Studies on flowering time and photoperiod sensitivity in domesticated and wild amaranth species (*Amaranthus* spp.)

Dissertation to obtain the doctoral degree of Agricultural Sciences

(Dr. sc. agr.)

Faculty of Agricultural Sciences

University of Hohenheim

Institute of Plant Breeding, Seed Science and Population Genetics

Submitted by

Ali Baturaygil

from Ankara, Turkey

2023

This thesis was accepted as a doctoral thesis (Dissertation) in fulfilment of the regulations to acquire the doctoral degree “Doktor der Agrarwissenschaften” by the Faculty of Agricultural Sciences at the University of Hohenheim, on 31st of May 2023.

Date of the oral examination: 26th of July 2023

Examination Committee:
Chairperson of the oral examination: Prof. Dr. Stefan Böttinger
Supervisor and Reviewer: Prof. Dr. Karl Schmid
Co-Reviewer: Prof. Dr. Christian Jung
Additional examiner: Prof. Dr. Uwe Ludewig

“When you see a good move, look for a better one.”

Emanuel Lasker

Contents

Abbreviations

BVP	Basic vegetation period
CH_4	Methane
CMS	Cytoplasmic male sterile
CO_2	Carbon dioxide
GA_3	Gibberelin
GDD	Growing degree days
GWAS	Genome-wide association study
KASP	Kompetitive allele specific PCR
LD	Linkage-disequilibrium
LSD	Least significant difference
MAF	Minor allele frequency
MAGIC	Multi-parent advanced generation inter-cross
Mbp	Mega base pair
NAM	Nested association mapping
PCA	Principal component analysis
PRR	Pseudo response regulator
PSP	Photoperiod sensitive period
QTL	Quantitative trait locus
RIL	Recombinant inbred line
SNP	Single nucleotide polymorphism
WGS	Whole-genome sequencing

General Abstract

Flowering time plays fundamental roles in the local adaptation and agricultural productivity of the crops. Photoperiodic response regulates the time of flowering by adjusting the response of plant circadian rhythm to environmental signals. Amaranth (*Amaranthus* spp.) is a short-day crop native to Central and South America, and mainly used as grain and vegetable. Hence, photoperiod sensitivity is a pivotal trait for grain amaranths in Central Europe climatic and long-day conditions, as it determines the local adaptability and the cultivation purpose of the crop i.e., grain or biomass production. However, the knowledge on the different aspects such as breeding, domestication history and adaptation genetics is very limited in grain amaranths. In this project, we studied such different aspects of grain amaranths by addressing the elucidative photoperiod sensitivity trait.

In the first study, the phenotypic evaluation of biomass yield components revealed two distinct growth types. Of those, our ten biomass genotypes showed mild to high photoperiod sensitivity, flowered late or completely rejected flowering, reached long final plant heights and low dry matter content. In contrast, the only grain type variety showed photoperiod insensitivity, flowered early, and reached a short final plant height and a relatively higher dry matter content. Our results suggested that selection for both high dry matter yield and content requires a trade-off between photoperiod sensitivity and early flowering, due to the negative correlation between these traits.

In the second study, characterization of genebank accessions from the three major grain species (*A. caudatus, A. cruentus, A. hypochondriacus*) and their wild relative species (*A. hybridus and A. quitensis*) for adaptive traits such as flowering time and seed setting under long-day conditions discovered a larger photoperiodic variation in the Central American accessions ranging from insensitivity to high sensitivity, whereas South American accessions showed a more narrow variation, limited by mild sensitivity. This result suggests the Central American origin of the wild relative *A. hybridus*, which might have migrated from Central to South America, and potentially has been selected against high photoperiod sensitivity. Moreover, we studied the environmental variables that may influence seed setting.

Photoperiod insensitive accessions set seed regardless of their origin. However, mild photoperiod-sensitive accessions set seed, only if they were from warm center of origin.

In the third study, we investigated the genetic architecture of photoperiod sensitivity. The bimodal-like flowering time distributions, and the linkage and association mapping studies using three different populations revealed that photoperiod sensitivity trait is controlled in an oligogenic manner. In particular, all three populations consistently found the same '*consensus region*' that includes a very promising candidate gene called 'response regulator of two-component system*'*. The homologs of this candidate gene are responsible for photoperiodic response in a variety of different crops and the model species *Arabidopsis thaliana*. In addition, the phenotypic analyses, and the marker data (i) showed photoperiod sensitivity guided pleiotropic relationships between the traits, (ii) revealed a potential epistatic behavior of the genomic region controlling photoperiod sensitivity, and (iii) showed the dominance of photoperiod sensitivity over insensitivity in that region.

Allgemeine Zusammenfassung

Der Blühzeitpunk spielt eine grundlegende Rolle für die lokale Anpassung und die landwirtschaftliche Produktivität von Kulturpflanzen. Die photoperiodische Reaktion reguliert den Zeitpunkt der Blüte, indem sie den zirkadianen Rhythmus der Pflanzen auf Umweltsignale anpasst. Amaranth (*Amaranthus* spp.) ist eine in Mittel- und Südamerika beheimatete Kurztagpflanze, die hauptsächlich als Getreide und Gemüse verwendet wird. Daher ist die Sensitivität der Photoperiode ein zentrales Merkmal für Getreide Amaranth unter mitteleuropäischen Klima- und Langtagbedingungen, da sie die lokale Anpassungsfähigkeit und den Anbauzweck der Pflanze, d. h. die Getreide- oder Biomasseproduktion, bestimmt. Allerdings ist das Wissen über verschiedenen Aspekte wie Züchtung, Domestikationsgeschichte und Anpassungsgenetik bei Getreide Amaranth sehr begrenzt. In diesem Projekt untersuchten wir diese verschiedenen Aspekte von Amaranth, indem wir uns mit dem aufschlussreichen Merkmal der Sensitivität der Photoperiode befassten.

In der ersten Studie ergab die phänotypische Auswertung der Ertragskomponenten der Biomasse zwei unterschiedliche Wachstumstypen. Unsere zehn Biomasse Genotypen wiesen eine geringe bis hohe Sensitivität der Photoperiode auf, blühten spät oder verzichteten ganz auf die Blüte, erreichten große Endhöhen und niedrige Trockenmassegehalte. Im Gegensatz dazu zeigte die einzige Getreidesorte eine Unempfindlichkeit gegenüber der Photoperiode, blühte früh und erreichte eine kurze Endhöhe und einen relativ hohen Trockensubstanzgehalt. Unsere Ergebnisse deuten darauf hin, dass die Selektion auf hohe Trockenmasseerträge und -gehalte aufgrund der negativen Korrelation zwischen diesen Merkmalen einen Kompromiss zwischen der Sensitivität der Photoperiode und früher Blüte erfordert.

In der zweiten Studie wurde bei der Charakterisierung von Genbankakzessionen der drei wichtigsten Getreidearten (*A. caudatus, A. cruentus, A. hypochondriacus*) und ihrer wilden Verwandten (*A. hybridus* und *A. quitensis*) hinsichtlich adaptiver Merkmale wie Blütezeit und Samenbildung unter Langtagbedingungen eine größere photoperiodische Variation bei den mittelamerikanischen Akzessionen festgestellt, die von Unempfindlichkeit bis zu hoher

Empfindlichkeit reicht, während die südamerikanischen Akzessionen eine geringere Variation mit geringer Empfindlichkeit aufweisen. Dieses Ergebnis deutet auf den mittelamerikanischen Ursprung des wilden Verwandten *A. hybridus* hin, der möglicherweise von Mittel- nach Südamerika eingewandert ist und möglicherweise gegen eine hohe photoperiodische Empfindlichkeit selektiert wurde. Darüber hinaus untersuchten wir die Umwelteinflüsse, die die Samenbildung beeinflussen können. Akzessionen, die eine photoperiodische Unempfindlichkeit aufweisen, setzen unabhängig von ihrer Herkunft Samen an. Milde photoperiodische empfindliche Akzessionen setzten jedoch nur dann Samen an, wenn sie aus warmen Herkunftsgebieten stammten.

In der dritten Studie untersuchten wir die genetische Architektur der Sensitivität auf die Photoperiode. Die bimodale Verteilung der Blütezeit und die Linkage- und Assoziationsstudien mit drei verschiedenen Populationen zeigten, dass das Merkmal der Sensitivität der Photoperiode oligogen kontrolliert wird. Insbesondere wurde in allen drei Populationen durchweg dieselbe " *Konsensus Region*" gefunden, die ein vielversprechendes Kandidatengen namens ‘response regulator of two-component system’ enthält. Die Homologe dieses Kandidatengens sind für die photoperiodische Reaktion in einer Reihe von verschiedenen Kulturpflanzen und der Modelpflanze *Arabidopsis thaliana* verantwortlich. Darüber hinaus haben die phänotypischen Analysen und die Markerdaten (i) gezeigt, dass die photoperiodische Empfindlichkeit zu pleiotropen Beziehungen zwischen den Merkmalen führt, (ii) ein potenzielles epistatisches Verhalten der genomischen Region, die die Sensitivität der Photoperiode kontrolliert, aufgezeigt und (iii) die Dominanz der Sensitivität der Photoperiode gegenüber der Unempfindlichkeit in dieser Region gezeigt.

Özet

Çiçeklenme zamanı bitkilerin lokal adaptasyonu ve tarımsal üretkenliği açısından temel roller üstlenir. Fotoperiodik tepki sirkadiyen ritmin çevresel sinyallere verdiği tepkiyi ayarlayarak çiçeklenme zamanını düzenler. Amarant (*Amaranthus* spp.) Orta ve Güney Amerika kökenli bir kısa gün bitkisidir ve esas olarak tane ve sebze olarak tüketilir. Dolayısıyla, Orta Avrupa iklimi ve uzun gün koşullarında yetişen amarantlar için fotoperiyot hassasiyeti son belirleyici bir özelliktir çünkü lokal adaptasyonu ve bitkinin yetiştirilme amacını (tane ya da biyokütle üretimi) tayin etmektedir. Buna rağmen amarant bitkisinde ıslah, domestikasyon süreci ve adaptasyon genetiği gibi alanlardaki bilgi oldukça sınırlıdır. Bu projede tane amarantının bu alanlardaki bilinmeyen yönlerine ışık tutmak için bilgilendirici bir özellik olan fotoperiyot hassasiyetini ele aldık.

İlk çalışmamızda biyokütle verim bileşenlerinin fenotipik değerlendirmesi neticesinde birbirine zıt iki farklı büyüme tipi tespit ettik. Bunlardan ilkinde, 10 biyokütle tipi amarant genotipi hafif ile yüksek derece arasında fotoperiyot hassasiyeti gösterdi, ya geç çiçeklendi ya da hiç çiçeklenmedi, uzun bir boya ve düşük bir kuru madde içeriğine ulaştı. Fakat tane tipi olan tek amarant genotipi ise fotoperioyota tepkisizlik gösterdi, erken çiçeklendi, kısa bir hasat uzunluğuna ve biyokütke tipi amatantlara kıyasla daha yüksek bir kuru madde içeriğine ulaştı. Edindiğimiz sonuçlara göre hem kuru madde verimi hem de içeriği için yapılacak olan seleksiyonun fotoperiyot hassasiyeti ile erkencilik arasında bir dengeyi tutturması, bu iki özellik arasındaki negatif korelasyon sebebiyle gereklidir.

İkinci çalışmamızda üç ana tane amarant türü (*A. caudatus, A. cruentus, A. hypochondriacus*) ve bunların yabani akrabası olan iki türe (*A. hybridus ve A. quitensis*) ait genbankası genotiplerini, çiçeklenme zamanı ve tohum bağlama gibi adaptasyon özellikleri açısından uzun gün koşullarında fenotipik değerlendirmeye tabi tuttuk. Orta Amerika genotiplerinde fotoperiyota tepkisizlikle yüksek tepki arasında değişen büyük bir varyasyon gözlenirken, Güney Amerika genotiplerinde orta düzeyde hassasiyetle sınırlı olmak üzere daha düşük bir varyasyon gözlendi. Bu sonuç yabani bir akraba tür olan *A. hybridus*'un Orta Amerika kökenli olabileceğini, Orta Amerika'dan GüneyAmerika'ya göç etmiş olabileceğini, ve yüksek

seviyede fotoperiyodik hassasiyete karşı bir seleksiyona maruz kalmış olabileceğini ortaya koymaktadır. Bunun haricinde, tohum bağlamaya etki edebilecek çevresel değişkenleri çalıştık. Fotoperioyot hassasiyeti olmayan genotipler orijinlerinden bağımsız olarak tane bağladılar, fakat hafif derecede hassas olanlar sadece sıcak bölgelerden köken alıyorlarsa tane bağladılar.

Üçüncü çalışmamızda, fotoperiyot hassasiyeti özelliğinin genetik mimarisini çalıştık. Bimodal benzeri çiçeklenme zamanı dağılımları ve üç farklı popülasyon kullanılarak gerçekleştirilen genetik haritalama çalışmaları, fotoperiyot hassasiyeti özelliğinin oligogenik bir şekilde kalıtıldığını gösterdi. Özellikle üç popülasyon da 'response regulator of two-component system' isimli aday geni ihtiva eden aynı 'konsensüs bölgesini' tutarlı bir şekilde buldu. Bu aday genin homogları pek çok farklı tarımsal amaçlı kullanılan ve *Arabidopis thaliana* gibi model bitki türünde fotoperiyodik tepkiyi kontol etmektedir. Ayrıca fenotipik analizler ve markör datası (i) özellikler arasında fotoperiyot hassasiyeti kaynaklı pleiotropik ilişkileri, (ii) fotoperiyot hassasiyetini kontrol eden genomik bölgenin muhtemel epistatik karakterini, ve (iii) bu bölgede fotoperiot hassasiyetinin, fotoperiyot tepkisizliğine olan dominant karakterini ortaya çıkartmıştır.

Resumen general

El tiempo la floración desempeña un papel fundamental en la adaptación local y la productividad agrícola de los cultivos. La respuesta fotoperiódica regula el momento de la floración ajustando la respuesta del ritmo circadiano de la planta a las señales ambientales. El amaranto (*Amaranthus* spp.) es un cultivo de día corto originario de América Central y del Sur, que se utiliza principalmente como cereal y hortaliza. Por lo tanto, la sensibilidad al fotoperiodo es un rasgo fundamental para los amarantos de grano en las condiciones climáticas y de día largo de Europa Central, ya que determina la adaptabilidad local y el propósito de su cultivo, es decir, la producción de grano o biomasa. Sin embargo, el conocimiento sobre diferentes aspectos como la mejora genética, la historia de la domesticación y la genética de la adaptación es muy limitado en el amaranto. En este proyecto, estudiamos estos diferentes aspectos del amaranto elucidando el rasgo de sensibilidad al fotoperiodo.

En el primer estudio, la evaluación fenotípica de los componentes del rendimiento de biomasa reveló dos tipos de crecimiento distintos. De ellos, nuestros diez genotipos de biomasa mostraron una sensibilidad al fotoperiodo de leve a alta, florecieron tarde o no florecieron, alcanzaron una altura final de planta alta y bajos contenidos de materia seca. Por el contrario, la única variedad de tipo grano mostró insensibilidad al fotoperiodo, floreció pronto, y alcanzó una altura final de planta baja y un contenido en materia seca relativamente más alto. Nuestros resultados sugieren que la selección para un alto rendimiento y contenido de materia seca requiere un compromiso entre la sensibilidad al fotoperiodo y la floración temprana, debido a la correlación negativa entre estos rasgos.

En el segundo estudio, la caracterización de accesiones del banco de germoplasma de las tres principales especies de grano (*A. caudatus*, *A. cruentus*, *A. hypochondriacus*) y sus especies parientes silvestres (*A. hybridus* y *A. quitensis*) para rasgos adaptativos como el tiempo de floración y el establecimiento de las semillas en condiciones de día largo descubrió una mayor variación fotoperiódica en las accesiones centroamericanas, que va desde la insensibilidad a la

alta sensibilidad, mientras que las accesiones sudamericanas mostraron una variación más estrecha, limitada con una sensibilidad leve. Este resultado sugiere el origen centroamericano del pariente silvestre *A. hybridus*, que podría haber migrado de Centroamérica a Sudamérica, y potencialmente haber sido seleccionado contra la alta sensibilidad al fotoperiodo. Además, estudiamos las variables ambientales que pueden influir en el establecimiento de las semillas. Las accesiones sensibles al fotoperiodo se establecen independientemente de su origen. Sin embargo, las accesiones insensibles al fotoperiodo leve se establecen, sólo si proceden de centros de origen cálidos.

En el tercer estudio, investigamos la arquitectura genética de la sensibilidad al fotoperiodo. Las distribuciones del tiempo de floración de tipo bimodal y los estudios de mapeo de ligamiento y asociación utilizando tres poblaciones diferentes revelaron que el rasgo de sensibilidad al fotoperiodo está controlado de forma oligogénica. En particular, las tres poblaciones encontraron sistemáticamente la misma "región de consenso" que incluye un gen candidato muy prometedor denominado ‘response regulator of two-component system’ regulador de respuesta del sistema de dos componentes". Los homólogos de este gen candidato son responsables de la respuesta fotoperiódica en una variedad de cultivos diferentes y en la especie modelo *Arabidopsis thaliana*. Además, los análisis fenotípicos y los datos de los marcadores (i) mostraron relaciones pleiotrópicas guiadas por la sensibilidad al fotoperiodo entre los rasgos, (ii) revelaron un potencial comportamiento epistático de la región genómica que controla la sensibilidad al fotoperiodo, y (iii) mostraron la dominancia de la sensibilidad al fotoperiodo sobre la insensibilidad en esa región.

1. General Introduction

1.1.Need for alternative crops

The global population has been projected to increase between 2 to 2.4 billion by 2050 (Delgado *et al.*, 2011; Massawe *et al.*, 2016) and the rate of current food production does not meet the rate of population growth (Abberton *et al.*, 2016). In addition, the adverse effects of climate change on the natural resources and the increasing agricultural input expenses undermine agricultural productivity and threatens food security. The current conventional agricultural systems are the extension of Green Revolution, which is characterized by high-yielding cultivars and narrow crop diversity in the farmers' fields limited with a few major cereal crops (Pingali, 2012; Massawe *et al.*, 2016). As a result of the Green Revolution, rapid yield increments were achieved by the intensive use of input such as fertilizers and irrigation, however, that dependence on high input and the vulnerability of low crop diversity to biotic and abiotic stressors hinders the sustainability of agricultural productivity and food security. Such an intensification in the agricultural production systems have additional negative impacts on the environment including water shortages, soil degradation, pollution, and loss of biodiversity (Ebert, 2014). Even though the required production increase can be met with intensive agricultural production systems, that would require a remarkable amount of land clearing (Ebert, 2014). Altogether, Ebert (2014) concludes that intensification in agricultural production systems brings negative environmental impacts and dependency on major crops for meeting the required agricultural production, which is not sustainable in the long term. In contrast, alternative agricultural systems should be less dependent on the input use, more tolerant to biotic and abiotic stress factors and promote crop diversification to meet with the increasing food demand (Abberton *et al.*, 2016).

An effective strategy to achieve this goal is the inclusion of underutilized crops into such alternative agricultural production systems. Ebert (2014) suggests that the proliferation of underutilized crops would establish temporal and spatial heterogeneity in the agricultural production systems, and such diversification would result in higher resilience against environmental stress factors. At present, the daily calorie intake is supplied by a very low

number of crops and restricted to very narrow crop diversity. For instance, the diversity in the farmers' field has been reduced by 75 % in the last century and only 20 out of approximately 300 crops available in the market supply 90 % of the daily calorie intake, of which only three major cereal crops (maize, wheat, and rice) are the biggest contributors of our diets (Massawe *et al.*, 2016). Most of the staple crops are adopted to the high-input agricultural systems and more vulnerable to the environmental stresses (Mayes *et al.*, 2012). Contrariwise, some underutilized crops display resilient physiological attributes that ease adaptation to harsh environmental conditions such as drought and flood (Massawe *et al.*, 2016), where high-input systems are not practiced, and staple crops cannot be cultivated in a feasible manner (Ebert, 2014). In addition, many underutilized crops are the cultural components of their country of origin, and in many cases, they exhibit high nutritional qualities (Ebert, 2014; Massawe *et al.*, 2016). Ordinary practices in the modern plant breeding of the major crops have been based on the rearrangement of a limited number of useful alleles due to a narrowed genetic base in the elite germplasm (Gepts, 2004; Abberton *et al.*, 2016), which do not permit large genetic gains and also increase the cost of improvement (Abberton *et al.*, 2016). On the other hand, the increasing number of reference genomes in underutilized crops and the decreasing sequencing costs may permit the improvement and dispersal of minor crops rapidly through the genomics-assisted breeding thanks to the availability of large untapped variation in these crops (Abberton *et al.*, 2016). Hence, development of underutilized crop varieties, broadening their use by improving their adaptive skills, especially if they are more competent over their major crop alternatives is a fundamental strategy to ensure food security (Abberton *et al.*, 2016).

1.2. Flowering time as an adaptive trait

Flowering time adaptation is a key process between domestication and genetic improvement of crops. It is a strategical decision made by plants to adapt to the local conditions and to warrant agricultural productivity in case of crop cultivation. More specifically, plants intend to sense the favorable season of the year to assure the availability of environmental conditions suitable for seed production. In this regard, two major parameters are photoperiod and temperature (Izawa, 2007). Photoperiod means the duration of a day with sunlight and is a

function of latitude. Photoperiod sensitivity indicates the dependency of a plant to a specific photoperiod to flower. For instance, short-day plants such as maize, sorghum and rice originated from tropical to subtropical zones and need short-day to flower, whereas long-day plants such as wheat, barley and lentil originated from the Fertile Crescent and need long-day to flower (Zohary *et al.*, 2012; Nakamichi, 2014). A third group consists of day-neutral plants such as tomato, eggplant, and potato, that do not need a specific photoperiod to flower (Nakamichi, 2014). Photoperiod changes in a certain location during the year because of the annual rotation of the earth around the sun (Jackson, 2009), however, it is more stable across years. Differently, temperature determines the rate of plant development (Craufurd and Wheeler, 2009) but it is more unstable across years (Lagercrantz, 2009). Hence, photoperiod is a more reliable indicator of the season relative to temperature. Under ideal growth conditions, late flowering leads to reserve sources for higher seed production, however, early flowering is a preferred strategy under sub-optimal conditions such as the short duration of favorable temperature or in case of an approaching terminal drought because plants prioritize to produce seeds before the stress factors act (Roux *et al.*, 2006).

Plants adjust their photoperiodic responses for local adaptation. Interestingly, contrasting photoperiodic responses can both contribute to the agricultural productivity. For example, photoperiod sensitivity is an adaptive mechanism for sorghum and pearl millet cultivation in West Africa that synchronizes flowering time despite different sowing dates due to the highly variable time of rain onset in the region (Haussmann *et al.*, 2012). In contrast, photoperiod insensitivity is a desired trait especially in short-day crops grown in the high latitudes and temperate climates and establishes a constant length of vegetation period (Gaudinier and Blackman, 2019), thus, ensures a uniform seed production despite varying photoperiods of different growing environments (Ceccarelli, 1994).

1.3. Use of short-day crops for bioenergy

Several short-day crops such as maize, rice, soybean, and sorghum delay or reject flowering under long-day conditions as their specific photoperiod need for flowering is not fulfilled but prolong vegetative growth and accumulate biomass (Jung and Müller, 2009; Matsubara *et al.*, 2011; Murphy *et al.*, 2014; Zhang *et al.*, 2020). Plant biomass is used as substrate in bioenergy production. Biogas is a renewable energy source that is used in heat and electricity production and obtained by the biochemical decomposition (anaerobic digestion) of biomass substrate in biogas plants (Weiland, 2010). Biogas is composed of methane (CH_4) and carbon dioxide (CO_2), of which the methane yield is the ultimate product in the energy production and is primarily determined by dry matter yield. Currently, silage maize is the most popular biogas substrate in Germany due to high methane yield (Brauer-Siebrecht *et al.*, 2016). In 2017, biogas crops were cultivated in 1.374.000 hectare and approximately 0.9 million hectare area – corresponding to 36 % of the total maize cultivation area – was allocated to the biogas maize cultivation in Germany (FNR, 2019). However, conventional maize cultivation raise the following concerns: (i) silage maize has been dominating crop rotations and its successive cultivation disturbs agricultural diversity, (ii) the occupation of the arable land with biogas maize reduces the available land for food crop cultivation and therefore increases food prices and threatens food security and (iii) cultivation of biogas maize in the marginal areas increase the soil erosion risk and eventually these factors exhibit a need for alternative biogas crops (Brauer-Siebrecht *et al.*, 2016; Vogel *et al.*, 2016).

1.4. Amaranth as an alternative crop

Amaranth is an ancient crop and originated from Central and South America (Sauer, 1967). It is one of the oldest crops of the Americas, that had been consumed as a major crop together with maize and beans by the ancient civilizations of that geography such as the Aztecs, the Mayas and the Incas in the pre-Columbian era (Das, 2014). Furthermore, amaranth was an important element of the Aztec customs and had been used in religious ceremonies (Myers, 1996). The family of *Amaranthaceae* consists of 60-70 species, most of which are known as weeds in the agricultural production areas (Costea and Tardif, 2004; Chaudhari *et al.*, 2017). Minor crops such as quinoa, buckwheat, and amaranth that are consumed as grain. Since they

are taxonomically not members of *Poaceae* family they are called “pseudo-cereals”. Amaranth is consumed as a pseudo-cereal more frequently in the Americas and Asia, and as a vegetable in a variety of tropic and temperate regions, in addition to its use as an ornamental plant (National Research Council, 1984). Some of the major grain amaranth producers are Mexico, Russia, China, India, Nepal, Argentina, and Kenya (Aderibigbe *et al.*, 2020), however, conventional amaranth production is very limited, and no information is available in FAOSTAT database for 2019, due to its minor crop status. The production and marketing of amaranth has been re-initiated by the initiative of the Rodale Institute in the mid-70s (Valcárcel-Yamani and Lannes, 2012). The three major grain amaranth species are *A. cruentus, A. caudatus and A.hypochondriacus.* Grains of these species contain high protein and lysine concentrations and present an alternative source of carbohydrate without gluten to patients with celiac disease (Becker *et al.*, 1981; Stallknecht and Schulz-Schaeffer, 1993; Venskutonis and Kraujalis, 2013). The major species consumed as a vegetable are *A. cruentus, A. dubius, A. hybridus, A. lividus, and A. tricolor,* and those species have been integrated into the regional human diets in many different parts of the world (National Research Council, 1984). Principally, their nutritional qualities are similar to the other green vegetables, however, present a higher amount of nutrients per portion due to the higher dry matter content of their leaves compared to many other green vegetables (National Research Council, 1984).

1.5. Biological features of Amaranth

Amaranth is a predominately self-fertilizing plant, but out-crossing rates between 3 and 32% have been reported (Jain *et al.*, 1982; Hauptli and Jain, 1985), which frequently leads to spontaneous hybridization and gene flow events in the natural populations. Due to its self-fertilizing nature, breeding methods applied for self-pollinating crops are also applicable for amaranth. For instance, line selection from segregating breeding populations have been used as the main breeding strategy for long years, especially in the developing countries (Joshi *et al.*, 2018). On the other hand, making crosses is possible among the grain species as long as the genetic barriers that are driven by taxonomic restrictions permit, and up to 88 % biomass heterosis was reported in amaranth (Lehmann *et al.*, 1991), which demonstrates the potential applicability of hybrid breeding. Furthermore, the current availability of a cytoplasmic male

sterile seed stock (Brenner, 2019), and the dissection of its genetic basis in the future may pave the way for hybrid breeding in grain amaranths. Amaranth is a rare dicotyledonous crop that uses C_4 photosynthetic pathway. This allows a more economical water use compared to many C_3 plants and makes amaranth more tolerant to drought and eventually a good alternative for marginal and drylands (Liu and Stützel, 2004; Zhang *et al.*, 2018).

1.6. Amaranth for grain and biomass production in Europe

Amaranth is a short-day crop as many tropical plants, however, also displays a wide photoperiodic variation as its wide latitudinal distribution across the globe suggests (Wu *et al.*, 2000; Andini *et al.*, 2020). Similar to many short-day crops that extend vegetation period under the temperate climates, some amaranths (potentially photoperiod sensitive) also delay flowering and accumulate a high amount of biomass under the long-day conditions of Europe. In other words, amaranth has a potential for biomass cultivation under long-day conditions because of short-day genes, whereas photoperiod insensitive amaranths can be grown for grain production. Accordingly, the potential of amaranth for biomass production under temperate Europe conditions and the suitability of amaranth biomass as a biogas substrate was investigated in several studies (Balodis *et al.*, 2011; Seppälä *et al.*, 2013; von Cossel *et al.*, 2017). Some studies compared amaranth and maize for biomass and biogas related parameters including dry matter content and yield, and maize surpassed amaranth as expected due to the factors such as the long hybrid breeding history of maize and the presence of well-determined heterotic groups. Besides, some studies focused on the adaptation potential of amaranth for biomass and grain production under the diverse climatic conditions of Europe such as Denmark (Hoidal *et al.*, 2019) and Italy (Rivelli *et al.*, 2008) and showed that amaranth is a promising alternative crop that can diversify crop rotations and can be successfully grown under such contrasting environments.

1.7. Hybridus complex and the domestication history of the grain amaranth species

Three major grain amaranth species and their two wild putative ancestors (*A.hybridus L.* and *A. quitensis* Kunth) form the hybridus complex, which is a taxonomic sub-unit of *Amaranthus* subgenera. Among the grain species, *A. cruentus* originated from Central America and is characterized by photoperiod insensitivity with a wide geographical distribution (Brenner *et al.*, 2000). *A. hypochondriacus* is another Central American species, however, no specific information is available in the literature about its photoperiodic behavior. *A. caudatus* grows typically in the highlands of the Andean region of South America, which is characterized by high altitudes and respective low temperatures (Sauer, 1967; Brenner *et al.*, 2000). *A. hybridus* is mainly accepted as the common wild ancestor of the three major grain species since it shows the largest genetic variation within the hybridus complex and the largest geographical distribution from North to South America (Sauer, 1967; Kietlinski *et al.*, 2014; Stetter *et al.*, 2020). *A. quitensis* is another wild putative ancestor that shows distribution across South America, similar to *A. caudatus*. Therefore, its potential role in the evolution of *A. caudatus* has been still under discussion (Kietlinski *et al.*, 2014). The domestication history of grain amaranths has not been completely revealed. However, the factors such as the availability of high-quality reference genome and small genome size (~500 mbp) (Lightfoot *et al.*, 2017) along with the decreasing sequencing costs has been provided recent molecular evidence, which suggest that the independent evolution events of the three grain species from the wild putative ancestor *A. hybridus* seems the most probable scenario (Stetter and Schmid, 2017; Stetter *et al.*, 2020).

1.8. Genetic mapping of the useful variation

Dissection of the genetic architecture of the adaptive traits is a fundamental step to understand the genetics of adaptation and to introgress the useful alleles into the breeding programs. The recent advances in the omics technologies allow rapid production of high-quality data. Of those, the decreasing cost of sequencing permits a high amount of sequencing data (Zargar *et al.*, 2015) and the advancements in machine learning have been revolutionized the phenomics area (Esposito *et al.*, 2020). The principle used in the genetic mapping studies is based on the

detection of significant marker-trait associations and linkage-disequilibrium (LD) is the key concept in the mapping studies that indicates the tendency of the physically close markers to inherit together. Accordingly, mapping studies aim to find the significant marker-trait associations based on the LD between the causal variant(s) underlying the trait and the markers. Two major methods used in the mapping studies are linkage mapping (conventional QTL mapping) and association mapping (Genome-Wide Association) and they differ in the types of populations used.

Linkage mapping uses bi-parental or multi-parental populations which naturally represent a lower number of recombination events and a narrower scale of phenotypic variation limited with two or a few parents. A low number of recombination events is characterized by large linkage blocks, especially, in the F_2 generation of bi-parental mapping populations and associates these very large blocks with the causal variant(s) instead of a small and specific region, a term known as low resolution. To cope with this problem, several strategies can be used such as the production of recombinant inbred lines (RILs) by selfing the recombinants from F_2 generation until they reach uniformity, which would increase the number of recombination events with each selfing. Alternatively, multi-parental mapping populations can be used such as nested association mapping (NAM) or multi-parent advanced generation inter-cross (MAGIC) design to increase the number of recombination events and the phenotypic variation in the populations for increased statistical power in the detection of causal variants (Scott *et al.*, 2020)

Association mapping is the other mapping method and accommodates a large number of independent or relative individuals with different evolutionary histories and large number of recombination events, which increase the resolution of mapping. In addition, assembling a natural population for association mapping studies would be more practical than generating bi-parental or multi-parental populations in terms of time, complex crossing procedures and financial costs. However, association mapping has its own challenges such as population structure, multiple testing problem and minor allele frequencies (MAF).

Population structure problem occurs when there is a genetic separation between usually geographically distant sub-populations that display different phenotypic performances. In this case, the genetic markers that are not in LD with the causative variants are correlated to the sub-populations due the similar evolutionary histories of the individuals within the same sub-population. This may lead to a spurious association between the phenotypic performance of a sub-population and the neutral markers that are correlated with the sub-population, instead of the association between the markers and the causal variant(s). A commonly used solution of this problem is the use of linear mixed models accommodating the *Q* and *K* matrices, that takes the population structure – generally using the output of a principal component analysis – and the kinship – that is responsible for the similar phenotypic performances of the individuals stemmed from their genetic similarity – into account, respectively (Xu *et al.*, 2017). Multiple-testing problem is another challenge because of a large number of the performed tests, which results in the detection of a number false-positive variants. Approaches such as Bonferroni correction, permutations, and false discovery rates are used to cope with this issue, however, the decision of the correction method should be given according to the research objectives (Xu *et al.*, 2017). MAF is another common problem confronted in the association mapping studies and hampers the detection of a rare variant with a low allele frequency unless it has a major effect. However, rare variants are valuable sources in plant breeding and their detection is important (Xu *et al.*, 2017). Rare variants are hard to detect due to their low frequencies in association mapping populations. Alternatively, a bi-parental mapping population can be generated where one of the parents possesses that rare allele and the other does not, which would increase the frequency of the rare allele ($p=0.5$ in F_2 generation) and facilitates its detection (Bernardo, 2016). A detailed review on genetic mapping of the agricultural traits can be found in Xu *et al.* (2017).

1.9. Objectives of the study

Kulakow and Jain (1985) reported that three major genes control flowering time in amaranth, of which one gene is responsible for reduced vegetative growth and the other two genes are responsible for photoperiodic response. However, these adaptive genes have not been cloned

and characterized yet, which is of crucial importance for their introgression into the breeding programs using modern molecular tools such as marker-assisted selection and genomic selection. The superiority of maize over amaranth in biomass and biogas relevant parameters have been proven, but how can amaranth be improved by breeding for such an alternative use has not been investigated up to date. In most of the studies in amaranth, sources of variation such as latitudinal groups and species have been represented with a very low number of genotypes. Flowering time is an important evolutionary trait and the characterization of the hybridus complex for this trait can inform us about the amount and the distribution of the genetic variation, that can be harnessed in the breeding programs (Gaudinier and Blackman, 2019; Turner-Hissong *et al*., 2020).

The particular objectives of this study were to

(i) characterize phenotypic variation in the biomass yield components and assess the composition of this variation using quantitative genetics measures to determine selection strategies for improved dry matter yield,

(ii) characterize phenotypic variation in adaptive traits and environmental variables in terms of latitudinal groups and species, to understand how the environmental variables affect the phenotypic variation, and how these findings would contribute to the establishment of amaranth breeding programs and our knowledge about the domestication history and local adaptability of grain amaranths,

(iii) dissect the genetic basis of the adaptive traits such as photoperiod sensitivity, seed setting, flowering time and plant height to gain a better understanding of adaptation genetics of grain amaranths and their wild relatives.

We achieved these objectives using genomics approaches. We performed field experiments to produce phenotypic datasets and used the whole-genome sequencing (WGS) data of the accessions from the hybridus complex to map the adaptive traits and to study the diversity patterns between Central and South American accessions.

2. Breeding Amaranth for Biomass: Evaluating Dry Matter Content and Biomass Potential in Early and Late Maturing Genotypes[1]

Abstract

Amaranth (*Amaranthus* spp.) is a promising biomass crop for silage and biogas production. Under long-day conditions, it exhibits prolonged vegetative growth. To evaluate the breeding potential of amaranth for biomass production, we characterized phenotypic variation in biomass yield components, quantitative genetic parameters, and the relationships between traits. We conducted field trials of 10 biomass-type genotypes exhibiting a 'giant' growth habit derived from spontaneous hybridization between genetically diverse parents and used the variety "Bärnkrafft" as check. We observed two contrasting growth patterns: Bärnkrafft is a variety for grain production and was characterized by a short vegetative growth followed by a long seed ripening. In contrast, the biomass genotypes displayed a long vegetative growth followed by a short seed ripening. We observed strong correlations between dry matter content and stem diameter ($r = -0.78$, $p < 0.01$) and between plant height and biomass score ($r = 0.95$, $p < 0.001$). High values for broad-sense heritability of stem diameter ($H^2 = 0.88$) and plant height ($H^2 = 0.92$) suggest that the dry matter content and yield can be improved by indirect phenotypic selection. We hypothesize that selection for dry matter content and yield implies a trade-off between earliness and photoperiod sensitivity. Hence, dry matter content should be improved first by recurrent selection, which can be then combined with short-day genes to improve dry matter yield. Overall, this work provides an avenue to the breeding of biomass amaranth.

[1] This chapter is published as:
Baturaygil, A.; Stetter, M.G.; Schmid, K. Breeding Amaranth for Biomass: Evaluating Dry Matter Content and Biomass Potential in Early and Late Maturing Genotypes. *Agronomy* **2021**, *11*, 970. https://doi.org/10.3390/agronomy11050970

Keywords: amaranth, biomass, quantitative genetics, photoperiod sensitivity, dry matter yield.

2.1. Introduction

The production of bioenergy is an important component in efforts to reduce dependence on fossil fuels. One way to produce bioenergy is the anaerobic digestion of plant material in bioreactors and a consecutive conversion of the resulting biogas into electricity and heat through a generator (Weiland, 2010). In Germany, biogas production from energy crops has grown rapidly with 9,200 biogas plants that produce 4.2 GigaWatts as of 2016 (Myrna *et al.*, 2019). Maize silage is the most popular biogas substrate in Germany, with a mass-based contribution of about 70% among energy crops (Daniel-Gromke *et al.*, 2018). Given that methane yield is mainly determined by dry matter yield, high dry matter yield is the primary breeding objective in biogas crops (Grieder *et al.*, 2012; Herrmann and Rath, 2012). Maize has become the predominant biogas crop because it combines high dry matter yield and content (Balodis *et al.*, 2011). However, potential negative impacts of maize monoculture such as increased risk of soil erosion and a decrease in agrobiodiversity create a demand for alternative energy crops (Brauer-Siebrecht *et al.*, 2016; Vogel *et al.*, 2016).

Amaranth is a possible alternative bioenergy crop. The genus *Amaranthus* harbors more than 60 species, of which several species are cultivated as grain crops, leaf vegetables or ornamental plants (Brenner *et al.*, 2000). In Central and South America, grain amaranth species are ancient crops (Sauer, 1967; National Research Council, 1984) that have been rediscovered in the last decades due to their favorable nutritional qualities (Stallknecht and Schulz-Schaeffer, 1993). However, most *Amaranthus* species are undomesticated and weeds in agricultural production areas (Sauer, 1950; Costea and Tardif, 2004). Amaranths are C4 plants and therefore able to photosynthesize at high temperatures in conjunction with a higher water use efficiency than many C3 plants (Assad *et al.*, 2017). These characteristics facilitate the introduction of amaranth as a crop to dry and marginal zones (Myers, 1996; Liu and Stützel, 2004). In the temperate Central European climate amaranth showed better drought stress tolerance than maize in a comparison of both crops for biogas suitability (von Cossel *et*

al., 2017). In addition, amaranth can be used as forage crop (Sleugh *et al.*, 2001; Svirskis, 2009; Seguin *et al.*, 2013).

Amaranthus species differ in their photoperiodic response (Wu *et al.*, 2000), which can be utilized to develop varieties suitable for biomass production. Many amaranth species need a short daylight below 12 hours for flowering induction. Among the three cultivated grain amaranths (*A. caudatus* L., *A. cruentus* L. *and A. hypochondriacus* L.), *A. caudatus*, requires less than eight hours (Fuller, 1949; National Research Council, 1984; Assad *et al.*, 2017). In contrast, *A. cruentus* is the most photoperiod insensitive amaranth species (Brenner *et al.*, 2000). Under long-day conditions of Central Europe, short-day crops delay flowering and prolong biomass accumulation (Jung and Müller, 2009). Since most amaranths are short-day plants, amaranth was considered as potential biomass and biogas crop for cultivation in Europe.

Amaranth is mainly self-pollinating with an out-crossing rate between 3-32% (Jain *et al.*, 1982; Hauptli and Jain, 1985), which makes it an attractive species for plant breeding because it allows breeding methods used for both autogamous and allogamous crops. Even though up to 88% mid-parent biomass heterosis was observed in F_1 generation hybrids (Lehmann *et al.*, 1991), an efficient large-scale method for hybrid seed production does currently not exist for amaranth. However, methods for experimental crosses have been successfully applied (Stetter *et al.*, 2016) and the existence of cytoplasmic male sterile (CMS) line and the restorer line (*A. hypochondriacus* L.) may allow a large-scale production of F_1 seeds and may serve to exploit biomass heterosis commercially in the future (Brenner, 2019).

