Genetic analysis of metal hyperaccumulation and hypertolerance of *Noccaea caerulescens*

Jitpanu Yamjabok

Propositions

1. Affordable high-throughput whole genome sequencing methods greatly improve QTL detection in *Noccaea caerulescens*, though identification of the causal variant remains challenging.
(this thesis)

2. Removing the vernalization requirement is key to obtain a successful transformation protocol for *Noccaea caerulescens*.
(this thesis)

3. The draft whole genome sequences of *Durio zibethinus*, durian, are crucial for the development of new varieties through genomic selection.
(Teh et al., 2017, Nat Genet 49(11): 1633-1641; et al., 2023, Front Plant Sci 14:1137077)

4. Coding literacy is a crucial skill for data analysis in scientific research nowadays.

5. People often overlook how politics function until they experience an increase in their cost of living.

6. Embracing optimism is a practical way to adjust to the frequently unfavourable weather in the Netherlands.

Propositions belonging to the thesis entitled:

"Genetic analysis of metal hyperaccumulation and hypertolerance of *Noccaea caerulescens*"

Jitpanu Yamjabok

Wageningen, 24 April 2024

Genetic analysis of metal hyperaccumulation and hypertolerance of *Noccaea caerulescens*

Jitpanu Yamjabok

Thesis committee

Promotors

Prof. Dr. M.G.M. Aarts

Personal chair at the Laboratory of Genetics

Wageningen University & Research

Co-promotor

Dr. J. van den Heuvel

Assistant Professor, Laboratory of Genetics

Wageningen University & Research

Other members

Prof. Dr. M.E. Schranz, Wageningen University & Research

Dr. A.G.L. Assunção, Universidade do Porto, Portugal

Prof. Dr. S. Clemens, University of Bayreuth, Germany

Dr. P Vergeer, Wageningen University & Research

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences

Genetic analysis of metal hyperaccumulation and hypertolerance of *Noccaea caerulescens*

Jitpanu Yamjabok

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday April 24, 2024
at 16:00. p.m.

Jitpanu Yamjabok
Genetic analysis of metal hyperaccumulation and hypertolerance of Noccaea caerulescens, 177 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

DOI https://doi.org/10.18174/651745

Dedicated to my beloved grandfather Dee Yamjabok.

Contents

Chapter 1

General introduction 9

Chapter 2

Development of an immortal bi-parental mapping population for QTL analysis of heavy metal accumulation and tolerance in *Noccaea caerulescens* 23

Chapter 3

Genetic analysis of the cadmium exposure response in a *Noccaea caerulescens* recombinant inbred line population 47

Chapter 4

Genetic variation in natural populations of *Noccaea caerulescens* in Europe 77

Chapter 5

An efficient flower dipping transformation method for the metal hyperaccumulator *Noccaea caerulescens* 113

Chapter 6

General Discussion 131

References 141

English summary 166

CHAPTER

General introduction

Metal contamination in soil

Soil is a non-renewable natural resource that forms the foundation for the life of numerous terrestrial organisms. Importantly, it supports ecosystems and human communities by serving as a medium for agricultural production and providing habitats for several animals and plants (Cassidy et al., 2013; Gao et al., 2022; Li et al., 2022). Unfortunately, the rapid expansion of human urbanization and advancements in industry have released a substantial amount of various kinds of hazardous and harmful contaminants into soils, including inorganic toxic metals. There are nine metals and metalloids commonly found in contaminated soils, including arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), mercury (Hg), nickel (Ni), lead (Pb), selenium (Se), and zinc (Zn) (Evanko and Dzombak, 1997; FAO and UNEP, 2021; Su, 2014). Four of these metals (Cu, Ni, Zn) are essential elements required in only small concentrations for proper plant growth, but high-level exposure can be toxic for plants. The rest are non-essential elements and highly toxic for plants even at low concentrations. In addition, these contaminated metals are non-biodegradable and persistent in nature (FAO and UNEP, 2021). Due to these properties, metal contamination has been recognized as a serious concern for soil resources and human health worldwide (European Commission, 2006; Gao et al., 2022; Sarwar et al., 2010).

These metals naturally originate from the mineralization of parental rocks and their erosion (Kabata-Pendias, 2004; Wuana and Okieimen, 2011), and are typically found at non-toxic concentrations (below 1000 ppm), as shown in Table 1. Moreover, volcanic eruptions and hot spring activities are reported to contribute to metal contamination in soils (Scott et al., 1996). An excess amount of metals can be observed in naturally metal-enriched sites owing to the presence of bedrock containing high levels of metals. These sites are rare to find and are often located remotely from industrial areas and traffic routes. For instance, Pb-enriched soils are found in National Parks in Switzerland located at high altitudes. High concentrations of Pb, Zn, and Cd are detected in soils in the Mendip region in the UK. Substantial levels of As are found in soils in Bangladesh, Pakistan, and India as a result of the mineralization of As-containing bedrock (Khalid et al., 2017). The metal-enriched sites are distributed patchily in Central Europe, the Pyrenees, the Alps (Baker et al., 2010; and ref. therein), Central Africa and New Caledonia (Ernst, 2006; and ref. therein).

Besides, soils are enriched with metals predominantly through various anthropogenic activities. Globally, mining and smelting activities are the major sources of discharging metals (Cd, Pb, As, Hg) into soils. The long-term use of fertilizers and chemicals (e.g., pesticides, fungicides, herbicides) to sustain crop production contributes significantly to the soil contamination of metals (As, Cd, Pb, Se, Zn), especially in developing countries (Atafar et al., 2010; Rashid et al., 2023; Srivastava et al., 2022). Moreover, industrial activities (e.g., electroplating, textile and petrochemical, industrial waste disposal of municipal wastes) and fuel combustion also deposit large amounts of metals into soils (Hou et al., 2020; Islam et al., 2021; Khan et al., 2008; Su, 2014)

A recent study has reported that there are around five million metal-contaminated sites worldwide, showing metal concentrations exceeding regulatory levels (Liu et al., 2018). In the USA, approximately 600,000 hectares (mostly in brownfield sites) are enriched with metals (Khalid et al., 2017). In Europe, data from the LUCAS Topsoil Survey indicates that around 6% of arable lands (13.7 million hectares) are potentially contaminated with metals and require local assessment (Tóth et al., 2016). In China, 80 million hectares of land (~16% of arable lands) show elevated levels of metals and are considered metal-contaminated areas (He et al., 2015). Based on the chemical composition, the metal-enriched soils from both natural and anthropogenic origins are categorized as calamine, ultramafic, and seleniferous soils. Calamine soils are mainly enriched with Cd, Pb, and Zn, sometimes Cu and sulfur. Ultramafic, also called serpentine,

predominantly consists of elevated levels of Ni, Cr, and Cobalt (Co). High concentrations of iron (Fe) and magnesium (Mg) are found in ultramafic soils as well. Seleniferous soils are characterized by abnormal levels of Cu and Co (Reeves, 2006).

Table 1. *Metal concentration reported in non-metallicolous and metal-enriched soils adapted from* (Kabata-Pendias and Mukherjee, 2007; and ref. therein)

Metal	Concentration range in non-metallicolous soils (mg kg^{-1})	Maximum concentration reported in metal-enriched soils (mg kg^{-1})
Arsenic (As)	0.5 - 2.5	20,000
Cadmium (Cd)	0.1 - 0.2	1,781
Chromium (Cr)	126 - 185	100,000
Copper (Cu)	25 - 27	5,241
Mercury (Hg)	0.02 - 0.06	17,100
Nickel (Ni)	19 - 22	26,000
Lead (Pb)	15	20,000
Selenium (Se)	0.05 - 0.5	360
Zinc (Zn)	10 - 300	10547

These metals in soils have deleterious effects on human health and plant growth. In plants, metals are acquired via roots and subsequently translocated to the shoot through the xylem transport system. The metals are able to reduce plant growth through four main mechanisms: a) competition with nutrient ions with similar properties to metals for root adsorption; b) interaction with sulfhydryl and carboxylic groups on proteins that cause negative impacts on structure and activities; c) induction of oxidative stress via the production of reactive oxygen species (ROS); and d) substitution of crucial cofactor ions in enzymes and signaling proteins that inhibit enzymatic activities and interfere with gene expression (Dalcorso et al., 2013; Krämer, 2010). The detrimental effects of metals on plants are described in Table 2. These metals can accumulate in the edible parts of plants and subsequently enter the food chain. For instance, As accumulation has been reported in rice grains from south and southeast Asia (Srivastava et al., 2022). High Cd concentration has been detected in carrots above the safety threshold (Sarwar et al., 2017). Several vegetables such as basil, turmeric, lemongrass, and ginger accumulate elevated levels of Pb and Cd (Jolly et al., 2013).

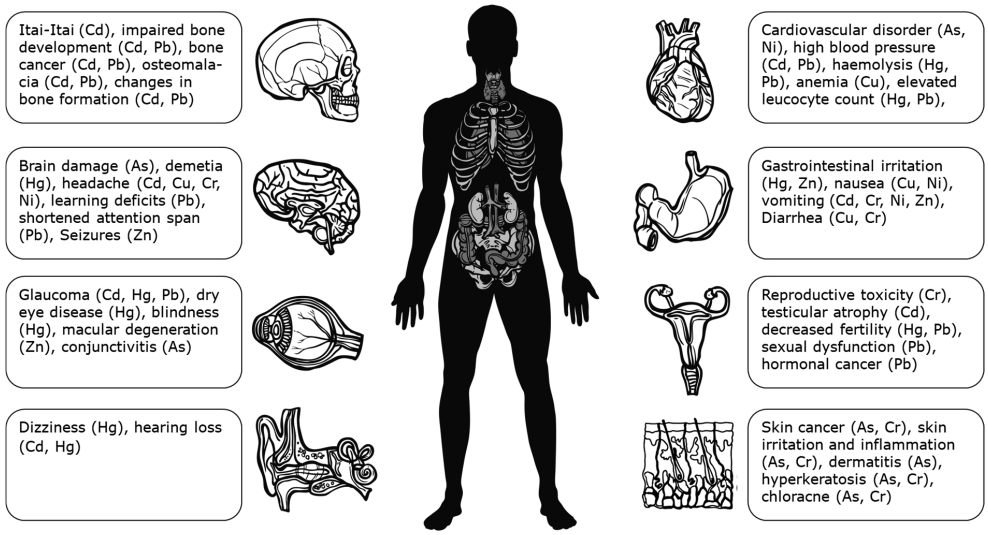
Table 2. *Harmful effects of different metals on plants* (Dalcorso et al., 2013).

Detrimental effects	Metals and their salt/ionic forms
Competition for root uptake from soils	As (arsenate, AsO_4^{3-}), Se (selenate, SeO_4^{2-}), Cr (chromate, CrO_4^{2-})
Reaction with sulfhydryl groups (-SH) and carboxyl groups (-COOH)	As, Cd, Hg, Pb
ROS generation and oxidative stress induction	Cd, Cu, Hg, Zn, Pb,
Substitution of crucial cofactor ions in enzymes and signalling proteins	As (arsenate, AsO_4^{3-}), Cu, Cr (Cr(III), Cr(VI)), Ni, Zn

In humans, these metals enter the human body mainly through the consumption of food contaminated with metals. The minor pathways are dermal contact and inhalation (Scutarașu and Trincă, 2023; Tsai and Lee, 2013; Xiong et al., 2016). Several health issues are associated with metal toxicity that depend on the type of metal and exposure concentration. The well-known disorder caused by long exposure to metal is Itai-itai disease. This disease was first reported in farming areas contaminated with high levels of Cd in

Japan. Patients suffer from severe bone pain and express pain continuously as 'Ouch.' This term later became the name for this disease, which is categorized as one of the harmful effects of Cd toxicity (Aoshima, 2016; Chang et al., 1996; Friberg, 2018). The harmful effects of metals are described in Figure 1.

Figure 1. *Harmful effects of metals on human health adapted from* (Ayangbenro and Babalola, 2017; FAO and UNEP, 2021).



The detoxification of contaminated toxic metals can significantly alleviate the risk to human health and restore contaminated lands for agricultural production. Many remediation approaches have been developed to clean up contaminated soils by either removing contaminants or reducing the bioavailability of contaminants (Rajendran et al., 2022). Remediation approaches can be classified based on the nature of the approach as physical, chemical, and biological; or based on the movement of contaminants as *in-situ* and *ex-situ* (Gomes et al., 2013).

Physical remediation is suitable for applying to vast contaminated areas with high levels of metal. Chemical remediation can clean up contaminated soils rapidly and effectively, yet it strongly depends on soil types and metals. Both techniques have been proven to efficiently detoxify metals in soils, but these techniques are costly (up to billions of dollars), have temporary effects, and sometimes even result in the deposition of new contaminants into soils (Khalid et al., 2017; Manea et al., 2013).

Biological remediation using plants, namely phytoremediation, has been proposed as an alternative technique that is considered ecologically and environmentally friendly, viable, and cost-effective for cleaning up contaminated soil (Chaney et al., 1997; Fasani et al., 2018; Hou et al., 2020; Wójcik et al., 2017). Several cases of phytoremediation have been conducted on actual metal-contaminated sites across the globe; for a comprehensive review, refer to (Fasani et al., 2018; Hou et al., 2020; Pandey and Bajpai, 2019; Wang et al., 2021). A recent study has demonstrated that phytoremediation costs are lower than other remediation techniques with the possibility for cost reduction (Wan et al., 2016). The development of phytoremediation requires an understanding of the genetic basis underlying heavy metal accumulation

and tolerance in plants, especially those that survive on contaminated soil, called hyperaccumulators (Baker et al., 2010; Brooks, 1994; Rascio and Navari-Izzo, 2011; Wójcik et al., 2017).

Metal hyperaccumulation and hypertolerance in plants

Metal hyperaccumulators are plants exhibiting an extraordinary concentration of metal in the aerial part and withstanding elevated metal-enriched soils as well (Pollard et al., 2014; van der Ent et al., 2013). This phenomenon was first observed in *Alyssum bertolonii*, which survives on ultramafic soils and accumulates Ni content at 10,000 μg g^{-1} of dry biomass (Minguzzi and Vergnano, 1948). In the 1970s, the term "hyperaccumulator" was coined by Brooks et al., (1977) for eleven plant species showing Ni concentration in the shoot, regardless of their soil metal contents. To date, this term has been widely used in more than 700 studies and has also expanded to other metals with different thresholds (Table 3).

Several attempts to explain the evolutionary origin of metal hyperaccumulation in plants have been made since 1992 (Manara et al., 2020). The accepted hypothesis has been proposed by Boyd (2007, 2012), namely the defensive enhancement hypothesis. This hypothesis suggests that plants accumulate a substantial concentration of metals in the shoot to defend against herbivores and pathogens both directly (via metal toxicity) and indirectly (via signaling pathways) (Manara et al., 2020).

Table 3. *Metal hyperaccumulators of relevant metals and Rare Earth Elements adapted from* (Manara et al., 2020; Reeves et al., 2017).

Metals	Threshold to define hyperaccumulators (mg kg^{-1})	Family number	Species number	Major genera	Recorded accumulation range (mg kg^{-1})[a]
Arsenic (As)	> 1,000	1	5	*Pteris*, *Pityrogramma*	8,350 - 23,000
Cadmium (Cd)	> 100	7	8	*Noccaea*, *Sedum*	154 - 938
Copper (Cu)	> 300	20	53	*Anisopappus, Crepidorhopalon, Glochidion, Phyllanthus, Persicaria*	300 - 2,278
Nickel (Ni)	> 1,000	54	532	*Alyssum, Buxus, Berkheya, Glochidion, Geissois, Homalium, Hybanthus, Phyllanthus, Leucocroton, Senecio, Xylosma*	1,000 - 66,660
Lead (Pb)	> 1,000	7	9	*Noccaea*	1,600 - 28,370
Rare earth elements	> 1,000	2	2	*Dicranopteris*	-
Selenium (Se)	> 100	7	41	*Astragalus, Stanleya*	-
Zinc (Zn)	> 3,000	9	21	*Arabidopsis, Noccaea, Sedum*	6,690 - 43,710

[a] *Metal accumulation ranges are based on the record of hyperaccumulators from natural habitats* (Xu et al., 2020).

These hyperaccumulators can accumulate elevated concentrations of metals in their aerial tissues, achieved by the enhancement of physiological processes, including metal uptake from the soil to the root, xylem loading, and root-to-shoot translocation. These metals are then stored in the aerial tissues, mostly in leaves. To keep toxic metals away from the photosynthetic machinery, these plants employ two mechanisms with enhancement: vacuolar sequestration or intracellular chelation (Verbruggen et al., 2009). With a great ability for metal detoxification, these hyperaccumulators can tolerate high levels of

metals in soils, a phenomenon frequently called metal hypertolerance (Baker, 1987). Another group of plants endemic to metal-enriched soil displays metal hypertolerance but does not accumulate metals in above-ground tissues. These plants are called hypertolerant non-hyperaccumulators or excluders (Manara et al., 2020; Verbruggen et al., 2009). The excluders can withstand high levels of metals by keeping metals out of their roots and preventing root-to-shoot translocation of metals (Verbruggen et al., 2009). Both metal hyperaccumulation and hypertolerance involve several families of transporters (Table 4).

Metal hyperaccumulation is a rare trait in the plant kingdom. A small fraction (approximately 0.2%) of known vascular plants have been reported for this extraordinary trait. To date, a total of 791 plant species, belonging to 52 plant families, are recognised as hyperaccumulators worldwide, as described in Table 4. In general, the hyperaccumulation trait is widely reported in a broad range of unrelated plant families. A large proportion of hyperaccumulators belongs to Brassicaceae family (104 species) and Phyllanthaceae family (130 species). With regard to metals, Ni hyperaccumulators are the most abundant (532 species), such as different Alyssum species from the Mediterranean area, *Phyllanthus pallidus* discovered on ultramafic soils in Cuba, *Pycnandra acuminata* (blue sap tree) from New Caledonia, and *Phyllanthus balgooyi* from Malaysia (Jaffré et al., 1976; Manara et al., 2020; Reeves et al., 1999; van der Ent and Mulligan, 2015).

Zn hyperaccumulators are reported in the Brassicaceae family, including *Arabidopsis halleri*, *Noccaea caerulescens*, and other Noccaea species, and in the Crassulaceae family, e.g., *Sedum alfredii* (Bert et al., 2000; Küpper et al., 2000; Peer et al., 2003; Yang et al., 2006). Noticeably, some Zn hyperaccumulators (e.g., *N. caerulescens*) are able to accumulate Cd and Pb due to their similar chemical properties (Krämer, 2010). For the hyperaccumulation of the non-essential element Cd, a few plant species are exhibiting this trait, including *A. halleri*, *N. caerulescens*, *Noccaea praecox*, *Viola baoshanensis*, *Solanum nigrum*, *S. alfredii*, and *Phytolacca americana* (Manara et al., 2020). These hyperaccumulators are mainly discovered using elemental analysis of leaf samples, primarily Atomic Absorption Spectrophotometry or Inductively Coupled Plasma-Atomic Emission Spectroscopy. Both methods are time-consuming, destructive, and use harmful chemicals and flammable gas (Wilschefski and Baxter, 2019). The recent development of X-ray fluorescence spectroscopy (XRF) has allowed non-destructive and high-throughput elemental analysis for both living samples and herbarium specimens (Invernón et al., 2021; van der Ent et al., 2022). This new analytical approach facilitates and speeds up the discovery of hyperaccumulators, thus the number is expected to increase in the near future (Jaffré et al., 1976; Manara et al., 2020; Reeves et al., 1999; van der Ent and Mulligan, 2015).

Table 4. *Reported genes involved in metal hyperaccumulation and hypertolerance adpated from* (Chen et al., 2018; De Abreu-Neto et al., 2013; Seregin and Kozhevnikova, 2023; Skuza et al., 2022; Zhang et al., 2020).

Gene family [a]	Description	Example genes from different plant species		
		Gene	Plant	Metal [b]
CAX family	Calcium/cation antiporters play a role in metal detoxification.	*AtCAX2, AtCAX4*	*Arabidopsis thaliana*	Cd
		OsCAX	*Oryza sativa*	Cd, Mn
CDF family	Divalent cation transporters contribute to metal vacuolar sequestration and endomembrane transport.	*AtMTP*	*Arabidopsis thaliana*	Mn, Zn
		ShMTP	*Stylosanthes hamata*	Mn
		OsMTP11	*Oryza sativa*	Mn
HIPP family	Metal-binding metallochaperones are important for metal homeostasis and detoxification.	*AtHIPPs*	*Arabidopsis thaliana*	Cd
		OsHIPPs	*Oryza sativa*	Cd
		TaHIPPs	*Triticum aestivum*	Cd
		SlHIPPs	*Solanum lycopersicum*	Cd
		CsHIPPs	*Camellia sinensis*	Cd
HMA family	Monovalent/divalent cation transporters contribute to xylem loading, vacuolar sequestration of metals, and metal remobilisation.	*NcHMA4*	*Noccaea caerulescens*	Cd
		AhHMA	*Arabidopsis halleri*	Zn
		AtHMA2, AtHMA4	*Arabidopsis thaliana*	Cd, Zn
		OsHMA2	*Oryza sativa*	Cd, Zn
NAS family	Genes encode metal chelators that facilitate metal transport, detoxification and sequestration.	*NcNAS2, NcNAS4*	*Noccaea caerulescens*	Cd, Fe, Ni, Zn
		AhNASs	*Arabidopsis halleri*	Fe, Zn
		AtNAS1, AtNAS4	*Arabidopsis thaliana*	Cu, Fe, Zn
		SaNASs	*Sedum alfredii*	Cd, Zn
NRAMP family	Divalent cation transporters involve metal remobilisation.	*AtNRAMPs*	*Arabidopsis thaliana*	Cd, Cu, Fe, Mn
		OsNRAMP3, OsNRAMP5	*Oryza sativa*	Cu, Mn
		MtNRAMP1	*Medicago truncatula*	Fe
YSL family	Metal-nicotianamide complexes transporters play a role in xylem loading and long-distance transport.	*NcYSL3*	*Noccaea caerulescens*	Fe, Ni
		AhYSLs	*Arabidopsis halleri*	Cu
		AtYSLs	*Arabidopsis thaliana*	Cu, Fe
		OsYSLs	*Oryza sativa*	Cu, Fe , Mn
		ZmYSL1, ZmYSL2	*Zea mays*	Cu, Cd, Fe, Ni, Zn
ZIP family	Divalent cation transporters are involved in cellular metal uptake and xylem loading.	*NcZNT1*	*Noccaea caerulescens*	Zn
		AhIRT3	*Arabidopsis halleri*	Fe
		AtZIPs, AtIRTs	*Arabidopsis thaliana*	Cd, Cu, Fe, Mn, Zn
		MtZIPs	*Medicago truncatula*	Fe, Mn

[a] *CAX: Cation exchange, CDF: Cation Diffusion Facilitator, HIPP: Heavy metal-associated Isoprenylated Plant Proteins, HMA: P1B-ATPase subfamily Heavy Metal Associated, NAS: Nicotianamine Synthase, NRAMP: Natural Resistance Associated Macrophage Proteins, YSL: Yellow Stripe 1-Like, ZIP: Zinc-regulated, Iron regulated transporter-like Proteins;* [b] *Fe: Iron, Mn: Manganese*

Since the discovery of the Ni hyperaccumulator *A. bertolonii*, several studies have investigated different hyperaccumulators to understand the genetics and physiology underlying metal hyperaccumulation and hypertolerance, as well as the evolutionary perspective (Hanikenne and Nouet, 2011). This knowledge can be applied to developing phytoremediation technology, as described above, and also two additional technologies: biofortification, which refers to the increasing level of essential metals (particularly Zn) in

crop plants, especially for cereal grain crops (Palmgren et al., 2008); and phytomining, which involves the extraction of profitable metals (especially Ni) from contaminated sites or naturally metal-enriched soil using plants (Brooks et al., 1998; Chaney et al., 2004; Ent et al., 2015). For a few decades, two particular hyperaccumulators, *A. halleri* (Macnair et al., 1999) and *N. caerulescens* (Assunção et al., 2003a; Milner and Kochian, 2008), have been intensively studied as model species for hyperaccumulation because of their close relatedness to *Arabidopsis thaliana*, a model system for plant study (Koornneef and Meinke, 2010). There is substantial available genomic, transcriptomic, and proteomic information in *A. thaliana* that has been transferred to *A. halleri* and *N. caerulescens*, and used for dissecting the molecular mechanisms of metal hyperaccumulation and hypertolerance (Krämer, 2010).

Noccaea caerulescens, a model species for metal hyperaccumulation and hypertolerance

Noccaea caerulescens J. and C. Presl. is a Zn/Ni/Cd/Pb hyperaccumulator belonging to the Brassicaceae family (Assunção et al., 2003b; Dinh et al., 2018; Escarré et al., 2013; Kozhevnikova et al., 2020; Milner and Kochian, 2008; Mohtadi et al., 2012; Sterckeman et al., 2017). Their occurrences are found dispersed across Europe, particularly in France, Scandinavia, and the United Kingdom, abundantly observed throughout these areas (Banásová et al., 2006; M. Koch et al., 1998; Krämer, 2010; Sterckeman et al., 2017). Recent studies also reported the occurrence of *N. caerulescens* in western Russia (Kravchenko et al., 2021) and Scandinavia (Hilmo et al., 2021; Møller et al., 2020; Pagad and Wong, 2020), recognized as an invasive species.

Regarding soil types, *N. caerulescens* is endemic to metal-enriched soils (ultramafic and calamine soils), as well as non-metallicolous or normal soils (Assunção et al., 2003b; Gonneau et al., 2017; Sterckeman et al., 2017), as showed in Figure 2, which is used to define ecotypes for *N. caerulescens* in several studies. Originally, *N. caerulescens* was first discovered as a Zn hyperaccumulator in 1865 when its scientific name was *Thlaspi calaminare* (Risse in Sachs, 1865). The local population occurring on calamine soils situated close to the border between Germany and Belgium displays an elevated level of Zn in above-ground tissues (17% of dry biomass). Subsequently, several studies have reported that *N. caerulescens* is able to accumulate Cd (Baker, 1989), Ni Reeves and Brooks, 1983), and Pb (Shimwell and Laurie, 1972), since it was called *Thlaspi alpestre*. Several studies have hypothesized that Zn hyperaccumulation is reported to be a species-wide trait in *N. caerulescens*, whereas Cd, Ni, and Pb hyperaccumulation are rather population-specific traits. This hypothesis, excluding Pb hyperaccumulation, has been confirmed through accumulation tests of two collections *of N. caerulescens* local populations (Kozhevnikova et al., 2020; Sterckeman et al., 2017). t is noticeable that calamine populations are distinct from the other two ecotypes (ultramafic and non-metallicolous populations) for metal hyperaccumulation capacities. For example, the calamine ecotype exhibits higher Cd accumulation in the shoot than the other two ecotypes, up to 4 to 4.3 times (Sterckeman et al., 2017).

Regarding metal hypertolerance, sequential tolerance tests have reported that Ni and Zn hypertolerance are reported as species-wide traits in *N. caerulescens*, with the calamine ecotype exhibiting higher Zn tolerance than ultramafic and non-metallicolous ecotypes (Kozhevnikova et al., 2020). Unlike Zn and Ni, Cd hypertolerance is an exclusive trait of the calamine ecotype (Krämer, 2010). Different metal excluder populations have been studied as well, which often exhibit a high degree of hypertolerance to metals, regarding their soil metal content (Kozhevnikova et al., 2020).

N. caerulescens is characterized as a small compact rosette plant that reproduces through self-pollination and cross-pollination between different compatible populations (Besnard, 2009; Peer et al., 2003). Early studies considered *N. caerulescens* a biennial species, requiring two winter seasons to break seed dormancy in the first winter and promote flowering in the second winter. Currently, *N. caerulescens* is able to complete its life cycle within one winter season as an annual under frost-free greenhouse conditions (Figure 2), requiring at least 10 weeks of cold treatment to induce flowering (Wang et al., 2020). Its flowers are raceme-type, consisting of 15 – 90 flowers, and its fruit shape is obcordate, containing 4 – 6 seeds per locule (Al-Shehbaz, 2014), resulting in a relatively high seed yield per plant.

Regarding the genome of *N. caerulescens*, there are seven pairs of chromosomes, making it a diploid species (2n = 14) (Mandáková et al., 2015), with a total size of 312 megabases, based on a calamine *N. caerulescens* accession from Ganges, France accession from Ganges, France (Severing, Wang, van den Heuvel, Aarts et al., unpublished). Considering these practical characteristics together makes *N. caerulescens* a prominent model species for metal hyperaccumulation and hypertolerance, especially for Cd, which is an extreme trait among hyperaccumulators (Assunção et al., 2003a; Milner & Kochian, 2008). Additionally, the genetic analyses of *N. caerulescens* can make use of a wide range of genetic tools and resources, particularly the genome database of *A. thaliana*, due to the high similarity (88%) of the protein-coding region between these two species (Rigola et al., 2006).

Figure 2. *N. caerulescens growing in nature and greenhouse conditions. (A) A well-studied calamine population occurs in La Calamine, Belgium, situated in the historical Zn mining area of Europe. (B) and (C) Non-metallicolous populations grow on roadside and dirt roads of Swiss Alps in Fontana, Switzerland. (D) A population is endemic to the Nieuwe Waterweg riverbank of Rozenburg, The Netherlands, characterized as calamine soils (Henk Schat, personal communication). E) Inbred N. caerulescens lines are propagated on peat-mixed soil pots in Nergena frost-free greenhouse at Wageningen University (The Netherlands).*

According to the Scopus citation database, *N. caerulescens* has been the experimental subject in more than 400 articles during the past five decades. These studies have developed various forward and reverse genetic tools for *N. caerulescens*, including inter-ecotypic segregation populations (Assunção et al., 2006; Deniau et al., 2006; Frérot et al., 2003; Richau and Schat, 2009; Zha et al., 2004); genetic maps (Assunção et al., 2006b; Deniau et al., 2006); genome assembly (at least one available on GenBank (Benson et al., 2013)); gene expression profiles and transcriptomics (Halimaa et al., 2014; Hammond et al., 2006; Milner et al., 2014; Mortel et al., 2008); proteomics (Schneider et al., 2013; Tuomainen et al., 2010); genetic transformation protocols (Blande et al., 2017; Guan et al., 2008; Y. F. Lin et al., 2016; Peer et al., 2006), generation of early-flowering lines (Lochlainn et al., 2011; Wang et al., 2020); Targeting Induced Local Lesions In Genomes (TILLING) population (Wang et al., 2022); and collections of different *N. caerulescens* local populations (Besnard, 2009; Gonneau et al., 2017; M. Koch et al., 1998; van der Zee, Corzo Remigio, et al., 2021).

The studies on *N. caerulescens* have indeed identified several genes and quantitative trait loci (QTL) associated with metal hyperaccumulation and hypertolerance. *NcZNT1,* a member of ZIP family, a plasma membrane transporter is involved in Cd and Zn hyperaccumulation by enhancing xylem loading of Cd and Zn (Y. F. Lin et al., 2016). Another member of ZIP family, *NcIRT1*, is involved in Cd accumulation in Cd hyperaccumulating *N. caerulescens* accession, but its function is still under investigation (Y. F. Lin and Aarts, 2012; Lombi et al., 2002). *NcZTP1 or NcMTP1* is a tonoplast-localized transporter, belonging to CDF family, responsible for Zn vacuolar sequestration (Küpper and Kochian, 2010). Two members of HMA family are studied in *N. caerulescens* including *NcHMA3* and *NcHMA4*. *NcHMA3* plays a crucial role in Cd sequestration into the vacuole in leaf cells (Ueno et al., 2011), while *NcHMA4* is a Cd/Zn efflux transporter responsible for Cd root-to-shoot translocation in root cells and Zn/Cd remobilization in above-ground tissues (Craciun et al., 2012; Mishra et al., 2017). Three transporters of MT family (*NcMT1*, *NcMT2*, and *NcMT3*) are highly expressed in *N. caerulescens*, conferring enhanced tolerance to metals such as Cd, Cu, and Zn (Roosens, et al., 2005; Roosens, et al., 2005). Three NAS genes including *NcNAS1*, *NcNAS2* and *NcNAS3*, are highly expressed in *N. caerulescens*, which is likely to enhance the production of Nicotinamide (NA) (Hammond et al., 2006). The NA can form complexes with metals, implying metal detoxification. Besides, the metal-NA complex can be transported across cells by YSL transporters (Mari et al., 2006). Elevated expression of YSLs (*NcYSL3*, *NcYSL5*, and *NcYSL7*) are observed, which contribute to long-distance translocation of Ni in Ni-NA complex form, at least for *NcYSL3* (Milner and Kochian, 2008). NRAMP family plays an important role in metal hyperaccumulation and tolerance. Three members of NRAMP family are investigated in *N. caerulescens*, that includes *NcNRAMP1*, *NcNRAMP3*, and *NcNRAMP4*. NcNRAMP1 is involved in xylem loading and root-to-shoot translocation of Cd (Milner et al., 2014), while *NcNRAMP3* and *NcNRAMP4* are Cd efflux transporters located on tonoplast, conferring Cd hypertolerance in *N. caerulescens* (Oomen et al., 2009). Although previous studies have identified QTLs from inter-ecotypic bi-parental populations, the candidate genes directly associated with metal hyperaccumulation and hypertolerance have not been established yet.

Despite the advantages of studying *N. caerulescens* as model species, there are a number of drawbacks to consider. In *A. thaliana*, a wide range of genetic tools enables efficient validation of candidate genes and identification of their functions, such as collection of insertional or ethyl methanesulfonate (EMS) mutants for most of *A. thaliana* genes, high-resolution mapping populations, loss-of-function mutant generation using CRISPR-Cas9 or artificial microRNAs, and public genome databases (e.g. TAIR database) (Alonso-Blanco et al., 2006; Cantó-Pastor et al., 2021; Lamesch et al., 2010). Few mapping populations have been developed in *N. caerulescens* for identification of QTLs associated with metal hyperaccumulation

and hypertolerance. However, it is less feasible for *N. caerulescens* than for *A. thaliana* due to its longer life cycle and larger plant size. For example, an F_2 mapping population can be developed within half a year for *A. thaliana*, whereas it takes at least two years for *N. caerulescens*. The *A. thaliana* genome database (e.g., TAIR) has often been used for identifying the function of candidate genes in *N. caerulescens* (Assunção et al., 2003a; Krämer, 2010). It would be a great resource for *N. caerulescens* to have its public genome database and predicted annotation. There is only one scaffold-level genome assembly (GCA_900406465.1) based on whole-genome shotgun sequencing, available on the NCBI repository. Its use is relatively challenging due to a substantial number of scaffolds (~19,000). Moreover, reverse genetic tools (e.g., genetic transformation protocol) are limited and less efficient in *N. caerulescens*, which has hampered the identification of gene function in this species. An attempt to generate TILLING lines, comparable to EMS lines, was recently made (Wang et al., 2022), offering potential utility for studying metal hyperaccumulation and hypertolerance. Noticeably, the cost of "-omics" technologies has been reduced over the past few years, which makes these technologies, especially whole-genome sequencing, affordable and extremely useful tools for *N. caerulescens* and non-model organisms.

The scope of this thesis

The objective of this thesis is to develop forward and reverse genetic tools, using to-date NGS technology, to overcome the limitations of the metal hyperaccumulator plant species *N. caerulescens*. These genetic tools are employed to uncover the molecular basis of metal hyperaccumulation and hypertolerance. I delve further into exploring population diversity at the species level in *N. caerulescens* and trying to understand the evolutionary dynamics shaping metal accumulation and tolerance under ecological and evolutionary pressures in this species.

Chapter 2 describes the generation of a recombinant inbred line (RIL) population derived from parental accessions Lellingen (LE) and Ganges (GA), exhibiting contrasting Zn/Cd accumulation capacities and tolerance. Whole-genome sequencing provides genotypic information used to construct a genetic map. The validation of this population is described through QTL analyses of different life history traits, including the well-studied trait of flowering time.

After demonstrating in **Chapter 2**, the LE x GA RIL population is immortal population and feasible for further QTL mapping of metal-related traits, **Chapter 3** delves into the genetic basis of Cd response. QTL mapping is employed for multiple traits derived from three distinct experiments (Cd accumulation test, Cd tolerance test, and growth response to Cd). A comprehensive set of syntenic orthologues is generated from whole genome alignment between *A. thaliana* and *N. caerulescens*. Strong association between genotype and Cd accumulation is observed, with potential candidate genes being discussed.

In **Chapter 4**, a diversity panel of N. caerulescens accessions from a large geographic area of Europe is generated. Leveraging the NGS approach, whole-genome genetic variation is obtained for this panel, enabling the assessment of genetic relationships among each accession. Clustering analysis provides evidence on the evolutionary history of this species and the population structure within this panel. In addition, I attempt to identify genetic variation associated with flowering time and Zn/Ni accumulation through the first genome-wide association study (GWAS) performed in this species.

Chapter 5 describes the development of a versatile stable transformation system *for N. caerulescens*, which combines high fecundity early flowering plants with a simple *Agrobacterium tumefaciens* transformation protocol. The resulting system is anticipated to function as an efficient reverse genetic

tool for *N. caerulescens*, laying the groundwork for the future development of advanced techniques, particularly genome editing methods.

The main findings described in each experimental chapter and the future perspectives for *N. caerulescens* research are discussed in **Chapter 6**.

CHAPTER

2

Development of an immortal bi-parental mapping population for QTL analysis of heavy metal accumulation and tolerance in *Noccaea caerulescens*

Jitpanu Yamjabok, Joost van den Heuvel, Henk Schat, Mark G.M. Aarts

Laboratory of Genetics, Wageningen University and Research,
Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

Financial source:
This research is financial supported by Royal Thai government (Thailand).

Abstract

Quantitative trait loci (QTL) and Candidate genes of metal-related traits identified from previous studies in *Noccaea caerulescens* have all been generated segregating F_2 and F_3 populations. This makes it cumbersome to validate and compare QTLs, as due to segregation, different genotypes have to be used for each experiment. In order to generate a segregating population with fixed genotypes, we generated a F_7 recombinant inbred line (RIL) population for *N. caerulescens*, from the cross between accessions Ganges (GA) and Lellingen (LE). GA is a calamine accession, with high zinc (Zn)/cadmium (Cd) accumulation capacity and high Zn/Cd tolerance. LE is a non-metallicolous accession, with high Zn/Cd accumulation capacity and moderate Zn/Cd tolerance. This population was genotyped and used for QTL mapping of six life history traits. A genetic map was constructed based on 443 SNP markers on 232 F_7 RILs, which identified seven linkage groups (LG) spanning over 820 cM. A total of seven QTLs were identified using an interval mapping model. Of these, shared QTLs associated with flowering time (qFT2019 and qFT2020-1) were detected in two consecutive years, explaining 14.95 and 16.91 % of phenotypic variance with LOD scores of 7.18 and 8.40. Novel candidates genes were identified for flowering time QTLs. Based on these findings we conclude that this RIL population is large enough and sufficiently densely genotyped, covering all seven *N. caerulescens* chromosomes in a representative and balanced way, suitable for QTL mapping. The RIL population will be a useful genetic tool for deciphering QTLs underlying metal accumulation and tolerance in *N. caerulescens* in the future.

Introduction

Soil contamination has been recognised as one of major threats to soil security worldwide (Carré et al., 2017). Petroleum hydrocarbons, pesticides, organic solvents and metals are the most common chemical soil contaminants that are introduced into soil via various anthropogenic activities such as mining and smelting (Ashraf et al., 2014). In Europe, four non-nutritional, toxic metal(loid)s: Arsenic (Ar), Cadmium (Cd), Lead (Pb) and Mercury (Hg), have been found at elevated levels in the soil (Tóth et al., 2016), posing a serious threat to human health and to food production (Montanarella et al., 2015). Consumption of food contaminated with non-nutritional metal can cause several severe toxicity symptoms in humans. For instance, the long-term consumption of Pb- and Cd-contaminated crops can cause upper gastrointestinal cancer (Järup, 2003). In plants, metals can inhibit fundamental physiological processes by enhancing the production of reactive oxygen species (ROS) and competing with essential metals (Rascio and Navari-Izzo, 2011). The increase of ROS induces oxidative stress to cells, leading to macromolecule degradation, lipid peroxidation, DNA-strand breakup and, consequently cell death (Kumar, 2015; Y. F. Lin and Aarts, 2012). A combination of these effects leads to reduced yields in crops, which has an impact on food security. The removal of contaminating toxic metals from soil can significantly alleviate the risk to human health and restore contaminated land into arable land for growing crops. Conventional remediation approaches (physical and chemical techniques) can be carried out to clean-up soil, which are found to be highly efficient (Khalid et al., 2017). Unfortunately, these techniques are often not feasible due to the high-costs, inefficiency for low level contaminants and irreversible alteration of remediated soil properties (Yan et al., 2020). Phyto-remediation has been proposed as an ecologically and environmentally friendly, viable, and cost-effective option to clean-up contaminated soil (Brooks et al., 1998; R. L. Chaney and Baklanov, 2017). Phyto-remediation is a method using plants to either remove contaminants or reduce bioavailability of contaminants (Berti and Cunningham, 2000). The efficient development of phyto-remediation technology requires a thorough understanding of the genetic basis underlying metal accumulation and tolerance in plants, especially of

those metallophyte species that hyperaccumulate metal (Chaney and Baklanov, 2017; Fasani et al., 2018).

Metal hyperaccumulation was first described for *Alyssum bertolonii*. This species can survive on Nickel (Ni)-enriched soil and exhibits elevated Ni concentrations at 1% in dry biomass (Minguzzi and Vergnano, 1948). To date, 721 plant species from 52 families are described as hyperaccumulator, most of these are Ni hyperaccumulators and some species exhibit accumulation of more than one element (Reeves et al., 2017). *Noccaea caerulescens* J. and C. Presl. (previously known as *Thlaspi caerulescens*) is a diploid (2n = 14) zinc (Zn)/Ni/Cd/Pb hyperaccumulator from the *Brassicaceae* family. Different accessions of *N. caerulescens* show different metal preferences for accumulation and tolerance, which is rare among hyperaccumulators (Assunção et al., 2003b, 2006; Deniau et al., 2006; Escarré et al., 2000; Frérot et al., 2003, 2005; Lombi et al., 2000; Richau and Schat, 2009; Sterckeman et al., 2017). The metal preference is expected to be comparable in *N. caerulescens* accessions originating from the same edaphic types, that are divided into three groups or ecotypes: calamine (from Zn/Cd/Pb-enriched soil), ultramafic (previously called serpentine, from Ni-enriched soil) and non-metallicolous (from normal soil). The early studies sometimes define calamine and ultramafic as the metallicolous ecotype. This unique intraspecific variation makes bi-parental segregating populations of *N. caerulescens* a valuable genetic resource for quantitative trait locus (QTL) analysis (Pollard et al., 2014). *N. caerulescens* shares 88% DNA sequence similarity in coding regions with *Arabidopsis thaliana* (Rigola et al., 2006), the well-studied plant model species with substantial available genomic information (Koornneef and Meinke, 2010). This permits the identification of the function of *N. caerulescens* genes based on the function of the *A. thaliana* homologue located in the same syntenic region (Assunção et al., 2003a). In addition *Arabidopsis halleri* is another Zn/Cd hyperaccumulator closely related to *A. thaliana*, albeit less versatile with respect to Ni and Pb accumulation and tolerance (Peer et al., 2003). With these advantages, *N. caerulescens* and *A. halleri* have been proposed as plant model species for metal hyperaccumulation and tolerance (Assunção et al., 2003a; Krämer, 2010; Milner and Kochian, 2008a; Pauwels et al., 2012).

Previous studies have reported on the development of bi-parental segregation populations QTL analysis of metal accumulation and tolerance traits in *N. caerulescens* (Assunção et al., 2003a, 2006; Deniau et al., 2006; Escarré et al., 2000; Frérot et al., 2003, 2005; Lombi et al., 2000; Richau and Schat, 2009; Zha et al., 2004). These bi-parental populations were often used for co-segregation analysis, which was used to decipher the pre-liminary genetic basis (e,g, monogenic or polygenic) of metal-related traits and relationships between those traits (Frérot et al., 2003, 2005; Richau and Schat, 2009; Zha et al., 2004). To specifically identify QTLs behind metal-related traits, analyses were conducted in different F_2 and F_3 mapping populations which revealed ten QTLs associated with shoot and root Zn/Cd accumulation, explaining 9.6% - 54.4% of phenotypic variance (Assunção et al., 2006; Deniau et al., 2006). While it is relatively easy and straightforward to generate bi-parental F_2 and F_3 populations, the F_2 plants or F_3 lines are still largely heterozygous and as such replicate or additional experiments testing the same genetic material again or in different environments is not possible, or very challenging at least. Instead, an immortal bi-parental population would be needed, either based on homozygous doubled haploid lines (F. Han et al., 1997) or recombinant inbred lines (Jansen, 2004). Such an 'immortal' population can be genotyped once and thereafter used for phenotyping different traits under different environments for multiple times. It can be easily propagated and maintained through seeds, as the progeny will have the same genotype as the previous, genotyped, generation. If available, it will become a valuable genetic tool to dissect complex traits in *N. caerulescens*.

The aim of this study was to develop an immortal recombinant inbred line (RIL) population for phenotyping metal hyperaccumulation and tolerance traits in *N. caerulescens*. This RIL population was generated from the inter-ecotypic cross between the calamine accession GA and the non-metallicolous accession LE. The progeny of this cross is expected to segregate for these traits. It was further developed to a RIL population via single seed descent. This RIL population was genotyped for single nucleotide polymorphism (SNP) markers using the GA draft whole genome sequence as a reference. A set of representative SNPs was selected for linkage map construction. Flowering time, self-fertility and plant growth were traits measured for QTLs analysis. The QTLs associated with flowering time were investigated for co-location with a homologue of *A. thaliana* genes known to be involved in flowering initiation, based on syntenic blocks.

Methodology

Plant material and population development

A recombinant inbred line (RIL) population was generated through single seed descent from a cross between the non-metallicolous accession 'Lellingen' (LE), originating from Luxembourg (49°59'1.83"N 05°59'39.00"E), and the calamine accession 'Ganges' (GA), originating from the Cevennes, France (43° 56' 11.2'' N 03° 40' 17.2'' E). LE was used as the mother. This RIL population was propagated during September to June annually, to generate the F_7 generation. In total, it consists of 232 RILs.

To propagate the lines, seeds of each line were imbibed on filter papers moistened with modified half-strength Hoagland's solution containing 3 mM KNO_3, 2 mM $Ca(NO_3)_2.4H_2O$, 1 mM $NH_4H_2PO_4$, 0.5 mM $MgSO_4.7H_2O$, 1 µM KCl, 25 µM H_3BO_3, 2 uM $MnSO_4.4H_2O$, 2 µM $ZnSO_4.7H_2O$, 0.1 µM $CuSO_4.5H_2O$, 0.1 µM $(NH_4)_6Mo_7O_{24}.4H_2O$ and 20 uM Fe-EDDHA (N,N'-ethylenediamine-di(O-hydroxyphenylacetic acid). This nutrient solution was set at pH = 5.7, and buffered with 2 mM MES (2-[N-morpholino]-ethanesulfonic acid). To break dormancy, 10 µM Gibberellins 4 + 7 (Globachem, Belgium) was used and samples were stratified at 4 °C for three days. After radicle emergence, the seeds were sown on a moist peat-based substrate. The plants were grown in frost-free, insect-free greenhouse (min. 5 °C) at the Wageningen University campus (51°59'46.4"N 5°39'29.4"E), from September onwards, to seed harvest in June of the next year.

Plant phenotyping

This population was phenotyped for three traits: flowering time, in 2019 and 2020; self-fertility, in 2020; and plant growth (plant height and rosette diameter), in 2019. Flowering time was defined as the number of days between sowing in the greenhouse until the appearance of the first flower with an open anther. This to also include the few RILs of which the first flowers were apetalous. Plants were exposed to natural vernalisation in winter (except for protection against frost), which was needed to induce flowering. In February 2019 and 2020, this population was monitored three times and five times a week to record flowering time 2019 (FT2019) and 2020 (FT2020), respectively.

Self-fertility was defined as the ratio of the number of fertilized siliques to the total number of siliques. To determine this, the plants were allowed to self-pollinate in the greenhouse in April to May, 2020. In June 2020, self-fertility of primary (SFP) and the most developed secondary inflorescence (SFS) were recorded for one plant per RILs.

To determine plant growth, a photo was taken from 1 to 3 plants per RIL in the population in a fixed mini photo studio in April 2019. Plant height (PH) and rosette diameter (RD) was determined on these photos with ImageJ Fiji (Schindelin et al., 2012).

Whole genome sequencing and SNP calling

DNA was extracted from young inflorescences, collected from one plant of each RIL in the population and stored in liquid nitrogen. The inflorescences were ground in liquid nitrogen and transferred to a 96-deepwell plate. DNA extraction was employed using solid phase reversible immobilization (SPRI) beads method (see supplementary information). For DNA library preparation, the DNA sample was quantified using SYBR green (Invitrogen, USA) and diluted to 0.2 - 0.25 ng/µL for each sample. 2 µL diluted DNA was tagmented and adapter-ligated in a 3-µL reaction volume using an Illumina Nextera Kit (Illumina, USA). Libraries were amplified by Illumina TruSeq primers and Robust 2G enzyme (Kapa biosystems, Cape Town, South Africa). The libraries were selected for 300 – 500-bp fragments using SPRI beads-based size selection. The selected libraries were paired-end sequenced to an average depth of 5x whole-genome coverage by Novogene Europe (Cambridge, United Kingdom). The raw reads were demultiplexed and evaluated for quality by Novogene Europe. The QC-passed reads were processed through a SNP-calling pipeline. Briefly, the reads were trimmed for adapter sequence and filtered for low quality reads using Trimmomatic v. 0.32 (Bolger et al., 2014). The trimmed, filtered reads were aligned to the *N. caerulescens* 'Ganges' reference genome assembly (Aarts, unpublished data) with BWA (Li and Durbin, 2009) using the default parameter. SAMtools (Li et al., 2009) and BCFtools (Danecek et al., 2021) were used to generate an VCF file containing the SNP genotype data.

SNP filtering and linkage map construction

Raw SNP data were filtered to remove SNPs which are either found at lower than 5x coverage, or found at over 2 times the average sequencing depth per line. The filtered SNP genotypic data was converted into allelic marker format using the following code: "A" when representing the GA allele; "B" when representing the LE allele; and "H" when heterozygous; the so-called ABH format.

In case of missing data, loci missing genotype information for more than 50% of the RILs, and 22 RILs missing more than 20% of the markers, were excluded. leaving 210 RILs for mapping. Then, all but one of the loci with an identical ABH score for all RILs were excluded, as determined using findDupMarker in R/qtl (Broman et al., 2003). As this still resulted in a large number of markers, which would lead to unpractically long computer running times when constructing and using the linkage map. The genotypic data was selected for representative SNPs, well distributed across entire reference genome assembly (Aarts, unpublished data).

For the linkage map construction, loci were grouped into one linkage group with a minimum logarithm of the odds (LOD) score of 12 using the Kosambi mapping function (Kosambi, 1943) of Joinmap version 4.0 (Van Ooijen, 2006). MapChart was used to draw the corresponding linkage map (Voorrips, 2002).

Quantitative trait loci (QTL) analyses

QTL analyses were performed based on interval mapping, as implemented in R/qtl (Broman et al., 2003). LOD significance thresholds were determined using a 1000 x-permutation test of each trait. QTLs with LOD scores higher than the LOD significant threshold were declared significant. To determine the confidence interval for each QTL, a 1.5-LOD support interval was applied, excluding loci with LOD score below the

significant threshold. The peak LOD score was used to estimate the phenotypic variance explained by each QTL. The effect of QTLs was investigated on the loci at the peak of each QTL.

Identification of candidate gene by microsynteny

Since *N. caerulescens* and *A. thaliana* are both in the Brassicaceae family, it is possible to identify candidate gene in *N. caerulescens* based on the synteny with known genes in *A. thaliana*. To identify candidate genes underlying QTLs for flowering time, microsyntenies were generated using the alignment between the *N. caerulescens* 'Ganges' reference genome (Aarts, unpublished data) and the *A. thaliana* Col-0 reference genome, TAIRv10 (Berardini et al., 2015), which was conducted using Minimap2 (Li, 2018). A custom R script was used to filter poor alignments, which have poor mapping quality (mapq = 0) and short alignment length (alen < 100), and elucidate *A. thaliana* genes localized within each microsynteny. All candidate genes in QTLs associated with flowering time were annotated with the information and functional description of presumed *A. thaliana* orthologues.

Statistical analysis

Descriptive statistical analysis was performed using R (Team, 2008). Broad-sense heritability (H^2) was estimated for each trait as the ratio between genetic variance and phenotypic variance using a mixed linear model, modified from Flood et al. (2016). Pairwise Pearson's correlation (r) was calculated for all traits using psych version 2.3.9 implemented in R (Revelle, 2018).

Result

Phenotypic variation of life history traits

A RIL population of *N. caerulescens* was made from the F_2 progeny of the cross between accessions 'Lellingen' (mother) and 'Ganges' (father), propagated by single seed descent to F_7. In 2019 and 2020, the whole population (232 lines) was assessed for a few life-history traits: FT2019, FT2020, SFP, SFS, PH and RD (Table 1 and Figure 1). The average FT2019 of GA, LE and RILs were 175.3, 168.2, and 170.9 Days After Sowing (DAS), respectively (Figure 1A). In 2020, plants flowered slightly earlier, as the average FT2020 were 168.2, 165.5 and 164.7 DAS for GA, LE and RILs (Figure 1B). Although the flowering time differed in both years, a positive correlation between years was observed (r = 0.7). The broad sense heritabilities of FT2019 and FT2020 were 78.7% and 77.5%. This high H^2 indicated that the variation of flowering time is largely due to genotype as it is commonly found in several plant species.

The averages found for SFP and SFS were 61% and 69%, respectively. A low rate (< 50%) was found in 54 RIL lines (32% of RIL population) for SFP and 34 RIL lines (20% of RIL population) for SFS (Figure 1C and 1D). This suggests that this RIL population can be propagated by self-pollination, nevertheless additional hand pollination is required to avoid losing RILs with low SFP and SFS. A positive correlation was found between SFP and SFS (r = 0.6).

For plant growth under greenhouse conditions, the average PH and RD was 26.20 cm and 15.47 cm, respectively (Figure 1E and 1F). PH showed high broad sense heritability (H^2 = 77.59), while it was lower for RD (H^2 = 45.57). PH and RD are poorly correlated (r = 0.2).

Table 1. *Descriptive statistics of life history traits in RIL population.*

Life history traits	Min	Max	Average	Standard deviation	H^2 (%)
FT2019 (DAS)	118.7	190.0	170.90	6.7	78.7
FT2020 (DAS)	133.3	187.7	164.8	7.4	77.5
SFP (%)	1	100	61	24	ND
SFS (%)	13	100	69	23	ND
PH (cm)	10.6	36.5	26.2	5.7	77.5
RD (cm)	11.2	20.1	15.4	1.7	45.5

ND: Not determined; Self-fertility is referred to the ratio between the number of fertilized siliques to the total number of siliques on the primary inflorescence (SFP) and the most developed secondary inflorescence (SFS); Plant height (PH) and rosette diameter (RD) is measured from image of RIL lines recorded in April, 2019; H^2 for SFP and SFS were not determined since only one measurement was made per RIL.

Pairwise correlation between flowering time, self-fertility and plant growth was determined to explore relationship between traits. Negative correlations were found for flowering time and self-fertility (r < 0.17); and flowering time and PH (r < 0.46). A weak positive correlation was observed between PH and self-fertility (r < 0.18). RD was uncorrelated to any traits, except for PH.

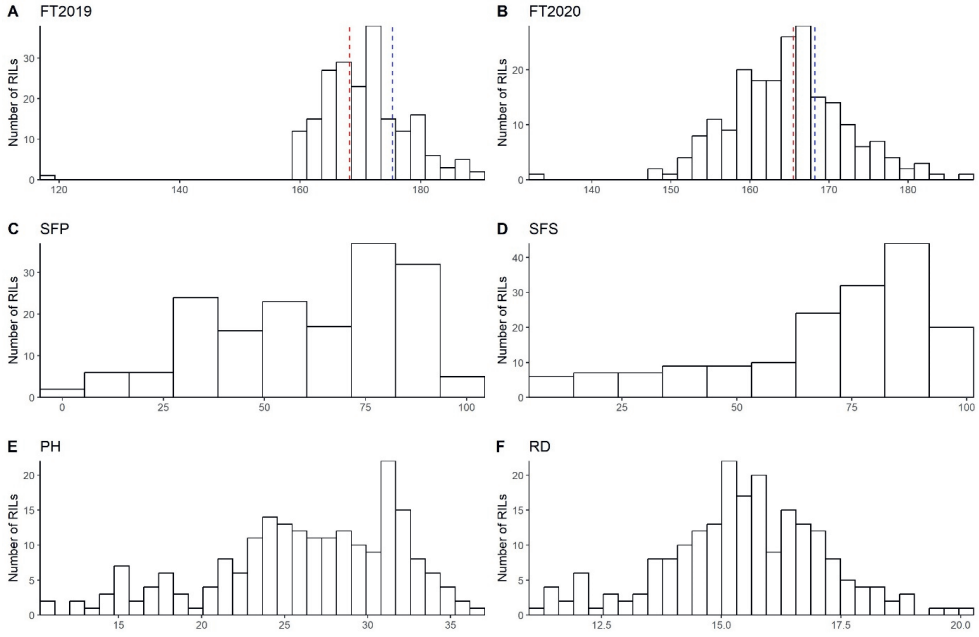


Figure 1. *distribution of life history traits of RILs. Red and blue dashed line indicate average flowering time of RIL parental accessions GA and LE, respectively.*

SNP genotyping and linkage map construction

The RIL population was genotyped by low-coverage, whole genome sequencing. SNP calling was conducted using the *N. caerulescens* GA draft whole-genome sequence (Aarts, unpublished data) as reference. A total of 406,038 SNPs were obtained and converted into the ABH format of allelic annotation, resulting in 406,038 potential SNP markers. After filtering for missing data, 94,377 SNPs and 210 RILs were retained. To reduce redundancy of multiple SNPs representing the same locus on the linkage map, the SNPs data were filtered for identical ABH genotypes, resulting in 16,315 SNPs that could be used for construction of a high-density linkage map. As the use of so many markers would take excessive computation time, this number was further reduced. Analysis of the markers locating at one contig of the reference whole genome sequence, showed that a physical sequence distance of 1 Mb corresponds to a linkage mapping distance of ~ 5 cM. Based on this ratio, 433 representative, well-distributed SNP markers were selected and used for linkage map construction. The mapping resulted in seven linkage groups (LG), which spanned 840 cM. with an average distance of 1.93 cM between markers (Figure 2 and table S1). The numbering of LGs was assigned arbitrarily by the Joinmap software. LG6 was the longest LG with four gaps of 21 - 20 cM, while LG7 was shortest one. The fractions of the LG lengths were respectively 17.56 %, 13.24 %, 15.95 %, 11.29 %, 12.69 %, 19.40 % and 9.88 % of the total map length (from LG1-LG7).

