mM each). Reaction was incubated 5 min at 65°C and quick chill on ice before the 5 minutes ends while temperature in thermal cycler was lowered to 42°C. 1 unit of *Superscript II* RT was added to the mix and 0.9 μL of distilled water to a final volume of 11 μL. The mixture was returned to the thermal cycler and kept at 42°C for 60 min more, then the temperature was raised to 70°C for 15 min to inactivate the enzyme. Specific primers used for RT-qPCR are shown in Table A.4.

Primers validation

Using RNA from 14 days old B73 seedlings as template. once confirmed the specificity, the standard curve was performed: Five serial dilutions (1:5) were aliquoted from the synthesized cDNA (1:5, 1:25, 1:125, 1:625, 1:3125). The standard curve was represented as the antilogarithm regression of the CT mean value of each dilution against the logarithm of the cDNA amount used at each point. Here the slope should accomplish two parameters: value -3.32 $\pm$10%, and correlation index (R^2 = 0.99); these parameters indicate a close adjust between the lineal regression and the pattern curve and the CT individual values from the pattern samples and confirms that the relative gene expression is inside the adequate dynamic range.

Finally, to validate (Confirm that our gene amplification is comparable with the house keeping selected) a CT plot was constructed (interest gene CT – house keeping gene CT), plotting against the logarithms of the cDNA amount; the slope should be between 0.1 to -0.1, as indication that amplification efficiency of both genes are comparable. Three technical replicates were performed twice with no template control (NTC) to verify the quality of the amplification.

Quantitative Real-Time PCR Analysis

RT-qPCR

In order to confirm the expression pattern of the selected genes, the synthesized cDNA was used as template for the qPCR performed in the Step One Plus Real-Time PCR System (Applied Biosystems). SYBR Green PCR Master Mix (KAPA Biosystems) was used for all reactions in a 10 μL reaction volume: 5 μL Master Mix (2X), 0.2 μL primer forward (10 μM), 0.2 μL primer reverse (10 μM), 0.2 μL sybr ROX dye as passive reference, 0.9 μL milliQ water and 3.5 μL of cDNA (1:10) as template, following next conditions: 95°C for 10min, 40 cycles of 95°C for 15s, 60°C for 1min, and a final melt curve stage from 60°C to 95°C. Control reactions were carried out using specific primers for the βtubuline gene (GRMZM2G108766).

Gene expression was normalized by subtracting the CT value of the control gene (HK) from the CT value of the gene of interest. Average expression ratios were obtained from the equation 2–ΔΔCT, where ΔΔCT represents ΔCT (gene of interest)–ΔCT (gene of interest at T_0), according to the protocol reported by Czechowski *et al.* in 2004 [14]. Specific primers used for RT-qPCR are shown in Table A.4.

Appendix D

Derived publications

- Allele specific expression analysis identifies regulatory variation associated with stress-related genes in the Mexican highland maize landrace Palomero Toluqueño