The biomass and biogas potential of amaranth was investigated in several studies (Rivelli *et al.*, 2008; Mursec *et al.*, 2009; Pospišil, 2009; Svirskis, 2009; Balodis *et al.*, 2011; Seppälä *et al.*, 2013; Sitkey *et al.*, 2013; von Cossel *et al.*, 2017). Comparative studies evaluated whether amaranths are competitive with maize as a bioenergy crop and revealed that maize is superior to amaranth due to its high performance in both dry matter yield and content. Such an advantage of maize is expected because it has been improved by long-running commercial hybrid breeding programs that utilized heterosis (Melchinger and Gumber, 1998; Duvick,

2005; Fu *et al.*, 2014). As a consequence, a large number of high yielding biogas type maize varieties have been released. In contrast, breeding efforts in grain amaranths have been restricted to the selection of individual genotypes from landrace populations (Joshi *et al.*, 2018). In vegetable amaranth, breeding efforts have been limited to the acclimatization of a small number of lines in India (Shukla *et al.*, 2003, 2004), but quantitative genetic parameters estimated in trials indicated a positive potential for future improvement of vegetable amaranth by breeding (Shukla *et al.*, 2006). Overall, a lack of breeding activities likely contributes to the current position of amaranths as minor crop.

Although amaranth genebank accessions and landraces were evaluated for their suitability as biomass crops and for biogas production, no variety for biomass production was released to date (von Cossel *et al.*, 2017). The necessity of additional breeding efforts to improve amaranth as potential biogas crop was recognized (von Cossel *et al.*, 2017). So far, no study investigated the plant breeding potential of amaranth as bioenergy crop and estimated quantitative-genetic parameters relevant for breeding. In this study we evaluate the potential of breeding for biomass amaranth by (i) characterizing phenotypic variation in biomass yield components, (ii) determining the components of phenotypic variation and detecting correlations between traits, and (iii) proposing a breeding strategy for amaranth with high dry matter yield. Our results suggest that amaranth could become a suitable addition to existing biomass crops by targeted breeding programs.

2.2. Materials and Methods

2.2.1. Plant material

We focused on ten genotypes from our biomass amaranth breeding pool, whose ancestors include putative F_1 generation hybrids derived from spontaneous outcrossing events between Bärnkrafft, Puerto Moutt (*A. cruentus*), C6 (*A. caudatus*) and Pastewny (*A. hybridus*) that occurred during field trials in 2012, as well as multiple genebank accessions cultivated with these four genotypes. These species are diploid, and their chromosome number is $2n = 32$

(Assad *et al.*, 2017). We selected putative hybrid plants based on a giant growth habit and excessive plant height in 2013. Seeds of selected individuals were sown in a greenhouse after harvest to obtain F_2 generation plants. In the next growing season of 2014, we tested F_2 generation seeds originating from the 2013 field trial, as well as F_3 generation seeds produced in a greenhouse, and 120 genebank accessions from which individuals with gigantic growth habit in a field trial were selected. From all these populations which we harvested seeds of putative biomass type individuals. Due to the large number of plants, they were not covered with bags, and for this reason outcrossing was possible that may contribute to heterogeneity in the next generation. Collected seeds were planted and the best ten individual plants were selected for intensive evaluation in repeated trials in 2015. Seeds of these individuals were used in our study. In addition, we included Bärnkrafft, which was the only amaranth variety registered in Germany at that time (Bundessortenamt-Deleted varieties, 2016) as check variety. Bärnkrafft is a grain amaranth variety and has a stable phenotype that has been selected for cultivation in Central European climates.

2.2.2. Experimental design and phenotypic evaluation of field trials

In 2016, we tested the eleven genotypes at the Heidfeldhof agricultural station (48° 42’ N, 395 m a.s.l.) and the Eckartsweier agricultural station (48° 52’ N, 140 m a.s.l.) of the University of Hohenheim. These two locations differ in the distribution of precipitation and temperature during the vegetation period. Eckartsweier is a suitable growth environment for amaranth due to its high temperature and Heidfeldhof was used for receiving higher precipitation (**Figure 1**). The soil type was silty loam in both locations. The field trial had a randomized complete block design with three blocks per location that each contained the eleven genotypes in individual plots. Double-row plots had a length of 5 m with a distance of 0.75 m between rows. Plots within blocks were separated by 75 cm and between blocks by 1 m distance. Each experiment was surrounded by a check variety to prevent border effects. Sowing and thinning were conducted manually by leaving 10 cm distance between plants. The two experiments were planted on 4 May 2016 and 9 May 2016 and harvested on 12 Oct. 2016 and 11 Oct. 2016 in Heidfeldhof and Eckartsweier, respectively. Weed control was carried out manually and

mechanically, and no irrigation or fertilization was applied. We recorded five biomass yield components: plant height (in cm), dry matter content (as the percentage) and stem diameter (in mm). Plant height was measured from the ground surface to the top of the inflorescence and was recorded ten times during the growing season at Heidfeldhof and eight times at Eckartsweier. For this, 15 plants were randomly selected at the young plant stage approximately a month after the sowing in each plot and were labeled. All further measurements were taken from these individuals. For plant height, only the last measurements taken at harvest time in each location were used in the statistical analysis. Stem diameter was measured from 10 cm above the ground surface with a caliper and recorded on the same 15 plants at harvest time. Dry matter content was estimated on five randomly selected individuals per plot at harvest time. Harvested plants were cut above the ground surface and their roots were left in the soil. A sample from the mixture of ground tissue from these five individuals was weighted (fresh weight). After drying in a ventilated oven at 110°C for 72 hours, samples were again weighed (dry weight). The ratio of dry weight to fresh weight was considered as dry matter content. We scored the plots visually for biomass and inflorescence volume at harvest time using the 1-9 scale, where 1 refers to the most inferior and 9 to the most superior performance based on the visual volume of the plots for the respective traits.

2.2.3. Statistical Analysis

Location-specific and adjusted genotype means were estimated for each trait with a linear and a linear mixed model, respectively. In the estimation of adjusted genotype means, only genotype effect was taken as fixed and all other effects as random, whereas all effects in both the linear and linear mixed models were taken as random in the estimation of variance components. A Friedman test was used to estimate genotype means in the location-specific analyses of the score traits (Bewick *et al.*, 2004). In the linear model, linear mixed models and Friedman test, the significance of the genotype effect was evaluated with a type 3 test of fixed effects. Pairwise comparisons of genotype means were conducted with Fisher's least significant difference (LSD) test at a significance level of 0.05.

We used the following linear mixed model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + b_{jk} + e_{ijk} \quad (1)$$

where y_{ijk} is the observation value of response variable obtained from *i*-th genotype in *k*-th block in *j*-th location, μ is the overall mean, α_i the effect of *i*-th genotype, β_j the effect of *j*-th location, $(\alpha\beta)_{ij}$ the interaction of the *i*-th genotype and *j*-th location, b_{jk} the effect of *k*-th block nested within *j*-th location and e_{ijk} the error associated with y_{ijk}.

The linear model was

$$y_{ij} = \mu + b_j + \alpha_i + e_{ij} \quad (2)$$

where y_{ij} is the observation value of response variable obtained from *i*-th genotype in *j*-th block, μ is the overall mean, b_j is the effect of *j*-th block, α_i is the effect of *i*-th genotype and e_{ij} the error associated with y_{ij}.

Broad-sense heritability H^2 was calculated as (Piepho and Möhring, 2007)

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gxe}^2}{m} + \frac{\sigma^2}{rm}} \quad (3)$$

and plot-based repeatability as (Pucher *et al.*, 2015)

$$w = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma^2}{r}} \quad (4)$$

where σ^2_g is the genetic variance, σ^2_{gxe} is the genotype by environment interaction variance, σ^2 is the residual error variance, *m* is the number environments and *r* is the number of replicates per environment.

Relationships between traits were studied based on adjusted genotype means using Pearson's correlation. We performed statistical analyses with linear mixed and linear models, and estimations of variance components using the MIXED procedure of SAS 9.4. (SAS Institute Inc, 2013). The SAS/IM macro %MULT, developed by Piepho (2012) was used for ensuring the robust notation of the significant differences among the pairwise comparisons in linear mixed and linear model analyses. The Friedman test was performed with the agricolae package (de Mendiburu, 2015) and the correlation analysis was performed using the GGally package of the R statistical environment (Schloerke *et al.*, 2014).

2.3. Results

2.3.1. Differences between biomass types and grain types

For several quantitative and morphological traits, we observed heterogeneity within lines due to residual genetic segregation (**Figure 2**). Lines with residual heterogeneity were most easily recognized in qualitative color traits such as inflorescence, leaf, and stem color, whereas plant height and flowering time were quantitative traits with a high heterogeneity. To minimize the effect of variation within lines, we did not take averages for each plot, but collected observations from 15 randomly selected individuals per plot. The mean number of individuals recorded per plot was 13.70 (SD: 1.66) for plant height, and 13.85 (SD: 1.33) for stem diameter, respectively. Losses were mainly caused by insufficient emergence of two or three genotypes at Heidfeldhof and rarely by plant lodging, which was equally distributed over plots.

The extent of heterogeneity varied among genotypes. The biomass genotypes differed in their heterogeneity e.g., were more heterogeneous in quantitative traits such as plant height and stem diameter (**Figure 2A and B**) than Bärnkrafft, which was essentially uniform. Overall, the grain type variety Bärnkrafft differed strongly from the biomass genotypes in all three quantitative traits and was characterized by a lower plant height, smaller stem diameter, but higher dry matter content (**Figure 2A-C**). In addition, Bärnkrafft reached its final height after

approximately 100 days in contrast to the other genotypes, which spent a longer time in the vegetative growth phase before switching to the generative growth phase (**Figure 2D**). Due to these differences, we conducted the following statistical analyses both with including and without Bärnkrafft to evaluate the effect of this distinct variety on parameter estimates.

2.3.2. Variation in biomass yield components

We estimated ranges for trait values based on adjusted genotype means of two locations (**Table 1 and File S1**). A wide phenotypic range was obtained for plant height (109 to 253 cm) and stem diameter (14 to 23 mm), and a narrower range for dry matter content (19 to 24%). The traits with scores (biomass, inflorescence volume) were normally distributed in the joint analysis and therefore directly used without data transformation. Biomass score was highly variable between genotypes (1.50 to 8.17), but the range was narrower for inflorescence volume score (3.67 to 7.67).

In the joint analyses of both locations, Bärnkrafft reached the highest dry matter content in the experiment, and it was different than other genotypes, except a single genotype (*p=0.0227,* **File S1**), whereas we observed no difference among biomass genotypes when Bärnkrafft was excluded (*p=0.0966*). In the location-specific analyses for dry matter content, Bärnkrafft was also different from the other genotypes (**Table S1, File S1**). However, the biomass genotypes were not different from each other in Heidfeldhof, but were different from each other in Eckartsweier, when we performed the analyses without Bärnkrafft (**Table S1, File S1**).

Phenotypic trait means differed between the two locations. Larger trait means were estimated in Heidfeldhof for dry matter content and in Eckartsweier for plant height in location-specific analyses (**Table S1**). Since the traits biomass and inflorescence volume score were not normally distributed in location-specific analyses, we performed a non-parametric Friedman test to compare genotype means and found genotypes to be different in both traits and in both locations (**Table S1**). However, a Friedman test did not allow to estimate location-specific trait means, as it is a rank-based test and characterized by sample size. Therefore, we

estimated median and interquartile range in these traits for the location-specific trait comparisons (**Table S1**).

2.3.3. Quantitative genetics parameters and relationships between traits

We estimated two different broad-sense heritability and plot-based repeatability values for each quantitative trait both with and without Bärnkrafft (**Table 1**). Heritabilities were generally high but showed lower values when Bärnkrafft was excluded. The lowest heritability was estimated for dry matter content ($H^2 = 0.74$). Similarly, large values were obtained for plot-based repeatability (**Table 1**). Like heritability estimates, repeatability values were lower when Bärnkrafft was excluded. Dry matter content showed the biggest difference in repeatability with a value of 0.78 with and 0.44 without Bärnkrafft in the Heidfeldhof field trial.

We also correlated trait values based on the adjusted genotype means by including and excluding Bärnkrafft (**Figure 3**). The correlation between biomass score and plant height and between dry matter content and stem diameter were significant in both analyses. There was a strong positive correlation between biomass score and plant height with ($r = 0.95, p < 0.001$) and without Bärnkrafft ($r = 0.88\ p < 0.001$). Dry matter content and stem diameter were negatively correlated with ($r = -0.78, p < 0.01$) and without Bärnkrafft ($r = -0.72, p < 0.05$). Dry matter content was negatively correlated with plant height ($r = -0.71, p < 0.05$) and biomass score ($r = -0.71\ p < 0.05$). Finally, plant height showed a positive correlation with stem diameter ($r = 0.64, p < 0.05$) when Bärnkrafft was included.

2.4. Discussion

2.4.1. Variation in biomass yield components

We observed a strong difference in average dry matter content between the ten biomass lines (average 21.7 %) and the single grain type variety Bärnkrafft (24.5 %, **File S1**). A higher dry

matter content in Bärnkrafft is expected because of its earlier maturity. This difference explained also the significance in genotype effect for dry matter content in the joint analysis of all 11 genotypes and the non-significance in the 10 biomass genotypes. Dry matter content should be at least 28 % for a satisfactory ensiling process in biogas production (Herrmann and Rath, 2012). In our study, the grain type variety Bärnkrafft reached 26.6 % dry matter content at the Heidfeldhof site and von Cossel *et al*. (2017) reported up to 27.6 %. These results suggest that amaranth has the potential to meet this requirement by further breeding. Furthermore, we note that amaranth has smaller particle sizes for seed and chaff than maize. Franco *et al*. (2016) suggested that a smaller particle size improves methane yield production and a dry matter content threshold of 28% for maize may be lower for amaranth.

We analyzed inflorescence volume score as an indirect measure of grain yield and found significant genotype effect when Bärnkrafft was both included and excluded. This significance presumably originated from variation in number of days required to reach physiological maturity among biomass genotypes. We also found genotypes to be different in the traits related to vegetative growth such as plant height and stem diameter probably because genotypes had a longer time until harvest to demonstrate their differences. Although we could not obtain dry matter yield values for comparison, a highly significant genotype effect in biomass score is a positive indicator for the improvement of dry matter yield (**Table 1**).

2.4.2. Trade-off between earliness and photoperiod sensitivity

We did not phenotype flowering time in our study due to residual segregation for this trait, however, the time point at which genotypes achieve a constant plant height can be used as a proxy for flowering time, when plants switch from vegetative to generative growth (Li and Xu, 2017). According to this definition, the difference in days for the beginning of flowering is quite large between Bärnkrafft and the biomass genotypes, but much smaller among the ten biomass genotypes (**Figure 2**). We suggest that the differences in dry matter content and plant height between Bärnkrafft and the biomass genotypes are mainly caused by variation in flowering time, as early flowering leads longer seed ripening phase and improved dry matter

content, whereas late flowering due to short-day genes leads longer vegetative growth and higher plant height.

Although the exact parents of biomass genotypes are unknown, they include *A. caudatus* and *A. cruentus* accessions, which are photoperiod sensitive and insensitive, respectively, that were involved in spontaneous crossing events (Fuller, 1949; Brenner *et al.*, 2000; Assad *et al.*, 2017). Therefore, we hypothesize that short-day genes were introgressed into the biomass genotypes that lead to prolonged vegetative growth of the biomass genotypes by a combination of two factors: (i) the presence of short-day genes responsible for photoperiodic response and (ii) long-day conditions during the growth phase that caused delayed flowering in the presence of short-day genes (Kulakow and Jain, 1985; Wu *et al.*, 2000; Jähne *et al.*, 2020). According to their genetic architecture of flowering time, individuals delayed or completely withheld flowering and continued to accumulate biomass throughout the cultivation period. Therefore, a widely known pattern in energy crops – delayed flowering leading to higher biomass yield – appears to hold true in amaranths as in other crops (Fernandez *et al.*, 2009; Jung and Müller, 2009; Schwartz *et al.*, 2010; Jensen *et al.*, 2012).

However, composition of dry matter yield i.e., the contribution of grains to total dry matter yield is of crucial importance to secure sufficient dry matter content. In forage maize, cob to total dry matter yield ratio is around 50% (Lynch *et al*., 2010), and the main breeding objective is digestibility, which is determined by dry matter content (Grieder *et al*., 2012). It promotes the use of earliness genes for higher grain yield and restricts the use of short-day genes. In contrast, the main objective of biogas maize breeding is high dry matter yield, and the exploitation of short-day genes is more flexible, provided that 28 % dry matter content is secured (Grieder *et al*., 2012). Similarly, in biomass sorghum, use of short-day genes increases dry matter yield, but a low dry matter content requires to prioritize increased panicle contribution to total dry matter yield. Windpassinger *et al*. (2015) propose the use of silage type sorghum for biogas production, whose ratio of panicle to total dry matter yield ranges

between 40-50 %, and therefore outperforms biomass type sorghum in methane yield and dry matter content.

Similar to the situation in these grass species, our study indicates a trade-off between earliness genes, for increased dry matter content, and short-day genes, for increased dry matter yield. We therefore propose that variation in flowering time is required to select for both earliness and photoperiod-sensitivity. Selection for a defined flowering time and the introduction of short-day genes may improve dry matter yield. More specifically, a flowering time interval within dry matter yield is maximized should be determined, since the inclusion of short day genes may likely influence such an interval across a set of environments. To achieve maximum dry matter yield, the optimal inflorescence to biomass ratio and its dependence on flowering time have to be evaluated. As a first step, selection for improved dry matter content should be prioritized and short-day genes can subsequently used to improve dry matter yield, provided the requirement of a high dry matter content is fulfilled.

2.4.3. Quantitative genetics parameters and relationships between traits

In the first step, a suitable base population for dry matter content improvement can be generated by using photoperiod insensitive *A. cruentus* accessions, which have a panicle to total dry matter yield ratio of around 50 % (Rivelli *et al.*, 2008). In the second step, high dry matter content can be combined with short-day genes to improve dry matter yield, by making crosses between photoperiod insensitive and sensitive genotypes. Accordingly, genotypes combining high dry matter content and prolonged vegetative growth can be selected from such populations with a large segregation variance. A similar approach succeeded in an energy maize breeding program in Germany, by combining photoperiod sensitivity genes from exotic Peruvian and Mexican populations, high grain yield potential from Italian populations and cold-tolerance genes from German populations (Schmidt, 2003). In our study, stem diameter and plant height were highly heritable and also were strongly correlated to dry matter content and yield, respectively (**Figure 3**). Therefore, these traits seem promising to be used in an

indirect phenotypic selection of the target traits. However, future studies should re-examine the selection efficiency of these traits by testing more grain type genotypes. Consistent with our study, moderate to strong positive correlations between plant height and dry matter yield (*r=0.81* and *0.71*) were also reported in biogas maize and sorghum, respectively (Grieder *et al.*, 2012; Windpassinger *et al.*, 2015).

The residual heterogeneity within plots was also a source of genetic variance but is explained by the residual error term in the mixed model. This heterogeneity may cause an underestimation of genetic variance and an overestimation of the residual error term, which then results in an underestimation of broad-sense heritability and plot-based repeatability parameters. Since the residual error variance is larger than the genetic variance for dry matter content, heritability and repeatability may be underestimated for this trait, but its effect on genetic variance cannot be estimated with our design. Furthermore, the execution of multi-environment field trials across several years with a higher number of target environments would allow more accurate parameter estimations in future studies.

2.4.4. Future Prospects

Future breeding efforts in biomass amaranth should primarily address the genetic characterization of flowering time and photoperiod sensitivity because of the role of variation in the trade-off of these two traits. Such a goal can be achieved because of the availability of a high-quality reference genome (*A. hypochondriacus* L.) and crossing methods to generate mapping populations (Lehmann, 1995; Brenner and Widriechner, 1998; Stetter *et al.*, 2016; Lightfoot *et al.*, 2017). Here, we focused on primary biomass traits, but traits like ligno-cellulose, sugar, protein and lipid content as well as nutrients and trace elements can be alternative selection targets for optimized biochemical composition of biomass amaranths (Lübken *et al.*, 2010). Overall, application of novel breeding methods such as genomic selection combined with speed breeding may rapidly improve the selection gain in the desired traits and promote the use of this minor crop as a resilient alternative to current biomass crops

that is suitable for cultivation in marginal areas and thereby reduces competition for food and feed.

Author Contributions

AB, MGS and KS designed the experiment. MGS developed the genotypes. AB collected and analyzed the data. AB and KS wrote the manuscript. All authors edited, read and approved the final manuscript.

Funding

This work was supported by the F. W. Schnell endowed Professorship of the Stifterverband and the Hohenheim-Tübingen Regional Alliance of the Ministry of Science and Culture (MWK) of Baden Württemberg.

Acknowledgments

We thank Viola Abraham and the staff of the Hohenheim experimental stations for help with the field experiments and Hans-Peter Piepho for statistical advice.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data Availability Statement

The phenotypic data and the supplementary files including File S1 are available from Figshare (10.6084/m9.figshare.c.5421621).

2.5. References

Assad R, Reshi ZA, Jan S and Rashid I (2017) Biology of Amaranths. *Botanical Review* 83: 382–436.

Balodis O, Bartuševics J and Gaile Z (2011) Biomass yield of different plants for biogas production. *Vide, Tehnologija, Resursi - Environment, Technology, Resources* 1: 238–245.

Bewick V, Cheek L and Ball J (2004) Statistics review 10: Further nonparametric methods. *Critical Care* 8: 196–199.

Brauer-Siebrecht W *et al*. (2016) Silage Maize and Sugar Beet for Biogas Production in Rotations and Continuous Cultivation: Dry Matter and Estimated Methane Yield. *Agronomy* 6: 2.

Brenner DM *et al*. (2000) Genetic resources and breeding of *Amaranthus*. *Plant Breeding Reviews* 19: 227–285.

Brenner DM (2019) Registration of DB 199313, Cytoplasmic Male Sterile Grain Amaranth Genetic Stock. *Journal of Plant Registrations* 0: 0.

Brenner DM and Widriechner MP (1998) *Amaranthus* seed regeneration in plastic tents in greenhouses. *Plant Genetic Resources Newsletter* 1–4.

Bundessortenamt-Deleted varieties (2016) See http://www.bundessortenamt.de/internet30/index.php?id=141 (accessed 18 November 2016).

von Cossel M, Möhring J, Kiesel A and Lewandowski I (2017) Methane yield performance of amaranth (*Amaranthus hypochondriacus* L.) and its suitability for legume intercropping in comparison to maize (*Zea mays* L.). *Industrial Crops and Products* 103: 107–121.

Costea M and Tardif FJ (2004) The biology of Canadian weeds. 130. Amaranthus retroflexus L., *A. powellii* S. Watson and *A. hybridus* L. (update). *Canadian Journal of Plant Science* 84: 631–668.

Daniel-Gromke J *et al*. (2018) Current Developments in Production and Utilization of Biogas and Biomethane in Germany. *Chemie Ingenieur Technik* 90: 17–35.

Duvick DN (2005) The Contribution of Breeding to Yield Advances in maize (*Zea mays* L.). *Advances in agronomy* 86: 83–145.

Fernandez MG., Becraft PW, Yin Y and Lübberstedt T (2009) From dwarves to giants? Plant height manipulation for biomass yield. *Trends in Plant Science* 14: 454–461.

Fu D *et al*. (2014) Utilization of crop heterosis: A review. *Euphytica* 197: 161–173.

Fuller HJ (1949) Photoperiodic Responses of *Chenopodium quinoa* Willd. and *Amaranthus caudatus* L . *American Journal of Botany* 36: 175–180.

Grieder C, Dhillon BS, Schipprack W and Melchinger AE (2012) Breeding maize as biogas substrate in Central Europe: II. Quantitative-genetic parameters for inbred lines and correlations with testcross performance. *Theoretical and Applied Genetics* 124: 981–988.

Hauptli H and Jain S (1985) Genetic variation in outcrossing rate and correlated floral traits in a population of grain amaranth (*Amaranthus cruentus* L.). *Genetica* 66: 21–27.

Herrmann A and Rath J (2012) Biogas Production from Maize: Current State, Challenges, and Prospects. 1. Methane Yield Potential. *Bioenergy Research* 5: 1027–1042.

Jähne F, Hahn V, Würschum T and Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theoretical and Applied Genetics*.

Jain SK, Hauptli H and Valdya KR (1982) Outcrossing rate in grain amaranths. *Journal of Heredity* 73: 71–72.

Jensen E *et al*. (2012) Flowering induction in the bioenergy grass *Miscanthus sacchariflorus* is a quantitative short-day response, whilst delayed flowering under long days increases biomass accumulation. *Journal of Experimental Botany* 64: 541–552.

Joshi DC *et al*. (2018) From zero to hero: the past, present and future of grain amaranth breeding. *Theoretical and Applied Genetics* 131: 1–17.

Jung C and Müller AE (2009) Flowering time control and applications in plant breeding. *Trends in Plant Science* 14: 563–573.

Kulakow PA and Jain SK (1985) The inheritance of flowering time in Amaranthus species. *Journal of Genetics* 64: 85–100.

Lehmann JW (1995) Rapid cycling of grain amaranths. *Legacy* 8: 15–17.

Lehmann JW, Clark RL and Frey KJ (1991) Biomass Heterosis and Combining Ability in Interspecific and Intraspecific Matings of Grain Amaranths. *Crop science* 31: 1111–1116.

Lemenkova P (2019) Regression Models by GRETL and R Statistical Packages for Data Analysis in Marine Geology. *International Journal of Environmental Trends* 3: 39–59.

Li Y and Xu M (2017) *CCT* family genes in cereal crops: A current overview. *Crop Journal* 5: 449–458.

Lightfoot DJ *et al*. (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology* 15: 1–15.

Liu F and Stützel H (2004) Biomass partitioning, specific leaf area, and water use efficiency of vegetable amaranth (*Amaranthus* spp.) in response to drought stress. *Scientia Horticulturae* 102: 15–27.

Lübken M, Gehring T and Wichern M (2010) Microbiological fermentation of lignocellulosic biomass: Current state and prospects of mathematical modeling. *Applied Microbiology and Biotechnology* 85: 1643–1652.

Melchinger AE and Gumber RK (1998) Overview of Heterosis and Heterotic Groups in Agronomic Crops. *Concepts and Breeding of Heterosis in Crop Plants.* pp. 29–44.

de Mendiburu F (2015) Statistical Procedures for Agricultural Research.

Mursec B, Vindis P, Janzekovic M, Brus M and Cus F (2009) Analysis of different substrates for processing into biogas Manufacturing and processing. *Journal of Achievements in Materials and Manufacturing Engineering* 37: 652–659.

Myers RL (1996) Amaranth: New crop opportunity. *Progress in new crops* 207–220.

Myrna O, Odening M and Ritter M (2019) The Influence of Wind Energy and Biogas on Farmland Prices. *Land* 8.

National Research Council (1984) *Amaranth Modern Prospects for an Ancient Crop*. National Academy Press.

Piepho H-P (2012) A SAS macro for generating letter displays of pairwise mean comparisons. *Communications in Biometry and Crop Science* 13: 456–466.

Piepho H-P and Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. *Genetics* 177: 1881–1888.

Pospišil A (2009) Yield and Quality of Forage Sorghum and Different Amaranth Species (*Amaranthus* spp.) Biomass. *Agriculturae Conspectus Scientificus* 74: 85–89.

Pucher A, Høgh-jensen H, Gondah J, Hash CT and Haussmann BIG (2015) Micronutrient Density and Stability in West African Pearl Millet — Potential for Biofortification. 1709–1720.

Rivelli AR, Gherbin P, de Maria S and Pizza S (2008) Field evaluation of *Amaranthus* species for seed and biomass yields in southern Italy. *Italian Journal of Agronomy* 3: 225–229.

SAS Institute Inc (2013) Base SAS® 9.4 Procedures Guide: Statistical Procedures, Second Edition.

Sauer JD (1950) The Grain Amaranths : A Survey of Their History and Classification. *Annals of the Missouri Botanical Garden* 37: 561–632.

Sauer JD (1967) The grain amaranths and their relatives: A revised taxonomic and geo-graphic survey. *Ann. Missouri Bot. Gard.* 54: 102-137.

Schloerke B, Crowley J, Cook D and Hofmann H (2014) Ggally: Extension to ggplot2. R package version 0.5. 0.

Schmidt W (2003) Hybridmaiszüchtung bei der KWS SAAT AG. In: Bericht über die 54. Tagung 2003 der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs. *BAL Gumpenstein* 1–6.

Schwartz CJ *et al*. (2010) Natural variation of flowering time and vernalization responsiveness in *Brachypodium distachyon*. *Bioenergy Research* 3: 38–46.

Seguin P, Mustafa AF, Donnelly DJ and Gélinas B (2013) Chemical composition and ruminal nutrient degradability of fresh and ensiled amaranth forage. *Journal of the Science of Food and Agriculture* 93: 3730–3736.

Seppälä M, Laine A and Rintala J (2013) Screening of novel plants for biogas production in northern conditions. *Bioresource Technology* 139: 355–362.

Shukla S *et al*. (2003) Nutritional contents of different foliage cuttings of vegetable amaranth. *Plant Foods for Human Nutrition* 58: 1–8.

Shukla S, Bhargava A, Chatterjee A and Singh SP (2004) Estimates of genetic parameters to determine variability for foliage yield and its different quantitative and qualitative traits in vegetable amaranth (*A. tricolor*) [India]. *Journal of Genetics and Breeding* 169–176.

Shukla S, Bhargava A, Chatterjee A, Srivastava A and Singh SP (2006) Genotypic variability in vegetable amaranth (*Amaranthus tricolor* L.) for foliage yield and its contributing traits over successive cuttings and years. *Euphytica* 151: 103–110.

Sitkey V, Gaduš J, Kliský Ľ and Dudák A (2013) Biogas Production From Amaranth Biomass. *Acta Regionalia et Environmentalica* 10: 59–62.

Sleugh BB *et al*. (2001) Forage nutritive value of various amaranth species at different harvest dates. *Crop Science* 41: 466–472.

Stallknecht GF and and Schulz-Schaeffer JR (1993) Amaranth Rediscovered. *In: J. Janick and J.E. Simon (eds.)* 211–218.

Stetter MG *et al*. (2016) Crossing Methods and Cultivation Conditions for Rapid Production of Segregating Populations in Three Grain Amaranth Species. *Frontiers in plant science* 7.

Svirskis A (2009) Prospects for non-traditional plant species cultivated for forage in Lithuania. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 37: 215–218.

Teixeira Franco R, Buffière P and Bayard R (2016) Ensiling for biogas production: Critical parameters. A review. *Biomass and Bioenergy* 94: 94–104.

Vogel E, Deumlich D and Kaupenjohann M (2016) Bioenergy maize and soil erosion - Risk assessment and erosion control concepts. *Geoderma* 261: 80–92.

Weiland P (2010) Biogas production: Current state and perspectives. *Applied Microbiology and Biotechnology* 85: 849–860.

Windpassinger S, Friedt W, Frauen M, Snowdon R and Wittkop B (2015) Designing adapted sorghum silage types with an enhanced energy density for biogas generation in temperate Europe. *Biomass and Bioenergy*.

Wu H *et al*. (2000) Field evaluation of an *Amaranthus* genetic resource collection in China. *Genetic Resources and Crop Evolution* 47: 43–53.

FIGURES

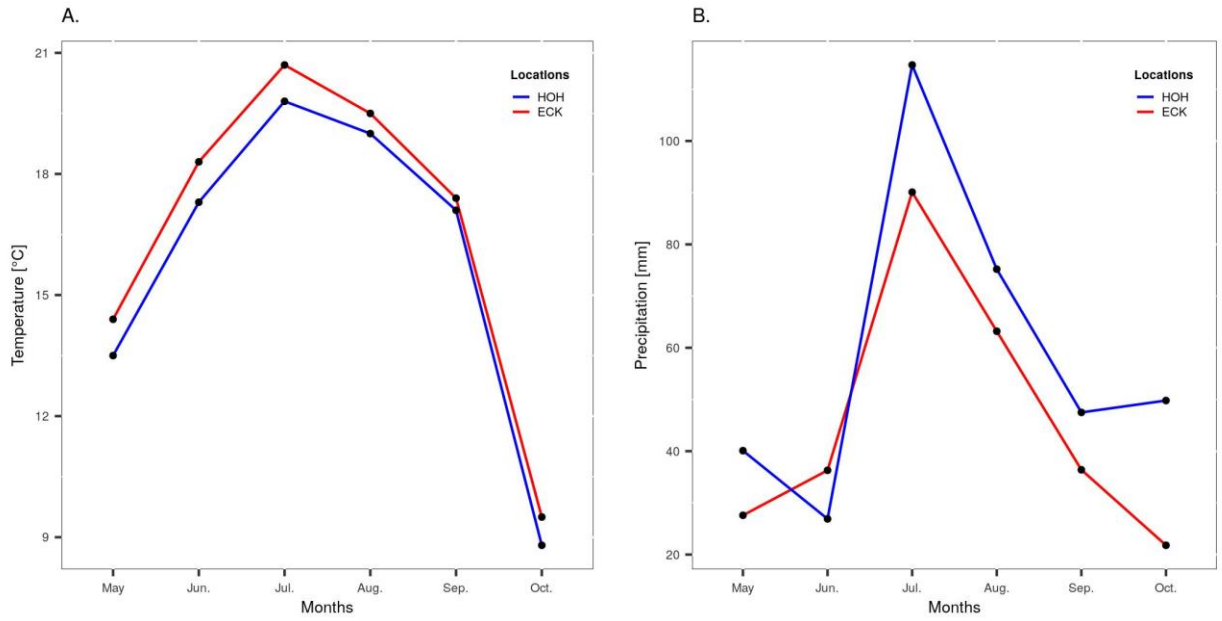

Figure 1. Monthly mean values of (**A**) temperature (°C) and (**B**) precipitation (mm) belong to the experimental locations between May–October 2016 (Agrarmeteorologie Baden-Württemberg, 2016).

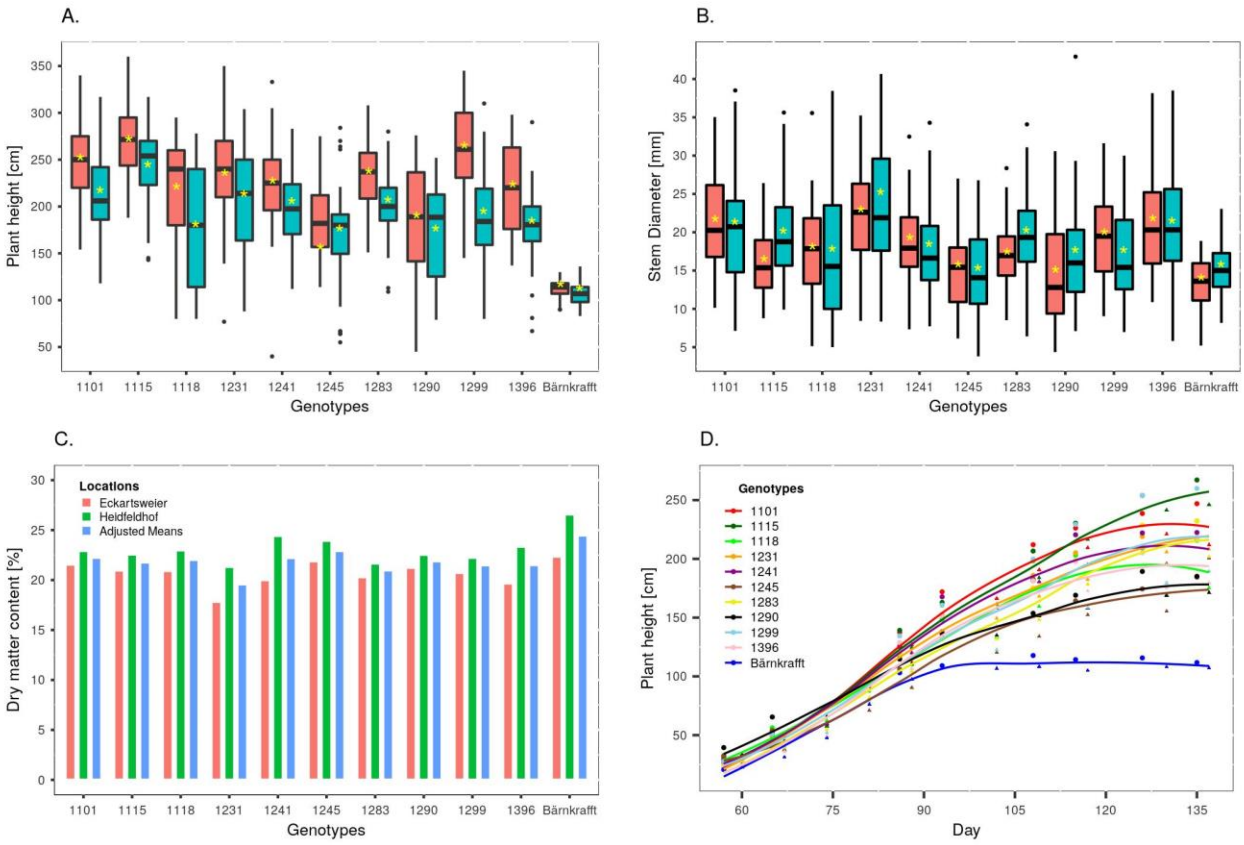

Figure 2. Box plots of genotypes for **(A)** plant height and **(B)** stem diameter, based on individual plant observations across three blocks per location. Red and turquoise colors indicate Eckartsweier and Heidfeldhof, respectively. Yellow asteriks indicate location-specific least square means of the genotypes. **(C)** Bar plot of dry matter content. Red and green colors indicate location-specific least square means of the genotypes from Eckartsweier and Heidfeldhof, respectively. Blue color indicates the adjusted least square means estimated over both locations. **(D)** Repeated measurements of plant height based on both locations. Dot and triangle point shapes represent Eckartsweier and Heidfeldhof, respectively. Each observation point indicates the mean value of a genotype based on the labeled individuals over three blocks, whereas the corresponding value of an observation point on the X axis indicates the observation day. The graph was generated using local weighted polynomial regression model (Lemenkova, 2019).

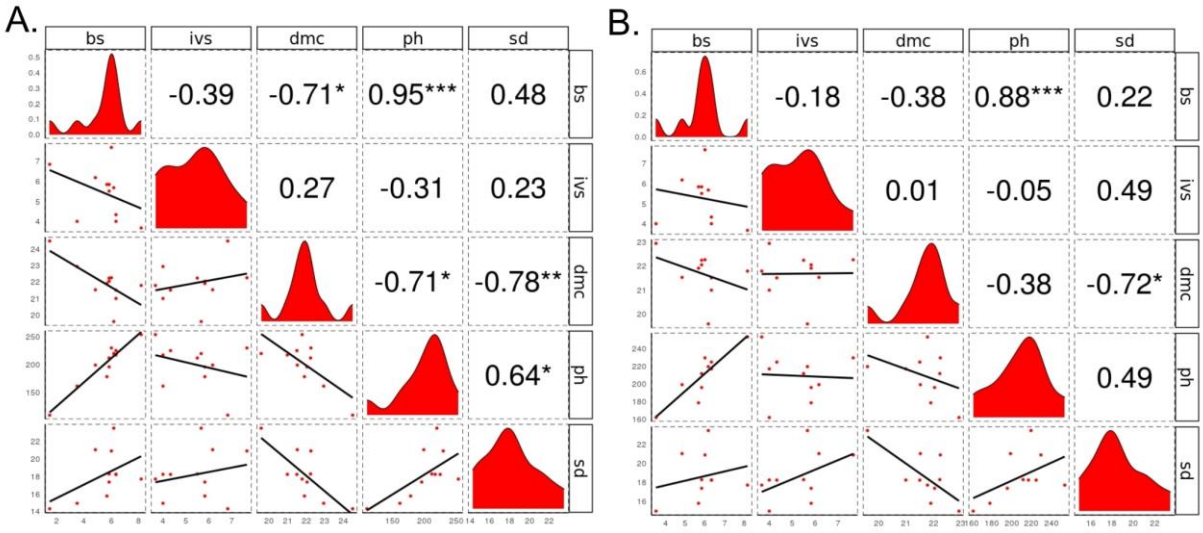

Figure 3. Pearson's correlation coefficients between biomass yield components, based on the adjusted genotype means over two locations. Bärnkrafft is (**A**) included and (**B**) excluded. Traits are bs, biomass score; ivs, inflorescence volume score; dmc, dry matter content; ph, plant height and sd, stem diameter. $^{*, **, ***}$ significant at $p<0.05$, $p<0.01$ and $p<0.001$ respectively.

TABLES

Table 1. Trait means, associated standard errors and ranges based on adjusted genotype means, mixed model analyses (Type 3 tests of fixed effects) with biomass yield components as dependent variables and genotype as fixed effect, and broad-sense heritability and plot-based repeatability values estimated in Heidfeldhof and Eckartsweier, including and excluding Bärnkrafft.

	Trait values (Including Bärnkrafft)						Type 3 tests for genotype effect (Including Bärnkrafft)				Broad-sense heritability and plot-based repeatability (Including Bärnkrafft)		
Traits	Mean	±	SE	Range			Num DF	Den DF	F-Value	Pr>F	H^2	w_H	w_E
Biomass score	5.47	±	0.52	1.50	-	8.17	10	10	13.51	0.0002***	0.93	0.91	0.97
Inflorescence volume score	5.41	±	0.36	3.67	-	7.67	10	10	6.08	0.0043**	0.84	0.74	0.76
Dry matter content (%)	21.93	±	0.38	19.59	-	24.48	10	10	3.83	0.0227*	0.74	0.78	0.79
Plant height (cm)	199.90	±	11.80	109.46	-	253.36	10	10	11.87	0.0003***	0.92	0.90	0.95
Stem diameter (mm)	18.18	±	0.83	14.29	-	23.46	10	10	8.23	0.0013**	0.88	0.78	0.80
	Trait values (Excluding Bärnkrafft)						Type 3 tests for genotype effect (Excluding Bärnkrafft)				Broad-sense heritability and plot-based repeatability (Excluding Bärnkrafft)		
Traits	Mean	±	SE	Range			Num DF	Den DF	F-Value	Pr>F	H^2	w_H	w_E
Biomass score	5.87	±	0.37	3.50	-	8.17	9	9	5.94	0.007**	0.83	0.82	0.93
Inflorescence volume score	5.27	±	0.40	3.67	-	7.67	9	9	5.60	0.0086**	0.82	0.74	0.72
Dry matter content (%)	21.67	±	0.28	19.59	-	22.93	9	9	2.47	0.0966ns	0.60	0.44	0.75
Plant height (cm)	208.95	±	8.37	161.36	-	253.36	9	9	5.41	0.0096**	0.82	0.75	0.89
Stem diameter (mm)	18.57	±	0.81	14.91	-	23.46	9	9	6.63	0.0047**	0.85	0.75	0.74

*, **, *** significant at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively, whereas ns shows non-significance.
w_H and w_E reprsesents plot-based repeatability values estimated in Heidfeldhof and Eckartsweier, respectively.