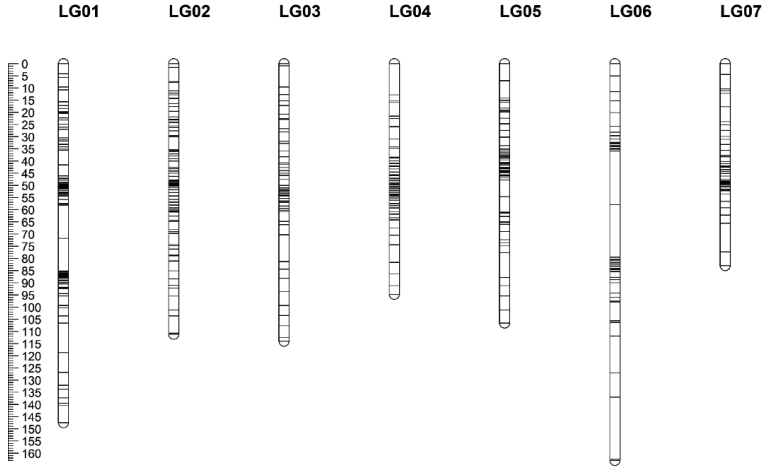


Figure 2. *Linkage map of N. caerulescens. The linkage map based on RILs was constructed with Joinmap version 4.1* (Van Ooijen, 2006). *The linkage group names are indicated at the top. A ruler indicating genetic distances in cM is shown on the left. SNP markers are indicated as black horizontal bars.*

Of 443 SNP markers, 73 segregation distorted markers were observed in all LG, except LG2 and LG7 (Table S2). Of these, 57 distorted markers displayed an over-representation of GA alleles, with 44 markers being mapped to LG4 within the range of 16 to 95 cM. Additionally, 11 markers with over-represented GA alleles were observed on LG1, with seven markers located at the end of LG1 between 118.7 to 147.5 cM, and four markers detected at 35.5, 48.5, 85.9, and 86.7 cM on LG1. The remaining distorted markers were distributed at 66, 73.7, and 74.9 cM on LG5, and at 15 and 20 cM on LG6. Furthermore, 15 distorted markers with an overrepresentation of LE alleles were exclusively found on LG3, positioned between 9.6 to 32.8 cM and at random locations including 47.6, 64.8, and 66.1 cM. Most

of distorted markers showed ratio between GA allele to LE allele close to 60/30 or 30/60 (Table S2). The serious distortion ratio such as ratio of GA allele to LE allele of 90/10 or 10/90 was not observed in the distorted markers.

QTLs associated with target traits

QTL analyses were carried for the six traits of interest, based on interval mapping in R/qtl. In total, 7 QTLs were identified, for five of the six traits (Table 2). No significant QTL was found for SFP (Figure 4B).

Table 2. *QTLs detected in this study*

Trait	LOD threshold	QTL	LG	QTL peak marker	LOD	% PVE	Trait enhancing allele
FT2019	2.26	qFT2019	2	Contig21_483829	7.18	14.95	LE
FT2020	2.32	qFT2020-1	2	Contig21_483829	8.40	16.91	LE
		qFT2020-2	5	Contig94_2743579	3.14	6.68	GA
SFS	2.37	qSFS	7	Contig12_2496096	3.35	8.80	LE
PH	2.49	qPH-1	3	Contig102_492796	3.06	6.49	LE
		qPH-2	5	Contig94_1499252	5.15	10.68	GA
RD	2.44	qRD	2	Contig5_499375	5.42	11.21	LE

Likelihood of Difference (LOD) thresholds are determined by permutation testing (n = 1000) as implemented in R/qtl. LG: linkage group; LOD: maximum LOD score; % PVE: percentage explained phenotypic variance.

For FT2019, a major QTLs, qFT2019, was detected on LG2, which located between markers Contig337_87051 (30.15 cM) and Contig34_1188586 (47.92 cM) (Figure 4A). By adopting a 1.5-LOD confidence interval, the qFT2019 spanned over 17.77 cM with an LOD score of 7.17, explaining 14.95% of the total phenotypic variance. The same locus on LG2 was identified for FT2020, now called qFT2020-1, explaining 16.9 % of the total phenotypic variance. In addition, a new QTL was found on LG5, qFT2020-2 (Figure 4A), explaining 6.7% of the total phenotypic variance. This qFT2020-2 reached a LOD score of 3.14, and mapped between Contig94_499416 (0 cM) and Contig351_283777 (15.84 cM) on LG5. qFT2020-1 covered a narrower range than qFT2019, which will be beneficial for fine mapping this QTL. The GA alleles for qFT2020-1 and qFT2020-2 affected flowering time in different directions. The GA allele of qFT2020-1 increased flowering time, while qFT2020-2 decreased it (Figure 3B and 3C).

One major QTL, qSFS, was identified for SFS on LG7, with a LOD score of 3.35 and an explained variance of 8.8 % of the total phenotypic variance of (Figure 4B). The qSFS QTL covers 3.44 cM, between Contig12_1499740 (62.19 cM) and Contig12_2496096 (65.63 cM). The LE allele conferred higher SF (Figure 3D).

Two QTLs associated with PH were detected on LG3 and LG5 (Figure 3C). The major QTL, qPH-2, with a LOD score of 5.15, located on LG5, covering 18.3 cM, flanked by Contig94_499416 (0 cM) and Contig253_697428 (18.31 cM). A minor QTL, qPH-1, located to LG3 between Contig257_873005 (0 cM) and Contig102_492796 (9.65 cM), with a LOD score of 3.06. The two QTLs explained 10.68 % and 6.49 % of PH, respectively. For RD, the major QTL qRD was detected on LG2, which was flanked by Contig233_261338 (69.68 cM) and Contig52_498228 (95.49 cM), covering 25.81 cM (Figure 4C). The qRD explained 11.21 % of the total variance for RD.

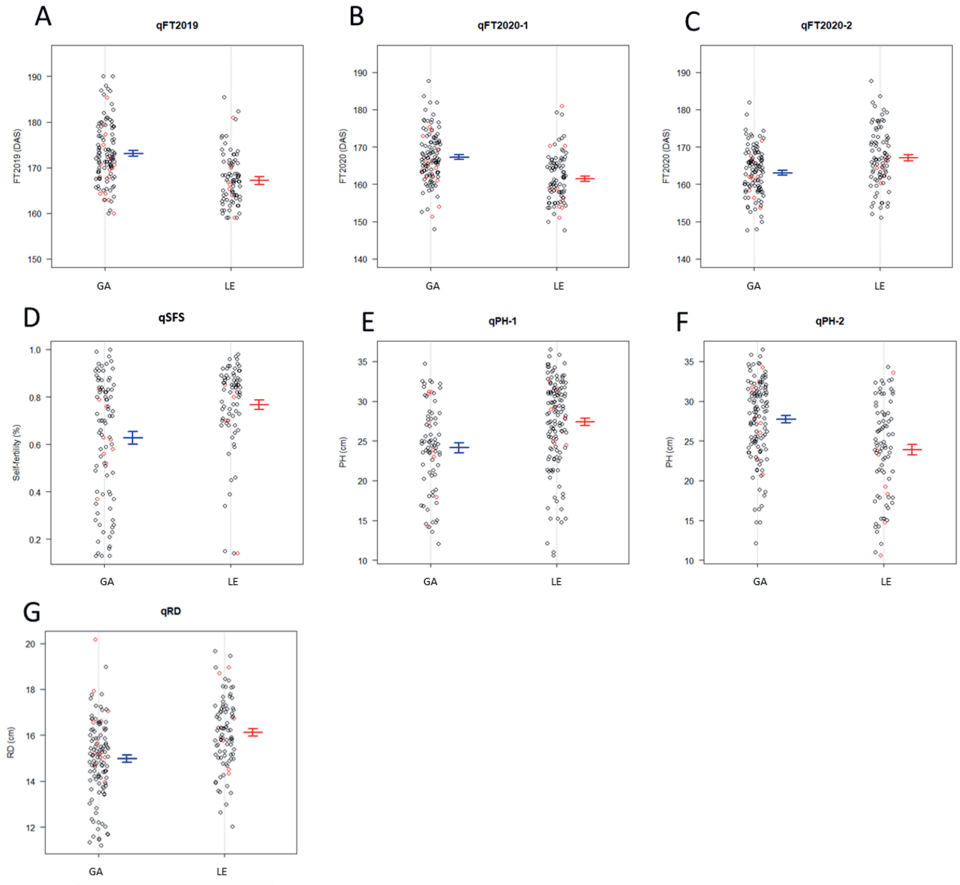


Figure 3. *QTL effect plots of life history traits from the RIL population on their significant SNPs, which includes QTLs for flowering time, FT2019 on Contig21_483829 (A); FT2020 on Contig21_483829 (B) and Contig94_2743579 (C); self-fertility of secondary inflorescence, SFS on Contig12_2496096 (D); plant height, PH on Contig102_492796 (E) and Contig94_1499252 (F); and rosette diameter, RD on Contig5_499375. Each dot represents a phenotypic value. GA = allelic effect from parental accession GA. LE = allelic effect from parental accession LE. Blue and red bars indicated average of phenotypic values ± 1 standard error for GA and LE.*

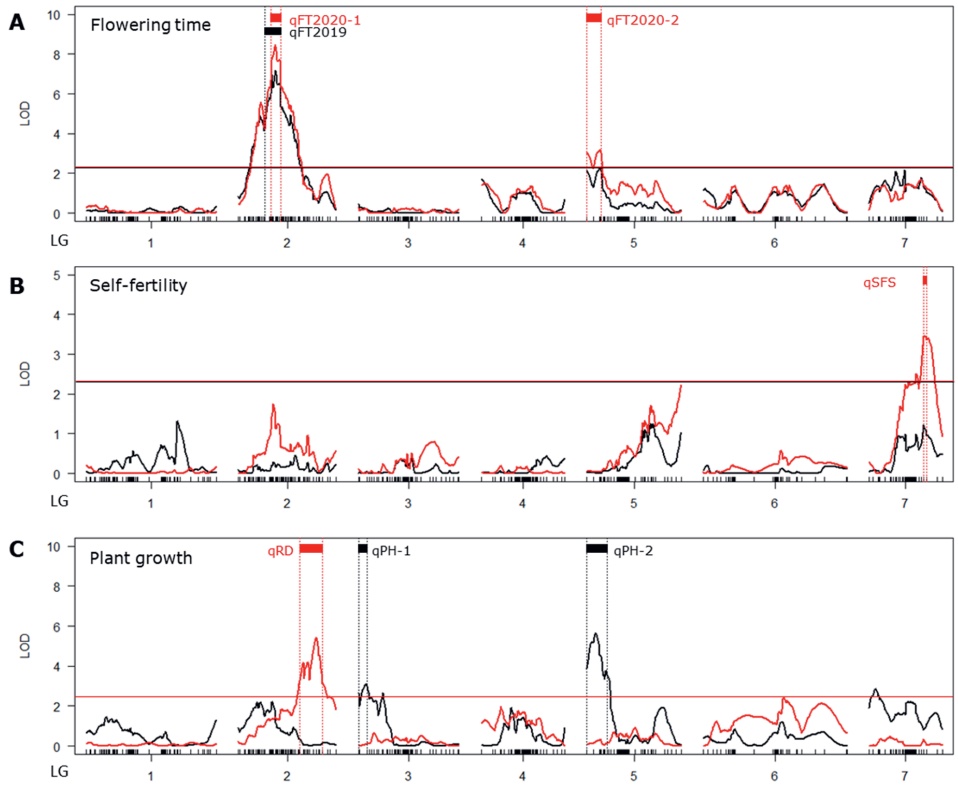


Figure 4. *Seven significant QTLs detected in this study. Flowering time is the number of days between sowing until the appearance of the first flower with an open anther, measured in 2019 (FT2019) and 2020 (FT2020). Self-fertility is the ratio between the number of fertilized siliques to the total number of siliques on the primary inflorescence (SFP) and the most developed secondary inflorescence (SFS). Plant height (PH) and rosette diameter (RD) is measured from image of RIL lines. (A) The distribution of LOD scores for FT2019 (black) and FT2020 (red). (B) The distribution of LOD scores for SFP (black) and SFS (red). (C) The distribution of LOD scores for plant height (PH) (black) and rosette diameter (RD)(red). QTLs are indicated with red or black horizontal bars. Vertical dashed lines indicate the positions of the flanking markers for each QTL, based on a 1.5-LOD confidence interval. LG numbers are indicated below the panels. Horizontal line indicates LOD score thresholds corresponding to colour of each trait per plots.*

Candidate genes controlling flowering time in *N. caerulescens*

For this study, the attempt was made to identify a candidate gene underlying one of the flowering time QTLs. In comparison to other traits (self-fertility and plant growth), flowering time is well-studied plant trait that is controlled by a relatively low number of well-characterized genes in the Brassicaceae (Bouché et al., 2016; Koornneef et al., 1998; Putterill et al., 2004). The genomic regions between the flanking markers of qFT2019, qFT2020-1 and qFT2020-2 were further examined for known candidate genes that could underly the QTL, based on microsynteny between *A. thaliana* and *N. caerulescens*. qFT2020-1 and qFT2019 co-located on LG2 and shared the same peak marker, but qFT2020-1 was narrower than qFT2019. qFT2020-1 was covered by contig 277 and contig 21 of the *N. caerulescens* reference genome

(Figure 5). This region largely aligned to two regions on chromosome 4 of *A. thaliana*: between 7988058 bp to 9047449 bp; and between 9581586 bp to 9726828 bp, which contained 150 and 27 protein coding genes, respectively. Of all these genes, only the *N. caerulescens* orthologue of At4g15880, *EARLY IN SHORT DAYS 4* (*ESD4*), was found as the single known candidate gene for flowering time. This gene maps closely to the peak of qFT2020-1. *ESD4* was reported to be involved in the autonomous flowering pathway in *A. thaliana* (Kong et al., 2017; Reeves et al., 2002; Son et al., 2014; Villajuana-Bonequi et al., 2014). The minor QTL, qFT2020-2, was located between contig 94 and contig 351 of *N. caerulescens*. This region aligned to *A. thaliana* chromosome 5 between 24469587 bp to 26970641 bp, spanning over 2.5 Mb, containing several 100s of predicted genes. Due to the pronounced alignment size, qFT2020-2 was not investigated for candidate genes.

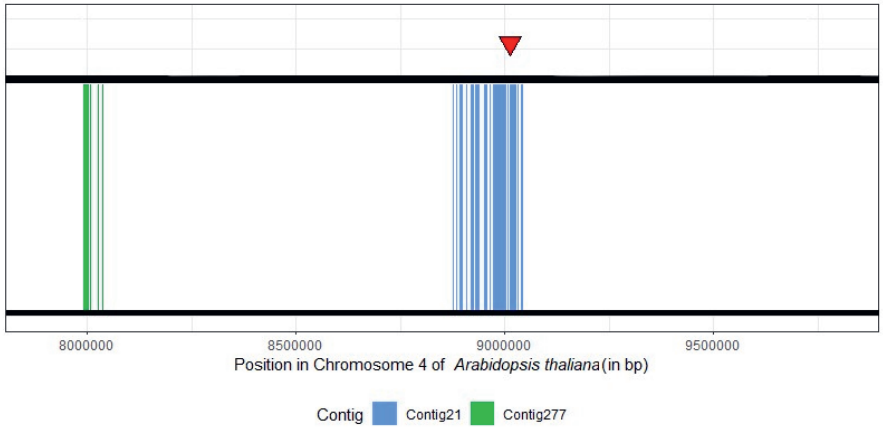


Figure 5. *The N. caerulescens orthologue of A. thaliana EARLY IN SHORT DAYS 4 (ESD4) gene maps to the major qFT2020-1 flowering time QTL. qFT2020-1 is covered by contigs 227 and 21 of the N. caerulescens draft genome sequence. These contigs align with the regions from 7,988,058 to 9,047,449 bp and 9,581,586 to 9,726,828 bp of A. thaliana chromosome 4. The candidate flowering time gene ESD4 locates to positions 9012569 to 9016179 bp on chromosome 4, covered by contig21 of N. caerulescens (indicated with a red triangle). Box represents chromosome 4 of A. thaliana. Vertical blue and green lines represent aligned regions between N. caerulescens contig 21 and contig 277 of qFT2020-1 and A. thaliana.*

Discussion

An improved SNPs-based linkage map of *N. caerulescens*

Genetic linkage maps have provided the foundation for dissecting the genetic basis of various traits in plant species, both crops and models and lead to identifying candidate causal genes affecting the traits. This study presents the construction of a new genetic linkage map for *N. caerulescens*, consisting of 443 well-distributed SNP markers, on an immortal mapping population of homozygous RILs. This is a great improvement over previous maps and mapping populations made for *N. caerulescens*, which used AFLP and PCR-based markers on largely heterozygous F_2/F_3 lines Assunção et al., 2006; Deniau et al., 2006). We used low-coverage whole genome sequencing to identify these SNP markers. This was only possible due to the recent advances in sequencing technology, along with decreasing sequencing costs, and the

availability of a draft whole genome reference sequence. The present work discovered 16,315 biallelic polymorphic SNP markers distributed across 300 contigs of the reference genome sequence, which illustrates the advantage of the current genotyping techniques and other genomic tools. The representative set of 443 SNP markers that was selected, simplified the ordering of markers and reduced the computation time to a practical level. If required for local fine mapping, this current linkage map can easily act as the backbone for high-density mapping by adding a selected set of local SNP markers from the 16,315 SNP dataset to the QTL analysis. This will fasten the speed of computation time because it is only the calculation of recombination frequency between the new markers, while the ordering of markers is not needed. The new markers can be placed in LGs by referring to their original physical position on a contig of the *N. caerulescens* whole genome sequence.

Our linkage map exhibits a resemblance to the previously reported linkage maps of *N. caerulescens* (Assunção et al., 2006; Deniau et al., 2006). Unfortunately, the present map displays some map inflation as the total map length is ~ 1.6 – 1.8 times longer than those reported before. This inflation may be caused by the presence of segregation distortion and possible genotyping errors (Hackett and Broadfoot 2003). Segregation distorted markers were observed in the linkage map, but most of them tend to be mild distortion. Segregation distortion is a natural phenomenon, which may arise from gametic selection (Lyttle, 1991). Several studies tend to exclude distorted markers from the dataset to avoid complications during linkage map construction. In this present map, no obvious problem was shown in LG4. The large gaps located in LG1 and LG6 were also distant from distorted markers. Segregation showed no negative impact on the linkage map of the RIL population. A previous study reported that segregation distortion has minimal effect on marker order and LG length when the distance between adjacent markers is ~ 10 cm (Hackett and Broadfoot, 2003). The present linkage map displayed an average genetic distance below 4 cm, implying that segregation distortion would not contribute to map inflation. Next to segregation distortion, genotyping errors are recognised as a source of map inflation (Cartwright et al., 2007; Hackett and Broadfoot, 2003; Shields et al., 1991). These errors may arise from the use of low-coverage high-throughput genotyping technology. Low-coverage sequencing reduces genotyping costs drastically, compared to PCR-based markers, but it may raise uncertainty with respect to SNP calling if missing genotypes are accidently being called, of if markers are called as heterozygote, when in a homozygous region in largely homozygous lines (Bilton et al., 2018; Petter et al., 2023). Such errors would increase the length of linkage maps (Cartwright et al., 2007; Hackett and Broadfoot, 2003), and might occur in this study. We tried to detect and correct for genotyping errors as much as possible, to alleviate map inflation, using different bioinformatic tools such as SMOOTH (Van Os et al., 2005) and Lep-MAP (Rastas et al., 2013). Nevertheless, if error rates are unknown, these procedures are not straightforward. One obvious indication of erroneous markers contributing to map inflation would be those that result in large gap sizes flanking one marker. Examples of such are especially found for LG1 and LG6 (Figure 1, Table S1). These regions are indeed associated with poor quality SNP markers (e.g. high rate of missing values, and deviating segregation patterns from flanking markers) either of the central marker in the gap, or those adjacent to it. These poor SNP markers should be replaced with better ones obtained from a full genotypic dataset of 16,315 SNP markers. So far, automated tools for this procedure have not been developed, thus requiring manual implementation. Fortunately, this task is manageable, given the limited presence of only a few large gaps in the linkage map. In addition to the gaps on LG1 and LG6, there are also several clusters of markers mapped at high density observed on five LGs (Figure 1). These are likely to indicate the positions of the centromeres. In centromeres, the recombination frequency is strongly suppressed, meaning a short genetic distance in cM is covered by a large genomic region in megabasepairs (Hey, 2004). This has been reported in linkage maps of many other plant species (for

detailed review in Fernandes et al., 2019)). While in itself not a problem for QTL analysis, a few markers could be removed from these regions still, as they are likely to not be very informative. Overall, the current linkage map could undergo some further refinement through the use of various bioinformatic tools and higher-quality SNP markers from our full SNP dataset, but it is certainly useful for QTL analysis in its current state.

In comparison to the cytogenic map (Mandakova and Lysak, 2008), the current genetic map shows some degree of similarity in terms of the length of each chromosome/LG. The cytogenic map was constructed by comparative chromosome painting method and well characterized for metal homeostatic genes from *A. thaliana* (Mandáková et al., 2015). These genes provided reference loci for assigning LGs of the present linkage map to their chromosomes reported in the cytogenetic map (Mandáková et al., 2015). The integration of the current genetic linkage map and the reported cytogenetic map (Mandáková et al., 2015) can be established as described in other species such as *A. thaliana* (Chang et al., 2001) and *Pisum sativum* (Ellis and Poyser, 2002). Such integrated map will serve as a valuable tool, facilitating functional genomics analysis for *N. caerulescens*.

The candidate genes controlling flowering time in *N. caerulescens*

The induction of flowering has been the subject of extensive research, particularly in the well-studied model species *A. thaliana*. To date, 306 genes have been reported to be involved in seven regulatory networks controlling inflorescence development in *A. thaliana* (Bouché et al., 2016). In most of Europe, *A. thaliana* is a winter-annual species, requiring vernalisation, a prolonged period of low temperatures, typical for the winter season in temperate regions, to induce flowering. So far, two genes involved in vernalisation regulatory network including *FRIGIDA* (*FRI*) and *FLOWERING LOCUS C* (*FLC*) were identified (Amasino, 2010; Caicedo et al., 2004; Shindo et al., 2005; Werner et al., 2005). This regulatory network is mainly regulated by *FLC*, a central repressor of inflorescence development (Simpson and Dean, 2002). The expression of *FLC* is induced by *FRI* before vernalisation (Simpson and Dean, 2002). Once *A. thaliana* is exposed to prolonged cold of winter, it suppresses expression of *FLC* and promotes inflorescence development in spring. The requirement of vernalisation is also observed in *N. caerulescens* (Wang et al., 2020). As it is a winter annual from the same Brassicaceae family as *A. thaliana* it is anticipated that orthologues of the genes controlling variation in flowering time in *A. thaliana* would do the same in *N. caerulescens*, meaning the will display natural genetic variation and thus play a role in controlling the variation in flowering time in *N. caerulescens* through similar regulatory pathways. Currently, only mutants for *FLC* and *SHORT VEGETATIVE PHASE* (*SVP*), as well as for another unknown gene, have been described to affect flowering time in *N. caerulescens*, but no natural variants for flowering time have been analysed as yet (Wang et al., 2020).

This study does not identify either *FLC* or *FRI* as candidates underlying any of the two flowering time QTLs, while these are the major genes affecting variation in flowering time in *A. thaliana*. *FRI* and *FLC* are respectively located on contig 0, between 3490774 to 3519508 bp, and on contig 31, between 168310 to 297689 bp of the draft *N. caerulescens* reference genome (Figure S1). While these genes reside on different chromosomes in *A. thaliana*, both contigs are mapped to LG4, in which no significant QTLs of flowering time were observed. A previous study showed that the decrease in *NcFLC* expression is comparable in GA and LE, reducing it to very low levels after 12 weeks of vernalization (Wang et al., 2020). As this time would be comparable to the vernalisation period the RILs experienced in the unheated greenhouse over winter, it may explain why *FLC* is not detected in this RIL population. Also different from the situation in *A. thaliana* is that early flowering *N. caerulescens* accessions, in which *FRI*

and *FLC* are mutated and non-functional, to-date have not been observed in nature, all accessions examined so far need substantial vernalization time (chapter 4). All these clues suggest that *FLC* and *FRI* are less important in determining adequate flowering in their natural environment for *N. caerulescens* as they do for *A. thaliana*, of which early flowering accessions exist in nature.

Instead of *FLC* and *FRI*, we identify *ESD4* as a plausible candidate gene affecting variation in flowering time in *N. caerulescens*. The *N. caerulescens* homologue of *AtESD4* was located on contig 21 of the *N. caerulescens* draft reference genome, close to the peak marker of the most prominent flowering time QTL, and the only known flowering-time-related gene found in the region. The trait-enhancing allele was contributed by the LE background. Previous studies reported that *ESD4* participates in the autonomous regulatory network of flowering time in *A. thaliana* (Cheng et al., 2017). In *A. thaliana*, the *esd4* mutant shows severe early flowering particularly under short days, indicating that *ESD4* functions as a flowering repressor (Murtas et al., 2003; Reeves et al., 2002). *ESD4* is possibly regulating *FLC* by encoding a small ubiquitin-like modifier (SUMO) protease, variation of which affects stability of the FLC protein (Son et al., 2014). Even though *ESD4* has so far not found to be involved in natural genetic variation for flowering time in *a*is proposed as a candidate gene that is worth further investigations.

A new immortal mapping population for analysis of metal accumulation and tolerance

Different types of mapping populations have been widely used for forward genetic research in order to identify quantitative trait loci underlying traits of interest in plant model species and crops. Immortal mapping populations are considered to be very useful in forward genetics, as the lines can be phenotyped for different traits for multiple times, while they only need to be genotyped once. Even though *N. caerulescens* is described as a short-living species, being mainly found as a winter annual, it requires substantial vernalisation time to induce flowering and complete its life cycle. Because of the size of the population, with over 200 lines, the propagation of the RILs (often by growing five plants per line) needs to rely on natural vernalization, over winter, which allows for only single generation per year to be produced, with plants growing from fall to the early summer next year. Generation of the RIL population has therefore not been trivial, requiring eight years to develop it. The successful analysis of flowering time QTLs acts as an instructive demonstration of the genetic power of the population for a trait with a high H^2 (Table 1).

The restricted genetic diversity deriving from two parental genotypes may impede the QTL detection in RIL populations. Enhancing genetic variation through the use of multiple-parental populations (Klasen et al., 2012) or a diverse set of natural populations (Atwell et al., 2010) has been recognised as an effective way to exploit genetic resources for QTL identification in *A. thaliana*. The diversity panels of natural accessions are indeed more attractive compared to multiple-parental populations, which demand extended population development similar to RILs. Recently, a diversity array of 86 natural *N. caerulescens* accessions has been reported (van der Zee et al., 2021), which could be used for QTL detection in such high genetic diversity population through genome-wide association analysis (further elaborated on in **Chapter 4**).

Even though we do not detect the initially expected genes underlying major flowering time QTLs in *A. thaliana*, the high heritability and the reliability in scoring flowering time, made it an attractive exemplary trait to score and analyse here. This may be more challenging for traits that are strongly influenced by environmental factors (leading to lower H^2), in particular for metal accumulation capacities and

tolerance. Our findings demonstrated that this RIL population is a robust mapping population for genetic studies in *N. caerulescens*, which is promising to use to explore metal adaptation traits in the future.

Conclusion

This study presents the first 'immortal' RIL population for quantitative genetic analysis of the Zn/Ni/Cd hypertolerant and hyperaccumulating species *N. caerulescens*. The population is of sufficiently large size to provide a large number of recombination events. It is genotypically characterized by a large set of bi-allelic polymorphic SNP markers providing a reliable linkage map, all contributing to high-resolution genetic mapping. This population is robust and ready-to-use, and will be a powerful resource for dissecting quantitative traits involving metal accumulation and tolerance. As proof of concept, this study revealed a major flowering time locus for *N. caerulescens*, including *ESD4* as a novel candidate to underly a flowering time QTL.

Supplementary information

Table S1. *Summary of the single nucleotide polymorphism (SNP)-based linkage map, displaying seven linkage groups (LGs). This map is constructed in Joinmap version 4* (Van Ooijen, 2006) *based on 443 SNP markers obtained for the N. caerulescens LE x GA RIL population.*

LG	# markers	Total map length (cM)	Average marker distance (cM)	Maximum gap (cM)
1	94	147.48	1.60	13.65
2	76	111.23	1.48	7.13
3	55	133.98	2.11	10.93
4	58	94.83	1.66	12.82
5	58	106.57	1.87	10.33
6	46	163.01	3.62	25.47
7	47	82.96	1.80	11.71
Overall	433	840.06	1.93	25.47

Table S2. *Segregation distorted markers observed in this study*. Chi-square test identified a total of 57 markers skewed toward particular parental lines across five linkage group (LG). The ratio between GA allele to LE allele (GA/LE) was calculated as percentage relative to the total RIL lines.

Marker	LG	Position (cM)	Missing genotype	GA allele	LE allele	GA/LE (%)
Contig362_38124	1	35.491	113	69	28	33/13
Contig43_298358	1	49.482	96	78	36	37/17
Contig286_111110	1	85.891	101	80	29	38/14
Contig175_210302	1	86.721	54	103	53	49/25
Contig107_498172	1	118.651	14	127	69	60/33
Contig307_450125	1	126.757	16	124	70	59/33
Contig308_565226	1	132.126	24	125	61	60/29
Contig223_213417	1	133.694	38	113	59	54/28
Contig407_87162	1	139.52	57	107	46	51/22
Contig239_333322	1	140.502	33	122	55	58/26
Contig145_496950	1	147.482	14	126	70	60/33
Contig102_492796	3	9.647	16	69	125	33/60
Contig229_490162	3	12.742	19	67	124	32/59
Contig38_2739117	3	15.321	18	68	124	32/59
Contig38_2493031	3	17.211	20	67	123	32/59
Contig38_1499883	3	20.75	21	62	127	30/60
Contig248_497046	3	22.606	24	60	126	29/60
Contig248_526692	3	22.912	25	60	125	29/60
Contig231_426774	3	26.666	34	58	118	28/56
Contig105_472484	3	28.139	34	59	117	28/56
Contig256_129442	3	31.903	65	49	96	23/46
Contig104_491594	3	32.83	27	64	119	30/57
Contig104_520803	3	32.83	27	64	119	30/57
Contig66_445441	3	47.604	74	43	93	20/44
Contig152_444152	3	64.768	30	64	116	30/55
Contig160_476477	3	66.095	58	50	102	24/49
Contig4275_3272	4	16.045	93	83	34	40/16
Contig320_474224	4	21.621	26	120	64	57/30
Contig57_1917825	4	22.518	30	120	60	57/29
Contig57_1499191	4	25.911	26	120	64	57/30
Contig291_499667	4	34.178	35	118	57	56/27
Contig291_508109	4	34.673	32	122	56	58/27
Contig106_531379	4	38.407	34	124	52	59/25
Contig106_499770	4	38.729	34	125	51	60/24
Contig103_499486	4	39.855	19	131	60	62/29
Contig103_907282	4	40.987	21	128	61	61/29
Contig318_116332	4	41.875	43	112	55	53/26
Contig227_204519	4	42.342	50	111	49	53/23
Contig183_415908	4	43.302	39	113	58	54/28
Contig114_309038	4	44.486	48	106	56	50/27
Contig3_476793	4	45.773	29	117	64	56/30

Marker	LG	Position (cM)	Missing genotype	GA allele	LE allele	GA/LE (%)
Contig200_429089	4	47.083	24	123	63	59/30
Contig228_184177	4	47.898	55	105	50	50/24
Contig88_375398	4	48.712	45	111	54	53/26
Contig84_258765	4	49.844	45	112	53	53/25
Contig92_403162	4	49.844	44	111	55	53/26
Contig169_284860	4	50.342	60	100	50	48/24
Contig191_436083	4	50.836	40	118	52	56/25
Contig110_443062	4	52.081	46	109	55	52/26
Contig72_493364	4	52.702	28	122	60	58/29
Contig72_801534	4	52.948	28	124	58	59/28
Contig30_1292085	4	53.822	15	128	67	61/32
Contig30_499797	4	54.199	18	124	68	59/32
Contig261_164355	4	54.579	59	100	51	48/24
Contig230_214874	4	56.062	39	111	60	53/29
Contig153_600767	4	57.459	21	122	67	58/32
Contig153_498519	4	58.318	20	125	65	60/31
Contig53_488713	4	59.255	14	127	69	60/33
Contig355_132393	4	61.835	38	112	60	53/29
Contig418_138599	4	62.247	37	114	59	54/28
Contig113_498547	4	62.302	16	124	70	59/33
Contig0_6495932	4	67.316	20	124	66	59/31
Contig0_5498985	4	70.54	23	123	64	59/30
Contig0_3498962	4	81.842	21	122	67	58/32
Contig0_2497575	4	86.419	23	127	60	60/29
Contig0_1499902	4	91.363	24	127	59	60/28
Contig0_499734	4	94.99	20	133	57	63/27
Contig26_495316	5	65.976	22	122	66	58/31
Contig195_370468	5	73.669	42	110	58	52/28
Contig116_1130176	5	74.855	17	124	69	59/33
Contig2_3499276	6	15.318	18	125	67	60/32
Contig2_4496404	6	20.12	39	111	60	53/29

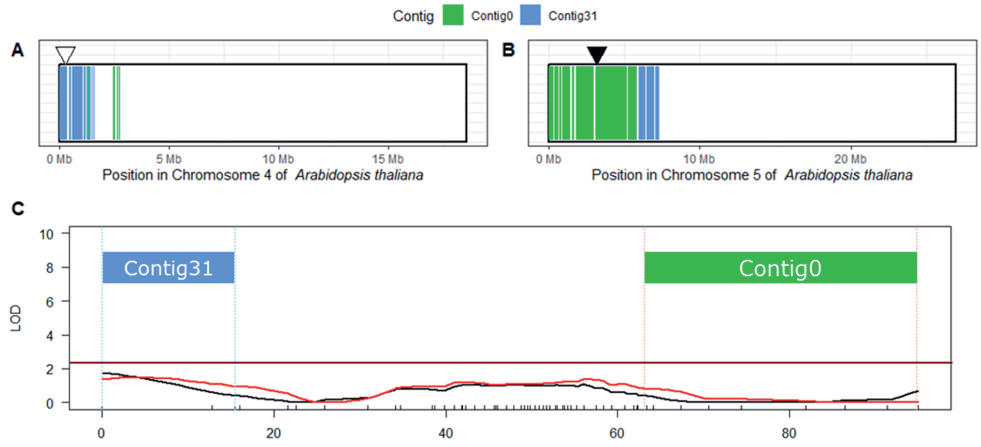


Figure S1. *Locations of flowering time controlling genes, FRIGIDA (FRI) and FLOWERING TIME LOCUS C (FLC), on the Noccaea caerulescens reference genome. (A) FRI is located on chromosome 4 of Arabidopsis thaliana between 268900 to 271503 bp, (white triangle) which aligns to contig 0 of N. caerulescens. (B) FLC is located on chromosome 5 of A. thaliana between 3173496 to 3179448 bp (black triangle), which aligns to contig 31 of N. caerulescens. Green and blue vertical lines depict alignments between A. thaliana and N. caerulescens contig 0 and contig 31, respectively. (C) LOD score distribution of FT2019 (red) and FT2020 (black) on linkage group 4, to which SNP markers of contig 0 (green box) and contig 31 (blue box) are mapped. The contig 0 SNP markers are located between 64.22 cM to 94.825 cM. The contig 31 SNP markers are located between 0 cM to 15.438 cM. Vertical black and red lines indicate LOD score threshold for FT2019 and FT2020, respectively. No significant SNP markers within contig 0 and contig 31 show LOD score above the permutation-determined significance threshold.*

DNA extraction using SPRI beads modified from Pederson (2021)

Extraction Buffer (need approximately 50 ml/ 96 well plate)

200 mM Tris-HCl

25 mM EDTA

1% SDS

Potassium acetate solution (KAc)

Add 98.14 g of potassium acetate (KAc) to 160 mL H_20 (5 M)

Add 3.5 mL Tween

qs to 200 mL

Equipment:

- Waterbath
- Plate centrifuge (Eppendorf)
- Multichannel pipets from 50-300 uL range
- Shaking table
- Promega 96 well magnet (https://nld.promega.com/products/biochemicals-and-labware/tips-and-accessories/magnabot-96-magnetic-separation-device/?catNum=V8151)
- SPRI beads (see protocol for Homemade SPRI beads)
- Speedbead Magnetic carboxylate modified particles Sigma GEHE45152105050250 (for Europe, VWR is the supplier
- 96 well plates + lids used for sampling, grinding and waterbath steps (step 2-5) and SPRI beads steps (step 11-17)
- Greiner deepwell 1 ml
- Greiner mats for deepwell 1 ml
- 96 wells plates used for KAc steps (step 6-10):
- Greiner 2ml deepwell plates
- Mats for deepwell 2ml
-

Sampling:

Bring the plates with zirconia (3-4) beads already added, to the greenhouse/climate chamber.

A flower head of *N. caerulescens* is used for higher yield of total DNA.

Protocol

1. Prepare 10 mL Extraction Buffer by adding 40 uL of 20 mg/mL RNase A.

2. Tape the mats of the deepwell plates used for sampling to prevent leaking during the shaking and waterbath steps. Make sure mats are sealed tight! Spin 4 minutes after shaking to bring down tissue dust.

Important: freeze sample at -20 °C and spin down quickly after grinding and before adding extraction buffer to bring down debris.

3. Add 500 uL of extraction buffer to each sample

4. Invert plates a few times every 15 mins. Prepare plates with SPRI beads and 80 % ethanol during this hour.

5. Flash spin to pellet debris (3000 x g for 5 min).

6. Add 130 uL KAc plus Tween to empty deep-well plate.

7. Transfer 400 uL lysate to KAc plate.

8. Cover with seal, push seals/mats hard to properly close and invert to mix (1-2 min)

9. Incubate on ice 10 min.

10. Centrifuge at 3000 x g for 5 min.

11. Transfer 400 uL of supernatant to new plate containing 1.1 vol SPRI (beads diluted 1:1 in PEG buffer). Put plate on a shaking table for 30 mins to 1 hour.

12. Allow beads to settle ~5 mins (or until beads are drawn to magnet).

13. Remove supernatant was removed by inverting the plate in the sink, while still on the magnet.

14. Wash beads 3 times with 500 uL 80% ethanol and vortex carefully on the first 2 washes do not let beads 'escape' from the deep well during vortexing.

15. After the last wash, carefully remove all traces of EtOH and let the beads dry (10 – 15 min). Do not let the beads get too dry (generally you can tell the beads are too dry when cracks start to appear).

16. Resuspend in 50 uL Tris-HCl or TE or TE-1, pH 8 and let the beads sit in the buffer for 30-60 mins for good resuspension.

17. Seal the plates tightly and mix by hand or use the Qiagen mix mate at 2000 rpm for 1 min

18. Place on plate magnet for 5 mins

19. Transfer eluate to fresh PCR plate, being careful not to transfer the beads.

Acknowledgement

The authors are thankful to Corrie Hanhart propagating the RIL population until the F_5 generation; to Frank Becker for providing protocols and valuable suggestion on high throughput DNA extraction and library preparation; to the greenhouse staff of Unifarm for taking excellent care of the plants in every reproductive cycle, enabling high plant survival and high viable seed yield; and to the Royal Thai government (Thailand) for financial support for this research.

CHAPTER

3

Genetic analysis of the cadmium exposure response in a *Noccaea caerulescens* recombinant inbred line population

Jitpanu Yamjabok, Siqi Duan, Laurens van Oostrom, Joost van den Heuvel, Henk Schat, Mark G.M. Aarts

Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

Financial source:
This research is financial supported by Royal Thai government (Thailand).

Abstract

Cadmium accumulation in crops poses risks to human health both in short-term and long-term. An understanding of the genetic architecture underlying Cd accumulation and tolerance in a Cd hyperaccumulator species, can be used for developing methods to rehabilitate Cd polluted land by phytoremediation. To identify candidate genes involved with those traits, a recombinant inbred line (RIL) population of the cross between Ganges (a Cd hypertolerant genotype) and Lellingen (an intermediate Cd tolerant genotype) was phenotyped for Cd accumulation, Cd tolerance and growth in response to Cd exposure at 1 µM $CdSO_4$, 15 µM $CdSO_4$ and a series from 50 to 350 µM $CdSO_4$, respectively. Quantitative Trait Loci (QTL) analysis of 158 traits was performed. In total 27 significant QTLs were detected, which explained 5.75 to 31.24% of the phenotypic variance in the RIL population. Colocalizations of different QTLs were observed on different linkage groups. Overall, a QTL of Cd root accumulation, qRT-Cd-1, exhibited a small mapping interval with the most significant Logarithm of Odds (LOD) score of 16.3, colocalized with a QTL of Cd translocation factor, qTF-Cd on linkage group (LG) 2. The narrow QTL significance interval was aligned to a region chromosome 4 of *Arabidopsis thaliana* containing the *HEAVY METAL ATPASE 3* (*HMA3*) gene, encoding a member of the P_{1B}-ATPase superfamily of metal transporters, targeted to the vacuolar membrane. This gene plays a major role in Cd sequestration in *A. thaliana*, *Oryza sativa* and *Sedum plumbizincicola*. The findings in this study are expected to provide a valuable resource for the further identification of genes involved in plant Cd accumulation, tolerance and response.

Introduction

Cadmium (Cd) is a non-essential rare element that is harmful to micro-organisms, plants and animals, as well as humans. Cd is found in the earth's surface at approximately 0.1 part per million (ppm). In nature, Cd is not found as a pure element, but as an impurity, mainly in zinc (Zn), lead (Pb) and phosphate ores. Upon mining rock phosphate and upon mining and smelting of Zn and Pb, Cd can be released into environment. Moreover, Cd is also introduced into the environment via various anthropogenic activities, including incineration of household waste, fossil fuel combustion and the use of phosphate fertilizer (Hayat et al., 2018). Of these, phosphate fertilizer is recognized as a major source for Cd enrichment of agricultural soils (Dharma-Wardana, 2018; Roberts, 2014). This is particularly the case when the phosphate fertilizers are produced from sedimentary phosphate rocks containing high Cd levels (could be up to 507 mg kg^{-1}), as can be found in phosphate deposits in Morocco, Togo, Senegal and Idaho (USA) (Mar and Okazaki, 2012). Globally, Cd is now widely spread in soil, water and atmosphere. In Europe, Geochemical Mapping of Agricultural and Grazing Land Soil of Europe (GEMAS) demonstrates the distribution of Cd across Europe (Birke et al., 2016), where an elevated level of Cd is observed in Croatia, Slovenia, Germany, UK, Greece, France, Ireland and Ukraine. Next to soil contamination, Cd is also detected in water streams in Germany and Slovakia (Pan et al., 2010). In Asia, environmental Cd pollution has been investigated in many countries. For example, ~ 20 million hectares of arable land in China is contaminated with Cd (Cai et al., 2019). In Thailand, enrichment of Cd is observed in paddy fields in the Northern regions with concentrations between 0.31 to 13.9 ppm of total Cd in the soil (Sriprachote et al., 2012).

The spread of Cd pollution in soil has become a serious concern in many countries because Cd is a highly toxic element and a carcinogenic agent. In plants, Cd adversely affects growth, visible at morphological and physiological levels, which subsequently reduces crop yields (Genchi et al., 2020; He et al., 2017). Cd can accumulate in various crop species including major cereals (rice, wheat and maize) and

vegetables. Ingestion of Cd-contaminated crops or food can cause a variety of disorders and health issues in different organs including the respiratory system (pneumonitis), kidneys (kidney stone), the skeletal system (e.g. causing the Itai-Itai disease) and the reproductive system (testicular necrosis) (Hayat et al., 2018). Since metals cannot be degraded, the accumulation of Cd in edible crops has to be prevented to alleviate the harm of Cd contamination. This could be achieved by preventing Cd to be taken up by plants when grown on Cd-containing soil, or by reducing the level of Cd in soil with phytoremediation. Phytoremediation is an ecological-friendly and cost-effective method to absorb harmful metals, including Cd, by using metal hyperaccumulating plants. However, detailed knowledge underlying Cd uptake and detoxifying in plant is required for development and optimization of prevention of Cd uptake and Cd phytoremediation (Clemens et al., 2013).

In nature, a few plant species are reported to survive exposure to high concentrations of metal and accumulate them in their aerial parts without showing toxic effects. Originally, these species are found on either Zn/Cd/Lead (Pb)-enriched calamine soils and/or Ni-enriched ultramafic or serpentine soils (Pollard et al., 2014; Reeves et al., 2017). Among the metal hyperaccumulators, the Ni hyperaccumulators are the majority (532 species), while Cd hyperaccumulators are rare (only 7 species), including *Arabidopsis halleri, Noccaea caerulescens*, and *Sedum alfredii*/*Sedum plumbizincicola*. These hyperaccumulators are not only able to accumulate substantial amounts of metals, they are also able to tolerate elevated concentrations of these metals. The Cd hyperaccumulating species exhibit a wide range of Cd accumulation and tolerance in the field and in experimental conditions. Previous studies have reported that *A. halleri* is able to accumulate ~400 to 600 μg g^{-1} in the shoot and ~135 to 400 μg g^{-1} in the root, depending on the Cd exposure levels (Bert et al., 2003; Corso et al., 2018; Kozhevnikova et al., 2020). In *N. caerulescens*, the accumulation of Cd to as much as 15.7 to 348.5 μg g^{-1}, or even 634 μg g^{-1}, in shoots many genotypes (Kozhevnikova et al., 2020; Sterckeman et al., 2017). Root accumulation of Cd is quantified for few genotypes, mostly for parental lines of bi-parental populations including Ganges (3.2 to 12.4 μmol g^{-1}) and La Calamine (0.2 to 1.8 μmol g^{-1}) (Deniau et al., 2006).

N. caerulescens and *A. halleri* have been investigated and compared to *Arabidopsis thaliana*, the well-known plant model species, for the molecular mechanism underlying accumulation and tolerance of Cd. This identified several genes involved in uptake, translocation and sequestration of Cd (reviewed in Lin and Aarts, 2012; Sterckeman and Thomine, 2020). In these species, Cd is adsorbed into root cells as a divalent cation (Cd^{2+}) by advection. Cd ions can be transported to endodermal cells either by diffusion through apoplast or using transport proteins through the symplast. With the Casparian strip barrier at the intercellular space between endodermal cells, Cd ions need to be transported symplastically or by cellular import and export using transmembrane transporters and chelating agents via cytosol. Cd ions are loaded into xylem and bounded with organic ligands to form complexes. Subsequently, these complexes are transported to aboveground plant parts and distributed to different intracellular compartments in cells of different tissues, depending on species. There are several transport proteins involved in these processes including members of the Zinc-regulated transporter, Iron-regulated transporter Protein (ZIP) family, the P_{1B}-ATPase subfamily of Heavy Metal Associated (HMA) transporters, and the Natural Resistance-Associated Macrophage Protein (NRAMP) families. Transport proteins of the ZIP family are plasma membrane transporters of divalent cations (Fe^{2+}, Zn^{2+}, Cd^{2+}, Co^{2+} and Mn^{2+}), which mediate Cd ions to enter into cytosol. In *N. caerulescens*, *NcZNT1*, a homologue of *A. thaliana ZIP4*, and *NcIRT1* appear to be involved in Cd uptake (Lombi et al., 2000; Pence et al., 2000). Three *HMA* genes, including *HMA2*, *HMA3* and *HMA4*, encode P-type ATPases, exporting Cd outside the cytosol, contributing to Cd root-to-shoot translocation, Cd sequestration in vacuoles and Cd xylem loading,

respectively (Chao et al., 2012; Liu et al., 2017; Mills et al., 2003; Papoyan and Kochian, 2004; Verret et al., 2004; Wong et al., 2009; Wong and Cobbett, 2009). In *A. thaliana*, HMA3 is located in the tonoplast or vacuolar membrane and participates in Cd sequestration (Morel et al., 2009). Similar functions are described for the orthologues of *HMA3* in *N. caerulescens* (Ueno et al., 2011), *O. sativa* (Sasaki et al., 2014) and *S. plumbizincicola* (Liu et al., 2017); however, *HMA3* is expressed in different tissue in these species. *OsHMA3* is expressed in the root cells, which reduces the accumulation of Cd in the leaves and grain of *O. sativa* (Ueno et al., 2010). The number of copies of *NcHMA3* is higher than that of *HMA3* in *A. thaliana*. NcHMA3 is involved in Cd sequestration in leaves of *N. caerulescens* (Ueno et al., 2011). In *A. halleri*, three copies of *AhHMA4* are identified, which play major roles in Cd/Zn hypertolerance and Zn hyperaccumulation (Hanikenne et al., 2008). Likewise, four copies of *NcHMA4* are also reported in *N. caerulescens* (Lochlainn et al., 2011). The *NcHMA4* copy number varies among different *N. caerulescens* accessions, contributing to their different Cd accumulation capacities and tolerance (Craciun et al., 2012). Four proteins of the NRAMP family are reported to be involved in Cd transport in plants. In *A. thaliana*, NRAMP3 and NRAMP4 function as cytosol efflux transporters, that import Cd into the vacuole and contribute to Cd tolerance (Lanquar et al., 2005, 2010). Similarly, orthologues of both *NRAMP3* and *NRAMP4* exhibit similar functions in *N. caerulescens*; however, their expression levels are higher and they confer Cd tolerance in *N. caerulescens* (Oomen et al., 2009). *NcNRAMP1* is characterized to encode a Cd transporter involved in root-to-shoot translocation of Cd, which contributes to increased Cd tolerance in Cd hypertolerant *N. caerulescens* accession Ganges (Milner et al., 2014). In rice, *OsNRAMP1* contributes to Cd uptake as well as iron uptake (Fe) (Takahashi et al., 2011), while *OsNRAMP5* encodes the main transporter for manganese (Mn) and Cd in roots (Ishimaru et al., 2012).

N. caerulescens is an attractive plant model species for the accumulation and tolerance of several metals, including Cd (Assunção et al., 2003a; Milner and Kochian, 2008). When comparing to other hyperaccumulators, *N. caerulescens* shows considerable intraspecific variation of Cd accumulation and tolerance. These traits segregate in a bi-parental population and can be investigated for identification of candidate genes with Quantitative Trait Loci (QTL) analysis (Pollard et al., 2014). Previous studies have developed several bi-parental segregation populations analysis of accumulation and tolerance traits, although only one population, a Ganges (GA) x La Calamine (LC) F_2 population, is studied for QTL of Cd accumulation (Assunção et al., 2006; Deniau et al., 2006; Frérot et al., 2003; Richau and Schat, 2009; Zha et al., 2004). A genetic map of seven linkage group has previously been constructed with Amplified fragment length polymorphism (AFLP) markers and codominant markers based on Expressed Sequence Tags (ESTs) of orthologues of *A. thaliana* genes associated with metal homeostasis such as *ZIP4*, *IRT3* and *NRAMP4*. As a result, two QTLs of Cd shoot accumulation and one QTL of Cd root accumulation have been identified, which explain a substantial proportion of the phenotypic variance (9.6 – 33.1%) (Deniau et al., 2006). However, for none of these QTLs has a candidate gene been identified that is directly linked to metal accumulation or tolerance traits, not even one of the EST markers. An F_2 population has a high level of heterozygosity, which makes it very difficult to re-evaluate or confirm the findings. Instead, a population with fixed, homozygous, genotypes would be required, that allows such population to be used repeatedly for QTL mapping under different conditions or at different locations. In chapter 2, we report the development of a Recombinant inbred line (RIL) population between GA and Lellingen (LE) that segregates for Cd accumulation and tolerance. This RIL population has been genotyped for single nucleotide polymorphism (SNP) markers, that were mapped into seven linkage groups. Moreover, this population has been measured for flowering time under greenhouse conditions. QTL mapping of flowering time identified the orthologue of *EARLY IN SHORT DAYS 4* (*ESD4*) as a likely candidate to underly the major QTL on linkage group 2, explaining 16.91 % of total variance of phenotype. This demonstrated

that the RIL population is suitable for QTL mapping and can be used to map QTL of other traits including Cd accumulation and tolerance (**Chapter 2**).

This present study aims to identify QTLs and candidate genes involved in accumulation and tolerance of Cd in the RIL population. To this end, the RIL population was phenotyped in three experiments for Cd accumulation, for Cd tolerance and for plant growth in response to Cd. The mapping of all traits has been performed using the available genotypic data and linkage map (**Chapter 2**). The detected QTLs were investigated for homologues of protein coding *A. thaliana* genes to identify candidate genes in order to dissect genetic architecture underlying Cd accumulation and tolerance of *N. caerulescens*.

Methodology

Plant material and genotyping data

An F_7 recombinant inbred line (RIL) population of the cross between the calamine, high Zn/Cd tolerant and hyperaccumulating accession Ganges (GA) and the non-metallicolous, medium Zn/Cd tolerant and hyperaccumulating accession Lellingen (LE) was developed as described in **chapter 2**. LE was used as mother. In total, 232 RILs were used in this study. This population was genotyped for SNPs as described in **Chapter 2**. In brief, the raw SNPs were filtered and 433 representative SNPs were used to create a genetic map of seven linkage groups (LG), corresponding to the seven chromosomes of *N. caerulescens*. These representative SNPs were used for QTL analysis in this experiment.

Quantification for Cd accumulation

The seeds were gas-sterilised for three hours upon addition of 2.9% HCl to a common household bleach solution in a sealing jar. The seeds were sown 600-µl PCR tubes of which the bottom had been cut off, filled with modified half-strength Hoagland's (3 mM KNO_3, 2 mM $Ca(NO_3)_2.4H_2O$, 1 mM $NH_4H_2PO_4$, 0.5 mM $MgSO_4.7H_2O$, 1 µM KCl, 25 µM H_3BO_3, 2 µM $MnSO_4.4H_2O$, 2 µM $ZnSO_4.7H_2O$, 0.1 µM $CuSO_4.5H_2O$, 0.1 µM $(NH_4)6Mo_7O_{24}.4H_2O$ and 20 µM Fe-EDDHA (N,N'-ethylenediamine-di(O-hydroxyphenylacetic acid), solidified with 0.3% Gelrite (Duchefa Biochemie, The Netherlands), and suspended in the same modified half Hoagland's nutrient solution. To break any seed dormancy, 7 µL of 10 µM Gibberellins 4+7 (Globlachem, Belgium) was applied to the seeds on every tube, and seeds were stratified at 4 °C for three days. Thereafter, the germination boxes were transferred to a climate-controlled growing chamber (set at 12hr/12hr light/dark, 21/17°C for day/night and 70% RH). After roots of germinated seeds had penetrated the Gelrite and reached the nutrient solution, the samples were transferred to aerated 9-L hydroponic trays contained modified half-strength Hoagland's solution supplemented with 1 µM $CdSO_4$. The nutrient solution was renewed weekly. The plants were grown in a completely randomized block design (RCBD) without replication. After six weeks, all plants were sampled for shoot and root separately. The shoot and root samples were oven-dried at 65 °C for four days. Dry weight (DW) was determined for the shoot and root samples. Ionomics profiles (element concentrations) were determined for 19 elements: Boron (B), Na (Sodium), Magnesium (Mg), Aluminium (Al), Phosphorus (P), Sulphur (S), Potassium (K), Calcium (Ca), Iron (Fe), Manganese (Mn), Cobalt (Co), Nickel (Ni), Copper (Cu), Zinc (Zn), Arsenic (As), Rubidium (Rb), Strontium (Sr), Molybdenum (Mo) and Cadmium (Cd), with an inductively coupled plasma-mass spectrometer (ICP-MS) at the Baxter laboratory of the Donald Danforth Plant Science Centre (St. Louis, Missouri, USA).

The translocation factor (TF) was calculated for 19 elements as shown below.

$$TF = \frac{[shoot]}{[root]}$$

Where [shoot] and [root] is the elemental concentration in shoot and root, respectively.

Assessment of Cd tolerance

A sequential exposure test, adapted from (Schat and ten Bookum (1992), was conducted to determine the effective concentration 100 (EC100). The EC100 was defined as the lowest concentration of Cd that completely inhibited the root growth. Powdered charcoal staining was applied to roots to track root growth. The samples were sown following the same procedure as above. In addition, the samples were pre-cultured on modified half-strength Hoagland's solution without Cd for two weeks to avoid toxicity upon early growth stage. The root system of each plant was dipped several times in an active charcoal suspension that was stirred continuously to stain the roots black. The samples were exposed to the series of $CdSO_4$ as follows: 50 µM, 100 µM, 150 µM, 200 µM, 250 µM, 300 µM, 350 µM. After a week of exposure, the roots were observed and the EC100 was recorded for the sample with completely black primary and secondary roots. Then, the roots were stained again with active charcoal suspended in water and plants were exposed to the higher Cd concentration. The experiment was continued until 350 µM and the samples were harvested at the end of experiment. Shoot and root were dried at 65 $^{\circ}$C and measured for dry weight (DW). A completely randomized block design (RCBD) with two replications was applied to this experiment.

Quantification for growth response to Cd

The projected leaf area (PLA) was measured as a suitable growth parameter for response to Cd. The seeds of the RIL population were sown on filter paper soaked with modified half-strength Hoagland's solution and 10 µL of 10 µM Gibberellins 4+7 (Globlachem, Belgium). The sowing plates were stratified for 3 days and transferred to a climate-controlled growing chamber (set at 12hr/12hr light/dark, 21/17°C for day/night and 70% RH). After radicle emergence (at 9 DAS), the samples were sown on Rockwool blocks soaked with Hyponex nutrient solution containing no Cd (control) or 15 µM $CdSO_4$ (Cd treatment). The samples were grown in a climate-controlled chamber (set at 12hr/12hr light/dark, 70% RH, 20/14°C for day/night and 200 µmol m^2 s^{-1} Photon) for six weeks. This experiment was designed as RCBD with four blocks per treatment and four replicates. The solution was renewed at week 2 and week 4. The image of each block was taken using µEye cameras model UI-1490LE (IDS, Germany) at 9.00, 12.00, 15.00 and 18.00 h daily. The image of each sample was cropped from the block image and measured for PLA as a pixel by Cutimage, an automated python script designed specifically for this growth chamber.

Quantitative trait loci (QTL) analyses

All traits were investigated for QTL analyses, excluding PLA responding to the control treatment. For each trait, QTLs were detected using interval mapping analyses implemented in R/qtl (Broman et al., 2003). Logarithm of Odds (LOD) significant thresholds were calculated using a 1000-permutation test to determine candidate QTL for each trait. The QTLs which had LOD scores above the LOD significant threshold were declared as significant. For each significant QTL, markers with the highest LOD score were used as co-factors in multiple-QTL mapping (MQM) implemented in R/qtl (Arends et al., 2010) to enhance the accuracy of QTL detection. The LOD score threshold was re-calculated again for MQM mapping input data to determine significant QTLs. A 1.5-LOD support interval was applied to determine the confidence interval for each QTL. The loci with a LOD score below the significance threshold were not included in further

analysis. The markers with the highest LOD score were used to estimated phenotypic variance explained by each QTL. The effect of the QTLs was investigated considering the peak of each QTL. The detected QTLs mapped across linkage map were visualised using LinkageMapView (Ouellette et al., 2018).