SUPPLEMENTARY TABLES

Table S1: Location-specific variance analyses (Type 3 tests of fixed effects) for dry matter content, plant height and stem diameter and Friedman tests for biomass and inflorescence volume score, with biomass yield components as dependent variables and genotype as fixed effect, including and excluding Bärnkrafft. Locations are Heidfeldhof and Eckartsweier, respectively.

Including Bärnkrafft

	Heidfeldhof							Eckartsweier						
	Type 3 Tests for Genotype Effect				Descriptive Statistics			Type 3 Tests for Genotype Effect				Descriptive Statistics		
Traits	Num DF	Den DF	F-Value	Pr>F	Mean	±	SE	Num DF	Den DF	F-Value	Pr>F	Mean	±	SE
Dry matter content (%)	10	20	4.51	0.002**	23.16	±	0.44	10	20	4.70	0.0016**	20.69	±	0.37
Plant height (cm)	10	20	10.27	<.0001***	186.92	±	10.14	10	20	19.06	<.0001***	212.89	±	14.10
Stem diameter (mm)	10	20	4.25	0.0029**	18.56	±	0.86	10	20	5.02	0.0011**	17.79	±	0.90

	Friedman Tests for Genotype Effect			Friedman Tests for Genotype Effect		
	DF	Chi-sq	P-Value	DF	Chi-sq	P-Value
Biomass score	10	23.0547	0.0105466*	10	27.431	0.002225102**
Inflorescence volume score	10	20.1911	0.0274963*	10	18.935	0.04109846*

Excluding Bärnkrafft

	Heidfeldhof							Eckartsweier						
	Type 3 Tests for Genotype Effect				Descriptive Statistics			Type 3 Tests for Genotype Effect				Descriptive Statistics		
Traits	Num DF	Den DF	F-Value	Pr>F	Mean	±	SE	Num DF	Den DF	F-Value	Pr>F	Mean	±	SE
Dry matter content (%)	9	18	1.78	0.1418ns	22.82	±	0.30	9	18	3.78	0.0079**	20.52	±	0.37
Plant height (cm)	9	18	3.97	0.0062**	194.90	±	6.92	9	18	9.44	<.0001***	223.00	±	10.87
Stem diameter (mm)	9	18	3.78	0.0071**	18.89	±	0.88	9	18	3.82	0.0075**	18.23	±	0.86

	Friedman Tests for Genotype Effect			Friedman Tests for Genotype Effect		
	DF	Chi-sq	P-Value	DF	Chi-sq	P-Value
Biomass score	9	18.4923	0.0298726*	9	23.84	0.004561565**
Inflorescence volume score	9	18.0756	0.03430616*	9	15.161	0.08660893ns

*, **, *** significant at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively, whereas ns shows non-significance.

3. Characterization of Flowering Time in Genebank Accessions of Grain Amaranths and Their Wild Relatives Reveals Signatures of Domestication and Local Adaptation[2]

Abstract

Grain amaranths (*Amaranthus* spp.) are ancient crops from the Americas that are consumed as pseudo-cereals and vegetables. The two grain amaranths *A. cruentus* and *A. hypochondriacus* originated in Central America and *A. caudatus* in South America. Flowering time variation plays a central role in their uses as seed, vegetable, and biomass crops. We characterized phenotypic variation for plant height, flowering time and seed setting among 253 genebank accessions including three grain and two wild ancestor species (*A. hybridus* and *A. quitensis)* in the temperate climatic and long-day conditions of Germany. Among grain amaranths, *A. cruentus* flowered early and 88 % accessions set seed. *A. hypochondriacus* accessions were mildly or highly photoperiod sensitive with a lower proportion of seed setting (31 %). *A. caudatus* accessions were mildly photoperiod sensitive and failed seed production. Photoperiod insensitive accessions set seed regardless of their origin, and mildly photoperiod sensitive accessions set seed if they originated from regions with higher temperatures. Overall, Central American accessions of both wild and domesticated amaranths showed large variation in flowering time and photoperiod sensitivity, whereas variation among South American wild and domesticated amaranths was limited to mild photoperiod sensitivity. This observation is

[2] This chapter is published as:
Baturaygil, A.; Schmid, K. Characterization of Flowering Time in Genebank Accessions of Grain Amaranths and Their Wild Relatives Reveals Signatures of Domestication and Local Adaptation. *Agronomy* **2022**, *12*, 505. https://doi.org/10.3390/agronomy12020505

consistent with a model of independent domestication in Central and South America and suggests a potential Central American origin of *A. hybridus* followed by migration to and selection against high photoperiod sensitivity in South America. Our results provide useful information for the design of breeding programs for different uses and provide insights into grain amaranth domestication by considering flowering time as an adaptive trait.

Keywords: Flowering, amaranth, photoperiod, domestication, adaptation, characterization.

3.1. Introduction

Flowering time plays a central role in the environmental adaptation and agricultural productivity of crops. Two main factors that determine flowering timing are photoperiod and temperature (Izawa, 2007). Photoperiod is a more useful indicator of crop season compared to temperature because it is more stable across years (Lagercrantz, 2009), whereas temperature controls the rate of development (Craufurd and Wheeler, 2009) and is more variable across years. For short-day crops such as maize and soybean, early flowering varieties with little or no photoperiod sensitivity are grown for grain production at high latitudes such as northern Europe and North America because of the short duration of favorable temperatures (Jung and Müller, 2009; Zhang *et al.*, 2020). Insensitivity to photoperiod ensures uniform phenological development across a range of different growing environments (Ceccarelli, 1994), while short-day cultivars with high photoperiod sensitivity delay flowering under long photoperiods, resulting in prolonged vegetative growth and higher biomass accumulation. This trait is exploited in the breeding and cultivation of bioenergy crops such as maize and sorghum to

achieve high dry matter yield for improved methane yield for heat and power generation in biogas plants (Herrmann and Rath, 2012). Overall, there is a close relationship between flowering time and the different uses of a crop, such as grain or biomass yield, especially for annual field crops.

The genus *Amaranthus* harbors more than 60 species of which the majority are wild species and well known as weeds in crop production areas. In addition, there is a small number of cultivated amaranth species for grain and vegetable use (Sauer, 1967; National Research Council, 1984; Costea and Tardif, 2004; Chaudhari *et al.*, 2017). In the recent past, interest in amaranths has been growing because of their high nutritional qualities, their capability of growing under dry and marginal areas, and their potential as silage and bioenergy crops. The three species of domesticated grain amaranths are ancient crops in South (*A. caudatus* L.) and Central America (*A. cruentus* L., and *A. hypochondriacus* L.*)*. Their grains are characterized by high protein and lysine contents, and they are of interest as alternative diet for patients with celiac disease because they are gluten-free (Becker *et al.*, 1981; Stallknecht and Schulz-Schaeffer, 1993). Other amaranth species such as *A. tricolor* and *A. dubius* are commonly used as a vegetable in tropical regions of the Americas, Africa, and Asia (National Research Council, 1984; Brenner *et al.*, 2000). The grain amaranth species use the C4 photosynthetic pathway, and their economical water use provides an advantage in coping with drought stress (Liu and Stützel, 2004; von Cossel *et al.*, 2017). A comparison of maize and amaranth for biogas production in the temperate Central European climate showed that amaranth exhibited a higher drought tolerance, even though maize was superior in biogas yield components (von Cossel *et al.*, 2017). The grain amaranths are mainly self-fertilizing plants with a rate of out-crossing between 3 to 32 % (Jain *et al.*, 1982; Hauptli and Jain, 1985).

Grain amaranths are predominately short-day crops because of their provenance in the tropics. However, a large photoperiodic variation was reported in parallel with their large distribution area (Wu *et al.*, 2000). Under the long-day conditions of temperate Central Europe, photoperiod sensitive amaranth genotypes delay flowering and display elongated vegetative growth. This observation prompted the evaluation of the biomass potential of amaranth in the context of biogas production, which revealed a promising performance and a potential for future improvement (Mursec *et al.*, 2009; Pospišil, 2009; Svirskis, 2009; Balodis *et al.*, 2011; Seppälä *et al.*, 2013; Sitkey *et al.*, 2013; von Cossel *et al.*, 2017). In addition, amaranth can be used as a silage crop in both tropical and temperate regions, for which a high biomass is advantageous (Brenner *et al.*, 2000). Recently, we reported two contrasting growth patterns in amaranth under long-day conditions; a grain-type amaranth cultivar showed early flowering and short plant height, suggesting photoperiod insensitivity, while biomass-type amaranths showed delayed or no flowering and reached a tall plant height, suggesting the effect of short-day genes (Baturaygil *et al.*, 2021). Kulakow and Jain (1985) postulated that flowering time in amaranth is controlled by three major genes, with a single gene controlling reduced vegetative growth and two genes controlling photoperiodic response in multiple backcross-derived populations between *A. cruentus* and the wild relative *A. retroflexus.* Wu *et al*. (2000) tested 229 accessions of 20 amaranth species at two contrasting sites in China, and Andini *et al*. (2020) tested 69 accessions of nine amaranth species at a single site in Japan under four different growing seasons with different photoperiods. Both studies found substantial photoperiodic variation in their diversity panels, each representing a large number of countries of origin and including the three grain amaranths as well as the wild putative ancestor species *A. hybridus*. A few studies examined the agricultural production and adaptive potential of

amaranth for these different uses such as seed and biomass production, under a variety of contrasting environments in Europe, ranging from the Mediterranean to temperate climates, but often using a very limited number of genotypes (Rivelli *et al.*, 2008; Hoidal *et al.*, 2019). In a study looking at the dual-use of amaranth as seed and vegetable, all of seven tested cultivars of the three major grain species set seed in field experiments that were conducted in different environments with very distant latitudes such as Denmark (55°N) and Mexico (17°N) (Hoidal *et al.*, 2019). Overall, these studies showed that amaranth has potential for both seed and biomass production in high-contrast environments in Europe.

The *Amaranthus* subgenus of *Amaranthus* genus includes the hybridus complex, which consists of the three grain type amaranths (*A. caudatus* L., *A. cruentus* L., and *A. hypochondriacus* L.) and their two putative wild ancestor species (*A. hybridus* L. and *A. quitensis* Kunth) (Kietlinski *et al.*, 2014; Stetter and Schmid, 2017; Stetter *et al.*, 2020). Of the species within the hybridus complex, *A. cruentus* and *A. hypochondriacus* originate from Central America, the former being considered the most photoperiod insensitive grain amaranth species (Brenner *et al.*, 2000). *A. caudatus* is a short-day species native to the high altitudes of the Andean highlands in South America. It has become adapted to lower temperatures and tends to mature late in temperate zones (Sauer, 1967; National Research Council, 1984; Brenner *et al.*, 2000; Escobedo-López *et al.*, 2014). Among the wild ancestor species, *A. hybridus* shows a very wide distribution across North, Central, and South America, while *A. quitensis* is restricted to South America and might have been involved in the domestication of *A. caudatus* (Sauer, 1967).

The availability of a *A. hypochondriacus* reference genome (Lightfoot *et al.*, 2017) and subsequent population genetic analyses provided new evidence on the evolutionary history and domestication of grain amaranths. A first model with support from genetic polymorphism data postulated that *A. hybridus* is the most diverse wild ancestor and at least two grain amaranth species evolved from this wild ancestor (Kietlinski *et al.*, 2014). Kietlinksi *et al*. (2014) studied genetic diversity of the hybridus complex using 11 microsatellite markers and suggested that all three grain amaranth species have evolved from *A. hybridus* based on its pattern and levels of nucleotide diversity. In addition, they postulated two scenarios in which (i) *A. hybridus* was domesticated once in Mesoamerica or in the Andes and a subsequent geographic expansion and separation of that domesticate led to the origins of *A. caudatus* and *A. hypochondriacus*, or (ii) a single *A. hybridus* lineage ranging from Mesoamerica to the Andes was domesticated twice, leading to independent origins of *A. caudatus* and *A. hypochondriacus*. Stetter and Schmid (2017) and Stetter *et al*. (2020) also reported a high nucleotide diversity in *A. hybridus.* Population structure inference of the hybridus complex using genotyping-by-sequencing and whole-genome sequencing, respectively, supported the hypothesis that the three grain amaranth species independently domesticated from different subpopulations of *A. hybridus* (Stetter and Schmid, 2017; Stetter *et al.*, 2020). The population structure of both the wild ancestor and the domesticates is more strongly determined by geographic than taxonomic separation because each of the three grain species clustered with the wild ancestors from the same geographical region based on SNP markers in principal component analysis (Stetter and Schmid, 2017; Stetter *et al.*, 2020).

Although previous studies demonstrated flowering time variation among species of the hybridus complex, its extent and distribution according to taxonomic or geographical aspects

is unknown. Potential sources of useful genetic variation such as specific latitudinal regions or species within the hybridus complex are undetermined, which may be useful for an enhanced local adaptation and further crop improvement. Amaranth can be adapted to a wide range of environmental conditions in temperate climates, but how the multiple environmental factors at the center of origin shape adaptive variation and influence the plant adaptation in temperate climates has not yet been studied. In addition, the evolutionary processes associated with variability in photoperiodic responses during flowering time adaptation have not been characterized in amaranth, which may provide information on the domestication history of grain amaranths. In this study, we conducted a field trial in Central Europe with a large panel of genetically diverse genebank accessions of grain amaranths and their wild relatives. Our objectives were to (i) identify genebank accessions suitable for grain or biomass breeding programs, (ii) characterize the phenotypic traits and environmental variables in terms of categorically determined latitudinal groups and species, (iii) study the relationships between phenotypic traits and the environmental variables to understand how the environmental variables affect phenotypic variation, and (iv) address how these findings contribute to an understanding of the domestication history of grain amaranths and the design of amaranth breeding programs for temperate climates.

3.2. Materials and Methods

3.2.1. Plant material

We tested 253 genebank accessions obtained from USDA-ARS genebank in a field trial, including three major grain species (*A. caudatus* L., *A. cruentus* L., and *A. hypochondriacus* L.), their putative wild ancestors (*A. hybridus* L. and *A. quitensis* Kunth) and accessions that were 'hybrids' in the passport data (**Table S2**). Furthermore, we included the grain type variety Bärnkrafft (*A. cruentus*), which has been developed for cultivation in Central Europe.

3.2.2. Field experiment

In 2019 we performed a field trial in a single location, the Heidfeldhof experimental station (48° 42' N, 9° 11' E, 395 m a.s.l.) of Hohenheim University, Germany (**Figure S1**). The experiment was conducted with an augmented design of 280 plots distributed into 10 blocks, and Bärnkrafft was included as check variety in each block to be used in the correction of unreplicated entries. We used an augmented design since we had a large number of accessions with limited amount of seeds to test, which is common in the early stages of breeding programs (Kahriman *et al.*, 2020). Plot length was 2 m, the distance between the adjacent plots was 75 cm, whereas the distance between blocks was 1 m. Each plot consisted of a single row. Seeds were sown manually on 17th of May and the experiment was finished on 23rd of October. Weeds were manually and mechanically removed when needed. Thinning was manually performed by leaving 10 cm between the plants and no fertilization and irrigation was applied.

During the period of the field trial, we retrieved the monthly mean temperate and precipitation data from http://wetter-bw.de for the Heidfeldhof agricultural field station at which the trial

was performed (**Figure S1**). In the estimation of photoperiod during the vegetation period, we retrieved the day length data for each day from the sunrise and sunset calculator at http://www.timeanddate.com in the format of hour-minute-second (hms) for the Stuttgart region. Using the *lubridate* R package (Grolemund and Wickham, 2011), we first converted the hms format into seconds using period_to_seconds function and then converted the respective seconds to hours by dividing them by 3600. Finally, we estimated the mean photoperiod values for each month.

3.2.3. Phenotypic traits

We phenotyped each plot for plant height, flowering time, and seed setting. Plant height was recorded at harvest time as distance (cm) between the soil level and the top of the inflorescence. Flowering time was recorded as number of days from sowing until pollen shedding, and seed setting was recorded as binary trait at harvest time. Plant height and seed setting were recorded based on three randomly selected plants, whereas flowering time was recorded plot-wise. Growing degree days indicate the heat accumulation from sowing until flowering and is formulated as

$$\frac{(T_{\max}+T_{\min})}{2} - T_{\text{base}} \qquad (5)$$

where T_{max} and T_{min} are the maximum and minimum temperatures of a day, respectively, whereas T_{base} is the base temperature we used for amaranth. If T_{base} is larger than T_{min}, T_{min} is replaced by T_{base} in the formula. The formula was applied for each day and the cumulative sum

of each day's estimated values from sowing until flowering gives the total heat accumulation of that particular flowering date. We took the T_{base} value as 10°C (Horak and Loughin, 2000) and estimated growing degree days using the *pollen* R package (Nowosad, 2019).

3.2.4. Environmental variables

Coordinates of geographic origin of accessions were retrieved from their passport data (USDA ARS). For the majority of accessions coordinates were missing in the passport data and were retrieved with Google Maps using the location name. We obtained the climatic data from the WorldClim database version 2.1 with the resolution of 30 seconds (~ 1 km^2) using the collection coordinates of the accessions (Fick and Hijmans, 2017). We extracted and processed the data using the *raster* R package (Hijmans, 2020). The variables we used were monthly average temperature (°C), precipitation (mm), solar radiation (kJ m^{-2} day^{-1}), and elevation (m). As the tested accessions mainly originate from Central and South America and the sowing dates and the vegetation periods of amaranth vary largely across this geographic range (Kauffman, 1992), we used the annual means of these climatic variables in our analyses.

3.2.5. Statistical analyses

We grouped accessions according to their latitudinal distribution and species and characterized them by comparing these group means in phenotypic traits and the environmental variables at the geographic origin of accessions. We separated 226 accessions with complete coordinates into three latitudinal groups based on their latitude of origin (**Figure 4**). First, we assigned all

accessions from Central America into a single group (n = 60). Second, we estimated the median latitude of South American accessions (-10.76°) and accordingly divided the South American accessions into South America-I (n = 84) and South America-II (n = 82) groups. Of the 280 plots in the field trial, 10 plots were planted with accessions from a different biomass breeding population to fill empty plots, but these genotypes were not included in the statistical analysis because they do not belong to any latitudinal group and do not have species information. In addition, we discarded seven accessions with insufficient field emergence from further analysis. Hence, the comparison of species and latitudinal groups were performed with 254 and 226 accessions, respectively. We compared groups using a generalized linear model due to the heterogeneity of variance among the compared groups (**File S1**). We used a binomial distribution with a logit link function in seed setting and normal distribution in the other traits and variables. We performed a likelihood ratio test by comparing our model with the default null model – assuming a constant mean across all groups – using a chi-square test, since a generalized linear model does not produce a p-value. In case of significant differences between comparisons, we compared the groups using a Fisher's least significant difference (LSD) test (alpha = 0.05). Analyses with generalized linear models were performed using *stats* package, the likelihood ratio test was carried out with the *lmtest* package (Zeileis and Hothorn, 2002), the LSD test with the *agricolae* package (de Mendiburu, 2015), correlation analyses with the *Hmisc* package (Harell, 2018) and plotting of the correlations with the *corrplot* package (Wei and Simko, 2017) within the R environment (R Core Team, 2020).

3.3. Results

3.3.1. Variation in phenotypic traits and environmental variables

We observed a very large variation in flowering time among accessions. The number of days to flowering ranged between 50 and 160 days with a mean of 106.3 days. Six accessions from Central America did not reach the flowering stage at all (**Table S3**). All grain species (*A. caudatus, A. cruentus,* and *A. hypochondriacus*) and wild species (*A. hybridus* and *A. quitensis*) differed from each other in mean flowering time (**Figure 5A**). *A. cruentus* flowered the earliest with a mean of 79.44 days, while *A. caudatus* flowered later, with a mean of 110.28 days. *A. hypochondriacus* accessions flowered the latest with a mean of 122.33 days and showed a large range of 89 days (71-160 days). The 'hybrid'-labeled accessions showed the widest range in flowering time with 110 days (50-160 days). Three strong outlier observations of this group consisting of one accession flowering after 50 days and two accessions flowered after 160 days strongly affected this estimate (**Figure 5A**). Removal of these three outliers reduced the flowering time range to 60 days (65-125 days).

In contrast to the grouping by species, the mean flowering time of the three latitudinal groups was 107.95 days and the three groups did not differ from each other (**Figure 5B**). The Central America group including the earliest flowering species *A. cruentus* and the latest flowering species *A. hypochondriacus* showed a wide range in flowering time (70-160 days). Among the three latitudinal groups, South America-II showed the widest range with 110 days (50-160 days) similar to the 'hybrid'-labeled accessions in the species-based comparisons. However, this range estimate was also biased by several strong outliers (**Figure 5B**). Therefore, we report the inter-quartile ranges of these three groups as this statistic is less prone to outliers. The inter-quartile ranges of the Central America, South America-I and South America-II

groups are 62, 8.5 and 11 days, respectively, which indicates that Central American accessions have a much larger flowering time variation than the South American accessions.

Growing degree days is determined by flowering time and daily temperature, and we observed very similar results between this trait and flowering time in the species-based comparisons (**Figure 5C and D**). *A. cruentus* and *A. hypochondriacus* reached the lowest and the highest growing degree days means, respectively, but 'hybrid'-labeled accessions showed the largest range (416 – 1,196° days; **Figure 5C**), although the three latitudinal groups did not differ from each other (**Figure 5D**). Similar to flowering time, the Central America group showed the largest inter-quartile range (450° days) among the three latitudinal groups. For the trait plant height, accessions ranged between 45.2 and 359.9 cm with a mean of 220.3 cm, when all accessions were considered. At the species level, only *A. cruentus* was shorter than all other accessions with a mean height of 170.7 cm (**Figure 5E**). Among latitudinal groups, the Central American group had a mean plant height of 208.73 cm, which was shorter than the South American groups (**Figure 5F**).

We recorded seed setting as binary trait where '0' represents no seed setting and '1' represents seed setting. Considering all accessions, the mean proportion of seed setting was only 19 % (n = 49). Species and latitudinal groups showed distinct patterns for this trait (**Figure 5G and H**). Among species, proportion of seed setting was highest in *A. cruentus* with a mean of 88 % (n = 30), followed by *A. hypochondriacus* with 31% (n = 11), whereas no *A. caudatus* accession managed to set seed. The wild amaranth species were not different from each other, and their proportion of seed setting was approximately 12 % (n = 2, n = 3 and n = 3 for the 'hybrid' group, *A. hybridus*, *A. quitensis* respectively) in seed setting. Among latitudinal

groups, 43.3% (*n* = 26) of Central American accessions showed seed setting. Most South American accessions failed to produce seeds with a proportion of 1% (*n* = 1) accessions with seeds in the South America-I group and 7.3% (*n* = 6) in South America-II group.

To put environmental variation into context with phenotypic variation, we characterized variation in six environmental variables that represent the collection sites of the genebank accessions (**Figure S2**). We detected differences in the distribution of all environmental variables in the comparison of species and latitudinal groups, except for the variable precipitation among species. The two Central American species *A. cruentus* and *A. hypochondriacus* were more similar to each other with respect to environmental parameters than the other species. In contrast, the three latitudinal groups were different from each other in all environmental variables including precipitation. The Central American group received a higher amount of precipitation compared to the two South American groups, which did not differ from each other.

3.3.2. Relationship between phenotypic traits and environmental variables

We calculated correlations between phenotypic traits and environmental variables with four different datasets using the accessions from (i) all countries, (ii) South America, (iii) Central America, and (iv) *A. hypochondriacus* from Central America only. We estimated the correlation matrices using Spearman's rank correlation since seed setting is coded as a binary trait and a major interest was to investigate the traits and variables affecting seed setting. However, we studied the relationship between flowering time and plant height also using

Pearson's correlation to investigate the linear trend between these traits. Furthermore, we used the absolute values of latitude and longitude in the correlation analyses.

In the dataset including all accessions (n = 254), all traits and environmental variables were weakly to moderately but always highly significantly correlated to seed setting ($p < 0.001$), except precipitation (**Figure 6A**). Among the traits and environmental variables, flowering time, growing degree days, plant height, and elevation were negatively correlated, whereas temperature and solar radiation were positively correlated to seed setting. Likewise, latitude and longitude were weakly but positively correlated to seed setting. In addition, we observed strong correlations between several pairs of variables (**Figure 6A**). Flowering time and growing degree days were positively (rho = 1, $p < 0.001$) and elevation and temperature were negatively correlated (rho = -0.90, $p < 0.001$). Plant height and flowering time (n = 224) showed a moderate positive correlation (r = 0.43, $p < 0.001$; **Fig 4A**). Since plant height increased together with the flowering time between 45 and 120 days but remained constant with later flowering dates, the inclusion of a polynomial term to the model improved the fit by increasing the r-squared from 0.18 to 0.33.

The collection sites of the South American accessions are located more closely to each other than to the Central American sites (**Figure 4**). We then repeated the correlation analysis by excluding the Central American accessions to detect their potential outlier effects. In the analysis with the second dataset (n = 181), all traits and variables except precipitation and solar radiation were significantly but only weakly correlated to seed setting (**Figure 6B**).

We also conducted the analysis with the Central American accessions only (n = 67), since this region is the center of the variation for flowering time and seed setting (**Figure 5B and H, Figure 7C**). Seed setting was strongly and negatively correlated to flowering time and growing degree days (*rho* = -0.81, $p < 0.001$), also negatively correlated with latitude (*rho* = -0.36, $p < 0.01$) and positively correlated with temperature and solar radiation ($p < 0.01$; **Figure 6C**). Among environmental variables, only elevation and temperature were strongly (*rho* = -0.86, $p < 0.001$) correlated.

Finally, we repeated the analysis with the fourth set (n = 32) that includes only *A. hypochondriacus* accessions from Central America, due to the large intraspecific variation in flowering time (**Figure 7D**). Flowering time and growing degree days were moderately to strongly and negatively (*rho* = -0.73, $p < 0.001$) and elevation was weakly but negatively (*rho* = -0.34, $p < 0.05$), whereas temperature and solar radiation were moderately and positively correlated to seed setting (**Figure 6D**). In addition, we observed strong correlations between flowering time and growing degree days (*rho* = 1, $p < 0.001$) and elevation and temperature (*rho* = -0.90, $p < 0.001$).

3.3.3. Factors influencing seed setting

We focused on the flowering time distribution of the Central American accessions to understand the factors determining seed setting. *A. cruentus* accessions display little flowering time variation and include predominantly photoperiod insensitive but only a few late-flowering accessions (**Figure 7C**). On the other hand, *A. hypochondriacus* accessions show a

very wide variation and include moderate to highly photoperiod sensitive accessions. Furthermore, correlation analyses within Central American accessions revealed that temperature and solar radiation are moderately correlated climatic variables to seed setting in datasets three and four. Accordingly, we separated the Central American accessions into seed setting (22 x *A. cruentus,* 8 x *A. hypochondriacus,* 3 x *A. hybridus*) and non-seed setting (24 x *A. hypochondriacus,* 4 x 'hybrids', 3 x *A. cruentus and* 3 x *A. hybridus*) groups to compare their differences in phenotypic traits and environmental variables based on their geographical origin.

In these comparisons, these two groups from Central America significantly differed in flowering time, growing degree days, plant height, temperature and solar radiation (**Figure S3, File S1**). We repeated the same comparisons between seed setting (8) and non-seed setting (24) Central American *A. hypochondriacus* accessions only, due to their high variation in time to flowering and seed setting (**Figure 7D**). In the Central American *A. hypochondriacus,* the two groups significantly differed in flowering time, growing degree days, temperature, solar radiation and in elevation (**Figure S4, File S1**).

The temperature at the geographic origin of accessions was significantly higher (**File S1**) for (i) the Central American group (18.49°C) than the South American groups (13.68-14.86°C) (**Figure S2**), (ii) seed setting accessions (19.82°C) compared to non-seed setting ones (17.47°C) within Central American accessions (**Figure S3**), and finally (iii) for seed setting accessions (20.87°C) compared to non-seed setting ones (16.95°C) within Central American *A. hypochondriacus* accessions (**Figure S4**). These results suggest that (i) seed setting and non-seed setting accessions clearly diverge in many phenotypic traits and environmental

variables and (ii) an accession's temperature of origin should reach a certain threshold for seed setting in the temperate climate of Germany.

3.4. Discussion

3.4.1. Categorization of photoperiodic response

Flowering time response is a quantitative trait that depends on the growth environment in which ecological factors like photoperiod and temperature act together (Liu *et al.*, 2020). The photoperiodic response of genotypes adapted to a short-day environment can be quantified if (i) the material is photoperiod sensitive and (ii) long day conditions are present during the cultivation period. We therefore categorized the photoperiodic responses of the accessions roughly into three groups. The first group is composed of photoperiod insensitive accessions that flowered between 50 and 80 days (n = 31) and mostly set seed (n = 29). The majority of accessions in this group originated from Central America and the main representative species of this group is *A. cruentus,* which is characterized by short plant height and a high temperature at the geographical origin. A few accessions not originating from Central America also showed photoperiod insensitivity and set seed (**Table S4**). The second group is composed of mildly photoperiod sensitive accessions that flowered between 80 and 140 days (n = 200) and rarely set seed (n = 19). The majority of accessions in this group are *A. caudatus* or originated from locations in South America with high altitudes and cool temperatures. However, this group also includes *A. hypochondriacus* accessions from Central America. We separated the second group from the third group based on the geographical origins of the

accessions. The accessions in the third group (n = 18) are highly photoperiod sensitive and all originated from Central America, compared to the second group accessions that mostly originated from South America. This group includes *A. hypochondriacus* accessions that flowered after 140 days and *A. hybridus*, *A. hypochondriacus and* 'hybrid' accessions that did not flower (**Table S3**). All accessions in this group failed to set seed in our experiment.

3.4.2. Relationship between phenotypic traits and environmental variables

The correlation between flowering time and seed setting varied strongly in the different datasets defined by geographic groups. In the joint analysis of Central and South American accessions correlations were moderate to strong. Given the Central American accessions contributed the highest proportion of variation for these two traits whereas South American accessions contributed a neglectable amount of variation for seed setting, a strong correlation was observed between flowering time and seed setting within Central American accessions, and the exclusion of the Central American accessions remarkably reduced the correlation between flowering time and seed setting within the South American accessions.

The correlation analysis of latitude of origin with flowering times in the dataset consisting of South and Central American accessions was constrained (**Figure 6A**) because accessions from both regions were not evenly distributed along latitude but clustered in two regions (**Figure 4**). However, these two groups included a very wide range of photoperiodic responses i.e., the Central American accessions ranged from photoperiod insensitive to highly sensitive, whereas the South American accessions were mainly moderately sensitive. The narrow latitudinal

variation with a very wide flowering time variation reduced the correlation between these variables.

3.4.3. Variation in seed setting

Seed setting under temperate climate and long-day conditions of Europe is a primary measure of success for local adaptation and grain yield. Among all tested accessions, mainly those of Central American origin set seed. However, some accessions originating from South America also managed to set seed (n = 16). These accessions can be categorized in two groups based on their photoperiodic responses. The first group included 10 photoperiod insensitive accessions that flowered between 50 and 76 days (**Table S4**). The second group included six mild photoperiod sensitive accessions that flowered between 82 and 110 days (**Table S5**). All accessions in the second group belong to *A. cruentus* and *A. hypochondriacus*, which are native species to Central America.

We observed variation in seed setting among accessions that flower at similar dates (**Figure 7B**), i.e., there was no single flowering time threshold representing all species at which accessions ceased seed setting. In contrast, a different threshold existed for each species (**Figure 7B**). These threshold days were 76 for *A. quitensis*, 89 for 'hybrids', 111 for *A. hybridus*, 113 for *A. cruentus,* and 124 days for *A. hypochondriacus*. However, such a threshold might be sensitive to strong outliers. For example, a single accession flowered at day 110 and changed the threshold from 102 to 111 in *A. hybridus*. Similarly, another type of outlier, an accession that behaves differently than the latitudinal group or species it belongs to, such as the photoperiod insensitive accessions that are of non-Central America origin may also

cause such a misinterpretation. The most visible example of this pattern was *A. caudatus* because many *A. cruentus* and *A. hypochondriacus* accessions that flowered later than *A. caudatus* accessions set seed, whereas no *A. caudatus* accession managed to set seed regardless of the flowering time. Hoidal *et al*. (2019) reported an *A. caudatus* accession that set seed in a set of field experiments in the high latitude of Denmark (55°), which contradicts the results with *A. caudatus* in this study. Rivelli *et al*. (2008) reported photoperiod sensitivity and associated late flowering in a small number of accessions of wild and cultivated amaranth species under the Mediterranean climatic conditions of Italy, which are characterized by higher annual mean temperature than Germany. Under these conditions all accessions managed to set seed. These results suggest that photoperiodic response play a major role in flowering time, whereas high temperature has a strong effect on seed setting. Since a certain temperate threshold in site of origin appears to be reached for seed setting under temperate Europe climatic conditions and almost only Central America originated accessions met that requirement (**Figure S2B**), seed setting variation was the highest among the Central American accessions. Altogether, these results show that photoperiod insensitive accessions set seeds in Europe regardless of their origin, whereas the accessions with mild photoperiod sensitivity may set seed provided that their site of origin reaches a certain temperature threshold, such as the mild photoperiod sensitive accessions from Central America.

3.4.4. Selection of early flowering for local adaptation

Roux *et al*. (2006) suggests that late flowering is an ancestral character in *Arabidopsis thaliana* and the flowering time response can be altered by the down or up-regulation of the

regulatory factors such as floral repressive or promotive genes. Early flowering is a desired characteristic in maize that is cultivated at higher latitudes for a local adaptation i.e., to be able to produce seeds and complete filling phase before the onset of unfavorable environmental conditions such as frost (Jung and Müller, 2009). In high latitudes, selection for short-day plants for earlier flowering has led the convergent evolution from an ancestral high photoperiod sensitivity to a reduced photoperiod sensitivity in several crops such as maize, sorghum, and rice (Matsubara *et al.*, 2011; Hung *et al.*, 2012; Olsen and Wendel, 2013; Murphy *et al.*, 2014; Gaudinier and Blackman, 2019). Early flowering facilitates local adaptation also in high altitudes for similar reasons. Recently, Wang *et al*. (Wang *et al.*, 2021) reported the parallel evolution of early flowering phenotypes in four highland maize populations compared to two lowland populations that exhibited later flowering. This study also showed that many flowering time genes were selected during highland adaptation, including genes from photoperiod-pathway and circadian clock. Among natural populations of the long-day plant *Arabidopsis thaliana* that were collected from close latitudes, populations from lower altitudes showed higher photoperiodic variation than populations from higher altitudes (Lewandowska-Sabat *et al.*, 2017), which suggests that environmental variables associated with altitude impose natural selection on photoperiod sensitivity.

In our study, we observed a similar separation in the Central and South American accessions in the context of photoperiodic variation i.e., a very large photoperiodic variation existed in Central America ranging from photoperiod insensitive to highly sensitive accessions, whereas a lower range of variation existed among accessions from South America that mostly consisted of mildly photoperiod sensitive accessions, in addition to a few exceptions. Particularly, we observed the same pattern between the Central and South American *A. hybridus* accessions.

We tested 31 *A. hybridus* accessions, 25 of which originated from South America and six originated from Central America. The 23 South American accessions showed mild sensitivity, one showed high sensitivity and one accession did not flower, whereas three of the Central American accessions did not flower indicating a high photoperiod sensitivity. Accordingly, South American *A. hybridus* accessions can be categorized as mild photoperiod sensitive, whereas a limited number of Central American *A. hybridus* accessions showed a wider variation in flowering time. Altogether, our results suggest a selection for reduced photoperiod sensitivity in (both grain and wild species) South American accessions including the wild putative ancestor *A. hybridus*.

Furthermore, we hypothesize that high photoperiod sensitivity of Central American *A. hybridus* accessions might have been eliminated in a potential migration event to South America in exchange to adapt to a decreasing latitude and/or increasing altitude to secure seed production under the colder temperature of South America. Stetter and Schmid (2017) reported a population structure between the Central and South American lineages of *A. hybridus* and Swarts *et al*. (2021) proposed that flowering time in maize is not only correlated to population structure but also differentiates populations during the process of local adaptation. Similarly, such a differentiation between the Central and South American *A. hybridus* lineages may reflect local adaptation to South America.

3.4.5. Domestication hypotheses in the light of photoperiodic variation

Similar levels of flowering time variation among Central American *A. hybridus* and the Central American grain amaranths *A. cruentus* and *A. hypochondriacus*, and among South American *A. hybridus* accessions and South American grain species *A. caudatus*, respectively, agrees with a domestication model of an independent domestication of the three grain amaranth species from different subpopulations of *A. hybridus* therefore supporting population genetic analyses (Kietlinski *et al.*, 2014; Stetter *et al.*, 2020).

The domestication model by Sauer (1967) suggested that *A. cruentus* evolved from *A. hybridus*, *A. hypochondriacus* from a natural hybridization between *A. cruentus* and *A. powellii*, and *A. caudatus* from a natural hybridization between *A. cruentus* and *A. quitensis*, based on morphological characters. In our study, *A. hypochondriacus* accessions showed a high photoperiod sensitivity not observed in the other species. Therefore, mapping alleles responsible for this trait in *A. hypochondriacus* may allow to test this domestication model. Alleles for high photoperiod sensitivity should be present in *A. powellii* populations, but not in a photoperiod insensitive *A. cruentus* progenitor. To test this hypothesis, photoperiodic variation of the Central American *A. powellii* accessions needs to be characterized. Espitia-Rangel *et al*. (2010) studied the geographic distribution of *A. cruentus and A. hypochondriacus* originating from Mexico, and of their putative wild ancestors *A. hybridus* and *A. powellii* with approximately 3,000 geo-referenced accessions using passport data. They concluded that the domestication model of Sauer is supported because (i) *A. hybridus* showed the widest geographic distribution which overlaps the distribution of *A. cruentus*, (ii) *A. hypochondriacus* showed a similar latitudinal variation pattern with *A. powellii* and altitudinal variation pattern with *A. cruentus*. In contrast, *A. powellii* was not closely related to grain amaranths in a phylogenetic analysis of 35 *Amaranthus* species (Stetter and Schmid, 2017).

However, the two *A. powellii* accessions used in this study did not originate from Central and South America. We therefore conclude that based on our data the first model of independent domestication from *A. hybridus* is more strongly supported.

3.4.6. Future Prospects

The large variation in flowering time and photoperiodic response can be harnessed in amaranth breeding programs for improved grain and biomass yield in Central Europe. We propose to use *A. cruentus* for grain amaranth breeding programs in temperate regions under long-day conditions. It is photoperiod insensitive and not strongly affected by the variable photoperiods in different areas of cultivation. The species is also suitable for mechanical harvest because of a shorter plant height than the other grain amaranths. Breeding programs for biomass amaranth should consider a trade-off between earliness for high dry matter content and photoperiod sensitivity for improved dry matter yield (Baturaygil *et al.*, 2021), which can be combined in crosses between these two types of amaranths that differ in photoperiodic response. To this end, earliness can be selected from *A. cruentus* and South American accessions as donors of mild photoperiod sensitivity since the delay in flowering time of highly photoperiod sensitive accessions did not additionally contribute to plant height (**Figure 7A**). Genetic characterization of the grain amaranths and their putative wild ancestor species for adaptive traits plays an important role in utilizing the available variation in breeding programs. Therefore, testing of the species in the hybridus complex with (i) proper taxonomic and geographical sampling, (ii) under contrasting environments with varying photoperiods, and (iii) with more replicates contributes their more accurate photoperiodic characterization. Finally, the efficiency of speed breeding in amaranth has been demonstrated and this novel

technology can be used in breeding programs and flowering time studies and may contribute to the spread of this minor crop as an alternative to cope with the adverse effects of climate change (Jähne *et al.*, 2020).

Acknowledgements

This work was supported by the F. W. Schnell endowed Professorship of the Stifterverband and the Hohenheim-Tübingen Regional Alliance of the Ministry of Science and Culture (MWK) of Baden-Württemberg. We thank the staff of the Heidfeldhof experimental station of Hohenheim University for help with the field experiment.

Author Contributions

AB and KS designed the experiment. AB collected and analyzed the data and wrote the first version of the manuscript. Both authors edited, read and approved the final manuscript.

Data Availability Statement

The phenotypic and climatic data, and File S1 are available from Figshare (10.6084/m9.figshare.19071812).

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.5. References

Andini R, Sulaiman MI, Moulana R, Hmon KPW and Ohsawa R (2020) Application of principle component analysis in differentiating the three types of *Amaranthus* based on their photoperiodic flowering response. *IOP Conference Series: Earth and Environmental Science* 425: 0–11.

Balodis O, Bartuševics J and Gaile Z (2011) Biomass yield of different plants for biogas production. *Vide, Tehnologija, Resursi - Environment, Technology, Resources* 1: 238–245.

Baturaygil A, Stetter MG and Schmid K (2021) Breeding Amaranth for Biomass: Evaluating Dry Matter Content and Biomass Potential in Early and Late Maturing Genotypes. *Agronomy* 11(5): 920.

Becker R *et al*. (1981) A Compositional Study of Amaranth Grain. *Journal of Food Science* 46: 1175–1180.

Brenner DM *et al*. (2000) Genetic resources and breeding of *Amaranthus*. *Plant Breeding Reviews* 19: 227–285.

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. *Euphytica* 77: 205–219.

Chaudhari S *et al*. (2017) Biology and Management of Glyphosate-Resistant and Glyphosate-Susceptible Palmer Amaranth (*Amaranthus palmeri*) Phenotypes from a Segregating Population. *Weed Science* 65: 755–768.

von Cossel M, Möhring J, Kiesel A and Lewandowski I (2017) Methane yield performance of amaranth (*Amaranthus hypochondriacus* L.) and its suitability for legume intercropping in comparison to maize (*Zea mays* L.). *Industrial Crops and Products* 103: 107–121.