Candidate gene identification

Available knowledge on *A. thaliana* homologues (orthologues) was used to characterise the function of *N. caerulescens* genes. To identify such *A. thaliana* homologous genes, syntenic blocks were generated using the whole-genome alignment between *N. caerulescens* reference 'Ganges' genome and *A. thaliana* "TAIR10" reference genome (Berardini et al., 2015b), which was conducted using Minimap2 (Li, 2018). The syntenic block covering the contigs where the significant SNPs located to, were used to specify the syntenic region and identify the *A. thaliana* genes from "TAIR10". Homologues of genes involved in metal homeostasis and other relevant processes were selected based on searches with terms including "metal", " ion binding", "ion transport", "ion transmembrane", "Cadmium", "Zinc", "homeostasis", "detoxification", "transmembrane transporter activity" as candidate genes (Halimaa et al., 2014). All the candidate genes in QTLs associated with metal homeostasis were annotated by referring to reported functional descriptions of *A. thaliana* homologues, as well as Araport11 predicted functions based on protein sequences (Cheng et al., 2017). In addition, the candidate genes were compared with reported genes involved in metal homeostasis from other species.

Statistical analysis

Descriptive statistical analysis was performed using R (R Core Team, 2008). Before analysis, all elemental concentrations were log-transformed. Principle component analysis (PCA) was performed for ionomic profile of accumulation test and PLA as a respond to Cd using FactoMineR version 2.9 package implemented in R (Lê et al., 2008). The EC100 and PLA of RILs exposed to 15 µM Cd were estimated for broad-sense heritability (H^2) as ratio between genetic variance to the phenotypic variance using mixed linear model modified from (Flood et al., 2016). Pairwise Pearson's correlation (r) was calculated for element ionomics profiles of the accumulation test. The correlation matrix was drawn using ggcorrplot package implemented in R.

Result

Accumulation test

To phenotype metal accumulation, the RIL population of 195 lines was exposed to 1 µM $CdSO_4$ for six weeks and measured for shoot and root concentration of 19 elements based on ICP-MS analysis. The result exhibited a broad range of concentrations from low to high (Figure 1). In shoot, the concentration of Zn and Cd ranged from 31.19 to 603.6 and 1.65 to 47.22 mg kg^{-1}, respectively (Table 1). The average Zn concentration was higher in shoot than in root, while Cd showed a higher average concentration in root than in shoot (Table 1). Transgression was observed for all elements. For instance, some RILs showed root Cd accumulation levels five-fold higher than that of the highest parental line, GA. The shoot DW was not significantly different between GA and LE ($p > 0.05$). The shoot DWs of RILs ranged from 0.041 to 0.778 g. The TFs of Na, Al, Fe, Co, Cu, As, Sr and Mo distributed in a narrow window below 1 (Figure 1C). Moreover, TFs of Cd and Zn were widely distributed with exceptional maximum values reaching over 12. The TF of Cd ranged from 0.03 to 12.24, with an average of 1.91, very similar to the TF of Zn, ranging from 0.09 to 12.17 and an average of 2.94 (Table 1).

Significant positive correlation coefficients were identified among pairs of shoot element concentrations, ranging from r = 0.16 to r = 0.99, with the exception of Ni, Al, and Mn, which frequently exhibited no significant correlation with shoot elements ($p > 0.05$) (Figure 2). Herein, the shoot concentration of element X was denoted as "Shoot X," while the root concentration of element X was denoted as "Root X." Notably, Shoot Al displayed no significant correlation coefficients with any elements, whereas Shoot Ni and Shoot Mn exhibited a few significant correlations: Shoot Ni and Shoot As, Shoot Mn and Shoot Cu, Shoot Mn and Shoot Cd, Shoot Mn and Shoot Rb, and Shoot Mn and Shoot Sr. Two pairs demonstrated strong positive correlations: Shoot K and Shoot Rb (r = 0.99) and Shoot Ca and Shoot Sr (r = 0.99). In the root, correlation coefficients varied, ranging from -0.47 to 0.89. Similar to the shoot, root Al, root Ni, and root Mn were largely uncorrelated with other root elements ($p > 0.05$). Significant correlation coefficients were identified for Root Ni and Root Sr, Root Mn and Root Co, Root Mn and Root Zn, and Root Mn and Root Rb, while Root Al exhibited no significant correlation coefficients with any element accumulations in the root. Root Cd and Root Zn correlated significantly with most element accumulations in the root, except for Root K (r = -0.16). Root Ca strongly correlated with Root Co (r = 0.88) and Root Sr (r = 0.89). Overall, shoot and root element concentrations were mostly uncorrelated ($p > 0.05$), although significant weak positive and negative correlations were observed ($r < 0.4$). Notably, strong significant correlations were discovered between shoot Mn concentration and root Fe concentration (r = -0.58) and between Shoot Mn and Root Mn concentration (r = 0.63). For Al, only a single significant correlation was observed between Shoot Al and Root Mo (r = 0.15). This robust correlation of elemental concentration within organs was also evident in the Principal Component Analysis (PCA) of the ionomic profile (Figure 3A). The first axis was predominantly influenced by element concentration in the shoot, while the second axis was attributed to concentration in the root (Figure 3B and 3C).

Table 1 *Overview of the variation for shoot and root element concentrations in ppm (mg per kg dry weight) and the element translocation factors in the N. caerulescens RIL population. Min – minimum; Max – maximum; SD – standard deviation*

Element	Shoot concentration (mg kg^{-1})			Root concentration (mg kg^{-1})			Translocation factor		
	Min	Max	Average ± SD	Min	Max	Average ± SD	Min	Max	Average ± SD
B	28.99	130.93	66.34 ± 19.43	34.86	408.29	81.28 ± 43.31	0.23	2.91	0.98 ± 0.42
Na	4.42	75.68	20.4 ± 12.57	52.2	1218.28	303.6 ± 176.52	0.01	1.12	0.1 ± 0.12
Mg	950.97	3739.67	2143.07 ± 535.71	904.33	3571.28	1420.38 ± 324.19	0.64	3.35	1.54 ± 0.45
Al	0.57	142.2	2.85 ± 10.33	5	5767.26	58.47 ± 442.45	0.0002	15.26	0.28 ± 1.28
P	1237.85	5542.36	2878.23 ± 747.18	3840.76	18347.15	6369.37 ± 1369.78	0.22	0.89	0.48 ± 0.15
S	4902.19	23242.61	12408.56 ± 3612.05	9994.56	112613.68	24541.83 ± 11936.9	0.16	1.54	0.58 ± 0.24
K	15161	78944.68	40119.18 ± 11374.93	4744.85	46903.01	23305.96 ± 9409.98	0.7	11.15	1.98 ± 1.05
Ca	2658.08	12819.27	7060.87 ± 2051.35	1763.58	10072.13	3392.79 ± 1186.28	0.52	5.56	2.4 ± 1.10
Fe	8.54	134.06	28.77 ± 15.14	120.59	1792.95	405.14 ± 244.26	0.01	0.39	0.09 ± 0.06
Mn	10.53	343.58	70.96 ± 50.07	8.68	1699.48	142.31 ± 191.53	0.05	4.09	0.9 ± 0.64
Co	0.0006	0.005	0.0022 ± 0.001	0.005	0.04	0.01 ± 0.005	0.03	0.78	0.22 ± 0.12
Ni	0.03	12.46	0.28 ± 0.92	0.17	60.59	0.94 ± 4.67	0.002	25.56	0.79 ± 2.29
Cu	0.52	3.97	1.52 ± 0.61	4.74	49.65	13.86 ± 5.69	0.03	0.51	0.13 ± 0.07
Zn	31.19	603.56	150.57 ± 77.82	17.2	864.28	100.18 ± 121.42	0.09	12.17	2.94 ± 2.5
As	0.0008	0.01	0.002 ± 0.001	0.0007	0.09	0.01 ± 0.01	0.05	4.08	0.34 ± 0.35
Rb	0.16	0.77	0.4 ± 0.11	0.14	0.84	0.29 ± 0.08	0.58	3.54	1.44 ± 0.44
Sr	0.44	1.72	0.96 ± 0.25	1.1	13.31	3.04 ± 1.59	0.08	1.27	0.42 ± 0.24
Mo	2.26	21.8	8.29 ± 3.69	14.02	98.72	41.34 ± 15.29	0.05	0.69	0.23 ± 0.11
Cd	1.65	47.22	15.54 ± 7.45	1.32	203.7	20.65 ± 20.77	0.03	12.25	1.99 ± 2.5

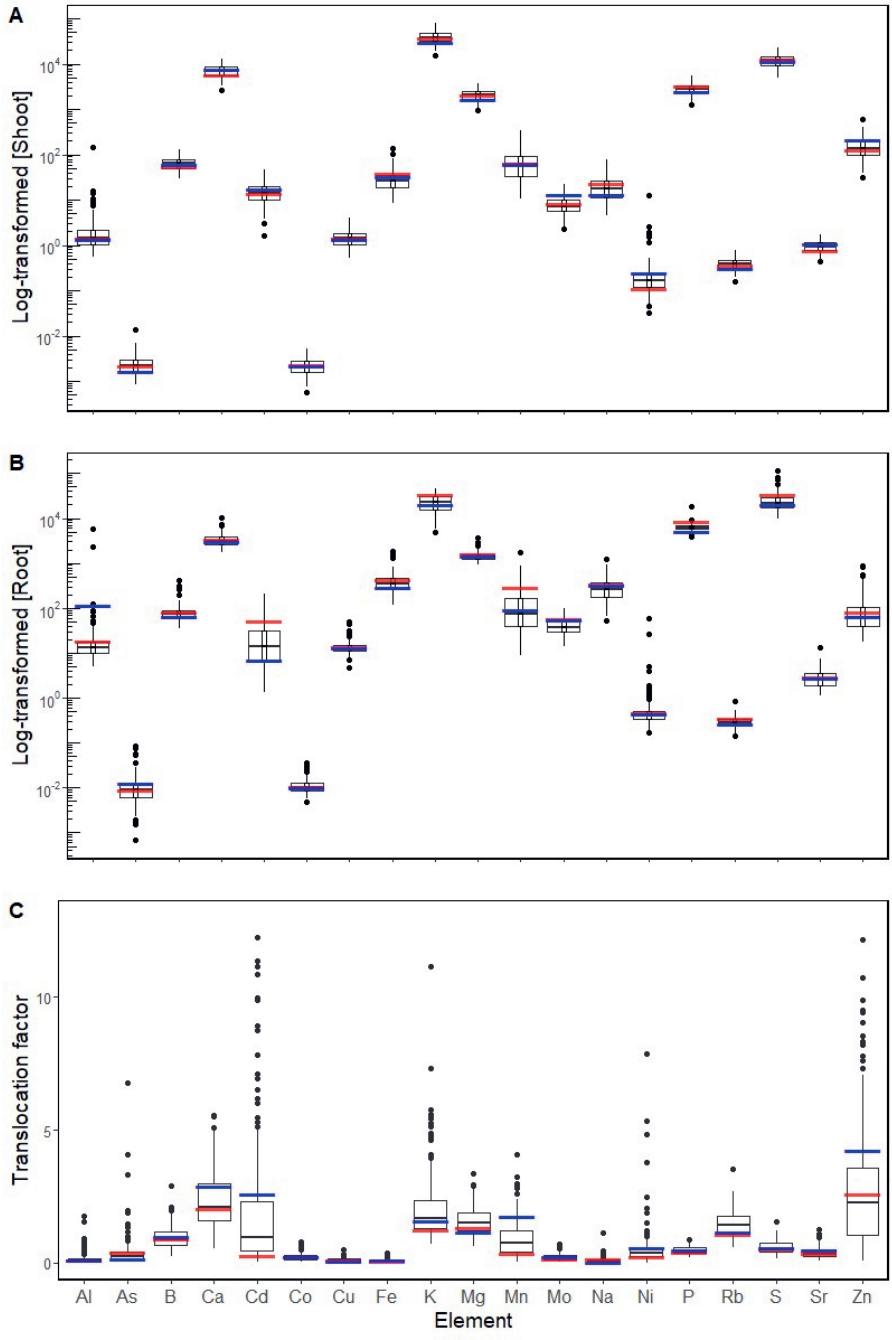


Figure 1. *Frequency distributions of log-transformed ionome element profiles and translocation factors (TF) observed in the RIL population exposed to 1 µM $CdSO_4$. (A), (B) Boxplots illustrating ionome profiles in shoot and root. (C) Boxplot showing TF ionome profile. Red and blue bars indicate average phenotypic values of RIL parental lines Ganges and Lellingen.*

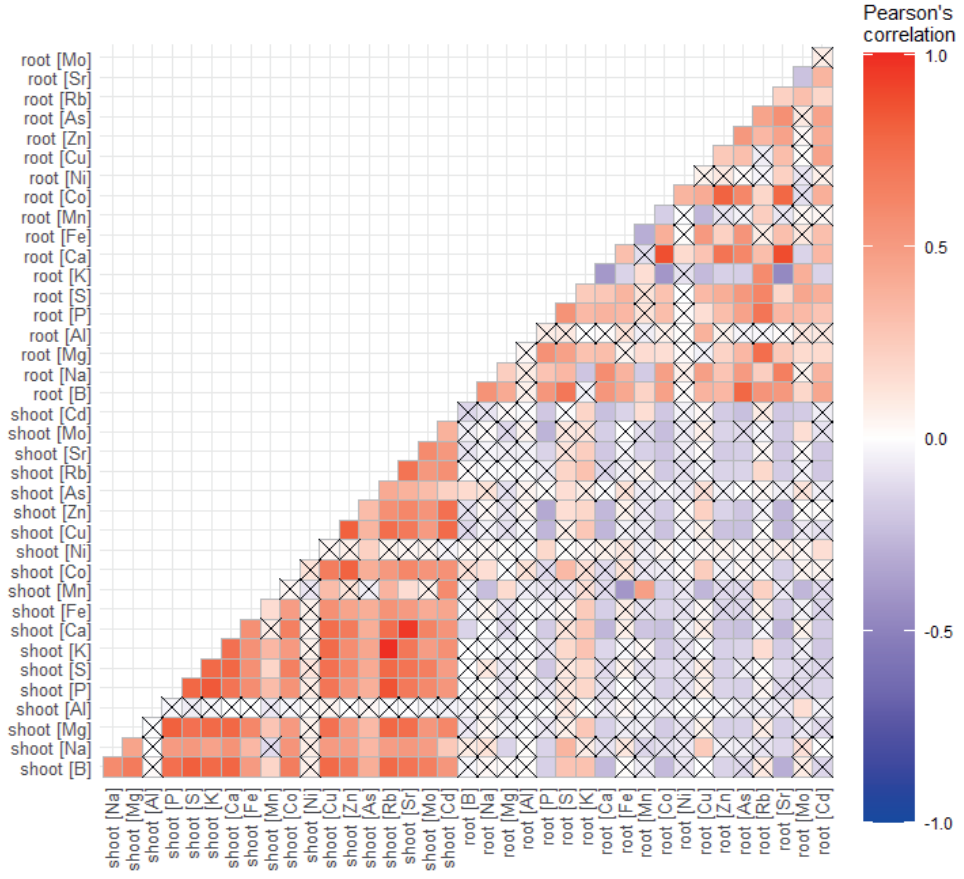


Figure 2. *Pairwise correlation matrix for shoot and root concentrations of 19 elements from the Cd accumulation test. The Pearson's correlation coefficient is indicated using a coloured heat map, with blue boxes indicating a negative correlation, and red boxes indicating a positive correlation. The lighter tone is the lower coefficient value. The crossed cells indicate non-significant correlations at p > 0.05.*

Cd tolerance test

The sequential exposure test of plants to 50 – 350 µM $CdSO_4$ was conducted to determine the degree of tolerance, expressed as the EC100. The average EC100 of GA and LE were 219.23 and 185.71 µM $CdSO_4$, respectively. The EC100 in the RIL population ranged between the full range of 50 to 350 µM $CdSO_4$, displaying a continuous and normal distribution (Figure 4B). Transgression beyond the parental lines on both directions is obvious. 47 RIL lines were more sensitive than LE (EC100 < 185.71 µM $CdSO_4$), while 136 RIL lines showed higher tolerance than GA (EC100 > 219.23 µM $CdSO_4$) (Figure 4A). The broad sense heritability of EC100 was calculated to be 20%.

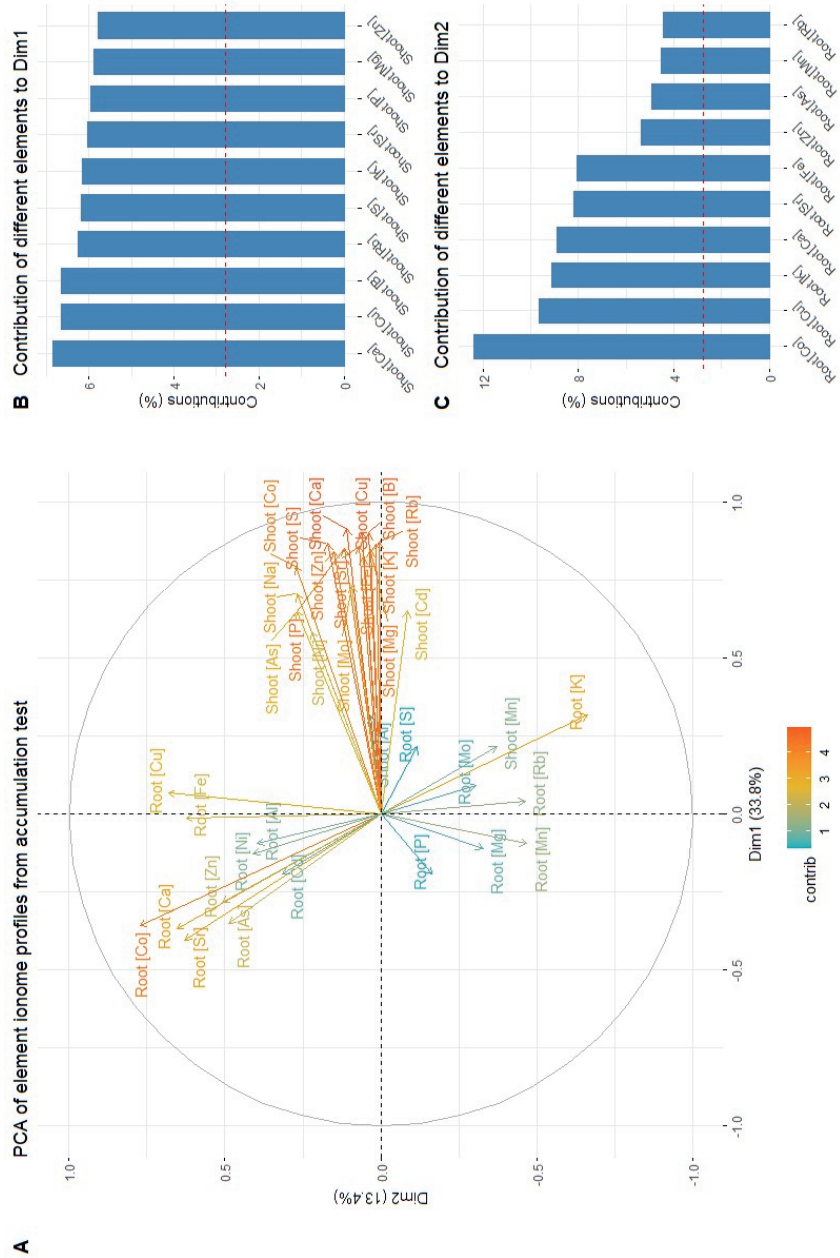


Figure 3. *Principle component analysis (PCA) of all log-transformed element concentrations in shoot and root of the Cd accumulation test. (A) PCA plot, of which each arrow is labeled with the percentage of contribution to dimension (Dim)1 and Dim2. (B), (C) Contribution of element ionome profile to Dim1 and Dim2, respectively.*

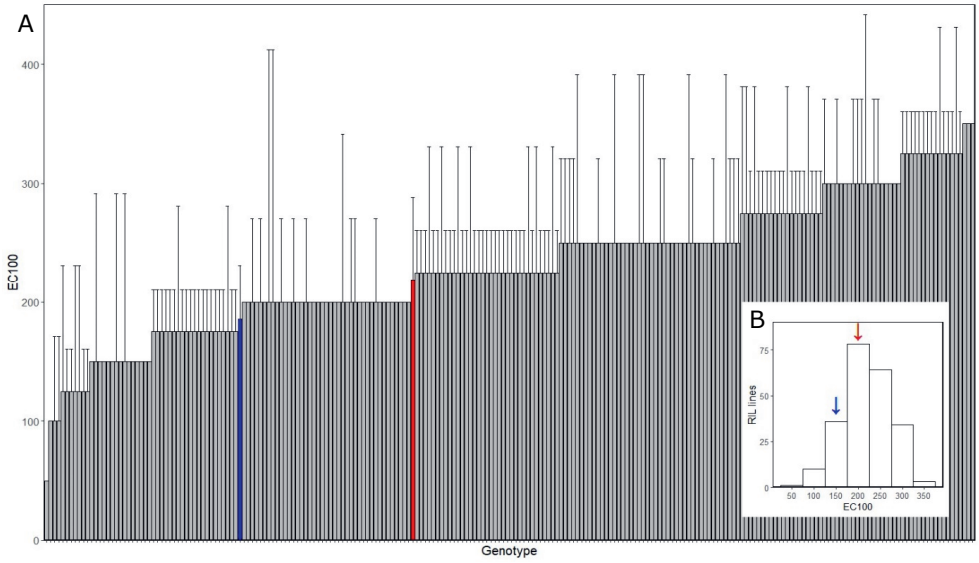


Figure 4. *Variation of the EC100 as reference of Cd tolerance for root growth in the RIL population. (A) Bar plot shows EC100 per genotype, which is sorted according to an increasing EC100 (from left to right). Error bars indicate standard deviations. Red and blue bars represent RIL parental lines Ganges and Lellingen. (B) Frequency distribution of EC100, which the majority of the RILs displaying EC100 values above 200 µM Cd. Red and blue arrows indicate RIL parental lines Ganges and Lellingen*

RIL Growth responding to Cd

The RIL population was exposed to 15 µM $CdSO_4$ to observe the temporal effect of Cd on plant growth. The average PLA of the Cd treatment was lower than that of the no-Cd control, although these differences were not significant (Figure 5A). Other than the PLA reduction, no toxicity symptoms were detected. After 26 DAS, a pattern of reduced PLA in the evening was observed, which is due to leaf movement, leaves move to an erect stature at night. This PLA reduction became most pronounced towards the end of the experiment. The daily maximum PLA was most often observed at noon (12 AM). The broad-sense heritability of PLA (H^2-PLA) broadly ranged from 35.5 to 73.6% (Figure 5B), and were most often above 50%. Some drops in H^2-PLA values were found on 19 DAS, 30 DAS and 37 DAS. The first two incidences (at 19 DAS and 30 DAS) occurred before the dates of watering plants. During the end of the experiment, the PLA often exceeded 10000 pixels, with few plants already overlapping. This is why the heritability values declined after 35DAS (Figure 5B).

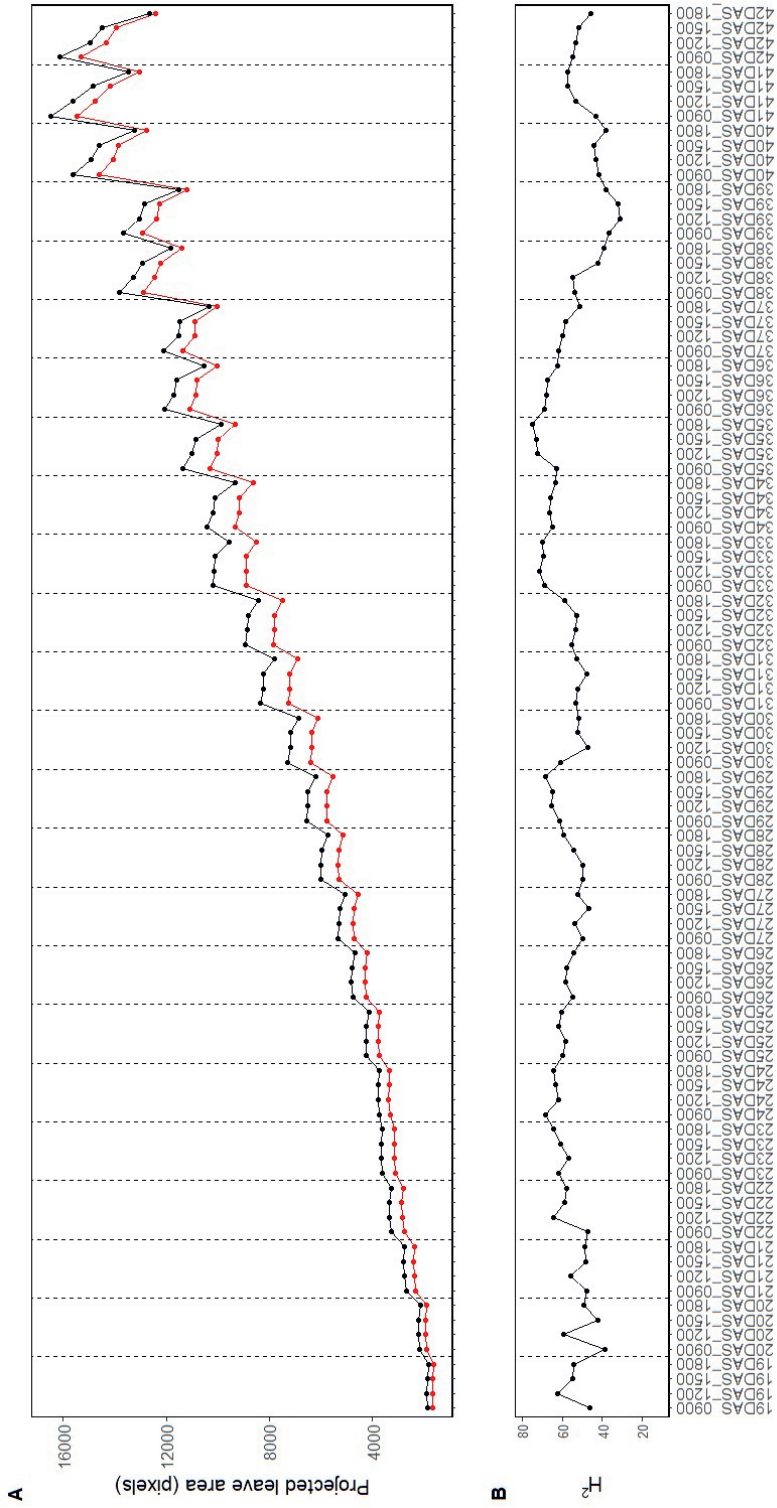


Figure 5. *RIL projected leaf area (PLA) in response to Cd exposure. (A) The average PLA of the RIL population growing from 19 days after sowing (DAS) to 42 DAS, either under control nutrient supply (black line) or with additional 15 µM $CdSO_4$ (red line). (B) Broad-sense heritability of PLA (H^2-PLA) in the RIL population over time. Each day is separated by a dashed line. Timepoints of both figures are labeled at the bottom.*

QTL mapping

QTL analysis for all traits was carried out using interval mapping and MQM approaches based on the 433 well-distributed SNP markers. The interval mapping approach was employed to preliminary scan for potential QTLs across the genome for all traits. Permutation testing determined LOD score significance thresholds, mostly close to 2.6 or above (Figure 5). The interval mapping approach identified a total of 25, 1 and 4 QTLs, associated with traits derived from the accumulation test (e.g. element concentrations and TFs), tolerance test and growth response experiment, respectively (Table 2). Thereafter, the MQM mapping approach, utilizing the most significant true markers as cofactors, was implemented to enhance QTL detection accuracy and to narrow QTL regions for the traits with significant QTLs. The new LOD score thresholds were lower than the threshold of the interval mapping approach (Figure 5).

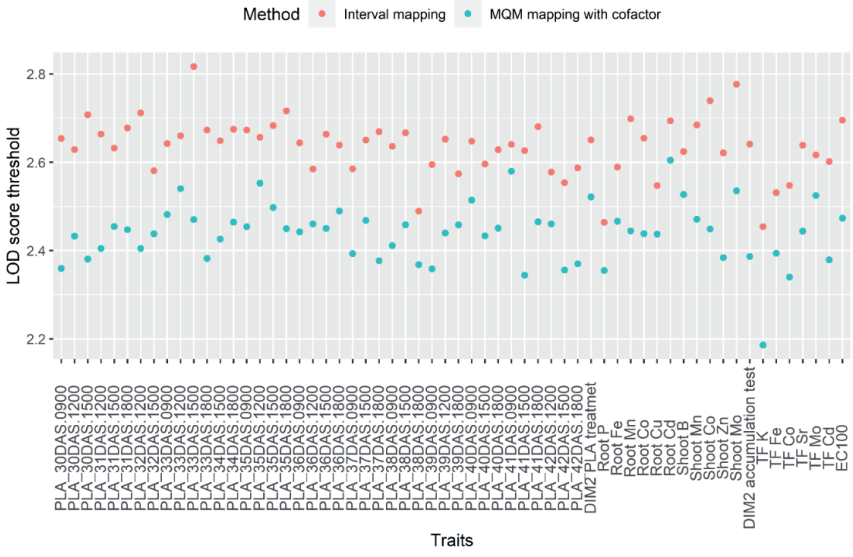


Figure 5. *LOD thresholds to determine significant QTLs in interval mapping (orange) and MQM mapping approaches (blue). The thresholds are based on 1000 permutations. PLA – projected leaf area; TF – translocation factor; LOD – Logarithm of Odds; Root – root concentration; Shoot – shoot concentration.*

The MQM mapping approach detected a total of 27 QTLs across three experiments (Table 2), excluding three QTLs (qRT-Co, qRT-Cu-2 and qSH-Mn-2) identified in the interval mapping approach. In addition, the MQM mapping approach detected two additional QTLs (qRT-Cd-2 and qPC2-CONC-3). In comparison to the interval mapping, changes of LOD scores were observed in MQM mapping-identified QTL. For example, the LOD score of qRT-Fe-3 reduced from 4.4 to 3.8, while the LOD score of qRT-Cd-1 increased from 13.3 to 16.3 (Table 2). The LOD scores of QTLs identified by MQM mapping ranged between 2.5 to 16.3, explaining 5.7 to 31.2 % of phenotypic variance (Table 2). Of these QTLs, a total of 22 significant QTLs were mapped for different element concentration traits (Table 2). The most significant QTL, with a LOD score of 16.3, was a major Root Cd QTL, qRT-Cd-1, located between 90 to 92.2 cM on LG2, which contributed to 31.2 % of Root Cd variation. Another Root Cd QTL, qRT-Cd-2, was mapped to LG1 showing a lower LOD score, referred after as minor Root Cd QTL. A strong QTL was found for shoot Mn with a LOD score of 10, located on LG 3. PCA analysis, focusing on PC values explaining the variance of log-transformed shoot and root concentrations, revealed that the first two dimensions (DIM1 and DIM2)

accounted for most variances in the Cd accumulation test. No QTLs were detected for the PC value of DIM1, while three QTLs, qPC2-CONC-1, qPC2-CONC-2 and qPC2-CONC-3, were identified for DIM2 on LG2, LG5 and LG3 (Table 2). In the Cd tolerance test, a single QTL, qEC100, was identified with a LOD score of 4.9, located between 36.1 to 62.6 cM on LG2 (Table 2). Allelic effects were observed, with the GA allele enhancing certain traits such as Root Co, Root Fe, Root Mn, Shoot P, Shoot Mn, and EC100. The LE allele had enhancing effects on Root Mo, Shoot B, Shoot Mo, Shoot Zn, and all TF traits (Table 2). The major and minor Root Cd QTL was influenced by different parental alleles on each QTL. The GA allele of qRT-Cd-1 increased Root Cd, while the LE allele of qRT-Cd-2 increased Root Cd (Table 2).

For growth response to Cd, three QTL colocalizations were detected during the period from 30 DAS to 42 DAS (Figure 6). These colocalized QTLs were designated as qPLACd-1, qPLACd-2, and qPLACd-3, located on LG2, LG4, and LG6 (Figure 6, Table 2). These colocalized QTLs accounted for 6.9 – 11.7 % of PLA variance at each timepoint. Overall, the LOD scores for these QTLs ranged from 2.4 to 4.5. The PCA analysis for PLA responding to Cd revealed one QTL, qPC2-Cd with LOD scores of 3.6, identified between 15 to 60.9 cM on LG4.

Apart from growth response to Cd, colocalizations of QTLs were also observed for other traits in this study. Of 27 QTLs, only five (qRT-Cd1, qPLACd-3, qRT-P, qPC2-CONC-2, and qPC2-CONC-3) did not overlap with any other QTL (Figure 7). QTLs associated with the same element traits (either TF or shoot/root concentration) were found to colocalize on the same LGs. The QTLs of Shoot Mo and Root Mo colocalized on LG6, a colocalized QTL of Shoot Mn and Root Mn was observed on LG3, and qRT-Cd-1 colocalized with qTF-Cd on LG2 (Figure 7).

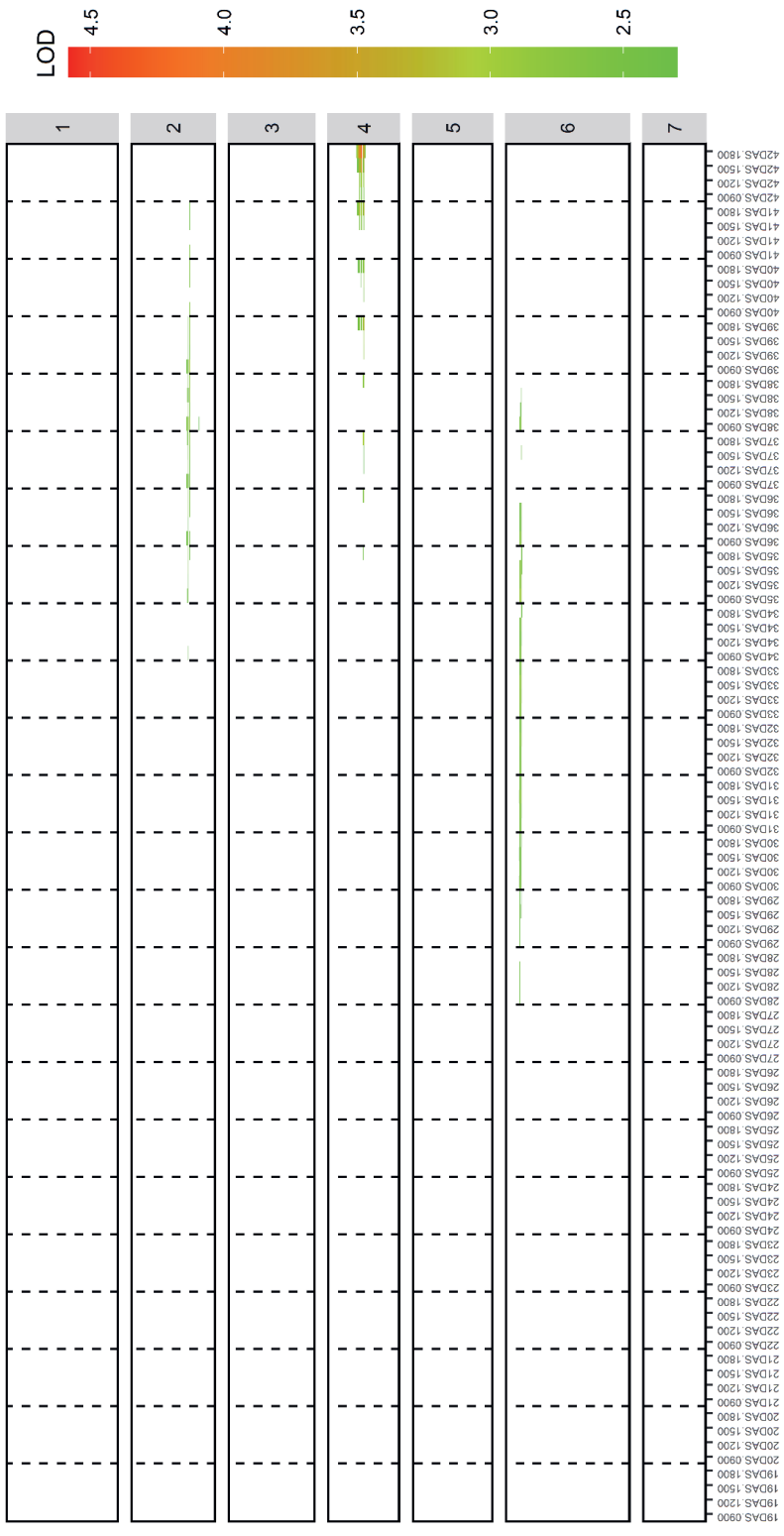


Figure 6. *Heatmap showing QTLs detected for plant growth response to Cd, based on projected leaf area determined at four timepoints per day, using the interval mapping approach. Rows represent chromosomal locations, based on SNP markers mapped to each of the seven N. caerulescens linkage groups. Columns represent the timepoints. Dash lines indicate the breaks between days after sowing (DAS). Significant QTLs are indicated with a colour scale corresponding to the QTL LOD score.*

Table 2. *QTLs identified for different traits in three experiments using the interval mapping and MQM mapping approaches. A total of 28 and 27 QTLs were detected by these two approaches, respectively. Of these, 27 QTLs were detected first by interval mapping, then narrowed down by the MQM mapping approach.*

Trait	QTL	LG	Nearest Marker	Interval mapping approach		MQM mapping approach		% PVE	Trait enhancing allele
				LOD score	QTL cM (CI)	LOD score	QTL cM (CI)		
Cd accumulation test									
Root Cd	qRT-Cd-2	1	Contig216_973907	ND	ND	3	105 (103.6-106.6)	6.6	LE
	qRT-Cd-1	2	Contig270_912010	13.3	91 (89-93)	16.3	90 (90-92.2)	31.2	GA
Root Co	qRT-Co	5	Contig61_351449	2.7	45.7 (45.7-46)	ND	ND	-	GA
Root Cu	qRT-Cu-1	2	Contig5_499375	2.7	86 (85-90)	3.4	90 (85-95.5)	7.8	GA
	qRT-Cu-2	3	Contig7_2499709	2.7	66.1 (64.8-67)	ND	ND	-	LE
	qRT-Cu-3	5	Contig5_499375	4.4	24.7 (10-82)	3.8	25 (10-46.1)	8.5	GA
Root Fe	qRT-Fe-1	2	Contig35_742550	4.4	74.8 (42.5-82)	4.3	50 (44.1-76.2)	9.8	GA
Root Fe	qRT-Fe-2	5	Contig7_4231558	3.7	31 (21-56)	2.5	30 (27.5-30.2)	5.7	GA
Root Mn	qRT-Mn	3	Contig256_129442	4.7	28.1 (24-36)	5.2	30 (26.7-35)	10.6	GA
Root Mo	qRT-Mo	7	Contig12_1499740	2.9	59.2 (57-77.3)	2.8	70 (60-75)	5.9	LE
Root P	qRT-P	7	Contig273_45152	2.9	42 (37.7-52.2)	2.5	50 (45-50.7)	5.8	GA
Shoot B	qSH-B	4	Contig291_508109	2.7	38.4 (36-42)	2.6	35 (33.2-35.6)	6.1	LE
Shoot Mn	qSH-Mn-1	3	Contig105_472484	10.3	28.1 (26-31)	10	30 (26.7-32.8)	19.9	GA
	qSH-Mn-2	5	Contig58_1159183	3.1	57 (34-60)	ND	ND	-	LE
Shoot Mo	qSH-Mo	7	Contig12_3499035	4	71 (62-82)	4.3	70 (65-83)	8.8	LE
Shoot Zn	qSH-Zn	4	Contig0_1499902	2.8	93 (91-94)	2.6	90 (90-91.4)	5.8	LE
TF Cd	qTF-Cd	2	Contig270_912010	5.9	91 (82-98)	6.9	90 (85-95)	10.1	LE
TF Co	qTF-Co	4	Contig0_6495932	3.7	66 (57.5-95)	3.8	70 (56.1-95)	8.7	LE
TF Fe	qTF-Fe	2	Contig189_508757	4	74.8 (47-83)	4	75 (48.3-81.1)	9.2	LE
TF K	qTF-K	2	Contig270_496202	2.7	95 (85.1-96)	2.9	95 (88.4-95.5)	6.4	LE
TF Mo	qTF-Mo	2	Contig52_498228	3	95.5 (95-97)	2.6	95 (95-95.5)	5.9	LE
TF Sr	qTF-Sr	4	Contig53_1275806	2.8	60.9 (60.9-62.3)	2.8	60 (58.3-65)	6.3	LE
PC value DIM2	qPC2-CONC-1	2	Contig52_498228	4.2	85.1 (47.9-102)	4.5	95 (74.8-101.2)	10.2	LE

Continued

Trait	QTL	LG	Nearest Marker	Interval mapping approach		MQM mapping approach		% PVE	Trait enhancing allele
				LOD score	QTL cM (CI)	LOD score	QTL cM (CI)		
PC value DIM2	qPC2-CONC-3	3	Contig152_444152	ND	ND	2.8	65 (64.8-66.1)	6.4	LE
PC value DIM2	qPC2-CONC-2	5	Contig58_499084	3	52 (32-60)	2.5	60 (60-62.8)	5.8	GA
Cd tolerance test									
EC100	qEC100	2	Contig35_742550	4.8	48.8 (35-63)	4.9	50 (36.1-62.6)	10.6	GA
Growth response to Cd									
PLA treatment	qPLACd-1	2	Contig312_199473	2.7 - 3	45 (35-44.1)	2.9 - 3.2	45 (35-48.3)	6.9 - 8.7	GA
	qPLACd-2	4	Contig0_5498985	2.7 - 3.6	70.5 (17-71)	3.1 - 4.5	70 (67.3-70.5)	7.2 - 11.7	LE
	qPLACd-3	6	Contig49_751580	2.7 - 3	54 (46-56)	2.4 - 3.3	50 (45-56)	7.5 - 8.3	LE
DIM2 PLA treatment	qPC2-Cd	4	Contig103_499486	3.9	40 (14-73)	3.6	40 (15-60.9)	10.1	GA

ND – not detected; QTL – quantitative trait locus; LG - linkage group; LOD - maximum LOD score; % PVE - percentage phenotypic variation explained; CI – confidential interval; LOD - logarithm of odds; cM – centimorgan; PLA – projected leaf area; GA - Ganges; LE - Lellingen; SH- shoot; RT - root; TF – translocation factor; PC -principle component; Shoot x – shoot x concentration; Root x – root x concentration.

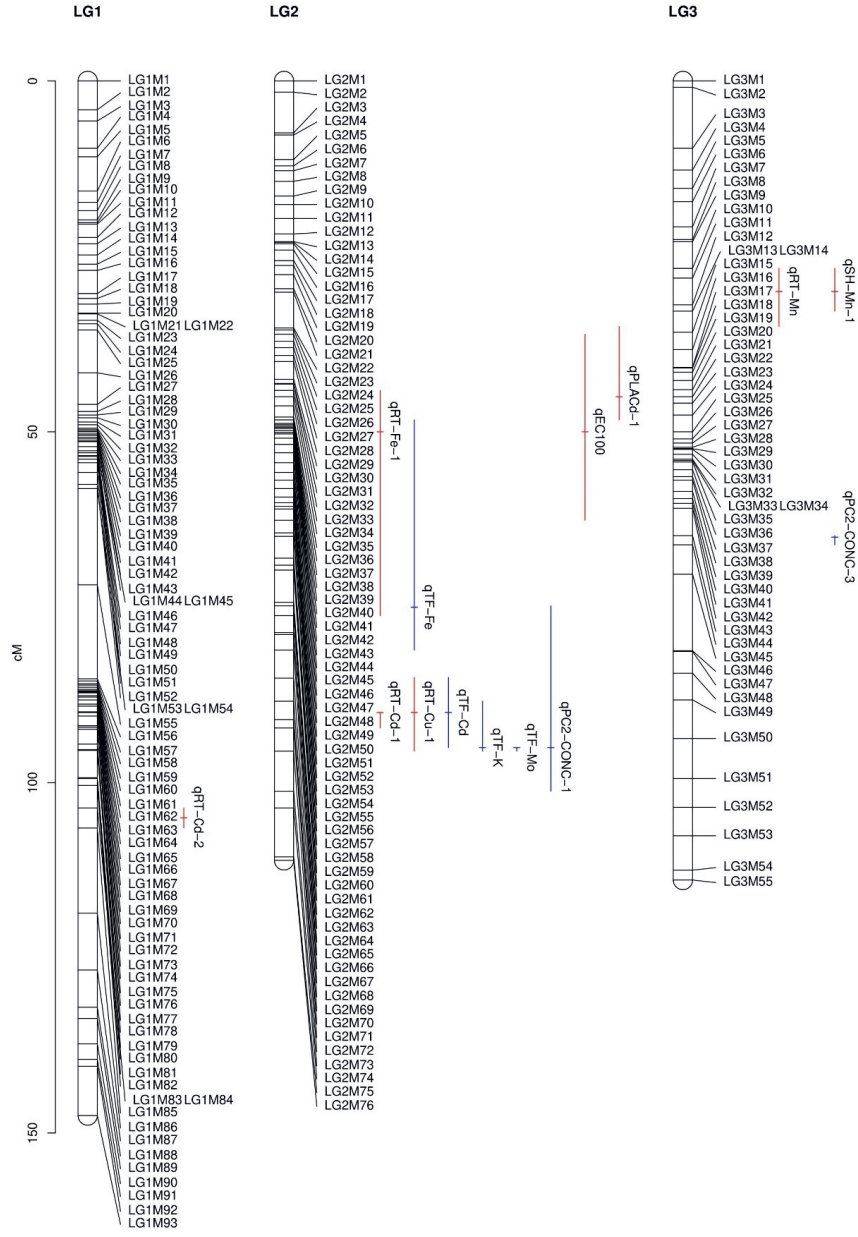
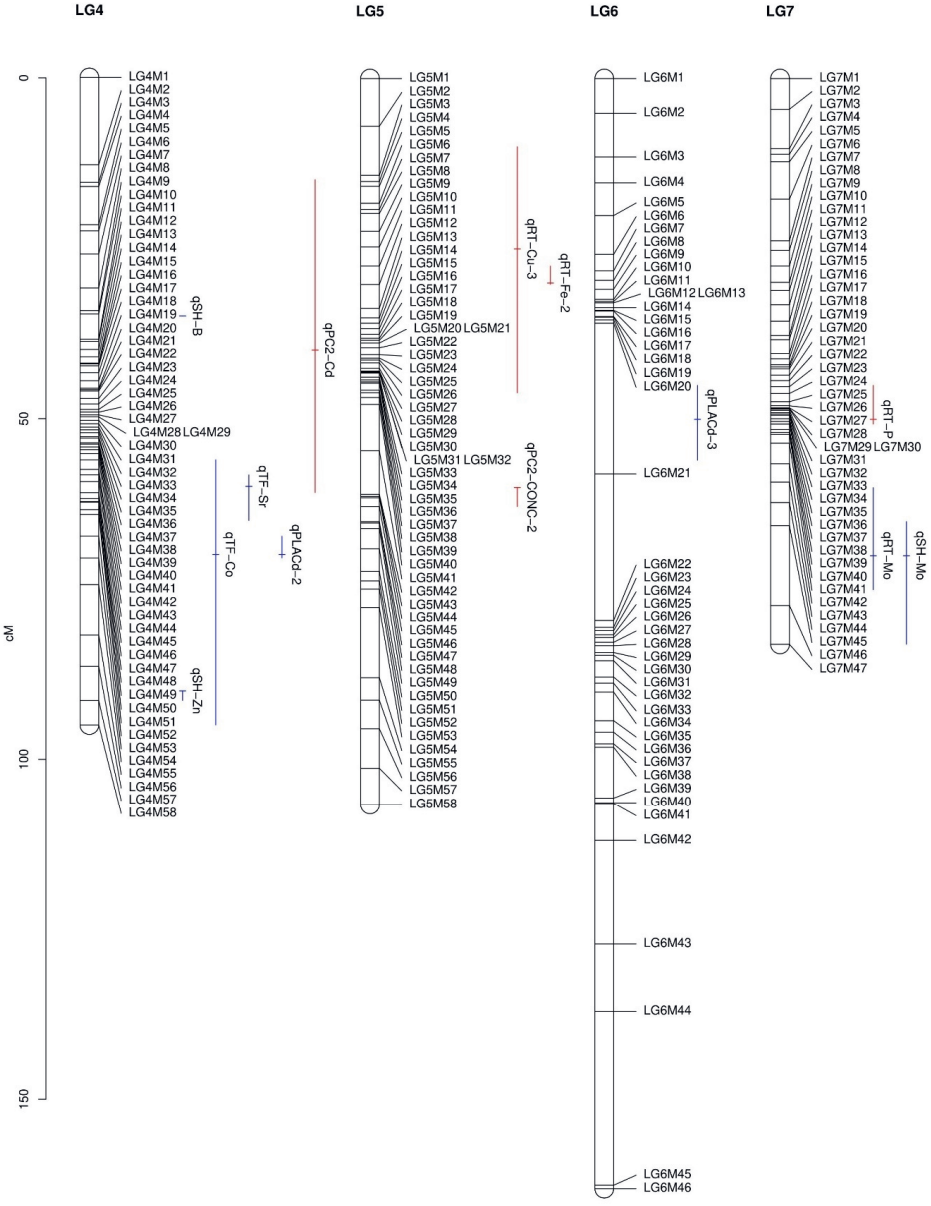


Figure 7. *Genetic map of N. caerulescens showing an overview of 27 QTLs detected by MQM mapping approach in the recombinant inbred line population. Colocalization of the QTLs are detected in each linkage group (LG), except LG1 and LG6. The ruler indicates the genetic distances for the seven LGs in cM. Vertical bars indicate confidential intervals of each QTL. Horizontal bars indicate peaks of each QTL. The QTL bars are coloured corresponding to the origin of the trait-enhancing allele, either Ganges (red) or Lellingen (blue). LG1 – LG3 (Top); LG4 – LG7 (Below).*


LG4
LG5
LG6
LG7
cM
0
50
100
150
LG4M1
LG4M2
LG4M3
LG4M4
LG4M5
LG4M6
LG4M7
LG4M8
LG4M9
LG4M10
LG4M11
LG4M12
LG4M13
LG4M14
LG4M15
LG4M16
LG4M17
LG4M18
LG4M19
LG4M20
LG4M21
LG4M22
LG4M23
LG4M24
LG4M25
LG4M26
LG4M27
LG4M28 LG4M29
LG4M30
LG4M31
LG4M32
LG4M33
LG4M34
LG4M35
LG4M36
LG4M37
LG4M38
LG4M39
LG4M40
LG4M41
LG4M42
LG4M43
LG4M44
LG4M45
LG4M46
LG4M47
LG4M48
LG4M49
LG4M50
LG4M51
LG4M52
LG4M53
LG4M54
LG4M55
LG4M56
LG4M57
LG4M58
qSH–B
qPC2–Cd
qTF–Sr
qPLACd–2
qTF–Co
qSH–Zn
LG5M1
LG5M2
LG5M3
LG5M4
LG5M5
LG5M6
LG5M7
LG5M8
LG5M9
LG5M10
LG5M11
LG5M12
LG5M13
LG5M14
LG5M15
LG5M16
LG5M17
LG5M18
LG5M19
LG5M20 LG5M21
LG5M22
LG5M23
LG5M24
LG5M25
LG5M26
LG5M27
LG5M28
LG5M29
LG5M30
LG5M31 LG5M32
LG5M33
LG5M34
LG5M35
LG5M36
LG5M37
LG5M38
LG5M39
LG5M40
LG5M41
LG5M42
LG5M43
LG5M44
LG5M45
LG5M46
LG5M47
LG5M48
LG5M49
LG5M50
LG5M51
LG5M52
LG5M53
LG5M54
LG5M55
LG5M56
LG5M57
LG5M58
qRT–Cu–3
qRT–Fe–2
qPC2–CONC–2
LG6M1
LG6M2
LG6M3
LG6M4
LG6M5
LG6M6
LG6M7
LG6M8
LG6M9
LG6M10
LG6M11
LG6M12 LG6M13
LG6M14
LG6M15
LG6M16
LG6M17
LG6M18
LG6M19
LG6M20
LG6M21
LG6M22
LG6M23
LG6M24
LG6M25
LG6M26
LG6M27
LG6M28
LG6M29
LG6M30
LG6M31
LG6M32
LG6M33
LG6M34
LG6M35
LG6M36
LG6M37
LG6M38
LG6M39
LG6M40
LG6M41
LG6M42
LG6M43
LG6M44
LG6M45
LG6M46
qPLACd–3
LG7M1
LG7M2
LG7M3
LG7M4
LG7M5
LG7M6
LG7M7
LG7M8
LG7M9
LG7M10
LG7M11
LG7M12
LG7M13
LG7M14
LG7M15
LG7M16
LG7M17
LG7M18
LG7M19
LG7M20
LG7M21
LG7M22
LG7M23
LG7M24
LG7M25
LG7M26
LG7M27
LG7M28
LG7M29 LG7M30
LG7M31
LG7M32
LG7M33
LG7M34
LG7M35
LG7M36
LG7M37
LG7M38
LG7M39
LG7M40
LG7M41
LG7M42
LG7M43
LG7M44
LG7M45
LG7M46
LG7M47
qRT–P
qRT–Mo
qSH–Mo

Candidate genes associated with metal accumulation, Cd tolerance and Cd response

The whole-genome alignment of the 300 contigs covering the seven LGs of *N. caerulescens* and the *A. thaliana* "TAIR10" reference genome was employed to identify the *N. caerulescens* homologues of *A. thaliana* genes. In total, 13225 microsyntenic blocks containing 23044 *A. thaliana* genes were identified, which are distributed across the five chromosomes of *A. thaliana* with some gaps (Figure 8).

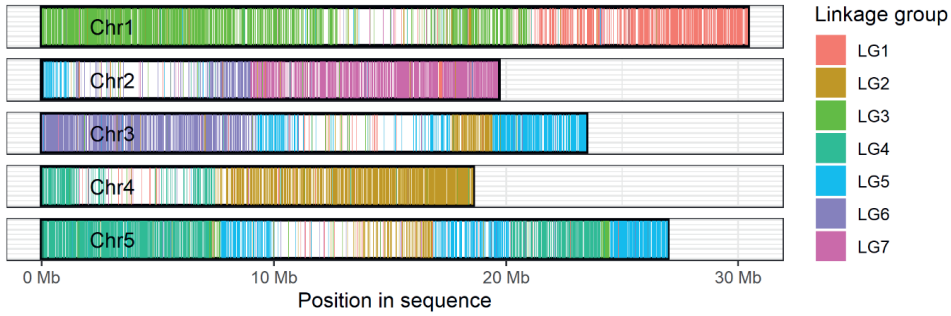


Figure 8. *Whole genome alignment of the seven linkage groups (LG) of N. caerulescens and the five A. thaliana chromosomes. The five bars represent the five chromosomes of A. thaliana. Colours indicate the seven LGs.*

To elucidate candidates for further functional analysis, the detected QTLs (Figure 7, Table 2) were examined for Arabidopsis homologues and their GO terms. A total of 9504 protein-encoding homologues were obtained. The GO terms of these homologues were filtered using relevant key words, such as ion binding, homeostasis and ion transport, to select for potential candidate genes related to the response to Cd. As a result, 1195 homologues were identified, of which 867 homologues were well described for their functions in *A. thaliana*. These include well known metal-related genes, such as *HMAs*, *NRAMPs*, *ZIPs*, *IRT1*, *HEAVY METAL-ASSOCIATED ISOPRENYLATED PLANT PROTEIN* (*HIPP*), K^+ *UPTAKE PERMEASE* (*KUP8*), *YELLOW STRIPE LIKEs* (*YSL*), *ZINC INDUCED FACILITATOR 1* (*ZIF*1) and *MAGNESIUM TRANSPORTER 7* (*MGT7*).

While well-known genes and proteins related to metal response were recognized within the detected QTLs, identification for candidate genes for each trait with certainty proved challenging due to the broad genomic regions covered by most QTLs. Consequently, a detailed examination was conducted in the major Root Cd QTL, qRT-Cd-1, which displayed both the highest LOD score and a small confidence interval. A total of 16 potential candidate genes were identified, nine of which were well documented, including *HMA3*, *ATP-BINDING CASSETTE E3* (*ABCE3*), H^+*-ATPase2* (*HA2*), *LEUCYL AMINOPEPTIDASE 2* (*LAP2*), *CYTOCHROME P450 FAMILY 71 SUBFAMILY B POLYPEPTIDE 2* (*CYP71B2*), *MITOCHONDRIAL RNA EDITING FACTOR 29* (*MEF29*), *REI1-LIKE 1* (*REIL1*), *RS-CONTAINING ZINC FINGER PROTEIN 22* (*RSZ22*), and *SENESCENCE-ASSOCIATED RECEPTOR-LIKE KINASE* (*SARK*). Through comparison with reported metal-related genes from other studies, a well-documented Cd-related gene, *HMA3*, localized to contig270, was identified for Cd accumulation in the root (Figure 9). In addition, *HMA3* was in the QTL colocalizing Root Cd and TF Cd, suggesting that HMA3 contributed to root-to-shoot translocation of Cd as well (Figure 9).