Costea M and Tardif FJ (2004) The biology of Canadian weeds. 130. *Amaranthus retroflexus* L., *A. powellii* S. Watson and *A. hybridus* L. (update). *Canadian Journal of Plant Science* 84: 631–668.

Craufurd PQ and Wheeler TR (2009) Climate change and the flowering time of annual crops. *Journal of Experimental Botany* 60: 2529–2539.

Escobedo-López D, Núñez-Colín CA and Espitia-Rangel E (2014) Adaptation of Cultivated Amaranth (*Amaranthus* spp.) and Their Wild Relatives in Mexico. *Journal of Crop Improvement* 28: 203–213.

Espitia Rangel E, Mapes Sánchez E, Núñez-Colín CA and Escobedo López D (2010) Geographical distribution of cultivated species of *Amaranthus* and their wild relatives in Mexico Eduardo. *Revista mexicana de ciencias agrícolas* 1: 427–437.

Fick SE and Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.

Gaudinier A and Blackman BK (2019) Evolutionary processes from the perspective of flowering time diversity. *New Phytologist* 225: 1883–1898.

Grolemund G and Wickham H (2011) Dates and Times Made Easy with {lubridate}. *Journal of Statistical Software* 40: 1–25.

Harell FE (2018) Hmisc: Harrell Miscellaneous. R package version 4.1-1.

Hauptli H and Jain S (1985) Genetic variation in outcrossing rate and correlated floral traits in a population of grain amaranth (*Amaranthus cruentus* L.). *Genetica* 66: 21–27.

Herrmann A and Rath J (2012) Biogas Production from Maize: Current State, Challenges, and Prospects. 1. Methane Yield Potential. *Bioenergy Research* 5: 1027–1042.

Hijmans RJ (2020) raster: Geographic Data Analysis and Modeling.

Hoidal N, Díaz Gallardo M, Jacobsen SE and Alandia G (2019) Amaranth as a dual-use crop for leafy greens and seeds: Stable responses to leaf harvest across genotypes and environments. *Frontiers in Plant Science* 10.

Horak MJ and Loughin TM (2000) Growth Analysis of Four *Amaranthus* Species. *Weed Science* 48: 347–355.

Hung HY *et al*. (2012) *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences of the United States of America* 109.

Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *Journal of Experimental Botany* 58: 3091–3097.

Jähne F, Hahn V, Würschum T and Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theoretical and Applied Genetics*.

Jain SK, Hauptli H and Valdya KR (1982) Outcrossing rate in grain amaranths. *Journal of Heredity* 73: 71–72.

Jung C and Müller AE (2009) Flowering time control and applications in plant breeding. *Trends in Plant Science* 14: 563–573.

Kahriman F, Aktaş F, Songur U, Şerment M and Egesel CÖ (2020) Screening Turkish maize landraces for kernel oil content and oil quality traits. *Plant Genetic Resources: Characterisation and Utilisation* 18: 278–286.

Kauffman CS (1992) Realizing the potential of grain amaranth. *Food Reviews International* 8: 5–21.

Kietlinski KD *et al*. (2014) Relationships between the weedy *Amaranthus hybridus* (*Amaranthaceae*) and the grain amaranths. *Crop Science* 54: 220–228.

Kulakow PA and Jain SK (1985) The inheritance of flowering time in *Amaranthus* species. *Journal of Genetics* 64: 85–100.

Lagercrantz U (2009) At the end of the day: a common molecular mechanism for photoperiod responses in plants? *Journal of Experimental Botany* 60: 2501–2515.

Lewandowska-Sabat AM, Fjellheim S, Olsen JE and Rognli OA (2017) Local Populations of *Arabidopsis thaliana* Show Clear Relationship between Photoperiodic Sensitivity of Flowering Time and Altitude. *Frontiers in Plant Science* 8.

Lightfoot DJ *et al*. (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology* 15: 1–15.

Liu F and Stützel H (2004) Biomass partitioning, specific leaf area, and water use efficiency of vegetable amaranth (*Amaranthus* spp.) in response to drought stress. *Scientia Horticulturae* 102: 15–27.

Liu L, Zhang Y and Yu H (2020) Florigen trafficking integrates photoperiod and temperature signals in *Arabidopsis*. *Journal of Integrative Plant Biology* 62: 1385–1398.

Matsubara K *et al*. (2011) *Ehd3*, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *Plant Journal* 66: 603–612.

de Mendiburu F (2015) Statistical Procedures for Agricultural Research.

Murphy RL *et al*. (2014) *Ghd7* (*Ma 6*) Represses Sorghum Flowering in Long Days: *Ghd7* Alleles Enhance Biomass Accumulation and Grain Production. *The Plant Genome* 7: plantgenome2013.11.0040.

Mursec B, Vindis P, Janzekovic M, Brus M and Cus F (2009) Analysis of different substrates for processing into biogas Manufacturing and processing. *Journal of Achievements in Materials and Manufacturing Engineering* 37: 652–659.

National Research Council (1984) *Amaranth Modern Prospects for an Ancient Crop*. National Academy Press.

Nowosad J (2019) pollen: Analysis of Aerobiological Data.

Olsen KM and Wendel JF (2013) Crop plants as models for understanding plant adaptation and diversification. *Frontiers in Plant Science* 4.

Pospišil A (2009) Yield and Quality of Forage Sorghum and Different Amaranth Species (*Amaranthus* spp.) Biomass. *Agriculturae Conspectus Scientificus* 74: 85–89.

R Core Team (2020) R: A Language and Environment for Statistical Computing.

Rivelli AR, Gherbin P, de Maria S and Pizza S (2008) Field evaluation of *Amaranthus* species for seed and biomass yields in southern Italy. *Italian Journal of Agronomy* 3: 225–229.

Roux F, Touzet P, Cuguen JL and Le Corre V (2006) How to be early flowering: an evolutionary perspective. *TRENDS in Plant Science* 11.

Sauer JD (1967) The grain amaranths and their relatives: A revised taxonomic and geo-graphic survey. *Ann. Missouri Bot. Gard.* 54: 102-137.

Seppälä M, Laine A and Rintala J (2013) Screening of novel plants for biogas production in northern conditions. *Bioresource Technology* 139: 355–362.

Sitkey V, Gaduš J, Kliský Ľ and Dudák A (2013) Biogas Production from Amaranth Biomass. *Acta Regionalia et Environmentalica* 10: 59–62.

Stallknecht GF and and Schulz-Schaeffer JR (1993) Amaranth Rediscovered. *In: J. Janick and J.E. Simon (eds.)* 211–218.

Stetter MG and Schmid KJ (2017) Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution* 109: 80–92.

Stetter MG, Vidal-Villarejo M and Schmid KJ (2020) Parallel Seed Color Adaptation during Multiple Domestication Attempts of an Ancient New World Grain. *Molecular biology and evolution* 37: 1407–1419.

Svirskis A (2009) Prospects for non-traditional plant species cultivated for forage in Lithuania. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 37: 215–218.

Swarts K *et al*. (2021) Joint analysis of days to flowering reveals independent temperate adaptations in maize. *Heredity* 126: 929–941.

Wang L *et al*. (2021) Molecular Parallelism Underlies Convergent Highland Adaptation of Maize Landraces. *Molecular Biology and Evolution* 1–14.

Wei T and Simko V (2017) R package “corrplot”: Visualization of a Correlation Matrix.

Wu H *et al*. (2000) Field evaluation of an *Amaranthus* genetic resource collection in China. *Genetic Resources and Crop Evolution* 47: 43–53.

Zeileis A and Hothorn T (2002) Diagnostic Checking in Regression Relationships. *R News* 2: 7–10.

Zhang L *et al*. (2020) Principles and practices of the photo-thermal adaptability improvement in soybean. *Journal of Integrative Agriculture* 19: 295–310.

FIGURES

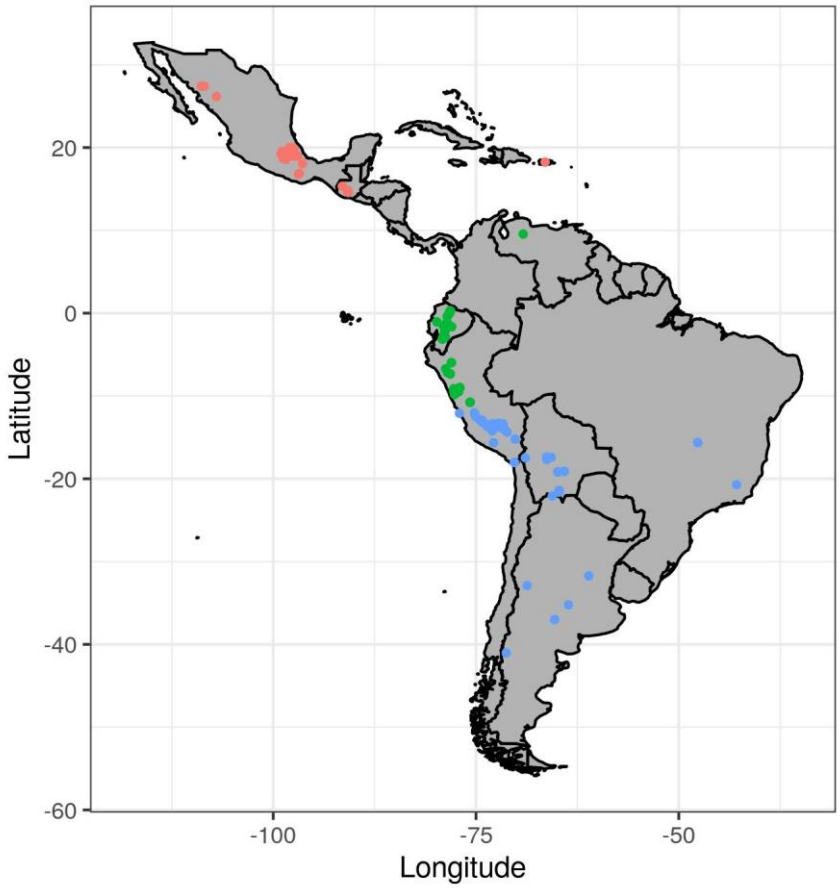

Figure 4. Geographical distribution of the accessions used in the study (*n*=226). Dot colors indicate the categorically determined latitudinal groups of the accessions. Red color indicates Central America group, green color indicates South America-I group and blue color indicates South America-II group.

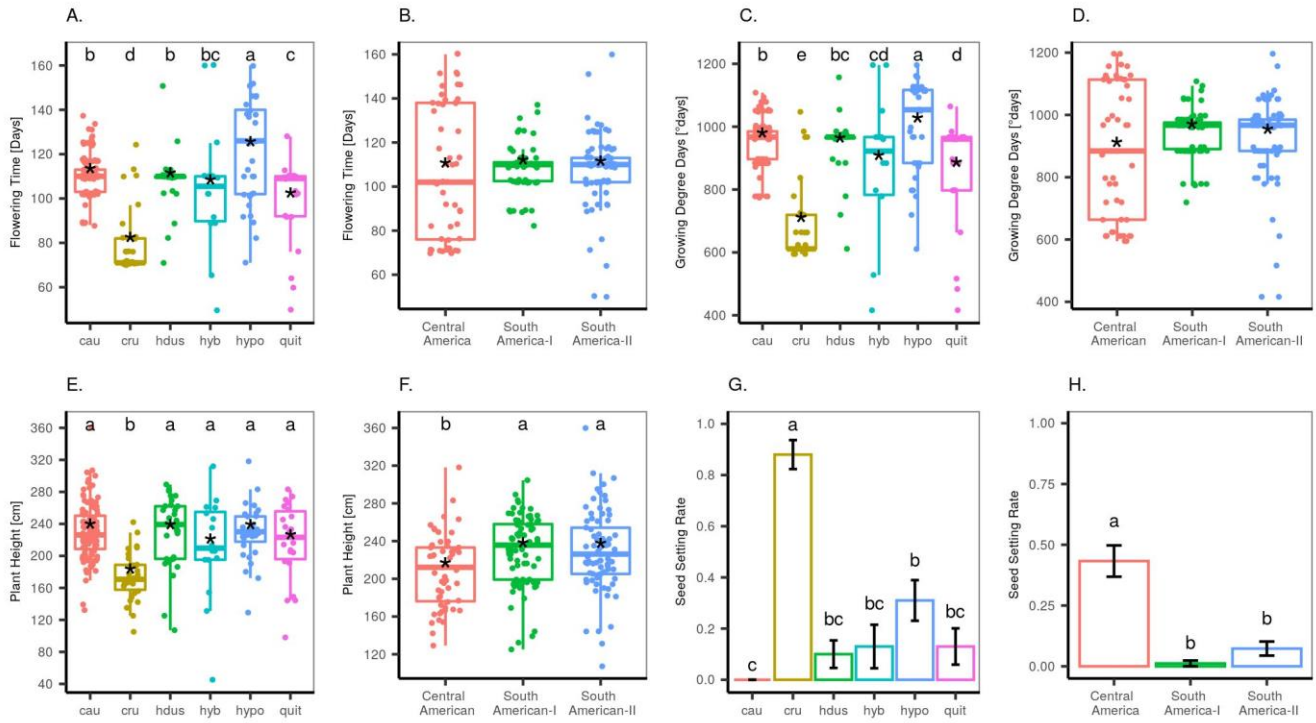

Figure 5. Box and bar plots of four phenotypic traits grouped by species and latitudinal groups, in each box and bar plot the respective groups were compared using a least significant difference (LSD) test and the groups with the different letters are significantly different at alpha=0.05. No letters if there is no significant difference among the compared groups. Black asterisks indicate the mean values of each group in the box plots. Species are cau, *A. caudatus*; cru, *A. cruentus*; hdus, *A. hybridus*; hyb, the 'hybrid' group; hypo, *A. hypochondriacus* and quit, *A. quitensis*.

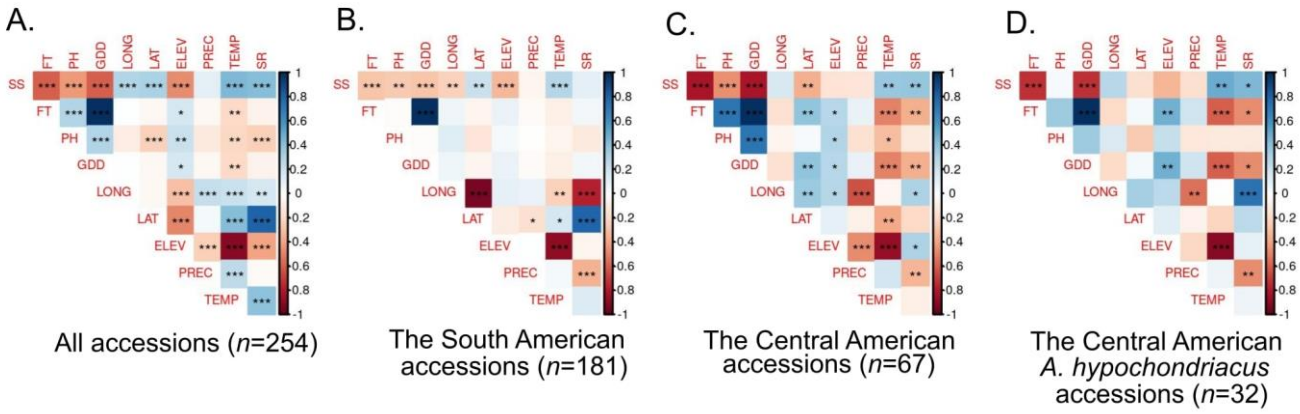

Figure 6. Spearman's correlation matrix of (**A**) all accessions, (**B**) the South American accessions, (**C**) the Central American accessions and (**D**) the Central American *A. hypochondriacus* accessions. Phenotypic traits and environmental variables are SS, seed setting; FT, flowering time; PH, plant height; GDD, growing degree days; LONG, longitude; LAT, latitude; ELEV, elevation; PREC, precipitation and TEMP, temperature. In the correlation matrices *, **, *** indicate significance at $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.

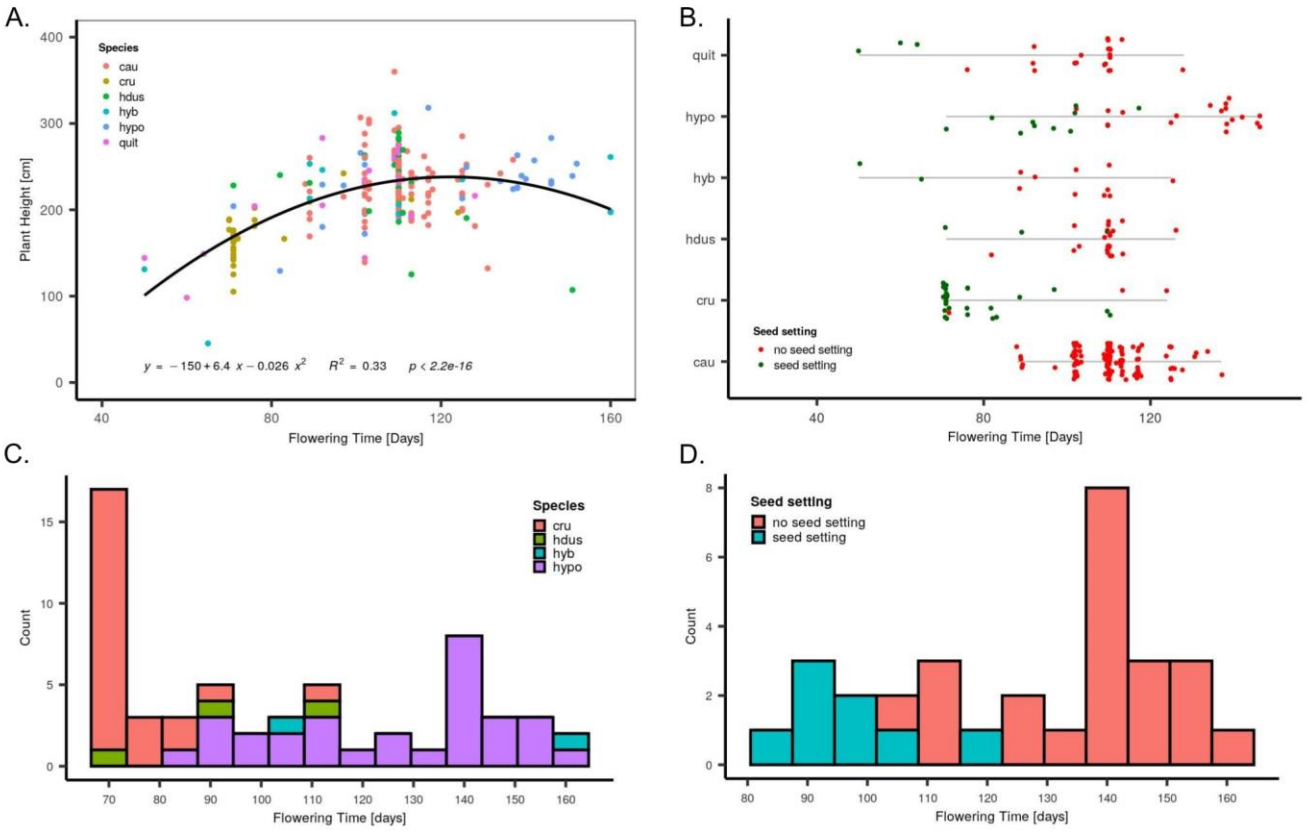

Figure 7. (**A**) The polynomial relationship between flowering time and plant height. (**B**) Dotplot of flowering time. Y axis indicates the species and X axis indicates the flowering days. Each dot corresponds to an accessions flowering time and the color of the dot indicates if the flowering event occurred on the corresponding day resulted in seed setting (green color) or no seed setting (red color). (**C**) Flowering time histogram of the Central American accessions grouped by species. (**D**) Flowering time histogram of the Central American *A. hypochondriacus* accessions grouped by seed setting behavior.

SUPPLEMENTARY FIGURES

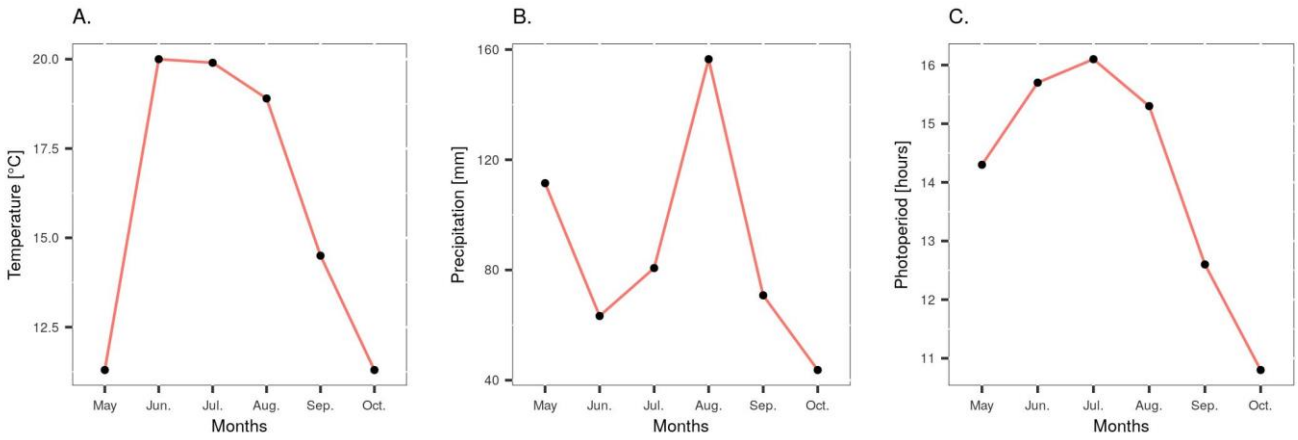

Figure S1. Monthly mean values of the three environmental variables belong to the experimental location Heidfeldhof between May – October 2019.

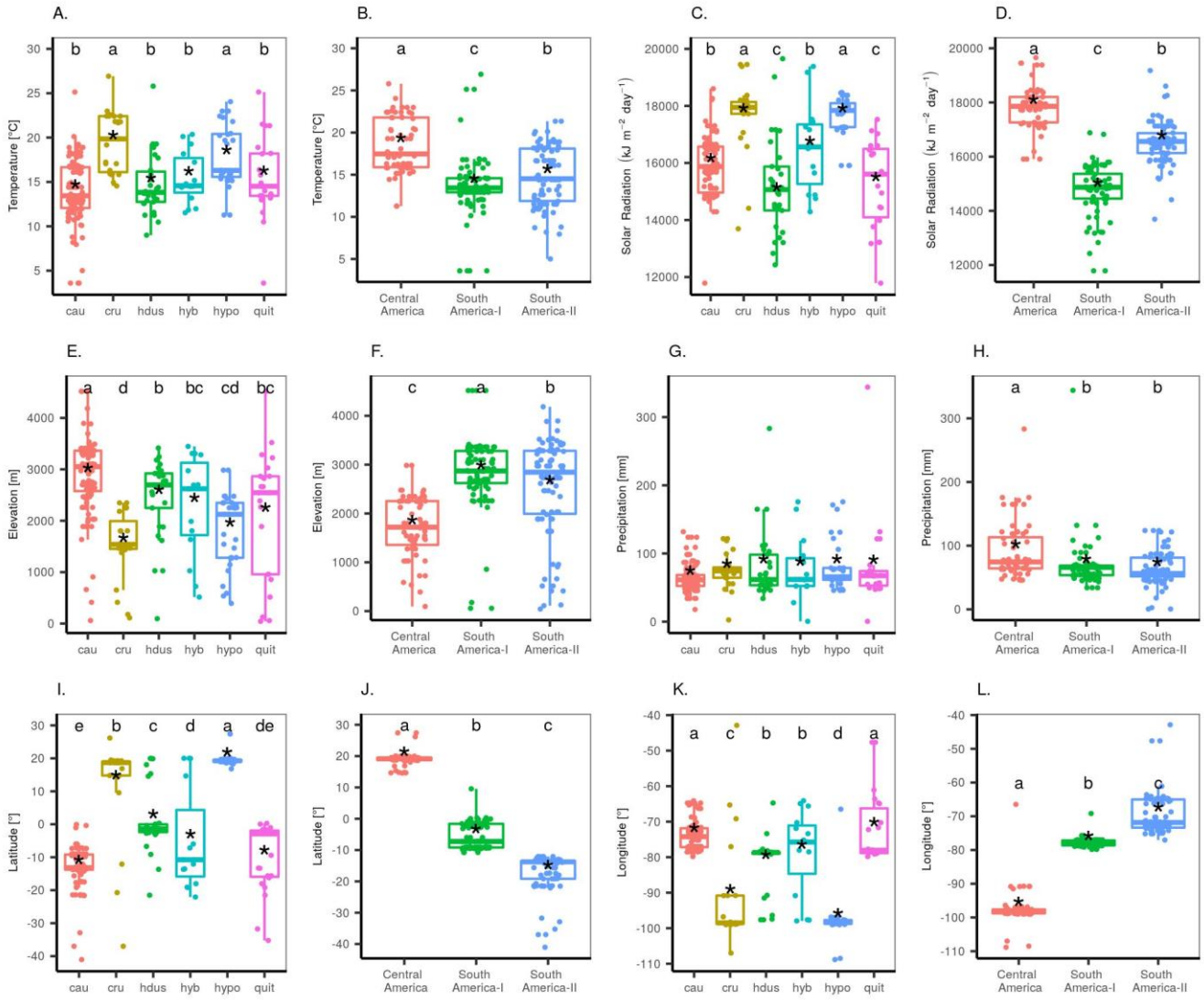

Figure S2. Box and bar plots of six environmental variables grouped by species and latitudinal groups, in each box and bar plot the respective groups were compared using a least significant difference (LSD) test and the groups with the different letters are significantly different at alpha=0.05. No letters if there is no significant difference among the compared groups. Black asterisks indicate the mean values of each group in the box plots. Species are cau, *A. caudatus*; cru, *A. cruentus*; hdus, *A. hybridus*; *hyb, the* 'hybrid' group; hypo, *A. hypochondriacus* and quit, *A. quitensis*.

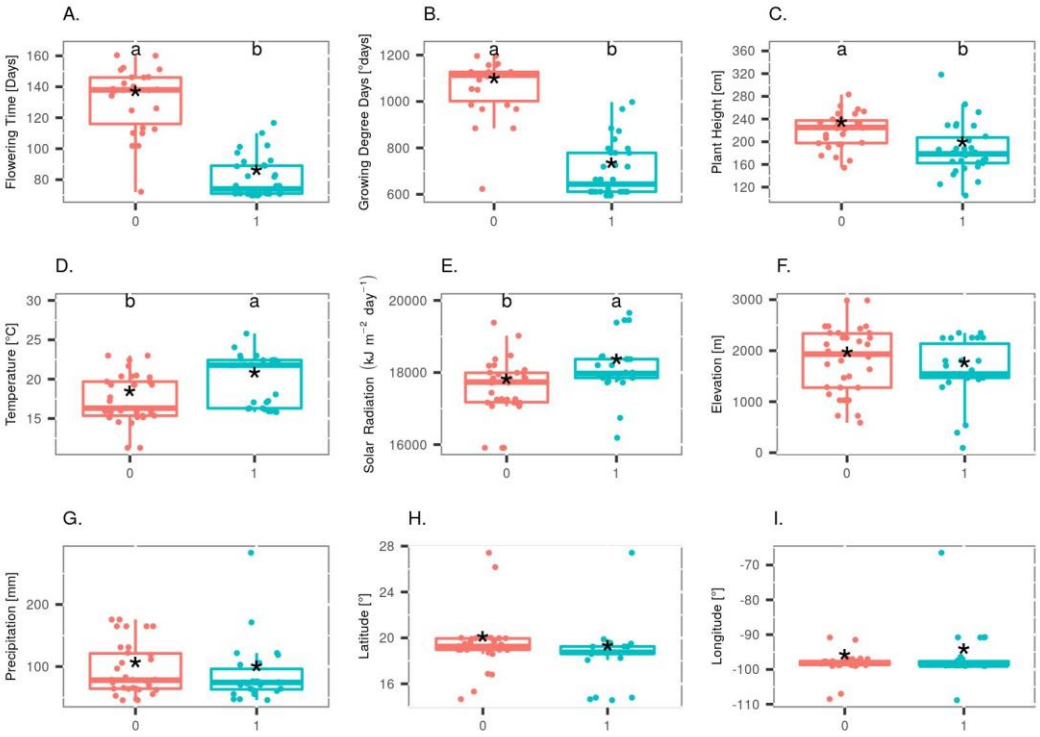

Figure S3. Box plots of three phenotypic traits and six environmental variables among the Central American accessions that were grouped by seed setting '1' or no seed setting '0'. In each box plot the two groups were compared using a least significant difference (LSD) test and the groups with different letters are significantly different at alpha=0.05. No letters if there is no significant difference among the two groups. Black asterisks indicate the mean values of each group in the box plots.

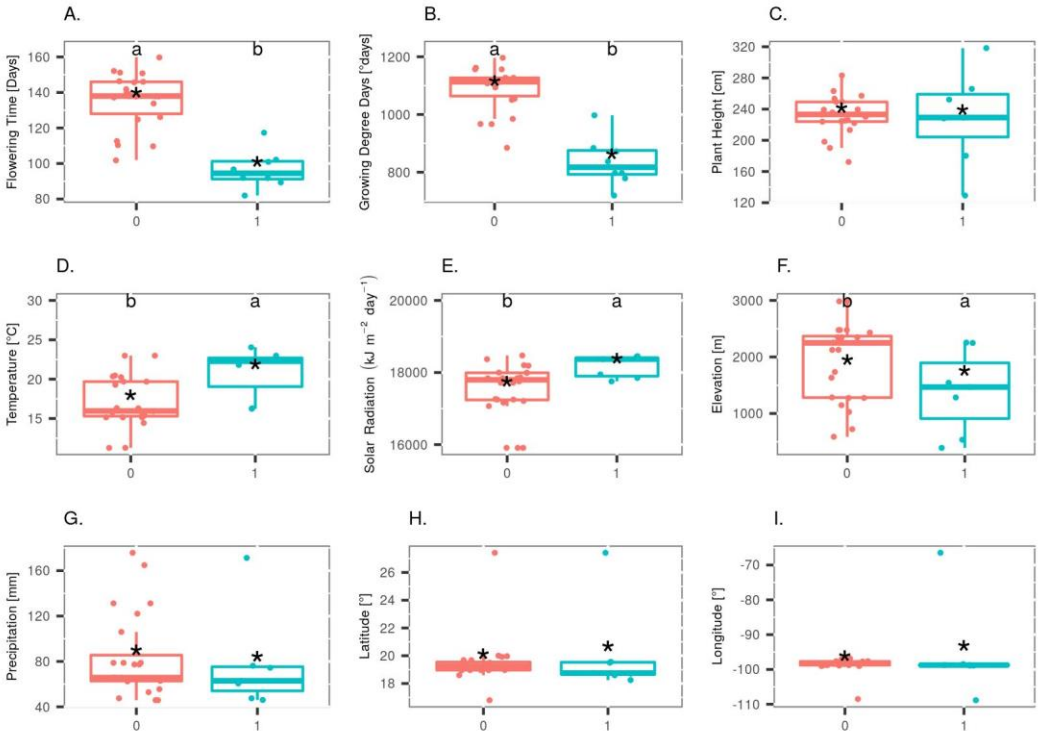

Figure S4. Box plots of three phenotypic traits and six environmental variables among the Central American *A. hypochondriacus* accessions that were grouped by seed setting '1' or no seed setting '0'. In each box plot the two groups were compared using a least significant difference (LSD) test and the groups with different letters are significantly different at alpha=0.05. No letters if there is no significant difference among the two groups. Black asterisks indicate the mean values of each group in the box plots.

SUPPLEMENTARY TABLES

Table S2. List of the accessions

No	Species	Country	Accession Number	Source
1	*A. caudatus*	Peru	PI 649242	USDA ARS
2	*A. caudatus*	Peru	PI 490565	USDA ARS
3	*A. caudatus*	Peru	PI 490445	USDA ARS
4	*A. caudatus*	Peru	PI 649227	USDA ARS
5	*A. caudatus*	Peru	PI 511686	USDA ARS
6	*A. caudatus*	Peru	PI 649222	USDA ARS
7	*A. caudatus*	Peru	PI 490552	USDA ARS
8	*A. caudatus*	Peru	PI 490551	USDA ARS
9	*A. caudatus*	Peru	PI 649240	USDA ARS
10	*A. caudatus*	Bolivia	PI 490459	USDA ARS
11	*A. caudatus*	Peru	PI 490519	USDA ARS
12	*A. caudatus*	Peru	PI 490562	USDA ARS
13	*A. caudatus*	Peru	PI 481960	USDA ARS
14	*A. caudatus*	Peru	PI 490559	USDA ARS
15	*A. caudatus*	Peru	PI 490538	USDA ARS
16	*A. caudatus*	Bolivia	PI 669838	USDA ARS
17	*A. caudatus*	Peru	PI 490536	USDA ARS
18	*A. caudatus*	Peru	PI 511687	USDA ARS
19	*A. caudatus*	Bolivia	Ames 15157	USDA ARS
20	*A. caudatus*	Bolivia	PI 490462	USDA ARS
21	*A. caudatus*	Bolivia	PI 642741	USDA ARS
22	*A. caudatus*	Peru	PI 490458	USDA ARS
23	*A. caudatus*	Peru	PI 490537	USDA ARS
24	*A. caudatus*	Peru	PI 490518	USDA ARS
25	*A. caudatus*	Bolivia	PI 490604	USDA ARS
26	*A. caudatus*	Peru	PI 490490	USDA ARS
27	*A. caudatus*	Peru	PI 481950	USDA ARS
28	*A. caudatus*	Peru	PI 649235	USDA ARS
29	*A. caudatus*	Peru	PI 649217	USDA ARS
30	*A. caudatus*	Peru	PI 490521	USDA ARS
31	*A. caudatus*	Peru	PI 490546	USDA ARS
32	*A. caudatus*	Peru	PI 490612	USDA ARS
33	*A. caudatus*	Peru	PI 490515	USDA ARS
34	*A. caudatus*	Peru	PI 511696	USDA ARS
35	*A. caudatus*	Peru	PI 490450	USDA ARS
36	*A. caudatus*	Argentina	PI 511679	USDA ARS

37	*A. caudatus*	Peru	PI 490552	USDA ARS
38	*A. caudatus*	Peru	PI 511683	USDA ARS
39	*A. caudatus*	Peru	Ames 15155	USDA ARS
40	*A. caudatus*	Peru	PI 490569	USDA ARS
41	*A. caudatus*	Peru	PI 490547	USDA ARS
42	*A. caudatus*	Peru	Ames 15139	USDA ARS
43	*A. caudatus*	Bolivia	PI 478403	USDA ARS
44	*A. caudatus*	Peru	PI 490423	USDA ARS
45	*A. caudatus*	Peru	PI 490534	USDA ARS
46	*A. caudatus*	Bolivia	PI 490579	USDA ARS
47	*A. caudatus*	Peru	PI 490539	USDA ARS
48	*A. caudatus*	Peru	PI 511690	USDA ARS
49	*A. caudatus*	Peru	PI 511704	USDA ARS
50	*A. caudatus*	Bolivia	PI 511681	USDA ARS
51	*A. caudatus*	Peru	PI 490553	USDA ARS
52	*A. caudatus*	Peru	PI 649245	USDA ARS
53	*A. caudatus*	Bolivia	PI 490582	USDA ARS
54	*A. caudatus*	Peru	PI 490540	USDA ARS
55	*A. caudatus*	Peru	PI 490535	USDA ARS
56	*A. caudatus*	Bolivia	PI 490583	USDA ARS
57	*A. caudatus*	Peru	PI 649236	USDA ARS
58	*A. caudatus*	Peru	PI 511688	USDA ARS
59	*A. caudatus*	Peru	PI 511689	USDA ARS
60	*A. caudatus*	Peru	PI 490524	USDA ARS
61	*A. caudatus*	Peru	PI 649228	USDA ARS
62	*A. caudatus*	Peru	PI 511753	USDA ARS
63	*A. caudatus*	Peru	PI 649234	USDA ARS
64	*A. caudatus*	Peru	PI 511694	USDA ARS
65	*A. caudatus*	Peru	PI 490511	USDA ARS
66	*A. caudatus*	Peru	PI 481957	USDA ARS
67	*A. caudatus*	Peru	Ames 15141	USDA ARS
68	*A. caudatus*	Peru	PI 511695	USDA ARS
69	*A. caudatus*	Peru	PI 490558	USDA ARS
70	*A. caudatus*	Ecuador	PI 511712	USDA ARS
71	*A. caudatus*	Peru	PI 490431	USDA ARS
72	*A. caudatus*	Peru	PI 490549	USDA ARS
73	*A. caudatus*	Peru	PI 490573	USDA ARS
74	*A. caudatus*	Peru	PI 490439	USDA ARS
75	*A. caudatus*	Peru	PI 649230	USDA ARS
76	*A. caudatus*	Peru	PI 490555	USDA ARS
77	*A. caudatus*	Peru	PI 649219	USDA ARS

78	*A. caudatus*	Bolivia	PI 568137	USDA ARS
79	*A. caudatus*	Argentina	Ames 15178	USDA ARS
80	*A. caudatus*	Bolivia	PI 490456	USDA ARS
81	*A. caudatus*	Bolivia	PI 568150	USDA ARS
82	*A. caudatus*	Peru	PI 490526	USDA ARS
83	*A. caudatus*	Peru	PI 481970	USDA ARS
84	*A. caudatus*	Bolivia	PI 490580	USDA ARS
85	*A. caudatus*	Peru	PI 511706	USDA ARS
86	*A. caudatus*	Peru	PI 649220	USDA ARS
87	*A. caudatus*	Peru	PI 490486	USDA ARS
88	*A. caudatus*	Bolivia	PI 490606	USDA ARS
89	*A. caudatus*	Bolivia	Ames 15156	USDA ARS
90	*A. caudatus*	Bolivia	PI 608018	USDA ARS
91	*A. caudatus*	Peru	PI 649244	USDA ARS
92	*A. caudatus*	Argentina	PI 511680	USDA ARS
93	*A. caudatus*	Peru	PI 481947	USDA ARS
94	*A. caudatus*	Peru	PI 481951	USDA ARS
95	*A. caudatus*	Peru	PI 490569	USDA ARS
96	*A. caudatus*	Peru	PI 490575	USDA ARS
97	*A. caudatus*	Peru	PI 481959	USDA ARS
98	*A. caudatus*	Peru	PI 649231	USDA ARS
99	*A. caudatus*	Peru	PI 511701	USDA ARS
100	*A. caudatus*	Peru	PI 481967	USDA ARS
101	*A. caudatus*	Bolivia	Ames 13860	USDA ARS
102	*A. caudatus*	Peru	Ames 5231	USDA ARS
103	*A. caudatus*	Argentina	PI 490491	USDA ARS
104	*A. caudatus*	Peru	PI 511693	USDA ARS
105	*A. caudatus*	Peru	PI 649226	USDA ARS
106	*A. caudatus*	Bolivia	PI 490457	USDA ARS
107	*A. caudatus*	Peru	PI 490561	USDA ARS
108	*A. caudatus*	Bolivia	PI 490461	USDA ARS
109	*A. caudatus*	Peru	PI 649233	USDA ARS
110	*A. caudatus*	Peru	PI 490447	USDA ARS
111	*A. caudatus*	Ecuador	PI 608019	USDA ARS
112	*A. caudatus*	Peru	PI 649237	USDA ARS
113	*A. caudatus*	Bolivia	PI 568144	USDA ARS
114	*A. caudatus*	Peru	PI 490488	USDA ARS
115	*A. caudatus*	Peru	PI 490452	USDA ARS
116	*A. cruentus*	Mexico	PI 477913	USDA ARS
117	*A. cruentus*	Peru	PI 511713	USDA ARS
118	*A. cruentus*	Mexico	PI 576482	USDA ARS

119	*A. cruentus*	Mexico	PI 643037	USDA ARS
120	*A. cruentus*	Mexico	PI 477914	USDA ARS
121	*A. cruentus*	Mexico	PI 511723	USDA ARS
122	*A. cruentus*	Mexico	PI 649509	USDA ARS
123	*A. cruentus*	Mexico	PI 606798	USDA ARS
124	*A. cruentus*	Mexico	PI 649514	USDA ARS
125	*A. cruentus*	Brazil	PI 667165	USDA ARS
126	*A. cruentus*	Guatemala	PI 451826	USDA ARS
127	*A. cruentus*	Mexico	Ames 5552	USDA ARS
128	*A. cruentus*	USA	PI 606767	USDA ARS
129	*A. cruentus*	Mexico	PI 511876	USDA ARS
130	*A. cruentus*	Venezuela	PI 665286	USDA ARS
131	*A. cruentus*	Argentina	PI 636182	USDA ARS
132	*A. cruentus*	Mexico	PI 662284	USDA ARS
133	*A. cruentus*	Mexico	PI 643058	USDA ARS
134	*A. cruentus*	USA	PI 515959	USDA ARS
135	*A. cruentus*	Mexico	PI 643039	USDA ARS
136	*A. cruentus*	Mexico	PI 649524	USDA ARS
137	*A. cruentus*	Guatemala	PI 511717	USDA ARS
138	*A. cruentus*	Peru	PI 511714	USDA ARS
139	*A. cruentus*	Mexico	Ames 15191	USDA ARS
140	*A. cruentus*	USA	PI 658731	USDA ARS
141	*A. cruentus*	Mexico	PI 658728	USDA ARS
142	*A. cruentus*	Mexico	PI 643042	USDA ARS
143	*A. cruentus*	Mexico	PI 643063	USDA ARS
144	*A. cruentus*	Mexico	PI 576481	USDA ARS
145	*A. cruentus*	Mexico	PI 649609	USDA ARS
146	*A. cruentus*	Mexico	Ames 2240	USDA ARS
147	*A. cruentus*	Guatemala	PI 433228	USDA ARS
148	*A. cruentus*	Mexico	PI 643049	USDA ARS
149	*A. cruentus*	Germany	Baernkrafft	Commercial
150	‘hybrid’ group	Peru	PI 511684	USDA ARS
151	‘hybrid’ group	USA	PI 568179	USDA ARS
152	‘hybrid’ group	Mexico	PI 604564	USDA ARS
153	‘hybrid’ group	Peru	PI 490430	USDA ARS
154	‘hybrid’ group	Peru	PI 490453	USDA ARS
155	‘hybrid’ group	Bolivia	PI 490586	USDA ARS
156	‘hybrid’ group	Peru	PI 490424	USDA ARS
157	‘hybrid’ group	Mexico	PI 604566	USDA ARS
158	‘hybrid’ group	Peru	PI 511752	USDA ARS
159	‘hybrid’ group	Mexico	PI 604571	USDA ARS