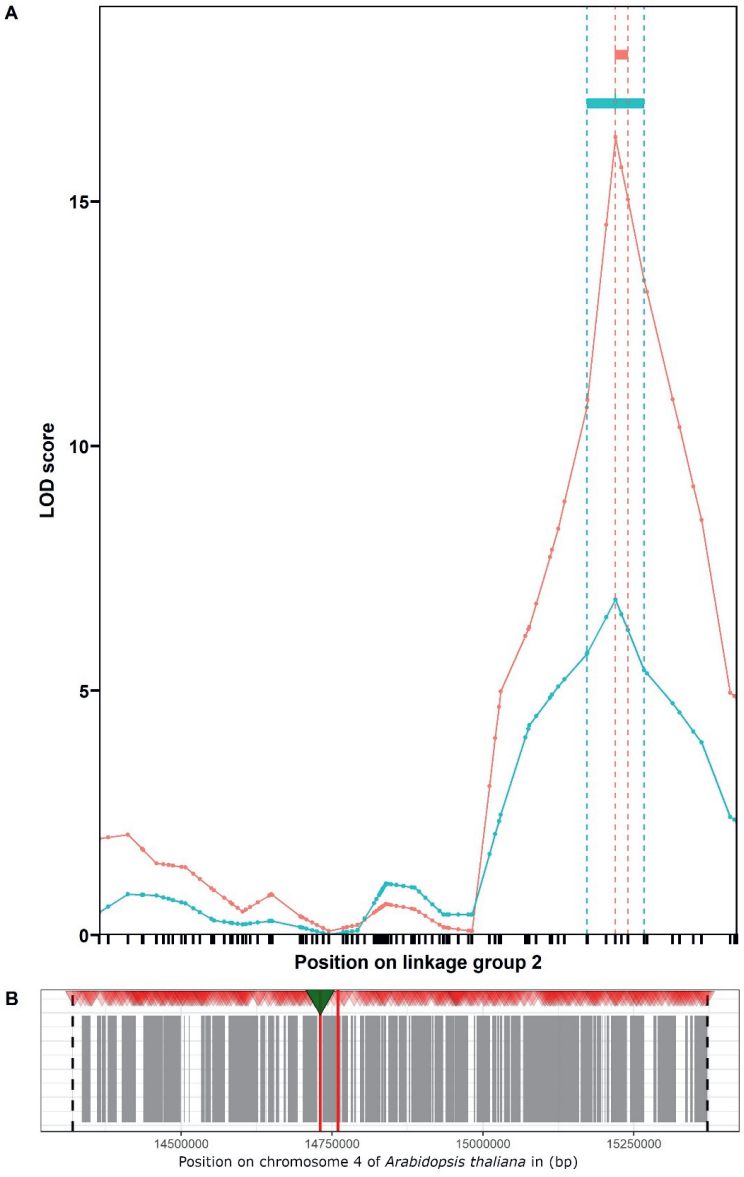


Figure 9. *The N. caerulescens homologue of the A. thaliana HMA3 gene is identified to associate with colocalized QTLs for Root Cd (qRT-Cd-1) and TF Cd (qTF-Cd). (A) Distribution of LOD scores for Root Cd (red) and TF Cd (blue) on linkage group 2 (LG2). Red and blue boxes correspond to each QTL, with red and blue vertical dashed lines indicating the markers delineating the significance intervals for each QTL. (B) Syntenic blocks of both qTF-Cd and qRT-Cd-2 are aligned to chromosome 4 of A. thaliana between 14320201 to 15372328 bp. Grey boxes indicate alignment between A. thaliana and N. caerulescens genome sequences. Red triangles indicate the positions of A. thaliana genes. The homologue of HMA3 (green triangle) is detected in the syntenic block corresponding to the end of contig 270 (between red vertical lines) of N. caerulescens.*

Discussion

Cd accumulation in the RIL population

The accumulation and root-to-shoot translocation of a suite of elements was studied under Cd exposure in the *N. caerulescens* RIL population, including the parental lines GA and LE. A relatively low Cd concentration of 1 µM $CdSO_4$ was applied, which corresponds to Cd concentrations found in mildly Cd-contaminated sites (Muller, 1969). In this way, we expected the accumulation test to reveal as much variation of Cd accumulation as possible, and avoid unwanted toxicity responses coming from the Cd-sensitive LE genetic background. Indeed our analysis indicated that GA accumulated more Cd than LE, in both shoot and root tissues. GA accumulated more Cd in the roots than in the shoots (Table S1), which is similar to a previous study (Deniau et al., 2006). In comparison to our study, Deniau et al. (2006) exposed samples to a higher Cd level (5 µM $CdSO_4$), resulting in much higher Cd accumulation (7.52 µmol g^{-1} or 845.33 mg kg^{-1} in root, 3.41 µmol g^{-1} or 383.32 mg kg^{-1} in shoot). In our experiment, LE accumulated more Cd in shoots than in roots, as was observed before (Assunção et al., 2003b), but the concentrations we found were several magnitudes lower than reported before (~ 10 µmol g^{-1} or 1121.14 mg kg^{-1} in both shoot and root). Likewise, also the RILs exhibited lower concentrations of Zn and especially Cd than in the previous studies. These large differences are likely to be caused by differences in plant size and differences in Cd exposure concentration. In the reported studies, *N. caerulescens* was pre-grown in moist peat for three weeks before exposing plants to Cd treatment. This leads to much larger plants at the start of the experiment than when growing plants on agar media as we did. Furthermore, the Cd concentration used to challenge the plants in the reported studies was five times higher than what we used. Both combined, provided a much larger capacity to accumulate Cd than in our experiment.

LE accumulated more Cd in shoot than root, which is opposite to that of GA. Consequently, the Cd TF was higher in LE (259.33 %) than GA (31.16 %). This suggests there to be a main difference between both accessions on root Cd sequestration and root-to-shoot Cd translocation. A previous study of Cd influx in GA protoplasts has shown that foliar Cd accumulation may be controlled by root transport processes (Cosio et al., 2004). Moreover, Lombi et al. (2001) provided evidence that Cd transporters are upregulated under low Cd in GA. Thus it seems likely to assume that in GA, low Cd induces the expression of transporters involved in the influx of Cd into the vacuoles of root cells. This will keep Cd in the roots and will limit Cd root-to-shoot translocation and accumulation in the shoots. The Cd accumulation in LE is likely to be performed by a Cd uptake system with a lower affinity to Cd than in GA (Assunção et al., 2003b). Our results show that LE accumulates more Zn in shoot than GA, while shoot Cd accumulation is not significantly different between GA and LE ($p > 0.05$). As Cd and Zn may be competing for the same transporters, the higher affinity for Zn could restrict Cd root-to-shoot translocation at low Cd supply, in agreement with the findings of Assunção et al. (2003b). The identification of *HMA3* as the major candidate underlying the qRT-Cd-1 QTL of root Cd concentration, further discussed below, suggests this gene to be one of the components contributing to this difference between GA and LE.

The phenotypic assessment of the RIL population for element concentrations, showed that when the shoot and root element concentration data were log-transformed, they mostly displayed a continuous and normal distribution, very much in accordance with a polygenic quantitative nature of the variation. For both roots and shoots, the element concentrations are mostly significantly positively correlated. For example, there were significant correlations for shoot Zn and Cd concentrations ($r = 0.73$); and shoot Mn

and Cd concentrations (r = 0.59), which agrees with previous findings (Assunção et al., 2003a; Deniau et al., 2006; Zha et al., 2004). However, when comparing the concentrations of the same elements in root and shoot, the correlation was often not significant. So far, largely positive correlations among elements within organ have not been reported in other metal hyperaccumulators and non-hyperaccumulator. In *A. halleri*, both positive and negative correlations were reported between Cd and three elements (Fe, Mg, and Ca) when exposing to 5 µM Cd (Meyer et al., 2015). In the non-hyperaccumulator species *A. thaliana*, significant correlations have been reported between some shoot elements, but not for all (Buescher et al., 2010; Campos et al., 2021; Ghandilyan et al., 2009). This suggests that the element uptake and sequestration, but not the root-to-shoot translocation, relies on common systems for a multitude of elements. The root and shoot concentrations of Ni and Al appeared to be largely uncorrelated to those of any other elements, which could be due to the asymmetric frequency distribution (Figure 1), with a few genotypes with much higher concentrations than the average.

Transgressive segregation was observed in the RIL population, especially for Cd root-to-shoot translocation (TF Cd). 43 RILs showed a higher TF Cd than that of the parent with the highest TF, LE (TF = 2.6). The highest TF Cd found among the RILs was 12.2, which is almost six-fold that of LE. This pattern of transgressive segregation can be observed between parents with small differences in trait values (Rieseberg et al., 1999). The large proportion of segregants with extreme phenotypes is likely to be caused by differences in several QTLs with relatively small effects. In addition, transgressive segregation in both directions indicated that the alleles from one of the parents confer opposite effects to Cd tolerance (in this case, Cd sensitive alleles). Overdominance and epistasis may also contribute to this transgressive segregation (Rieseberg et al., 1999).

Cd tolerance and response of the RIL population

The root-growth based analysis, resulting in the EC100 value as a measure for Cd tolerance, revealed extensive variation in the RIL population. Previously, a shoot-chlorosis based measure showed that the difference in Cd tolerance between GA and LE was rather narrow (50 µM Cd in GA and 5 µM Cd in LE) (Assunção et al., 2003b). Although the Cd concentrations differ for roots, the

differences we observed based on the EC100 values are also close for both accessions (220 µM Cd in GA and 185 µM Cd in LE). Despite these small differences between both parents, the RILs exhibited a wide range of EC100 values. Evidently, this resulted in ample transgressive segregation in both directions (Figure 4A). Most RILs (~ 60%) outperformed GA, the most Cd tolerant parent.

Nevertheless, this wide variation did not result in a high heritability, with a H^2 of only about 20%. Main reason for this is probably that only two replicates could be grown for each RIL, and that the EC100 is sensitive to developmental differences affecting root architecture and interpretation issues, making the method laborious, less suited to be performed by more than one observer, and thus requiring more replicates to get reliable values reflecting the genetic component of the phenotypic variance.

QTL detection in the RIL population

QTL detection was performed using two approaches, the interval mapping and the MQM mapping approach. The results showed that MQM mapping approach can refine QTL detection identified in the interval mapping in this study, by reducing the QTL intervals for most traits (Table 2). In addition, this

method reduced the LOD score of the weaker QTLs, that are less likely to be associated with the traits of interest, subsequently leading to the removal of less relevant QTLs, such as qRT-Cu-2 and qRT-Co.

Our results identified three colocalized QTLs for PLA in response to Cd during to DAS, differing in the timing of expression (Figure 6). Such dynamic, temporal expression of QTLs has been observed more often in high-throughput phenotyping experiments, e.g. of *A. thaliana* (Meyer et al., 2021) and *Brassica napus* (Knoch et al., 2020). It shows that time-series experiments are important to dissect the genetics of developmental traits.

The strong correlations between different element concentrations were shown in Figure 2. In general, highly correlated traits are likely to be influenced by shared genetic factors (Lynch and Walsh, 1998). The QTLs of correlated element concentrations may colocalize in same genomics regions, as observed for some elements in this study (Figure 7). This QTL colocalizations could indicate the presence of a metal transporter influencing accumulation of the same element in different tissues. The root Mn concentration correlated well with shoot Mn concentration and their QTLs co-located on LG3. Likewise, metal transporters could be involved in shoot and root accumulation of Mo, Cd and Fe, as evidenced by their element-specific QTL colocalization (Figure 7).

Candidate genes for Cd accumulation and tolerance in *N. caerulescens*

DNA sequence homology based on synteny between *A. thaliana* and *N. caerulescens* (Figure 8) was used to identify candidate *N. caerulescens* genes underlying QTLs identified in the RIL population analysis. In total 9504 homologues of *A. thaliana* genes were localized in the significance intervals of 27 QTLs. According to GO terms and Araport11 information, 1195 homologues were selected as candidate genes for Cd accumulation and tolerance in *N. caerulescens*, including many well-known genes for metal transporters such as *NRAMP4*, *IREG3*, *ZIP3*, *HMA2* and *YSL3*. Fine-mapping is required to narrow these QTLs down to smaller regions. This however is a very time-consuming effort in *N. caerulescens*. Since most genotypes require stratification and vernalization to germinate and flower, it is often not possible to get more than two generations a year. Fine-mapping was therefore beyond the scope of a PhD project and could not be performed. Introgression of the early flowering *flc* mutation (Wang et al., 2020) into the LE and GA background could facilitate this better in the future, but also this is long term effort. Given these restrictions, we could only identify promising candidate genes for those colocalized QTLs with a high LOD score and a small significance interval (Table 2, Figure 7).

The most prominent QTL we identified was qRT-Cd-1, for the root Cd concentration, which showed the highest LOD score. This QTL colocalized with qTF-Cd, spanning only a small region on LG2 (Figure 9). As discussed above, common metal transporter(s) responsible for both traits are expected. In this QTL colocalization significance interval, syntenic orthologues for six A. *thaliana* genes encoding proteins that are associated with metal homeostasis were found: *HMA3*, *ABCE3*, *HA2*, *LAP2*, *CYP71B2*, *RSZ22* and *SARK*. The most promising candidate gene was *HMA3*. *HMA3* encodes a well-documented Cd transporter involved in Cd sequestration and extreme Cd tolerance, that has been investigated extensively in many species, including *A. thaliana*, *N. caerulescens* and *O. sativa* (Chao et al., 2012; Gravot et al., 2004; Halimaa et al., 2014; Liu et al., 2017; Morel et al., 2009; Ueno et al., 2011). *AtHMA3* has been reported as the sole Cd transporter, variation of which can increase Cd accumulation in shoot in *A. thaliana* (Chao et al., 2012). Similarly, variation of *OsHMA3* in rice, is attributed to reduce Cd accumulation in rice shoots (and grain) by enhancing Cd sequestration into root cell vacuoles (Ueno et al., 2010). In *N. caerulescens*, *NcHMA3* was higher expressed in GA root than in shoot, when plants were exposed to 1 µM

Cd (Ueno et al., 2011). *NcHMA3* expression was higher in GA than in Prayon (PR), a low-Cd-accumulating *N. caerulescens*, under 1 and 10 µM Cd (Ueno et al., 2011). In the absence of Cd-exposure, GA exhibited tenfold higher expression of *HMA3* than LC and Monte Prinzera (MP), a Ni accumulating accession from ultramafic soil (Halimaa et al., 2014). GA was found to contain more copies of *HMA3* than PR, as well as LC and MP, which is likely to explain part of the higher expression of *HMA3* in GA. Immunostaining has revealed that NcHMA3 is mainly localized to vacuoles of epidermal and mesophyll cells in leaves, and endodermis and pericycle cells in roots of *N. caerulescens* (Ueno et al., 2011). Ueno et al. (2011) reported that NcHMA3 also contributes to Cd sequestration in leaf cell vacuoles, and thus contributes to enhanced Cd accumulation in *N. caerulescens* shoots. In the root, *NcHMA3* is suggested to confer Cd tolerance to plants, which facilitates transport of Cd to the shoot of *N. caerulescens*. In our study, HMA3 appears to play a role in the accumulation of Cd in root cells through vacuolar sequestration. This leads to a higher concentration of Cd in the root, subsequently reducing the translocation factor (TF) for Cd. This suggests the participation of HMA3 in regulating TF Cd. HMA3 is indeed involved in shoot Cd accumulation, even though no QTLs were detected for shoot Cd accumulation in this study. This indicates that shoot Cd accumulation is controlled by many individual QTLs, each with small effects, under low Cd exposure. Moreover, genetic interactions among genes may play a role in Cd accumulation in the shoot, which may not be discerned through the one-dimensional approach employed in this study (Broman, 2010). While this one-dimensional approach is effective for detecting QTL(s) with large effects, exploring small-effect QTLs and QTL interactions could be achieved through a two-dimensional approach using a two-QTL model in R/qtl. In *A. thaliana*, two-dimensional approach has revealed multiple epistatic interactions for four element concentrations under low Fe condition (Buescher et al., 2010). Adopting a two-dimensional approach may yield in discovery of small effect QTLs and gene interaction controlling shoot Cd accumulation in *N. caerulescens*.

A second, minor, QTL for root Cd concentration, qRT-Cd-2, was located to LG1. This region contains a homologue of *HIPP20*, which was GO characterised as involved with Cd detoxification. In *A. thaliana*, *HIPP20* is strongly expressed in vascular tissue of shoot and root (Tehseen et al., 2010). Moreover, the *hipp20/21/22* triple mutant was less tolerant to Cd and accumulated less Cd than wild type. *HIPP20* appears to be involved in sensing Cd status, and may thus contribute to Cd accumulation in *N. caerulescens*. Further analysis of the allelic variation for both these genes will be needed to further substantiate their roles in the molecular mechanism of Cd accumulation in *N. caerulescens* roots. Regarding Cd tolerance, 17 candidate genes were localized in the qEC100 QTL, including very promising candidates such as *NRAMP5* and *CATION EXCHANGERS 3* (*CAX3*). In rice, the *NRAMP5* orthologue has been reported to be involved in Cd accumulation and uptake (Sasaki et al., 2012). A recent study reported that the functional knockout of *OsNRAMP5* can alleviate Cd toxicity by reducing Cd uptake and enhancing Cd translocation to shoots in rice (Tang et al., 2022). In the case of *CAX3*, the overexpression of this gene was found to increase Cd tolerance in *A. thaliana* by decreasing the Cd-induced production of reactive oxygen species (ROS) (Modareszadeh et al., 2021). If such is also the case in *N. caerulescens* will need to be investigated.

Conclusion

In summary, a total of 27 QTLs were detected in this study for Cd accumulation, Cd tolerance, and growth response to Cd, revealing intricate genetic associations and potentially a suite of candidate genes influencing these traits. Future work can adopt functional analysis approaches such as comparative gene

expression analysis and floral dipping transformation (**Chapter 5**) to verify the roles of potential candidates in Cd accumulation and tolerance in *N. caerulescens*.

Supplementary information

Table S1. Ionomics profile of RIL parental lines Ganges (GA) and Lellingen (LE) growing on 1 µM $CdSO_4$. Both GA and LE are grown alongside with the RIL population in the Cd accumulation test. Eight plants are used for each parent.

Element	Average shoot concentration (mg kg^{-1})		Average root concentration (mg kg^{-1})		Average TF	
	GA	LE	GA	LE	GA	LE
B	55.03	61.10	82.23	65.36	0.87	1.00
Na	23.03	13.07	340.19	337.67	0.14	0.05
Mg	2062.25	1710.18	1650.70	1470.93	1.30	1.18.67
Al	1.46	1.38	18.58	114.05	0.13	0.13
P	3234.93	2451.39	8603.89	5181.71	0.40	0.49
S	13017.80	11681.20	33515.65	20153.33	0.58	0.60
K	38867.51	30759.54	33752.51	20431.24	1.21	1.60
Ca	5845.32	7959.32	3463.93	3036.33	2.02	2.89
Fe	38.84	32.45	450.54	290.36	0.09	0.11
Mn	63.98	63.18	287.36	93.94	0.34	1.76
Co	0.00	0.00	0.01	0.01	0.23	0.24
Ni	0.11	0.25	0.47	0.45	0.26	0.59
Cu	1.43	1.39	14.01	13.55	0.11	0.10
Zn	131.02	216.58	81.58	67.25	2.63	4.26
As	0.00	0.00	0.01	0.01	0.35	0.15
Rb	0.37	0.31	0.36	0.26	1.07	1.19
Sr	0.77	1.08	2.91	2.88	0.35	0.49
Mo	8.64	13.46	58.16	55.23	0.16	0.24
Cd	14.36	17.63	53.85	7.15	0.31	2.59

Acknowledgement

The authors appreciate Tom P.J.M. Theeuwen, René Boesten, and Roel F.H.M. van Bezouw (Wageningen University) for recommendations on statistical and QTL analysis; and the Royal Thai government (Thailand) for financial support for this research.

CHAPTER

Genetic variation in natural populations of *Noccaea caerulescens* in Europe

Jitpanu Yamjabok, Laurens van Oostroom, Joost van den Heuvel, Henk Schat, Mark G.M. Aarts

Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

Financial source:
This research is financially supported by Royal Thai government (Thailand).

Abstract

Noccaea caerulescens is a cadmium (Cd)/lead (Pb)/nickel (Ni)/zinc (Zn) hyperaccumulating *Brassicaceae* species, that is widely investigated in order to understand local adaptation to metal exposure, expressed as metal accumulation and tolerance. Previous studies have investigated the population structure of *N. caerulescens* local populations. However, these studies are limited by either a small number of local populations from larger areas of Europe or by several local populations from a limited area. To overcome these issues, a diversity panel of 109 *N. caerulescens* accessions collected from much of its geographical area was developed and investigated. The genetic variation among the diversity panel was assessed by whole genome sequencing, which generated 733,912 filtered SNPs, based on which the population structure of *N. caerulescens* was determined. A total of six genetically distinct subgroups were identified which are similar to subgroups described in a previous study. Furthermore, a new refugial population from the Iberian mountains and the Apennines was identified. The population structure suggested that the accessions from British and Scandinavia are likely to have originated from Western Europe. We hypothesize that human activities were involved in the spread of *N. caerulescens* to these regions. The diversity panel was also used for genome-wide association analyses (GWAS), which was partly successful in identification of quantitative trait loci (QTLs) for metal accumulation and flowering time. The strong population structure masked much of the variation though. Further optimization or expansion of the panel will be required to enhance the success of GWAS for detection of QTLs.

Introduction

Noccaea caerulescens is a plant species that tolerates and accumulates a diversity of metals, such as nickel (Ni), zinc (Zn), cadmium (Cd) and lead (Pb) in its leaves (Assunção et al., 2003a, 2003b, 2008; Dinh et al., 2018; Escarré et al., 2013; Kozhevnikova et al., 2020; Milner and Kochian, 2008; Mohtadi et al., 2012; Sterckeman et al., 2017). This enables the species to grow on metal-enriched soils like calamine soil (CAL) with elevated concentrations of Zn and/or Cd and Pb; and ultramafic soil (ULT) with elevated concentrations of Ni and Mg. It also occurs widely on non-meticulous soils (NM) though. Genetic variation exists among accessions from different soil types in the concentrations of metals they can accumulate and tolerate. This heritable intraspecific variation for metal hyperaccumulation and tolerance of *N. caerulescens* represents an example of local adaption in plants (Westerband et al., 2021). Gonneau et al. (2014) suggested that metal-adapted populations appear to be recruited locally, for NM populations, which seems to the case for Pontaut and Vall de Varados populations as well (Wang, 2016). Unravelling the genetic basis of population differences in *N. caerulescens* not only enriches our understanding of plant evolutionary dynamics but also has practical implications for phytoremediation, phytomining, and biofortification of crops with essential nutrients.

The relationship between genotypes and phenotypes among populations can be the result of local adaptation, but also non-adaptive changes such as mutation and drift. To understand whether local adaptation has led to genetic differences in *N. caerulescens*, knowledge of the evolutionary history and population structure of this species is key. The population history and structure have been previously examined in three different sets (CAL, ULT, NM) of accessions (Besnard, 2009; Gonneau et al., 2017; M. Koch et al., 1998). *N. caerulescens* can be found throughout Europe. Populations from the British Isles are proposed to likely be endemic to Britain as old relict flora rather than originating from Central Europe such as Germany (Koch et al., 1998). Scandinavian populations are suggested to be recently introduced by human activity (Koch et al., 1998; and references therein). Local populations from France (the Massif Central and the Pyrenees), Belgium-Luxembourg (Ardennes) and Switzerland (Jura and the Alps) have

been clustered into three distinct genetic subunits (SU) (SU1, SU2 and SU3) with one admixture group (AZ) (Gonneau et al., 2017). This illustrates the high degree of genetic diversity on a relatively small geographic scale for *N. caerulescens*, although this may be quite common for wild Brassicaceae species (Baumgarten et al., 2023; M. Koch and Bernhardt, 2004; Parisod and Besnard, 2007; Stein et al., 2017; Sunar et al., 2016). Approximate Bayesian computation analysis has revealed that SU1 accessions, from southeast of the Massif Central, diverged first from SU2/SU3 at 31.8 thousand years Before Present (kyrBP), then the separation between SU2 and SU3 is relatively recent at 11.7 kyrBP. Still, the estimated time of the divergence among groups is clearly before establishment of human mining activities. This indicates that metal accumulation and tolerance traits are developed prior to the local adaption to specific metalliferous habitats (Gonneau et al., 2017). What remains unknown is to what extent specific genetic changes have resulted in the difference between genotypes and phenotypes between the *N. caerulescens* populations in Europe.

Here we present a study in which we performed whole genome sequencing on a new panel of 109 *N. caerulescens* diverse accessions from larger geographical areas of Europe, compared to previous studies. This data allow us to identify patterns of population structure among the accessions. In addition, we performed phenotyping, allowing for the first genome wide association study for *N. caerulescens* between genotype and ecologically relevant phenotypes. The traits of interest were flowering time and Zn and Ni accumulation under Zn/Ni treatment. This study deepens our fundamental understanding of local adaptation and tests whether the current panel of *N. caerulescens* enables the discovery of new candidate genes for hyperaccumulation of metals, which has practical implications for phytoremediation and phytomining.

Methodology

Plant material and population development

A set of 109 *Noccaea caerulescens* natural accessions, referred to as the "Diversity panel", was collected from across Europe (Table S1) or kindly donated by Alan Baker, Catherine Sirguey, Dario Galanti, Enrica Roccotiello, Guillaume Besnard, Guillaume Echevarria, Nausicaa Noret, Ohana Barriuta, Radim J. Vašut, Sylvian Merlot, Terezie Mandáková, Thibault Sterckerman, Sophie Laurent, Fangjie Zhao. The accessions of this panel were grouped according to geographic locality, including the Alps (9), Apennines (2), Ardennes (9), Bohemian Massif (1), British Isles (5), Carpathians (9), Finland (1), Iberian mountains (7), Jura (17), Dutch (3), Massif Central (25), Pyrenees (5), Sweden (2), and Vosges (14). The accessions of this panel were collected from different soil types, which categorizes them in three different ecotypes: Calamine (CAL) accessions (35 in total) originating from soils enriched with zinc (Zn), cadmium (Cd), and lead (Pb), Ultramafic (ULT) accessions (11) from soils enriched with nickel (Ni), and Non-metallicolous (NM) accessions (63) from regular soils without metal enrichment. To enhance the genetic uniformity and phenotypic consistency within each accession, the diversity panel was propagated by self-pollination for at least three generations during September 2007 to June 2019 in the Unifarm greenhouses at the campus of Wageningen University, The Netherlands.

Single Nucleotide polymorphism (SNP) genotyping and filtering

DNA was extracted from frozen young inflorescences using solid phase reversible immobilization (SPRI) beads as described in **Chapter 2**. For DNA library preparation, DNA was diluted to 0.2 - 0.25 ng/uL and tagmented using Illumina Nextera Kit (Illumina, USA). Libraries were amplified by Illumina TruSeq primer and Robust 2G enzyme (Kapa biosystems, South Africa). The 300-500-bp libraries were selected for paired-

end sequencing to an average depth of 30x whole-genome coverage at Novogene Europe (Cambridge, United Kingdom). The raw reads were trimmed for adapter sequences and filtered for low-quality bases and reads using Trimmomatic v. 0.32 (Bolger et al., 2014). The filtered reads were aligned to the so far unpublished *N. caerulescens* 'Ganges' reference genome assembly (Severing and Aarts, unpublished data) with Burrows-Wheeler Aligner (BWA) (Li et al., 2009) using default parameters. Variant calling was conducted using Freebayes (Garrison and Marth, 2012) with custom parameters (-use-best-n-alleles 4 -min-alternate-count 2 -limit-coverage 100). The variant output was read and filtered using vcfR package (Knaus and Grünwald, 2017) and a custom R script (R Core Team, 2008) with these parameters: variants should be SNPs; maximum allele frequency (non-reference sequencing depth / total sequencing depth (DP)) difference between any accessions should exceed 0.8; variants should be biallelic; no missing genotypes were allowed; DP should be higher than 4 for all accessions; maximum DP should be less than three times of mean average DP for each accession; and minor allele frequency (MAF) over all accessions should be larger than 0.05.

Population structure analysis

The genetic distances of all combinations between two accessions were calculated as absolute allele frequency differences (AFD) as described in Berner (2019). Based on AFD, hierarchical cluster analysis was conducted to divide this diversity into different genetic units called subgroups (SG). To determine the optimal number of subgroups, Gap statistic was conducted for K from 1 to 10, with ten replicates for each K using FactomineR version 2.9 (Lê et al., 2008). A dendrogram of this panel was illustrated using ggtree version 3.1.0.0 (Yu et al., 2017).

Quantification for flowering time and Zn/Ni accumulation

The diversity panel was phenotyped for flowering time as described in **Chapter 2**. For, A subset of 90 accessions of the diversity panel was assessed for Zn/Ni accumulation, as described by van der Zee et al. (2021).

Genome-wide association study trial

Genome-wide association (GWA) between SNP genotypes and traits of interest including flowering time and Zn/Ni accumulation was analysed using a mixed linear model in GEMMA version 0.98.5 (Zhou and Stephens, 2012). The pairwise AFD matrix was included in the model to correct for population structure. The thresholds for this analysis were set to -log *P* of > 7.17 at significance level of 5%, based on Bonferroni correction. Manhattan plots were drawn using a custom R script and ggplot2 package version 3.4.4 (Villanueva and Chen, 2019). Percentage variance explained (PVE) was calculated per single SNP for each trait of interest in GEMMA. To simulate a level chromosome map of *N. caerulescens*, each contig was anchored and oriented into seven pseudomolecules based on the genetic map of the Ganges x Lellingen recombinant inbred lines population (**Chapter 2**).

To define regions of interest, squared correlation coefficients (r^2) were calculated between significant SNPs and other SNPs located on the same contig using Plink version (Purcell et al., 2007,)--allow-extra-chr --allow-no-sex --ld-window 300 --ld-window-kb 100 --ld-window-r2 0 --maf 0.05 --r2 --silent). To identify candidate genes, the regions of interest were investigated for orthologues of *Arabidopsis thaliana* as described in **Chapter 2**.

Statistical analyses

Statistical analysis was performed using R (R Core Team, 2008). All mineral concentrations were log-transformed before analysis. Principle component analysis (PCA) was conducted using FactoMineR version 2.9. The board-sense heritability was calculated as a ratio between genetic variance to the phenotypic variance using linear mix model, with genotype as a random effect. Pearson's correlations were determined for Zn and Ni concentration at shoot and root. Topological maps were drawn using ggmap (Kahle and Wickham, 2013).

Result

SNP dataset of Diversity panel

A total of 3,839,001,870 150-bp pair-end reads were obtained from 109 accessions of the *N. caerulescens* diversity panel with an average of 35,220,201 reads per accessions resulting in an average coverage of 21.76 fold. After trimming and quality filtering, a total of 3,702,829,873 pair-end reads were retained. A total of 18,320,212 raw variants were obtained from the SNP calling procedure using Freebayes. Setting the minimal DP above 5 was too strict, resulting in far too low number of filtered SNPs. The minimal DP was set to be above 4 and a total of 733,912 filtered SNPs was obtained after variant-filtering procedures (Figure S1).

Population clusters of *Noccaea caerulescens*

To explore the population structure of *Noccaea caerulescens*, pairwise genetic distances were estimated from AFD between all 109 accessions. The AFD between accessions ranged from 0.0089 to 0.3936 with an average of 0.2559. Gap analysis identified six subgroups, as the best model for hierarchical clustering of the 109 accessions. The accessions were grouped as SG1 (37), SG2 (13), SG3 (15), SG4 (17), SG5 (16), and SG6 (11) (Figure 1 and 2, Table S1). CAL and NM accessions were assigned to all subgroups, while the 11 ULT accessions were assigned to SG1 (3), SG2 (1) and SG3 (7) (Figures 1 and 2). The average geographical distance between accessions was 675 km, while the maximum distance was 3222 km, between Kuopio (KUO) from Finland and Cira (CIR) from Spain (Figure 3). Principle component analysis (PCA) was conducted on the AFD values. The results showed that the first four principal components (PCs) explained over 80% of the AFD variance. PC1 distinguished SG5 from the other subgroups, PC3 separated SG6 from the others as well, while PC2 and PC4 together divided the diversity panel into six subgroups with few SG1 members locating in SG4 space (Figure 3).

The different subgroups of *N. caerulescens accessions* resulting from the clustering analysis were associated to particular localities. The accessions from Apennines, Ardennes, British, Dutch, Iberian mountains and Pyrenees were each uniquely assigned within a single subgroup, while accessions from the Alps, Carpathians, Jura, Massif Central, Sweden and the Vosges were allotted to separated subgroups. A small number of accessions, from the Carpathians (1) and Vosges (2), were found to be an outgroup and showed membership to different subgroups (Figure 1 and S3).

SG1 was the largest subgroup consisting of accessions of all three ecotypes (21 NM, 13 CAL, and 3 ULT), originated from Ardennes (9), British (5), Dutch (3), Jura (5), Vosges (12), Alps (1), and Carpathian (1). The Carpathian NM accession "Moravsko-slezské Beskydy Mountains" (MSB) located apart from the rest of the SG1 members (Figure S3A), yet MSB was closely related to two CAL accessions from two distant locations: Plombieres (PB, Belgium), and Priddy Mineries (PDY, UK) (Figure 2). SG2 included accessions from the Pyrenees region (5), the Alps (2), Massif Central (5), and Sweden (1), which includes one of a

ULT accessions, Puy de Wolf (PDW). Interestingly, two members of SG2 from the Alps (Vallauria mine (VLM) and Casterino (CAS)) were collected from the Western Alps lying between Massif Central and Apennines, separating SG3 and SG5 (Figure 1). Members from SG3 originated from the Alps (2), Apennines (2), Iberian mountains (7) and Massif Central (4), which roughly split into two groups a northeast group (Alps, Apennines and Massif Central) and southwest group (Iberian mountains). Seven SG3 members were ULT including Pontboset (POB), Monte Prinzera (MP), Valle Gargassino (VGG), Bandeira (BAN), Puente Basadre (BAS), Cira (CIR) and Eidian (EID), considering a majority of ULT of this panel. The fourth subgroup (SG4) consisted of 16 NM accessions from the Alps (4), Jura (11) and Vosges (1), and one CAL accession "Plancher-BAS" (PLA), also from the Vosges. SG5 was the most genetically distant from the other subgroups (Figure 2 and S2). SG5 included mainly accessions from the southern Massif central (15) and one from further west in France (Viviez (VIV)), which were 10 CAL accessions and 6 NM accessions. This subgroup showed the smallest geographical area, surrounded by members of SG2 and SG3 (Figure 1 and S3). The CAL accession VIV, located from a Zn smelter very close (7.24 km) to the ultramafic mountain hosting PDW, an ULT member of SG2, but they are clearly not genetically close, with a genetic distance $\Delta G = 0.295$. The last subgroup (SG6) was the smallest, but it occupied the largest geographic area. The majority of SG6 originated from the Carpathians (8) in Slovakia, while the rest were from the Bohemian Massif (1), Finland (1) and Sweden (1). The two accessions originating from Sweden, Höör (HOO) and Uppsala-4 (UPP4), were genetically distant from each other, with HOO to be closely related to the Finnish accession Kuopio (KUO) (Figure 2).

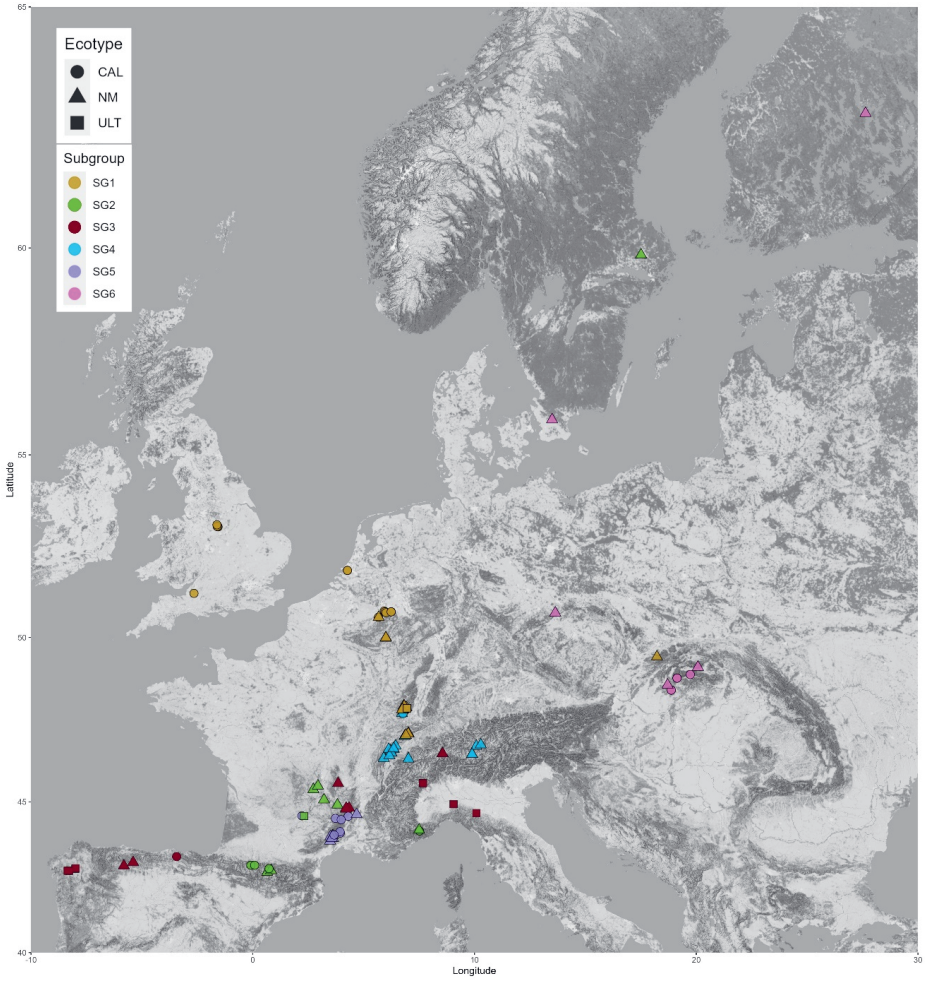


Figure 1. *A geographical map indicating the locations across Europe of the 109 N. caerulescens accessions of the diversity panel. Six subgroups are distinguished following hierarchical clustering (see Figure 2). These are highlighted with different colours (gold = SG1, green = SG2, bordeaux = SG3, blue = SG4, violet = SG5, pink = SG6). Ecotypes are labelled with different shapes, Calamine - circle, Non-metallicolous - triangle, Ultramafic - square.*

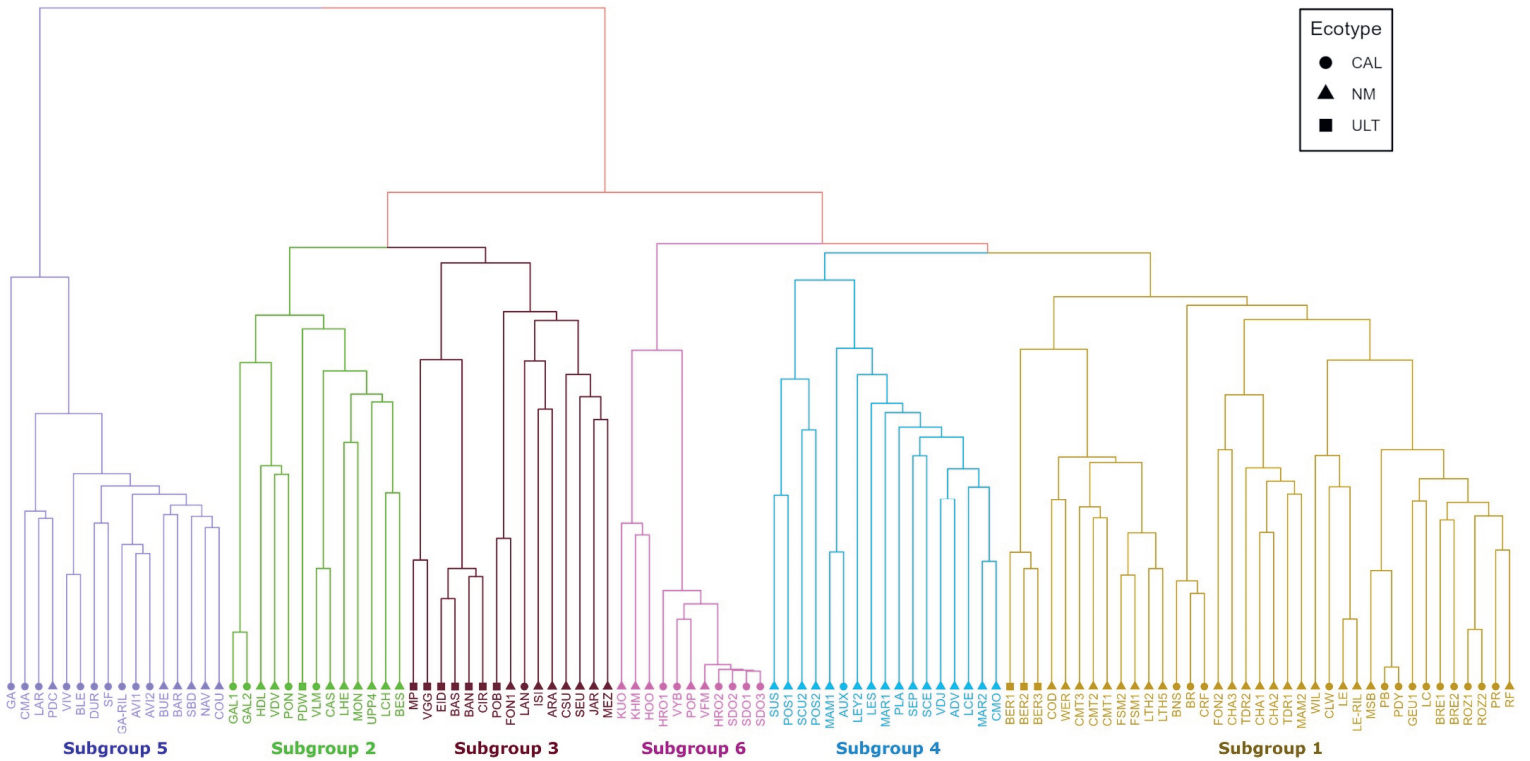


Figure 2. *Dendrogram showing hierarchical clustering of the 109 Noccaea caerulescens accessions in the diversity panel. The clusters (referred to as subgroups) are determined with Gap Statistic and labelled with different colours. The ecotype of each accession is indicated with a different shape, Calamine (CAL) - circle, Non-metallicolous (NM) - triangle, Ultramafic (ULT) – square.*

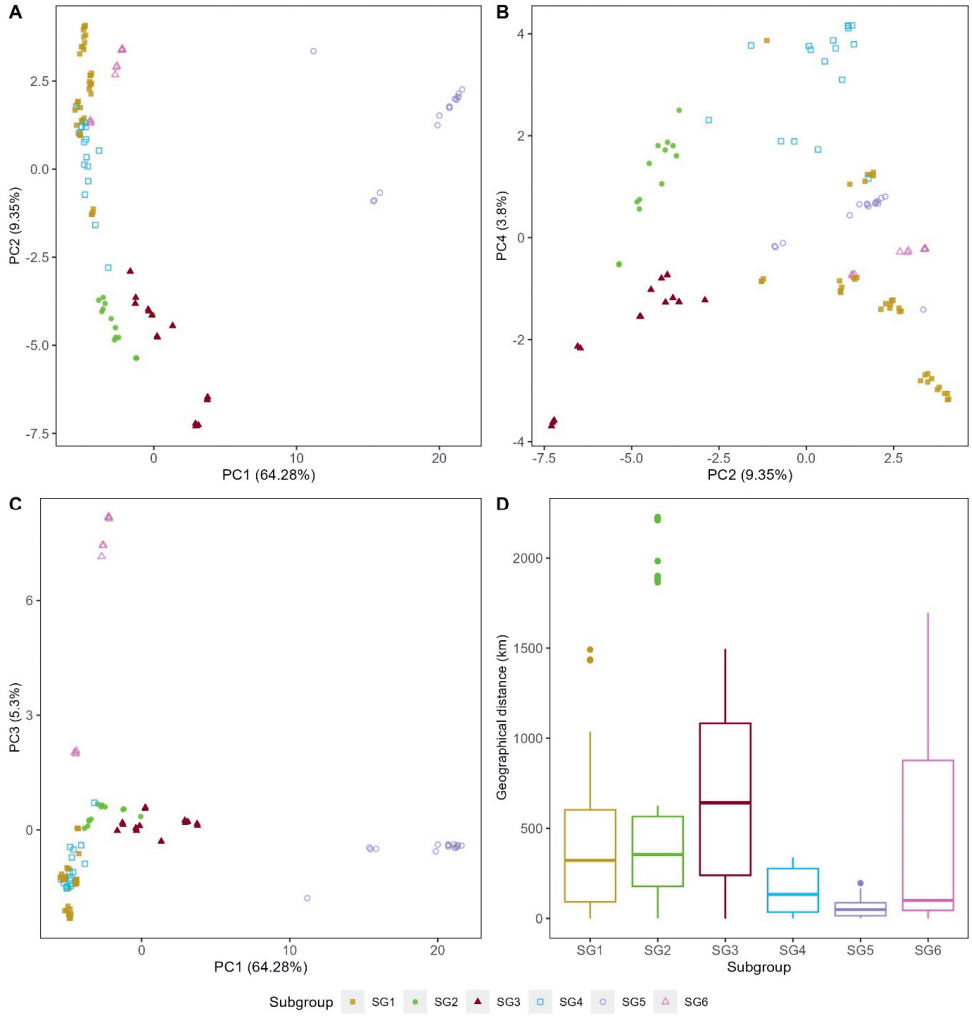


Figure 3. *Clustering analysis of Noccaea caerulescens diversity panel using genetic distances ΔG. Principal component analysis (PCA) showing genetic variation in the Noccaea caerulescens accessions with: (A) PC1 and PC2, (B) PC2 and PC4, and (C) PC1 and PC3, the colour and shape was defined by subgroup (gold = SG1, green = SG2, bordeaux = SG3, blue = SG4, violet = SG5, pink = SG6). (D) boxplot shows the differences of pairwise geographical distance between accessions within subgroups. The pairwise geographical distances were calculated based on great circle distance (Haversine) between GPS coordinates using R package "Geosphere".*

Phenotypic variation

To phenotype flowering time (FT), plants from the diversity panel were grown in a frost-free greenhouse over winter. FT was scored as the time until the date when the first anther released pollen. FT varied across the diversity panel of 109 accessions from 166 days after sowing (DAS) to 204 DAS, with an average of 170.0 DAS (Figure 4A). The distribution of FT was normal (test-stat = Shapiro-Wilk's method, $p < 0.05$) with three accessions, Black Rocks (BR), Bandeira (BAN), and Cromford (CRF), showing extreme late flowering at 204 DAS. The broad-sense heritability (H^2) for FT was 0.79, indicating that a large proportion of the variance among plants is explained by genetic differences. Significant differences in FT were found between the CAL and NM ecotypes (test-stat = ANOVA, $p < 0.001$) (Figure 4C). In comparison, among subgroups, SG2 exhibited the most delayed FT with an average of 179 ± 9.1 DAS (Figure 4B). Although the FT of SG2 differed significantly only from SG1, SG2 showed distinctly higher FT than the other subgroups of which average FT ranged from 167 to 172 DAS. Moreover, the result indicated that population structure (subgroup) associated with FT significantly at $p = 0.0005$, based on ANOVA model using subgroup as fixed variable.

To explore metal accumulation within the diversity panel, plants were grown hydroponically under a combined elevated Zn/Ni treatment. The shoot and root concentrations of Zn and Ni were determined using ICP-AES. Zn and Ni concentrations showed similar degrees of variation between accessions (coefficients of variation (COV) varying between 0.32 and 0.61). There were small differences between the average shoot and root concentrations of Ni and Zn (Table 1). Of the four accessions with elevated shoot Ni concentration (above 2000 µg/mg), one was an ULT accession: Puy de Wolf (PDW); and three were NM accessions: Le Thillot-5 (LTH), St.Cergeus (SCE), and Vallée de Joux (VDJ) (Figure S4). High shoot Zn concentrations (above 10,000 µg/mg) were observed in 14 accessions, including VDJ, Leysin-2 (LEY2), Le Suchet (LES), Le Thillot-5 (LTH5), Tête de Ran (TRD), Tête de Ran-2 (TRD2), Au Dent de Vaulion (ADV), Vyšná Boca (VYB), Col du Mollendruz (CMO), Lellingen parent RIL (LE-RILPT), St.Cergeus (SCE), Bergenbach-1 (BER1), PDW and Popradské pleso (POP), of which the majority were also belonging to the NM ecotype. A high root Zn concentration (above 10,000 µg/mg) or root Ni concentration (above 2000 µg/mg) were found in seven (3 ULT, 2 NM and 2 CAL) and two accessions (1 CAL and 1 NM), respectively.

The metal Translocation Factor (TF) was calculated as the ratio of the shoot metal concentration to the root mineral concentration. Approximately one-third (34) and half (41) of this panel displayed a TF above 1 for Ni and Zn, respectively (Figure 5). A high TF of Ni above 2 was detected in six accessions, especially in the ULT accession PDW (0.51). In contrast, a low TF of Ni (< 0.5) was observed in 14 accessions. Of these, accession Largentiere (LAR) exhibited the lowest TF of Ni (0.05). In case of Zn, 11 accessions showing elevated TF (2) were found among all three ecotypes including 1 ULT, 1 CAL and 9 NM. Among these accessions, LES showed the highest Zn TF (3.24). Low Zn TFs (<0.5) were mostly observed in 15 accessions (2 ULT, 2 NM, 11 CAL), of which two (Galine-1 (GAL1) and LAR) had a Zn TF below 0.1. Zn and Ni accumulation and the TFs upon Zn/Ni treatment were genetically controlled, as we estimated H^2 to be above 0.50. The shoot Zn concentration showed an extremely high H^2 of 0.95. Correlation analysis revealed that the Zn concentration correlated significantly positively with the Ni concentration for both within or between plant parts. Strong correlations were detected for Zn shoot concentration vs Ni shoot concentration ($r = 0.64$), and for Zn root concentration vs Ni root concentration ($r = 0.66$), based on Pearson's correlation analysis.

When comparing ecotypes, CAL displayed significantly lower Zn and Ni shoot and root concentrations than ULT and NM, based on ANOVA at $p < 0.05$ (Figure 4F, 3I, 3L and 3O). In shoot tissues, Zn and Ni

concentrations were not different between ULT and NM, and the same for root Ni concentrations. When categorizing the diversity panel into three groups based on shoot Zn concentration, a high shoot Zn, intermediate shoot Zn and low shoot Zn concentration group was distinguished (Figure 4K). The intermediate shoot Zn group included most of the subgroups (SG1, SG2, SG3 and SG6). SG4 largely contained high shoot Zn accessions, of which the majority were NM, from the Jura and Vosges regions, while SG5 generally exhibited the lowest shoot Zn concentrations. A similar division was observed for shoot Ni concentrations (Figure 4E). With respect to root metal concentrations, SG1 and SG5 exhibited low Zn concentrations, while SG1 showed low Ni concentrations (Figure 4H). For both plant parts, SG4 showed high Zn and Ni concentrations. In general, distinct differences between subgroups were observed with respect to shoot metal concentrations, while for root concentrations there was less difference between subgroups.

Table 1. *Descriptive statistics for five traits determined for the Noccaea caerulescens diversity panel. Minimum (Min), maximum (Max), standard deviation (SD), coefficient of variance (COV), and broad sense heritability (H^2) calculated for diversity panel across biological replication (n = 3). The translocation factor is calculated as the ratio of shoot concentration to root concentration.*

Trait	Average ± SD	Min	Max	COV	H^2
Flowering time (DAS)	170.03 ± 11.14	145.00	204.00	0.07	0.79
Shoot Ni concentration (µg/mg)	1048.51 ± 423.08	106.67	2430.00	0.40	0.83
Root Ni concentration (µg/mg)	1253.14 ± 398.91	485.33	2119.00	0.32	0.82
Ni translocation factor	0.97 ± 0.64	0.05	5.06	1.53	0.86
Shoot Zn concentration (µg/mg)	5872.15 ± 3550.86	63.00	15126.67	0.60	0.95
Root Zn concentration (µg/mg)	5546.27 ± 3388.63	465.00	18973.33	0.61	0.88
Zn translocation factor	1.20 ± 0.70	0.05	3.24	1.7	0.75

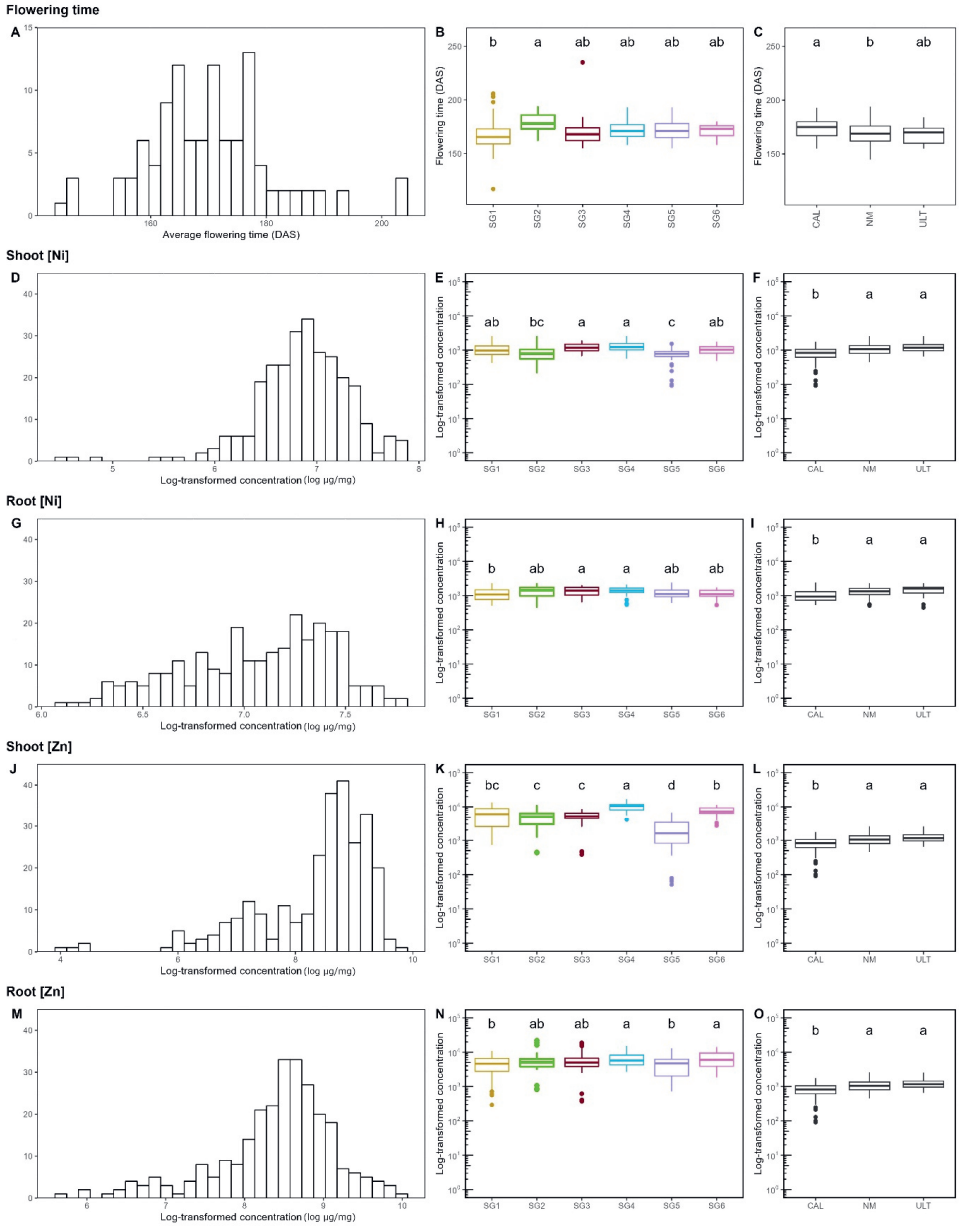


Figure 4. *Phenotypic variation for flowering time, shoot and root Ni concentration ([Ni]), and shoot and root Zn concentration ([Zn]) in the N. caerulescens diversity panel (A, D, G, J, M), of which horizontal bars indicate number of accessions. Histograms for traits are indicated in the middle column (panels B, E, H, K, N). Y axes indicated number of accessions. Each box is coloured according to subgroup (gold = SG1, green = SG2, bordeaux = SG3, blue = SG4, violet = SG5, pink = SG6). The third column box plots (C, F, I, L, O) show the differences between the three ecotypes Calamine (CAL), Non-Metallicolous (NM) and Ultramafic (ULT). Letters represent significant differences based on Tukey's HSD.*

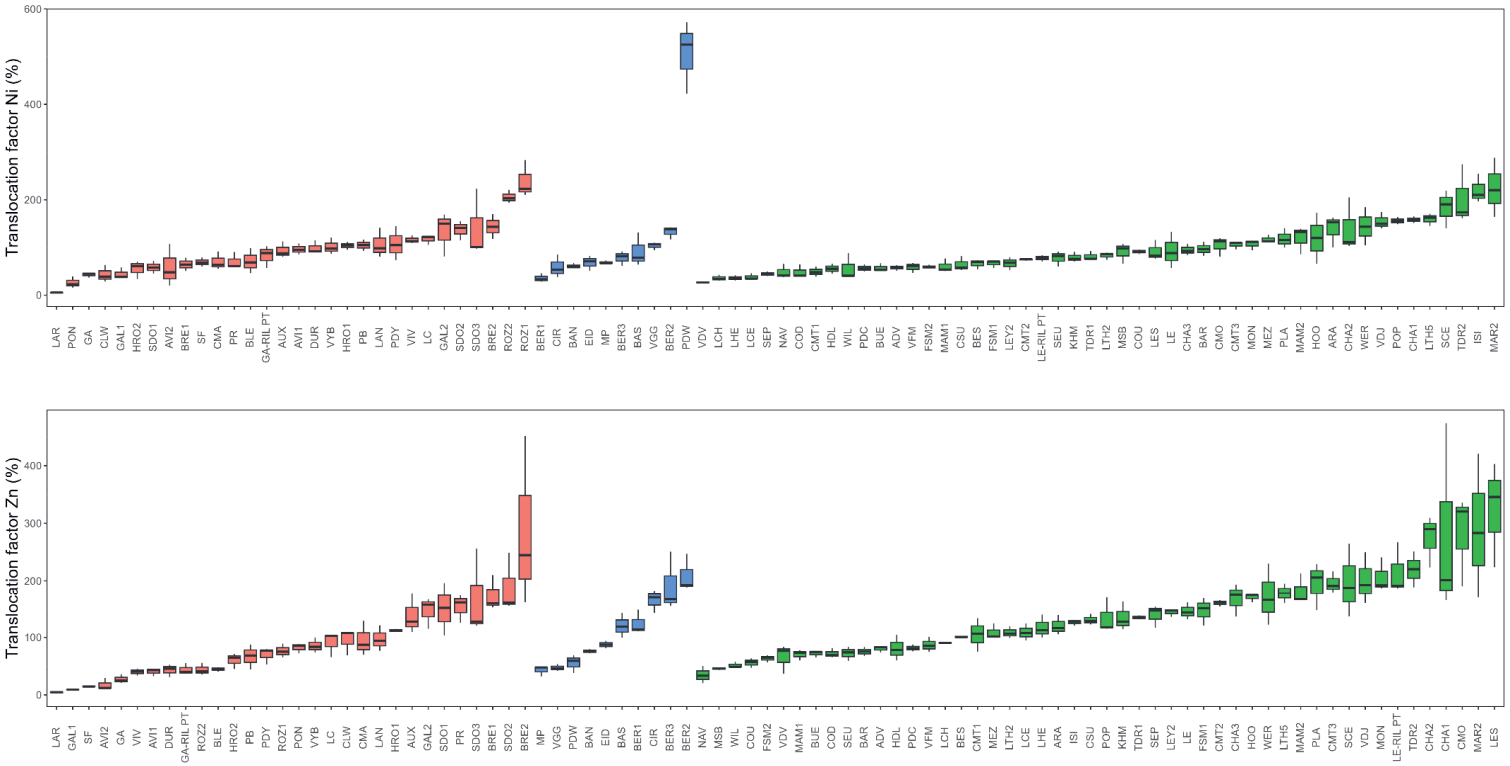


Figure 5. *Translocation factors of Ni and Zn of 90 Noccaea caerulescens accessions upon Zn/Ni treatment. Box plots are drawn per genotype, that are ranked from low to high value for Ni (top) and Zn (bottom). The black lines indicate the average per accession. Each plot is coloured according to ecotypes including Calamine (red), Non-metallicolous (green), and Ultramafic (blue).*

***N. caerulescens* genome wide association studies**

Genome wide association studies (GWAS) were conducted for the determined traits flowering time (FT) and Zn and Ni concentrations in shoot and root. The genetic distance (AFD) was used to correct for population structure. Significant SNPs, exceeding the Bonferroni-corrected -log(p) threshold value, were found for all traits (Figure 6).

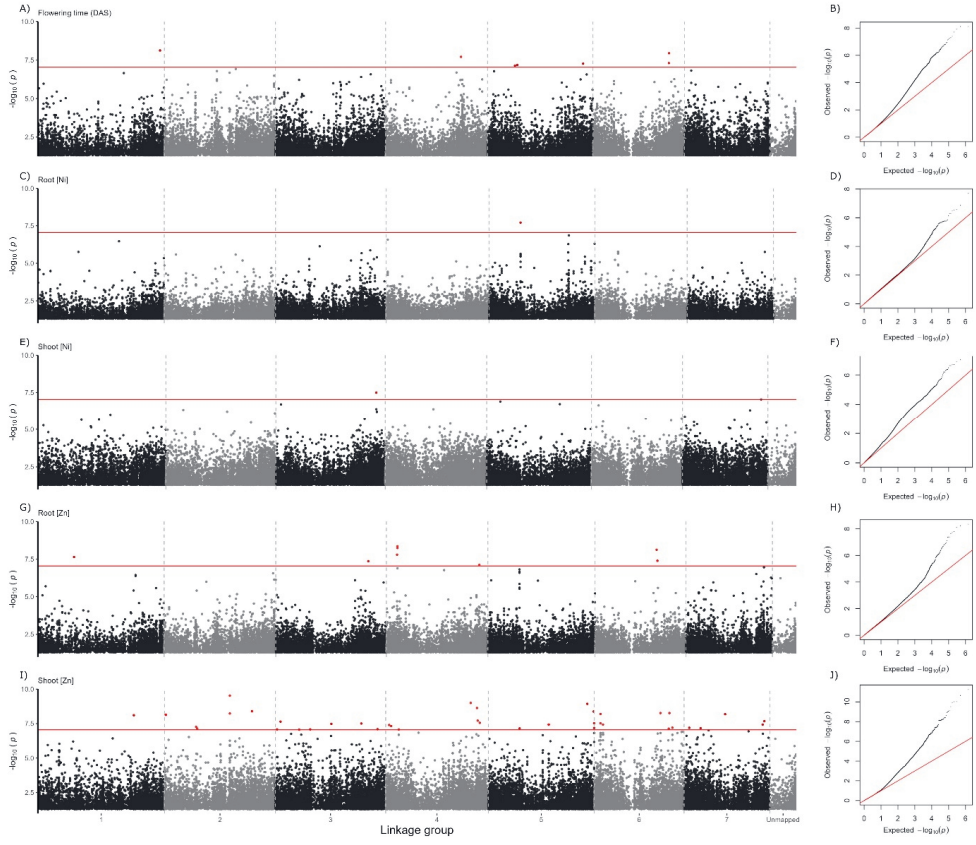


Figure 6. *Genome wide association study of the N. caerulescens diversity panel. On the left, Manhattan plots of five traits are shown: FT (A), root and shoot Ni concentration (C, E), and root and shoot Zn concentration (G, I). The red line indicates Bonferroni-corrected significance thresholds at -log$_{10}$(p) = 7.17. Red dots above the threshold represent significant SNPs. The right panel shows Q-Q plots corresponding to traits on the left (B, D, F, H, I).*

A total of 7, 1, 1, 9 and 52 SNPs significantly associated with FT, root and shoot Ni concentration, and root and shoot Zn concentration, respectively. The significant SNPs associated with FT explained a large proportion of trait variance with the maximum PVE of 15.24%, while the PVE for the SNPs associated with Zn and Ni concentration was quite low (< 1%) (Table S2). No SNPs were detected that significantly associated with multiple traits. We determined the r^2 for all significant SNPs on both sides to identify the SNPs in linkage disequilibrium (LD) with the associated SNP(s), delimiting the region of the genome mostly likely to contain the allelic sequence variation underlying the observed phenotypic variation. The majority

of the significant SNPs had r^2 below 0.8 to surrounding SNPs (Figure 7B and 8B). We proceeded by identifying syntenic blocks in the *A. thaliana* genome, around the significant SNPs to identify candidate genes. Based on whole genome alignment (see **Chapter 3)**, syntenic blocks were identified for 61 significant SNPs. For nine SNPs, no match to *A. thaliana genome* was found. Thes syntenic blocks contained *N. caerulescens* orthologues to 285 *A. thaliana* genes, of which most were not obviously showing a predicted function affecting the associated trait of interest (Table S2).

It is worth noting that two homologues with GO terms related to the phenotypes were observed including AT2G18735 (*FLOWERING-ASSOCIATED INTERGENIC INCRNA* (*FLAIL*)) for FT and AT5G59520 (*ZINC-REGULATED TRANSPORTER /IRON-REGULATED TRANSPORTER-LIKE PROTEIN 2* (*ZIP2*)) for shoot Zn concentration. For FT, the *FLAIL* homologue was located 14kb downstream of SNP Contig16:117310 on LG6. This is an Adenine (A)/Cytosine (C) SNP with a 25 / 84 accession allelic variation over the diversity panel (Figure 7A). The accessions with A allele flowered at 180 DAS on average, which was 12 days later than those with C allele (Figure 7C). For shoot Zn concentration, a significant SNP, Contig320:118716, was found in a syntenic block corresponding to *A. thaliana* chromosome 5 between 23985785 to 24007676 bp. This region contains the *ZIP2* gene (AT2G28670), which belongs to the ZRT/IRT-like protein family of metal transporter genes. The *N. caerulescens* homologue was detected at approximately 8 kb downstream of the associated SNP (Figure 8). This SNP distinguishes a thymine (T) or C allele, in respectively 100 and 9 of the accessions (Figure 8A, B). The accessions with the C allele accumulated a much lower average Zn concentration in shoots (1002 µg/mg) compared to the ones with the T allele (6421 µg/mg) (Figure 8C).

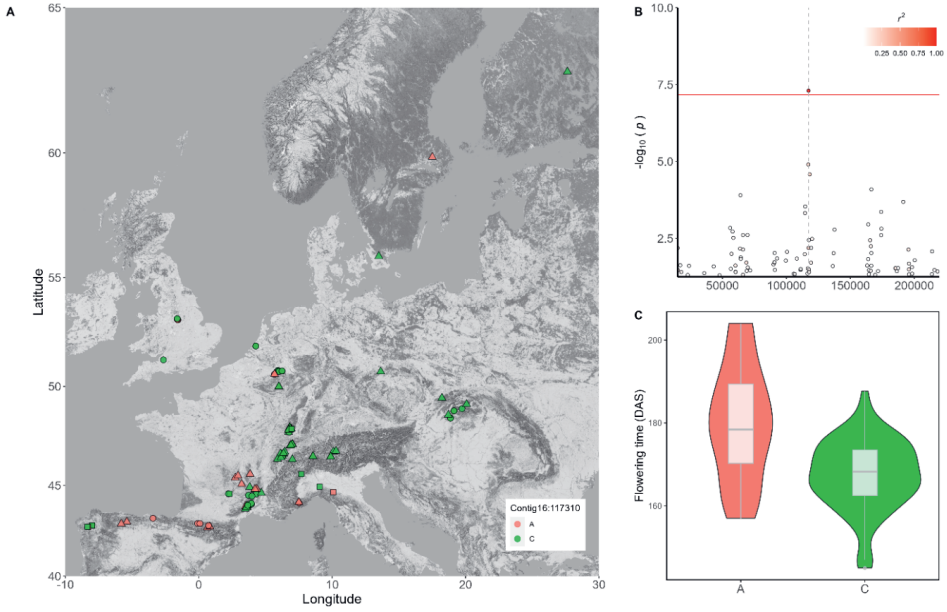


Figure 7. *The distribution and significance of allelic variation at SNP Contig16:117310 associated with flowering time. (A) The geographic locations of Noccaea caerulescens accessions that contain the cytosine (C, green) or adenine (A, red) allele. (B) Manhattan plot for flowering time shows a 0.1-Mb flanking region of Contig16:117310. The colour of each point indicates r^2 value with Contig16:117310 as shown in the colour intensity scale at top-right. (C) Violin plot illustrates the flowering time differences between accessions carrying the A or C allele. Horizontal bars lines indicate the average per allele.*

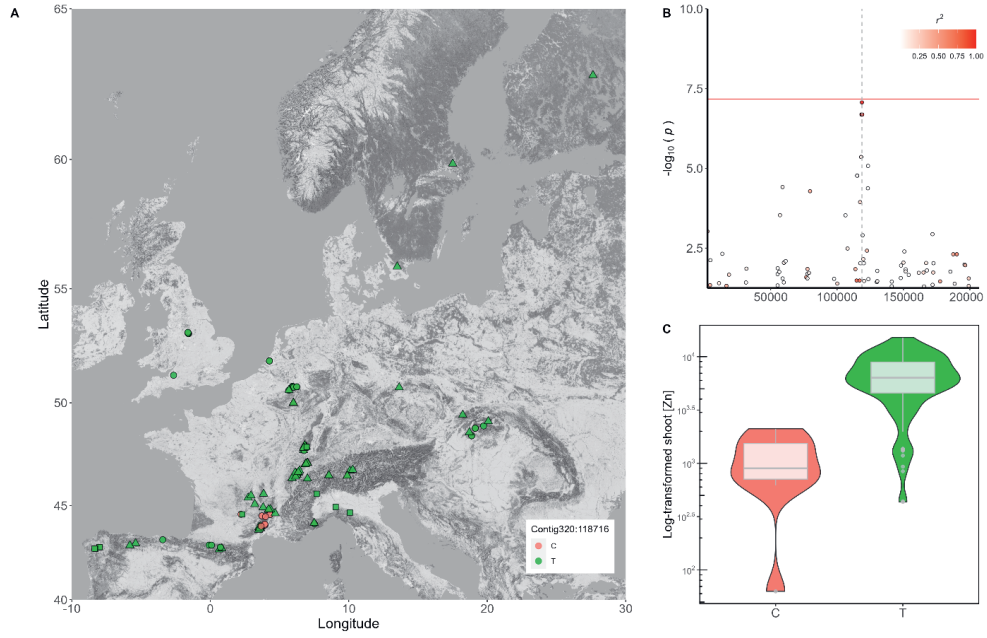


Figure 8. *The distribution and significance of allelic variation at SNP Contig320:118716 associated with shoot Zn concentration. (A) The geographic locations of Noccaea caerulescens accessions that contain the cytosine (C, green) or adenine (A, red) allele. (B) Manhattan plot for flowering time shows a 0.1-Mb flanking region of Contig320:118716. The colour of each point indicates r^2 value with Contig320:118716 as shown in the colour intensity scale at top-right. (C) Violin plot illustrates the flowering time differences between accessions carrying the C or T allele. Horizontal bars lines indicate the average per allele.*

Discussion

Preliminary information on evolutionary history of *N. caerulescens*

The evolutionary history of *N. caerulescens* has previously been extensively described by Gonneau et al. (2017), who revealed three genetic distinct groups (SU1, SU2, SU3) and one admixture group (AZ) of accessions from France (Massif Central and the Pyrenees), Belgium-Luxembourg (Ardennes) and Switzerland (Jura and the Alps). To extend this effort, this study investigated the population structure of a larger diversity panel than studied by Gonneau et al. (2017). This panel contains accessions also studied by Gonneau et al. (2017), and additional accessions, particularly from the UK, northern Europe (Scandinavian), Spain (Iberian mountains), Italy and central Europe (the Carpathians and the Bohemian massif). Our findings are largely coherent with those of Gonneau et al. (2017). We identify 16 accessions in SG5, from the south-eastern Massif Central and Viviez, to be the most genetically distant from the other accessions (figure 2). SG5 largely resembles SU1 identified by Gonneau et al. (2017). A total 8 accessions of 12 SU1 members are found in SG1. Similarly, SG1 and 4 align with SU3, and SG2 and 3 with SU2 (figure 1). Of 22 accessions of SU3, 17 accessions are observed in SG1 (10) and SG4 (7). One third of accessions in SU2 (7 out of 21) are identified in SG2 (5) and SG3 (2). We thus confirm the separation between accessions from the north-western Massif Central, the Pyrenees and the western Alps (SU2) and accessions from the Vosges, Jura, and Ardennes (SU3).

Due to the larger number of accessions, from a larger geographic region, our study gives a more detailed analysis of the population history of *N. caerulescens* than reported before. Next to the accessions in SU1-SU3/SG1-5, our study distinguishes SG6 in central Europe, as a separate SG. We also show a much less distinct geographic separation between the different SU/SGs than suggested by Gonneau et al. (2017). For instance, four accessions from the north-eastern Massif Central are closely related to accessions from the Apennines and the Iberian mountains (all in SG2), and in close proximity to not only accessions of SG3, which is most close to SG2, but also to accessions of SG4 and SG5. These findings largely support the hypothesis of Gonneau et al. (2017) on the origin of *N. caerulescens* populations in Western Europe to be post-glacially recolonised from refuges on the Iberian Peninsula or Italy, or even from an in situ survival in the south-eastern Massif Central, as appears to be the case for SG5. The accessions in SG5 are most distinct from the others, and this SG appears to have been separated for longer from the SG2 and SG3 accessions, that cluster together, and the SG1, 4 and 6 accessions. While SG6 is distinguished from SG1 and 4, they cluster together, and there is no reason to assume SG6 belongs to a different *N. caerulescens* subspecies, subsp. *caerulescens*, which has been described to reside in Eastern Central Europe, than the Western Central European subspecies, subsp. *sylvestris* (Koch and German, 2013) that is assumed to comprise all accessions in SG1-5 (Gonneau et al., 2017). The northward recolonisation of refugial populations from southern Europe during post-glacial period has been observed in several temperate species including *A. thaliana* (Hewitt, 2004; Lee et al., 2017).Assuming SG2 and 3 originate from Iberian and Italian refuges, and SG5 from a local, south-eastern Massif Central refuge, that survived the glacial period in situ, it is tempting to postulate that SG1, 4 and 6 recolonized Central Europe from a more eastern refuge, which could be the Balkans. This study is the first report of the relationship between accessions from Western Europe, and accessions from Central Europe (the Carpathians) and the eastern Swiss Alps. Within this Swiss Alps, two accessions: FON1 and FON2, located in Fontana, showed to fit in SG1 and SG3, unlike the surrounding Swiss Alps accessions. Although these physical distance between FON1 and FON2 was ~ 1.2 km, they were genetically quite distant. This suggests that FON1 and FON2 may have been introduced to the site by human activities. This is also the most likely explanation for six SG1 accessions from the north-eastern Jura (CHA1, CHA2, CHA3, MAM1, MAM2 and TDR2) and two SG4 accessions from the Vosges (AUX and PLA), that are genetically distant to accessions located in their proximity. The accessions from the eastern Swiss Alps (POS1, POS2, SCU2, SUS) belong to SG4, which are closely related to the ones from Vosges and Jura, that have been classified as SU3 in Gonneau et al. (2017). There are two scenarios for the origin of SG4: either this SG4 originated from east of the Alps and moved west or this SG4 originated from the Jura and Vosges and moved east. Previous studies reported the westward recolonisation from populations located in the eastern area of the Alps for e.g. *A. thaliana* (Lee et al., 2017) and *Arabidopsis halleri* (Pauwels et al., 2005). To validate whether this is also true for *N. caerulescens*, the collection of accessions from north-eastern Alps, between the areas of SG4 and SG6, and from further southeast, from Austria, north-eastern Italy and the Balkan countries, will be useful to resolve this further.

In general, geographic distance can act as a barrier to suppress gene flow between subgroups in their natural habitats. Furthermore, natural barriers such as mountains, deserts, rivers, lakes, sea and other geographical features can limit genetic exchange between populations (Holderegger and Wagner, 2006). When clustering of genetically distant accessions is observed despite geographic separation, this suggests human involvement. The colonisation of the British Isles likely has an anthropogenic origin, leading to the introduction from Western Europe. Even though we tested only few British accessions, they fit very well within SG1, the major Central European group. *N. caerulescens* seeds likely hitchhiked with miners and mining equipment from mining sites in Western Europe, such as from the former La Calamine mining site in Belgium and adjacent mining sites in Germany, where Zn ores may have been exposed to the surface

for some time (Koch et al., 1998). Thereafter, seeds could easily be dispersed over British mining sites by travelling miners (Ingrouille and Smirnoff 1986). This is in disagreement with the hypothesis of Koch et al. (1998), who assumed the British populations to represent an ancient relict of native British flora. However, this was based on only few isozyme markers, which, in retrospect, are likely too few to substantiate this conclusion. An anthropogenic origin has been suggested widely for Scandinavian *N. caerulescens*, as well (e.g. Koch et al., 1998, and refs. therein). Surprisingly though, although we only tested three Scandinavian accessions, they associate with two very different SGs, SG2 (south-western Massif Central) and SG6 (Carpathians). While this indeed suggests that the Scandinavian accessions are introduced, such introduction seems to have happened twice. According to our analysis, two Scandinavian accessions: KUO (Finland) and HOO (Sweden) likely originated from Central Europe, while UPP4 (Sweden) originated from the south-western Massif Central. More sampling in Scandinavia would allow for a more complete explanation of origin and later mixing of Scandinavian *N. caerulescens*, especially if CAL or ULT accessions could be tested, which are rare in Scandinavia. One potential CAL accession could be occurring at the Røros Copper Mine in Norway, which is so far the only Scandinavian accession growing on metallicolous soil (Bayçu et al., 2017).

This present study provided preliminary information for the evolutionary history of a comprehensive set of diverse *N. caerulescens* accessions from a substantial geographical area of Europe. Further population genetic analyses are needed to validate these hypotheses and elucidate accurate demographic scenarios for divergence of this *N. caerulescens* diversity panel.

The first GWAS in *N. caerulescens*

GWAS has been used to dissect complex traits in various plant species, especially *A. thaliana* with several association studies for different traits (review in Alseekh et al., 2021). In this study, we present the first GWAS for *N. caerulescens* which was previously hampered by scarcity of diverse accessions. The recent increase in sampling of accessions of *N. caerulescens* has allowed this study to conduct the first GWAS analysis for this species. Our study demonstrated that GWAS can detect significant SNPs for both life history trait (FT) and metal-related traits (Zn and Ni concentration). Very importantly, variation in loci associated with the traits is not limited to certain localities within Europe, but variants are spread and found in different subgroups of populations, allowing for separation of local adaptation and population history.

Our results reported that PVEs were much lower (< 1%) for significant SNPs associated with Zn/Ni treatment than FT. Assuming the significant SNPs are not false positives (but see below), this suggests that FT is likely to be controlled by a small number of genes with relatively large effects, while Zn/Ni accumulation might be controlled by many loci with relatively small effects (Table S2). In general, traits with polygenic underpinning with each variant having a small effect would need to large sample sizes to improve the statistical power for significant detection (Korte and Farlow, 2013). GWAS for Zn/Ni accumulation may be improved by collecting additional *N. caerulescens* accessions from distant metalliferous habitats.

Most of the significant SNPs were not in LD with surrounding SNPs (Figure 7B and 8B). When compared to *A. thaliana* GWAS, this would be unusual, although when compared to human GWAS it is not (Ke et al., 2008). Of course it could be that the significant SNPs we identified are false positives, despite obeying to the very stringent Bonferroni correction threshold, due to population structure artifacts. However, even for plants, such SNPs can be true positives either as true singletons (Ke et al., 2008) or as they located in

transposon elements (TEs) and actually correspond to other genomic locations than assumed now. Previous GWAS studies have identified TEs as true positives associated with traits of interest in tomato (Domínguez et al., 2020) and rice (Akakpo et al., 2020; Castanera et al., 2023). In tomato, 40 TE insertion polymorphisms has been identified as significant variants associated with 17 agronomic traits such as fruit colour and foliar morphology using GWAS. Unlike tomato and rice, structural genomic variation was not examined and described in *N. caerulescens*. Validation of TEs in *N. caerulescens* genome can be helpful for filtering false positives in the GWAS results of *N. caerulescens*. Another way to confirm these significant SNPs and candidate genes is to compare this GWAS result with QTLs result from bi-parental population derived from accessions of this panel (**Chapter 2**). The main limitation is that the identified significant SNPs/QTLs have to segregate in the bi-parental population.

Our findings identified two candidate genes, *FLAIL* and *ZIP2*, for FT and shoot Zn concentration, respectively. The alternative alleles of *FLAIL* were observed in late flowering accessions, that distributed across Europe and showed membership to all subgroups, except SG6 (Figure 7A). This supports the validity of this SNP association as a true positive. A recent study revealed that *FLAIL* corresponds to a long non-coding RNA, repressing flowering by activating *LACCASE 8* (*LAC8*) (Jin et al., 2023), which supports *FLAIL* as a candidate gene for FT in *N. caerulescens*. For *ZIP2*, allele contributed to low Zn shoot concentration was only observed in CAL accessions of SG5 distributed in small area of Central Massif (Figure 8A). This indicated that this association is a false positive as a result of confounding effect from population structure. It is unfortunate, but *ZIP2* is less likely to be candidate for Zn accumulation in *N. caerulescens.*

Several genes involving metal accumulation related traits have been reported in *N. caerulescens* using gene expression experiments and conventional QTL analysis in bi-parental populations (Assunção et al., 2010; Gendre et al., 2007; Halimaa et al., 2014; Halimaa et al., 2014; Milner et al., 2014; Mortel et al., 2008; Oomen et al., 2009; Papoyan and Kochian, 2004; Plaza et al., 2007; Rigola et al., 2006; Ueno et al., 2011; van de Mortel et al., 2006; Wang et al., 2020), however, none of these genes were detected in the GWAS. The same is the case for FT. For instance, *FLOWERING LOCUS C* (*FLC*) and *SHORT VEGETATIVE PHASE* (*SVP*) have been described as major regulators of FT in *N. caerulescens* (Wang et al., 2020). Similar as described in **Chapter 3**, also GWAS does not pick up genetic variation associated with these genes. Other than that these genes are not under adaptive selection in *N. caerulescens* as suggested in **chapter 3**, it could be that GWAS is ineffective when target traits are strongly associated with population structure (Korte and Farlow, 2013). Such may be partially the case here. Therefore, this diversity panel can be improved by removing genetically very closely related accessions, collected at geographically close sites, such as from AVI, GAL, SDO, BER, CMT, FSM, LTH, CHA, BRE and ROZ, and to add accessions from other locations. In this way, trait values might become less related to population structure (Brachi et al., 2011). This can improve the power of GWAS, which has been successful in *A. thaliana*, such as RegMap populations (Horton et al., 2012).

Conclusion

This study reveals six genetically distinct subgroups of a comprehensive *N. caerulescens* diversity panel from a large geographical area of Europe. A preliminary comprehensive evolutionary history of *N. caerulescens* is described, which is the first report for *N. caerulescens* from Iberian mountains, the Carpathians and UK. Moreover, this study demonstrated that this *N. caerulescens* panel can be used for association studies, as this study identified *FLAIL* as a candidate gene for FT. Still, further optimization, e.g. adding accessions and collection of more geographically distant metallicolous accessions, can enhance

and improve accuracy for QTL detection. This genotyped diversity panel of *N. caerulescens* will be useful for future genetic analysis of heavy metal hyperaccumulation and tolerance in plants.

Supplementary information

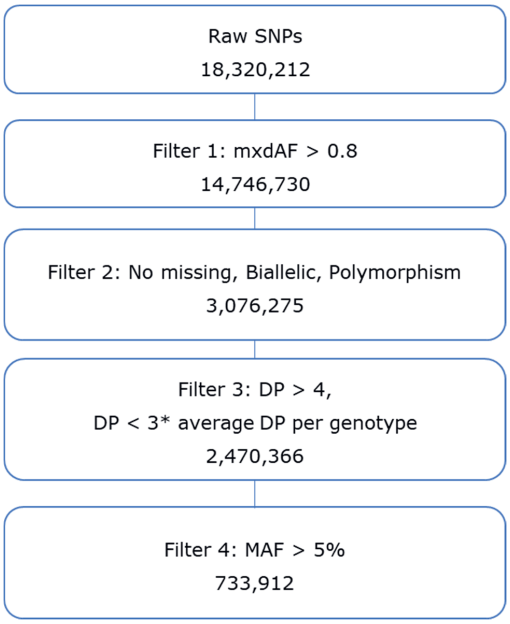


Figure S1. *Pipelines showing the number of SNPs during each filtering procedures. Different softwares were used for each filtering: filter 1 and 3 using R software; filter 2 using VCFtools with these parameter (--min-alleles 2 --max-alleles 2 --max-missing 1.0) and BCFtools with this parameter (-a -c 1); and filter 4 using GEMMA with these parameters (-maf 0.05).*

Table S1. *Information of the Noccaea caerulescens diversity panel including: accession number (used for seed storage of the Laboratory of Genetics, Wageningen University, Netherlands); accession name and acronym; collection site information (edaphic type (ecotype), country, locality, GPS coordinates) and Subgroup (SG) of each accession identified in this study.*

No.	Accession number	Accession name	Acronym	Ecotype	Country	Locality	Latitude	Longitude	Subgroup
1	190002	Rue du Foret	RF	NM	Belgium	Ardennes	50.5801	5.6788	SG1
2	170003	La Calamine	LC	CAL	Belgium	Ardennes	50.7108	6.0104	SG1
3	170004	St. Felix de Pallières	SF	CAL	France	Massif Central	44.0445	3.9383	SG5
4	170005	Lellingen	LE	NM	Luxemburg	Ardennes	49.9838	5.9942	SG1
5	170006	Monte Prinzera	MP	ULT	Italy	Apennines	44.6424	10.0847	SG3
6	170007	Krušné hory Mountains	KHM	NM	Czech	Bohemian Massif	50.7046	13.6431	SG6
7	170008	Moravsko-slezské Beskydy Mountains	MSB	NM	Czech	Carpathians	49.4359	18.2297	SG1
8	170009	Durfort	DUR	CAL	France	Massif Central	44.0003	3.9511	SG5
9	170010	Plombieres	PB	CAL	Belgium	Ardennes	50.7345	5.9638	SG1
10	170012	Priddy Mineries	PDY	CAL	UK	British	51.2598	-2.6500	SG1
11	170017	Ganges	GA	CAL	France	Massif Central	NA	NA	SG5
12	170018	Clough Wood	CLW	CAL	UK	British	53.1527	-1.6157	SG1
13	170019	Wilwerwilz	WIL	NM	Luxemburg	Ardennes	49.9913	6.0030	SG1
14	170021	Vall de Varrados	VDV	NM	Spain	Pyrenees	42.7763	0.8275	SG2
15	170022	Pontaut	PON	CAL	Spain	Pyrenees	42.8342	0.7334	SG2
16	170023	Lanestosa	LAN	CAL	Spain	Iberian mountains	43.2311	-3.4327	SG3
17	170024	Navacelles	NAV	NM	Spain	Massif Central	43.8875	3.5091	SG5
18	170025	Prayon	PR	CAL	Belgium	Ardennes	50.5813	5.6668	SG1
19	170026	Mezilhac	MEZ	NM	France	Massif Central	44.8078	4.3458	SG3
20	170027	Jean Arsac	JAR	NM	France	Massif Central	44.7743	4.2037	SG3
21	170029	Sainte Eulalie	SEU	NM	France	Massif Central	44.8036	4.2131	SG3
22	170030	Les Avinières-1	AVI1	CAL	France	Massif Central	43.9364	3.6714	SG5
23	170034	Puerto de Aralla	ARA	NM	Spain	Iberian mountains	42.9294	-5.8121	SG3
24	170035	Bandeira	BAN	ULT	Spain	Iberian mountains	42.7626	-8.2842	SG3
25	170037	Baraquette	BAR	NM	France	Massif Central	43.9169	3.6152	SG5
26	170038	Puente Basadre	BAS	ULT	Spain	Iberian	42.8462	-7.9870	SG3
27	170039	Bergenbach-1	BER1	ULT	France	Vosges	47.9044	6.9568	SG1

No.	Accession number	Accession name	Acronym	Ecotype	Country	Locality	Latitude	Longitude	Subgroup
28	170043	Le Blémard	BLE	CAL	France	Massif Central	44.4714	3.7286	SG5
29	170044	Cira	CIR	ULT	Spain	Iberian mountains	42.7667	-8.3500	SG3
30	170045	Col d'Oderen	COD	NM	France	Vosges	47.9229	6.9256	SG1
31	170047	Coulet	COU	NM	France	Massif Central	43.8291	3.5616	SG5
32	170049	Spania Dolina-1	SDO1	CAL	Slovakia	Carpathians	48.8094	19.1327	SG6
33	170050	Eidian	EID	ULT	Spain	Iberian mountains	42.8315	-8.0034	SG3
34	170052	La Galène-1	GAL1	CAL	France	Pyrenees	42.9457	-0.0861	SG2
35	170053	Kuopio	KUO	NM	Finland	Finland	62.8969	27.6348	SG6
36	170055	Saint Baudille	SBD	NM	France	Massif Central	43.7446	3.4861	SG5
37	170056	Höör	HOO	NM	Sweden	Sweden	55.9000	13.5000	SG6
38	170057	Puerto de San Isidro	ISI	NM	Spain	Iberian mountains	43.0394	-5.3939	SG3
39	170058	Viviez	VIV	CAL	France	Massif Central	44.5537	2.2190	SG5
40	170059	Werschmatt	WER	NM	France	Vosges	47.9187	6.7807	SG1
41	170157	Vyšná Boca	VYB	CAL	Slovakia	Carpathians	48.9155	19.7338	SG6
42	170158	Popradské pleso	POP	NM	Slovakia	Carpathians	49.1316	20.0772	SG6
43	170159	Horná Roveň-1	HRO1	CAL	Slovakia	Carpathians	48.4487	18.8755	SG6
44	170160	Hospice de Luchon	HDL	NM	France	Pyrenees	42.7152	0.6503	SG2
45	170187	Valle Gargassino	VGG	ULT	Italy	Apennines	44.9275	9.0561	SG3
46	170191	Col du Mas de l'Ayre	CMA	CAL	France	Massif Central	44.4303	3.9878	SG5
47	180001	Auxelles	AUX	CAL	France	Vosges	47.7393	6.7766	SG4
48	180003	Largentiere	LAR	CAL	France	Massif Central	44.5406	4.3051	SG5
49	180011	Pic de Chenavari	PDC	NM	France	Massif Central	44.5997	4.6845	SG5
50	180012	L'Herm	LHE	NM	France	Massif Central	44.9067	3.8195	SG2
51	180013	Montchamp	MON	NM	France	Massif Central	45.0686	3.2126	SG2
52	180014	La Chavignee	LCH	NM	France	Massif Central	45.3968	2.7331	SG2
53	180015	Besse	BES	NM	France	Massif Central	45.5071	2.9406	SG2
54	180016	Col des Supeyres	CSU	NM	France	Massif Central	45.5900	3.8487	SG3
55	180018	Le Cernois	LCE	NM	France	Jura	46.6533	6.1333	SG4
56	180020	Sepmoncel	SEP	NM	France	Jura	46.3753	5.9077	SG4
57	180021	Fresse sur Moselle-1	FSM1	NM	France	Vosges	47.8718	6.7915	SG1

No.	Accession number	Accession name	Acronym	Ecotype	Country	Locality	Latitude	Longitude	Subgroup
58	180023	Cornimont-3	CMT3	NM	France	Vosges	47.9843	6.8181	SG1
59	180024	Bueges	BUE	NM	France	Massif Central	43.8298	3.6185	SG5
60	180028	Vel'ka Fatra Mountain	VFM	NM	Slovakia	Carpathians	48.6039	18.7069	SG6
61	180031	Mont d'Amin-1	MAM1	NM	Switzerland	Jura	47.0866	6.9019	SG1
62	180032	Tête de Ran	TDR1	NM	Switzerland	Jura	47.0556	6.8564	SG4
63	180035	Au Dent de Vaulion	ADV	NM	Switzerland	Jura	46.6800	6.3550	SG4
64	180037	Breiniger-1	BRE1	CAL	Germany	Ardennes	50.7370	6.2450	SG1
65	180038	Breiniger-2	BRE2	CAL	Germany	Ardennes	50.7370	6.2420	SG1
66	180040	Col du Marchairuz 2	MAR2	NM	Switzerland	Jura	46.5453	6.2511	SG4
67	180041	Chasseron-1	CHA1	NM	Switzerland	Jura	47.1290	7.0122	SG1
68	180042	Chasseron-2	CHA2	NM	Switzerland	Jura	47.1290	7.0122	SG1
69	180043	Chasseron-3	CHA3	NM	Switzerland	Jura	47.1325	7.0325	SG1
70	180044	Col du Mollendruz	CMO	NM	Switzerland	Jura	46.6456	6.3572	SG4
71	180045	Cornimont-1	CMT1	NM	France	Vosges	47.9670	6.8299	SG1
72	180046	Cornimont-2	CMT2	NM	France	Vosges	47.9820	6.8199	SG1
73	180047	Fresse sur Moselle-2	FSM2	NM	France	Vosges	47.8718	6.7915	SG1
74	180048	Horná Roveň-2	HRO2	CAL	Slovakia	Carpathian	48.4514	18.8738	SG6
75	180049	Le Suchet	LES	NM	Switzerland	Jura	46.7533	6.4458	SG4
76	180050	Leysin-2	LEY2	NM	Switzerland	Jura	46.3553	7.0200	SG4
77	180051	Le Thillot-2	LTH2	NM	France	Vosges	47.8646	6.7544	SG1
78	180052	Le Thillot-5	LTH5	NM	France	Vosges	47.8627	6.7374	SG1
79	180053	Mont d'Amin-2	MAM2	NM	Switzerland	Jura	47.0856	6.9019	SG1
80	180054	Plancher-BAS	PLA	NM	France	Vosges	47.7511	6.7311	SG4
81	180055	St.Cergeus	SCE	NM	Switzerland	Jura	46.4556	6.1625	SG4
82	180056	Špania Dolina-2	SDO2	CAL	Slovakia	Carpathians	48.8078	19.1350	SG6
83	180057	Tête de Ran-2	TDR2	NM	Switzerland	Jura	47.0542	6.8542	SG1
84	180058	Vallée de Joux	VDJ	NM	Switzerland	Jura	46.6067	6.2150	SG4
85	180059	Rotterdam Rozenburg-1	ROZ1	CAL	Netherland	Dutch	51.9025	4.2677	SG1
86	180060	Rotterdam Rozenburg-2	ROZ2	CAL	Netherland	Dutch	51.9012	4.2705	SG1
87	180094	Les Avinières-2	AVI2	CAL	France	Massif Central	43.9364	3.6714	SG5

No.	Accession number	Accession name	Acronym	Ecotype	Country	Locality	Latitude	Longitude	Subgroup
88	180102	Bergenbach-2	BER2	ULT	France	Vosges	47.9061	6.9613	SG1
89	180110	Špania Dolina-3	SDO3	CAL	Slovakia	Carpathians	47.0542	6.8542	SG6
90	180112	La Galène-2	GAL2	CAL	France	Pyrenees	42.9467	0.0867	SG2
91	180164	Ganges parent RIL	GA-RIL	CAL	France	Massif Central	43.9364	3.6714	SG5
92	180166	Lellingen parent RIL	LE-RIL	NM	Luxemburg	Ardennes	49.9913	6.0030	SG1
93	180101	Bergenbach-3	BER3	ULT	France	Vosges	47.9064	6.9571	SG1
94	180115	Puy de Wolf	PDW	ULT	France	Massif Central	44.5497	2.3100	SG2
95	190102	Fontana-2	FON2	NM	Switzerland	Alps	46.5155	8.5483	SG1
96	190104	Scuol-2	SCU2	NM	Switzerland	Alps	46.7950	10.2849	SG4
97	190106	Susch	SUS	NM	Switzerland	Alps	46.7502	10.0817	SG4
98	190107	Postresina-1	POS1	NM	Switzerland	Alps	46.4977	9.8907	SG4
99	190108	Postresina-2	POS2	NM	Switzerland	Alps	46.4965	9.8922	SG4
100	190109	Geuldal-1	GEU1	CAL	Netherland	Dutch	50.7608	5.9313	SG1
101	190112	Uppsala-4	UPP4	NM	Sweden	Sweden	59.8453	17.5100	SG2
102	190113	Casterino	CAS	NM	Italy	Alps	44.1002	7.5046	SG2
103	190114	Pontboset	POB	ULT	Italy	Alps	45.5937	7.6801	SG3
104	190116	Vallauria mine	VLM	CAL	France	Alps	44.0703	7.5166	SG2
105	190117	Bonsall	BNS	CAL	UK	British	53.1122	-1.6056	SG1
106	190118	Cromford	CRF	CAL	UK	British	53.0986	-1.5728	SG1
107	190120	Black Rocks	BR	CAL	UK	British	53.0978	-1.5650	SG1
108	190122	Fontana-1	FON1	NM	Switzerland	Alps	46.5187	8.5630	SG3
109	180029	Col du Marchairuz	MAR1	NM	Switzerland	Jura	46.5408	6.2558	SG4

NA: Not available

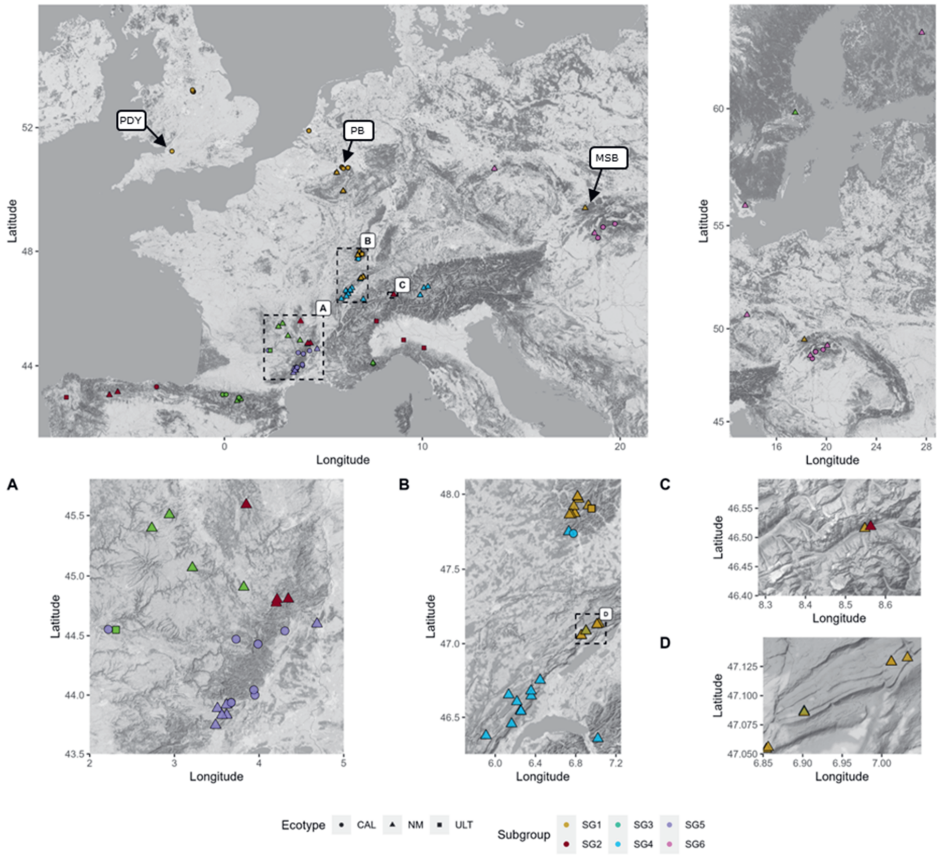


Figure S3. *A topographic map indicates locations of 109 Noccaea caerulescens accessions across Europe with zooms into four regions including Massif Central (A); Vosges and Jura (B); Fontana (C); and northeast Jura (D). Each subgroup is labeled with different colours (Yellow = SG1, Red = SG2, Cyan = SG3, Blue = SG4, Purple = SG5, Pink = SG6). Shapes represent three different ecotypes including Calamine (circle), Non-metallicolous (triangle), and Ultramafic (rectangular).* Three SG1 accessions are indicated by black arrows including Moravsko-slezské Beskydy Mountains (MSB), Plombieres (PB) and Priddy Mineries (PDY).

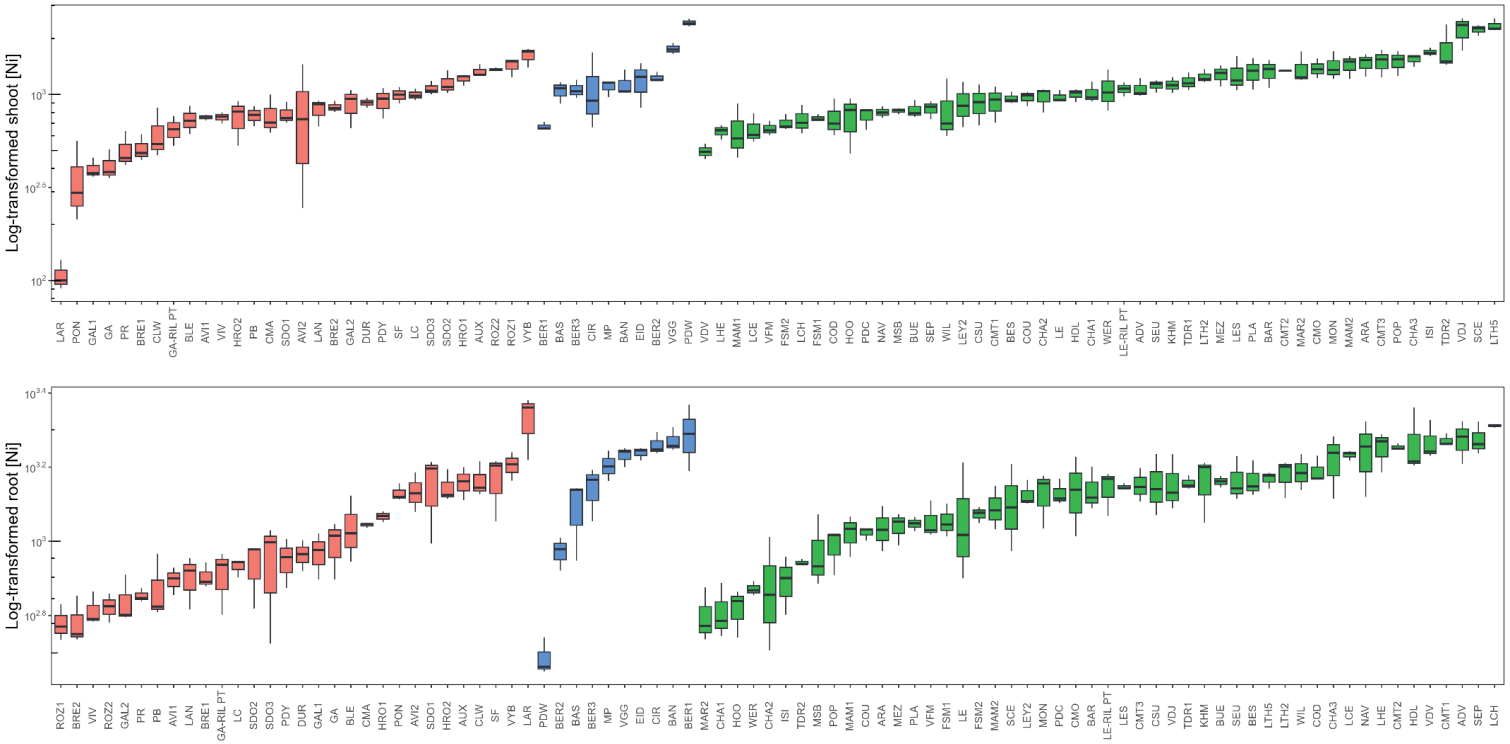


Figure S4. *Ni accumulation of 90 Noccaea caerulescens accessions upon Zn/Ni treatment. Box plots are drawn per genotype, that are ranked from low to high concentration for shoot (top) and root (bottom). The black lines indicate the average per accession. Each plot is coloured according to ecotypes including Calamine (red), Non-metallicolous (green), and Ultramafic (blue).*

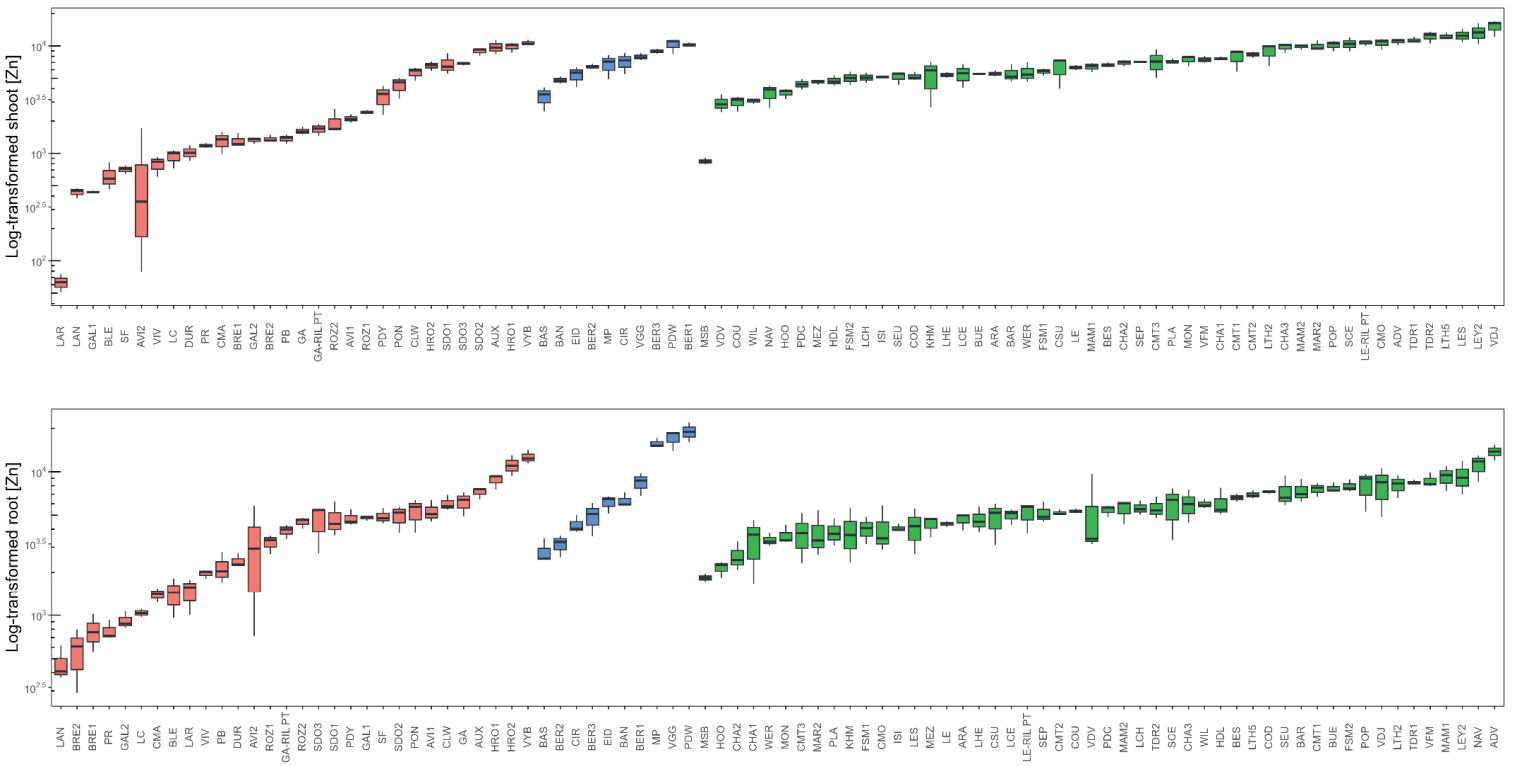


Figure S5. *Zn accumulation of 90 Noccaea caerulescens accessions upon Zn/Ni treatment. Box plots are drawn per genotype, that are ranked from low to high concentration. The black lines indicate the average per accession. Each plot is coloured according to ecotypes including Calamine (red), Non-metallicolous (green), and Ultramafic (blue).*

Table S2. *Significant single nucleotide polymorphisms detected in GWAS trial for flowering time, shoot and root Ni concentration of Ni and Zn.*

Trait	Significant SNP	-log10(p)	LG	Allele	AF	PVE	Arabidopsis syntenic block (chromosome number: bp…bp)	Arabidopsis genes in the syntenic block
FT	Contig145:453391	8.12	1	A/T	0.10	14.95	Chr1:29962929...29999159	AT1G79640, AT1G79650, AT1G79660, AT1G79670, AT1G79680, AT1G79690, AT1G79700, AT1G79710, AT1G79720
	Contig0:1531473	7.71	4	G/A	0.10	8.78	Chr5:1428788...1441568	AT5G04900, AT5G04910, AT5G04920
	Contig7:3678132	7.13	5	C/T	0.11	12.91	-	-
	Contig115:97125	7.18	5	T/G	0.14	13.21	Chr3:9529063...9564858	AT3G26085, AT3G26090, AT3G26100, AT3G26110, AT3G26115, AT3G26120, AT3G26125, AT3G26130, AT3G26140
	Contig56:1767611	7.27	5	T/A	0.06	13.65	Chr3:20715031...20744999	AT3G55840, AT3G55850, AT3G55860, AT3G55870, AT3G55880, AT3G55890, AT3G55900, AT3G55910, AT3G55920
	Contig16:117310	7.30	6	A/C	0.21	8.00	Chr2:8096877...8128273	AT2G18690, AT2G18700, AT2G18710, AT2G18720, AT2G18721, AT2G18730, AT2G18735, AT2G18740, AT2G18750
	Contig16:167351	7.94	6	T/A	0.04	15.23	Chr4:17341505...17344611	AT4G36808, AT4G36810
Root [Ni]	Contig115:499251	7.71	5	G/T	0.19	0.12	Chr3:9295286...9303421	AT3G25585, AT3G25590
Shoot [Ni]	Contig11:4545454	7.50	3	A/C	0.03	0.29	Chr1:2200685...2284377	AT1G07175, AT1G07180, AT1G07190, AT1G07200, AT1G07210, AT1G07220, AT1G07230, AT1G07240, AT1G07250, AT1G07260, AT1G07270, AT1G07280, AT1G07290, AT1G07300, AT1G07310, AT1G07320, AT1G07330, AT1G07340, AT1G07350, AT1G07360, AT1G07370, AT1G07380, AT1G07390, AT1G07400, AT1G07410, AT1G07420, AT1G07430
Root [Zn]	Contig54:697756	7.64	1	T/A	0.26	0.26	-	-
	Contig11:2108766	7.36	3	T/C	0.16	0.23	Chr1:4278219...4296469	AT1G12580, AT1G12590, AT1G12600, AT1G12610, AT1G12620
	Contig31:2868620	7.80	4	G/A	0.05	0.42	Chr5:6958605...6968100	AT5G20590, AT5G20600
	Contig31:2983198	8.36	4	A/T	0.05	0.41	Chr5:7061327...7138352	AT5G20850, AT5G20852, AT5G20854, AT5G20856, AT5G20858, AT5G20860, AT5G20870, AT5G20885, AT5G20890, AT5G20900, AT5G20910, AT5G20920, AT5G20930, AT5G20935, AT5G20940, AT5G20950, AT5G20960, AT5G20970, AT5G20980, AT5G20990, AT5G21005, AT5G21010

Trait	Significant SNP	-log10(p)	LG	Allele	AF	PVE	Arabidopsis syntenic block (chromosome number: bp...bp)	Arabidopsis genes in the syntenic block
	Contig31:2983781	8.24	4	T/A	0.06	0.40	Chr5:7061327...7138352	AT5G20850, AT5G20852, AT5G20854, AT5G20856, AT5G20858, AT5G20860, AT5G20870, AT5G20885, AT5G20890, AT5G20900, AT5G20910, AT5G20920, AT5G20930, AT5G20935, AT5G20940, AT5G20950, AT5G20960, AT5G20970, AT5G20980, AT5G20990, AT5G21005, AT5G21010
	Contig0:6344338	7.12	4	C/G	0.13	0.25	Chr5:5483219...5496798	AT5G16710, AT5G16715, AT5G16720
	Contig48:278921	8.12	6	A/G	0.12	0.29	Chr3:5896231...5897774	AT3G17261
	Contig71:197407	7.39	6	G/C	0.06	0.34	Chr3:5399074...5418357	AT3G15960, AT3G15970, AT3G15980
	Contig71:200042	7.39	6	G/C	0.06	0.34	Chr3:5399074...5418357	AT3G15960, AT3G15970, AT3G15980
	Contig135:212090	8.11	1	C/T	0.10	0.45	-	-
Shoot [Zn]	Contig211:324014	8.14	2	A/T	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig211:324091	8.14	2	A/G	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig211:325130	8.14	2	G/T	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig211:326265	8.14	2	A/C	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig211:326270	8.14	2	A/C	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig211:326522	8.14	2	C/T	0.06	1.04	Chr5:16311287...16316622	AT5G40770
	Contig201:308294	7.26	2	C/T	0.11	0.38	Chr4:10021984...10032336	AT4G18050, AT4G18060, AT4G18070
	Contig201:625143	7.14	2	G/T	0.13	0.45	Chr4:9876426...9908816	AT4G17780, AT4G17785, AT4G17788, AT4G17790, AT4G17800, AT4G17810
	Contig154:151734	8.23	2	A/T	0.16	0.36	Chr4:10322053...10334090	AT4G18810, AT4G18815, AT4G18820
	Contig154:151743	9.52	2	C/G	0.09	0.41	Chr4:10322053...10334090	AT4G18810, AT4G18815, AT4G18820
	Contig5:2352035	8.39	2	A/G	0.12	0.37	Chr4:12954016...12987660	AT4G25320, AT4G25330, AT4G25340, AT4G25350, AT4G25360, AT4G25370, AT4G25380, AT4G25386, AT4G25390, AT4G25400, AT4G25410
	Contig5:2352044	8.39	2	G/T	0.12	0.37	Chr4:12954016...12987660	AT4G25320, AT4G25330, AT4G25340, AT4G25350, AT4G25360, AT4G25370, AT4G25380, AT4G25386, AT4G25390, AT4G25400, AT4G25410
	Contig257:444532	7.08	3	A/T	0.06	0.45	Chr1:20632673...20676063	AT1G55320, AT1G55325, AT1G55330, AT1G55340, AT1G55350, AT1G55360, AT1G55365, AT1G55370

Trait	Significant SNP	-log10(p)	LG	Allele	AF	PVE	Arabidopsis syntenic block (chromosome number: bp…bp)	Arabidopsis genes in the syntenic block
	Contig102:578478	7.64	3	C/T	0.49	1.56	Chr1:19499199...19511414	AT1G52370, AT1G52380
	Contig105:206401	7.06	3	T/C	0.08	0.39	Chr1:11738857...11751996	AT1G32470, AT1G32480, AT1G32490, AT1G32500
	Contig28:2166541	7.08	3	G/C	0.06	0.45	-	-
	Contig23:374828	7.48	3	G/C	0.06	0.40	Chr1:11033743...11043731	AT1G30950, AT1G30960, AT1G30970
	Contig100:167258	10.02	3	T/C	0.02	0.69	Chr1:8527768...8535544	AT1G24110, AT1G24120, AT1G24130
	Contig11:344595	7.51	3	C/A	0.06	0.44	Chr1:5554911...5604343	AT1G16240, AT1G16250, AT1G16260, AT1G16270, AT1G16280, AT1G16290, AT1G16300, AT1G16310, AT1G16320, AT1G16330, AT1G16340, AT1G16350, AT1G16360, AT1G16370, AT1G16380, AT1G16390
	Contig11:4739478	7.09	3	T/A	0.06	0.39	Chr1:2032821...2081683	AT1G06645, AT1G06650, AT1G06660, AT1G06670, AT1G06680, AT1G06690, AT1G06700, AT1G06710, AT1G06720, AT1G06730, AT1G06750, AT1G06760, AT1G06770
	Contig31:934491	7.40	4	G/A	0.13	0.38	Chr4:897976...905473	AT4G02050, AT4G02060, AT4G02055
	Contig31:1473097	7.32	4	T/G	0.11	0.49	Chr4:1319896...1334703	AT4G02990, AT4G03000, AT4G03010, AT4G03020
	Contig320:118716	7.07	4	C/T	0.11	0.40	Chr5:23985785...24007676	AT5G59500, AT5G59505, AT5G59510, AT5G59520, AT5G59530, AT5G59540, AT5G59550, AT5G59560, AT5G59570, AT5G59580
	Contig0:3913141	9.00	4	T/C	0.06	0.43	Chr5:3567215...3570563	-
	Contig0:5588140	10.65	4	G/A	0.35	0.32	Chr5:4677448...4786595	AT5G14510, AT5G14520, AT5G14530, AT5G14540, AT5G14550, AT5G14545, AT5G14560, AT5G14565, AT5G14570, AT5G14580, AT5G14590, AT5G14600, AT5G14602, AT5G14610, AT5G14620, AT5G14640, AT5G14650, AT5G14660, AT5G14670, AT5G14680, AT5G14690, AT5G14700, AT5G14710, AT5G14720, AT5G14730, AT5G14740, AT5G14750, AT5G14760, AT5G14770, AT5G14780, AT5G14790, AT5G14800
	Contig0:5588192	8.63	4	C/T	0.35	0.30	Chr5:4677448...4786595	AT5G14510, AT5G14520, AT5G14530, AT5G14540, AT5G14550, AT5G14545, AT5G14560, AT5G14565, AT5G14570,

Trait	Significant SNP	-log10(p)	LG	Allele	AF	PVE	Arabidopsis syntenic block (chromosome number: bp…bp)	Arabidopsis genes in the syntenic block
								AT5G14580, AT5G14590, AT5G14600, AT5G14602, AT5G14610, AT5G14620, AT5G14640, AT5G14650, AT5G14660, AT5G14670, AT5G14680, AT5G14690, AT5G14700, AT5G14710, AT5G14720, AT5G14730, AT5G14740, AT5G14750, AT5G14760, AT5G14770, AT5G14780, AT5G14790, AT5G14800
	Contig0:5790034	7.72	4	T/G	0.13	0.33	Chr5:4958684...4977090	AT5G15280, AT5G15290, AT5G15300, AT5G15310
	Contig0:6342065	7.56	4	T/C	0.13	0.39	-	-
	Contig115:510909	7.15	5	A/G	0.12	0.56	-	-
	Contig123:119671	7.43	5	G/A	0.09	0.41	Chr2:518258...522003	-
	Contig123:138198	7.43	5	C/G	0.09	0.41	Chr2:511898...512084	AT2G02080
	Contig123:138242	7.43	5	T/A	0.09	0.41	Chr2:511898...512084	AT2G02080
	Contig123:147495	7.43	5	A/C	0.09	0.41	Chr2:487860...510594	AT2G02050, AT2G02060, AT2G02061, AT2G02070
	Contig56:2612746	8.93	5	C/T	0.06	0.45	Chr3:19945590...19977148	AT3G53850, AT3G53860, AT3G53870, AT3G53880, AT3G53890, AT3G53900, AT3G53910, AT3G53920, AT3G53930, AT3G53940, AT3G53950
	Contig199:1063368	8.38	5	A/G	0.03	0.62	Chr3:22416004...22426373	AT3G60647, AT3G60650, AT3G60660, AT3G60670
	Contig199:1213418	7.52	5	T/C	0.09	0.41	Chr3:22297525...22306370	AT3G60328, AT3G60330, AT3G60340
	Contig199:1213448	10.03	5	C/T	0.14	0.36	Chr3:22297525...22306370	AT3G60328, AT3G60330, AT3G60340
	Contig199:1213852	7.18	5	G/T	0.16	0.32	Chr3:22297525...22306370	AT3G60328, AT3G60330, AT3G60340
	Contig2:1711450	7.52	6	G/C	0.09	0.41	Chr3:1466357...1478364	AT3G05170, AT3G05180, AT3G05190, AT3G05193, AT3G05200
	Contig2:1722790	8.20	6	G/T	0.11	0.44	Chr3:1476480...1478527	AT3G05200
	Contig2:2392120	7.43	6	T/C	0.09	0.41	Chr3:1917340...1934331	AT3G06335, AT3G06340, AT3G06350, AT3G06360, AT3G06370
	Contig96:270640	8.26	6	C/T	0.08	0.43	Chr2:2055225...2056626	AT2G05580
	Contig16:108876	7.13	6	C/A	0.06	0.35	Chr2:8096877...8128273	AT2G18690, AT2G18700, AT2G18710, AT2G18720, AT2G18721, AT2G18730, AT2G18735, AT2G18740, AT2G18750
	Contig16:297049	8.26	6	C/A	0.07	0.42	Chr2:7928489...7956045	AT2G18230, AT2G18240, AT2G18245, AT2G18250, AT2G18260, AT2G18280,

Trait	Significant SNP	-log10(p)	LG	Allele	AF	PVE	Arabidopsis syntenic block (chromosome number: bp…bp)	Arabidopsis genes in the syntenic block
								AT2G18270, AT2G18290, AT2G18300, AT2G18310
	Contig16:1100640	7.20	6	T/A	0.11	0.40	-	-
	Contig95:1388937	7.20	7	G/A	0.09	0.37	Chr4:13032907...13035740	-
	Contig122:1157278	7.17	7	G/A	0.03	0.46	Chr2:16452715...16454570	-
	Contig178:298719	8.18	7	A/C	0.07	0.76	-	-
	Contig12:139985	11.26	7	A/T	0.12	0.53	Chr2:12295431...12309470	AT2G28671, AT2G28670, AT2G28680, AT2G28690
	Contig12:2536392	7.43	7	G/C	0.09	0.41	Chr2:10437634...10460486	AT2G24580, AT2G24590, AT2G24592, AT2G24600, AT2G24610
	Contig12:2989908	7.68	7	G/T	0.11	0.32	-	-

LG - Linkage group; AF - Allele frequency; PVE - Percentage variance explained

Acknowledgement

The authors thank Guillaume Echevarria, Thibault Sterckeman and Catherine Sirguey (INRA, France) for providing seeds of *Noccaea caerulescens* accessions Auxelles, Largentiere, Pic de Chenavari, L'Herm, Montchamp, La Chavignee, Besse, Col des Supeyres, Le Cernois, Sepmoncel, Fresse sur Moselle, Cornimont, Bueges, Vel'ka Fatra Mountain, Mont d'Amin, Tête de Ran, and Au Dent de Vaulion; Sylvian Merlot is acknowledged for providing seeds of Bergenbach; Nausicaa Noret (ULB, Belgium) is acknowledged for providing seeds of Rue du Foret, Casterino, Pontboset, and Vallauria mine; Sophie Laurent (ULB, Belgium) is acknowledged for seeds from Black Rocks, Bonsall, and Cromford; Roel van Bezouw is acknowledged for providing the seeds of Rozenburg; Enrica Roccotiello (University of Genoa, Italy) is acknowledged for providing seeds of Valle Gargassino; Dario Galanti (University of Tübingen, Germany) is acknowledged for providing seeds of Uppsala; Terezie Mandáková (CEITEC, Czech Republic) is acknowledged for providing seeds of Höör, Špania Dolina and La Galène; Radim J. Vašut (Palacký University Olomouc, Czech Republic) is acknowledged for providing seeds of Krušné hory Mountains and Moravsko-slezské Beskydy Mountains; Authors thank Ben Auxier (Wageningen University) for helpful supports on SNP calling procedure; Tom Theeuwen (Wageningen University) for recommendations on GWAS analysis; and the Royal Thai government (Thailand) for funding this research.