160	'hybrid' group	Guatemala	Ames 21996	USDA ARS
161	'hybrid' group	Peru	PI 490449	USDA ARS
162	'hybrid' group	Peru	PI 490441	USDA ARS
163	'hybrid' group	Bolivia	PI 511734	USDA ARS
164	'hybrid' group	Bolivia	PI 511735	USDA ARS
165	'hybrid' group	Peru	PI 511733	USDA ARS
166	*A. hybridus*	Ecuador	PI 490681	USDA ARS
167	*A. hybridus*	Ecuador	PI 490689	USDA ARS
168	*A. hybridus*	Peru	Ames 5232	USDA ARS
169	*A. hybridus*	Peru	PI 490740	USDA ARS
170	*A. hybridus*	Ecuador	PI 490716	USDA ARS
171	*A. hybridus*	Ecuador	PI 490682	USDA ARS
172	*A. hybridus*	Ecuador	PI 490731	USDA ARS
173	*A. hybridus*	Ecuador	PI 490739	USDA ARS
174	*A. hybridus*	Ecuador	PI 490684	USDA ARS
175	*A. hybridus*	Ecuador	PI 490715	USDA ARS
176	*A. hybridus*	Ecuador	PI 490676	USDA ARS
177	*A. hybridus*	Guatemala	PI 667158	USDA ARS
178	*A. hybridus*	Mexico	PI 604574	USDA ARS
179	*A. hybridus*	Ecuador	PI 490664	USDA ARS
180	*A. hybridus*	Ecuador	PI 490685	USDA ARS
181	*A. hybridus*	Peru	PI 490489	USDA ARS
182	*A. hybridus*	Ecuador	PI 490722	USDA ARS
183	*A. hybridus*	Mexico	PI 511724	USDA ARS
184	*A. hybridus*	Ecuador	PI 490679	USDA ARS
185	*A. hybridus*	Ecuador	PI 490670	USDA ARS
186	*A. hybridus*	Ecuador	PI 667156	USDA ARS
187	*A. hybridus*	Ecuador	PI 490703	USDA ARS
188	*A. hybridus*	Ecuador	PI 490721	USDA ARS
189	*A. hybridus*	Ecuador	PI 511754	USDA ARS
190	*A. hybridus*	Ecuador	PI 490678	USDA ARS
191	*A. hybridus*	Mexico	PI 604568	USDA ARS
192	*A. hybridus*	Guatemala	Ames 22001	USDA ARS
193	*A. hybridus*	Ecuador	PI 490677	USDA ARS
194	*A. hybridus*	Bolivia	Ames 5335	USDA ARS
195	*A. hybridus*	Mexico	PI 604582	USDA ARS
196	*A. hybridus*	Ecuador	PI 490680	USDA ARS
197	*A. hypochondriacus*	Mexico	PI 604595	USDA ARS
198	*A. hypochondriacus*	Mexico	PI 649607	USDA ARS
199	*A. hypochondriacus*	Mexico	PI 649623	USDA ARS
200	*A. hypochondriacus*	Mexico	PI 649565	USDA ARS

201	*A. hypochondriacus*	Mexico	PI 643070	USDA ARS
202	*A. hypochondriacus*	Mexico	PI 649633	USDA ARS
203	*A. hypochondriacus*	Mexico	PI 649529	USDA ARS
204	*A. hypochondriacus*	Mexico	PI 643041	USDA ARS
205	*A. hypochondriacus*	Mexico	PI 477916	USDA ARS
206	*A. hypochondriacus*	Chile	Ames 5355	USDA ARS
207	*A. hypochondriacus*	USA	PI 568127	USDA ARS
208	*A. hypochondriacus*	Mexico	PI 649537	USDA ARS
209	*A. hypochondriacus*	Mexico	PI 643067	USDA ARS
210	*A. hypochondriacus*	Mexico	PI 649532	USDA ARS
211	*A. hypochondriacus*	Mexico	Ames 2215	USDA ARS
212	*A. hypochondriacus*	Mexico	PI 649559	USDA ARS
213	*A. hypochondriacus*	Mexico	PI 649575	USDA ARS
214	*A. hypochondriacus*	Mexico	PI 649595	USDA ARS
215	*A. hypochondriacus*	Puerto Rico	Ames 5149	USDA ARS
216	*A. hypochondriacus*	Mexico	PI 633589	USDA ARS
217	*A. hypochondriacus*	Mexico	Ames 5457	USDA ARS
218	*A. hypochondriacus*	Mexico	PI 649544	USDA ARS
219	*A. hypochondriacus*	Mexico	PI 604587	USDA ARS
220	*A. hypochondriacus*	Mexico	PI 643036	USDA ARS
221	*A. hypochondriacus*	Mexico	PI 619247	USDA ARS
222	*A. hypochondriacus*	Brazil	Ames 5690	USDA ARS
223	*A. hypochondriacus*	Mexico	PI 649602	USDA ARS
224	*A. hypochondriacus*	Mexico	PI 649535	USDA ARS
225	*A. hypochondriacus*	Mexico	Ames 2085	USDA ARS
226	*A. hypochondriacus*	Mexico	PI 649602	USDA ARS
227	*A. hypochondriacus*	Mexico	PI 643074	USDA ARS
228	*A. hypochondriacus*	Mexico	PI 604559	USDA ARS
229	*A. hypochondriacus*	Mexico	PI 604581	USDA ARS
230	*A. hypochondriacus*	Mexico	PI 649587	USDA ARS
231	*A. hypochondriacus*	Mexico	PI 511731	USDA ARS
232	*A. quitensis*	Ecuador	PI 511747	USDA ARS
233	*A. quitensis*	Brazil	PI 652428	USDA ARS
234	*A. quitensis*	Ecuador	PI 490720	USDA ARS
235	*A. quitensis*	Ecuador	PI 490673	USDA ARS
236	*A. quitensis*	Peru	PI 490466	USDA ARS
237	*A. quitensis*	Ecuador	PI 511737	USDA ARS
238	*A. quitensis*	Ecuador	PI 511739	USDA ARS
239	*A. quitensis*	Brazil	PI 652422	USDA ARS
240	*A. quitensis*	Ecuador	PI 490705	USDA ARS
241	*A. quitensis*	Peru	PI 649246	USDA ARS

242	*A. quitensis*	Ecuador	PI 511745	USDA ARS
243	*A. quitensis*	Bolivia	PI 511736	USDA ARS
244	*A. quitensis*	Peru	Ames 5342	USDA ARS
245	*A. quitensis*	Argentina	Ames 21666	USDA ARS
246	*A. quitensis*	Ecuador	PI 511738	USDA ARS
247	*A. quitensis*	Peru	PI 511751	USDA ARS
248	*A. quitensis*	Brazil	PI 652426	USDA ARS
249	*A. quitensis*	Ecuador	PI 511749	USDA ARS
250	*A. quitensis*	Argentina	Ames 5334	USDA ARS
251	*A. quitensis*	Ecuador	Ames 5247	USDA ARS
252	*A. quitensis*	Bolivia	PI 568154	USDA ARS
253	*A. quitensis*	Ecuador	PI 490709	USDA ARS
254	*A. quitensis*	Ecuador	PI 511741	USDA ARS

Table S3. Accessions that did not flower

No	Species	Country	Accession Number	Plot
1	*A. hypochondriacus*	Mexico	PI 649602	714
2	'hybrid' group	Mexico	PI 604571	725
3	*A. hybridus*	Mexico	PI 604568	1010
4	*A. hypochondriacus*	Mexico	PI 649587	1014
5	*A. hybridus*	Guatemala	Ames 22001	1016
6	*A. hybridus*	Mexico	PI 604582	1027

Table S4. Photoperiod insensitive accessions that are of non-Central American origin but set seed

No	Species	Country	Accession Number/Name	Plot	Flowering time (day)
1	*A. quitensis*	Brazil	PI 652428	118	64
2	*A. cruentus*	Germany	Bärnkrafft	125	71
3	'hybrid' group	USA	PI 568179	216	65
4	*A. hypochondriacus*	USA	PI 568127	315	71
5	*A. cruentus*	USA	PI 606767	424	76
6	*A. quitensis*	Brazil	PI 652422	428	60
7	*A. cruentus*	Argentina	PI 636182	517	71
8	'hybrid' group	Peru	PI 511752	717	50
9	*A. cruentus*	Peru	PI 658731	811	71
10	*A. quitensis*	Peru	PI 511751	817	50

Table S5. Mild photoperiod sensitive accessions that are of non-Central American origin but set seed

No	Species	Country	Accession Number	Plot	Flowering time (day)
1	*A. hypochondriacus*	Chile	Ames 5355	309	110
2	*A. cruentus*	Brazil	PI 667165	407	97
3	*A. cruentus*	Venezuela	PI 665286	508	110
4	*A. cruentus*	USA	PI 515959	609	82
5	*A. hypochondriacus*	Brazil	Ames 5690	711	102
6	*A. cruentus*	Peru	PI 511714	718	110

4. Linkage and Association Mapping in Grain Amaranth Populations Reveal the Genetic Basis of Photoperiod Sensitivity

Abstract

Flowering time plays an important role in the adaptation of plants to their environments. The photoperiodic pathway helps plants to fine-tune flowering time by regulating the response of plant internal mechanism – circadian clock – to the external cues. Amaranth (*Amaranthus* spp.) is a short-day crop that is used as grain and vegetable thanks to the high nutritional qualities. The genetic basis of photoperiod sensitivity in grain amaranths is of fundamental importance to utilize the existing adaptive variation in breeding efforts, however our knowledge about the genetic basis of this trait is very limited. We therefore phenotyped three different types of grain amaranth populations including (i) a bi-parental family, (ii) a population derived from spontaneous hybridization events between a number of cultivars, and (iii) a genebank population which consists of traditional varieties for adaptive traits such as flowering time, plant height and seed setting under the long-day conditions of Germany and investigated their genomic basis using association and linkage mapping. We identified the same consensus genomic region on chromosome 10 in all three populations that separates photoperiod sensitive from insensitive accessions, suggesting an oligogenic inheritance of photoperiod sensitivity. The consensus region hosts a promising candidate gene 'response regulator of two-component system', the homologs of which regulate photoperiodic response in many crop and model plant species including rice and *Arabidopsis thaliana*. The marker data of the F_2 bi-parental family revealed a dominance effect of photoperiod sensitivity over insensitivity in the consensus region. In addition, this consensus region was identified for all traits studied i.e., flowering time, plant height and seed setting, demonstrating that pleiotropic relationships caused by photoperiod sensitivity exist, consistent with a bimodal-like distribution of phenotypic traits. Finally, we identified an epistatic relationship between the consensus region and another locus, as some individuals with the photoperiod insensitivity allele showed a photoperiod sensitive phenotype. In summary, we characterized the genomic background of the adaptive photoperiod sensitivity trait in grain amaranth species to better

understand the adaptation genetics and the utilization of genetic resources in the breeding programs of grain amaranths.

Keywords: adaptation, amaranth, flowering, mode of inheritance, photoperiod sensitivity, pleiotropy, epistasis.

4.1. Introduction

Photoperiod sensitivity is an adaptive trait that adjusts flowering time in response to day-length and plays key roles in reproductive success and agricultural productivity. Tropical short-day plants typically flower in shorter time in tropical regions under short days, but they delay or even may reject flowering under long-day conditions at higher latitudes (Nakamichi, 2014). This pattern is well established in several tropical grass crops such as sorghum, rice and maize, and leads increased biomass accumulation due to the action of photoperiod sensitivity genes under long-day conditions (Fujino, 2003; Jung and Müller, 2009; Grieder *et al.*, 2012; Murphy *et al.*, 2014; Windpassinger *et al.*, 2015), which has been exploited in breeding programs of energy crops. In contrast, successful grain production requires early flowering as it may extend the grain filling phase and help to escape from unfavorable growth conditions such as cold or terminal drought stress (Jung and Müller, 2009; Haussmann *et al*., 2012). Photoperiod insensitivity provides earliness and also secures uniform flowering across diverse production environments (Ceccarelli, 1994). Hence, photoperiodic response variation determines the use of such crops for different purposes.

The genetic architecture of photoperiod sensitivity and its delaying effect under long days vary across different species, however, is generally more complex – managed by many genes with small effects – in out-crossing species compared to self-fertilizing species due to a higher number of recombination events (Hung *et al.*, 2012). For example, 14 QTLs were reported to control photoperiodic response in maize, the strongest of which delayed flowering only by three days under long days (Buckler *et al.*, 2009; Hung *et al.*, 2012). In contrast, only the *Ma1* locus in sorghum delayed flowering by 60 days and the additive action of the dominant alleles of *Ma1* and *Ma6* loci extended flowering time to 175 days under long days, where the

combination of the recessive alleles caused an average flowering time of 70 days (Murphy *et al.*, 2011, 2014). In rice, the major grain number, plant height and heading date QTL (*Ghd7.1*) alone delayed heading by up to 12.5 days under long days, suggesting a similar, but less complex genetic control of photoperiodic response in a self-fertilizing crop (Liu *et al.*, 2013). Such alleles with large effects may also lead remarkable morphological changes in the plant architecture due to pleiotropic effects. In sorghum, a combination of recessive alleles at the *Ma1* and *Ma6* loci leads early flowering plants used for grain production, whereas the combination of dominant alleles delays flowering which allows to cultivate such genotypes as bioenergy crops (Olson *et al.*, 2012; Murphy *et al.*, 2014). In rice, several major QTLs controlling photoperiodic response were also reported with pleiotropic effects on heading date, grain yield and plant height (Yan *et al.*, 2011, 2013; Liu *et al.*, 2013; Wang *et al.*, 2019).

Amaranth is a minor pseudo-cereal crop that originated from Central and South America (Sauer, 1950, 1967). In recent decades, grain amaranth species (*A. cruentus, A. caudatus, and A. hypochondriacus*) have regained importance due to their high nutritional qualities such as high protein and lysine content (Becker *et al.*, 1981; Stallknecht and Schulz-Schaeffer, 1993; Assad *et al.*, 2017). Furthermore, amaranth seeds contain no gluten, which offers an alternative diet to people with celiac disease (Guerra-Matias and Arêas, 2005). *Amaranthus* is a diverse genus that comprises more than 60 species (Assad *et al.*, 2017). While cultivated amaranths are used as ornamental plant, grain crop, and leaf vegetable crop in tropical regions, most amaranth species are wild species and many of them are weeds in agricultural production areas (NRC, 1984; Costea and Tardif, 2004). Although amaranths mostly self-pollinate, they also show out-crossing with a frequency between 3-32 % (Jain *et al.*, 1982; Hauptli and Jain, 1985). Being a dicot plant that uses the C4 photosynthetic pathway, amaranth utilizes water efficiently and shows a wide adaptation range, including dry and marginal zones, which highlights amaranth as a resilient alternative in efforts to cope with the challenges of climate change (Myers, 1996; Liu and Stützel, 2004).

Amaranths originated from tropical regions and most amaranth species show photoperiod sensitivity, but photoperiod insensitivity also exists in the genus (Brenner *et al.*, 2000). Wu *et al*. (2000) reported photoperiodic variation in 229 amaranth accessions from 20 A*maranthus*

species in a characterization of agro-morphological traits. Recently, we characterized genebank accessions that belong to the three major grain amaranth species (*A. caudatus* L., *A. cruentus* L., *A. hypochondriacus* L.) and their two putative wild relative species (*A. hybridus* L. *and A. quitensis* Kunth) for flowering time under the long-day conditions of Europe and classified them into three groups according to their photoperiodic responses (Baturaygil and Schmid, 2022). Among the two grain amaranth species originating from Central America, *A. cruentus* was photoperiod insensitive, whereas *A. hypochondriacus* accessions ranged from photoperiod insensitive to highly sensitive. *A. caudatus*, which is native to South American highlands, showed mild photoperiod sensitivity. We also cultivated a diverse set of 11 amaranth genotypes under long-day conditions of Germany and observed two distinct growth patterns: A grain type variety of amaranth flowered early, had a short stature, and was photoperiod insensitive, whereas, other genotypes flowered late, had a high plant height and were photoperiod sensitive (Baturaygil *et al*., 2021). Based on these results, we proposed that variation in photoperiod sensitivity mainly differentiates these two types by controlling a large proportion of flowering time variation.

Reduced plant height and early flowering are the two main breeding objectives in grain amaranths and photoperiod insensitivity contributes to both objectives. It also synchronizes flowering time across different production environments with different photoperiods. Kulakow and Jain (1985) studied the genetic background of the flowering time in several backcross populations derived from a cross between *A. cruentus* and the weedy species *A. retroflexus*. They proposed a three-gene model, where one allele is responsible for reduced vegetative growth and two alleles are responsible for photoperiodic response. Altogether, photoperiod sensitivity showed a simple pattern of inheritance and seems to be controlled by a few major genes (Kulakow and Jain, 1985). This trait is important for amaranth breeding because it separates early flowering grain types from late flowering biomass types under long day conditions. These differences make amaranth an interesting model species for flowering time studies, similar to rice and sorghum, and a suitable candidate crop for marker assisted selection (MAS) of this trait because of a potentially simple genetic control of photoperiodic response (Bernardo, 2008). Breeding of grain amaranth has been largely limited to line selection of genebank populations (Joshi *et al.*, 2018). Meanwhile, breeding of biomass type

amaranth has been started (Baturaygil *et al.*, 2021) because various studies have shown sufficient potential of amaranth for bioenergy production in terms of dry matter yield and content (Mursec *et al.*, 2009; Pospišil, 2009; Svirskis, 2009; Balodis *et al.*, 2011; Seppälä *et al.*, 2013; Sitkey *et al.*, 2013; von Cossel *et al.*, 2017). In addition, a number of genomic resources have been created, such as a reference genome, which facilitate breeding (Clouse *et al.*, 2016; Lightfoot *et al.*, 2017; Stetter and Schmid, 2017). Recently, Lin *et al.* (2022) carried out genome-wide association studies (GWAS) using two different diversity panels including (i) grain amaranth species and (ii) *A. tricolor*, a major vegetable amaranth species, and reported some homolog genes that regulate flowering time in the model species *Arabidopsis thaliana*. Furthermore, two mapping studies reported the QTLs underlying some morphological traits such as flower and seed coat color (Lightfoot *et al.*, 2017; Stetter *et al.*, 2020) in grain amaranth species. However, no other agronomically important trait has been studied yet, according to our knowledge.

Given the importance of photoperiod sensitivity and flowering time for adaptation of grain amaranths in temperate regions, our objectives were to (i) characterize the genetic architecture of flowering time, seed setting and plant height using different types of populations and mapping methods to elucidate if photoperiod sensitivity has a simple inheritance that is controlled by a few genes, (ii) investigate the relationships between the underlying QTL to examine if photoperiod sensitivity guided pleiotropic effects exist, and (iii) examine the photoperiod sensitivity oriented potential epistatic relationships.

4.2. Materials and Methods

4.2.1. Mapping populations

We used three different mapping populations; (i) F_2 and F_3 generation families of a bi-parental cross using *A. hypochondriacus* accessions, which we name ‘the hypochondriacus family’, (ii) F_2 and F_3 generation families of some putative hybrids of unknown parental origin, which we name ‘the giant population’ because of their large plant height caused by heterosis and late

flowering, and (iii) a genebank population (**Table S6**) that includes accessions from the USDA genbank, eight lines from our breeding population for biomass amaranth, and the Bärnkrafft variety used for grain production (Baturaygil and Schmid, 2022; **Table 2**).

4.2.1.1. Generation of bi-parental mapping populations

We considered two criteria in parent selection for the generation of the bi-parental mapping population; (i) parents showing contrasting photoperiodic responses i.e. photoperiod sensitive and insensitive parents should have been crossed and (ii) since red stem color has a monogenic dominance over green stem color, a female parent with green stem and a male parent with red stem should have been crossed, where progeny of successful crosses are expected to have a red stem color in the F_1 generation (Kulakow and Jain, 1985). In this study, we conducted a greenhouse experiment using diverse genebank accessions. We set the photoperiod to eight hours, maintained a temperature of 30°C during light exposure and 25°C during darkness, and used 11 x 11 cm pots filled with Substrat 5 (Klasmann-Deilmann, Germany) as the growing medium. We identified parent candidates that met the stem color requirement and flowered at the same time and paired them by placing their pots together and covering their inflorescences with a breathable plastic bag (Sealed Air, Germany) to prevent pollen contamination. We daily shook the bags to promote pollen dispersal. One of the pair combinations (the hypochondriacus family; **Table 3**) met both criteria, and its F_2 population was generated by selfing its F_1 hybrids in a greenhouse. Additionally, we tested 10 genebank accessions, including the parents of the hypochondriacus cross, in speed breeding chambers under a 10-hour photoperiod (as described in Jähne *et al.*, 2020) to record their flowering times under short-day conditions (**Figure 8I**). In 2018, we grew hybrids from this F_2 family for visual assessment and seed harvest in two locations: Hohenheim and Oberer Lindenhof (**Table 4**). The F_3 population tested in the 2019 experiment was derived from the early flowering plants of the F_2 generation grown in the 2018 experiments. Therefore, we analyze only the 2019 field trial in this study and mention the 2018 experiments as the source of our F_3 population.

4.2.1.2. Origin of the biomass type 'giant' population

In 2012 and 2013, the Institute of Crop Science at Hohenheim University conducted a set of field experiments at a single location near Stuttgart to test the agronomic potential of ten amaranth genotypes using an identical experimental setup. In 2013, the seeds harvested from the 2012 experiments were used and some plants with large plant height ('giants') were observed. These were hypothesized to be the result of spontaneous crossing events between the tested genotypes in the 2012 experiments and to show strong heterosis. Only the giant plants that appeared in the plots of the Bärnkrafft (*A. cruentus*), C6 (*A. caudatus*), and Puerto Moutt (*A. cruentus*) varieties were harvested in 2013. These harvested plants established our biomass amaranth pool, were considered as F_1 generation hybrids, and harvested seeds from some of these F_1 individuals were grown in the greenhouse to produce F_2 generation hybrids, and then selfed again to produce the F_3 generation (Baturaygil *et al.*, 2021). Furthermore, 10 advanced lines selected from this base population were also tested for biomass yield components in our previous work (Baturaygil *et al.*, 2021). In 2019, we phenotyped 14 F_2 and 5 F_3 families derived from these giant plants for flowering time and plant height (**Table 2**).

4.2.1.3. Genebank accessions

In our set of genebank accessions, we included 163 of the 254 genebank accessions representing traditional cultivars from the centers of origin that we phenotyped in our previous study (Baturaygil and Schmid, 2022). These accessions comprise the three species of grain amaranths (*A. caudatus* L., *A. cruentus* L., and *A. hypochondriacus* L.), their two wild putative ancestor species (*A. hybridus* L. and *A. quitensis* Kunth), and accessions described as 'hybrids' in their passport data. We also included eight genotypes from our biomass amaranth breeding pool (Baturaygil *et al.*, 2021) and the commercial grain variety 'Bärnkrafft' (*A. cruentus*), that was developed for Central Europe climatic conditions (Baturaygil and Schmid, 2022).

4.2.2. Field trials

In 2019, we conducted field experiments at the Heidfeldhof experimental station of Hohenheim University, Germany, to phenotype three populations for flowering time, plant height, and seed setting (**Table 2**). We only evaluated seed setting in the genebank population. The experimental procedures for the genebank population were described in Baturaygil & Schmid (2022). For the giant and hypochondriacus populations, we tested 19 giant families with two replicates, for a total of 38 plots, and F_2 and F_3 generation hypochondriacus families in six and 19 plots, respectively. Each of the parental genotypes was also tested in a single plot. The experiments were performed in single-row plots of 5 meter length with 0.75 meter distance between rows. The experiments were surrounded by the grain variety Bärnkrafft. Seeds were manually sown, and plots were thinned manually in the seedling stage by leaving 10 cm distance between plants. We controlled weeds manually and mechanically when necessary. No fertilizer or irrigation was applied. We phenotyped all families for flowering time, and the F_2 generation hypochondriacus, giant families, and genebank population were also phenotyped for plant height. Plant height was measured at harvest as the distance between the ground surface and the highest level of the inflorescence in cm, and flowering time was recorded as the number of days between sowing and pollen spread in the inflorescence. We randomly selected and labeled several flowering individuals in both the giant and hypochondriacus populations to capture variation in flowering time throughout the vegetation period. We also measured plant height on these labeled individuals. In the genebank population, seed setting was recorded as a binary trait on three randomly selected individuals, where '0' represents no seed setting and '1' represents seed setting. We estimated monthly mean values for temperature, precipitation, and photoperiod during the field experiments for each location (**Table 4** and **Figure S5**) as described in Baturaygil and Schmid (2022). Temperature data was obtained from stations with the same names as the experiment locations from http://wetter-bw.de, and photoperiod data for the Hohenheim experiment was obtained from the Hohenheim location, while the data for the Oberer Lindenhof experiment was obtained from the Eningen unter Achalm location from http://www.timeanddate.com.

4.2.3. Whole genome sequencing experiments

A single leaf per individual was collected during the field trials after recording flowering time and dried with the aid of silica gel. DNA was extracted using the AX Gravity DNA extraction kit (A&A Biotechnology, Gdynia, Poland) following manufacturer's instructions. Purity and quality of DNA were controlled by agarose gel electrophoresis and concentration determined with a Qubit instrument using SYBR green staining. DNA sequencing libraries were constructed using the protocol of Baym *et al*. (2015). Whole-genome sequencing was done with short-read Illumina sequencing on an Illumina NovaSeq machine (Novogene). Raw sequences of some of the accessions (n=105) from the genebank population (**Table S6**) were taken from Stetter *et al.* (2020).

4.2.4. Data processing and filtering

Sequencing reads were trimmed with TrimGalore v0.6.7 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were then mapped to the *Amaranthus hypochondriacus* v.2.1 reference genome (Lightfoot *et al.*, 2017) with bwa mem v0.7.17 (Li and Durbin, 2009). Mapping summaries were calculated using Qualimap v2.2.1 (García-Alcalde *et al.*, 2012; Okonechnikov *et al.*, 2015). Duplicate reads were removed with picard MarkDuplicates v2.17 (http://broadinstitute.21572 github.io/picard/) and afterwards variants were discovered with bcftools v1.11 (Li, 2011), filtering out reads with mapping quality < 20 and base quality < 28 and generating one SNP matrix that include only variant positions. After SNP calling, we subsetted each of the populations from the major SNP matrix with only variant sites that includes all samples, filtered separately using vcftools with the given parameters in **Table S7,** and used them in our further analyses. Furthermore, we retained only biallelic loci in all three populations. We imputed missing data only in the giant population using PLINK v1.90b6.24 (Chang *et al.*, 2015). In particular, we used the non-imputed data in the principal component analysis (PCA) and the imputed dataset in GWAS in the giant population. LD pruning was performed using SNPRelate package version 1.12.1 in R environment (Zheng *et al.*, 2012) with a window size of 10.000 bp and LD threshold of 0.3 in all three populations. Finally, only SNPs located on the 16 largest scaffolds of the reference genome were retained.

4.2.5. Population genetic analyses

We performed principal component analyses in the genebank and giant populations using SNPRelate R package version 1.12.1. Furthermore, we estimated LD decay only in the genebank population using PopLDdecay (Zhang *et al.*, 2019), as the low coverage genetic data did not allow this analysis with the giant population.

4.2.6. Linkage and association mapping

We performed linkage mapping using the marker regression model in R/qtl package version 1.48.1 (Broman *et al.*, 2003) to investigate marker-trait associations in the hypochondriacus family. We converted the SNP matrix from VCF format to ABH, which is the default input format of R/qtl package, using the VCF2ABH script (Ogiso-Tanaka *et al.*, 2020). However, as the last individual in the ABH data format in each dataset seems to possess the alleles only of the male parent, we discarded that last individual from the analyses. We investigated the marker-trait associations in the genebank and giant populations using association mapping. In both populations, we used FarmCPU (Fixed and random model Circulating Probability Unification) model in GAPIT package version 3 (Lipka *et al.*, 2012) in the R environment. FarmCPU is an iterative method that improves the (i) false positive control and (ii) the confounding between co-factors and markers (Liu *et al.*, 2016). In the analyses of both populations, we took the population structure into account by inserting the pre-estimated Q matrix into the model, which includes the first three principal components estimated using SNPrelate package, since the Q matrices estimated by GAPIT are imputed and different from the matrices we estimated.

4.3. Results

4.3.1. Phenotypic variation

4.3.1.1. Hypochondriacus families

In the F_2 population, we observed two distinct groups with a bimodal distribution for flowering time, referred to as the early and late bulks (**Figure 8A**). The difference in flowering time between these groups was 50 days. Of the 154 individuals observed, 59 flowered between days 46 and 76, with a mean of 57.6 days (n=58). We named this group the 'early bulk' and it was composed of photoperiod insensitive individuals. The second group, which we named the 'late bulk,' consisted of 92 individuals that flowered between days 125 and 160 with a mean of 140.4 days (n=48) and also included individuals that did not flower (n=44). This group was composed of photoperiod sensitive individuals. In both bulks, we observed a transgressive segregation pattern for flowering time. In the early bulk, all individuals flowered earlier than the early flowering parent (female parent, 82 days), and in the late bulk, 67 individuals flowered later than the late parent (male parent, 138 days) or did not flower (**Figure 8A**). We also identified three individuals that flowered between days 102 and 113 and categorized them as intermediates. In the whole F_2 population, flowering time varied between 46 and 160 days among flowering and phenotyped individuals (n=109) with a mean of 95.3 and the median of 76 days. Additionally, only the early bulk individuals set seed. In the F_3 population, which was derived from the early bulk plants of the previous years' F_2 generation, all 113 individuals flowered between days 46 and 77 with a mean of 61.6 days, showed a similar transgressive pattern, and also set seed, which behaved almost identically to the early bulk plants of the F_2 generation (**Figure 8B**).

The trait distributions and scatter plot (**Figure 8A, Figure S6A,** and **D**) clearly illustrate that early bulk plants tend to flower early and reach a shorter final plant height, similar to the photoperiod insensitive female parent, while late bulk plants flower late or do not flower and reach a taller final plant height, similar to the photoperiod sensitive male parent. In addition, we compared if the early and late bulks differ for flowering time and plant height using a generalized linear model and the analyses revealed highly significant differences among the bulks for both traits (p<2.2e-16; **Figure 8G** and **H**).

We phenotyped 143 individuals for plant height in the F_2 generation and observed a bimodal pattern for plant height as well. However, the individuals in the early and late bulks were more

closely clustered than in flowering time (**Figure S6A**). The individuals ranged between 57 and 366 cm, with a mean of 194.6 cm. Of those 143 individuals, 46 were shorter than the short parent (female parent, 129 cm) and 61 were taller than the long parent (male parent, 235 cm). Furthermore, the early bulk plants ranged between 57 and 207 cm with a mean of 103.5 cm (*n*=56), while late bulk plants ranged between 75 and 366 cm, with a mean of 252 cm (*n*=48).

4.3.1.2. Giant families

In the giant families, the number of individuals phenotyped per family varied depending on the variation in flowering time observed within the families (**Table 2**). When considering all the giant families together, we observed a multinomial-like distribution in flowering time (*n*=166, **Figure 8C**). Flowering time ranged between 41 and 106 days, with a mean of 70.5 days (*n*=106), which is narrower compared to the range of the hypochondriacus F_2 family, due to the lack of individuals that flowered very late and did not flower at all. In contrast to flowering time, we observed a bimodal-like distribution for plant height (*n*=164) within giant families that ranged between 70 and 372 cm, with a mean of 225.5 cm (**Figure S6B**).

In both F_2 and F_3 generations of the hypochondriacus families, and in the giant population, we observed a specific phenotype (**Figure 8E**), which showed unique morphological attributes such as a large inflorescence, short plant height, a strong propensity for shoot branching and lodging. This phenotype is especially characterized by a very early flowering that is driven by photoperiod insensitivity. Accordingly, we refer it as the ‘dwarf’ phenotype. All of the early bulk individuals in the F_2 generation and all F_3 individuals showed transgressive segregation by flowering earlier and mostly reaching a shorter plant height than the female parent. However, only a certain fraction of the early bulk individuals – the ones that flowered between 40 and 60 days, based on our visual assessment – showed the morphological attributes of the ‘dwarf’ phenotype.

4.3.1.3. Genebank population

In the phenotypic analysis of the genebank population, we only included biomass genotypes in the investigation of trait distributions and the relationship between phenotypic traits. Furthermore, we divided the accessions into three latitudinal groups based on their collection sites (Central America, South America-I and II; **Figure S7**) and their species as described in (Baturaygil and Schmid, 2022). Accordingly, we compared the means of these groups in phenotypic traits to understand the factors responsible for divergence within the population, such as different flowering times, that potentially create the population structure using a generalized linear model as described in Baturaygil and Schmid, (2022). In contrast, we excluded biomass genotypes in the group comparisons as they do not belong to any species or have latitude information. We would like to remind that *A. cruentus* and *A. hypochondriacus* are from Central America, whereas *A. caudatus* is native to South America among the grain amaranth species. Therefore, we compared the species and the latitudinal groups with 163 and 140 accessions, respectively. In flowering time, we observed a wide variation that ranged between 50 and 160 days, and the mean flowering time was 103.7 days (n=160). In addition, six genotypes did not flower (**Table S3**). In the comparison of the species for flowering time, three major grain species were different from each other (**Figure S8A**), however, we detected no difference among the three latitudinal groups (**Figure S8B**).

In the analysis of plant height, the accessions ranged between 45.2 and 359.9 cm, with an average height of 218.8 cm (n=156). In the comparison between the species, *A. caudatus* was taller than *A. cruentus* by approximately 66 cm among the grain species (**Figure S8C**). Similarly, the South American groups were not different from each other, but taller than the Central American group in the comparison of latitudinal groups (**Figure S8D**). We recorded seed setting as a binary trait, where '0' represents no seed setting and '1' represents seed setting. The mean seed setting rate was only 32% within the genebank population (n=163). In the comparison of species, *A. cruentus* had the highest seed setting rate at 88%, followed by *A. hypochondriacus* at 31%, and no *A. caudatus* accessions set seed (**Figure S8E**). In the comparison of latitudinal groups, the Central American group had a seed setting rate of 44%, while the South America I and II groups had rates of 2% and 12% respectively (**Figure S8F**).

4.3.4. Relationships between phenotypic traits

In the F_2 generation of the hypochondriacus family, we observed a positive, moderate to strong correlation between flowering time and plant height (r^2=0.78, p<2.2e-16, n=104; **Figure S6D**). As the correlation between the traits was strong until the day 120, but decayed with the later flowering, a polynomial regression model explained the trend better than the linear regression model by 0.03 improvement in r-squared (**Figure S6D**).

In the giant population, we observed a moderate positive correlation between flowering time and plant height, which started to decay after the day 95 (r^2=0.39, p=3.944e-12, n=100; **Figure S6E**). The polynomial model explained the data better than linear model with the improvement in r-squared by 0.04 (**Figure S6E**). Hence, the presence of a bimodal-like distribution in both traits and the correlation analysis suggest that the separation between early flowering short plants and late flowering tall plants was weakly pronounced in the giant population, compared to the F_2 population of the hypochondriacus family.

In the genebank population, we observed a moderate positive correlation between flowering time and plant height, which showed a linear trend until day 100, but remained stagnant with later flowering (r^2=0.36, p=1.717e-14, n=144; **Figure S6F**). Similar to the other populations, inclusion of a polynomial term increased the r-squared, from 0.21 to 0.36. Seed setting was moderately and negatively correlated to flowering time (r=-0.68, p<2.2e-16, n=160), whereas weakly and negatively correlated to plant height (r =-0.38, p=1.38e-06, n=156).

4.3.5. Genetic Analyses

4.3.5.1. Whole genome sequencing of populations

The SNP matrix that includes all samples belonging to the three main mapping populations was produced with a single variant calling pipeline. Accordingly, the SNP matrix of each respective mapping population has been subsetted from that major SNP matrix. Even though a

reference genome for another grain amaranth species *A. cruentus* is also available (Ma *et al.*, 2021), we used *A. hypochondriacus* reference genome because we used an *A. hypochondriacus* bi-parental family.

We compared the proportion of the mapped reads to the reference genome from different populations using ANOVA to detect whether there are differences among the species in the genebank population or among the giant and hypochondriacus families. We also compared the groups using a t-test, when ANOVA yielded a significant difference. The proportion of the reads mapped to the reference genome ranged between 92.60% (*A. hybridus*) and 98.18% ('hybrids') with the mean of 95.74%, and ANOVA detected no significant differences among the compared species (p=0.5083). In contrast, the comparison of the giant population (98.57%) and the hypochondriacus family (98.37%) found a significant difference (p=0.02481). The sequence depth per individual differed also across the populations (**Figure S9**); the genebank population ranged between 0.15X (PI 643063, *A. cruentus*) and 39.08X (PI 642741, *A. caudatus*) (Mean: 5.96X), the giant population ranged between 0.01X and 17.6X (Mean: 1.91X), and the hypochondriacus family ranged between 0.001X and 8.86X (Mean: 0.67X). In the hypochondriacus family, the male parent (PI 649623) and the female parent (Ames 5149) showed a sequencing depth of 7.76X and 8.86X, respectively. In particular, we did not impute our populations for the inference of the population structure to avoid any potential imputation bias, as we have sufficient number of polymorphisms in all populations without imputation. For association mapping, we used an imputed dataset only in the giant population because of the low mean coverage. Since the imputation process increased the already existing high segregation distortion in the hypochondriacus family due to the relatively low coverages of the parental accessions, we did not use an imputed dataset in this population in linkage mapping.

4.3.5.2. Population Structure in the Giant and Genebank populations

We investigated the population structure in the giant and genebank populations using PCA to identify potential effects on association studies (**Figure 9A-B**). The analysis was conducted

with a non-imputed dataset of 582.146 SNP markers and did not yield a recognizable population structure in the giant population. In contrast, we observed a very strong differentiation in the genebank population using a non-pruned and non-imputed dataset of 8.574.262 SNP markers. The first two components explained approximately 45% of the total variation. The first component separated the South American-origin grain species *A. caudatus* from the Central American-origin species *A. cruentus* and *A. hypochondriacus*. Another Central American-origin wild relative, *A. quitensis*, also clustered with *A. caudatus*. However, *A. hybridus*, the wild putative ancestor species with the widest geographical distribution, clustered with all three grain species. The second component separated *A. hypochondriacus* from the other two grain species.

Given that all the three grain species showed a large divergence in both flowering time variation and in the PCA, we studied the correlation between the principal components and the adaptive traits flowering time and seed setting to investigate if local adaptation may play a role in the population differentiation. We obtained highly significant relationships between seed setting and PC 1 (r=0.69, p<2.2e-16), flowering time and PC 1 (r=-0.42, p=9.19e-08) and between flowering time and PC 2 (r=0.45, p=5.106e-09).

We observed a low genome-wide LD only in the South American species *A. caudatus* and *A. quitensis*, whereas the other species dropped to a constant level of r^2 between 0.12 and 0.2 (**Figure 9C**). Values of r^2 *A. caudatus* reached 0.1 within 1-1.5 kb, whereas in *A. hypochondriacus* r^2 dropped to 0.12 within 95-100 kb and in *A. cruentus to* 0.15 within 130-140 kb.

4.3.5.3. Linkage and association mapping and candidate gene investigation

We performed linkage and association mapping using three different mapping populations to unravel the genetic architecture of adaptive traits such as flowering time, seed setting and plant height to investigate if (i) photoperiod sensitivity is influenced by a few or many genes, (ii) there are pleiotropic effects, and (iii) epistatic relationships between such photoperiod

sensitivity related traits and loci exist. We used the F_2 and F_3 families of the 'hypochondriacus' cross of two parents with contrasting responses to photoperiod for linkage mapping. To increase statistical power, we merged the F_2 and F_3 families and conducted the analysis with 11.565 SNP markers. The F_3 population was selected from the early flowering individuals of the F_2 population and a correction for segregation distortion might have adversely affected the analysis. Therefore, we did not construct a genetic map and a used marker regression model implemented in the R/qtl package. Additionally, 44 individuals from the F_2 family of the hypochondriacus family and six individuals from the genebank population did not flower, and we assigned a flowering time of 190 days to these genotypes to be able to use in the mapping studies (**Table S3**).

Our analysis with the merged dataset of F_2 and F_3 families (n = 265) found a significant QTL for flowering time on chromosome 10 (position: 15.010.436, LOD: 21.19; **Figure 10A**) above the threshold determined by a permutation test (Critical LOD score: 5.74). Additionally, we performed linkage mapping for flowering time (n = 152) and plant height (n = 143) traits in the F_2 population (17.218 SNP markers) and only for flowering time in the F_3 population (19.231 SNP Markers) of the hypochondriacus family. The analyses for both flowering time and plant height in the F_2 population identified the same QTL on chromosome 10 (position: 13.643.228, $LOD_{flowering\ time}$: 9.41, $LOD_{plant\ height}$: 7.59; **Figure S10A and B**) above the critical scores ($LOD_{flowering\ time}$: 6.05, $LOD_{plant\ height}$: 6.02). We also identified a significant QTL in the F_3 population on chromosome 1 (position: 2.441.841, LOD: 6.64; **Figure S10C**) above the permutation based threshold (LOD: 6.01).

We used the giant and genebank populations for association mapping using FarmCPU and included a Q matrix that was calculated from first three principal components to account for population structure. For the giant population, we used an imputed dataset of 558.428 SNP markers, and we detected six and nine significant marker associations below the Bonferroni significance threshold for the traits flowering time and plant height, respectively. Of note, the same SNP polymorphism on chromosome 10 (Position 15.053.501; **Figure 10B** and **Figure S11A**) is the most significant marker for flowering time and the second most significant

marker for plant height, similar to the marker in the F_2 population of the hypochondriacus family.