CHAPTER

5

An efficient flower dipping transformation method for the metal hyperaccumulator *Noccaea caerulescens*

Jitpanu Yamjabok, Lies van der Heijden, Maarten Koornneef,
Henk Schat, Mark G.M. Aarts

Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1,
6708 PB Wageningen, the Netherlands

Financial source:
This research is financially supported by Royal Thai government (Thailand).

Abstract

Noccaea caerulescens is a metal hyperaccumulator plant species with the capacity to accumulate high concentrations of zinc (Zn), cadmium (Cd) and nickel (Ni). Many different natural populations can be found at different sites in Europe, displaying a wide range of metal tolerance and metal accumulation phenotypes, in addition to variation in plant architecture and plant growth. This variation makes it an attractive model to study plant adaptation to metal exposure, but also to develop it for application in phytomining and phytoremediation. Several candidate genes have so far been associated with accumulation and tolerance of these metals, but gene function analysis has been cumbersome, in the absence of an efficient transformation method for *N. caerulescens*. To address this issue, an identified mutation in the *FLOWERING LOCUS C*, conferring early flowering to the calamine accession St. Felix de Pallieres, was introgressed in the genetic background of five different genotypes, calamine accessions Clough Wood (CLW) and Le Bleymard (BLE), ultramafic accessions Cira (CIR), and non-metallicolous accessions St. Baudille (SBD) and Werschmatt (WER). The findings revealed that three of these introgression lines were successfully transformed via *Agrobacterium tumefaciens* floral dipping, achieving an average transformation efficiency > 0.29%. This study presented a reproducible floral dipping transformation method is presented here as an improvement on previously reported methods. The early flowering N. caerulescens *flc* mutation can be readily introgressed in different genetic backgrounds to facilitate genetic transformation studies, especially for gene function analysis.

Introduction

Noccaea caerulescens (J. and C. Presl) F.K. Mey., previously named *Thlaspi caerulescens*, is an annual, temperate plant species, that can grow and survive on soil rich in cadmium (Cd), lead (Pb) and zinc (Zn) (calamine soil) or in nickel (Ni) (ultramafic or serpentine soil), due to its very high metal tolerance (Assunção et al., 2003b; Baker et al., 1994; Mohtadi et al., 2012; Richau and Schat, 2009). In addition, this species is able to accumulate elevated amounts of Cd, Zn, Pb and Ni in its aerial parts, which is known as metal hyperaccumulation (Brooks et al., 1997). This has raised an interest to study *N. caerulescens* and explore the biological mechanisms and genetic architectures underlying metal hyperaccumulation and tolerance (Assunção et al., 2003a; Milner and Kochian, 2008). An understanding of these traits can have useful implications for application, such as for phytoremediation, using plant to clean metal-contaminated soil especially for arable land (Ashraf et al., 2010; Zhao and McGrath, 2009); for phytomining, using plants to extract valuable metals from anthropogenically metal-polluted or naturally metal-rich lands (Brooks et al., 1998; R. L. Chaney and Baklanov, 2017) or for biofortification, improving the nutritional value of food to combat human micronutrient deficiency (Zhao and McGrath, 2009). *N. caerulescens* is an attractive model metal hyperaccumulator species as it is a diploid plant (2n=14) that is fully self-compatible and easily out-crossed, which makes breeding and genetic research relatively easy (Assunção et al., 2003a). Being in the Brassicaceae family, with several well investigated species, notably the general plant model *Arabidopsis thaliana* with ~88% sequence identify (Rigola et al., 2006) and with transcriptome sequences available for *N. caerulescens* (Halimaa et al., 2014; Lin, 2014), there is also ample opportunity for gene identification.

N. caerulescens has been investigated for physiological, biochemical and genetic aspects of metal accumulation and tolerance, resulting in the identification of several genes and quantitative trait loci (QTL) involved with accumulation and tolerance of Zn, Cd and Ni (Assunção et al., 2006; Assunção et al., 2001; Deniau et al., 2006; Halimaa et al., 2014; Mortel et al., 2008; Rigola et al., 2006; van de Mortel et al.,

2006). However, these candidate genes and QTLs are rarely explored due to scarcity of reverse genetic tool in *N. caerulescens*.

Genetic transformation is a widely used reverse genetic tool in several species. Three transformation methods have been reported for *N. caerulescens*. The first one based on *Agrobacterium tumefaciens* mediated flower dipping (Peer et al., 2003), one on *A. tumefaciens*-mediated transformation upon co-cultivation of explants in tissue culture (Guan et al., 2008); and one based on hairy root transformation using *Agrobacterium rhizogenes* (Lin et al., 2016). Each method experiences different challenges. Flower dipping, initially developed and optimised for *A. thaliana* (Clough and Bent, 1998; Feldmann and David Marks, 1987) requires flowering plants, which is a lengthy procedure in *N. caerulescens* since it requires 2-3 months of vernalisation to flower. Transformation by tissue culture is apparently more complicated than the publication suggests, or requiring considerable skills, as both of these methods have only been reported once. Transformation by *A. rhizogenes* is more commonly used, but has the disadvantage that only chimeric transgenic plants can be produced, with a transgenic root system, and a non-transgenic shoot. Transgenic plants are therefore limited to one genetic generation, and difficult to maintain for a long time. Of the three methods, floral dipping transformation method seems to be the most attractive to develop for further use in future *N. caerulescens* studies, as, other than requiring flowering plants, it does not appear to require considerable technical skills, and will result in stable transgenic plants, that can be propagated by self-pollination. Especially if the generation of flowering plants, bearing with many inflorescences suitable for dip-inoculation, could be optimized, this could become a valuable tool for reverse genetic studies in *N. caerulescens*.

N. caerulescens is a winter bi-annual species, with a life cycle generally around 8 - 9 months. It first involves a vegetative phase of 2-3 months, a vernalization phase of 2-3 months, growing at a low temperature, to induce flowering, and a pollination and seed-ripening phase of 2-3 months, to produce mature seeds that can be stored for longer time. To generate stable, homozygous T_2 generation plants, will therefore need 12-18 months, and a climate-controlled growth room that can be set at ~ 5 °C, for vernalization. Experiments have been performed in the past, to generate mutants that do not require vernalisation to flower, a trait also referred to as 'early flowering'. Using fast neutron mutagenesis, Lochlainn et al. (2011) generated an early flowering mutant in the Ganges accession, and Wang et al. (2020) generated several early flowering mutants using an EMS-mutagenesis approach to generate a TILLING population. SF has been recommended as a promising genotype for studying metal hyperaccumulation (Peer et al., 2003, 2006). Among the different early flowering mutants identified by Wang et al. (2020), three were identified to contain a mutation in the *FLOWERING LOCUS C* (*FLC*) gene, and one in the *SHORT VEGETATIVE PHASE* (*SVP*) gene.

The *flc-1* TILLING mutant contains a G to A point mutation in the third exon of *the FLC*, that results in a splice variant and rearrangement of the third exon in the coding sequence (Wang et al., 2020). The *flc-1* mutant shows pronounced early flowering, at 51 day after sowing, without vernalization (Wang et al., 2020) and high fecundity when grown under well-fertilized conditions. Recently, a collection of 86 accessions of *N. caerulescens* has been described, which provides a wealth of genetic and phenotypic variation for metal related and morphological traits (van der Zee et al., 2021).These can be used to generate an interesting genetic resource to improve fecundity and self-fertility of *flc-1* via conventional intercrossing.

This study presents the generation of fertile, early flowering *N. caerulescens* inter-accession recombinants, and an optimized, simple protocol for the generation of transgenic plants from such recombinants through floral dipping.

Methodology

Plant material and generation of early flowering F_2/F_3 plants

20 *N. caerulescens* accessions, known for their fecundity and self-fertility, were used to cross with the early flowering *flc-1* mutant, originally identified in an EMS-mutagenized population in the St. Felix de Pallières (SF) background. All accessions originate from different populations in Europe and are inbred for at least 5 generations (Table S1). Seed propagation of all genotypes has been performed in a frost-free (≥ 5 °C) greenhouse (51°59'46.4 "N 5°39'29.4 "E) at the Wageningen University campus, during October to July in the past years.

The early flowering *flc-1* mutant was crossed as father to each accession. Before crossing, the seeds of the accessions were surface-sterilised using Cl_2 gas and pre-germinated in modified half-strength Hoagland's solution that had a nutrient composition of 3 mM KNO_3, 2 mM $Ca(NO_3)_2.4H_2O$, 1 mM $NH_4H_2PO_4$, 0.5 mM $MgSO_4.7H_2O$, 1 µM KCl, 25 µM H_3BO_3, 2 uM $MnSO_4.4H_2O$, 2 µM $ZnSO_4.7H_2O$, 0.1 µM $CuSO_4.5H_2O$, 0.1 µM $(NH_4)6Mo_7O_{24}.4H_2O$ and 20 uM Fe-EDDHA (*N,N'*-ethylenediamine-di(O-hydroxyphenylacetic acid) and was buffered at pH = 5.5 with 2 mM MES (2-[*N*-morpholino]-ethanesulfonic acid) solution. After stratification at 4 °C for four days, the seeds were transferred to a climate-controlled growth room, set at 20 °C, 15hr/9hr light/dark, and 70% humidity. After radicle emergence, the seeds were sown and seedlings were grown in pots with a mix of fertilised peat and sand (Table S2) in a frost-free (> 5 °C) greenhouse from October. Vernalized plants started to flower in March/April the next year, by which the crosses could be conducted. The F_1 seeds of each cross were harvested in July and sown again in October, to flower after vernalization. Flowers were self-pollinated to generate F_2 seeds following the same procedure as before.

To identify the early flowering plants among the segregating F_2 progeny of each cross, seeds were pre-germinated as described above and then sown on 4 x 4 cm^2 Rockwool block (Grodan, The Netherlands), weekly supplied with Hyponex (Table S3) nutrient solution. These plants were grown in a climate-controlled growth room set at 21 °C, 15hr/9hr light/dark, and 70% humidity. The flowering F_2 (~25% if the crosses were successful), displaying high fecundity, high fertility and segregation of morphological characteristics (e.g. rosette diameter and plant height) were selected for production of F_3 seeds by self-pollination.

Transformation by floral dipping

F_3 seeds from early flowering F_2 plants of the crosses with 25% flowering plants were pre-germinated as described above. These plants were grown in pots with a mix of fertilised peat and sand in a climate-controlled growth cabinet (Weiss Technik Nederland B.V., The Netherland) at 20 °C, 15hr/9hr light/dark and 70% humidity. Three plants per F_3 progeny were used for this experiment.

Silwet L-77 sensitivity test was performed on *N. caerulescens* inflorescences. For this, three concentrations of Silwet L-77 (Lehle seeds, USA): 0.02 % v/v, 0.05% v/v and 0.1 % v/v, were pipetted on the tested inflorescence twice, with a 7-day interval, to mimic a double-dipping transformation. The concentration that did no cause necrotic tissues on leaves and inflorescences was chosen for floral dipping transformation.

A. tumefaciens strain GV3101 carrying a modified plasmid was plasmid of fluorescence-accumulating seed technology with *OLEOSIN1*-Red Fluorescent Protein (*OLE1-RFP*), shortly called pFAST-R02 (Shimada et al., 2010), carrying enhanced Green Fluorescent Protein (eGFP), as shown in Figure S1, was used for this experiment. An infiltration medium was prepared as described by Clough and Bent (1998) with 5% sucrose and Silwet L-77 (Lehle seeds, USA) at the concentration from sensitivity test. For floral dipping, the inoculation medium was added to a 50-mL conical tube and the inflorescences with long peduncles were submerged in the medium. For the inflorescences with short peduncles, which remained very closed to the soil surface, the inoculation medium was pipetted directly on those inflorescences. Then, the dipped plants were placed in closed plastic bags to maintain humidity and covered with a black plastic bag overnight. On the next day, the bag was removed from dipped plants. These plants were dipped with the same procedure for the second time a week later to increase the transformation efficiency (Davis et al., 2009). The plants were grown for a further 10 weeks until siliques were ripe and dry. The dipped ripe seeds were harvested from siliques and stored in plastic bags. *A. thaliana* introgression line IL-A4 was used as a positive control for transformation. IL-A4 carries a small Col-0 introgression in a CSL32 genetic background, but largely resembles Ler (Wijnen, 2019).

Screening for transformed seeds and stable transformation

To identify transformed seeds, T_1 seeds was screened for a RFP signal using a fluorescence stereo microscope Axio Zoom.V16 (Carl Zeiss, Germany) equipped with a dsRED filter. The images were captured in Zeiss Zen 2.1 software (Carl Zeiss, Germany). The transformation efficiency is calculated as;

$$\text{Transformation efficiency} = \frac{\text{number of fluorescent seeds}}{\text{total number of seeds}} \times 100$$

Transformed T_1 seeds were grown on 0.3% modified half-strength Hoagland's agar plate. After three weeks, germinated seedlings were observed for GFP signal under Axio Zoom.V16 (Carl Zeiss, Germany) using GFP filter.

Result

Early flowering traits in fertile *N. caerulescens* backgrounds

The SF *flc-1* mutant is not very self-fertile, probably due to low efficiency in self-pollination (Wang et al., 2020), which is largely influenced by pot size. This drawback makes SF *flc-1* not a very attractive genotype to use to develop flower-dip transformation in *N. caerulescens*. Fortunately there are several *N. caerulescens* accessions that are highly self-fertile, producing many inflorescences and high seed yield. Therefore, to develop fertile *N. caerulescens* maternal backgrounds with the early flowering trait, crosses were made between early flowering mutant *flc-1* (Wang et al., 2020) and 23 fertile *N. caerulescens* accessions. F_1 plants were grown and vernalized to generate F_2 seeds. Of these, only 5 F_2 lines, with accessions Clough Wood (CLW), Le Blémard (BLE), Cira (CIR), Saint Baudille (SBD) and Werschmatt (WER) as paternal parent, showed the expected Mendelian ratio of approximately 3:1 for non-flowering to flowering plants, with early flowering plants setting abundant seeds. The flowering time of these F_2 plants ranged from 44.5 to 58.5 days after sowing (DAS). The SBD maternal genotype conferred the earliest average flowering time at 44.5 DAS (Table 1). The F_3 progeny seeds of these five lines were obtained upon natural self-pollination of individual plants, providing a sufficient number of seeds for further propagation and transformation.

Table 1. *Flowering characteristics of fertile F_2 progenies upon crossing with the SF flc-1 mutant*.

Maternal background	Acronym	Ecotype[a]	% Flowering plants	Ratio Non-flowering: flowering plants	Average flowering time (DAS)	Seed setting
La Calamine	LC	CAL	53.34	7:8	58.5 ± 3.5	No
Lellingen	LE	NM	0	15:0	0	
Monte Prinzera	MP	ULT	0	15:0	0	
Krušné hory Mountains	KHM	NM	80	3:12	55 ± 4.06	No
Moravsko-slezské Beskydy Mountains	MSB	NM	26.67	11:4	48.75 ± 6.76	No
Durfort	DUR	CAL	0	15:0	0	
Clough Wood	CLW	CAL	26.67	11:4	49.75 ± 5.76	Yes
Puerto de Aralla	ARA	NM	0		0	
Le Blémard	BLE	CAL	26.67	11:4	54 ± 6.48	Yes
Cira	CIR	ULT	26.67	11:4	55.75 ± 4.15	Yes
Copenhagen	COP	NM	0		0	
Kuopio	KUO	NM	6.67	14:1	57± 0	No
Höör	HOO	NM	26.67	11:4	55 ± 3.27	No
Werschmatt	WER	NM	60	6:9	53.33 ± 1.25	No
Horná Roveň	HRO	CAL	66.67	5:10	58± 2.94	No
Valle Gargassino	VGG	ULT	6.67	14:1	52 ± 0	No
Auxelles	AUX	CAL	13.33	13:2	55 ± 3	No
Fresse sur Moselle	FSM	NM	60	6:9	54.75 ± 2.68	No
Col du Marchairuz	MAR	NM	20	12:3	52.5 ± 3.5	No
Tête de Ran	TDR	NM	0	15:0		
Rozenburg	ROZ	Calamine	0	15:0		
Saint Baudille	SBD	NM	26.67	11:4	44.5 ± 1.5	Yes
Špania Dolina	SDO	CAL	0	15:0		

[a]*NM: Non-metallicolous; CAL = Calamine; ULT = Ultramafic.*

For the transformation experiment, F_3 plants were grown in 7-L pots with a mixture of fertilised peat and sand, to give abundant inflorescences and flowers. Flowering time ranged between 48 to 64 DAS and these plants took approximately 20 weeks after sowing to mature ripe seeds. The CLW maternal background was the earliest flowering at 48.3 DAS and also yielded the highest number of harvested seeds (Table 2). The CIR maternal background had long fragile stems, perhaps due to the relatively low irradiance, and a high rate of seed abortion (Figure 1D).

Table 2. *Flowering time and seed yield of selected F_3 populations.*

Maternal background	Average flowering time (DAS)	Average seed yield
CLW	48.3	545
BLE	64.3	361.7
CIR	64	188.3
SBD	55.2	121.7
WER	nd	250

nd: no data.

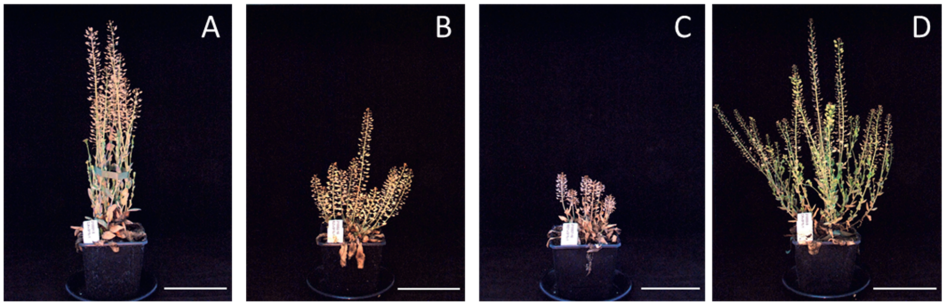


Figure 1. *N. caerulescens inter-crossed populations between flc-1 mutant and four maternal backgrounds: CLW (A), BLE (B), SBD (C) and CIR (D). Scale bar 10 cm.*

***N. caerulescens* inflorescence sensitivity to Silwet L-77**

For transformation of *A. thaliana* by flower dipping, the concentration of Silwet L-77 in the infiltration medium has impact on the transformation efficiency (Clough and Bent, 1998). A pilot experiment was therefore performed to evaluate the sensitivity of *N. caerulescens* inflorescences to Silwet L-77. The inflorescences did not show any sign of sensitivity (e.g. necrotic tissue) to any of these Silwet L-77 concentrations. However, on the leaves, necrosis was found at 0.05% v/v and 0.1 %v/v Silwet L-77 (Figure 2). Based on these results, we concluded that 0.02% v/v Silwet L-77 in the infiltration medium would be the best concentration to try for *N. caerulescens* transformation.

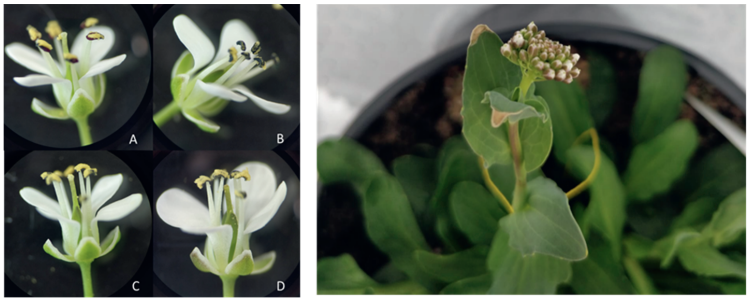


Figure 2. *N. caerulescens respond to Silwet L-77. Inflorescence did not show any necrotic tissue to Silwet L-77 (A: 0 v/v %, B: 0.05 v/v%, C: 0.02 v/v % and D: 0.1 v/v%.) below 0.1 % v/v (left). The leaves of N. caerulescens treated with 0.05 v/v % Silwet L-77 shows necrotic tissues on the edges. This image took seven days after the second Silwet L-77 pipetting (right).*

Successful floral dipping transformation in three maternal backgrounds

Floral dipping transformation of *N. caerulescens* was attempted using *A. tumefaciens* strain GV3101 carrying the pFAST-R02 eGFP. This modified plasmid contains two efficient selectable marker genes, suitable for non-destructive marker gene expression analysis: a) *pOLE1-RFP* (the original selectable marker of plasmid pFAST-R02) and b) *p35S-GFP*. The *pOLE1-RFP* allows screening for successful transformation in the seed stage. In addition, the *p35S-GFP* serves as additional marker for transformation, with GFP expression expected to be visible at most stages of plant development and growth.

A. thaliana was used as positive control for floral dip transformation. *A. thaliana* T_1 seeds exhibited RFP expression (Figure 3A) and the average transformation efficiency was established at 0.35%. For *N. caerulescens* flower dip transformation, the infiltration medium was applied using two approaches: a) by submerging inflorescences with a long peduncle into the infiltration medium; or b) by pipetting the infiltration medium directly on the inflorescence in case plants had a short peduncle (such as for BLE). The T_1 progeny of 3 maternal backgrounds: CLW, BLE and SBD provided seeds with red fluorescence, as evidence for stable transformation (Figure 3B). Progeny of plants with WER and CIR maternal backgrounds did not give any red fluorescent seeds. The average transformation efficiencies are 0.40% (CLW), 0.46% (BLE) and 0.29% (SBD) (Figure 3C). The GFP signal was also observed in T_1 seeds but only in seeds with a BLE maternal background. This indicated that GFP is not very efficient for selection in seeds for floral dip transformation, most likely as the 35S promoter was not very effective to drive expression of GFP in *N. caerulescens* seeds.

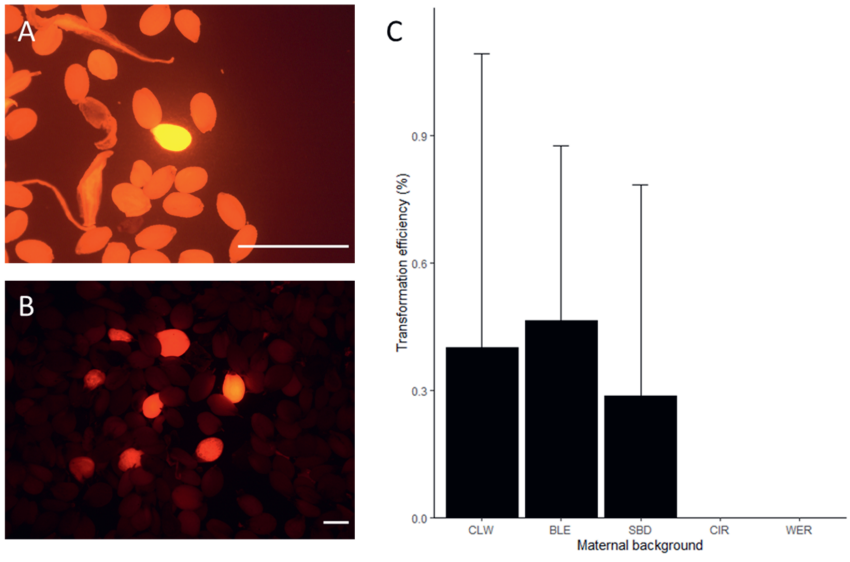


Figure 3. *Successful transformed A. thaliana (A) and N. caerulescens of CLW maternal background (B) seeds exhibiting RFP-positive. Scale bar 1 mm. Transformation efficiencies for different maternal backgrounds (C). The bars represent standard deviation.*

To further validate the stable expression of inserted genes integrated in the host genome upon transformation, the expression of p35S-GFP was determined in the germinated seedling. Indeed, T_1 seedlings originating from seeds selected for RFP expression, also showed green fluorescence. The intensity of green fluorescence was strongest in the meristematic regions (Figure 4). Overall, these results show that introduction of the *flc-1* early flowering mutation in different highly self-fertile genetic backgrounds, yielded three early flowering lineages suitable for flower-dip transformation of *N. caerulescens* with efficiencies up to 0.46%.

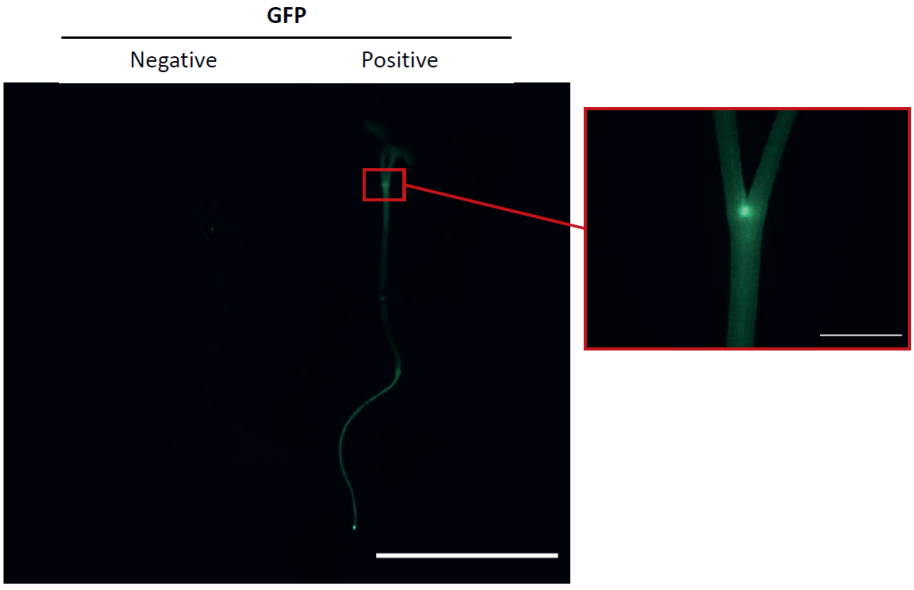


Figure 4. *An example of GFP-negative and GFP-positive in seedling of T1 CLW maternal background (scale bar 10 cm). The meristematic region exhibited high signal of GFP-positive (scale bar 1 cm).*

Discussion

N. caerulescens has been proposed as a model species for studying metal hyperaccumulation and hypertolerance for nearly two decades (Assunção et al., 2003a). However, so far, reverse genetic studies on this species were very difficult to perform, in the absence of a simple transformation method giving stable transformants that can be genetically propagated through seeds. While stable genetic transformation of *N. caerulescens* has been reported twice (Guan et al., 2008; Peer et al., 2003), somehow these methods were not considered to be attractive or successful, as no follow-up use has been reported until now. It was not clear if this was due to the difficulties in accomplishing the right regenerations and selection conditions in tissue culture, or the very lengthy procedure needed for flower dip transformation, requiring a 2-3 months vernalization period for plants to induce flowering, or if there were further complications, perhaps caused by differences in transformation ability between different *N. caerulescens* genotypes. The recent isolation of well-characterized early flowering mutants, such as the *flc* and *svp* mutants, identified upon EMS-mutagenesis in the SF background, offered the opportunity to significantly reduce the floral-dipping transformation procedure, abolishing the need for vernalization for plants to flower. Unfortunately the SF genetic background is not optimal for this approach. While plants in the field show no sign of reduced fecundity, in a controlled growth environment, such as a climate

room or greenhouse, plants display low fecundity (Wang et al., 2020, 2022). This is not due to genetic defects, but more likely associated with the flower architecture shared by many accessions in the Cevennes region in France where this accession originates from (Mousset et al., 2016). Due to this flower architecture, anthers and stigma remain at a distance upon opening of flower buds, favouring cross-pollination by insects. Manual pollination could be used to overcome this, but it would make the transformation procedure undesirably labour intensive. We therefore set out to introduce the early flowering *flc-1* mutation in different, highly self-fertile, *N. caerulescens* accessions. Considering the poor self-fertility of SF, we decided to use the *flc-1* SF genetic background as father in the crosses, which in retrospect was not the best choice. Out of 23 crosses made, only 5 showed segregation of the early flowering trait, probably due to problems in emasculating the flowers of the mother plants. Considering the long time to induce flowering in wild-type accessions, we did not attempt to make the reciprocal crosses. There is no obvious restriction to introgress *flc-1* into other genetic backgrounds, but the advice is to use it as a mother, rather than a father in the crosses, which will reveal already in the F1 if the crosses have been successful, and not just in the F2, as we experienced.

Our study demonstrates that introducing the early flowering trait indeed facilitates successful and efficient genetic transformation of *N. caerulescens*. In the previous study reporting successful transformation of *N. caerulescens* by floral dipping, herbicide tolerance (conferred by the *bar* gene) was used as selectable marker along with GFP as visible expression marker for stable transformation. Overall transformation efficiencies reached approximately 0.6% (Peer et al., 2006), which is slightly higher than the transformation efficiencies we found, ranging from 0.29 – 0.46%. Among the accessions tried for floral dip transformation by (Peer et al., 2003) were also SF and BLE genotypes, likely to originate from the same, large, populations present at the sites. Interestingly, transformation by spray inoculation resulted in abortion of all flowers for BLE (Peer et al., 2003), something we did not observe at all for the BLE *flc-1* F_3 plants, with a transformation efficiency of 0.46%. Rather than spray-inoculation, we applied the inoculum by pipetting, which may explain for the differences in success. Peer et al. (2003) also used 0.02% Silwet, as we did. Of course there may also be an effect of genetic background. We used introgressions of *flc-1* in the BLE background, and colud, by chance, have selected for plants carrying alleles favouring floral dipping. We certainly selected for F_2 plants of each cross that were early flowering (a SF trait) and providing high fecundity (a non-SF trait). Although we did not further determine plant fertility, crossing SF with high-fecundity *N. caerulescens* backgrounds certainly yielded ample high-fecundity *flc-1* progeny, suggesting that the poor fertility observed for SF is not a dominant trait.

The vernalization requirement to induce flowering and the difficulties in stable transformations have been serious limitations in developing *N. caerulescens* as a convenient model for metal hyperaccumulation and tolerance (Assunção et al., 2003a; Lochlainn et al., 2011; Wang et al., 2020). Certainly compared to *A. thaliana*, which can be reproduced up to four generations in the 24-36 weeks it may take for wild-type *N. caerulescens* to complete one seed-to-seed generation. The F_2 and F_3 progeny of the inter-accession crosses we made, took ~ 20 weeks to complete one seed-to-seed generation. This is longer than reported for the original *flc-1* mutant (~ 16 weeks to complete its life cycle), except for the CLW maternal background, but considerably faster than their wild-type backgrounds. *FLC* is one of the key proteins regulating induction to flowering in Brassicaceae species (Michaels and Amasino, 1999). The increase in flowering time of *flc-1* progeny upon introgression of the mutant gene in two other genetic backgrounds, suggests that there is additional variation for flowering time that affects the *flc-1* phenotype, although the identity of the genes underlying this variation is not yet known. The generated *flc-1* F2 and F3 plants can complete their life cycle without any further special requirements such as

special temperatures or light regimes. This provides a versatile and cost-effective system facilitating transgenic studies in *N. caerulescens*. Establishing stable transformed F_2 lines can be achieved within one year.

Genetic transformation in plants may be influenced by genetic background, with genotype specificity to be known to play a crucial role to determine the success of a transformation experiment (Ghedira et al., 2013). Our results identified the CLW x SF *flc-1* progeny as the most promising resource for *N.* transformation, with high fecundity, fast early flowering and long peduncle. CLW originates from calamine soil, and exhibits Zn and Ni accumulation characteristics comparable to those of SF (van de Zee et al., 2020). The Cd accumulation properties of CLW are not known yet, but SF is a good Cd accumulator (Koshevnikova et al., 2020), which could provide good Cd accumulation in the different F2/F3 progenies. SBD, the other background we could transform by floral dipping, originates from non-metallicolous soil. It is known to have average Zn and Cd tolerance (among other *N. caerulescens* accessions), and high Ni tolerance, as well as average Zn and Ni accumulation and high Cd accumulation, exceeding that of SF or BLE (Kozhevnikova et al., 2020). BLE is known to show high Zn and Cd tolerance, with average Ni tolerance, and relatively low Zn and Cd accumulation, with average Ni accumulation. Unfortunately, we were unsuccessful to transform the CIR and WER backgrounds, which is particularly unfortunate as both of these are collected at or close to ultramafic outcrops, respectively in Galicia, Spain, and the Vosges, France. The CIR background turned out to be hard to transform as the peduncles of the plants were more rigid than those of others, and prone to break upon covering upon floral dipping, leading to progressive loss of inflorescences, and transformed seeds, upon handling. Furthermore, the CIR *flc-1* F_3 plants yielded a low average number of seeds, and may be more affected by the Silwet than other plants. However, further refinements in the method may prove to be successful in establishing also this background as suitable for floral-dip transformation. For WER, we are not sure why the transformation was unsuccessful. Seed yield and peduncle length or rigidity is not different from the successfully transformed plants. It may therefore be that somehow WER is less susceptible for *A. tumefaciens* infection than the others, which is also known for *A. thaliana* accessions (Clough and Bent, 1998; Ghedira et al., 2013). Since we did not test a large number of plants, increasing this may well show that also CIR and WER *flc-1* plants can be transformed.

From the work on *A. thaliana*, there are several factors known to contribute to the success of *Agrobacterium*-meditated floral dip transformation, which can be further optimised. We used the common *A. tumefaciens* strain GV3101 for transformation, which was more efficient than other strains, such as LBA4404, C58C1, LMG201 or LMG62, in *A. thaliana* Col-0 (Ghedira et al., 2013). We did not investigate other strains, but some of those may be more effective in *N. caerulescens*.

Conclusion

This study reported a substantial improvement of one of the few reports on successful transformation of the metal hyperaccumulator *N. caerulescens*, by abolishing the vernalization requirement to flower upon introgression of the early-flowering-conferring *flc-1* mutation in the genetic background of three accessions, CLW, BLE and SBD. This transformation system will be a useful resource for future reverse genetic study on metal accumulation and tolerance.

Supplementary information

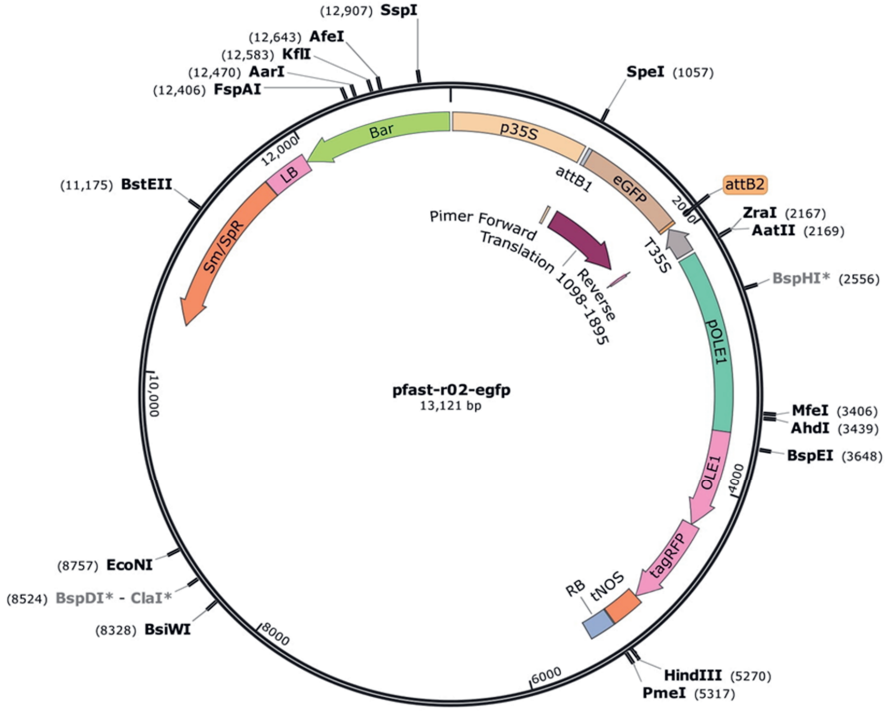


Figure s1. *Modified pFAST-R02 used in this study carrying two selectable markers: OLE1-RFP and 35s-GFP.*

Table S1. *Noccaea caerulescens maternal background collection site information.*

Accession name	Acronym	Origin	latitude	longitude
La Calamine	LC	Belgium	50°42'38.78"N	06° 0'37.42"E
Lellingen	LE	Luxemburg	49°59'1.83"N	05°59'39.00"E
Monte Prinzera	MP	Italy	44°38'32.56"N	10° 5'4.87"E
Krušné hory Mountains	KHM	Czech	50°42' 16.54'' N	13°38' 35.09'' E
Moravsko-slezské Beskydy Mountains	MSB	Czech	49°26.155 N	18°13.78 E
Durfort	DUR	France	44° 00' 01'' N	03° 57' 04'' E
Plombieres	PB	Belgium	50°44'4.31"N	05°57'49.52"E
Clough Wood, Darley Dale	CLW	UK	53° 9'9.77"N	01°36'56.38"W
Puerto de Aralla	ARA	Spain	42°53'45.79"N	05°48'43.51"W
Le Blémard	BLE	France	44° 28' 17'' N	03° 43' 43'' E
Cira	CIR	Spain	42°46'0.001"N	08°20'59.999"W
Copenhagen	COP	Denmark	55°40'50.68"N	12°32'32.98"E
Kuopio	KUO	Finland	62°53'48.94"N	27°38'5.44"E
Saint Baudille	SBD	France	43°44'40.7328"N	03°29'09.9600"E
Werschmatt	WER	France	47°55'7.46" N	06°46'50.458" E
Valle Gargassino	VGG	Italy	44°55'39''N	08° 63' 22'' E
Auxelles	AUX	France	47°44'21.52"N	06°46'35.84"E
Fresse sur Moselle	FSM	France	47°52'18.50"N	06°47'29.50"E
Tête de Ran	TDR	Switzerland	47°03'20.29"N	06°51'23.19"E
Col du Marchairuz	MAR	Switzerland	46°32'43.00"N	06°15'4.00"E
Tête de Ran	TDR	Switzerland	47° 3'15.00"N	06°51'15.00"E
Rozenburg	ROZ1	Netherland	51°54'08.8866"N	04°16'03.8692"E
Spania Dolina	SDO	Slovakia	48°48'28.001"N	19° 08'6"E

Table s2. *Growing material recipe for Noccaea caerulescens.*

pH = 5.7	EC = 0.80	EN-factor = 1.23	HWW = 154
Component	Volume		
Hortiklei	40.610 kg		
MD Zweeds veenmosveen	0.200 EN-m³		
Baltisch veen middel	0.300 EN-m³		
Tuinturf normaal	0.300 EN-m³		
StructuurBark	0.200 EN-m³		
Dolokal Extra potgrond (bulk)	3.300 kg		
PG-mix 15-10-20	0.810 kg		
Element	mmol/L	Solution	volume (L)
		Zwakal	25.2
NH4	1.1	BFK	44.2
K	5.11	Baskal	14.4
Ca	3	Ammitra	13.8
Mg	0.87	Magnitra	10.4
No3	7.79	Calsal	64
SO4	1	This is per 100,000 litre	
P	1.5		
EC = 1.45			

Table s3. *Nutrient solution Hyponex composition*

EC = 1.4		pH = 5.8	pH is adjusted by: K_20 or H_2SO_4				
Macro-elements in mmol/L		*100 liter water + salts in premade solution (YaraTera):*					*100 liter water + salts (in g):*
NH_4^+	1.7	21.3 ml Amnitra (ammonium nitrate)					47.2 gram $Ca(NO_3)_2.H2O$
K^+	4.1	36.8 ml BasKal (Potassium carbonate/hydroxide)					3.2 gram NH_4NO_3
Ca_2^+	2	42.1 ml Calsal (Calcium nitrate)					72.1 gram K_2SO_4
Mg_2^+	1.2	57.58 ml Magnesul (Magnesium sulphate)					30.6 gram $MgSO_4.7H_2O$
NO_3^-	4.3						15 gram $NH_4H_2PO_4$
SO_4^{2-}	3.3	52.24 ml Sulfakal (Potassium sulphate/sulphoric acid)					
$H2PO_4^-$	1.3	15.1 ml Phosphoric acid 59%					
Micro-elements in µmol/L		*Compound*	*g/Mol*	*µmol/L final*	*g/L final*	*g/10 L Stock solution*	
Mn	3.4	$MnCL_2.4H2O$	197.91	3.4	0.000673	6.72894	
Zn	4.7	$ZnSO_4.7H2O$	287.56	4.7	0.001352	13.51532	
B	14	H_3BO_3	61.83	14	0.000866	8.6562	
Cu	6.9	$CuSO_4.5H2O$	249.68	6.9	0.001723	17.22792	
Mo	0.5	$Na_2MoO_4.2H2O$	241.95	0.5	0.000121	1.20975	
Fe (Liquid)	21	Fe chelaat 50/50	1851.85	21	0.038889		
50% Fe-DTPA 3% 50% Fe- EDDHSA 3%							

Acknowledgement

The authors are thankful to Dr. Francisca Reyes Marquez (Wageningen University, The Netherlands) for providing pFAST-R02 eGFP plasmid; Guillaume Echevarria, Thibault Sterckeman and Catherine Sirguey (INRA, France) for providing seeds of *Noccaea caerulescens* accessions Auxelles, Fresse sur Moselle and Tête de Ran; Alan Baker is acknowledged for providing seeds of Clough Wood; Roel van Bezouw is acknowledged for providing the seeds of Rozenburg; Enrica Roccotiello is acknowledged for providing seeds of Valle Gargassino; Terezie Mandáková is acknowledged for providing seeds of Höör and Špania Dolina.

CHAPTER

6

General Discussion

Plants employ a strictly regulated metal homeostasis network to maintain the level of essential metals to ensure their normal growth and development. The major processes of plant metal homeostasis include metal transport, chelation and sequestration. These processes are used to control the uptake, accumulation, trafficking and detoxification of metals in plants (Dalcorso et al., 2013; Manara et al., 2020). When exposed to elevated levels of metals, most plants employ an excluder strategy to restrict metal translocation to above-ground tissues, particularly photosynthetically active tissues, to alleviate the toxicity of metals (Dalcorso et al., 2013; Manara et al., 2020). A small number of plants, called metal hyperaccumulators, have adapted to elevated levels of soil metals. These plants display strikingly high metal tolerance (so-called metal hypertolerance) and substantial metal accumulation in above-ground tissues (so-called metal hyperaccumulation) (Baker, 1987; Baker et al., 2010; Baker and Brooks, 1989; Pollard et al., 2014; van der Ent et al., 2013).

Studies in *Saccharomyces cerevisiae* (baker's yeast) have provided information for the basal cellular components of eukaryotic metal homeostasis and continue to serve as an alternative reverse genetics tool for *in planta* functional studies like yeast complementation assays (De Freitas et al., 2003; Gravot et al., 2004). Several genes involved in plant metal homeostasis have been comprehensively described in the model species *Arabidopsis thaliana*, including their tissue and organ-specific expression levels (Hussain et al., 2004; Lamesch et al., 2010; Roosens et al., 2008; Tehseen et al., 2010). Unlike non-hyperaccumulating *S. cerevisiae* and *A. thaliana*, metal hyperaccumulators have evolved a wide range of altered metal homeostasis processes, including changes in expression of genes involved in metal homeostasis such as metal chelators/transporters (Foroughi et al., 2014; Halimaa et al., 2014; van de Mortel et al., 2006), and transcription factors (Li et al., 2022; P. Zhang et al., 2019); higher copy numbers of metal transporter genes (Craciun et al., 2012; Hanikenne et al., 2008; Lochlainn et al., 2011); differential splicing of metal transporter transcripts (Dong et al., 2018; Persans et al., 2001), and modification of cell wall (Hasan et al., 2017; Lin and Aarts, 2012). These processes facilitate adaptation to harmful anthropogenically-caused or naturally occurring soil conditions by increasing metal accumulation capacities and tolerances in metal hyperaccumulators.

Among metal hyperaccumulators, two species have received considerable attention: *Noccaea caerulescens* and *Arabidopsis halleri,* which have been investigated intensively as model plants for metal hyperaccumulation and hypertolerance (Assunção et al., 2003a; Erlangung and Doktor, 2017; Milner and Kochian, 2008a; Pauwels et al., 2012; Schat and Aarts, 2003). The contribution of their distinct morphological and physiological characteristics to metal accumulation capacity and tolerance have been well-documented, at least for cadmium (Cd) and zinc (Zn). The rapid development of next-generation sequencing (NGS) has enabled the identification of genetic differences across the genome among large numbers of populations of interest (da Fonseca et al., 2016; Russell et al., 2017). For instance, the NGS system Novoseq X series can sequence 1 Giga base for less than 2 US dollars (Novoseq X system, https://www.illumina.com/systems/sequencing-platforms/novaseq-x-plus.html). Since the first launch of a commercial NGS platform in 2005, genome-wide expression profile approaches have been adopted in a handful of studies of *N. caerulescens*, such as expression arrays (Mortel et al., 2008; Plessl et al., 2005) and RNA-seq (Halimaa et al., 2014), which mainly emphasized a few well-characterised accessions such as accessions La Calamine, Lellingen, Ganges and Monte Prinzera. Whole-genome re-sequencing DNA information has been developed for a small number of accessions as well (Wang, 2016).

In this thesis I generated two new genetic tools for the metal hyperaccumulator plant model species *N. caerulescens*, which include a Lellingen (LE) x Ganges (GA) recombinant inbred line (RIL) mapping

population (**Chapter 2**) and a diversity panel of a set of diverse *N. caerulescens* accessions from a large geographical area of Europe (**Chapter 4**), aiming to make progress in the comprehension of molecular mechanisms underlying metal hyperaccumulation/hypertolerance in *N. caerulescens.* By exploitation of a cost-effective NGS approach, the genome-wide variations are characterized for both *N. caerulescens* populations. In **Chapter 3**, I explored the genetic basis of Cd accumulation in the RIL population derived from the RIL parental accessions GA and LE displaying variation in this trait. This suggests that GA and LE may employ different mechanisms for Cd accumulation. This was confirmed by the identification of trait-enhancing alleles derived from both GA and LE at different Quantitative Trait Loci (QTL) for Cd accumulation. In addition to exhibiting different Cd accumulation capacities, GA and LE occurred on different soil types and geographically distant from each other, suggesting that GA and LE are genetically distant. In **Chapter 4**, I investigated the diversity panel to gain new insights into the population diversity of *N. caerulescens* on species level, using genome-wide genotypic information. The clustering pattern observed in the diversity panel confirms that GA and LE are phylogenetically distant, based on absolute allele frequency differences (AFD) (Berner, 2019). The accession LE is a non-metallicolous accessions from Luxemburg, assigned to the largest genetic subunit, subgroup 1 (SG1). While, accession GA is a Cd/Zn hyperaccumulating and hypertolerant calamine accessions from the Cevennes region of France, belonging to SG5. The dendrogram of the diversity panel clearly illustrates substantial genetic divergence between SG1 and SG5 (**Chapter 4**). The diversity panel is characterized for Zn/Ni accumulation capacities, revealing accessions with interesting traits such as extreme low Zn/Ni accumulating calamine accession "Largentiere" and elevated Zn/Ni accumulating non-metallicolous accession "Vallée de Joux" (**Chapter 4**). A thorough investigation of the sequence variation of these accessions generated in this thesis, could explain how ecological and evolutionary pressure shape this specific variation, leading to such distinct traits. Moreover, it also may result in identification of novel genes involved in Zn/Ni accumulation in this species, which could expand to investigation of other metal-related traits and life history traits in *N. caerulescens*, such as Cd accumulation and flowering time.

The LE x GA RIL population developed in this thesis displays the "immortal" feature (**Chapter 2**). An immortal mapping population enables multiple analyses, in particular for quantitative trait loci (QTL) mapping for traits of interest, on the same population (Keurentjes et al., 2011). This was not possible for previously reported *N. caerulescens* F_2 and F_3 mapping populations (Assunção et al., 2003a, 2006; Deniau et al., 2006; Frérot et al., 2005; Zha et al., 2004). Nevertheless, immortal mapping populations are constrained by the requirement of multiple rounds of self-propagation, implying several years for the obligately vernalized *N. caerulescens*. The preferential metal accumulation corresponding to parental genotype backgrounds limits the potential for causal gene discovery in this population. For instance, it is less likely to detect QTLs associated with Ni accumulation in this mapping population derived from accessions LE and GA displaying small variation in Ni accumulation. The limited number of alleles results in many monomorphic loci in immortal mapping population, rendering these loci unsuitable for QTL detection in that population. Missing or undetected QTLs could potentially be identified in mapping populations derived from multiple parents, such as Nested Association Mapping populations (NAM) and Multi-parent Advanced Generation Inter-Cross populations (MAGIC), as well as diversity panels. In practice, a diversity panel is more feasible to generate and analyse than developing multi-parental mapping populations, due to the lengthy life cycle of *N. caerulescens*.

In comparison to an immortal mapping population, a diversity panel displays a high degree of allelic variation and a broad range of metal-specific accumulation phenotypes. The clustering analysis provided in **Chapter 4** provided a preliminary species history of *N. caerulescens* from a larger geographical area

compared to (Gonneau et al., 2017). The population structure within subunits defined by (Gonneau et al., 2017) was observed when including additional accessions from eastern Alps and Spain (**Chapter 4**). Further sample collection from eastern Europe from Switzerland to the Czech Republic would offer a more complete overview of *N. caerulescens* evolutionary history. Besides clustering analysis, the first attempt for GWAS in *N. caerulescens* was made in this largest collection of this species to-date. The results indicate that the diversity panel can be used for association studies for identification of genes involved with traits of interest. It is important to note that well-known genes associated with flowering time and Zn/Ni accumulation were not detected. This could be because these traits and their causal SNPs were completely confounded by the population structure (Korte and Farlow, 2013), indicating the limitation of this diversity panel. Potential improvements could be achieved by re-constructing the panel composition to minimize population structure effect upon these traits, as discussed in **Chapter 4**. Another factor to consider is the possibility of insufficient SNPs used in this thesis. The SNP filtering parameters I employed might be too stringent, leading to the removal of over 90% of raw SNPs, some of which could be directly associated with traits or linked to significant singletons identified in **Chapter 4**. Applying less stringent filtering parameters may improve the discovery of significant loci, as observed in some studies (Hou et al., 2017). However, this approach comes with the risk of introducing misleading signals and false positives into GWAS (O'Leary et al., 2018). This was evident in a pilot experiment where less stringent parameters were applied in SNP filtering, resulting in a higher number of filtered SNPs compared to the SNP dataset used in **Chapter 4**. The pilot results revealed a random clustering of the diversity panel and numerous synthetic genotype-phenotype associations, visible as parallel horizontal lines of dots in Manhattan plots (data not shown), indicating that higher stringent parameters are required. To enhance the statistical power of GWAS for this panel in future works, an optimization of SNP filtering parameters is crucial. An unsupervised filtering algorithm, as described in (Pongpanich et al., 2010) could be considered for this purpose. In addition to adjusting filtering parameters, low-quality SNPs from a subset of the panel—specifically those from accessions with sequencing coverage below 15X (half of the planned coverage)—were excluded using the filtering parameter "no missing genotypes allowed." This introduced an issue of unequal sample sizes, complicating the GWAS analysis. Therefore, the use of SNP imputation for this panel may be considered, as it has the potential to improve the statistical power of GWAS (Jiao et al., 2011; Marchini and Howie, 2010).

In higher plants, flowering too early or too late can drastically decrease reproductive success and survival chances, especially when growing in unfavorable environmental conditions (Inouye, 2008; Munguía-Rosas et al., 2011; Sherrard and Maherali, 2006). The timing of inflorescence and thereafter flower development, often defined as 'flowering time', has been studied in various plant species, which includes plant model species (i.e. *A. thaliana* (Koornneef et al., 1998; Stinchcombe et al., 2004), *Capsella bursa-pastoris* (Kryvokhyzha et al., 2016)), or related wild species (i.e. *Arabidopsis arenosa* (Baduel et al., 2018) and *Arabidopsis lyrata* (Kuittinen et al., 2008)) but also crop plants (Blümel et al., 2015; Kamran et al., 2014; Molla, 2022; Xu et al., 2021). A complex gene regulatory network controlling flowering time has been described for *A. thaliana*, which encompass more than 300 genes responding to a wide range of internal and external cues: plant age, plant hormone, photoperiod, stress, ambient temperature, and vernalization (Bouché et al., 2016).

The work described in this thesis used flowering time as a control trait to validate the robustness of the mapping population and the diversity panel for genetic study. The results of the conventional QTL mapping and GWAS revealed several candidate loci controlling flowering in *N. caerulescens*, in particular for *EARLY SHORT DAY4* (*ESD4*) (**Chapter 3**) and flowering-associated intergenic long non-coding RNA

(*FLAIL*) (**Chapter 4**), which serve as a proof of principle for the mapping panels I describe. In comparison to the studies in *A. thaliana*, identification of the key flowering genes, *FRIGIDA* (*FRI*), *FLOWERING LOCUS C* (*FLC*), was expected for the *Brassicaceae* winter type annual species *N. caerulescens*, nevertheless this was not shown in this thesis. In *A. thaliana*, *FRI* is a positive regulator of *FLC*, which encodes a MADS-box transcription factor that suppresses flowering development. Similar roles of FLC have been reported for *A. arenosa* (Baduel et al., 2018), *A. lyrata* (Kuittinen et al., 2008), *Boechera stricta* (Lee et al., 2018), *Capsella rubella* (Guo et al., 2012), through examination of flowering time variation in a set of natural accessions/populations. Unlike in *A. thaliana*, early flowering accessions have not been observed in *N. caerulescens* so far, indicating that *FRI* and *FLC* are unlikely to contribute to variation in flowering time in natural populations *of N. caerulescens*, including the RIL parental accessions GA and LE. Additionally, *FLC* has been recently discovered from screening of the loss-of-function mutant library of *N. caerulescens*, and included in the proposed flowering regulation model in this species (Wang et al., 2020). In *Thellungiella halophila*, characterization of isolated *ThFLC* (Fang et al., 2006) and *ThFRI* (Fang et al., 2008) provided evident that both genes have the same function as described in *A. thaliana*.

Concerning the identified flowering candidate genes, the loss-of-function *A. thaliana esd4* and *flail* mutants display early flowering indicating that *ESD4* and *FLAIL* are flowering repressors in *A. thaliana* Jin et al., 2023; Reeves et al., 2002). These genes employ different regulatory pathways to alter flowering time. *FLAIL* promotes alternative splicing of the flowering gene *LACCASE 8* (*LAC8*), which reduces *LAC8* expression and inhibits flowering (Reeves et al., 2002). While *ESD4* regulates the flowering process by influencing post-translational modification by SUMOylation (Hermkes et al., 2011). The same mechanism may occur in *N. caerulescens* since it is closely related to A. thaliana and shares high similarity on coding sequence (Rigola et al., 2006), suggesting that *ESD4* and *FLAIL* are likely to be flowering-controlled genes in *N. caerulescens*. As discussed above, further inspection of variation in sequence of these candidate genes in fast/slow flowering accessions (**Chapter 4**) will provide better understanding on how the selection process influences flowering time in this species.

Cd is a non-essential metal that poses significant harm to non-hyperaccumulating plants, with even low levels capable of inducing severe hypersensitive responses and programmed cell death (Dalcorso et al., 2013). In contrast, hyperaccumulator *N. caerulescens* employs specific mechanisms, both direct (i.e., Cd vacuolar sequestration and Cd chelation) and indirect (i.e., reduction of Cd-induced Reactive Oxygen Species, ROS), to restrict Cd toxicity. The results presented in this thesis reveal 27 QTLs associated with Cd response traits, with their respective *A. thaliana* orthologs identified within these QTLs (**Chapter 3**. Of particular significance was the major QTL for Cd root concentration, qRT-Cd-1, which exhibits a high LOD score of 16.3 and a compact QTL interval of 2.2 cM. This QTL colocalizes with the QTL for the Cd translocation factor, qTF-Cd. Within this colocalization, the ortholog of HEAVY METAL ATPASE 3 (HMA3) emerges as a the most likely candidate gene governing Cd root accumulation and Cd root-to-shoot translocation in *N. caerulescens*. Given the well-established role of HMA3 as a metal transporter responsible for vacuolar sequestration of Zn and Cd ions in *N. caerulescens*, this finding strongly underscores the versatility of the immortal population as an effective mapping resource for *N. caerulescens*. In addition to HMA3, several metal-related orthologs were identified in other QTLs. Notably, *NATURAL RESISTANCE-ASSOCIATED MACROPHAGE PROTEIN 5* (*NRAMP5*) and *CATION/PROTON EXCHANGER 3* (*CAX3*) were localized in the QTL associated with EC100, while *HEAVY METAL-ASSOCIATED ISOPRENYLATED PLANT PROTEIN 20* (*HIPP20*) was identified in the minor QTL of Cd root concentration, qRT-Cd-2 (**Chapter 3**). The identified orthologs serve as valuable resources for

uncovering genes associated with Cd response and enhancing our understanding of the genetic basis of Cd accumulation in plants. This knowledge will be crucial for the development of crop plants with minimal Cd concentrations in their edible parts, such as grain, leaves, and fruits, which would prevent Cd accumulation and toxicity in humans through the food chain.

Functional analysis has been a challenge in *N. caerulescens* due to the lack of a simplified and reproducible transformation protocol and its obligate vernalisation. Early flowering *N. caerulescens* mutants have been developed in the GA accession (Lochlainn et al., 2011) and the Saint Felix de Pallières (SF) accession (Wang et al., 2020). Nevertheless, both mutants exhibit low selfing rates and subsequently low fecundity in a controlled growth environment without natural pollinators, which may be in part due to their stigma that elongate above the stamen during flowering. This specific floral morphology characteristic is genetically inherited, and has been observed in many accessions from the Cevennes region of France, where the mutant backgrounds originated from (Mousset et al., 2016). **Chapter 5** describes a strategy to mitigate the low fecundity of early flowering mutant by introgression the SF early flowering mutant allele into accessions displaying high self-fertility and fecundity in controlled growth conditions. By selecting high-fecundity plants in each selfing generation, F_3 plants exhibiting both early flowering (the mutant trait in SF background) and high self-pollinated seed production (a trait not originating from SF) were obtained (**Chapter 5**). Combined with the modified floral dipping transformation method initially developed for *A. thaliana* (Clough and Bent, 1998; Feldmann and David Marks, 1987), a reproducible reverse genetic approach was developed for *N.* caerulescens. This protocol allows the establishment of additional advanced genetic tools, specifically genome editing approaches (i.e. CRISPR-Cas9) (Basharat et al., 2018; Osakabe et al., 2016). The reproducible transformation protocol could be effectively utilized to generate loss-of-function mutants for candidate genes identified in this thesis, such as *ESD4* (**Chapter 2**), *HMA3* (**Chapter 3**) and *FLAIL* (**Chapter 4**). Additionally, a comparative transcriptome analysis of leaf and floral tissues during the floral transition will allow a comprehensive flowering time regulation network of this species to be established (Ai et al., 2012; Gretsova et al., 2023), which is more affordable with current NGS technologies. The reported transformation protocol could also be adopted to characterize metal-homeostasis regulating genes, such as the *BASIC LEUCINE-ZIPPER 19* (*bZIP19*) gene, which already has been target for mutant analysis, but unsuccessfully (Wang et al., 2022).

The transcription factor bZIP19, in conjunction with bZIP23, has been identified as a core regulator of the response to zinc deficiency and the maintenance of zinc homeostasis in *A. thaliana* (Assuncao et al., 2010; Lilay et al., 2021) and rice (Lilay et al., 2020). Both bZIP19 and bZIP23 contain a Cysteine (Cys)/Histidine (His)-rich motif at their N-terminus, which is important to induce the expression of specific zinc homeostasis genes, such as *ZIPs* and *NASs*, in response to zinc deficiency (Assuncao et al., 2010). The role of this motif as a zinc sensor has been recently reported (Lilay et al., 2021). In *N. caerulescens*, only bZIP19 is expressed (Lin et al., 2014). Wang et al. (2022) identified four different *bzip19* mutants, but none of these were the desired loss-of-function mutants that could help to resolve the question on whether bZIP19 would be essential for the Zn/Cd hyperaccumulation or tolerance traits in *N. caerulescens*. By adopting the protocol reported in this thesis, different reverse genetic approaches, such as RNA interference (H. Han, 2018), T-DNA insertional mutagenesis (O'Malley et al., 2015), and genome editing CRIPR/Cas9 system (Ran et al., 2013), could be employed to generate high fecundity *bzip19* mutants in calamine or ultramafic *N. caerulescens* backgrounds (**Chapter 5)**. The recommended target of modification in *NcbZIP19* will be the Zn-sensing Cys/His-rich motif. The modification/deletion of this motif induces constitutive expression of Zn-deficiency responsive genes, independent of Zn-supply,

in *A. thaliana* (Lilay et al., 2021). Likewise, modification of Cys/His-rich motif will provide new insights on the bZIP19 contribution to the Zn homeostasis regulatory network in metal hyperaccumulators. The insights gained from this research may aid in the production of crops with enhanced zinc content, which addresses nutritional deficiencies and promotes better health in low-income countries where Zn deficiency in human is prevalent.

Noccaea caerulescens is a well-characterized multiple metal hyperaccumulating species from the *Noccaea* genus (Al-Shehbaz, 2014). In field conditions, this species grows as a small rosette plant with a slow growth rate and low biomass ranging between 140 to 360 mg (Knight et al., 1997), which reduces the feasibility to use it as a phytoremediation plant, to remove toxic concentrations of Zn, Cd or Ni from soils, even though its metal accumulation capacities are outstanding. It has been reported that soil amendments, such as fertilizer (Jacobs et al., 2019; Monsant et al., 2008; Simmons et al., 2015; Sirguey et al., 2006; Xie et al., 2009), biochar (Rees et al., 2020) and beneficial microorganisms (Yung et al., 2021), are able to enhance the shoot biomass and phytoextraction efficiency of this species. Plant density also contributes to the determination of shoot biomass (Jacobs et al., 2018, 2019). However, the increased shoot biomass is still considered relatively low, which is directly tied to the time required for phytoremediation.

Although *N. caerulescens* display elevated metal accumulation capacities, its agronomics traits, in particular for low biomass and slow growth rate, are not suitable for use as phytoremediation plant. The knowledge of molecular mechanism controlling metal accumulation in *N. caerulescens* and *A. thaliana* could be exploited for phytoremediation purpose. Within *Noccaea* genus, there are various species with ploidy levels ranging from diploid to dodecaploid (Jiménez-Ambriz et al., 2007; M. Koch and Al-Shehbaz, 2004; Mandáková et al., 2015; Peer et al., 2003, 2006). For instance, *Noccaea jankae* (2n = 4x = 32), *Noccaea montana* (2n = 4x = 28), and *Noccaea goesingense* (2n = 8x = 56). This polyploid species within the *Noccae*a genus are anticipated to exhibit significantly higher biomass compared to diploid *N. caerulescens* (Corneillie et al., 2019; Husband et al., 2013; Sattler et al., 2015). This increased biomass makes them excellent candidates for phytoremediation purpose. The enhanced biomass can contribute to higher metal phytoextraction per unit area, expediting the phytoremediation process. While genetic transformation has been utilized to create transgenic plants with transgenes involved in metal accumulation and tolerance across various plant species (for a comprehensive review see Fasani et al. (2018)), the random insertion of these transgenes into the recipient genome can lead to silencing and unintended effects (Miglani, 2017). The adoption of genome editing approaches enables precise modifications at specific regions in the genome, minimizing unwanted or genotoxic side effects. Beyond insertions, genome editing allows for deletions and substitutions of DNA sequences of interest (Miglani, 2017), providing a more flexible and versatile system for modifying gene regulatory networks controlling metal accumulation and tolerance in *Noccaea spp.*, such as alteration of *bZIP19* expression. To assess the phenotypic outcomes, this modification can be initially tested in diploid *N. caerulescens*, which undesirable phenotypes may occur resulting from off-target editing (Chu and Agapito-Tenfen, 2022; Modrzejewski et al., 2019). Once desirable phenotypes, particularly enhanced metal accumulation capacity, are observed in genome-edited *N. caerulescens*, genomic editing can be pursued in polyploid *Noccaea spp*. This presents an opportunity to harness the potential of polyploid *Noccaea* species for creating efficient genome-edited plants for phytoremediation purpose.

This thesis reports the advancements in forward and reverse genetic tools to study the metal hyperaccumulator plant model species *N. caerulescens*. These tools represents a great new asset for the

plant metal scientific community, offering opportunities for fundamental research to enhance our understanding of the evolution of an adaptive trait, metal hyperaccumulation/hypertolerance, and the associated selection processes. Moreover, these tools hold promises for practical applications, including phytoremediation, biofortification, and phytomining.

References

Ai, X.-Y., Lin, G., Sun, L.-M., Hu, C.-G., Guo, W.-W., Deng, X.-X., & Zhang, J.-Z. (2012). A global view of gene activity at the flowering transition phase in precocious trifoliate orange and its wild-type [Poncirus trifoliata (L.) Raf.] by transcriptome and proteome analysis. *Gene*, *510*(1), 47–58.

Akakpo, R., Carpentier, M. C., Ie Hsing, Y., & Panaud, O. (2020). The impact of transposable elements on the structure, evolution and function of the rice genome. In *New Phytologist* (Vol. 226, Issue 1, pp. 44–49). Blackwell Publishing Ltd. https://doi.org/10.1111/nph.16356

Alonso-Blanco, C., Koornneef, M., & van Ooijen, J. W. (2006). QTL Analysis. In J. Salinas & J. J. Sanchez-Serrano (Eds.), *Arabidopsis Protocols* (pp. 79–99). Humana Press. https://doi.org/10.1385/1-59745-003-0:79

Alseekh, S., Kostova, D., Bulut, M., & Fernie, A. R. (2021). Genome-wide association studies: assessing trait characteristics in model and crop plants. In *Cellular and Molecular Life Sciences* (Vol. 78, Issue 15, pp. 5743–5754). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00018-021-03868-w

Al-Shehbaz, I. A. (2014). A synopsis of the genus Noccaea (Coluteocarpeae, Brassicaceae). *Harvard Papers in Botany*, *19*(1), 25–51. https://doi.org/10.3100/hpib.2014.n3

Amasino, R. (2010). Seasonal and developmental timing of flowering. *The Plant Journal*, *61*(6), 1001–1013.

Aoshima, K. (2016). Itai-itai disease: Renal tubular osteomalacia induced by environmental exposure to cadmium—historical review and perspectives. *Soil Science and Plant Nutrition*, *62*(4), 319–326.

Arends, D., Prins, P., Jansen, R. C., & Broman, K. W. (2010). R/qtl: High-throughput multiple QTL mapping. *Bioinformatics*, *26*(23), 2990–2992. https://doi.org/10.1093/bioinformatics/btq565

Ashraf, M. A., Maah, M. J., & Yusoff, I. (2014). Soil contamination, risk assessment and remediation. *Environmental Risk Assessment of Soil Contamination*, *1*, 3–56.

Ashraf, M., Ahmad, M. S. A., & Ozturk, M. (2010). Plant adaptation and phytoremediation. In *Plant Adaptation and Phytoremediation*. https://doi.org/10.1007/978-90-481-9370-7

Assunção, A. G. L. A. G. L., Pieper, B., Vromans, J., Lindhout, P., Aarts, M. G. M. M. G. M., & Schat, H. (2006). Construction of a genetic linkage map of Thlaspi caerulescens and quantitative trait loci analysis of zinc accumulation. *The New Phytologist*, *170*(1), 21–32. https://doi.org/10.1111/j.1469-8137.2005.01631.x

Assunção, A. G. L., Bleeker, P., ten Bookum, W. M., Vooijs, R., & Schat, H. (2008). Intraspecific variation of metal preference patterns for hyperaccumulation in Thlaspi caerulescens: Evidence from binary metal exposures. *Plant and Soil*, *303*(1–2), 289–299. https://doi.org/10.1007/s11104-007-9508-x

Assunção, A. G. L., Bookum, W. M. Ten, Nelissen, H. J. M., Vooijs, R., Schat, H., & Ernst, W. H. O. (2003a). A cosegregation analysis of zinc (Zn) accumulation and Zn tolerance in the Zn hyperaccumulator Thlaspi caerulescens. *New Phytologist*, *159*(2), 383–390. https://doi.org/10.1046/j.1469-8137.2003.00758.x

Assunção, A. G. L., Bookum, W. M., Nelissen, H. J. M., Vooijs, R., Schat, H., & Ernst, W. H. O. (2003b). Differential metal-specific tolerance and accumulation patterns among Thlaspi caerulescens populations originating from different soil types. *New Phytologist*, *159*(2), 411–419. https://doi.org/10.1046/j.1469-8137.2003.00819.x

Assunção, A. G. L., CostaMartins, P. Da, Folter, S. De, Vooijs, R., Schat, H., & Aarts, M. G. M. (2001). Elevated expression of metal transporter genes in three accessions of the metal hyperaccumulator Thlaspi caerulescens. *Plant, Cell and Environment*, *24*(2), 217–226. https://doi.org/10.1046/j.1365-3040.2001.00666.x

Assunção, A. G. L., Schat, H., & Aarts, M. G. M. (2003). Thlaspi caerulescens, an attractive model species to study heavy metal hyperaccumulation in plants. *New Phytologist*, *159*(2), 351–360. https://doi.org/10.1046/j.1469-8137.2003.00820.x

Assunção, A. G. L., Schat, H., & Aarts, M. G. M. (2010). Regulation of the adaptation to zinc deficiency in plants. *Plant Signaling & Behavior*, *5*(12), 1553–1555. https://doi.org/10.4161/psb.5.12.13469

Assuncao, A., Herrero, E., Lin, Y. F., Huettel, B., Talukdar, S., Smaczniak, C., Immink, R. G., van Eldik, M., Fiers, M., Schat, H., & Aarts, M. G. (2010). Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc Natl Acad Sci U S A*, *107*(22), 10296–10301. https://doi.org/10.1073/pnas.1004788107

Assunção, A., Pieper, B., Vromans, J., Lindhout, P., Aarts, M., & Schat, H. (2006). Construction of a genetic linkage map of Thlaspi caerulescens and quantitative trait loci analysis of zinc accumulation. *The New Phytologist*, *170*(1), 21–32. https://doi.org/10.1111/j.1469-8137.2005.01631.x

Assunção, A., Schat, H., & Aarts, M. (2003). Thlaspi caerulescens, an attractive model species to study heavy metal hyperaccumulation in plants. *New Phytologist*, *159*(2), 351–360. https://doi.org/10.1046/j.1469-8137.2003.00820.x

Atafar, Z., Mesdaghinia, A., Nouri, J., Homaee, M., Yunesian, M., Ahmadimoghaddam, M., & Mahvi, A. H. (2010). Effect of fertilizer application on soil heavy metal concentration. *Environmental Monitoring and Assessment*, *160*(1–4), 83–89. https://doi.org/10.1007/s10661-008-0659-x

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Muliyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*(7298), 627–631. https://doi.org/10.1038/nature08800

Ayangbenro, A. S., & Babalola, O. O. (2017). A new strategy for heavy metal polluted environments: a review of microbial biosorbents. *International Journal of Environmental Research and Public Health*, *14*(1), 94.

Baduel, P., Hunter, B., Yeola, S., & Bomblies, K. (2018). Genetic basis and evolution of rapid cycling in railway populations of tetraploid Arabidopsis arenosa. *PLoS Genetics*, *14*(7), e1007510.

Baker, A. J. M. (1987). Metal tolerance. *New Phytologist*, *106*, 93–111. https://doi.org/10.1111/j.1469-8137.1987.tb04685.x

Baker, A. J. M., & Brooks, R. R. (1989). Terrestrial higher plants which hyperaccumulate metallic elements - a review of their distribution, ecology and phytochemistry. *Biorecovery*, *1*(2), 81–126.

Baker, A. J. M., Ernst, W. H. O., van der Ent, A., Malaisse, F., & Ginocchio, R. (2010). Metallophytes: the unique biological resource, its ecology and conservational status in Europe, central Africa and Latin America. *Ecology of Industrial Pollution*, *18*, 7–40.

Baker, A. J. M., Reeves, R. D., & Hajar, A. S. M. (1994). Heavy metal accumulation and tolerance in British populations of the metallophyte Thlaspi caerulescens J. & C. Presl (Brassicaceae). *New Phytologist*, *127*(1), 61–68. https://doi.org/10.1111/j.1469-8137.1994.tb04259.x

Banásová, V., Horak, O., Čiamporová, M., Nadubinská, M., & Lichtscheidl, I. (2006). The vegetation of metalliferous and non-metalliferous grasslands in two former mine regions in Central Slovakia. *Biologia*, *61*(4), 433–439. https://doi.org/10.2478/s11756-006-0073-1

Basharat, Z., Novo, L., & Yasmin, A. (2018). Genome Editing Weds CRISPR: What Is in It for Phytoremediation? *Plants*, *7*(3), 51. https://doi.org/10.3390/plants7030051

Baumgarten, L., Pieper, B., Song, B., Mane, S., Lempe, J., Lamb, J., Cooke, E. L., Srivastava, R., Strütt, S., & Žanko, D. (2023). Pan-European study of genotypes and phenotypes in the Arabidopsis relative Cardamine hirsuta reveals how adaptation, demography, and development shape diversity patterns. *PLoS Biology*, *21*(7), e3002191.

Bayçu, G., Gevrek-Kürüm, N., Moustaka, J., Csatári, I., Rognes, S. E., & Moustakas, M. (2017). Cadmium-zinc accumulation and photosystem II responses of Noccaea caerulescens to Cd and Zn exposure. *Environmental Science and Pollution Research*, *24*(3), 2840–2850. https://doi.org/10.1007/s11356-016-8048-4

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis*, *53*(8), 474–485. https://doi.org/10.1002/dvg.22877

Berner, D. (2019). Allele frequency difference AFD-an intuitive alternative to FST for quantifying genetic population differentiation. *Genes*, *10*(4). https://doi.org/10.3390/genes10040308

Bert, V., Macnair, M. R., Laguerie, P. De, Saumitou-Laprade, P., & Petit, D. (2000). Zinc tolerance and accumulation in metallicolous and nonmetallicolous populations of Arabidopsis halleri (Brassicaceae). *New Phytologist*, *146*(2), 225–233. https://doi.org/10.1046/j.1469-8137.2000.00634.x

Bert, V., Meerts, P., Saumitou-Laprade, P., Salis, P., Gruber, W., & Verbruggen, & N. (2003). Genetic basis of Cd tolerance and hyperaccumulation in Arabidopsis halleri. In *Plant and Soil* (Vol. 249).

Berti, W. R., & Cunningham, S. D. (2000). Phytostabilization of metals. *Phytoremediation of Toxic Metals: Using Plants to Clean up the Environment. Wiley, New York*, 71–88.

Besnard, G. (2009). *Thlaspi caerulescens (Brassicaceae) population genetics in western Switzerland: is the genetic structure affected by natural variation of soil heavy metal concentrations?* 974–984. http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2008.02706.x/pdf

Bilton, T. P., Schofield, M. R., Black, M. A., Chagné, D., Wilcox, P. L., & Dodds, K. G. (2018). Accounting for errors in low coverage high-throughput sequencing data when constructing genetic maps using biparental outcrossed populations. *Genetics*, *209*(1), 65–76.

Birke, M., Reimann, C., Oorts, K., Rauch, U., Demetriades, A., Dinelli, E., Ladenberger, A., Halamić, J., Gosar, M., & Jähne-Klingberg, F. (2016). Use of GEMAS data for risk assessment of cadmium in European agricultural and grazing land soil under the REACH Regulation. *Applied Geochemistry*, *74*, 109–121. https://doi.org/10.1016/j.apgeochem.2016.08.014

Blande, D., Halimaa, P., Tervahauta, A. I., Aarts, M. G. M., & Kärenlampi, S. O. (2017). De novo transcriptome assemblies of four accessions of the metal hyperaccumulator plant Noccaea caerulescens. *Scientific Data*, *4*(January), 1–9. https://doi.org/10.1038/sdata.2016.131

Blümel, M., Dally, N., & Jung, C. (2015). Flowering time regulation in crops—what did we learn from Arabidopsis? *Current Opinion in Biotechnology*, *32*, 121–129.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bouché, F., Lobet, G., Tocquin, P., & Périlleux, C. (2016). FLOR-ID: an interactive database of flowering-time gene networks in Arabidopsis thaliana. *Nucleic Acids Research*, *44*(D1), D1167–D1171.

Boyd, R. S. (2007). The defense hypothesis of elemental hyperaccumulation: status, challenges and new directions. *Plant and Soil*, *293*, 153–176.

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

Brachi, B., Morris, G. P., Borevitz, J. O., Morris, G. P., & Brachi, J. O. B. B. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biology*, *12*(232), 232. https://doi.org/10.1186/gb-2011-12-10-232

Broman, K. W. (2010). A Guide to QTL Mapping with R. *Journal of Statistical Software, Book Reviews*, *32*(5), 1–3. https://doi.org/http://dx.doi.org/10.18637/jss.v032.b05

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, *19*(7), 889–890. https://doi.org/10.1093/bioinformatics/btg112

Brooks, R., Chambers, M. F., Nicks, L. J., & Robinson, B. H. (1998). Phytomining. *Trends in Plant Science*, *3*(9), 359–362. https://doi.org/https://doi.org/10.1016/S1360-1385(98)01283-7

Brooks, R. R. (1994). Plants that hyperaccumulate heavy metals. *Plants and the Chemical Elements: Biochemistry, Uptake, Tolerance and Toxicity*, 87–105.

Brooks, R. R. R., Lee, J., Reeves, R. D. D., & Jaffre, T. (1977). Detection of nickeliferous rocks by analysis of herbarium specimens of indicator plants. *Journal of Geochemical Exploration*, *7*, 49–57. https://doi.org/https://doi.org/10.1016/0375-6742(77)90074-7

Buescher, E., Achberger, T., Amusan, I., Giannini, A., Ochsenfeld, C., Rus, A., Lahner, B., Hoekenga, O., Yakubova, E., & Harper, J. F. (2010). Natural genetic variation in selected populations of Arabidopsis thaliana is associated with ionomic differences. *PLoS One*, *5*(6), e11081.

Cai, K., Yu, Y., Zhang, M., & Kim, K. (2019). Concentration, source, and total health risks of cadmium in multiple media in densely populated areas, China. *International Journal of Environmental Research and Public Health*, *16*(13). https://doi.org/10.3390/ijerph16132269

Caicedo, A. L., Stinchcombe, J. R., Olsen, K. M., Schmitt, J., & Purugganan, M. D. (2004). Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences*, *101*(44), 15670–15675.

Campos, A. C. A. L., van Dijk, W. F. A., Ramakrishna, P., Giles, T., Korte, P., Douglas, A., Smith, P., & Salt, D. E. (2021). 1,135 ionomes reveal the global pattern of leaf and seed mineral nutrient and trace element diversity in Arabidopsis thaliana. *The Plant Journal*, *106*(2), 536–554.

Cantó-Pastor, A., Mason, G. A., Brady, S. M., & Provart, N. J. (2021). Arabidopsis bioinformatics: tools and strategies. *The Plant Journal*, *108*(6), 1585–1596. https://doi.org/https://doi.org/10.1111/tpj.15547

Carré, F., Caudeville, J., Bonnard, R., Bert, V., Boucard, P., & Ramel, M. (2017). Soil contamination and human health: a major challenge for global soil security. *Global Soil Security*, 275–295.

Cartwright, D. A., Troggio, M., Velasco, R., & Gutin, A. (2007). Genetic mapping in the presence of genotyping errors. *Genetics*, *176*(4), 2521–2527.

Cassidy, E. S., West, P. C., Gerber, J. S., & Foley, J. A. (2013). Redefining agricultural yields: from tonnes to people nourished per hectare. *Environmental Research Letters*, *8*(3), 034015.

Castanera, R., Morales-Díaz, N., Gupta, S., Purugganan, M., & Casacuberta, J. M. (2023). Transposons are important contributors to gene expression variability under selection in rice populations. *ELife*, *12*. https://doi.org/10.7554/eLife.86324

Chaney, R., Angle, J. s, Baker, A., & Li, Y. M. (2004). Method for phytomining of nickel, cobalt and other metals from soil. *United States Patent*, *5*.

Chaney, R. L., & Baklanov, I. A. (2017). Phytoremediation and Phytomining: Status and Promise. *Advances in Botanical Research*, *83*, 189–221. https://doi.org/10.1016/bs.abr.2016.12.006

Chaney, R. L., Malik, M., Li, Y. M., Brown, S. L., Brewer, E. P., Angle, J. S., & Baker, A. J. M. (1997). Phytoremediation of soil metals. *Current Opinion in Biotechnology*, *8*(3), 279–284.

Chang, L. W., Magos, L., & Suzuki, T. (1996). *Toxicology of metals*. CRC Boca Raton, FL.

Chang, Y.-L., Tao, Q., Scheuring, C., Ding, K., Meksem, K., & Zhang, H.-B. (2001). An integrated map of Arabidopsis thaliana for functional analysis of its genome sequence. *Genetics*, *159*(3), 1231–1242.

Chao, D. Y., Silva, A., Baxter, I., Huang, Y. S., Nordborg, M., Danku, J., Lahner, B., Yakubova, E., & Salt, D. E. (2012). Genome-Wide Association Studies Identify Heavy Metal ATPase3 as the Primary Determinant of Natural Variation in Leaf Cadmium in Arabidopsis thaliana. *PLoS Genetics*, *8*(9). https://doi.org/10.1371/journal.pgen.1002923

Chen, S., Sahito, Z. A., Zhang, M., Feng, Y., Yang, Q., & Yang, X. (2018). Identification and characterization of four nicotianamine synthase genes in Sedum alfredii Hance. *Journal of Biobased Materials and Bioenergy*, *12*(6), 551–559.

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *Plant Journal*, *89*(4), 789–804. https://doi.org/10.1111/tpj.13415

Cheng, J.-Z., Zhou, Y.-P., Lv, T.-X., Xie, C.-P., & Tian, C.-E. (2017). Research progress on the autonomous flowering time pathway in Arabidopsis. *Physiology and Molecular Biology of Plants*, *23*, 477–485.

Chu, P., & Agapito-Tenfen, S. Z. (2022). Unintended genomic outcomes in current and next generation GM techniques: A systematic review. *Plants*, *11*(21), 2997.

Clemens, S., Aarts, M. G. M., Thomine, S., & Verbruggen, N. (2013). Plant science: the key to preventing slow cadmium poisoning. *Trends in Plant Science*, *18*(2), 92–99.

Clough, S. J., & Bent, A. F. (1998). Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. In *The Plant Journal* (Issue 6, pp. 735–743). www.stanford.edu/cgi-bin/biosci_arabidopsis

Corneillie, S., De Storme, N., Van Acker, R., Fangel, J. U., De Bruyne, M., De Rycke, R., Geelen, D., Willats, W. G. T., Vanholme, B., & Boerjan, W. (2019). Polyploidy affects plant growth and alters cell wall composition. *Plant Physiology*, *179*(1), 74–87.

Corso, M., Schvartzman, M. S., Guzzo, F., Souard, F., Malkowski, E., Hanikenne, M., & Verbruggen, N. (2018). Contrasting cadmium resistance strategies in two metallicolous populations of Arabidopsis halleri. *New Phytologist*, *218*(1), 283–297. https://doi.org/10.1111/nph.14948

Cosio, C., Martinoia, E., & Keller, C. (2004). Hyperaccumulation of Cadmium and Zinc in Thlaspi caerulescens and Arabidopsis halleri at the Leaf Cellular Level. *Plant Physiology*, *134*(2), 716–725. https://doi.org/10.1104/pp.103.031948

Craciun, A. R., Meyer, C.-L., Chen, J., Roosens, N., De Groodt, R., Hilson, P., & Verbruggen, N. (2012). Variation in HMA4 gene copy number and expression among Noccaea caerulescens populations presenting different levels of Cd tolerance and accumulation. *Journal of Experimental Botany*, *63*(11), 4179–4189. https://doi.org/10.1093/jxb/ers104

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3–13. https://doi.org/10.1016/j.margen.2016.04.012

Dalcorso, G., Manara, A., & Furini, A. (2013). An overview of heavy metal challenge in plants: From roots to shoots. In *Metallomics* (Vol. 5, Issue 9, pp. 1117–1132). https://doi.org/10.1039/c3mt00038a

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2), giab008.

Davis, A. M., Hall, A., Millar, A. J., Darrah, C., & Davis, S. J. (2009). Protocol: Streamlined sub-protocols for floral-dip transformation and selection of transformants in arabidopsis thaliana. *Plant Methods*, *5*(1). https://doi.org/10.1186/1746-4811-5-3

De Abreu-Neto, J. B., Turchetto-Zolet, A. C., De Oliveira, L. F. V., Bodanese Zanettini, M. H., & Margis-Pinheiro, M. (2013). Heavy metal-associated isoprenylated plant protein (HIPP): Characterization of a family of proteins exclusive to plants. *FEBS Journal*, *280*(7), 1604–1616. https://doi.org/10.1111/febs.12159

De Freitas, J., Wintz, H., Hyoun Kim, J., Poynton, H., Fox, T., & Vulpe, C. (2003). Yeast, a model organism for iron and copper metabolism studies. *Biometals*, *16*(1), 185–197. https://doi.org/10.1023/A:1020771000746

Deniau, A. X. X., Pieper, B., Bookum, W. M. M. Ten, Lindhout, P., Aarts, M. G. M. G. M., & Schat, H. (2006). QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator Thlaspi caerulescens. *Theoretical and Applied Genetics*, *113*(5), 907–920. https://doi.org/10.1007/s00122-006-0350-y

Dharma-Wardana, M. W. C. (2018). Fertilizer usage and cadmium in soils, crops and food. *Environmental Geochemistry and Health*, *40*(6), 2739–2759.

Dinh, N., van der Ent, A., Mulligan, D. R., & Nguyen, A. V. (2018). Zinc and lead accumulation characteristics and in vivo distribution of Zn2+in the hyperaccumulator Noccaea caerulescens elucidated with fluorescent probes and laser confocal microscopy. *Environmental and Experimental Botany*, *147*(October 2017), 1–12. https://doi.org/10.1016/j.envexpbot.2017.10.008

Domínguez, M., Dugas, E., Benchouaia, M., Leduque, B., Jiménez-Gómez, J. M., Colot, V., & Quadrana, L. (2020). The impact of transposable elements on tomato diversity. *Nature Communications*, *11*(1), 4058. https://doi.org/10.1038/s41467-020-17874-2

Dong, C., He, F., Berkowitz, O., Liu, J., Cao, P., Tang, M., Shi, H., Wang, W., Li, Q., & Shen, Z. (2018). Alternative splicing plays a critical role in maintaining mineral nutrient homeostasis in rice (Oryza sativa). *The Plant Cell*, *30*(10), 2267–2285.

Ellis, T. H. N., & Poyser, S. J. (2002). An integrated and comparative view of pea genetic and cytogenetic maps. *New Phytologist*, *153*(1), 17–25.

Ent, A. Van Der, Baker, A. J. M., Reeves, R. D., Chaney, R. L., Anderson, C. W. N., Meech, J. A., Erskine, P. D., Simonnot, M. O., Vaughan, J., Morel, J. L., Echevarria, G., Fogliani, B., Rongliang, Q., & Mulligan, D. R. (2015). Agromining: Farming for metals in the future? *Environmental Science and Technology*, *49*(8), 4773–4780. https://doi.org/10.1021/es506031u

Erlangung, D. Zur, & Doktor, G. (2017). *Functional analysis of the metal hyperaccumulation and hypertolerance in Arabidopsis halleri*. *June*.

Ernst, W. H. O. (2006). Evolution of metal tolerance in higher plants. In *For. Snow Landsc. Res* (Vol. 80).

Escarré, J., Lefèbvre, C., Frérot, H., Mahieu, S., & Noret, N. (2013). Metal concentration and metal mass of metallicolous, non metallicolous and serpentine Noccaea caerulescens populations, cultivated in different growth media. *Plant and Soil*, *370*(1–2), 197–221. https://doi.org/10.1007/s11104-013-1618-z

Escarré, J., Lefèbvre, C., Gruber, W., Leblanc, M., Lepart, J., Rivière, Y., & Delay, B. (2000). Zinc and cadmium hyperaccumulation by Thlaspi caerulescens from metalliferous and nonmetalliferous sites in the Mediterranean area: implications for phytoremediation. *New Phytologist*, *145*(3), 429–437. https://doi.org/DOI: undefined

European commission. (2006). *Thematic Strategy for Soil Protection*.

Evanko, C. R., & Dzombak, D. A. (1997). *Remediation of metals-contaminated soils and groundwater*. Citeseer.

Fang, Q., Liu, J., Xu, Z., & Song, R. (2008). Cloning and characterization of a flowering time gene from Thellungiella halophila. *Acta Biochimica et Biophysica Sinica*, *40*(8), 747–753.

Fang, Q., Xu, Z., & Song, R. (2006). Cloning, characterization and genetic engineering of FLC homolog in Thellungiella halophila. *Biochemical and Biophysical Research Communications*, *347*(3), 707–714.

FAO, & UNEP. (2021). *Global assessment of soil pollution: Report*. FAO and UNEP. https://doi.org/10.4060/cb4894en

Fasani, E., Manara, A., Martini, F., Furini, A., & DalCorso, G. (2018). The potential of genetic engineering of plants for the remediation of soils contaminated with heavy metals. *Plant, Cell & Environment*, *41*(5), 1201–1232.

Feldmann, K. A., & David Marks, M. (1987). Agrobacterium-mediated transformation of germinating seeds of Arabidopsis thaliana: a non-tissue culture approach. *Molecular and General Genetics MGG*, *208*(1–2), 1–9.

Fernandes, J. B., Wlodzimierz, P., & Henderson, I. R. (2019). Meiotic recombination within plant centromeres. *Current Opinion in Plant Biology*, *48*, 26–35.

Flood, P. J., Kruijer, W., Schnabel, S. K., Schoor, R., Jalink, H., Snel, J. F. H., Harbinson, J., & Aarts, M. G. M. (2016). Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals circadian and long-term fluctuations in heritability. *Plant Methods*, *12*(1), 1–14. https://doi.org/10.1186/s13007-016-0113-y

Foroughi, S., Baker, A. J. M., Roessner, U., Johnson, A. A. T., Bacic, A., & Callahan, D. L. (2014). Hyperaccumulation of zinc by Noccaea caerulescens results in a cascade of stress responses and changes in the elemental profile. *Metallomics*, *6*(9), 1671–1682. https://doi.org/10.1039/c4mt00132j

Frérot, H., Lefèbvre, C., Petit, C., Collin, C., Santos, A. Dos, & Escarré, J. (2005). Zinc tolerance and hyperaccumulation in F1 and F2 offspring from intra and interecotype crosses of Thlaspi caerulescens. *New Phytologist*, *165*(1), 111–119. https://doi.org/10.1111/j.1469-8137.2004.01227.x

Frérot, H., Petit, C., Lefèbvre, C., Gruber, W., Collin, C., & Escarré, J. (2003). Zinc and cadmium accumulation in controlled crosses between metallicolous and nonmetallicolous populations of Thlaspi caerulescens (Brassicaceae). *New Phytologist*, *157*(3), 643–648. https://doi.org/10.1046/j.1469-8137.2003.00701.x

Friberg, L. (2018). *Cadmium in the Environment*. CRC press.

Gao, J., Faheem, M., & Yu, X. (2022). Global research on contaminated soil remediation: A bibliometric network analysis. *Land*, *11*(9), 1581.

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. http://arxiv.org/abs/1207.3907

Genchi, G., Sinicropi, M. S., Lauria, G., Carocci, A., & Catalano, A. (2020). The effects of cadmium toxicity. In *International Journal of Environmental Research and Public Health* (Vol. 17, Issue 11). MDPI AG. https://doi.org/10.3390/ijerph17113782

Gendre, D., Czernic, P., Conéjéro, G., Pianelli, K., Briat, J. F., Lebrun, M., & Mari, S. (2007). TcYSL3, a member of the YSL gene family from the hyper-accumulator Thlaspi caerulescens, encodes a nicotianamine-Ni/Fe transporter. *Plant Journal*, *49*(1), 1–15. https://doi.org/10.1111/j.1365-313X.2006.02937.x

Ghandilyan, A., Ilk, N., Hanhart, C., Mbengue, M., Barboza, L., Schat, H., Koornneef, M., El-Lithy, M., Vreugdenhil, D., & Reymond, M. (2009). A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three Arabidopsis thaliana RIL populations. *Journal of Experimental Botany*, *60*(5), 1409–1425.

Ghedira, R., Buck, S. De, Nolf, J., & Depicker, A. (2013). The efficiency of Arabidopsis thaliana floral dip transformation is determined not only by the Agrobacterium strain used but also by the physiology and the ecotype of the dipped plant. *Molecular Plant-Microbe Interactions*, *26*(7), 823–832. https://doi.org/10.1094/MPMI-11-12-0267-R

Gomes, B., Calanzani, N., Curiale, V., McCrone, P., Higginson, I. J., & de Brito, M. (2013). Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. *Cochrane Database of Systematic Reviews*, *6*.

Gonneau, C., Genevois, N., Frérot, H., Sirguey, C., & Sterckeman, T. (2014). Variation of trace metal accumulation, major nutrient uptake and growth parameters and their correlations in 22 populations of Noccaea caerulescens. *Plant and Soil*, *384*(1–2), 271–287. https://doi.org/10.1007/s11104-014-2208-4

Gonneau, C., Noret, N., Godé, C., Frérot, H., Sirguey, C., Sterckeman, T., & Pauwels, M. (2017). Demographic history of the trace metal hyperaccumulator Noccaea caerulescens (J. Presl and C. Presl) F. K. Mey. in Western Europe. *Molecular Ecology*, *26*(3), 904–922. https://doi.org/10.1111/mec.13942

Gravot, A., Lieutaud, A., Verret, F., Auroy, P., Vavasseur, A., & Richaud, P. (2004). AtHMA3, a plant P1B-ATPase, functions as a Cd/Pb transporter in yeast. *FEBS Letters*, *561*(1–3), 22–28. https://doi.org/10.1016/S0014-5793(04)00072-9

Gretsova, M., Surkova, S., Kanapin, A., Samsonova, A., Logacheva, M., Shcherbakov, A., Logachev, A., Bankin, M., Nuzhdin, S., & Samsonova, M. (2023). Transcriptomic Analysis of Flowering Time Genes in Cultivated Chickpea and Wild Cicer. *International Journal of Molecular Sciences*, *24*(3), 2692.

Guan, Z. Q., Chai, T. Y., Zhang, Y. X., Xu, J., Wei, W., Han, L., & Cong, L. (2008). Gene manipulation of a heavy metal hyperaccumulator species Thlaspi caerulescens L. via Agrobacterium-mediated transformation. *Molecular Biotechnology*, *40*(1), 77–86. https://doi.org/10.1007/s12033-008-9065-4

Guo, Y.-L., Todesco, M., Hagmann, J., Das, S., & Weigel, D. (2012). Independent FLC mutations as causes of flowering-time variation in Arabidopsis thaliana and Capsella rubella. *Genetics*, *192*(2), 729–739.

Hackett, C. A., & Broadfoot, L. B. (2003). Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity*, *90*(1), 33–38.

Halimaa, P., Blande, D., Aarts, M. G. M., Tuomainen, M., Tervahauta, A., & KÃ¤renlampi, S. (2014). Comparative transcriptome analysis of the metal hyperaccumulator Noccaea caerulescens. *Frontiers in Plant Science*, *5*(May), 1–7. https://doi.org/10.3389/fpls.2014.00213

Halimaa, P., Lin, Y. F., Ahonen, V. H., Blande, D., Clemens, S., Gyenesei, A., Häikiö, E., Kärenlampi, S. O., Laiho, A., Aarts, M. G. M., Pursiheimo, J. P., Schat, H., Schmidt, H., Tuomainen, M. H., & Tervahauta, A. I. (2014). Gene expression differences between noccaea caerulescens ecotypes help to identify candidate genes for metal phytoremediation. *Environmental Science and Technology*, *48*(6), 3344–3353. https://doi.org/10.1021/es4042995

Hammond, J. P., Bowen, H. C., White, P. J., Mills, V., Pyke, K. A., Baker, A. J. M., Whiting, S. N., May, S. T., & Broadley, M. R. (2006). A comparison of the Thlaspi caerulescens and Thlaspi arvense shoot transcriptomes. *New Phytologist*, *170*(2), 239–260. https://doi.org/https://doi.org/10.1111/j.1469-8137.2006.01662.x

Han, F., Ullrich, S. E., Kleinhofs, A., Jones, B. L., Hayes, P. M., & Wesenberg, D. M. (1997). Fine structure mapping of the barley chromosome-1 centromere region containing malting-quality QTLs. *Theoretical and Applied Genetics*, *95*, 903–910.

Han, H. (2018). RNA interference to knock down gene expression. *Disease Gene Identification: Methods and Protocols*, 293–302.

Hanikenne, M., & Nouet, C. (2011). Metal hyperaccumulation and hypertolerance: A model for plant evolutionary genomics. In *Current Opinion in Plant Biology* (Vol. 14, Issue 3, pp. 252–259). https://doi.org/10.1016/j.pbi.2011.04.003

Hanikenne, M., Talke, I. N., Haydon, M. J., Lanz, C., Nolte, A., Motte, P., Kroymann, J., Weigel, D., & Krämer, U. (2008). Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. *Nature*, *453*(7193), 391–395.

Hasan, M. K., Cheng, Y., Kanwar, M. K., Chu, X.-Y., Ahammed, G. J., & Qi, Z.-Y. (2017). Responses of plant proteins to heavy metal stress—a review. *Frontiers in Plant Science*, *8*, 1492.

Hayat, M. T., Nauman, M., Nazir, N., Ali, S., & Bangash, N. (2018). Environmental Hazards of Cadmium: Past, Present, and Future. In *Cadmium Toxicity and Tolerance in Plants: From Physiology to Remediation* (pp. 163–183). Elsevier. https://doi.org/10.1016/B978-0-12-814864-8.00007-3

He, S., Yang, X., He, Z., & Baligar, V. C. (2017). Morphological and Physiological Responses of Plants to Cadmium Toxicity: A Review. *Pedosphere*, *27*(3), 421–438. https://doi.org/10.1016/S1002-0160(17)60339-4

He, Z., Shentu, J., Yang, X., Baligar, V. C., Zhang, T., & Stoffella, P. J. (2015). *Heavy metal contamination of soils: sources, indicators and assessment*.

Hermkes, R., Fu, Y.-F., Nürrenberg, K., Budhiraja, R., Schmelzer, E., Elrouby, N., Dohmen, R. J., Bachmair, A., & Coupland, G. (2011). Distinct roles for Arabidopsis SUMO protease ESD4 and its closest homolog ELS1. *Planta*, *233*, 63–73.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Hey, J. (2004). What's so hot about recombination hotspots? *PLoS Biology*, *2*(6), e190.

Holderegger, R., & Wagner, H. H. (2006). A brief guide to Landscape Genetics. In *Landscape Ecology* (Vol. 21, Issue 6, pp. 793–796). https://doi.org/10.1007/s10980-005-6058-6

Horton, M. W., Hancock, A. M., Huang, Y. S., Toomajian, C., Atwell, S., Auton, A., Muliyati, N. W., Platt, A., Sperone, F. G., Vilhjálmsson, B. J., Nordborg, M., Borevitz, J. O., & Bergelson, J. (2012). Genome-wide patterns of genetic variation in worldwide Arabidopsis thaliana accessions from the RegMap panel. *Nature Genetics*, *44*(2), 212–216. https://doi.org/10.1038/ng.1042

Hou, D., O'Connor, D., Igalavithana, A. D., Alessi, D. S., Luo, J., Tsang, D. C. W., Sparks, D. L., Yamauchi, Y., Rinklebe, J., & Ok, Y. S. (2020). Metal contamination and bioremediation of agricultural soils for food safety and sustainability. In *Nature Reviews Earth and Environment* (Vol. 1, Issue 7, pp. 366–381). Springer Nature. https://doi.org/10.1038/s43017-020-0061-y

Hou, L., Sun, N., Mane, S., Sayward, F., Rajeevan, N., Cheung, K., Cho, K., Pyarajan, S., Aslan, M., & Miller, P. (2017). Impact of genotyping errors on statistical power of association tests in genomic analyses: A case study. *Genetic Epidemiology*, *41*(2), 152–162.

Husband, B. C., Baldwin, S. J., & Suda, J. (2013). *The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes*. Springer.

Hussain, D., Haydon, M. J., Wang, Y., Wong, E., Sherson, S. M., Young, J., Camakaris, J., Harper, J. F., & Cobbett, C. S. (2004). P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. *The Plant Cell*, *16*(5), 1327–1339. https://doi.org/10.1105/tpc.020487

Ingrouille, M. J., & Smirnoff, N. (1986). Thlaspi caerulescens J. & C. Presl. (T. Alpestre L.) in Britain. *New Phytologist*, *102*(1), 219–233. https://doi.org/10.1111/j.1469-8137.1986.tb00812.x

Inouye, D. W. (2008). Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. *Ecology*, *89*(2), 353–362.

Invernón, V. R., Tisserand, R., Jouannais, P., Navarrete Gutiérrez, D. M., Muller, S., Pillon, Y., Echevarria, G., & Merlot, S. (2021). The Discovery of New Metal-Hyperaccumulating Plant Species in Herbaria. In *Natural History Collections in the Science of the 21st Century* (pp. 79–94). https://doi.org/https://doi.org/10.1002/9781119882237.ch6

Ishimaru, Y., Takahashi, R., Bashir, K., Shimo, H., Senoura, T., Sugimoto, K., Ono, K., Yano, M., Ishikawa, S., Arao, T., Nakanishi, H., & Nishizawa, N. K. (2012). Characterizing the role of rice NRAMP5 in Manganese, Iron and Cadmium Transport. *Scientific Reports*, *2*. https://doi.org/10.1038/srep00286

Islam, M. S., Kormoker, T., Idris, A. M., Proshad, R., Kabir, M. H., & Ustaoğlu, F. (2021). Plant–microbe–metal interactions for heavy metal bioremediation: a review. *Crop and Pasture Science*.

Jacobs, A., Brabandere, L. De, Drouet, T., Sterckeman, T., & Noret, N. (2018). Phytoextraction of Cd and Zn with Noccaea caerulescens for urban soil remediation: influence of nitrogen

fertilization and planting density. *Ecological Engineering*, *116*(March), 178–187. https://doi.org/10.1016/j.ecoleng.2018.03.007

Jacobs, A., Noret, N., Van Baekel, A., Liénard, A., Colinet, G., & Drouet, T. (2019). Influence of edaphic conditions and nitrogen fertilizers on cadmium and zinc phytoextraction efficiency of Noccaea caerulescens. *Science of the Total Environment*, *665*, 649–659.

Jaffré, T., Brooks, R. R., Lee, J., & Reeves, R. (1976). Sebertia acuminata: a hyperaccumulator of nickel from New Caledonia. *Science*, *193*(4253), 579–580.

Jansen, R. C. (2004). Quantitative trait loci in inbred lines. *Handbook of Statistical Genetics*.

Järup, L. (2003). Hazards of heavy metal contamination. *British Medical Bulletin*, *68*(1), 167–182.

Jiao, S., Hsu, L., Hutter, C. M., & Peters, U. (2011). The use of imputed values in the meta-analysis of genome-wide association studies. *Genetic Epidemiology*, *35*(7), 597–605.

Jiménez-Ambriz, G., Petit, C., Bourrié, I., Dubois, S., Olivieri, I., & Ronce, O. (2007). Life history variation in the heavy metal tolerant plant Thlaspi caerulescens growing in a network of contaminated and noncontaminated sites in southern France: Role of gene flow, selection and phenotypic plasticity. *New Phytologist*, *173*(1), 199–215. https://doi.org/10.1111/j.1469-8137.2006.01923.x

Jin, Y., Ivanov, M., Dittrich, A. N., Nelson, A. D., & Marquardt, S. (2023). LncRNA FLAIL affects alternative splicing and represses flowering in Arabidopsis . *The EMBO Journal*, *42*(11). https://doi.org/10.15252/embj.2022110921

Jolly, Y. N., Islam, A., & Akbar, S. (2013). Transfer of metals from soil to vegetables and possible health risk assessment. *SpringerPlus*, *2*(1), 1–8. https://doi.org/10.1186/2193-1801-2-385

Kabata-Pendias, A. (2004). Soil–plant transfer of trace elements—an environmental issue. *Geoderma*, *122*(2–4), 143–149.

Kabata-Pendias, A., & Mukherjee, A. B. (2007). *Trace elements from soil to human*. Springer Science & Business Media.

Kahle, D., & Wickham, H. (2013). *ggmap: Spatial Visualization with ggplot2*.

Kamran, A., Iqbal, M., & Spaner, D. (2014). Flowering time in wheat (Triticum aestivum L.): a key factor for global adaptability. *Euphytica*, *197*, 1–26.

Ke, X., Taylor, M. S., & Cardon, L. R. (2008). Singleton SNPs in the human genome and implications for genome-wide association studies. *European Journal of Human Genetics*, *16*(4), 506–515.

Keurentjes, J. J. B., Willems, G., Eeuwijk, F. Van, Nordborg, M., & Koornneef, M. (2011). A comparison of population types used for QTL mapping in Arabidopsis thaliana. *Plant Genetic Resources: Characterisation and Utilisation*, *9*(2), 185–188. https://doi.org/10.1017/S1479262111000086

Khalid, S., Shahid, M., Niazi, N. K., Murtaza, B., Bibi, I., & Dumat, C. (2017). A comparison of technologies for remediation of heavy metal contaminated soils. *Journal of Geochemical Exploration*, *182*, 247–268.

Khan, S., Cao, Q., Zheng, Y. M., Huang, Y. Z., & Zhu, Y. G. (2008). Health risks of heavy metals in contaminated soils and food crops irrigated with wastewater in Beijing, China. *Environmental Pollution*, *152*(3), 686–692.

Klasen, J. R., Piepho, H.-P., & Stich, B. (2012). QTL detection power of multi-parental RIL populations in Arabidopsis thaliana. *Heredity*, *108*(6), 626–632. https://doi.org/10.1038/hdy.2011.133

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Knight, B., Zhao, F. J., McGrath, S. P., & Shen, Z. G. (1997). Zinc and cadmium uptake by the hyperaccumulator Thlaspi caerulescens in contaminated soils and its effects on the concentration and chemical speciation of metals in soil solution. *Plant and Soil*, *197*, 71–78.

Knoch, D., Abbadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., Snowdon, R. J., & Altmann, T. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnology Journal*, *18*(1), 68–82. https://doi.org/10.1111/pbi.13171

Koch, M. A., & German, D. A. (2013). Taxonomy and systematics are key to biological information: Arabidopsis, Eutrema (Thellungiella), Noccaea and Schrenkiella (Brassicaceae) as examples. *Frontiers in Plant Science*, *4*, 267.

Koch, M., & Al-Shehbaz, I. A. (2004). Taxonomic and phylogenetic evaluation of the American "Thlaspi" species: Identity and relationship to the Eurasian genus Noccaea (Brassicaceae). *Systematic Botany*, *29*(2), 375–384. https://doi.org/10.1600/036364404774195566

Koch, M., & Bernhardt, K. (2004). Comparative biogeography of the cytotypes of annual Microthlaspi perfoliatum (Brassicaceae) in Europe using isozymes and cpDNA data: refugia, diversity centers, and postglacial colonization. *American Journal of Botany*, *91*(1), 115–124.

Koch, M., Mummenhoff, K., & Hurka, H. (1998). Systematics and evolutionary history of heavy metal tolerant Thlaspi caerulescens in Western Europe: evidence from genetic studies based on isozyme analysis. In *Biochemical Systematics and Ecology* (Vol. 26).

Kong, X., Luo, X., Qu, G. P., Liu, P., & Jin, J. B. (2017). Arabidopsis SUMO protease ASP1 positively regulates flowering time partially through regulating FLC stability. *Journal of Integrative Plant Biology*, *59*(1), 15–29. https://doi.org/10.1111/jipb.12509

Koornneef, M., Alonso-Blanco, C., Peeters, A. J. M., & Soppe, W. (1998). Genetic control of flowering time in Arabidopsis thaliana. In *Annu. Rev. Plant Physiol. Plant Mol. Biol* (Vol. 49, pp. 345–370). www.annualreviews.org

Koornneef, M., & Meinke, D. (2010). The development of Arabidopsis as a model plant. *Plant Journal*, *61*(6), 909–921. https://doi.org/10.1111/j.1365-313X.2009.04086.x

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis awith GWAS: a review. *Plant Methods*, *9*(29), 1–9.

Kosambi, D. D. (1943). The estimation of map distances from recombination values. *Annals of Eugenics*, *12*(1), 172–175. https://doi.org/https://doi.org/10.1111/j.1469-1809.1943.tb02321.x

Kozhevnikova, A. D., Seregin, I. V, Aarts, M. G. M., & Schat, H. (2020). Intra-specific variation in zinc, cadmium and nickel hypertolerance and hyperaccumulation capacities in Noccaea caerulescens. *Plant and Soil*, *452*(1–2), 479–498. https://doi.org/10.1007/s11104-020-04572-7

Krämer, U. (2010). Metal Hyperaccumulation in Plants. *Annual Review of Plant Biology*, *61*(1), 517–534. https://doi.org/10.1146/annurev-arplant-042809-112156

Kryvokhyzha, D., Holm, K., Chen, J., Cornille, A., Glémin, S., Wright, S. I., Lagercrantz, U., & Lascoux, M. (2016). The influence of population structure on gene expression and flowering time variation in the ubiquitous weed Capsella bursa-pastoris (Brassicaceae). *Molecular Ecology*, *25*(5), 1106–1121. https://doi.org/10.1111/mec.13537

Kuittinen, H., Niittyvuopio, A., Rinne, P., & Savolainen, O. (2008). Natural variation in Arabidopsis lyrata vernalization requirement conferred by a FRIGIDA indel polymorphism. *Molecular Biology and Evolution*, *25*(2), 319–329.

Kumar, M. S. J. (2015). Adaptation Strategies of Plants against Heavy Metal Toxicity: A Short Review. *Biochemistry & Pharmacology: Open Access*, *04*(02). https://doi.org/10.4172/2167-0501.1000161

Küpper, H., & Kochian, L. V. (2010). Transcriptional regulation of metal transport genes and mineral nutrition during acclimatization to cadmium and zinc in the Cd/Zn hyperaccumulator, Thlaspi caerulescens (Ganges population). *New Phytologist*, *185*(1), 114–129. https://doi.org/https://doi.org/10.1111/j.1469-8137.2009.03051.x

Küpper, H., Lombi, E., Zhao, F.-J., & McGrath, S. P. (2000). Cellular compartmentation of cadmium and zinc in relation to other elements in the hyperaccumulator Arabidopsis halleri. *Planta*, *212*(1), 75–84. https://doi.org/10.1007/s004250000366

Lamesch, P., Dreher, K., Swarbreck, D., Sasidharan, R., Reiser, L., & Huala, E. (2010). Using The Arabidopsis Information Resource (TAIR) to Find Information About Arabidopsis Genes. *Current Protocols in Bioinformatics*, *30*(1), 1.11.1-1.11.51. https://doi.org/https://doi.org/10.1002/0471250953.bi0111s30

Lanquar, V., Lelièvre, F., Bolte, S., Hamès, C., Alcon, C., Neumann, D., Vansuyt, G., Curie, C., Schröder, A., Krämer, U., Barbier-Brygoo, H., & Thomine, S. (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. *EMBO Journal*, *24*(23), 4041–4051. https://doi.org/10.1038/sj.emboj.7600864

Lanquar, V., Ramos, M. S., Lelièvre, F., Barbier-Brygoo, H., Krieger-Liszkay, A., Krämer, U., & Thomine, S. (2010). Export of vacuolar manganese by AtNRAMP3 and AtNRAMP4 is required for optimal photosynthesis and growth under manganese deficiency. *Plant Physiology*, *152*(4), 1986–1999. https://doi.org/10.1104/pp.109.150946

Lê, S., Josse, J., Rennes, A., & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. In *JSS Journal of Statistical Software* (Vol. 25). http://www.jstatsoft.org/

Lee, C. R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D., & Nordborg, M. (2017). On the post-glacial spread of human commensal Arabidopsis thaliana. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14458

Lee, C.-R., Hsieh, J.-W., Schranz, M. E., & Mitchell-Olds, T. (2018). The functional change and deletion of FLC homologs contribute to the evolution of rapid flowering in Boechera stricta. *Frontiers in Plant Science*, *9*, 1078.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, L., Han, L., Liu, A., & Wang, F. (2022). Imperfect but hopeful: New advances in soil pollution and remediation. In *International Journal of Environmental Research and Public Health* (Vol. 19, Issue 16, p. 10164). MDPI.

Li, S., Han, X., Lu, Z., Qiu, W., Yu, M., Li, H., He, Z., & Zhuo, R. (2022). MAPK cascades and transcriptional factors: Regulation of heavy metal tolerance in plants. *International Journal of Molecular Sciences*, *23*(8), 4463.