In the genebank population, we detected 11, 8, and 13 marker-trait associations below the Bonferroni significance threshold for the traits flowering time, plant height and seed setting, respectively. Interestingly, the second most significant marker found for seed setting located very closely to the most significant marker found in (i) QTL analysis for flowering time in the F_2 / F_3 populations and (ii) the GWAS for flowering time in the giant population. We therefore consider the genomic region as 'consensus' region among three mapping population for the control of photoperiod sensitivity.

Furthermore, we investigated the phenotypic distribution among the marker genotypes in the found QTLs of the ‘consensus region’ in all three populations, and also inferred the mode of inheritance of photoperiod sensitivity in the hypochondriacus family (**Figure 81**). In all three populations, we detected significant differences between the marker genotypes (alpha = 0.05). The flowering time distribution of the F_2 generation hypochondriacus family (with and without marker imputation) showed the dominance of photoperiod sensitivity over insensitivity at this QTL locus (**Figure 11A-B**).

4.4. Discussion

4.4.1. Genetic architecture of photoperiod sensitivity

Our analysis of three different mapping populations revealed the genetic basis of photoperiod sensitivity and flowering time control in grain amaranths and their wild relatives. The hypochondriacus and giant families showed a similar phenotypic distribution in both flowering time and plant height. The bimodal-like distributions in both populations for both traits suggest an oligogenic control of photoperiod sensitivity. In contrast, phenotypic trait values for the two traits followed a normal distribution in the genebank population, suggesting a more polygenic control. The contrasting photoperiodic response of the parents in the

hypochondriacus family is the most likely explanation of the bi-modal distribution, where only a few loci seem to segregate. Although we do not know which parents were involved in the spontaneous outcrossing leading to the giant population, a similar bi-modal like trait distributions suggest that parents of the giant plants selected for the analysis differed in their flowering time and/or photoperiod sensitivity. However, when compared to the hypochondriacus family, the giant's parents may have been more similar to each other due to the smaller segregation variances observed in the trait distributions of the giant population.

To dissect the genetic basis of photoperiod sensitivity in grain amaranth species, we performed linkage and association mapping. We found strong evidence for oligogenic inheritance of flowering time that includes the same major QTL controlling photoperiod sensitivity in all three populations in addition to additional, minor QTL. The hypochondriacus and giant populations found the same major QTL in flowering time due to the bi-modal distribution of photoperiod sensitive and insensitive segregants. In contrast, there was a higher number of (11) significant marker-trait associations for flowering time in the GWAS of the genebank population. Major reasons for a larger number of signals and a more complex flowering time architecture in the genebank population are likely allelic heterogeneity and the independent evolution of flowering architecture given that the population is highly diverse and includes multiple taxonomic units and a wide geographical distribution range. The genebank population did not only comprise photoperiod sensitive and insensitive individuals but accommodated a larger phenotypic diversity thanks to a larger number of historical recombination events leading to a higher allelic richness and haplotypes. Lin *et al*. (2022) reported 12 and seven significant associations for flowering time using two different diversity panels including *A. tricolor* and a grain amaranth complex, but our study did not find any of those associations. The discrepancy may be explained with differences in the allelic composition of mapping populations and environmental variation because the field trial of Lin *et al*. (2022) was conducted in Shanhua, Taiwan (~ 23° 7’ N, 120°, 17’ E) which is characterized by short days whereas our field trial was conducted under long-day conditions. Additionaly, phenotyping amaranth photoperiod sensitivity in short days likely limits the detection of genomic regions controlling the photoperiodic pathway in association studies.

Unexpectedly, the GWAS of the genebank population identified the same consensus region for seed setting but not for flowering time as with the other two linkage mapping populations. Our previous study (Baturaygil and Schmid, 2022) using a larger assembly of the genebank accessions including all accessions used in this study suggested that the most critically photoperiod insensitivity attitude, but to a lesser extent, a warm center of origin – if an accession is mildly photoperiod sensitive – permits seed setting. Accordingly, it appears that the association mapping for seed setting managed to find the 'consensus region' because the seed setting trait split the photoperiod sensitive accessions from the insensitive ones more successful than the flowering time trait. However, the unbalanced distribution of the allelic genotypes (*TT*:138, *TC*:5, *CC*:4) in the marker with the strongest signal (Chr:2, pos:35.617.013) – but still retained according to the 3% marker allele frequency filtering – does not allow a reliable statistical test. Therefore, the 'consensus region' still could be considered as a very strong candidate for seed setting, and the strongest QTL for seed setting should be validated in the future studies.

We investigated the candidate genes in the 'consensus region' based on the genome annotation of *A. hypochondriacus v*2.1 with the strongest SNP markers detected in each mapping population. Even though the association mapping of the merged hypochondriacus dataset (F_2 and F_3 populations) for flowering time indicates a very large region (~4 mb), a single gene (response regulator of two-component system) within the 'consensus region' consistently located very closed to the most significant SNP marker of each mapping population. In particular, the most significant SNP found in (i) the merged hypochondriacus dataset for flowering time located ~3 kb downstream, (ii) the giant population for flowering time located ~31 kb upstream to the gene, and finally (iii) the genebank population for seed setting located directly in the gene region. Pseudo-Response Regulator gene family (PRRs) members are the pivotal components of circadian clock systems (Farré and Liu, 2013), and play important roles in the regulation of photoperiodic response. PRR genes have been exposed to selection during plant breeding efforts in a variety of different crops (Hotta, 2022). Homologs of the response regulator of two-component system gene were reported in many plant species including *Arabidopsis thaliana* (Para *et al.*, 2007), rice (Murakami *et al.*, 2005), sorghum (Murphy *et al.*, 2011), maize (Wang *et al.*, 2011), barley (Turner *et al.*, 2005), and the close relative beet

(Omolade *et al.*, 2016). In our study, we detected the same ‘consensus region’ in three different genetic backgrounds, which represents solid evidence about the genetic architecture of the photoperiod sensitivity trait in grain amaranths. However, a validation for gene identification is still required. To this end, recombinant inbred lines (RILs) produced from different parental combinations can be used to increase the mapping resolution, that allows to narrow down the size of the linkage blocks accommodating the underlying region. Accordingly, the promising candidate genes can be validated using with the molecular approaches such as gene expression and gene editing (Alqudah *et al.*, 2020).

4.4.2. Pleiotropic effect of photoperiod sensitivity on other traits

Pleiotropy is a genetic phenomenon describing the control of more than one trait by a single gene, which establishes genetic correlation among these traits. It is considered as an evolutionary limitation because it obstructs the individual selection of the correlated traits (Auge *et al.*, 2019). We studied the pleiotropic relationships between the traits through phenotypic investigation and their mapping. In all three populations, later flowering was correlated with increased plant height. This effect was particularly prominent in the F_2 generation of the hypochondriacus family, which showed a bimodal distribution of flowering time and a corresponding distribution of early flowering plants with a small height and late flowering plants that were very large. The bimodal-like distributions in the hypochondriacus and giant populations for flowering time and plant height suggest that the same locus (or loci) are responsible for photoperiod sensitivity control, also controls a remarkable variation for flowering time and plant height. Consequently, our linkage and association mapping studies found the same genomic region – consensus region – for (i) flowering time of the merged dataset of the hypochondriacus family, (ii) flowering time and plant height of the giant population, and also for (iii) seed setting in the genebank population. Furthermore, our linkage mapping studies for both flowering time and plant height in the F_2 generation hypochondriacus family also found the same marker, which is in the same large QTL region found in the merged dataset of the hypochondriacus family (**Figure 10A**), that located ~1.4 Mb downstream to the ‘consensus region’. Hence, our results clearly demonstrate that the genomic

region controlling photoperiod sensitivity also controls flowering time, plant height and seed setting, and confirms the pleiotropic effect of photoperiod sensitivity on flowering time, plant height and seed setting.

4.4.3. Hypothetical *A* and *B* loci to explain the observed flowering time variation

Chang *et al*. (1969) separated the vegetative growth period of short-day plants such as rice into two sub-periods. Basic vegetation period (BVP) is the duration of the vegetative period under the ideal day-length to which the plant is adapted (short-day). In contrast, photoperiod sensitive period (PSP) is a deviation that delays the completion of the vegetative period due to imperfect light duration (long-day). Hence, the BVP can be measured as the flowering time under the ideal short day-length, whereas flowering time under long-day conditions gives the sum of BVP and PSP.

We recorded the flowering times of the parent genotypes of the hypochondriacus family under short (10 hours, in speed breeding chamber) and long-day conditions (approximately 14 hours, under field conditions, **Figure 8I**). Under short-day conditions, male parent (PI 649623) flowered more than 39 days earlier than female parent (Ames 5149) – we had to end the experiment in speed breeding chambers before female parent flowered but at the day 69, female parent emerged inflorescence, which signified the upcoming flowering (Jähne *et al.*, 2020) – and we attributed this variation to BVP. In addition, we considered the female parent as photoperiod insensitive as it showed a very little variation under different light durations. In contrast, the male parent showed a very large difference of 110 days under two different light durations, and we considered the male parent as photoperiod sensitive. Accordingly, we attributed this variation of the parents in photoperiod sensitivity to PSP. Subsequently, we attributed the variation of PSP and BVP to hypothetical '*A*' and '*B*' loci in the F_2 population, respectively.

4.4.4. Mode of inheritance of photoperiod sensitivity

A monogenic inheritance of photoperiod sensitivity being dominant over insensitivity with the expected 3:1 segregation ratio was reported in rice (Chang *et al.*, 1969; Poonyarit *et al.*, 1989). A similar single photoperiod sensitivity allele was reported in amaranth being dominant over insensitivity, and the level of sensitivity is suggested to be quantitatively regulated by another locus (Kulakow and Jain, 1985). Accordingly, we compared the phenotypic segregation ratios in flowering time to detect if a similar monogenic pattern exists in our F_2 population. A chi-square (X^2) goodness of fit test revealed that the observed distribution of the photoperiod sensitive and insensitive phenotypes was significantly different than that of the expected single gene model with 3:1 ratio (X-squared: 15.949, p=6.50e-05). The fixation of photoperiod insensitivity in the F_3 generation after a single selection suggests a recessive genetic effect of this phenotype and a dominant effect of photoperiod sensitivity. Thus, we hypothesize that the female parent may have photoperiod insensitivity alleles in *A* locus (*aa*) but possess lateness alleles in *B* locus, as the female parent flowered later under short-day conditions, where photoperiod sensitivity cannot manifest itself. Similarly, the male parent may possess photoperiod sensitivity alleles (*AA*) in *A* locus, but may have earliness alleles in *B* locus, as it flowered earlier than the female parent, under short-day conditions. The 'dwarf' phenotypes we observed may then have inherited a combination of earliness alleles of both loci i.e., photoperiod insensitivity alleles in the *A* locus and the earliness alleles in *B* the locus, which is consistent with an earlier flowering of the 'dwarf' phenotypes compared to their early parent. Likewise, genotypes with very late or no flowering may have inherited a combination of alleles for late flowering in both loci, consistent with delayed flowering compared to their late flowering parent.

We attributed 190 days as the flowering time to the individuals that did not flower in the mapping studies and note that this arbitrarily given flowering time may complicate the quantification of the dominance level i.e., later flowering times may suggest over-dominance, whereas earlier flowering times may propose incomplete dominance.

4.4.5. Epistatic relationships

As the number of photoperiod insensitive individuals were higher than the sensitive ones compared to the single gene model, we tested the duplicative recessive epistasis model in the F_2 population. This model is a sort of complementary gene action that entails an epistatic relationship among two loci, and the observed distribution in our F_2 population fit to the expected 9:7 segregation ratio between photoperiod sensitive and insensitive phenotypes as a result of the chi-square (X^2) goodness of fit test (X-squared: 1.3423, p=0.2466). According to the duplicative recessive epistasis model, expression of the alleles in one locus is masked by the recessive alleles from the other locus (Malaviya *et al.*, 2019).

We hypothesized that the epistatic interaction of the '*A*' locus and an unknown '*X*' locus – that might be the '*B*' locus or an independent '*C*' locus that is not responsible for the variation in BVP – might have led the observed 9:7 ratio between the photoperiod sensitive and insensitive phenotypes. According to this hypothesis, *A_X_* genotypes lead photoperiod sensitivity, whereas *aa* _ _ and _ _ *xx* genotypes lead photoperiod insensitivity. The 9:7 segregation ratio between the photoperiod sensitive and insensitive phenotypes was reported in different rice cross combinations between the parents with contrasting photoperiodic behaviours (Chang *et al.*, 1969; Poonyarit *et al.*, 1989). Chang *et al*. (1969) and Nwe and Mackill (1986) suggested that photoperiod insensitive parents with short BVP appear to possess recessive allele(s), which reconstitutes the segregation ratio of photoperiod sensitive and insensitive plants in favour of the insensitive phenotype by inhibiting the expression of photoperiod sensitivity. The authors emphasized that the locus responsible for BVP seems to be involved in that epistatic relationship with PSP locus, however, an alternative relationship between the locus responsible for PSP and a recessive allele from an independent locus may also explain the inhibition of photoperiod sensitivity, according to which we formulated our hypothesis. In addition, Uwatoko *et al*. (2008) reported numerous epistatic relationships between all possible combinations of three major loci responsible for photoperiodic response and vegetative growth using near isogenic lines in rice and considered such epistatic relationships as an adaptive mechanism that diversify the photoperiodic variation by adjusting CONSTANS (CO)

transcription factor, which is the central photoperiodic response regulator in *Arabidopsis thaliana*.

We encountered several major obstacles in the testing of this hypothesis with marker data. Firstly, we needed to impute the significant marker due to the high proportion of missing data (~71%, 44 of 153 individuals). However, imputation changed the allelic distribution of the significant marker from *AA*: 28, *AB*: 13, *BB*: 3 to *AA*: 65, *AB*: 71, *BB* :17, which deviates from the expected 1 : 2 : 1 ratio and we also observed a strong segregation distortion in the marker. Subsequently, the genetic data did not support the duplicative recessive epistasis model as some of the individuals that possess maternal insensitivity alleles (*aa*) showed photoperiod sensitivity (**Figure 11A**). Another obstacle is the absence of a hypothetical '*X* locus', which would have assisted us to test the potential epistatic relationship between the loci. Nonetheless, the numerous individuals possessing *aa* alleles showed photoperiod sensitivity indicates an epistasis effect (**Figure 11A)**. In contrast to the late flowering individuals in the F_2 generation with the same *aa* alleles, all individuals from the F_3 generation – all flowered early – only possessed *aa* alleles, when we used non-imputed marker data. These findings suggest that photoperiod insensitivity is not only controlled by the *A* locus, but alleles from another locus (or loci, i.e., *X* locus) that is epistatic to the *A* locus prevents the expression of photoperiod insensitivity alleles in some cases.

4.6. Conclusion

Our study showed that an oligogenic architecture controls the photoperiod sensitivity trait in grain amaranths. Our phenotypic and molecular results suggest that pleiotropic relationships exist among flowering time, seed setting and plant height traits, and photoperiod sensitivity is the driving factor behind these relationships, as evident in some other short-day plants such as sorghum and rice. We attributed the phenotypic flowering time variation observed in the hypochondriacus and giant populations to the hypothetical '*A*' and '*B*' loci, which is similar to the explanation of Kulakow and Jain (1985). Accordingly, we mapped the same – consensus region – consistently in all our three populations, that successfully separated the photoperiod

sensitive individuals from the insensitive ones and reported the ‘response regulator of two-component system’ candidate gene as a promising ‘*A*’ locus. In addition, we showed that photoperiod sensitivity is dominant to insensitivity in the reported ‘consensus region’, however, a more precise quantification of the degree of the dominance effect requires a more complete phenotypic and marker data. Finally, the duplicative recessive epistasis model appears to be a potential explanation of the observed 9:7 segregation ratio between the photoperiod sensitive and insensitive phenotypes in the F_2 family. In addition, marker data also confirms an epistasis pattern, however, does not conform with the particular duplicative recessive epistasis model. Our mapping efforts did not find a potential ‘*B*’ locus, as our phenotypic data suggests. Therefore, particular epistasis models and the allelic background of the ‘dwarf’ phenotype can be investigated only after the hypothetical ‘*B*’ locus is also found. To date, the genomic basis of photoperiod sensitivity in grain amaranths has not been investigated in the molecular level. Our results provide useful information in the utilization of this important trait in the breeding programs and will serve to understand the local adaptation process of the grain amaranths.

Author Contributions

A.B. and K.S. designed the experiment. A.B. collected and analyzed the data and wrote the first version of the manuscript. M.V.-V. processed the raw sequences and estimated the initial SNP matrix. All authors edited, read and approved the final manuscript.

Acknowledgements

We thank Elisabeth Kokai-Kota and Vanessa Haseneder for the assistance in DNA sequencing and staff of the Hohenheim University for the support with the field and greenhouse experiments. This work was supported by the F.W. Schnell endowed Professorship of the Stifterverband and the Hohenheim-Tübingen Regional Alliance of the Ministry of Science and Culture (MWK) of Baden-Württemberg.

4.7. References

Alqudah AM, Sallam A, Stephen Baenziger P and Börner A (2020) GWAS: Fast-forwarding gene identification and characterization in temperate Cereals: lessons from Barley – A review. *Journal of Advanced Research* 22: 119–135.

Assad R, Reshi ZA, Jan S and Rashid I (2017) Biology of Amaranths. *Botanical Review* 83: 382–436.

Auge GA, Penfield S and Donohue K (2019) Pleiotropy in developmental regulation by flowering-pathway genes: is it an evolutionary constraint? *New Phytologist* 224: 55–70.

Balodis O, Bartuševics J and Gaile Z (2011) Biomass yield of different plants for biogas production. *Vide, Tehnologija, Resursi - Environment, Technology, Resources* 1: 238–245.

Baturaygil A and Schmid K (2022) Characterization of Flowering Time in Genebank Accessions of Grain Amaranths and Their Wild Relatives Reveals Signatures of Domestication and Local Adaptation. *Agronomy* 12:505.

Baturaygil A, Stetter MG and Schmid K (2021) Breeding Amaranth for Biomass: Evaluating Dry Matter Content and Biomass Potential in Early and Late Maturing Genotypes. *Agronomy* 11(5): 920.

Becker R *et al*. (1981) A Compositional Study of Amaranth Grain. *Journal of Food Science* 46: 1175–1180.

Bernardo R (2008) Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science* 48: 1649–1664.

Brenner DM *et al*. (2000) Genetic resources and breeding of Amaranthus. *Plant Breeding Reviews* 19: 227–285.

Broman KW, Wu H, Sen Ś and Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19: 889–890.

Buckler ES *et al*. (2009) The genetic architecture of maize flowering time. *Science* 325: 714–718.

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. *Euphytica* 77: 205–219.

Chang CC *et al*. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 4: 1–16.

Chang TT, Li CC and Vergara BS (1969) Component analysis of duration from seeding to heading in rice by the basic vegetative phase and the photoperiod-sensitive phase. *Euphytica* 18: 79–91.

Clouse JW *et al*. (2016) The Amaranth Genome: Genome, Transcriptome, and Physical Map Assembly. *The Plant Genome* 9.

von Cossel M, Möhring J, Kiesel A and Lewandowski I (2017) Methane yield performance of amaranth (*Amaranthus hypochondriacus* L.) and its suitability for legume intercropping in comparison to maize (*Zea mays* L.). *Industrial Crops and Products* 103: 107–121.

Costea M and Tardif FJ (2004) The biology of Canadian weeds. 130. *Amaranthus retroflexus* L., *A. powellii* S. Watson and *A. hybridus* L. (update). *Canadian Journal of Plant Science* 84: 631–668.

Farré EM and Liu T (2013) The PRR family of transcriptional regulators reflects the complexity and evolution of plant circadian clocks. *Current Opinion in Plant Biology* 16: 621–629.

Fujino K (2003) Photoperiod sensitivity gene controlling heading date in rice cultivars in the northernmost region of Japan. *Euphytica* 131: 97–103.

García-Alcalde F *et al*. (2012) Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* 28: 2678–2679.

Grieder C, Dhillon BS, Schipprack W and Melchinger AE (2012) Breeding maize as biogas substrate in Central Europe: II. Quantitative-genetic parameters for inbred lines and correlations with testcross performance. *Theoretical and Applied Genetics* 124: 981–988.

Guerra-Matias AC and Arêas JAG (2005) Glycemic and insulinemic responses in women consuming extruded amaranth (Amaranthus cruentus L). *Nutrition Research* 25: 815–822.

Hauptli H and Jain S (1985) Genetic variation in outcrossing rate and correlated floral traits in a population of grain amaranth (*Amaranthus cruentus* L.). *Genetica* 66: 21–27.

Haussmann BIG *et al*. (2012) Breeding Strategies for Adaptation of Pearl Millet and Sorghum to Climate Variability and Change in West Africa.

Hotta CT (2022) The evolution and function of the *PSEUDO RESPONSE REGULATOR* gene family in the plant circadian clock. *Genetics and Molecular Biology* 45: 4–5.

Hung HY *et al*. (2012) *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proceedings of the National Academy of Sciences of the United States of America* 109.

Jähne F, Hahn V, Würschum T and Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theoretical and Applied Genetics*.

Jain SK, Hauptli H and Valdya KR (1982) Outcrossing rate in grain amaranths. *Journal of Heredity* 73: 71–72.

Joshi DC *et al*. (2018) From zero to hero: the past, present and future of grain amaranth breeding. *Theoretical and Applied Genetics* 131: 1–17.

Jung C and Müller AE (2009) Flowering time control and applications in plant breeding. *Trends in Plant Science* 14: 563–573.

Kulakow PA and Jain SK (1985) The inheritance of flowering time in *Amaranthus* species. *Journal of Genetics* 64: 85–100.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Lightfoot DJ *et al*. (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology* 15: 1–15.

Lin YP, Wu TH, Chan YK, van Zonneveld M and Schafleitner R (2022) *De novo* SNP calling reveals the genetic differentiation and morphological divergence in genus *Amaranthus*. *Plant Genome* 15: 1–19.

Lipka AE *et al*. (2012) Genetics and population analysis GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28: 2397–2399.

Liu F and Stützel H (2004) Biomass partitioning, specific leaf area, and water use efficiency of vegetable amaranth (*Amaranthus* spp.) in response to drought stress. *Scientia Horticulturae* 102: 15–27.

Liu T, Liu H, Zhang H and Xing Y (2013) Validation and characterization of Ghd7.1, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (*Oryza sativa* L.). *Journal of Integrative Plant Biology* 55: 917–927.

Liu X, Huang M, Fan B, Buckler ES and Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genetics* 12.

Ma X *et al*. (2021) A chromosome-level *Amaranthus* cruentus genome assembly highlights gene family evolution and biosynthetic gene clusters that may underpin the nutritional value of this traditional crop. *Plant Journal* 107: 613–628.

Malaviya DR, Roy AK, Kaushal P, Yadav A and Pandey DK (2019) Complementary gene interaction and xenia effect controls the seed coat colour in interspecific cross between *Trifolium alexandrinum* and *T. apertum*. *Genetica* 147: 197–203.

Murakami M, Matsushika A, Ashikari M, Yamashino T and Mizuno T (2005) Circadian-associated rice pseudo response regulators (*OsPRRs*): Insight into the control of flowering time. *Bioscience, Biotechnology and Biochemistry* 69: 410–414.

Murphy RL *et al*. (2011) Coincident light and clock regulation of pseudoresponse regulator protein 37 (*PRR37*) controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 108: 16469–16474.

Murphy RL *et al*. (2014) *Ghd7* (*Ma 6*) Represses Sorghum Flowering in Long Days: *Ghd7* Alleles Enhance Biomass Accumulation and Grain Production. *The Plant Genome* 7: plantgenome2013.11.0040.

Mursec B, Vindis P, Janzekovic M, Brus M and Cus F (2009) Analysis of different substrates for processing into biogas Manufacturing and processing. *Journal of Achievements in Materials and Manufacturing Engineering* 37: 652–659.

Myers RL (1996) Amaranth: New crop opportunity. *Progress in new crops* 207–220.

Nakamichi N (2014) Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant and Cell Physiology* 56: 594–604.

National Research Council (1984) *Amaranth Modern Prospects for an Ancient Crop*. National Academy Press.

Nwe KT and Mackill DJ (1986) Photoperiod sensitivity in three rice crosses. *Euphytica* 35: 751–760.

Ogiso-Tanaka E, Yabe S and Tanaka T (2020) IonBreeders: bioinformatics plugins toward genomics-assisted breeding. *Breeding Science Preview* 70.

Okonechnikov K, Conesa A and García-Alcalde F (2015) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32: btv566.

Olson SN *et al*. (2012) High biomass yield energy sorghum: Developing a genetic model for C4 grass bioenergy crops. *Biofuels Bioproducts Bioreining* 6: 640–655.

Omolade O, Müller AE, Jung C and Melzer S (2016) *BvPRR7* is a cold responsive gene with a clock function in beet. *Biologia Plantarum* 60: 95–104.

Para A *et al*. (2007) *PRR3* is a vascular regulator of *TOC1* stability in the *Arabidopsis* circadian clock. *Plant Cell* 19: 3462–3473.

Poonyarit M, Mackill DJ and Vergara BS (1989) Genetics of photoperiod sensitivity and critical daylength in rice. *Crop Science* 29: 647–652.

Pospišil A (2009) Yield and Quality of Forage Sorghum and Different Amaranth Species (*Amaranthus* spp .) Biomass. *Agriculturae Conspectus Scientificus* 74: 85–89.

Sauer JD (1950) The Grain Amaranths: A Survey of Their History and Classification. *Annals of the Missouri Botanical Garden* 37: 561–632.

Sauer JD (1967) The grain amaranths and their relatives: A revised taxonomic and geo-graphic survey. *Ann. Missouri Bot. Gard.* 54: 102-137.

Seppälä M, Laine A and Rintala J (2013) Screening of novel plants for biogas production in northern conditions. *Bioresource Technology* 139: 355–362.

Sitkey V, Gaduš J, Kliský Ľ and Dudák A (2013) Biogas Production from Amaranth Biomass. *Acta Regionalia et Environmentalica* 10: 59–62.

Stallknecht GF and and Schulz-Schaeffer JR (1993) Amaranth Rediscovered. *In: J. Janick and J.E. Simon (eds.)* 211–218.

Stetter MG and Schmid KJ (2017) Analysis of phylogenetic relationships and genome size evolution of the Amaranthus genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution* 109: 80–92.

Stetter MG, Vidal-Villarejo M and Schmid KJ (2020) Parallel Seed Color Adaptation during Multiple Domestication Attempts of an Ancient New World Grain. *Molecular biology and evolution* 37: 1407–1419.

Svirskis A (2009) Prospects for non-traditional plant species cultivated for forage in Lithuania. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 37: 215–218.

Turner A, Beales J, Faure S, Dunford RP and Laurie DA (2005) The pseudo-response regulator *Ppd-H1* provides adaptation to photoperiod in barley. *Science* 310: 1031–1034.

Uwatoko N *et al*. (2008) Epistasis among the three major flowering time genes in rice: Coordinate changes of photoperiod sensitivity, basic vegetative growth and optimum photoperiod. *Euphytica* 163: 167–175.

Wang P, Gong R, Yang Y and Yu S (2019) *Ghd8* controls rice photoperiod sensitivity by forming a complex that interacts with *Ghd7*. *BMC Plant Biology* 19: 1–14.

Wang X *et al*. (2011) Robust expression and association of *ZmCCA1* with circadian rhythms in maize. *Plant Cell Reports* 30: 1261–1272.

Windpassinger S, Friedt W, Frauen M, Snowdon R and Wittkop B (2015) Designing adapted sorghum silage types with an enhanced energy density for biogas generation in temperate Europe. *Biomass and Bioenergy*.

Wu H *et al*. (2000) Field evaluation of an *Amaranthus* genetic resource collection in China. *Genetic Resources and Crop Evolution* 47: 43–53.

Yan W *et al*. (2013) Natural variation in *Ghd7.1* plays an important role in grain yield and adaptation in rice. *Cell Research* 23: 969–971.

Yan WH *et al*. (2011) A major QTL, *Ghd8*, plays pleiotropic roles in regulating grain productivity, plant height, and heading date in rice. *Molecular Plant* 4: 319–330.

Zhang C, Dong SS, Xu JY, He WM and Yang TL (2019) PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35: 1786–1788.

Zheng X *et al*. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28: 3326–3328.

TABLES

Table 2. Population size, generation, and parental information of the mapping populations

Type	Population	Generation	Number of Individuals	Female Parent	Male Parent
Targeted-cross derived (*n*=267)					
	hypochondriacus	F_2	154	Ames 5149	PI 649623
	hypochondriacus	F_3	113	Ames 5149	PI 649623
Spontaneous-crossing derived (Giants, *n*=166)					
	Baernkrafft 11-1 (F1)	F_2	15	Baernkrafft	unknown
	Baernkrafft 11-2 (F1)	F_2	17	Baernkrafft	unknown
	C6-1-1 (F1)	F_2	8	C6	unknown
	C6-20-1 (F1)	F_2	11	C6	unknown
	C6-20-2 (F1)	F_2	6	C6	unknown
	C6-20-3 (F1)	F_2	12	C6	unknown
	C6-27-1 (F1)	F_2	8	C6	unknown
	C6-27-2 (F1)	F_2	8	C6	unknown
	Puerto Moutt 12-1 (F1)	F_2	4	Puerto Moutt	unknown
	Puerto Moutt 12-2 (F1)	F_2	4	Puerto Moutt	unknown
	Puerto Moutt 12-3 (F1)	F_2	9	Puerto Moutt	unknown
	Puerto Moutt 29-1 (F1)	F_2	7	Puerto Moutt	unknown
	Puerto Moutt 29-2 (F1)	F_2	14	Puerto Moutt	unknown
	Puerto Moutt 3-1 (F1)	F_2	12	Puerto Moutt	unknown
	C6-1-1 (S1)	F_3	11	C6	unknown
	C6-20-2 (S1)	F_3	3	C6	unknown
	Puerto Moutt 12-2-b (S1)	F_3	3	Puerto Moutt	unknown
	Puerto Moutt 12-2-w (S1)	F_3	6	Puerto Moutt	unknown
	Puerto Moutt	Parent	1		
	C6	Parent	1		
	Pastewny	Parent	1		
Genebank population (*n*=171)	Speices/source			number of accessions (*n*)	
	A. caudatus			46	
	A. cruentus			34	
	A. hybridus			19	
	"hybrids"			9	
	A. hypochondriacus			35	
	A. quitensis			20	
	biomass amaranth pool			8	

Table 3. Parental genotypes of the hypochondriacus family

ID	Accession name	Species	Country of Origin	Stem Color	Photoperiod Sensitivity Status*
149	Ames 5149	*A.hypochondriacus*	Puerto Rico	Green	Insensitive
174	PI 649623	*A.hypochondriacus*	Mexico	Red	Sensitive

*Based on flowering times of two genotypes under field (~14 hours) and spreed breeding conditions (10 hours, **Figure 8I**).

Table 4. Information on the field trials

Location	Year	Altitude	Mean Temperature [C]*	Mean day Length	Sowing Date	Harvest Date
HOH	2018	400m a.s.l	18.1	14h 31m	8 May	12 October
OLI	2018	700m a.s.l	15.4	14h 16m	9 May	22 October
HOH	2019	400m a.s.l	17	14h 14m	17 May	23 October

* Two meters above ground surface.

FIGURES

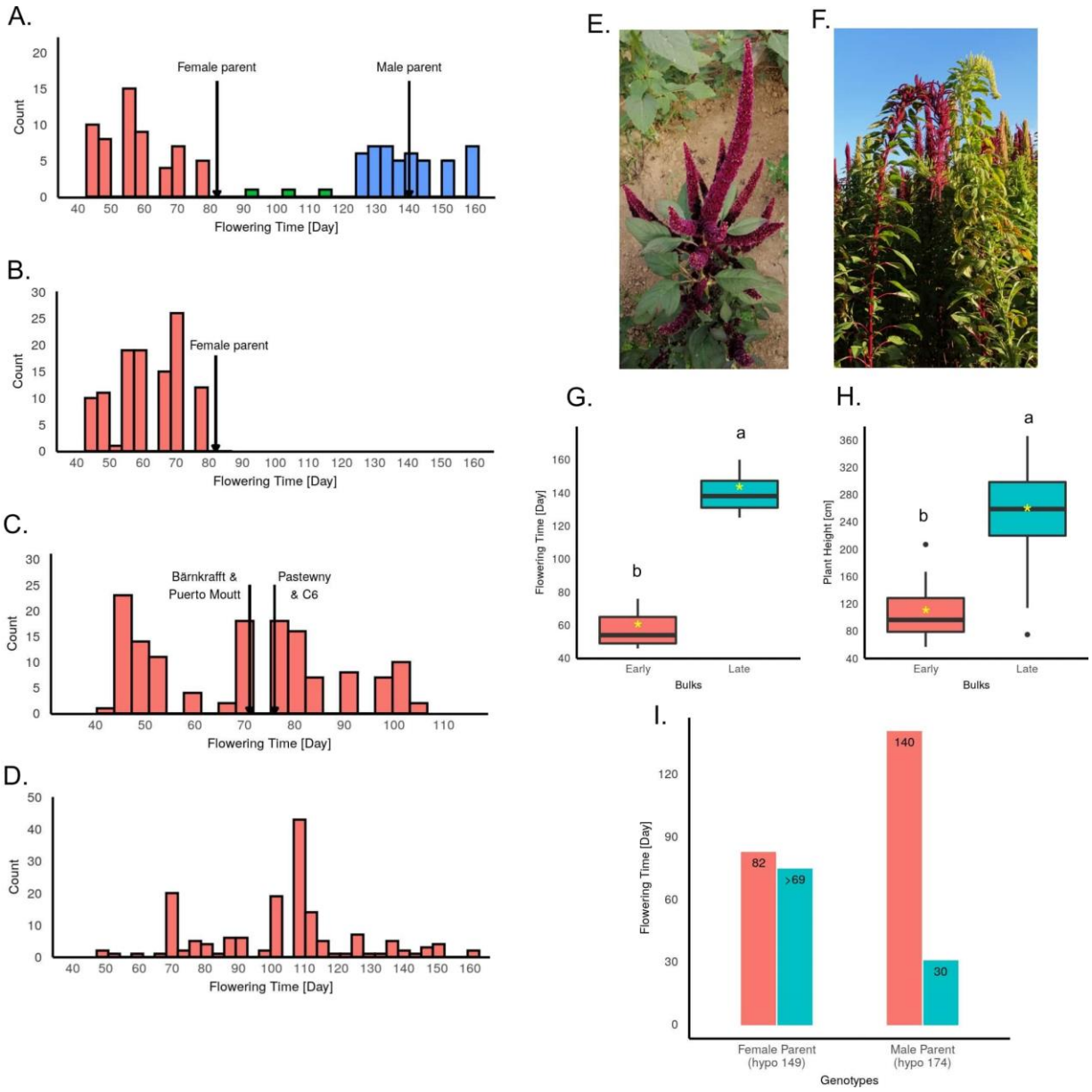

Figure 8. Phenotypic characteristics of the mapping populations. Histograms of flowering time in **(A)** the hypochondriacus F_2 population, where red, blue and green indicate early, late and intermediate bulks, respectively, **(B)** the hypochondriacus F_3 population, **(C)** the 'giant' families, **(D)** the genebank population. Representative photos of **(E)** early – dwarf phenotype – and **(F)** late bulk plants. Boxplots of **(G)** flowering time and **(H)** plant height in the hypochondriacus F_2 population. Yellow asterisks indicate least square means of the respective bulks for a given trait, whereas letters indicate highly significant differences between the bulks in both traits (p<2.2e-16). **(I)** Flowering time bar plots of the parents of the hypochondriacus populations under long-day and short-day conditions. Red indicates long-day (field) and turquoise indicate short-day (speed breeding chamber) conditions.

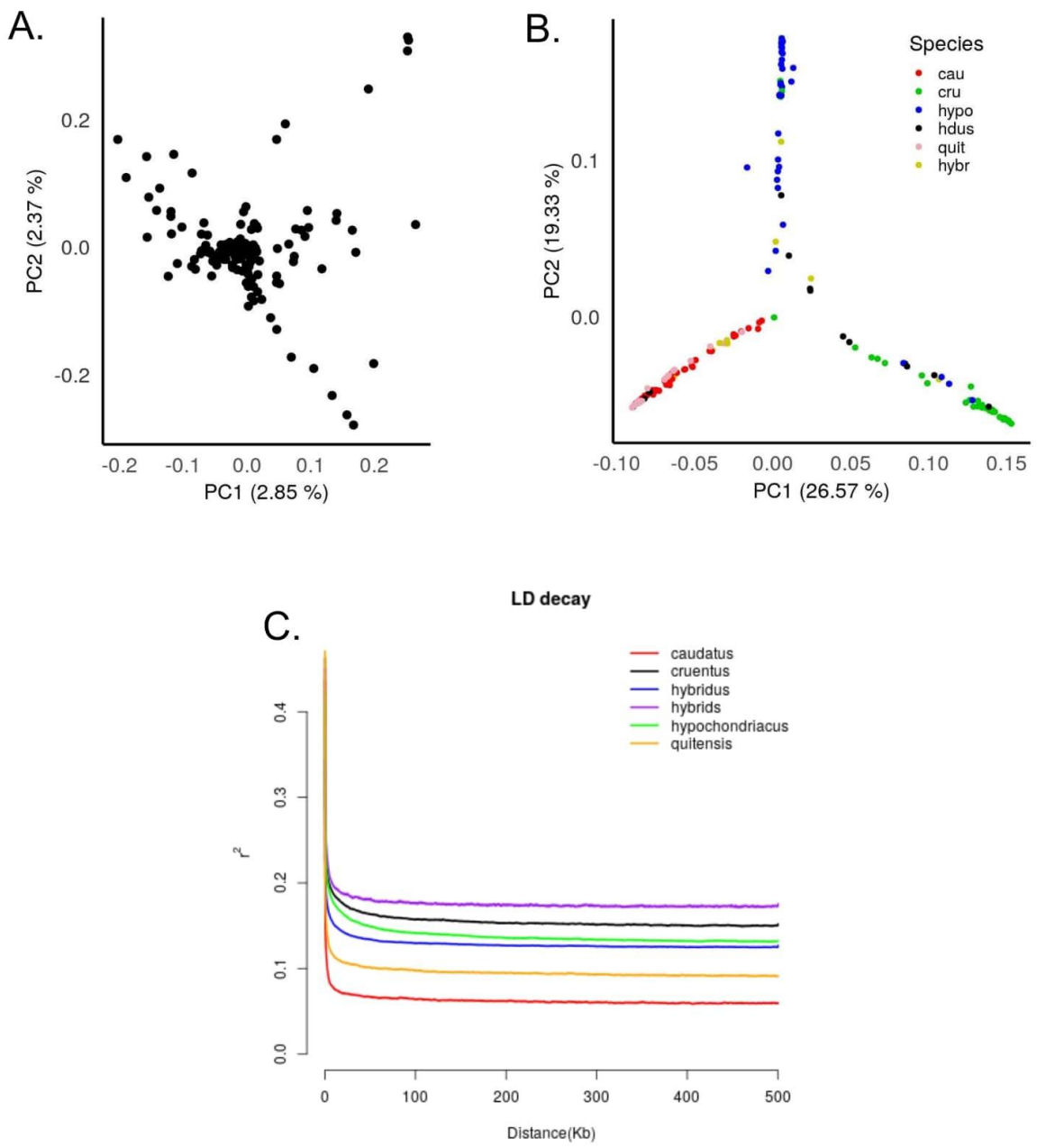

Figure 9. PCA of the **(A)** the offspring from the giant families (*n*=167) and **(B)** the genebank accessions (*n* =163). **(C)** LD decay plot of the three major grain amaranth species and their wild relatives. Sample sizes are; *A. caudatus* (*n*=46), *A. cruentus* (*n*=34), *A. hybridus* (*n*=19), 'hybrids' (*n*=9), *A. hypochondriacus* (*n*=35), *A. quitensis* (*n*=20).

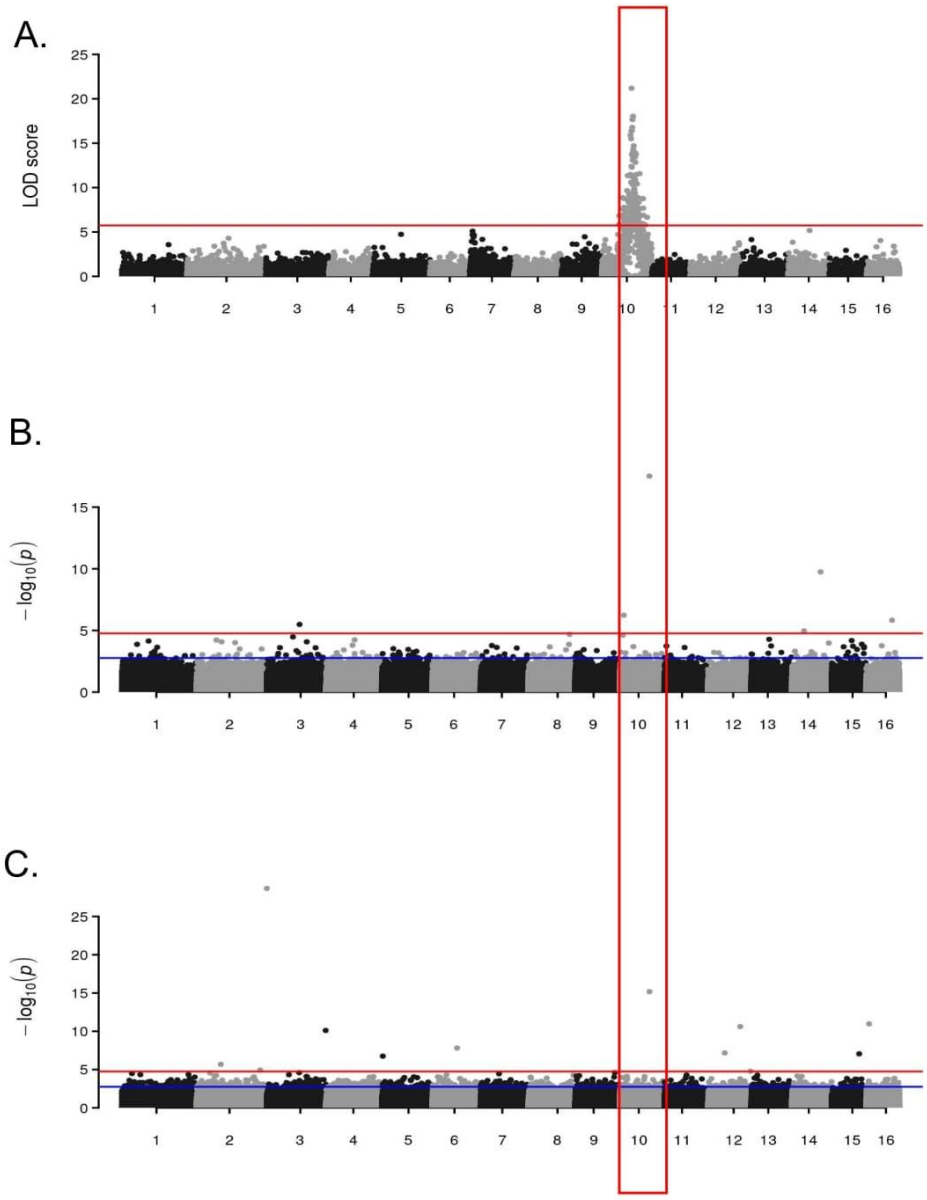

Figure 10. Manhattan plots of **(A)** linkage mapping for flowering time in a merged population of F_2 and F_3 generations of the hypochondriacus family (*n*=265), association mapping of **(B)** the 'giant' population for flowering time (*n*=142), and our (**C**) genebank population (*n*=163) for seed setting. The red line in **(A)** indicates the significance threshold based on permutation test, the blue line in **(B)** and **(C)** indicates the suggestive thresholds (1/ number of markers), and the red line shows the Bonferroni corrected thresholds for p=0.01. The genomic region shown in the red rectangle indicates the 'consensus region' harboring the significant associations across the three populations.

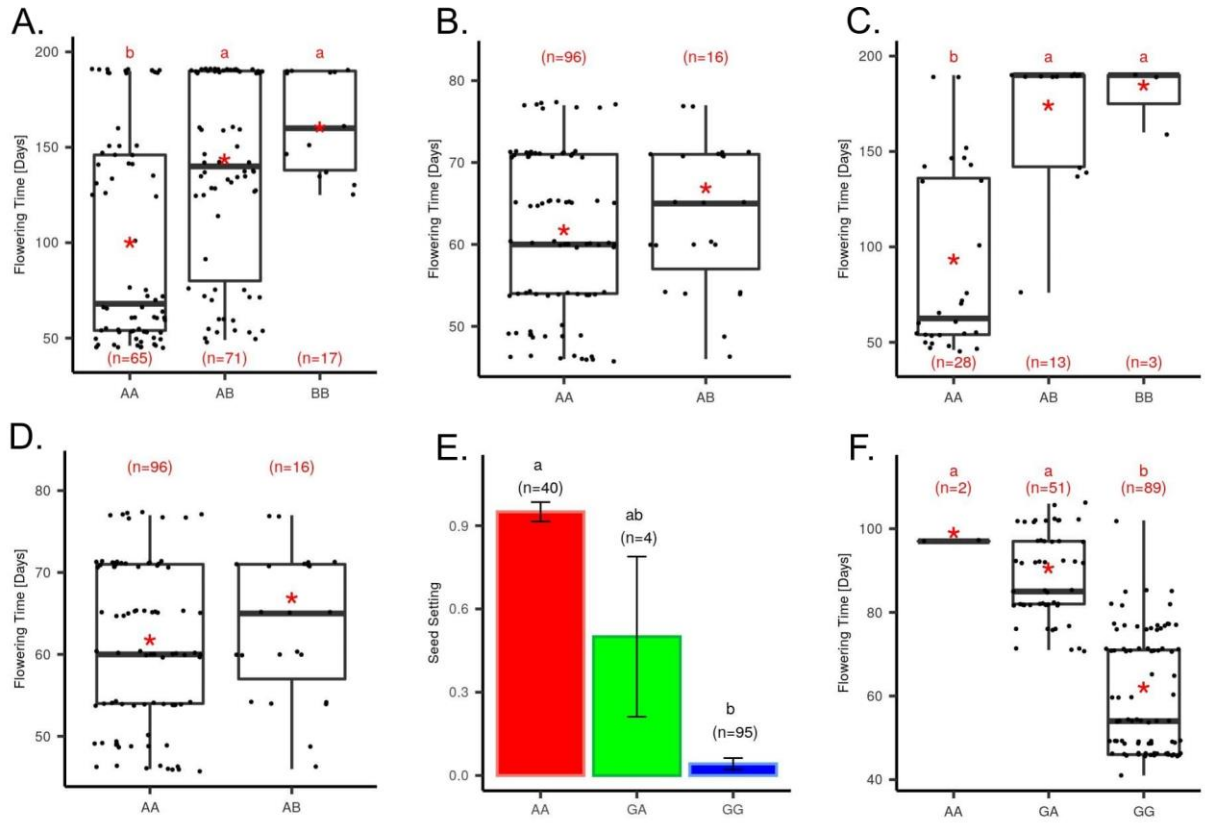

Figure 11. Box plots and the bar plot indicating the phenotpyic distributons of the genotypes from the most significant SNP markers from the 'consensus region'. Box plot showing **(A)** flowering time distribution of the genotypes of the significant trait - marker association (Chr:10, pos: 15.010.436) from the the linkage mapping with the imputed dataset of the F_2 generation hypochondriacus family, **(B)** flowering time distribution of the genotypes of the significant trait - marker association (Chr:10, pos: 15.010.436) from the linkage mapping with the imputed dataset of the F_3 generation hypochondriacus family, **(C)** flowering time distribution of the genotypes of the significant trait - marker association (Chr:10, pos: 15.010.436) from from the linkage mapping with the non-imputed dataset of the F_2 generation hypochondriacus family, **(D)** flowering time distribution of the genotypes of the significant trait - marker association (Chr:10, pos: 15.010.436) from the linkage mapping with the non-imputed dataset of the F_3 generation hypochondriacus family. **(E)** Bar plot showing seed setting rates of the genotypes of the significant trait - marker association (Chr:10, pos: 15.018.912) from the association mapping of the genebank population, and **(F)** box plot showing flowering time distribution of the genotypes of the significant trait - marker association (Chr:10, pos: 15.053.501) from the association mapping of the giant families. In figures **A-D**, A and B indicate the maternal and paternal alleles, respectively.

SUPPLEMENTARY TABLES

Table S6. List of the genebank population accessions

No	Species	Country	Accession number	Source
1	*A. caudatus*	Peru	PI 490445	USDA genebank
2	*A. caudatus*	Peru	PI 490552	USDA genebank
3	*A. caudatus*	Peru	PI 481960	USDA genebank
4	*A. caudatus*	Bolivia	PI 669838	USDA genebank
5	*A. caudatus*	Peru	PI 490515	USDA genebank
6	*A. caudatus*	Peru	PI 490450	USDA genebank
7	*A. caudatus*	Peru	PI 649245	USDA genebank
8	*A. caudatus*	Peru	PI 649234	USDA genebank
9	*A. caudatus*	Bolivia	PI 568137	USDA genebank
10	*A. caudatus*	Argentina	Ames 15178	USDA genebank
11	*A. caudatus*	Bolivia	PI 568150	USDA genebank
12	*A. caudatus*	Peru	PI 649220	USDA genebank
13	*A. caudatus*	Bolivia	PI 608018	USDA genebank
14	*A. caudatus*	Peru	PI 490569	USDA genebank
15	*A. caudatus*	Peru	PI 511701	USDA genebank
16	*A. caudatus*	Bolivia	Ames 13860	USDA genebank
17	*A. caudatus*	Peru	Ames 5231	USDA genebank
18	*A. caudatus*	Peru	PI 490447	USDA genebank
19	*A. caudatus*	Peru	PI 649237	USDA genebank
20	*A. caudatus*	Bolivia	PI 568144	USDA genebank
21	*A. caudatus*	Peru	PI 490488	USDA genebank
22	*A. cruentus*	Mexico	PI 477913	USDA genebank
23	*A. cruentus*	Peru	PI 511713	USDA genebank
24	*A. cruentus*	Mexico	PI 477914	USDA genebank
25	*A. cruentus*	USA	PI 606767	USDA genebank
26	*A. cruentus*	Venezuela	PI 665286	USDA genebank
27	*A. cruentus*	Argentina	PI 636182	USDA genebank
28	*A. cruentus*	Mexico	PI 662284	USDA genebank
29	*A. cruentus*	USA	PI 515959	USDA genebank
30	*A. cruentus*	Mexico	Ames 15191	USDA genebank
31	*A. cruentus*	USA	PI 658731	USDA genebank

32	*A. cruentus*	Mexico	PI 643063	USDA genebank
33	*A. cruentus*	Mexico	PI 649609	USDA genebank
34	*A. cruentus*	Mexico	Ames 2240	USDA genebank
35	'hybrid'	USA	PI 568179	USDA genebank
36	'hybrid'	Mexico	PI 604564	USDA genebank
37	'hybrid'	Peru	PI 490430	USDA genebank
38	'hybrid'	Peru	PI 490453	USDA genebank
39	'hybrid'	Peru	PI 490424	USDA genebank
40	'hybrid'	Peru	PI 511752	USDA genebank
41	'hybrid'	Guatemala	Ames 21996	USDA genebank
42	'hybrid'	Peru	PI 511733	USDA genebank
43	*A. hybridus*	Mexico	PI 604574	USDA genebank
44	*A. hybridus*	Guatemala	Ames 22001	USDA genebank
45	*A. hypochondriacus*	Mexico	PI 649633	USDA genebank
46	*A. hypochondriacus*	Mexico	PI 477916	USDA genebank
47	*A. hypochondriacus*	Chile	Ames 5355	USDA genebank
48	*A. hypochondriacus*	USA	PI 568127	USDA genebank
49	*A. hypochondriacus*	Mexico	PI 649532	USDA genebank
50	*A. hypochondriacus*	Puerto Rico	Ames 5149	USDA genebank
51	*A. hypochondriacus*	Mexico	PI 649544	USDA genebank
52	*A. hypochondriacus*	Mexico	PI 619247	USDA genebank
53	*A. hypochondriacus*	Brazil	Ames 5690	USDA genebank
54	*A. hypochondriacus*	Mexico	PI 649602	USDA genebank
55	*A. hypochondriacus*	Mexico	PI 649535	USDA genebank
56	*A. hypochondriacus*	Mexico	PI 643074	USDA genebank
57	*A. quitensis*	Bolivia	PI 568154	USDA genebank
58	*A. quitensis*	Ecuador	PI 490709	USDA genebank
59 *	*A. caudatus*	Peru	PI 649227	USDA genebank
60 *	*A. caudatus*	Peru	PI 511686	USDA genebank
61 *	*A. caudatus*	Bolivia	PI 490459	USDA genebank
62 *	*A. caudatus*	Peru	PI 511687	USDA genebank
63 *	*A. caudatus*	Bolivia	PI 642741	USDA genebank
64 *	*A. caudatus*	Peru	PI 490518	USDA genebank
65 *	*A. caudatus*	Bolivia	PI 490604	USDA genebank

66 *	*A. caudatus*	Peru	PI 649217	USDA genebank
67 *	*A. caudatus*	Peru	PI 490612	USDA genebank
68 *	*A. caudatus*	Peru	PI 511696	USDA genebank
69 *	*A. caudatus*	Argentina	PI 511679	USDA genebank
70 *	*A. caudatus*	Peru	PI 511690	USDA genebank
71 *	*A. caudatus*	Peru	PI 511704	USDA genebank
72 *	*A. caudatus*	Bolivia	PI 511681	USDA genebank
73 *	*A. caudatus*	Peru	PI 649228	USDA genebank
74 *	*A. caudatus*	Peru	PI 490511	USDA genebank
75 *	*A. caudatus*	Peru	PI 481957	USDA genebank
76 *	*A. caudatus*	Ecuador	PI 511712	USDA genebank
77 *	*A. caudatus*	Peru	PI 490431	USDA genebank
78 *	*A. caudatus*	Peru	PI 649230	USDA genebank
79 *	*A. caudatus*	Peru	PI 511706	USDA genebank
80 *	*A. caudatus*	Argentina	PI 511680	USDA genebank
81 *	*A. caudatus*	Argentina	PI 490491	USDA genebank
82 *	*A. caudatus*	Peru	PI 490561	USDA genebank
83 *	*A. caudatus*	Ecuador	PI 608019	USDA genebank
84 *	*A. cruentus*	Mexico	PI 576482	USDA genebank
85 *	*A. cruentus*	Mexico	PI 643037	USDA genebank
86 *	*A. cruentus*	Germany	Baernkrafft	Commercial
87 *	*A. cruentus*	Mexico	PI 511723	USDA genebank
88 *	*A. cruentus*	Mexico	PI 649509	USDA genebank
89 *	*A. cruentus*	Mexico	PI 606798	USDA genebank
90 *	*A. cruentus*	Mexico	PI 649514	USDA genebank
91 *	*A. cruentus*	Brazil	PI 667165	USDA genebank
92 *	*A. cruentus*	Guatemala	PI 451826	USDA genebank
93 *	*A. cruentus*	Mexico	Ames 5552	USDA genebank
94 *	*A. cruentus*	Mexico	PI 511876	USDA genebank
95 *	*A. cruentus*	Mexico	PI 643058	USDA genebank
96 *	*A. cruentus*	Mexico	PI 643039	USDA genebank
97 *	*A. cruentus*	Mexico	PI 649524	USDA genebank
98 *	*A. cruentus*	Guatemala	PI 511717	USDA genebank
99 *	*A. cruentus*	Peru	PI 511714	USDA genebank

100 *	*A. cruentus*	Mexico	PI 658728	USDA genebank
101 *	*A. cruentus*	Mexico	PI 643042	USDA genebank
102 *	*A. cruentus*	Mexico	PI 576481	USDA genebank
103 *	*A. cruentus*	Guatemala	PI 433228	USDA genebank
104 *	*A. cruentus*	Mexico	PI 643049	USDA genebank
105 *	'hybrid'	Mexico	PI 604571	USDA genebank
106 *	*A. hybridus*	Ecuador	PI 490689	USDA genebank
107 *	*A. hybridus*	Peru	Ames 5232	USDA genebank
108 *	*A. hybridus*	Peru	PI 490740	USDA genebank
109 *	*A. hybridus*	Ecuador	PI 490731	USDA genebank
110 *	*A. hybridus*	Ecuador	PI 490739	USDA genebank
111 *	*A. hybridus*	Ecuador	PI 490684	USDA genebank
112 *	*A. hybridus*	Guatemala	PI 667158	USDA genebank
113 *	*A. hybridus*	Ecuador	PI 490664	USDA genebank
114 *	*A. hybridus*	Peru	PI 490489	USDA genebank
115 *	*A. hybridus*	Mexico	PI 511724	USDA genebank
116 *	*A. hybridus*	Ecuador	PI 490679	USDA genebank
117 *	*A. hybridus*	Ecuador	PI 490670	USDA genebank
118 *	*A. hybridus*	Ecuador	PI 667156	USDA genebank
119 *	*A. hybridus*	Ecuador	PI 511754	USDA genebank
120 *	*A. hybridus*	Mexico	PI 604568	USDA genebank
121 *	*A. hybridus*	Bolivia	Ames 5335	USDA genebank
122 *	*A. hybridus*	Mexico	PI 604582	USDA genebank
123 *	*A. hypochondriacus*	Mexico	PI 604595	USDA genebank
124 *	*A. hypochondriacus*	Mexico	PI 649607	USDA genebank
125 *	*A. hypochondriacus*	Mexico	PI 649623	USDA genebank
126 *	*A. hypochondriacus*	Mexico	PI 649565	USDA genebank
127 *	*A. hypochondriacus*	Mexico	PI 643070	USDA genebank
128 *	*A. hypochondriacus*	Mexico	PI 649529	USDA genebank
129 *	*A. hypochondriacus*	Mexico	PI 643041	USDA genebank
130 *	*A. hypochondriacus*	Mexico	PI 649537	USDA genebank
131 *	*A. hypochondriacus*	Mexico	PI 643067	USDA genebank
132 *	*A. hypochondriacus*	Mexico	Ames 2215	USDA genebank
133 *	*A. hypochondriacus*	Mexico	PI 649559	USDA genebank

134 *	*A. hypochondriacus*	Mexico	PI 649575	USDA genebank
135 *	*A. hypochondriacus*	Mexico	PI 649595	USDA genebank
136 *	*A. hypochondriacus*	Mexico	PI 633589	USDA genebank
137 *	*A. hypochondriacus*	Mexico	Ames 5457	USDA genebank
138 *	*A. hypochondriacus*	Mexico	PI 604587	USDA genebank
139 *	*A. hypochondriacus*	Mexico	PI 643036	USDA genebank
140 *	*A. hypochondriacus*	Mexico	Ames 2085	USDA genebank
141 *	*A. hypochondriacus*	Mexico	PI 649602	USDA genebank
142 *	*A. hypochondriacus*	Mexico	PI 604559	USDA genebank
143 *	*A. hypochondriacus*	Mexico	PI 604581	USDA genebank
144 *	*A. hypochondriacus*	Mexico	PI 649587	USDA genebank
145 *	*A. hypochondriacus*	Mexico	PI 511731	USDA genebank
146 *	*A. quitensis*	Ecuador	PI 511747	USDA genebank
147 *	*A. quitensis*	Ecuador	PI 490720	USDA genebank
148 *	*A. quitensis*	Ecuador	PI 490673	USDA genebank
149 *	*A. quitensis*	Peru	PI 490466	USDA genebank
150 *	*A. quitensis*	Ecuador	PI 511737	USDA genebank
151 *	*A. quitensis*	Brazil	PI 652422	USDA genebank
152 *	*A. quitensis*	Ecuador	PI 490705	USDA genebank
153 *	*A. quitensis*	Peru	PI 649246	USDA genebank
154 *	*A. quitensis*	Ecuador	PI 511745	USDA genebank
155 *	*A. quitensis*	Bolivia	PI 511736	USDA genebank
156 *	*A. quitensis*	Peru	Ames 5342	USDA genebank
157 *	*A. quitensis*	Argentina	Ames 21666	USDA genebank
158 *	*A. quitensis*	Peru	PI 511751	USDA genebank
159 *	*A. quitensis*	Brazil	PI 652426	USDA genebank
160 *	*A. quitensis*	Ecuador	PI 511749	USDA genebank
161 *	*A. quitensis*	Argentina	Ames 5334	USDA genebank
162 *	*A. quitensis*	Ecuador	Ames 5247	USDA genebank
163 *	*A. quitensis*	Ecuador	PI 511741	USDA genebank

* Raw sequences of these accessions were taken from Stetter *et al.* 2020.

Table S7. Filtering parameters of the three populations

Population	maf	miss	qual	min_depth	max_depth
The hypochondriacus family	-	0.5	30	0.8	10
The landrace population	0.03	0.6	30	2.5	20
The giant population	-	0.3	30	0.5	10

SUPPLEMENTARY FIGURES

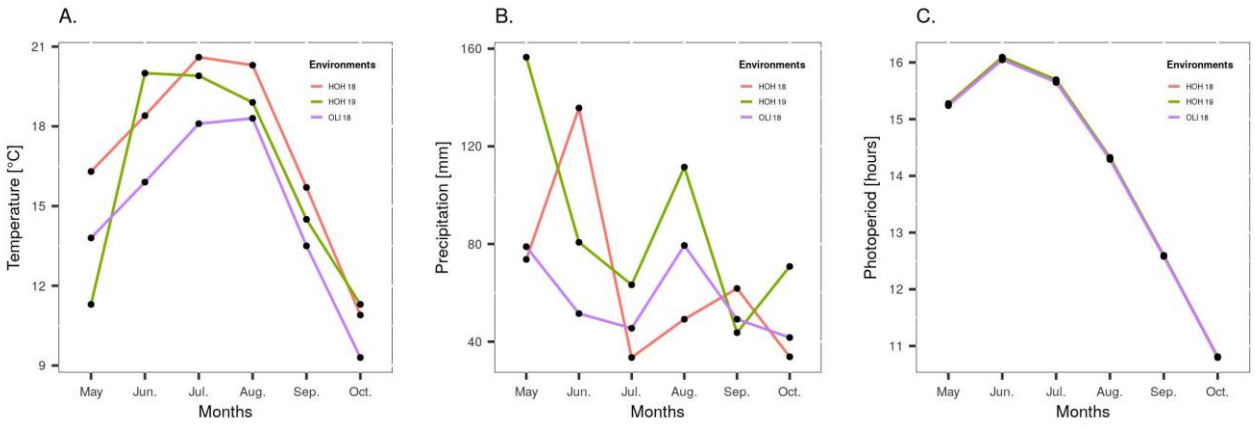

Figure S5. Monthly mean values of the three environmental variables belong to the three experiment environments (Hohenheim 2018-2019 and Oberer Lindenhof 2018) between May – October months.

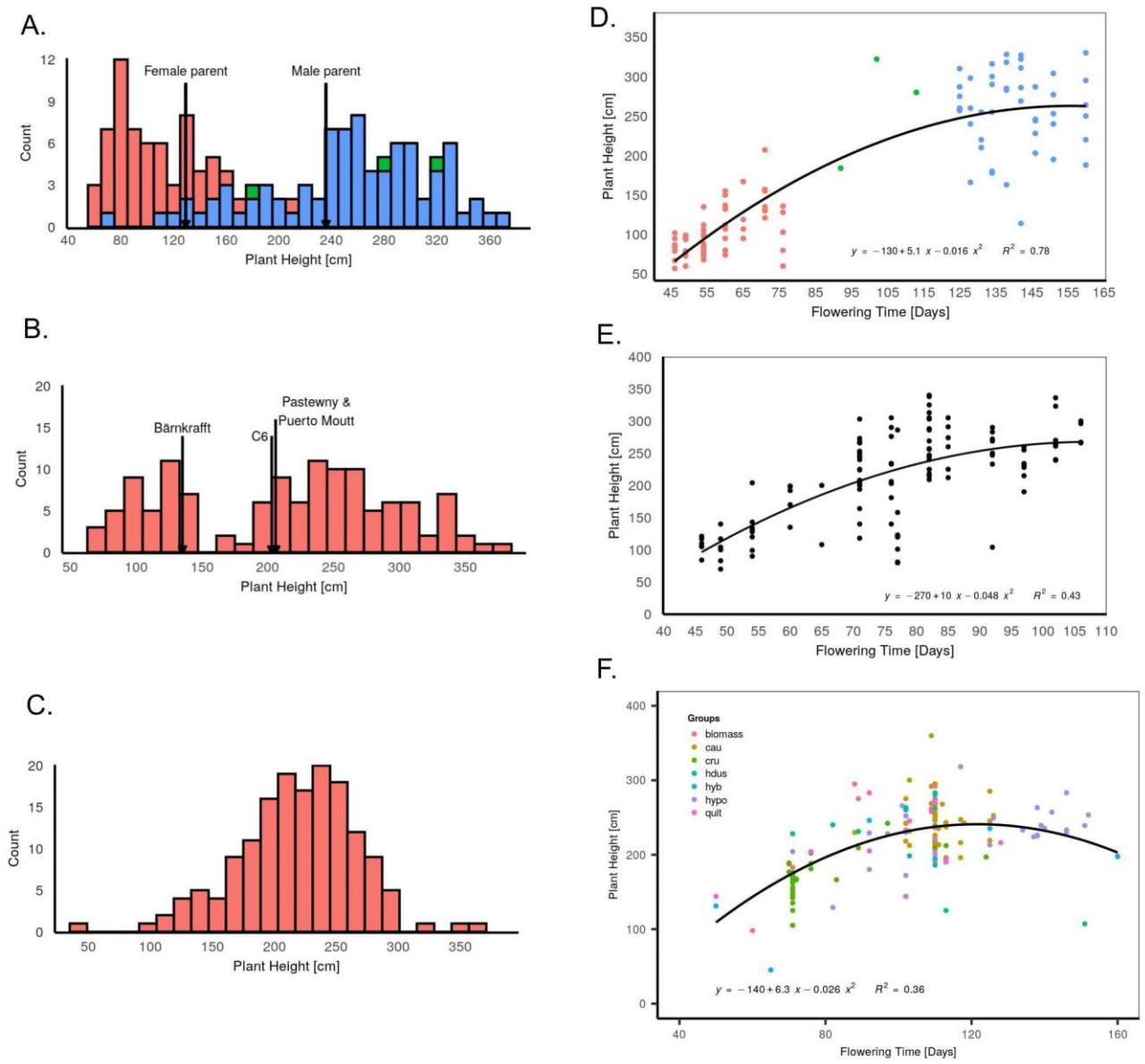

Figure S6. Plant height distributions and the relationship between flowering time and plant height in different mapping populations. Histograms of plant height in (**A**) *hypochondriacus* F_2 population, where red, blue and green indicate early, late and intermediate bulks, respectively, (**B**) the 'giant' families, (**C**) the genebank population. Arrows indicate the performances of the respective parent. The polynomial relationship between plant height and flowering time in (**D**) the *hypochondriacus* F_2 population, (**E**) the 'giant' families and in (**F**) the genebank population.

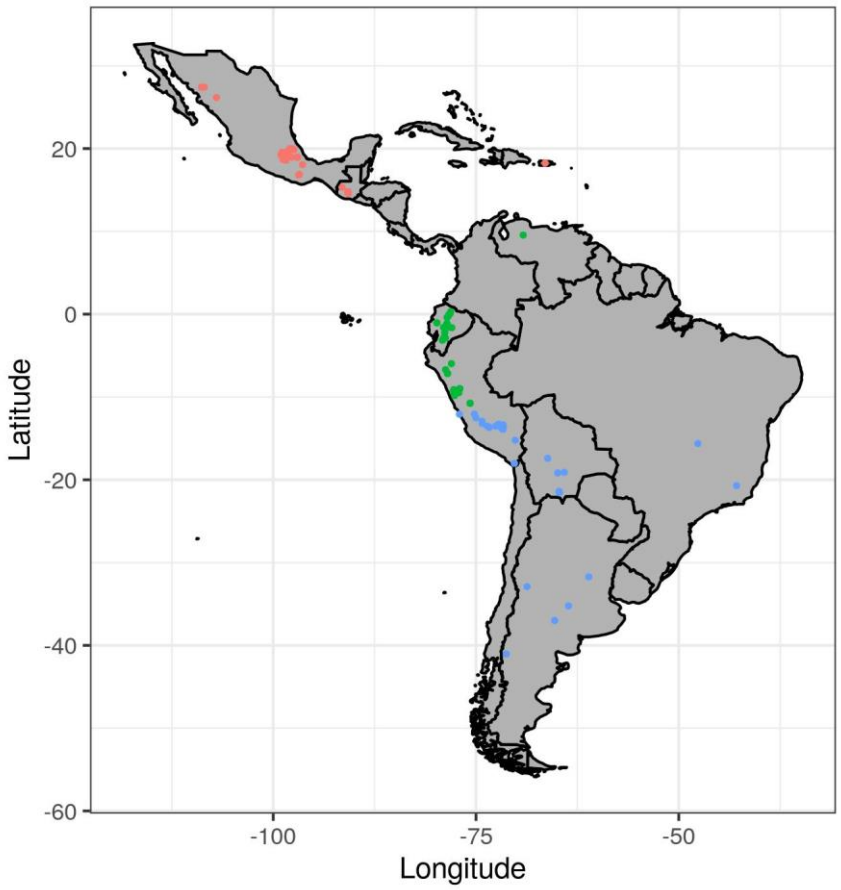

Figure S7. Geographical distribution of the accessions in the genebank population (n=140). Dot colors indicate the categorically determined latitudinal groups of the accessions. Red color indicates Central America group, green color indicates South America-I group and blue color indicates South America-II group.

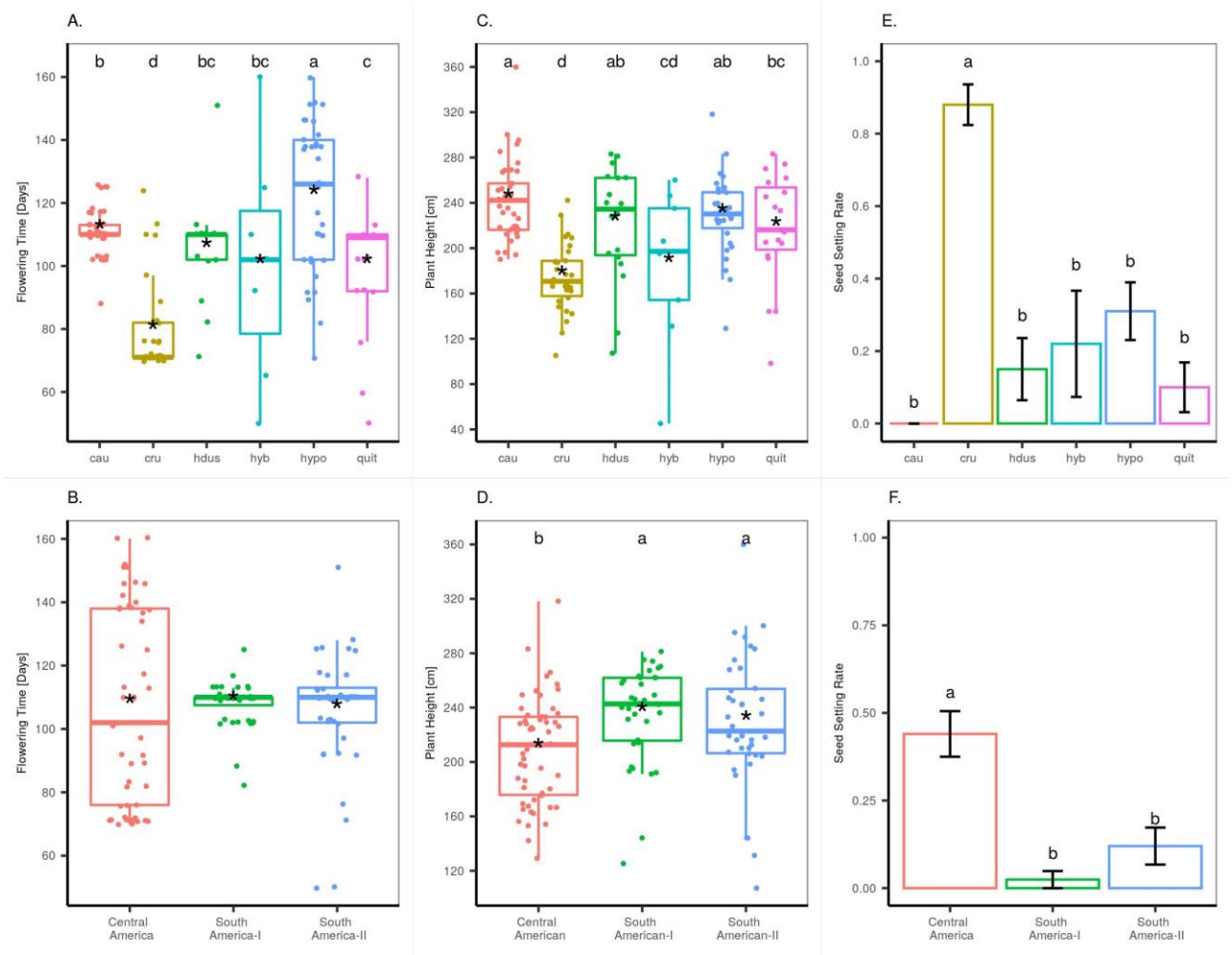

Figure S8. Comparison of the species and the latitudinal groups in the genebank population in terms of phenotypic traits. Box and bar plots of three phenotypic traits grouped by species and latitudinal groups, in each box and bar plot the respective groups were compared using a least significant difference (LSD) test and the groups with the different letters are significantly different at alpha=0.05. No letters if there is no significant difference among the compared groups. Black asterisks indicate the mean values of each group in the box plots. Species are cau, *A. caudatus*; cru, *A. cruentus*; hdus, *A. hybridus*; hyb, the 'hybrid' group; hypo, *A. hypochondriacus* and quit, *A. quitensis*.

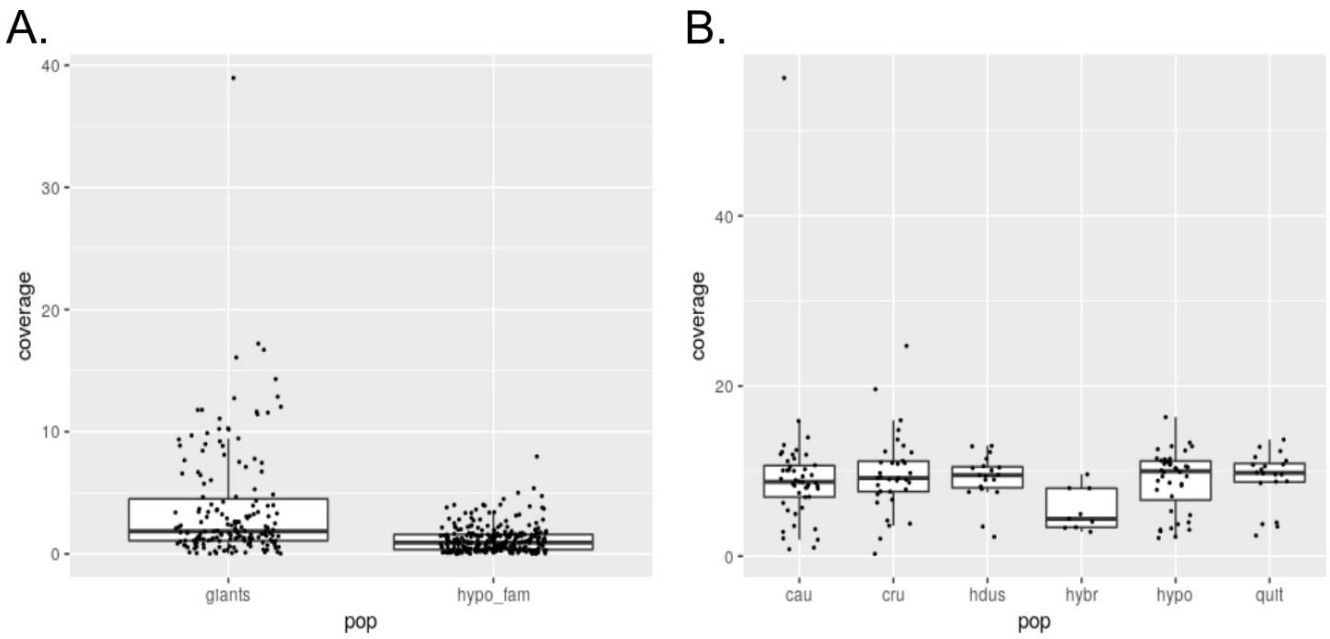

Figure S9. Boxplots of the mean sequencing X coverages of the populations. Each dot represents an individual. **(A)** The giant and hypochondriacus families. **(B)** The genebank population. Species are cau, *A. caudatus*; cru, *A. cruentus*; hdus, *A. hybridus*; hyb, the 'hybrid' group; hypo, *A. hypochondriacus* and quit, *A. quitensis*.

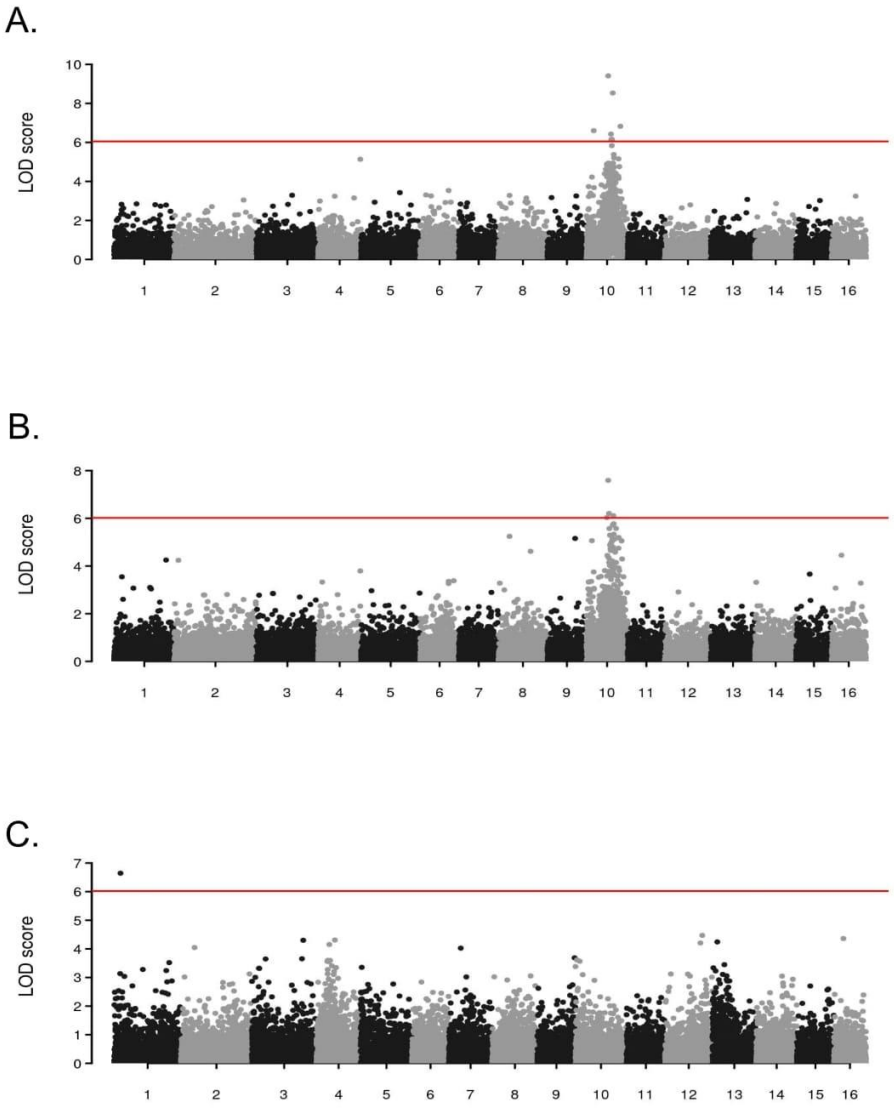

Figure S10. Manhattan plots of linkage mapping for **(A)** flowering time, **(B)** plant height in the F_2 generation, and for **(C)** flowering time of the F_3 generation of the hypochondriacus family. The red line indicates the significance threshold based on permutation test.

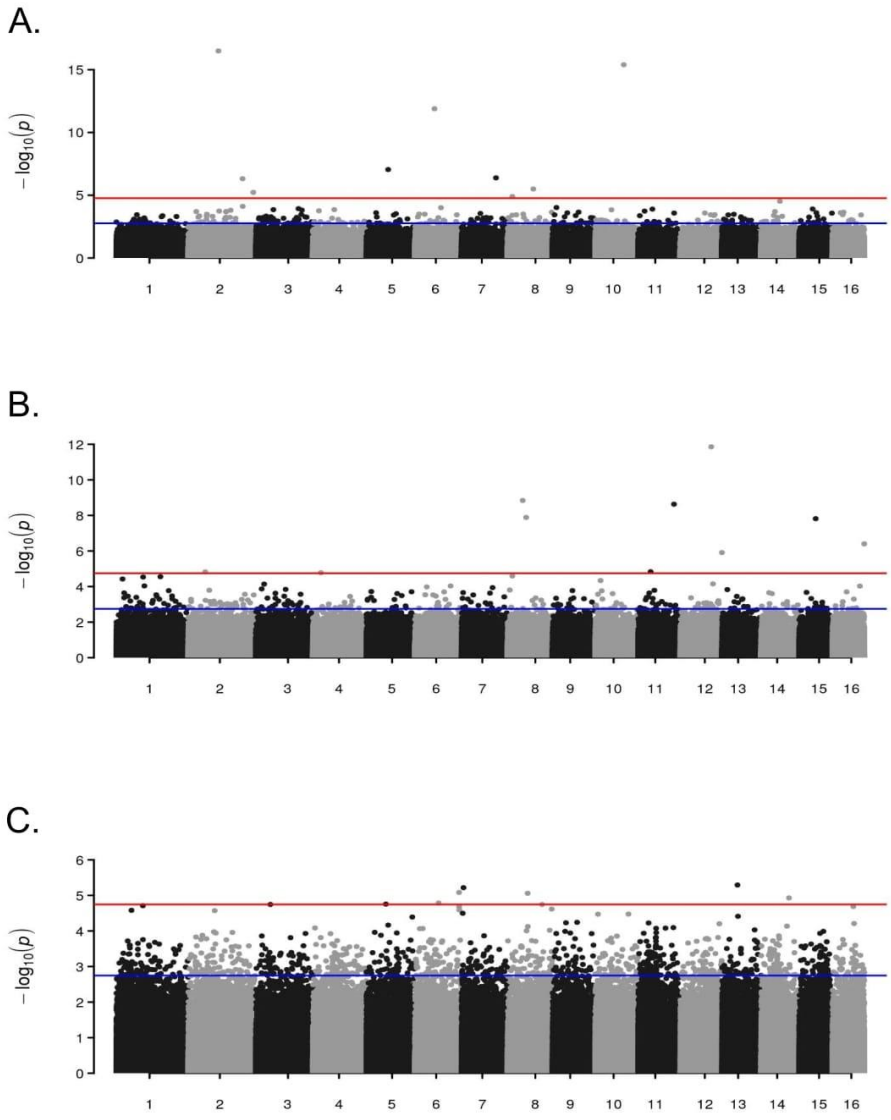

Figure S11. Manhattan plots of association mapping **(A)** for plant height in the giant families, for **(B)** flowering time and, **(C)** plant height in the genebank population, respectively. The blue line indicates the suggestive thresholds (1/ number of markers), and the red line shows the Bonferroni corrected thresholds for p=0.01.

5. General Discussion

Domestication, adaptation, and human-driven crop improvement are consecutive processes and may often overlap with each other. For example, loss of photoperiod sensitivity is considered as a domestication trait but at the same time allows the dispersal of short-day crops to higher latitudes and altitudes, hence, is also regarded as an adaptive trait. Moreover, it plays a paramount role in the breeding of amaranth for grain and biomass purposes. In this project, we studied photoperiod sensitivity in grain amaranths in terms of breeding, domestication history and adaptation genetics.

5.1. The novel findings

5.1.1. Chapter 2

In my first work, we showed that photoperiod sensitivity is responsible for two contrasting growth attitudes and separates grain types from the biomass type amaranths under long-day conditions. We proposed that a trade-off between earliness and photoperiod sensitivity is necessary for an optimized balance between dry matter content and yield to secure a feasible ensiling process. Dry matter content above 28% is required for a feasible ensiling procedure in the energy crops. Unfortunately, none of our tested genotypes managed to reach this threshold, in comparison to the strong competitors such as maize and sorghum. This results clearly showed that amaranth is not able to compete with well-established energy crops in a near future. I note that these crops have a long breeding history, and the modern breeding techniques have been consistently employed during this process. To his end, the use of modern breeding techniques such as marker-assisted selection and speed breeding can accelerate the genetic gain and may facilitate the use of amaranth as an alternative, only in the medium term, as discussed in the next sections.

5.1.2. Chapter 3

Given the determinant function of photoperiod sensitivity in breeding efforts, we studied the extent and the distribution of adaptive variation in different taxonomic and geographical units in the third chapter. We observed a higher photoperiodic variation in the Central American grain and wild species, in contrast to a narrower variation among the South American accessions that is limited with mild-photoperiod sensitivity. This result suggested a potential Central American origin of the wild relative *A. hybridus*, which might have migrated to South America and exposed to a selection for mild-photoperiod sensitivity. According to our interpretation, this was most likely an altitudinal selection that favors early flowering to escape from unfavorable cold temperatures during seed filling stage to ensure seed production, which was also reported in other plant species (Chapter 3). In addition, our result is in accordance with the model that proposing the independent domestication of the three grain amaranth species from *A. hybridus (Sauer, 1967)*. Subsequently, the consideration of adaptive photoperiod sensitivity and a potential altitudinal selection as a major potential instrument behind the genetic differentiation between the Central and South American *A. hybridus* species provided new insights into the domestication history of grain amaranth species, in contrast to previous studies that were mainly limited by phylogenetic approaches. We also studied the roles of environmental variables of origin in the adaptive success of the accessions and found that photoperiod sensitive accessions set seed independent from their origin, whereas mild-photoperiod accessions may also set seed if they originated from warm regions. This finding informs plant breeders which taxonomic and geographical units accommodate adaptive variation and provides insights on the local adaptability of grain amaranths.

5.1.3. Chapter 4

Finally in our fourth chapter, we consistently mapped the same genomic region, and identified the promising candidate gene '*response regulator of two-component systems*' as the putative *'A'* locus. Our analysis indicates an oligogenic architecture behind this trait, which is more frequently seen in the self-crossing crops such as rice and sorghum. Both phenotypic and

molecular results agreed with each other by showing the existence of photoperiod sensitivity centered pleiotropic effects. Our results also suggested the existence of a photoperiod sensitivity centered epistatic effect, however, phenotypic and molecular results disagreed with each other on the particular epistasis model, known as '*duplicative recessive epistasis*'. The marker data in the F_2 population showed a dominance effect of photoperiod sensitivity over insensitivity in the consensus region. Altogether, the exploration of this genomic region controlling photoperiodic variation across different populations has opened new avenues for a better understanding of the genetics of the flowering time adaptation, as well as a number of opportunities for a wider amaranth adaptation through breeding efforts.

5.2. Major constraints and their solutions

As semi-domesticated crops, grain amaranths have remarkably small seeds and are not very suitable for modern agriculture. In our field experiments, we had to sow the seeds manually, which was labor and time expensive. Due to the small size of the seeds, the young seedlings were highly vulnerable to environmental stresses, and most significantly to soil crusts emerging after heavy rains. Accordingly, we failed in more than 50 % of the field experiments during my studies and had to account for limitations in the experimental design and analysis such as a low number of replications. Insufficient number of seeds was another constraint behind such experimental setups. A minor difficulty we encountered was phenotypic heterogeneity in the field trials due to the residual genetic variation in the amaranth accessions used for the characterization of the biomass potential. As a result, phenotypic heterogeneity in the field limited mechanical harvesting and the respective dry matter yield estimation. Even though the visual biomass scoring was used as a proxy, a precise comparison between amaranths and the competitor crops such as sorghum and maize was no longer possible for this important trait. Another restriction during the biomass work was the insufficient representation of the grain type amaranths, because only a single variety (Bärnkrafft) was available. In future studies, a larger number of grain type varieties should be incorporated into the experimental setups for a more accurate dry matter content estimation.

In the third chapter, we observed a higher phenotypic variation in the Central American accessions in compared to the South American ones. This result suggests potential selection against high photoperiod sensitivity in the genomic region responsible for photoperiod sensitivity control. Similarly, Hotta *et al*. (2022) reported the breeding-driven selection events in the *PRR* gene family in many crops. Molecular footprints for the potential selection events were not investigated in our genebank population due to the sub-optimal sampling of the Central and South American *A. hybridus* accessions. A better representation of different taxonomic and geographical units should be ensured in the follow-up studies. In particular, I would like to stress the paramount role of *A. hybridus* in future research and breeding efforts of grain amaranths. Sinde it shows the widest geographical distribution (Sauer, 1967) and the highest nucleotide diversity within the hybridus complex (Kietlinski *et al.*, 2014), *A. hybridus* appears to be the major allelic resource against biotic and abiotic stresses in the future breeding efforts. The presence of a large number of genebank accessions for field testing (*n*=253) in our second work prevented the use of replicates across the consecutive years and different locations, despite the seed multiplication efforts in greenhouse for a limited number of accessions. Hence, future field experiments should take this restriction into account.

In our fourth chapter, a major constraint was the low quality of sequencing data for the 'hypochondriacus' population, which led to high segregation distortion and proportion of missing data, which complicated the detection of the particular epistasis model. In addition, 44 late bulk plants rejected flowering obstructed the precise estimation of the dominance and additive effects in the '*A*' locus. To address these issues, the parents and the progenies should be sequenced with higher coverage. As the segregation ratios between photoperiod sensitive and insensitive accessions differed across the different crossing combinations in the previously rice studies (Chang *et al.*, 1969; Nwe and Mackill, 1986; Poonyarit *et al.*, 1989), use of different combinations would provide more solid evidence about the underlying epistasis model.

RILs (Recombinant Inbred Line) can be easily generated using speed-breeding practices and would bring significant improvements to some of the major challenges we encountered during this work. More specifically, use of RILs would augment the number of tested individuals in

homozygous state and allow to test them in different locations and consecutive years, which increases the statistical power in linkage mapping. Moreover, RILs carry a larger number of recombination events and eventually their analysis would decrease the length of linkage blocks and increase the resolution in linkage mapping.

We attributed to the variation derived from the different juvenile period lengths to the hypothetical '*B*' locus, however, could not map it. An alternative approach to the mapping of this locus would be to phenotype our different populations under short-day conditions, where they cannot manifest their photoperiodic responses due to the deactivation of photoperiodic pathway. Hence, such an approach would allocate a higher statistical power to the detection of flowering time variation that is not controlled by photoperiodic pathway in the genetic mapping studies.

5.3. Aid of flowering time in the investigation of the parents of our base population

Another our interest in this project was to find out the putative parents involved in the spontaneous hybridization events in the 'giant' population. In 2014, Pietro Barbieri, a former master student in our group, tested putative hybrids from F_1, F_2 and F_3 generation families, and only Bärnkrafft, Puerto Moutt, C6 and Pastewny (*A. hybridus*) as the potential parents in field trials to observe segregation patterns in several agricultural and morphological traits, and some population genetic analyses were conducted in two gene sequences to verify the hybrid status of these families and to find out the parents of these families (Barbieri, 2014). As a result, occurrences of hybridization events were proven. However, parents were unable to be detected. In contrast, oligogenic control of photoperiod sensitivity resulted in distinct segregation patterns across the different populations and provided us with an alternative perspective by acting as a morphological marker, in this project.

In the targeted cross I did between the photoperiod sensitive and insensitive *A. hypochondriacus* accessions, I observed a 'dwarf' phenotype characterized by very early flowering, short plant height, a strong branching pattern and a resulting creeping stem.

Furthermore, we hypothesized that this ‘dwarf’ phenotype is a transgressive segregant that combines the earliness allele and the photoperiod insensitivity allele, which do not co-exist in a single parent (Chapter 4). This phenotype existed in the families I tested in 2016 for biomass traits and in 2019 field trials, in the F_2 and F_3 generations of the *A. hypocondriacus* family, and in the F_2 and F_3 generations of the putative hybrids, of which seeds were collected by Pietro Barbieri in 2013 (giant populations). Similarly, Pietro Barbieri also included a photograph of an F_3 family derived from Bärnkraft, where all individuals are ‘dwarf’ phenotype, and flowered very early similar to our records (~ 40 days), in his master thesis (**Figure 12B**). Therefore, I hypothesized that the ‘dwarf’ phenotype in our base population might have been resulted from the spontaneous crosses between photoperiod sensitive and insensitive varieties, same as in the *A. hypochondriacus* family.

Figure 12. Dwarf phenotypes. Photograph is taken (**A)** by Ali Baturaygil, (**B**) from the master thesis of Pietro Barbieri.

According to our phenotypic records, all the three potential female parents (Bärnkrafft, Puerto Moutt and C6) were photoperiod insensitive that flowered up to 76 days. Hence, the putative male parent(s) should have been photoperiod sensitive. Pietro Barbieri reported the flowering time performances of

the 10 genotypes that he phenotyped in 2013 in Kleinhohenheim location, all of which are the potential parents of the putative hybrids (**Figure 13**).

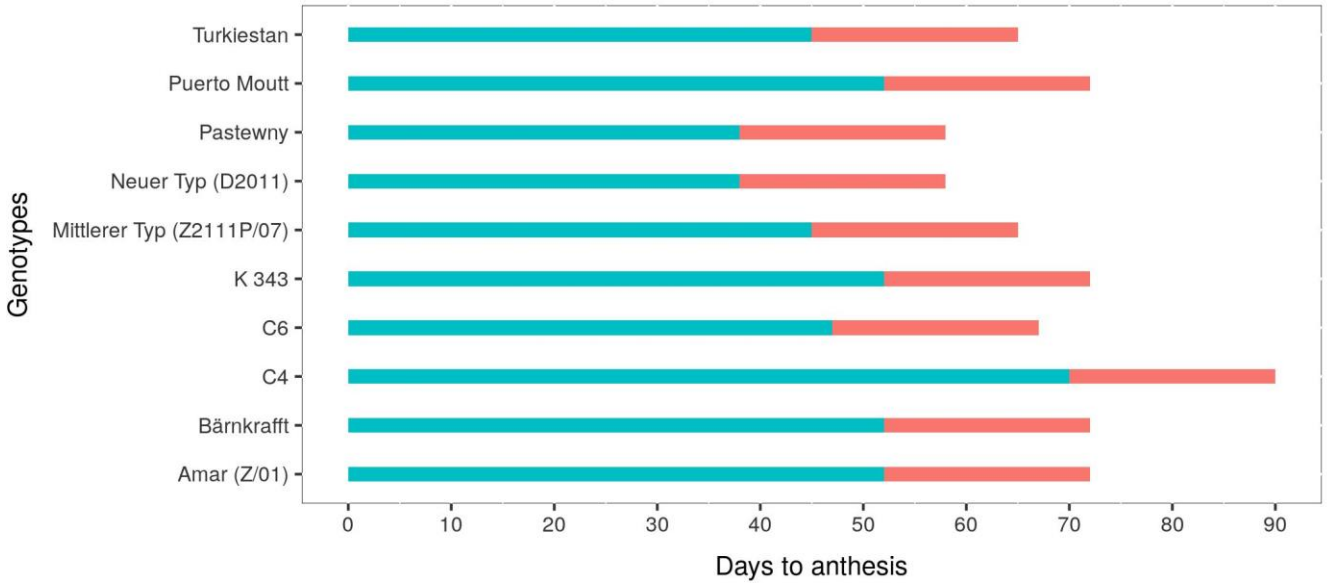

Figure 13. Approximate number of days to anthesis of the genotypes. Turquoise bars show the days from emergence to anthesis, and red bars show an approximate duration from sowing to emergence (taken as 20 days). Adapted from the report of Pietro Barbieri.

Of those potential parents, C4 variety (*A. caudatus*) appears to flower approximately in 90 days when the time between sowing and emergence taken into account, appears to be the only photoperiod sensitive genotype, and also similar to the early flowering mild-photoperiod sensitive *A. caudatus* genotypes, according to our photoperiodic-response classification. Therefore, we considered C4 as the most likely putative male parent and the photoperiod sensitivity donor. Unfortunately, we were unable to find the seeds of the C4 variety for sequencing and further population genetic analysis.

The potential participation of C4 genotype in the natural hybridization events with the genotypes from other species such as *A. cruentus* suggests that inter-specific hybridization between *A. caudatus* and *A. cruentus* is possible. Furthermore, such inter-specific hybridization events between *A. cruentus* and *A. caudatus* might have led to a higher extent of heterosis effect owing to a larger genetic distance between these species, which probably made

such gigantic plants more noticeable. In addition, our biomass lines – that hypothetically derived from *A. cruentus* x *A. caudatus* spontaneous hybridization – combined early flowering and long plant height. More specifically, four of them flowered earlier and reached higher plant height than the mean values of all the tested genotypes and set seed (**Figure 14**) in the experiment, where we tested the F_6-F_7 generations of a subset of our biomass genotypes together with the USDA genebank accessions (Chapter 3).

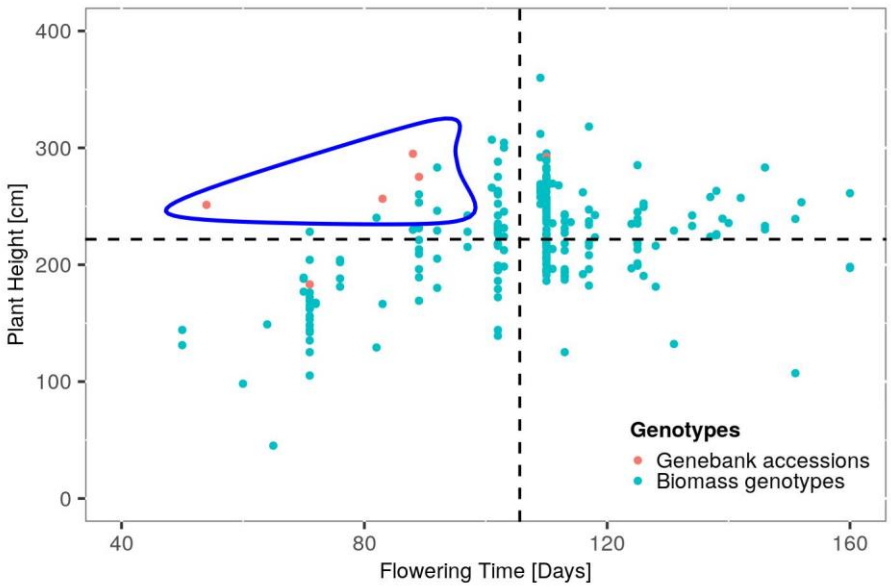

Figure 14. Scatter plot of the accessions tested in 2019; the genebank and the biomass genotypes. Dashed lines indicate the mean flowering time and plant height of the tested genotypes. The red observations in the blue cloud shows the biomass genotypes that flowered earlier and reached a taller plant height than the population means for these two traits.

This result suggests that the initial selection performed in 2013 managed to combine earliness that allows seed setting and a mild photoperiod sensitivity that allows above-average plant height, which was seldom among genebank accessions.

Segregation variances that increase in parallel with the differentiation of the parents gave us another clue about the potential involvement of C4 in the natural hybridization events. For

instance, the parents of *A. hypochnodriacus* cross differed in flowering by 56 days and the F_2 generation progeny showed a very wide range of flowering time from 46 days to 160 days, excluding the non-flowering ones, due to the different behaviors of the parents in this trait. In contrast, the flowering time range of the genotypes tested in 2012 was limited with only 20 days, and accordingly, their segregation variance ranged between 41 and 106 days, due to the more similar behaviors of the parents.

5.4. Future prospects of amaranth breeding

In our biomass work, dry matter content was insufficient in all our genotypes for a feasible ensiling process. Accordingly, we proposed a two-step strategy. Primarily, dry matter content should be improved with recurrent selection, and only after reaching a sufficient level, dry matter yield can be improved with the introgression of mild-photoperiod sensitivity alleles. An effective methodology would be a recurrent selection from the segregating families resulted from the experimental crosses between grain type amaranths. Recurrent selection can be applied as bulk selection combined with speed breeding. This approach is very similar to the RIL production, and the final selections can be performed under field conditions due to the complexity of the trait. A complementary strategy would be to explore the existing dry matter content variation in a subset of available germplasm. For example, Pietro Barbieri reported dry matter content above 30 % in his study, where he also reported 24 % dry matter content for Bärnkrafft, suggesting that his results are in line with our results.

Our results present new avenues for molecular breeding and wider adaptation possibilities for amaranth using marker-assisted selection or gene editing. Marker-assisted selection can be used to select for photoperiod sensitive or insensitive genotypes in variety of populations, and more importantly discriminate and prevent the selection of ‘dwarf’ phenotypes, which hypothetically combines photoperiod insensitivity and earliness alleles in the two major flowering time loci and are not suitable for commercial production. In this way, genotypes from variety of sources such as traditional varieties, crop wild relatives or populations resulted from natural or artificial hybridization can be easily detected and utilized in the breeding

programs without expensive and laborious field experiments. Ideally, KASP (Kompetitive Allele Specific PCR) markers (Semagn *et al.*, 2014) can be generated for the rapid selection of the alleles of choice. Therefore, KASP markers can be ideally used for the selection of the accessions combining high dry matter content and yield by eliminating photoperiod insensitive genotypes in the second stage of the above-mentioned selection strategy.

5.5. Re-domestication and *de novo* domestication possibilities through gene editing in domesticated and wild amaranths

Many orphan crop species have not completed their domestication syndrome and their large-scale cultivation is restricted because of their unfavorable wild attributes such as seed shattering, photoperiod sensitivity and branching. With the advancements in the plant genomics area, many major domestication genes were identified, and gene editing approaches allow the re-domestication of the crop species, or de-novo domestication of the wild species in a single generation (Kumar *et al.*, 2021). The most frequent application of gene editing in orphan crops is gene knock outs. That includes the loss of function, but novel genetic variation can be also generated by the induction of differential gene expression via promoter editing (Kumar *et al.*, 2021; Venezia and Creasey Krainer, 2021). In addition, gene editing allows the introduction of the alleles of interest without the undesirable linkage drag (Venezia and Creasey Krainer, 2021). Gene editing approaches primarily require a genomic target with its well-known function and a construct design to generate targeted mutations (Venezia and Creasey Krainer, 2021). No gene editing in amaranths has been performed to date (Venezia and Creasey Krainer, 2021), despite the availability of genomic resources such as high-quality reference genome, which is lacking in many orphan species. However, that may change in a near future with the major finding of our research and its follow ups. Similar to marker-assisted selection, a simple inheritance with major QTLs increases the applicability and success chance of gene editing (Bernardo, 2016), which makes our candidate gene a promising target. For example, a gene editing approach carried out the *de novo* domestication of wild tomato by combining four domestication syndrome phenotypes with simple inheritance including loss of photoperiod sensitivity (Li *et al.*, 2018).

Gene editing may offer interesting opportunities in amaranths. First, mild and high photoperiod sensitive genotypes from variety of species and latitudinal groups can be manipulated through targeted mutations to be photoperiod insensitive and may remarkably diversify the base populations in the high latitudes and altitudes for grain amaranth improvement. Higher dry matter content driven by improved grain yield would also contribute to the improvement of biomass amaranths. Such an approach may also allow the exploitation of commercial heterosis in grain amaranths if the genetic basis of cytoplasmic male sterility is dissected in a near future. More generally, targeted mutations in the photoperiod sensitivity QTL would alternate the photoperiodic responses of the cultivated and wild putative ancestor species. Hence, such a new and novel adaptive variation would contribute to the efforts against climate change by permitting a wider dispersal of these species. However, we note that gene editing technology has its own limitations such as off-target activity, and transformation and regeneration bottlenecks, that are independent from the species (Venezia and Creasey Krainer, 2021). More importantly, gene knock-outs via gene editing technology may result in unexpected phenotypes due to the poorly understood pleiotropic relationships in the '*A*' locus.

Our 'dwarf' phenotype combines early flowering with the morphological traits such as lodging and strong branching, which are wild characters. Kulakow and Jain (1985) also reported that flowering time was closely correlated with some plant architecture related domestication traits in a family of a cross between wild and cultivated species. A deeper understanding of the pleiotropic relationships between the '*A*' locus and such domestication traits may also contribute to the domestication of wild and cultivated amaranths via gene editing. For example, several desired phenotypes may be combined in a single individual, such as the early flowering 'dwarf' phenotype that is not lodging and branching.

5.6. Our speed breeding applications in grain amaranth

Especially in the last five years, speed breeding has become a popular tool in plant sciences (Watson *et al.*, 2018) and aims to increase the genetic gain by shortening the growth cycles of the crops to accelerate the selection process. Shortened generation time achieved by

manipulation of the growth conditions is a longstanding practice in plant breeding. Amaranth is an excellent organism for flowering time and photoperiod studies owing to its high photoperiodic variation. The 70 years old pioneer study of Fuller *et al*. (1949) reported that *A. caudatus* needs short day-length to flower, and also rejected flowering, and produced high dry weight under long day-length, which was a sort of basis for our studies. Several aspects of speed breeding practices have been implemented in amaranth for longer than 25 years (Lehmann, 1995; Brenner and Widriechner, 1998; Stetter *et al.*, 2016; Jähne *et al.*, 2020). Of those, more traditional approaches used day-length and temperature manipulations (Lehmann, 1995; Brenner and Widriechner, 1998; Stetter *et al.*, 2016), whereas the recent approach i.e., manipulation of light quality parameters was successfully applied by Jähne *et al*. (2020). In our research, I successfully benefited from speed breeding applications in several ways.

5.6.2. Making of the experimental crosses

At the beginning of the project, I grew two batches of 200 genotypes from the hybridus complex species by sowing the seeds with one month interval under the growth conditions as described in the Chapter 4. Photoperiod was controlled using time adjusted curtains. The natural flowering time variation of the genotypes and the use of the consecutive batches allowed me to make crosses between the genotypes that flower simultaneously (Chapter 4).

The success rate of open pollination was around 15 % and a high level of dormancy in the harvested seeds made difficult to produce sufficient number of F_1 hybrid plants for seed production. In the future experiments, I solved this problem with two different approaches. First, I exposed the plants to drought stress (Karimmojeni *et al.*, 2014) by gradually decreasing the amount of water I gave from the onset of the seed formation process, which helped to break their dormancy. In another successful approach, I germinated the dormant seeds using gibberelin (GA_3). Kepczynski *et al*. (1996) reported that 10^{-3} M GA_3 is very effective in breaking dormancy, which led a germination rate above 90 % in my experiment (no data available). However, this method was implemented in germination medium i.e., the

germinated seeds must be transplanted to the actual growth medium, which is laborious and time-consuming but successful.

5.6.1. Speed breeding through light quality manipulation

In 2019, I had a chance to perform a pre-experiment with a small subset of our genotypes with variable flowering times in the speed breeding chambers, which was published by Jähne *et al*. (2020). We tested the influence of red and far-red light on flowering time and plant height in that experiment. Particularly, two of the ten tested genotypes were the parents of the *A. hypochondriacus* cross that segregate for photoperiod sensitivity. To this end, this experiment helped me to observe the short-day performances of those parental genotypes. Furthermore, I compared the long and short-day performances of the genotypes that were tested under both this experiment and in the field experiment in 2019, which revealed the photoperiodic status of the genotypes (Chapter 4, **Figure 8I**).

Only, the photoperiod insensitive female parent did not flower in the experiment, however, emerged inflorescence at the day 69 under far-red light treatment and suggested that it was going to flower within several days. As a result, two treatments did not significantly differ in flowering time but genotypes under far-red light flowered four days earlier in average. More importantly, photoperiod sensitive accessions showed a very dramatic decrease in flowering time under short-day speed-breeding conditions in compared to the long-day field conditions, which was longer than 100 days in the male parent of the *A. hypochondriacus* cross. In contrast, a cross I made between two photoperiod insensitive *A. cruentus* accessions and grew the F_1 generation plants in a speeding breeding chamber under the photoperiod of 22 hours, and plants emerged inflorescence only within three weeks (no data available). Altogether, speed breeding through light quality manipulation has an outstanding potential in amaranth and an amaranth-specific speed breeding protocol is required.

5.7. References

Abberton M et al. (2016) Global agricultural intensification during climate change: A role for genomics. *Plant Biotechnology Journal* 14: 1095–1098.

Aderibigbe OR et al. (2020) Exploring the potentials of underutilized grain amaranth (*Amaranthus* spp.) along the value chain for food and nutrition security: A review. *Critical Reviews in Food Science and Nutrition* 0: 1–14.

Andini R, Sulaiman MI, Moulana R, Hmon KPW and Ohsawa R (2020) Application of principle component analysis in differentiating the three types of *Amaranthus* based on their photoperiodic flowering response. *IOP Conference Series: Earth and Environmental Science* 425: 0–11.

Balodis O, Bartuševics J and Gaile Z (2011) Biomass yield of different plants for biogas production. *Vide, Tehnologija, Resursi - Environment, Technology, Resources* 1: 238–245.

Barbieri P (2014) Analysis of a segregating population of amaranth (*Amaranthus* spp.) originated from spontaneous crossings leading to potential inter-specific hybrids. Master thesis, University of Hohenheim.

Becker R et al. (1981) A Compositional Study of Amaranth Grain. *Journal of Food Science* 46: 1175–1180.

Bernardo R (2016) Bandwagons I, too, have known. *Theoretical and Applied Genetics* 129: 2323–2332.

Brauer-Siebrecht W et al. (2016) Silage Maize and Sugar Beet for Biogas Production in Rotations and Continuous Cultivation: Dry Matter and Estimated Methane Yield. *Agronomy* 6: 2.

Brenner DM et al. (2000) Genetic resources and breeding of *Amaranthus*. *Plant Breeding Reviews* 19: 227–285.

Brenner DM (2019) Registration of DB 199313, Cytoplasmic Male Sterile Grain Amaranth Genetic Stock. *Journal of Plant Registrations* 0: 0.

Brenner DM and Widriechner MP (1998) *Amaranthus* seed regeneration in plastic tents in greenhouses. *Plant Genetic Resources Newsletter* 1–4.

Ceccarelli S (1994) Specific adaptation and breeding for marginal conditions. *Euphytica* 77: 205–219.

Chang TT, Li CC and Vergara BS (1969) Component analysis of duration from seeding to heading in rice by the basic vegetative phase and the photoperiod-sensitive phase. *Euphytica* 18: 79–91.

Chaudhari S et al. (2017) Biology and Management of Glyphosate-Resistant and Glyphosate-Susceptible Palmer Amaranth (*Amaranthus palmeri*) Phenotypes from a Segregating Population. *Weed Science* 65: 755–768.

von Cossel M, Möhring J, Kiesel A and Lewandowski I (2017) Methane yield performance of amaranth (*Amaranthus hypochondriacus* L.) and its suitability for legume intercropping in comparison to maize (*Zea mays* L.). *Industrial Crops and Products* 103: 107–121.

Costea M and Tardif FJ (2004) The biology of Canadian weeds. 130. *Amaranthus retroflexus* L., *A. powellii* S. Watson and *A. hybridus* L. (update). *Canadian Journal of Plant Science* 84: 631–668.

Craufurd PQ and Wheeler TR (2009) Climate change and the flowering time of annual crops. *Journal of Experimental Botany* 60: 2529–2539.

Das S (2014) Domestication, phylogeny and taxonomic delimitation in underutilized grain *Amaranthus* (*Amaranthaceae*) - a status review. *Feddes Repertorium* 123: 273–282.

Delgado JA et al. (2011) Conservation practices to mitigate and adapt to climate change. *Journal of Soil and Water Conservation* 66: 118–129.

Ebert AW (2014) Potential of Underutilized Traditional Vegetables and Legume Crops to Contribute to Food and Nutritional Security, Income and More Sustainable Production Systems. *Sustainability* 6: 319–335.

Esposito S, Carputo D, Cardi T and Tripodi P (2020) Applications and Trends of Machine Learning in Genomics and Phenomics for Next-Generation Breeding. *Plants* 9.

Federal Ministry FACP (2019) Bioenergy in Germany : Facts and Figures - Solid fuels, biofuels & biogas.

Fuller HJ (1949) Photoperiodic Responses of *Chenopodium quinoa* Willd. and *Amaranthus caudatus* L . *American Journal of Botany* 36: 175–180.

Gaudinier A and Blackman BK (2019) Evolutionary processes from the perspective of flowering time diversity. *New Phytologist* 225: 1883–1898.

Gepts P (2004) Plant Crop Domestication as a Long-Term Selection Experiment. *Breeding Reviews* 24 (Part 2: 1–44.

Hauptli H and Jain S (1985) Genetic variation in outcrossing rate and correlated floral traits in a population of grain amaranth (*Amaranthus cruentus* L.). *Genetica* 66: 21–27.

Haussmann BIG et al. (2012) Breeding Strategies for Adaptation of Pearl Millet and Sorghum to Climate Variability and Change in West Africa.

Hoidal N, Díaz Gallardo M, Jacobsen SE and Alandia G (2019) Amaranth as a dual-use crop for leafy greens and seeds: Stable responses to leaf harvest across genotypes and environments. *Frontiers in Plant Science* 10.

Hotta CT (2022) The evolution and function of the PSEUDO RESPONSE REGULATOR gene family in the plant circadian clock. *Genetics and Molecular Biology* 45: 4–5.

Izawa T (2007) Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *Journal of Experimental Botany* 58: 3091–3097.

Jackson SD (2009) Plant responses to photoperiod. *New Phytologist* 181: 517–531.

Jähne F, Hahn V, Würschum T and Leiser WL (2020) Speed breeding short-day crops by LED-controlled light schemes. *Theoretical and Applied Genetics*.

Jain SK, Hauptli H and Valdya KR (1982) Outcrossing rate in grain amaranths. *Journal of Heredity* 73: 71–72.

Joshi DC et al. (2018) From zero to hero: the past, present and future of grain amaranth breeding. *Theoretical and Applied Genetics* 131: 1–17.

Jung C and Müller AE (2009) Flowering time control and applications in plant breeding. *Trends in Plant Science* 14: 563–573.

Karimmojeni H, Bazrafshan AH, Majidi MM, Torabian S and Rashidi B (2014) Effect of maternal nitrogen and drought stress on seed dormancy and germinability of *Amaranthus retroflexus*. *Plant Species Biology* 29.

Kepczynski J, Corbineau F and Come D (1996) Responsiveness of *Amaranthus retroflexus* seeds to ethephon, 1-aminocyclopropane 1-carboxylic acid and gibberellic acid in relation to temperature and dormancy. *Plant growth regulation* 20: 259–265.

Kietlinski KD et al. (2014) Relationships between the weedy *Amaranthus hybridus* (*Amaranthaceae*) and the grain amaranths. *Crop Science* 54: 220–228.

Kulakow PA and Jain SK (1985) The inheritance of flowering time in *Amaranthus* species. *Journal of Genetics* 64: 85–100.

Kumar R et al. (2021) Understanding Omics Driven Plant Improvement and *de novo* Crop Domestication: Some Examples. *Frontiers in Genetics* 12: 1–25.

Lagercrantz U (2009) At the end of the day: a common molecular mechanism for photoperiod responses in plants? *Journal of Experimental Botany* 60: 2501–2515.

Lehmann JW (1995) Rapid cycling of grain amaranths. *Legacy* 8: 15–17.

Lehmann JW, Clark RL and Frey KJ (1991) Biomass Heterosis and Combining Ability in Interspecific and Intraspecific Matings of Grain Amaranths. *Crop science* 31: 1111–1116.

Li T et al. (2018) Domestication of wild tomato is accelerated by genome editing. *Nature Biotechnology* 36: 1160–1163.

Lightfoot DJ et al. (2017) Single-molecule sequencing and Hi-C-based proximity-guided assembly of amaranth (*Amaranthus hypochondriacus*) chromosomes provide insights into genome evolution. *BMC Biology* 15: 1–15.

Liu F and Stützel H (2004) Biomass partitioning, specific leaf area, and water use efficiency of vegetable amaranth (*Amaranthus* spp.) in response to drought stress. *Scientia Horticulturae* 102: 15–27.

Massawe F, Mayes S and Cheng A (2016) Crop Diversity: An Unexploited Treasure Trove for Food Security. *Trends in Plant Science* 21: 365–368.

Matsubara K et al. (2011) *Ehd3*, encoding a plant homeodomain finger-containing protein, is a critical promoter of rice flowering. *Plant Journal* 66: 603–612.

Mayes S et al. (2012) The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany* 63: 1075–1079.

Murphy RL et al. (2014) *Ghd7* (*Ma 6)* Represses Sorghum Flowering in Long Days: *Ghd7* Alleles Enhance Biomass Accumulation and Grain Production. *The Plant Genome* 7: plantgenome2013.11.0040.

Myers RL (1996) Amaranth: New crop opportunity. *Progress in new crops* 207–220.

Nakamichi N (2014) Adaptation to the local environment by modifications of the photoperiod response in crops. *Plant and Cell Physiology* 56: 594–604.

National Research Council (1984) *Amaranth Modern Prospects for an Ancient Crop*. National Academy Press.

Nwe KT and Mackill DJ (1986) Photoperiod sensitivity in three rice crosses. *Euphytica* 35: 751–760.

Pingali PL (2012) Green revolution: Impacts, limits, and the path ahead. *Proceedings of the National Academy of Sciences of the United States of America* 109: 12302–12308.

Poonyarit M, Mackill DJ and Vergara BS (1989) Genetics of photoperiod sensitivity and critical daylength in rice. *Crop Science* 29: 647–652.

Rivelli AR, Gherbin P, de Maria S and Pizza S (2008) Field evaluation of *Amaranthus* species for seed and biomass yields in southern Italy. *Italian Journal of Agronomy* 3: 225–229.

Roux F, Touzet P, Cuguen JL and Le Corre V (2006) How to be early flowering: an evolutionary perspective. *TRENDS in Plant Science* 11.

Sauer JD (1967) The grain amaranths and their relatives: A revised taxonomic and geo-graphic survey. *Ann. Missouri Bot. Gard.* 54: 102-137.

Scott MF et al. (2020) Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity* 125: 396–416.

Semagn K, Babu R, Hearne S and Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement.

Seppälä M, Laine A and Rintala J (2013) Screening of novel plants for biogas production in northern conditions. *Bioresource Technology* 139: 355–362.

Stallknecht GF and and Schulz-Schaeffer JR (1993) Amaranth Rediscovered. *In: J. Janick and J.E. Simon (eds.)* 211–218.

Stetter MG et al. (2016) Crossing Methods and Cultivation Conditions for Rapid Production of Segregating Populations in Three Grain Amaranth Species. *Frontiers in plant science* 7.

Stetter MG and Schmid KJ (2017) Analysis of phylogenetic relationships and genome size evolution of the *Amaranthus* genus using GBS indicates the ancestors of an ancient crop. *Molecular Phylogenetics and Evolution* 109: 80–92.

Stetter MG, Vidal-Villarejo M and Schmid KJ (2020) Parallel Seed Color Adaptation during Multiple Domestication Attempts of an Ancient New World Grain. *Molecular biology and evolution* 37: 1407–1419.

Valcárcel-Yamani B and Lannes SC da S (2012) Applications of Quinoa (*Chenopodium Quinoa* Willd.) and Amaranth (*Amaranthus* spp.) and their influence in the nutritional value of cereal based foods. *Food and Public Health* 2: 265–275.

Venezia M and Creasey Krainer KM (2021) Current Advancements and Limitations of Gene Editing in Orphan Crops. *Frontiers in Plant Science* 12.

Venskutonis PR and Kraujalis P (2013) Nutritional Components of Amaranth Seeds and Vegetables: A Review on Composition, Properties, and Uses. *Comprehensive Reviews in Food Science and Food Safety* 12: 381–412.

Vogel E, Deumlich D and Kaupenjohann M (2016) Bioenergy maize and soil erosion - Risk assessment and erosion control concepts. *Geoderma* 261: 80–92.

Watson A et al. (2018) Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* 4: 23–29.

Weiland P (2010) Biogas production: Current state and perspectives. *Applied Microbiology and Biotechnology* 85: 849–860.

Wu H et al. (2000) Field evaluation of an *Amaranthus* genetic resource collection in China. *Genetic Resources and Crop Evolution* 47: 43–53.

Xu Y, Li P, Yang Z and Xu C (2017) Genetic mapping of quantitative trait loci in crops. *Crop Journal* 5: 175–184.

Zargar SM et al. (2015) Recent advances in molecular marker techniques: Insight into QTL mapping, GWAS and genomic selection in plants. *Journal of Crop Science and Biotechnology* 18: 293–308.

Zhang H, Li Y and Zhu JK (2018) Developing naturally stress-resistant crops for a sustainable agriculture. *Nature Plants* 4: 989–996.

Zhang L et al. (2020) Principles and practices of the photo-thermal adaptability improvement in soybean. *Journal of Integrative Agriculture* 19: 295–310.

Zohary D, Hopf M and Weiss E (2012) *Domestication of Plants in the Old World*. Oxford Univeristy

Acknowledgments

I want to express my deep appreciation to the following people for their support during climbing my Everest.

First of all, I would like to thank to Karl for trusting me and giving me this chance. Furthermore, I want to thank you for your supervision, and especially for giving me the scientific freedom, that helped me to find my domain in the universe of plant breeding research.

Second, I want to thank to the previous and current members of the Schmid Lab, for the fruitful scientific discussions and explanations, paperwork, library preparation and many other things I cannot count here.

To Bärbel and Anja for helping me to survive against bureaucracy, to Elisabeth and Vanessa for the things we briefly describe in materials and methods but actually are the baseline of our findings – high quality DNA extraction and library preparation, to Fabian Freund, Katharina Böndel, and Christian Lampei for the explanatory discussions and time-saving hints, to Viola Abraham for a variety of support in greenhouse and field, to Mireia for estimating the ‘major SNP matrix’ and for her patient explanations to my ‘bioinformatics for dummies’ level questions.

Another round of thanks goes to the PhD students from our and neighboring groups. If the research was the shape of my PhD life, you were the color of it. Besides all the fun, symposiums, trips, 17.00 o’clock beers in Denkbar, and the scientific discussions, I would like to thank you for many different things. To Che-Wei Chang for troubleshooting my error messages with some of his outstanding programming skills, to Flavio Isla Lozano for many discussions and the Spanish translation of the abstract, to Thea Mi Weiß for her jogging companion, to Paul Gruner for helping me in bureaucratic problems, to Felix Jähne for discovering the speed breeding potential of amaranths with me, to Kenan Gençtürk for helping me to break the dormancy of recalcitrant amaranth seeds, to Muhammad Afzal and David

Sewordor Gaikpa for their encouragement, especially when I was feeling desperate about the project, and to Sandra Roller for the German translation of the abstract. In addition, I would like to thank some of the other colleagues during my journey for their companion; Julia, Sonja, Bianca, Max, Rodrigo, Moctar, Belén, Anna, Jan, Matthias, Felix, Johannes, Xintian, Pedro, Ana, Anurag, Khaoula, Cleo, Felicien, Lea, and Niharika.

I want to thank to Prof. Dr. Detlef Weigel, for his support in the project.

Without the valuable support of the staff from the experimental stations and the greenhouse of Hohenheim University, phenotypic data generation would have been nearly impossible… Thank you.

I would like to thank to the Hiwi students for their contributions at the different stages of the field trials, to Miguel, Azadeh, Nancy and Stefano. I also thank to Pietro Barbieri for sharing his previous work with me, and to Markus Stetter for contributing to the biomass paper.

I want to thank to the Hohenheim-Tübingen Regional alliance of the Ministry of Science and Culture (MWK) of Baden Württemberg for financing my project.

I would like to thank to my chess friends from the ‘Vaihingen-Rohr Chess Community’. The time I spent over the chessboard helped me to relieve myself from the pandemics and work-related stress, and hopefully contributed to me at being a better decision-maker.

Finally, without the life-long support of my family, it would have been impossible and meaningless.

And Fındık, you are my source of inspiration.

Annex 3

Declaration in lieu of an oath on independent work

according to Sec. 18(3) sentence 5 of the University of Hohenheim's Doctoral Regulations for the Faculties of Agricultural Sciences, Natural Sciences, and Business, Economics and Social Sciences

1. The dissertation submitted on the topic

"**Studies on flowering time and photoperiod sensitivity in domesticated and wild amaranth (*Amaranthus* spp.) species**".

is work done independently by me.

2. I only used the sources and aids listed and did not make use of any impermissible assistance from third parties. In particular, I marked all content taken word-for-word or paraphrased from other works.

3. I did not use the assistance of a commercial doctoral placement or advising agency.

4. I am aware of the importance of the declaration in lieu of oath and the criminal consequences of false or incomplete declarations in lieu of oath.

I confirm that the declaration above is correct. I declare in lieu of oath that I have declared only the truth to the best of my knowledge and have not omitted anything.

Stuttgart, 01.02.2023

Place, Date

Signature

Ali Baturaygil

Phone: (+49) 163 246 42 61

Address: Stuttgart, Germany

Email: alibaturaygil88@gmail.com

EXPERIENCE

May Agro Seed Corporation, Turkey
July 2012 - September 2012

Intern

Plant Breeding Chair, Technical University of Munich, Germany
July 2011 - October 2011

Intern

EDUCATION

University of Hohenheim, Germany
July 2017 - 2023

Ph.D. Candidate

Major: Plant Breeding

University of Hohenheim, Germany
October 2013 - December 2016

Master of Science

Major: Plant Breeding

University of Ankara, Turkey
September 2009 - June 2013

Bachelor of Science

Major: Field Crops

LANGUAGES

Turkish (Native), English (Good), German (Intermediate)

NATIONALITY

Turkish

Ali Baturaygil

Created using Resumonk - Online Resume Builder