Lilay, G. H., Persson, D. P., Castro, P. H., Liao, F., Alexander, R. D., Aarts, M. G. M., & Assunção, A. G. L. (2021). Arabidopsis bZIP19 and bZIP23 act as zinc sensors to control plant zinc status. *Nature Plants*, *7*(2), 137–143. https://doi.org/10.1038/s41477-021-00856-7

Lin, Y. F., & Aarts, M. G. M. (2012). The molecular mechanism of zinc and cadmium stress response in plants. In *Cellular and Molecular Life Sciences* (Vol. 69, Issue 19, pp. 3187–3206). https://doi.org/10.1007/s00018-012-1089-z

Lin, Y. F., Hassan, Z., Talukdar, S., Schat, H., & Aarts, M. G. M. (2016). Expression of the Znt1 zinc transporter from the metal hyperaccumulator noccaea caerulescens confers enhanced zinc and cadmium tolerance and accumulation to arabidopsis thaliana. *PLoS ONE*, *11*(3), 1–30. https://doi.org/10.1371/journal.pone.0149750

Lin, Y.-F. (2014). *An evolutionary perspective on differential regulation of zinc and cadmium homeostatis genes in Arabidopsis thaliana and Noccaea caerulescens*. Wageningen University and Research.

Lin, Y.-F., Severing, E. I., te Lintel Hekkert, B., Schijlen, E., & Aarts, M. G. M. (2014). A comprehensive set of transcript sequences of the heavy metal hyperaccumulator Noccaea caerulescens. *Frontiers in Plant Science*, *5*, 261.

Liu, H., Zhao, H., Wu, L., Liu, A., Zhao, F. J., & Xu, W. (2017). Heavy metal ATPase 3 (HMA3) confers cadmium hypertolerance on the cadmium/zinc hyperaccumulator Sedum plumbizincicola. *New Phytologist*, *215*(2), 687–698. https://doi.org/10.1111/nph.14622

Liu, L., Li, W., Song, W., & Guo, M. (2018). Remediation techniques for heavy metal-contaminated soils: Principles and applicability. In *Science of the Total Environment* (Vol. 633, pp. 206–219). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2018.03.161

Lochlainn, S. Ó., Bowen, H. C., Fray, R. G., Hammond, J. P., King, G. J., White, P. J., Graham, N. S., & Broadley, M. R. (2011). Tandem quadruplication of HMA4 in the zinc (Zn) and cadmium (Cd) hyperaccumulator Noccaea caerulescens. *PLoS ONE*, *6*(3). https://doi.org/10.1371/journal.pone.0017814

Lochlainn, S. Ó., Fray, R. G., Hammond, J. P., King, G. J., White, P. J., Young, S. D., & Broadley, M. R. (2011). Generation of nonvernal-obligate, faster-cycling Noccaea caerulescens lines through fast neutron mutagenesis. *New Phytologist*, *189*(2), 409–414. https://doi.org/10.1111/j.1469-8137.2010.03554.x

Lombi, E., Tearall, K. L., Howarth, J. R., Zhao, F. J., Hawkesford, M. J., & McGrath, S. P. (2002). Influence of iron status on cadmium and zinc uptake by different ecotypes of the hyperaccumulator Thlaspi caerulescens. *Plant Physiology*, *128*(4), 1359–1367. https://doi.org/10.1104/pp.010731

Lombi, E., Zhao, F. J., Dunham, S. J., & McGrath, S. P. (2000). Cadmium accumulation in populations of Thlaspi caerulescens and Thlaspi goesingense. *New Phytologist*, *145*(1), 11–20. https://doi.org/10.1046/j.1469-8137.2000.00560.x

Lombi, E., Zhao, F. J., McGrath, S. P., Young, S. D., & Sacchi, G. A. (2001). Physiological evidence for a high-affinity cadmium transporter highly expressed in a thlaspi caerulescens ecotype. *New Phytologist*, *149*(1), 53–60. https://doi.org/10.1046/j.1469-8137.2001.00003.x

Lyttle, T. W. (1991). Segregation distorters. *Annual Review of Genetics*, *25*(1), 511–581.

Macnair, M. R., Bert, V., Huitson, S. B., Saumitou-Laprade, P., & Petit, D. (1999). Zinc tolerance and hyperaccumulation are genetically independent characters. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1434), 2175–2179. https://doi.org/10.1098/rspb.1999.0905

Manara, A., Fasani, E., Furini, A., & DalCorso, G. (2020). Evolution of the metal hyperaccumulation and hypertolerance traits. In *Plant Cell and Environment* (Vol. 43, Issue 12, pp. 2969–2986). Blackwell Publishing Ltd. https://doi.org/10.1111/pce.13821

Mandakova, T., & Lysak, M. A. (2008). Chromosomal phylogeny and karyotype evolution in x= 7 crucifer species (Brassicaceae). *The Plant Cell*, *20*(10), 2559–2570.

Mandáková, T., Singh, V., Krämer, U., & Lysak, M. A. (2015). Genome Structure of the Heavy Metal Hyperaccumulator *Noccaea caerulescens* and Its Stability on Metalliferous and Nonmetalliferous Soils. *Plant Physiology*, *169*(1), 674–689. https://doi.org/10.1104/pp.15.00619

Manea, A., Dumitru, S., Dumitru, M., & Vrinceanu, N. (2013). Assessment of heavy metals contamination of soils in the Zlatna area using the multiple pollution index. *Journal of Environmental Protection and Ecology*, *14*(3), 875–881.

Mar, S. S., & Okazaki, M. (2012). Investigation of Cd contents in several phosphate rocks used for the production of fertilizer. *Microchemical Journal*, *104*, 17–21. https://doi.org/10.1016/j.microc.2012.03.020

Marchini, J., & Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, *11*(7), 499–511.

Mari, S., Gendre, D., Pianelli, K., Ouerdane, L., Lobinski, R., Briat, J.-F., Lebrun, M., & Czernic, P. (2006). Root-to-shoot long-distance circulation of nicotianamine and nicotianamine–nickel chelates in the metal hyperaccumulator Thlaspi caerulescens. *Journal of Experimental Botany*, *57*(15), 4111–4122. https://doi.org/10.1093/jxb/erl184

Meyer, C.-L., Juraniec, M., Huguet, S., Chaves-Rodriguez, E., Salis, P., Isaure, M.-P., Goormaghtigh, E., & Verbruggen, N. (2015). Intraspecific variability of cadmium tolerance and accumulation, and cadmium-induced cell wall modifications in the metal hyperaccumulator Arabidopsis halleri. *Journal of Experimental Botany*, *66*(11), 3215–3227.

Meyer, R. C., Weigelt-Fischer, K., Knoch, D., Heuermann, M., Zhao, Y., & Altmann, T. (2021). Temporal dynamics of QTL effects on vegetative growth in Arabidopsis thaliana. *Journal of Experimental Botany*, *72*(2), 476–490. https://doi.org/10.1093/jxb/eraa490

Michaels, S. D., & Amasino, R. M. (1999). FLOWERING LOCUS C encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant Cell*, *11*(5), 949–956.

Miglani, G. S. (2017). Genome editing in crop improvement: Present scenario and future prospects. *Journal of Crop Improvement*, *31*(4), 453–559.

Mills, R. F., Krijgert, G. C., Baccarini, P. J., Hall, J. L., & Williams, L. E. (2003). Functional expression of AtHMA4, a P1B-type ATPase of the Zn/Co/Cd/Pb subclass. *Plant Journal*, *35*(2), 164–176. https://doi.org/10.1046/j.1365-313X.2003.01790.x

Milner, M. J., & Kochian, L. V. (2008). Investigating heavy-metal hyperaccumulation using Thlaspi caerulescens as a model system. *Annals of Botany*, *102*(1), 3–13. https://doi.org/10.1093/aob/mcn063

Milner, M. J., Mitani-Ueno, N., Yamaji, N., Yokosho, K., Craft, E., Fei, Z., Ebbs, S., Clemencia Zambrano, M., Ma, J. F., & Kochian, L. V. (2014). Root and shoot transcriptome analysis of two ecotypes of Noccaea caerulescens uncovers the role of NcNramp1 in Cd hyperaccumulation. *Plant Journal*, *78*(3), 398–410. https://doi.org/10.1111/tpj.12480

Minguzzi, C., & Vergnano, O. (1948). Il contenuto di nichel nelle ceneri di Alyssum bertolonii Desv. *Atti Della Societa Toscana Di Science Naturali, Mem Ser A*, *55*, 49–77.

Mishra, S., Mishra, A., & Küpper, H. (2017). Protein Biochemistry and Expression Regulation of Cadmium/Zinc Pumping ATPases in the Hyperaccumulator Plants Arabidopsis halleri and Noccaea caerulescens. *Frontiers in Plant Science*, *8*(May), 1–13. https://doi.org/10.3389/fpls.2017.00835

Modareszadeh, M., Bahmani, R., Kim, D. G., & Hwang, S. (2021). CAX3 (cation/proton exchanger) mediates a Cd tolerance by decreasing ROS through Ca elevation in Arabidopsis. *Plant Molecular Biology*, *105*(1–2), 115–132. https://doi.org/10.1007/s11103-020-01072-1

Modrzejewski, D., Hartung, F., Sprink, T., Krause, D., Kohl, C., & Wilhelm, R. (2019). What is the available evidence for the range of applications of genome-editing as a new tool for plant trait modification and the potential occurrence of associated off-target effects: a systematic map. *Environmental Evidence*, *8*(1), 1–33.

Mohtadi, A., Ghaderian, S. M., & Schat, H. (2012). Lead, zinc and cadmium accumulation from two metalliferous soils with contrasting calcium contents in heavy metal-hyperaccumulating and non-hyperaccumulating metallophytes: A comparative study. *Plant and Soil*, *361*(1–2), 109–118. https://doi.org/10.1007/s11104-012-1320-6

Molla, K. A. (2022). *Flowering time and photoperiod sensitivity in rice: Key players and their interactions identified*. Oxford University Press.

Monsant, A. C., Tang, C., & Baker, A. J. M. (2008). The effect of nitrogen form on rhizosphere soil pH and zinc phytoextraction by Thlaspi caerulescens. *Chemosphere*, *73*(5), 635–642.

Montanarella, L., Pennock, D., Mckenzie, N., Alavipanah, S. K., Alegre, J., Alshankiti, A., Arrouays, D., Aulakh, M. S., Badraoui, M., & Costa, I. dos S. B. (2015). *The Status of the World's Soil Resources (Technical Summary)*. Food and agriculture organization of the united nations.

Morel, M., Crouzet, J., Gravot, A., Auroy, P., Leonhardt, N., Vavasseur, A., & Richaud, P. (2009). AtHMA3, a P1B-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in Arabidopsis. *Plant Physiology*, *149*(2), 894–904. https://doi.org/10.1104/pp.108.130294

Mortel, J. E. Van De, Schat, H., Moerland, P. D., Themaat, E. V. L. Van, Ent, S. Van Der, Blankestijn, H., Ghandilyan, A., Tsiatsiani, S., & Aarts, M. G. M. (2008). Expression differences for genes involved in lignin, glutathione and sulphate metabolism in response to cadmium in Arabidopsis thaliana and the related Zn/Cd-hyperaccumulator Thlaspi caerulescens. *Plant, Cell and Environment*, *31*(3), 301–324. https://doi.org/10.1111/j.1365-3040.2007.01764.x

Mousset, M., David, P., Petit, C., Pouzadoux, J., Hatt, C., Flaven, É., Ronce, O., & Mignot, A. (2016). Lower selfing rates in metallicolous populations than in non-metallicolous populations of the pseudometallophyte Noccaea caerulescens (Brassicaceae) in Southern France. *Annals of Botany*, *117*(3), 507–519. https://doi.org/10.1093/aob/mcv191

Muller, G. (1969). Index of geoaccumulation in sediments of the Rhine River. *Geojournal*, *2*, 108–118.

Munguía-Rosas, M. A., Ollerton, J., Parra-Tabla, V., & De-Nova, J. A. (2011). Meta-analysis of phenotypic selection on flowering phenology suggests that early flowering plants are favoured. *Ecology Letters*, *14*(5), 511–521.

Murtas, G., Reeves, P. H., Fu, Y. F., Bancroft, I., Dean, C., & Coupland, G. (2003). A Nuclear Protease Required for Flowering-Time Regulation in Arabidopsis Reduces the Abundance of SMALL UBIQUITIN-RELATED MODIFIER Conjugates. *Plant Cell*, *15*(10), 2308–2319. https://doi.org/10.1105/tpc.015487

O'Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). *These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists*. Wiley Online Library.

O'Malley, R. C., Barragan, C. C., & Ecker, J. R. (2015). A user's guide to the Arabidopsis T-DNA insertion mutant collections. *Plant Functional Genomics: Methods and Protocols*, 323–342.

Oomen, R. J. F. J., Wu, J., Lelièvre, F., Blanchet, S., Richaud, P., Barbier-Brygoo, H., Aarts, M. G. M., & Thomine, S. (2009). Functional characterization of NRAMP3 and NRAMP4 from the metal hyperaccumulator Thlaspi caerulescens. *New Phytologist*, *181*(3), 637–650. https://doi.org/10.1111/j.1469-8137.2008.02694.x

Osakabe, Y., Watanabe, T., Sugano, S. S., Ueta, R., Ishihara, R., Shinozaki, K., & Osakabe, K. (2016). Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Scientific Reports*, *6*(February), 1–10. https://doi.org/10.1038/srep26685

Ouellette, L. A., Reid, R. W., Blanchard, S. G., & Brouwer, C. R. (2018). LinkageMapView—rendering high-resolution linkage and QTL maps. *Bioinformatics*, *34*(2), 306–307. https://doi.org/10.1093/bioinformatics/btx576

Palmgren, M. G., Clemens, S., Williams, L. E., Krämer, U., Borg, S., Schjørring, J. K., & Sanders, D. (2008). Zinc biofortification of cereals: problems and solutions. In *Trends in Plant Science* (Vol. 13, Issue 9, pp. 464–473). https://doi.org/10.1016/j.tplants.2008.06.005

Pan, J., Plant, J. A., Voulvoulis, N., Oates, C. J., & Ihlenfeld, C. (2010). Cadmium levels in Europe: implications for human health. In *Environmental geochemistry and health* (Vol. 32, Issue 1, pp. 1–12). https://doi.org/10.1007/s10653-009-9273-2

Pandey, V. C., & Bajpai, O. (2019). Phytoremediation: from theory toward practice. In *Phytomanagement of polluted sites* (pp. 1–49). Elsevier.

Papoyan, A., & Kochian, L. v. (2004). Identification of Thlaspi caerulescens genes that may be involved in heavy metal hyperaccumulation and tolerance. Characterization of a novel heavy metal transporting ATPase. *Plant Physiology*, *136*(3), 3814–3823. https://doi.org/10.1104/pp.104.044503

Parisod, C., & Besnard, G. (2007). Glacial in situ survival in the Western Alps and polytopic autopolyploidy in Biscutella laevigata L.(Brassicaceae). *Molecular Ecology*, *16*(13), 2755–2767.

Pauwels, M., Saumitou-Laprade, P., Holl, A. C., Petit, D., & Bonnin, I. (2005). Multiple origin of metallicolous populations of the pseudometallophyte Arabidopsis halleri (Brassicaceae) in central Europe: the cpDNA testimony. *Molecular Ecology*, *14*(14), 4403–4414. https://doi.org/https://doi.org/10.1111/j.1365-294X.2005.02739.x

Pauwels, M., Vekemans, X., Godé, C., Frérot, H., Castric, V., & Saumitou-Laprade, P. (2012). Nuclear and chloroplast DNA phylogeography reveals vicariance among European populations of the model species for the study of metal tolerance, Arabidopsis halleri (Brassicaceae). *New Phytologist*, *193*(4), 916–928. https://doi.org/10.1111/j.1469-8137.2011.04003.x

Peer, W. A., Mahmoudian, M., Freeman, J. L., Lahner, B., Richards, E. L., Reeves, R. D., Murphy, A. S., & Salt, D. E. (2006). Assessment of plants from the Brassicaceae family as genetic models for the study of nickel and zinc hyperaccumulation. *New Phytologist*, *172*(2), 248–260. https://doi.org/10.1111/j.1469-8137.2006.01820.x

Peer, W. A., Mamoudian, M., Lahner, B., Reeves, R. D., Murphy, A. S., & Salt, D. E. (2003). Identifying model metal hyperaccumulating plants: Germplasm analysis of 20 brassicaceae accessions from a wide geographical area. *New Phytologist*, *159*(2), 421–430. https://doi.org/10.1046/j.1469-8137.2003.00822.x

Pence, N. S., Larsen, P. B., Ebbs, S. D., Letham, D. L. D., Lasat, M. M., Garvin, D. F., Eide, D., & Kochian, L. v. (2000). *The molecular physiology of heavy metal transport in the ZnCd hyperaccumulator Thlaspi caerulescens*. www.pnas.org

Persans, M. W., Nieman, K., & Salt, D. E. (2001). Functional activity and role of cation-efflux family members in Ni hyperaccumulation in Thlaspi goesingense. *Proceedings of the National Academy of Sciences*, *98*(17), 9995–10000.

Petter, E., Ding, Y., Hou, K., Bhattacharya, A., Gusev, A., Zaitlen, N., & Pasaniuc, B. (2023). Genotype error due to low-coverage sequencing induces uncertainty in polygenic scoring. *The American Journal of Human Genetics*, *110*(8), 1319–1329.

Plaza, S., Tearall, K. L., Zhao, F. J., Buchner, P., McGrath, S. P., & Hawkesford, M. J. (2007). Expression and functional analysis of metal transporter genes in two contrasting ecotypes of the hyperaccumulator Thlaspi caerulescens. *Journal of Experimental Botany*, *58*(7), 1717–1728. https://doi.org/10.1093/jxb/erm025

Plessl, M., Rigola, D., Hassinen, V., Aarts, M. G. M., Schat, H., & Ernst, D. (2005). *Transcription Profiling of the Metal-hyperaccumulator Thlaspi caerulescens (J. & C. PRESL)*. *60*(3–4), 216–223. https://doi.org/doi:10.1515/znc-2005-3-406

Pollard, A. J., Reeves, R. D., & Baker, A. J. M. (2014). Facultative hyperaccumulation of heavy metals and metalloids. *Plant Science*, *217–218*, 8–17. https://doi.org/10.1016/j.plantsci.2013.11.011

Pongpanich, M., Sullivan, P. F., & Tzeng, J.-Y. (2010). A quality control algorithm for filtering SNPs in genome-wide association studies. *Bioinformatics*, *26*(14), 1731–1737.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Putterill, J., Laurie, R., & Macknight, R. (2004). It's time to flower: the genetic control of flowering time. *Bioessays*, *26*(4), 363–373.

Rajendran, S., Priya, T. A. K., Khoo, K. S., Hoang, T. K. A., Ng, H.-S., Munawaroh, H. S. H., Karaman, C., Orooji, Y., & Show, P. L. (2022). A critical review on various remediation approaches for heavy metal contaminants removal from contaminated soils. *Chemosphere*, *287*, 132369.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, *8*(11), 2281–2308.

Rascio, N., & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, *180*(2), 169–181. https://doi.org/10.1016/j.plantsci.2010.08.016

Rashid, A., Schutte, B. J., Ulery, A., Deyholos, M. K., Sanogo, S., Lehnhoff, E. A., & Beck, L. (2023). Heavy Metal Contamination in Agricultural Soil: Environmental Pollutants Affecting Crop Health. In *Agronomy* (Vol. 13, Issue 6). MDPI. https://doi.org/10.3390/agronomy13061521

Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., & Auvinen, P. (2013). Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics*, *29*(24), 3128–3134.

Rees, F., Sterckeman, T., Morel, J. L., Louis, J., & Biochar, M. (2020). Biochar-assisted phytoextraction of Cd and Zn by Noccaea caerulescens on a contaminated soil: A four-year lysimeter study. *Science of the Total Environment*, *707*. https://doi.org/10.1016/j.scitotenv.2019.135654ï

Reeves, P. H., Murtas, G., Dash, S., & Coupland, G. (2002). *early in short days 4, a mutation in Arabidopsis that causes early flowering and reduces the mRNA abundance of the floral repressor FLC*.

Reeves, R. D. (2006). Hyperaccumulation of trace elements by plants. In *Phytoremediation of metal-contaminated soils* (pp. 25–52). Springer.

Reeves, R. D., Baker, A. J. M., Borhidi, A., & Berazaín, R. (1999). Nickel Hyperaccumulation in the Serpentine Flora of Cuba. *Annals of Botany*, *83*(1), 29–38. https://doi.org/10.1006/anbo.1998.0786

Reeves, R. D., Baker, A. J. M., Jaffré, T., Erskine, P. D., Echevarria, G., & van der Ent, A. (2017). A global database for plants that hyperaccumulate metal and metalloid trace elements. *New Phytologist*. https://doi.org/10.1111/nph.14907

Reeves, R. D., & Brooks, R. R. (1983). European species of Thlaspi L. (Cruciferae) as indicators of nickel and zinc. *Journal of Geochemical Exploration*, *18*(3), 275–283. https://doi.org/https://doi.org/10.1016/0375-6742(83)90073-0

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research*. Evanston, IL.

Richau, K. H., & Schat, H. (2009). Intraspecific variation of nickel and zinc accumulation and tolerance in the hyperaccumulator Thlaspi caerulescens. *Plant and Soil*, *314*(1–2), 253–262. https://doi.org/10.1007/s11104-008-9724-z

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). *Transgressive segregation, adaptation and speciation*.

Rigola, D., Fiers, M., Vurro, E., & Aarts, M. G. M. (2006). The heavy metal hyperaccumulator Thlaspi caerulescens expresses many species-specific genes, as identified by comparative expressed sequence tag analysis. *New Phytologist*, *170*(4), 753–766. https://doi.org/10.1111/j.1469-8137.2006.01714.x

Roberts, T. L. (2014). Cadmium and phosphorous fertilizers: the issues and the science. *Procedia Engineering*, *83*, 52–59.

Roosens, N. H., Bernard, C., Leplae, R., & Verbruggen, N. (2005). *Adaptative Evolution of Metallothionein 3 in the Cd/Zn Hyperaccumulator Thlaspi caerulescens*. *60*(3–4), 224–228. https://doi.org/doi:10.1515/znc-2005-3-407

Roosens, N. H. C. J., Willems, G., & Saumitou-Laprade, P. (2008). Using Arabidopsis to explore zinc tolerance and hyperaccumulation. *Trends in Plant Science*, *13*(5), 208–215. https://doi.org/10.1016/j.tplants.2008.02.006

Roosens, N. H., Leplae, R., Bernard, C., & Verbruggen, N. (2005). Variations in plant metallothioneins: The heavy metal hyperaccumulator Thlaspi caerulescens as a study case. *Planta*, *222*(4), 716–729. https://doi.org/10.1007/s00425-005-0006-1

Russell, J. J., Theriot, J. A., Sood, P., Marshall, W. F., Landweber, L. F., Fritz-Laylin, L., Polka, J. K., Oliferenko, S., Gerbich, T., Gladfelter, A., Umen, J., Bezanilla, M., Lancaster, M. A., He, S.,

Gibson, M. C., Goldstein, B., Tanaka, E. M., Hu, C. K., & Brunet, A. (2017). Non-model model organisms. *BMC Biology*, *15*(1). https://doi.org/10.1186/s12915-017-0391-5

Sarwar, N., Imran, M., Shaheen, M. R., Ishaque, W., Kamran, M. A., Matloob, A., Rehim, A., & Hussain, S. (2017). Phytoremediation strategies for soils contaminated with heavy metals: modifications and future perspectives. *Chemosphere*, *171*, 710–721.

Sarwar, N., Saifullah, Malhi, S. S., Zia, M. H., Naeem, A., Bibi, S., & Farid, G. (2010). Role of mineral nutrition in minimizing cadmium accumulation by plants. *Journal of the Science of Food and Agriculture*, *90*(6), 925–937.

Sasaki, A., Yamaji, N., & Ma, J. F. (2014). Overexpression of OsHMA3 enhances Cd tolerance and expression of Zn transporter genes in rice. *Journal of Experimental Botany*, *65*(20), 6013–6021.

Sasaki, A., Yamaji, N., Yokosho, K., & Ma, J. F. (2012). Nramp5 is a major transporter responsible for manganese and cadmium uptake in rice. *Plant Cell*, *24*(5), 2155–2167. https://doi.org/10.1105/tpc.112.096925

Sattler, M. C., Carvalho, C. R., & Clarindo, W. R. (2015). The polyploidy and its key role in plant breeding. *Planta*, *243*(2), 281–296. https://doi.org/10.1007/s00425-015-2450-x

Schat, H., & Aarts, M. G. M. (2003). Model Species To Study Heavy Metal Hyperaccumulation in Plants. *New Phytologist*, *1992*, 351–360. https://doi.org/10.1046/j.1469-

Schat, H., & ten Bookum, W. M. (1992). Genetic control of copper tolerance in silene vulgaris. *Heredity*, *68*(3), 219–229. https://doi.org/10.1038/hdy.1992.35

Schneider, T., Persson, D. P., Husted, S., Schellenberg, M., Gehrig, P., Lee, Y., Martinoia, E., Schjoerring, J. K., & Meyer, S. (2013). A proteomics approach to investigate the process of Z n hyperaccumulation in N occaea caerulescens (J & C. P resl) FK M eyer. *The Plant Journal*, *73*(1), 131–142.

Scott, W. E., Hoblitt, R. P., Torres, R. C., Self, S., Martinez, M. M. L., & Nillos, T. (1996). Pyroclastic flows of the June 15, 1991, climactic eruption of Mount Pinatubo. *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*, 545–570.

Scutaraşu, E. C., & Trincă, L. C. (2023). Heavy Metals in Foods and Beverages: Global Situation, Health Risks and Reduction Methods. *Foods*, *12*(18), 3340. https://doi.org/10.3390/foods12183340

Seregin, I. V, & Kozhevnikova, A. D. (2023). Nicotianamine: a key player in metal homeostasis and hyperaccumulation in plants. *International Journal of Molecular Sciences*, *24*(13), 10822.

Sherrard, M. E., & Maherali, H. (2006). The adaptive significance of drought escape in Avena barbata, an annual grass. *Evolution*, *60*(12), 2478–2489.

Shields, D. C., Collins, A., Buetow, K. H., & Morton, N. E. (1991). Error filtration, interference, and the human linkage map. *Proceedings of the National Academy of Sciences*, *88*(15), 6501–6505.

Shimada, T. L., Shimada, T., & Hara-Nishimura, I. (2010). A rapid and non-destructive screenable marker, FAST, for identifying transformed seeds of Arabidopsis thaliana: TECHNICAL ADVANCE. In *Plant Journal* (Vol. 61, Issue 3, pp. 519–528). https://doi.org/10.1111/j.1365-313X.2009.04060.x

Shimwell, D. W., & Laurie, A. E. (1972). Lead and zinc contamination of vegetation in the southern Pennines. *Environmental Pollution (1970)*, *3*(4), 291–301. https://doi.org/https://doi.org/10.1016/0013-9327(72)90024-9

Shindo, C., Aranzana, M. J., Lister, C., Baxter, C., Nicholls, C., Nordborg, M., & Dean, C. (2005). Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of Arabidopsis. *Plant Physiology*, *138*(2), 1163–1173.

Simmons, R. W., Chaney, R. L., Angle, J. S., Kruatrachue, M., Klinphoklap, S., Reeves, R. D., & Bellamy, P. (2015). Towards Practical Cadmium Phytoextraction with N occaea Caerulescens. *International Journal of Phytoremediation*, *17*(2), 191–199.

Simpson, G. G., & Dean, C. (2002). Arabidopsis, the Rosetta Stone of Flowering Time? *Science*, *296*(5566), 285–289. https://doi.org/10.1126/science.296.5566.285

Sirguey, C., Schwartz, C., & Morel, J. L. (2006). Response of Thlaspi caerulescens to nitrogen, phosphorus and sulfur fertilisation. *International Journal of Phytoremediation*, *8*(2), 149.

Skuza, L., Szućko-Kociuba, I., Filip, E., & Bożek, I. (2022). Natural molecular mechanisms of plant hyperaccumulation and hypertolerance towards heavy metals. *International Journal of Molecular Sciences*, *23*(16), 9335.

Son, G. H., Park, B. S., Song, J. T., & Seo, H. S. (2014). FLC-mediated flowering repression is positively regulated by sumoylation. *Journal of Experimental Botany*, *65*(1), 339–351.

Sriprachote, A., Kanyawongha, P., Ochiai, K., & Matoh, T. (2012). Current situation of cadmium-polluted paddy soil, rice and soybean in the mae sot district, tak province, thailand. *Soil Science and Plant Nutrition*, *58*(3), 349–359. https://doi.org/10.1080/00380768.2012.686435

Srivastava, S., Pathak, S., Ponsin, M., Hensawang, S., Chanpiwat, P., Yoeurn, C., & Phan, K. (2022). Sustainable solutions to arsenic accumulation in rice grown in south and south-east Asia. In *Crop and Pasture Science* (Vol. 73, Issue 2, pp. 149–159). CSIRO. https://doi.org/10.1071/CP21033

Stein, R. J., Höreth, S., de Melo, J. R. F., Syllwasschy, L., Lee, G., Garbin, M. L., Clemens, S., & Krämer, U. (2017). Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model Arabidopsis halleri. *New Phytologist*, *213*(3), 1274–1286.

Sterckeman, T., Cazes, Y., Gonneau, C., & Sirguey, C. (2017). Phenotyping 60 populations of Noccaea caerulescens provides a broader knowledge of variation in traits of interest for phytoextraction. *Plant and Soil*, *418*(1–2), 523–540. https://doi.org/10.1007/s11104-017-3311-0

Sterckeman, T., & Thomine, S. (2020). Mechanisms of Cadmium Accumulation in Plants. *Critical Reviews in Plant Sciences*, 10. https://doi.org/10.1080/07352689.2020.1792179ï

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Halldorsdottir, S. S., Purugganan, M. D., & Schmitt, J. (2004). A latitudinal cline in flowering time in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences*, *101*(13), 4712–4717.

Su, C. (2014). A review on heavy metal contamination in the soil worldwide: Situation, impact and remediation techniques. *Environmental Skeptics and Critics*, *3*(2), 24.

Sunar, S., Yildirim, N., Sengul, M., & Agar, G. (2016). Genetic diversity and relationships detected by ISSR and RAPD analysis among Aethionema species growing in Eastern Anatolia (Turkey). *Comptes Rendus Biologies*, *339*(3–4), 147–151.

Takahashi, R., Ishimaru, Y., Senoura, T., Shimo, H., Ishikawa, S., Arao, T., Nakanishi, H., & Nishizawa, N. K. (2011). The OsNRAMP1 iron transporter is involved in Cd accumulation in rice. *Journal of Experimental Botany*, *62*(14), 4843–4850. https://doi.org/10.1093/jxb/err136

Tang, L., Dong, J., Qu, M., Lv, Q., Zhang, L., Peng, C., Hu, Y., Li, Y., Ji, Z., Mao, B., Peng, Y., Shao, Y., & Zhao, B. (2022). Knockout of OsNRAMP5 enhances rice tolerance to cadmium toxicity in response to varying external cadmium concentrations via distinct mechanisms. *Science of the Total Environment*, *832*. https://doi.org/10.1016/j.scitotenv.2022.155006

Team, R. C. (2008). Computational Many-Particle Physics. In *R Foundation for Statistical Computing* (Vol. 739). https://doi.org/10.1007/978-3-540-74686-7

Tehseen, M., Cairns, N., Sherson, S., & Cobbett, C. S. (2010). Metallochaperone-like genes in Arabidopsis thaliana. *Metallomics*, *2*(8), 556–564. https://doi.org/10.1039/c003484c

Tóth, G., Hermann, T., Szatmári, G., & Pásztor, L. (2016). Maps of heavy metals in the soils of the European Union and proposed priority areas for detailed assessment. *Science of the Total Environment*, *565*, 1054–1062. https://doi.org/10.1016/j.scitotenv.2016.05.115

Tsai, C.-P., & Lee, C. T.-C. (2013). Multiple sclerosis incidence associated with the soil lead and arsenic concentrations in Taiwan. *PloS One*, *8*(6), e65911.

Tuomainen, M., Tervahauta, A., Hassinen, V., Schat, H., Koistinen, K. M., Lehesranta, S., Rantalainen, K., Häyrinen, J., Auriola, S., & Anttonen, M. (2010). Proteomics of Thlaspi caerulescens accessions and an inter-accession cross segregating for zinc accumulation. *Journal of Experimental Botany*, *61*(4), 1075–1087.

Ueno, D., Milner, M. J., Yamaji, N., Yokosho, K., Koyama, E., Clemencia Zambrano, M., Kaskie, M., Ebbs, S., Kochian, L. v., & Ma, J. F. (2011). Elevated expression of TcHMA3 plays a key role in the extreme Cd tolerance in a Cd-hyperaccumulating ecotype of Thlaspi caerulescens. *Plant Journal*, *66*(5), 852–862. https://doi.org/10.1111/j.1365-313X.2011.04548.x

Ueno, D., Yamaji, N., Kono, I., Huang, C. F., Ando, T., Yano, M., & Ma, J. F. (2010). Gene limiting cadmium accumulation in rice. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(38), 16500–16505. https://doi.org/10.1073/pnas.1005396107

van de Mortel, J. E., Villanueva, L. A., Schat, H., Kwekkeboom, J., Coughlan, S., Moerland, P. D., van Themaat, E. V. L., Koornneef, M., & Aarts, M. G. M. (2006). Large Expression Differences in Genes for Iron and Zinc Homeostasis, Stress Response, and Lignin Biosynthesis Distinguish Roots of Arabidopsis thaliana and the Related Metal Hyperaccumulator Thlaspi caerulescens. *Plant Physiology*, *142*(3), 1127–1147. https://doi.org/10.1104/pp.106.082073

van der Ent, A., Baker, A. J. M., Reeves, R. D., Pollard, A. J., & Schat, H. (2013). Hyperaccumulators of metal and metalloid trace elements: Facts and fiction. *Plant and Soil*, *362*(1–2), 319–334. https://doi.org/10.1007/s11104-012-1287-3

van der Ent, A., & Mulligan, D. (2015). Multi-element Concentrations in Plant Parts and Fluids of Malaysian Nickel Hyperaccumulator Plants and some Economic and Ecological Considerations. *Journal of Chemical Ecology*, *41*(4), 396–408. https://doi.org/10.1007/s10886-015-0573-y

van der Ent, A., Pillon, Y., Fogliani, B., Gei, V., Jaffré, T., Erskine, P. D., Echevarria, G., Spiers, K. M., Paul, A. L. D., & Isnard, S. (2022). Contrasting nickel and manganese accumulation and localization in New Caledonian Cunoniaceae. *Plant and Soil*, *475*(1–2), 515–534. https://doi.org/10.1007/s11104-022-05388-3

van der Zee, L., Corzo Remigio, A., Casey, L. W., Purwadi, I., Yamjabok, J., van der Ent, A., Kootstra, G., & Aarts, M. G. M. (2021). Quantification of spatial metal accumulation patterns in Noccaea caerulescens by X-ray fluorescence image processing for genetic studies. *Plant Methods*, *17*(1). https://doi.org/10.1186/s13007-021-00784-9

Van Ooijen, J. W. (2006). JoinMap 4. *Software for the Calculation of Genetic Linkage Maps in Experimental Populations. Kyazma BV, Wageningen, Netherlands*, *33*.

Van Os, H., Stam, P., Visser, R. G. F., & van Eck, H. J. (2005). SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. *Theoretical and Applied Genetics*, *112*, 187–194.

Verbruggen, N., Hermans, C., & Schat, H. (2009). Molecular mechanisms of metal hyperaccumulation in plants. *New Phytologist*, *181*(4), 759–776. https://doi.org/10.1111/j.1469-8137.2008.02748.x

Verret, F., Gravot, A., Auroy, P., Leonhardt, N., David, P., Nussaume, L., Vavasseur, A., & Richaud, P. (2004). Overexpression of AtHMA4 enhances root-to-shoot translocation of zinc and cadmium and plant metal tolerance. *FEBS Letters*, *576*(3), 306–312. https://doi.org/10.1016/j.febslet.2004.09.023

Villajuana-Bonequi, M., Elrouby, N., Nordström, K., Griebel, T., Bachmair, A., & Coupland, G. (2014). Elevated salicylic acid levels conferred by increased expression of ISOCHORISMATE SYNTHASE 1 contribute to hyperaccumulation of SUMO 1 conjugates in the Arabidopsis mutant early in short days 4. *The Plant Journal*, *79*(2), 206–219.

Villanueva, R. A. M., & Chen, Z. J. (2019). *ggplot2: elegant graphics for data analysis*. Taylor & Francis.

Voorrips, R. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity*, *93*(1), 77–78.

Wan, X., Lei, M., & Chen, T. (2016). Cost–benefit calculation of phytoremediation technology for heavy-metal-contaminated soil. *Science of the Total Environment*, *563*, 796–802.

Wang, L., Rinklebe, J., Tack, F. M. G., & Hou, D. (2021). A review of green remediation strategies for heavy metal contaminated soil. In *Soil Use and Management* (Vol. 37, Issue 4, pp. 936–963). John Wiley and Sons Inc. https://doi.org/10.1111/sum.12717

Wang, Y. (2016). *An evolutionary and functional genomics study of Noccaea caerulescens, a heavy metal hyperaccumulating plant species*. http://edepot.wur.nl/385510

Wang, Y., Salt, D. E., Koornneef, M., & Aarts, M. G. M. (2022). Construction and analysis of a Noccaea caerulescens TILLING population. *BMC Plant Biology*, *22*(1). https://doi.org/10.1186/s12870-022-03739-x

Wang, Y., Severing, E. I., Koornneef, M., & Aarts, M. G. M. (2020). FLC and SVP Are Key Regulators of Flowering Time in the Biennial/Perennial Species Noccaea caerulescens. *Frontiers in Plant Science*, *11*(November), 1–14. https://doi.org/10.3389/fpls.2020.582577

Werner, J. D., Borevitz, J. O., Uhlenhaut, N. H., Ecker, J. R., Chory, J., & Weigel, D. (2005). FRIGIDA-independent variation in flowering time of natural Arabidopsis thaliana accessions. *Genetics*, *170*(3), 1197–1207.

Wilschefski, S. C., & Baxter, M. R. (2019). Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects. *The Clinical Biochemist. Reviews*, *40*(3), 115–133. https://doi.org/10.33176/AACB-19-00024

Wójcik, M., Gonnelli, C., Selvi, F., Dresler, S., Rostański, A., & Vangronsveld, J. (2017). Metallophytes of Serpentine and Calamine Soils – Their Unique Ecophysiology and Potential for Phytoremediation. *Advances in Botanical Research*, *83*, 1–42. https://doi.org/10.1016/bs.abr.2016.12.002

Wong, C. K. E., & Cobbett, C. S. (2009). HMA P-type ATPases are the major mechanism for root-to-shoot Cd translocation in Arabidopsis thaliana. *New Phytologist*, *181*(1), 71–78. https://doi.org/10.1111/j.1469-8137.2008.02638.x

Wong, C. K. E., Jarvis, R. S., Sherson, S. M., & Cobbett, C. S. (2009). Functional analysis of the heavy metal binding domains of the Zn/Cd-transporting ATPase, HMA2, in Arabidopsis thaliana. *New Phytologist*, *181*(1), 79–88. https://doi.org/10.1111/j.1469-8137.2008.02637.x

Wuana, R. A., & Okieimen, F. E. (2011). Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. *ISRN Ecology*, *2011*, 1–20. https://doi.org/10.5402/2011/402647

Xie, H. L., Jiang, R. F., Zhang, F. S., McGrath, S. P., & Zhao, F. J. (2009). Effect of nitrogen form on the rhizosphere dynamics and uptake of cadmium and zinc by the hyperaccumulator Thlaspi caerulescens. *Plant and Soil*, *318*, 205–215.

Xiong, T., Dumat, C., Pierart, A., Shahid, M., Kang, Y., Li, N., Bertoni, G., & Laplanche, C. (2016). Measurement of metal bioaccessibility in vegetables to improve human exposure assessments: field study of soil–plant–atmosphere transfers in urban areas, South China. *Environmental Geochemistry and Health*, *38*, 1283–1301.

Xu, D., Li, X., Wu, X., Meng, L., Zou, Z., Bao, E., Bian, Z., & Cao, K. (2021). Tomato SlCDF3 delays flowering time by regulating different FT-like genes under long-day and short-day conditions. *Frontiers in Plant Science*, *12*, 650068.

Xu, W., Xiang, P., Liu, X., & Ma, L. Q. (2020). Closely-related species of hyperaccumulating plants and their ability in accumulation of As, Cd, Cu, Mn, Ni, Pb and Zn. *Chemosphere*, *251*. https://doi.org/10.1016/j.chemosphere.2020.126334

Yan, A., Wang, Y., Tan, S. N., Mohd Yusof, M. L., Ghosh, S., & Chen, Z. (2020). Phytoremediation: a promising approach for revegetation of heavy metal-polluted land. *Frontiers in Plant Science*, *11*, 359.

Yang, X. E., Li, T. Q., Long, X. X., Xiong, Y. H., He, Z. L., & Stoffella, P. J. (2006). Dynamics of zinc uptake and accumulation in the hyperaccumulating and non-hyperaccumulating ecotypes of Sedum alfredii Hance. *Plant and Soil*, *284*(1), 109–119. https://doi.org/10.1007/s11104-006-0033-0

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Yung, L., Sirguey, C., Azou-Barré, A., & Blaudez, D. (2021). Natural fungal endophytes from Noccaea caerulescens mediate neutral to positive effects on plant biomass, mineral nutrition and Zn Phytoextraction. *Frontiers in Microbiology*, *12*, 689367.

Zha, H. G., Jiang, R. F., Zhao, F. J., Vooijs, R., Schat, H., Barker, J. H. A., & McGrath, S. P. (2004). Co-segregation analysis of cadmium and zinc accumulation in Thlaspi caerulescens interecotypic crosses. *New Phytologist*, *163*(2), 299–312. https://doi.org/10.1111/j.1469-8137.2004.01113.x

Zhang, H., Zhang, X., Liu, J., Niu, Y., Chen, Y., Hao, Y., Zhao, J., Sun, L., Wang, H., & Xiao, J. (2020). Characterization of the heavy-metal-associated isoprenylated plant protein (HIPP) gene family from Triticeae species. *International Journal of Molecular Sciences*, *21*(17), 6191.

Zhang, P., Wang, R., Ju, Q., Li, W., Tran, L.-S. P., & Xu, J. (2019). The R2R3-MYB transcription factor MYB49 regulates cadmium accumulation. *Plant Physiology*, *180*(1), 529–542.

Zhao, F.-J., & McGrath, S. P. (2009). Biofortification and phytoremediation. *Current Opinion in Plant Biology*, *12*(3), 373–380.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. https://doi.org/10.1038/ng.2310

English summary

Genetic analysis of metal hyperaccumulation and hypertolerance of *Noccaea caerulescens*

Noccaea caerulescens is a plant species known for its remarkable adaptability to a wide range of cadmium (Cd), lead(Pb), nickel (Ni) and zinc (Zn) exposures. This species exhibits three distinct ecotypes — calamine, ultramafic, and non-metallicolous — corresponding to the soil types of their respective origins. *N. caerulescens* is widely distributed across Europe. Numerous studies have delved into the investigation of these prominent adaptive traits, aiming to unravel the complex mechanisms employed by the plant to cope with metal stress. The insights gained from such research hold potential applications in phytoremediation, phytomining, and biofortification. While previous studies have identified a few loci and several genes associated with metal accumulation and tolerance in *N. caerulescens*, these investigations were often limited to a few accessions and bi-parental F_2/F_3 populations, restricting the exploration of genetic variations present in other populations and accessions. Additionally, the absence of reverse genetic tools impeded the validation of these identified candidate genes in this species. The work in this thesis describes the development of new mapping populations of *N. caerulescens* for exploring the molecular basis of metal hyperaccumulation and hypertolerance using genome-wide genetic variation, and how these traits evolve and reshape under influence of ecological and evolutionary forces. A new method for successful genetic transformation of this species was also described in this thesis. Furthermore, the well-documented trait, flowering time, is investigated along with metal hyperaccumulation and hypertolerance as proof-of-concept for validation of genetic tools generated by the research described in this thesis.

Chapter 2 describes the generation of an F_7 recombinant inbred line (RIL) population for *N. caerulescens*, from the cross between accessions Ganges (GA), a calamine accession displaying high Zn/Cd accumulation/tolerance, and Lellingen (LE), a non-metallicolous accession displaying high Zn/Cd accumulation and moderate Zn/Cd tolerance. This population was genotyped for single nucleotide polymorphisms (SNPs) and used for genetic map construction. Six life history traits were measured and mapped to identify quantitative trait loci (QTLs). Seven QTLs were identified, including those associated with flowering time (qFT2019 and qFT2020-1), exhibiting strong Logarithm of Odds (LOD) scores of 7.18 and 8.40. These QTLs explained a significant portion of the phenotypic variance. Within the colocalized QTLs, an orthologue of an *Arabidopsis thaliana* gene involved in flowering time was identified, supporting the feasibility of this RIL population for QTL mapping and further analysis of metal-related traits in *N. caerulescens*.

Chapter 3 employs such QTL analysis to dissect the genetic basis underlying the response to Cd exposure in *N. caerulescens* using the LE x GA RIL population. This RIL population was assessed for 158 traits across three experiments: a Cd accumulation test (1 µM Cd), a Cd tolerance test (Cd series from 50 to 350 µM) and a growth test in response to Cd (15 µM Cd). The results reveal a total of 27 significant QTLs, explaining 5.75% to 31.24% of the phenotypic variance, depending on the trait. Colocalization of QTLs was observed, particularly for Cd translocation factor (qTF-Cd) and Cd root concentration (qRT-Cd-1), displaying a strong LOD score of 16.3 over a small region on linkage group 2. Within this region, an orthologue of *HEAVY METAL ATPASE 3* (*HMA3*), a well-known Cd transporter, was identified as a candidate gene for Cd accumulation in the root of *N. caerulescens*. Additionally, orthologues within each QTL were described. These findings will provide valuable resources for the discovery of novel genes contributing to Cd hyperaccumulation and hypertolerance in *N. caerulescens*.

Chapter 4 provides an account of the development of a diversity panel for *N. caerulescens*, comprising 109 accessions sourced from a large geographic area in Europe. Whole-genome genetic variation was assessed using next-generation sequencing, resulting in 733,912 filtered single nucleotide polymorphisms (SNPs). These filtered SNPs were then employed to investigate the population diversity and structure of *N. caerulescens*. The analysis revealed six genetically distinct subgroups, shedding light on the relationship between accessions from the Iberian mountains in Spain and the Apennines in Italy, and their connection to accessions from western Europe. Human activities were hypothetically implicated in the distribution of *N. caerulescens* from western Europe into the British and Scandinavian regions. Furthermore, this study conducted the first genome-wide association study (GWAS) in *N. caerulescens*, targeting flowering time and Zn/Ni accumulation. A total of 70 significantly associated SNPs with the traits of interest were identified. Among these, a significant SNP associated with flowering time (Contig16:117310) was found to be closely located to an orthologue of an *Arabidopsis thaliana* flowering time gene. This outcome underscores the potential utility of the developed diversity panel for GWAS analysis, although optimization is recommended to enhance statistical power. The diversity panel is anticipated to serve as a valuable resource for future development of genetic tools and for unravelling the signature of natural selection acting upon genes contributing to the variation of traits related to metal hyperaccumulation and hypertolerance in plants.

Chapter 5 describes a versatile transformation system for *N. caerulescens*, which involves the integration of early flowering plants with high fecundity and a straightforward *Agrobacterium tumefaciens* floral dipping transformation protocol. An identified mutation in *FLOWERING LOCUS C* from a St. Felix de Pallières (SF) mutant, which exhibits early flowering traits, was introgressed into five *N. caerulescens* accessions with high fecundity: Clough Wood (CLW), Le Bleymard (BLE), Cira (CIR), St. Baudille (SBD), and Werschmatt (WER). Of these, three backgrounds, CLW, BLE, and SBD, were successfully transformed with an average transformation efficiency of 0.29%. This new method serves as a versatile tool for functional analysis and contributes to the development of genome editing protocols in *N. caerulescens*.

Chapter 6 discusses the main findings and the future perspectives for *N. caerulescens* research.

Acknowledgements

Rome was not built in one day, and neither was my PhD thesis. After eight years of ups and downs, I have finally completed this challenging yet fulfilling journey. I am deeply grateful to all those who have accompanied me along the way, particularly the Royal Thai government and Wageningen University for generously sponsoring me financially throughout the entire PhD process. Additionally, I extend my heartfelt appreciation to the Laboratory of Genetics at Wageningen University for providing me with a multi-disciplinary, multi-cultural, supportive, and inspiring working environment.

I would like to express my deepest gratitude to my promotor and daily supervisor, Mark Aarts. Thank you for your unwavering guidance, empathy, straightforward suggestions, motivation, creative experimental ideas, and immense contribution to this thesis. I first encountered Mark during the plant transformation practical of the plant biotechnology course in my master's degree. Many students found Mark intimidating because he often responded to questions with further questions. Fortunately, having been trained similarly during my bachelor's, I found his active learning teaching style to be very effective and not as daunting. This positive experience led me to choose Mark's research group at the Laboratory of Genetics for my major master's thesis. During my master's thesis, I vividly remember rehearsing my final presentation for Mark, only to receive the feedback, "your project is lacking." Initially shocked, I quickly realized the value in his direct feedback as he explained the flaws in my project and provided constructive criticism on how to improve my experimental design and data visualization. This experience introduced me to the direct Dutch working culture, which I found incredibly useful and efficient. It inspired me to pursue my PhD thesis under your guidance. Although the first year of my PhD journey was challenging, marked by a series of failed experiments due to my attempts at independence, you patiently guided me and provided unwavering support. You once told me that if we thoroughly discussed experimental design and an experiment failed, it was not solely my failure but one we both shared equally. This encouraged me to persevere and continue conducting experiments based on our agreed-upon designs, which ultimately culminated in the completion of this PhD thesis. Beyond genetics, I deeply appreciate your extensive knowledge in other fields, particularly history. I thoroughly enjoyed listening to your insights on Dutch and European history during our road trip to attend a science conference in Lille, France. I am immensely grateful for this once-in-a-lifetime opportunity and for all that I have learned under your guidance. I am committed to utilizing this knowledge to produce research outcomes that will benefit my country and, hopefully, mankind.

I would like to extend my heartfelt gratitude to Joost van den Heuvel, my co-supervisor, for your invaluable suggestions and significant contribution to this thesis, particularly in the areas of bioinformatic analysis and population genetics. The SNP dataset used in this thesis would not have been possible without your expertise and the SNP calling pipeline that you developed. Despite bioinformatics not being my strong suit, you patiently and methodically explained concepts until I fully grasped them. I am especially grateful for your recommendation regarding Chapter 4 of my thesis. I spent over five months writing this chapter, continuously adding various analyses and figures until it exceeded 20 pages, yet I was only halfway through the discussion. However, after discussing the chapter with you, you advised removing non-relevant analyses and figures, which enabled me to streamline the chapter and complete it concisely within a month. Furthermore, I want to express my appreciation for your guidance on negotiation skills, which has proven to be incredibly valuable for me. Your advice in this regard has greatly enhanced my ability to navigate professional interactions effectively. Thank you for your unwavering support and invaluable contributions throughout this journey.

I wish to extend my heartfelt gratitude to Henk Schat, my external supervisor, for your unparalleled expertise in metal hyperaccumulation and hypertolerance in plants, as well as your invaluable contributions to experimental design. I have always been fascinated by your meticulous approach to statistical analysis, which you prefer to conduct manually by hand rather than relying on commercial software. Your insights and recommendations on metal exposure levels have been remarkably accurate, reflecting your extensive experience and deep understanding of the subject matter. Throughout my PhD journey, your wealth of knowledge and experience in the field of metal hyperaccumulators has been truly impressive. I am particularly grateful for your metal decontamination protocol, which has proven to be

exceptionally effective. Thanks to your protocol, I have not encountered any visible metal toxicity responses thus far, which has significantly contributed to the success of my research. Working with you during my PhD has been a tremendous learning experience, and I am grateful for the opportunity to benefit from your expertise and guidance. Thank you for your unwavering support and invaluable contributions to my academic and professional development.

I would like to show my thanks to the students who have contributed to the pilot experiments and research described in this thesis. Working with each of you—Cindy Laine, Meta van Ruijven, Meti Dhiba, Wouter Isarel, Siqi Duan, Lies van der Heijden, and Laurens van Ostroom—has been an enriching and rewarding experience. Your dedication, enthusiasm, and hard work have played a pivotal role in the progress and success of my research. Without your contributions, this thesis would have taken significantly longer to complete, and the successful transformation of N. caerulescens described in this thesis would not be possible.

I would like to extend my heartfelt gratitude to the three musketeers—René Boesten (the name maker), Roel van Bezouw (the joke creator), and Tom Theeuwen (the fashion leader)—who not only shared a working space with me but also countless beers and laughs during my PhD journey. Your support, especially in experimental design, statistics, and association analysis, has been invaluable to me. Your encouragement during moments of experiment failure kept me motivated until the completion of this thesis. I am deeply appreciative of René's creativity in coming up with memorable aliases like Panunu and Chipspanu, which made me easily recognizable and unforgettable. Roel, your humor and jokes not only brightened my days but also helped me better understand Western culture in an enjoyable way. And to Tom, thank you for inspiring me with your office costume and for sharing various R-scripts for SNP calling and heritability calculation. If Tom is reading this, he should know that my PhD dressing style was basically adopted after his. Without the three of you, I cannot imagine how different my PhD experience would have been, and I would never have discovered the wonders of Leffe blonde beer. I also want to express my gratitude to Ben Auxier for your assistance with SNP calling and Linux command line. Your expertise and quick solutions to the errors I encountered were immensely helpful. I would like to thank all my lab fellows: Alan, Alanna, Alex, Alexander, Anna, Bwalya, Cris, Erik, Eveline, Fatemeh, Francesco, Helena, Hylke, Jelle, Jiao, Kim, Krithi, Laavanya, Lennart, Louise, Majid, Mariana, Mariska, Mathijs, Max, Murambiwa, Nasim, Phuong, Qiong, Sabine, Sarah, Siva, and Suzette. Your assistance and camaraderie have made my time in the lab unforgettable. I would also like to express my sincere appreciation to Vrijmibo for transforming every Friday evening into a cherished and memorable occasion. The memories created during our Vrijmibo gatherings will stay with me forever, as they say, "What happens at Vrijmibo, stays at Vrijmibo."

The work described in this thesis would not have been possible without the invaluable technical support from the officers and lab technicians at the Laboratory of Genetics. I am deeply grateful to Wytske Nijenhuis and Marjan Kampinga for their assistance with official formalities. Special thanks to Corrie Hanhart for her expertise in seed storage, propagation, harvesting, tissue culture, and greenhouse management. I also extend my gratitude to Frank Becker for his assistance with whole-genome sequencing of N. caerulescens, to Jose van de Belt for his guidance on fluorescent microscopy, and to Francisca Reyes Marquez for providing the plasmid and valuable suggestions on genetic transformation. I would like to express my sincere appreciation to Bas Zwaan and all the staffs at the Laboratory of Genetics for their support, engaging academic discussions, and the camaraderie fostered through enjoyable team-building events and heartfelt festive gatherings. Your contributions have greatly enriched my research journey and made it a truly memorable experience.

I would like to extend my heartfelt gratitude to my Thai friends at Wageningen University for their companionship and support throughout my master's and PhD journey: Anupol (Jae, the great dancer and skillful chef), Arpavee (Tete), Bhakavadee (Plubplueng), Chakapon (P' Pop), Chalermpat (P' Aun+), Chaniga (P' Gahn), Chanoknun (Tangmo), Chatpavee (Suedow), Eakapat (Maithoe, the Le Cordon Bleu chef), Jidapa (P' Jugk), Kornphimol (P' Tom), Kornrawee (Wee), Kunalai (P' Aum), Marisa (Meen), Nalin (May), Napassorn (Jane, the basketball sister), Napat (Belle), Napon (Poon), Natapol (P' Aof), Natnicha (Natty), Nidhiprabha (Ming), Nutthaya (Peach), Patteela (Polly, the professional badminton trainer and

supermodel), Pichaya (Jaae), Pintip (Pin), Pornpipat (Paul), Pussamon (Ploy, the photographer and excellent chef), Rungnapa (P' Rung, the Thai traditional chef), Sasiwimon (P' Som), Sirapak (Kung, the photogenic Thai model), Songyos (Sun), Sureerat (Kaew), Susakul (P' Good), Tanin (P' Boat), Thanaporn (P' Nong), Wannida (Namtan), Waroonsiri (Tan, the traveler), Wasin (Lumluk, the boardgame master), and Yosapol (Ping, the academic networker). I am truly grateful for your assistance, conversations, and delicious dishes during our Thai monthly gatherings. Your friendship has brought color and joy to my life, even in the flatlands of Wageningen. I want to give a special thanks to P' Pop for your invaluable advice on navigating through my life's challenges. Your patience in listening to my problems and offering guidance has been immensely appreciated, no matter how daunting those challenges seemed.

I would like to convey my sincere gratitude to my teachers, lecturers, and supervisors at Mahidol Wittayanusorn Science High School and Kasetsart University for imparting their knowledge and inspiring me to pursue a career in science. A special thanks to Thongchai Chewprecha, the principal of Mahidol Wittayanusorn Science High School, for your insightful words on the importance of scientific development for Thailand and for encouraging me to pursue higher education at a top-ranking university in the world. I am deeply thankful to Julapark Chunwongse, my bachelor thesis supervisor at Kasetsart University, for your unwavering support, invaluable advice on both academic and life matters, and for giving me a second chance to rediscover my passion for science when I faced distractions in life. Without your guidance, I may have abandoned my scientific aspirations and pursued a different path. I would also like to express my gratitude to my Thai friends who supported and encouraged me to pursue my PhD journey: Itsarayot (Tim), Oranut (Orn), Parin (Bank), Pattarawan (Aon), Pornpimon (Miew), Tanachart (Gor), and all others whom I may not have mentioned here. Your presence has been instrumental in my long and challenging journey, and I am truly thankful for your support.

To my beloved family, ข้าพเจ้ากราบขอขอบคุณตา ดี แย้มจะบก ที่ปูพื้นฐานให้ความสำคัญกับการศึกษามาตั้งแต่รุ่นคุณแม่ ทำให้เป็นส่วนหนึ่งที่ส่งเสริมให้ข้าพเจ้ามาจนถึงในวันนี้ ถึงแม้วันนี้คุณตาดีจะล่วงลับไปแล้ว ข้าพเจ้าจะสืบทอดนามสกุล “แย้มจะบก” และพาไปให้ไกลที่สุดเท่าที่ข้าพเจ้าทำได้ ข้าพเจ้ากราบขอขอบคุณยายณี ที่ดูแลและไม่เคยบังคับเรื่องอาหารการกินกับข้าพเจ้าตั้งแต่ยังเด็ก ข้าพเจ้ากราบขอขอบคุณแม่จินดาที่เชื่อมั่นในทางที่ลูกเลือก ไม่ว่ามันจะแปลกและแตกต่างสักเพียงไหน วันนี้ลูกทำได้แล้วและลูกจะก้าวหน้าต่อไป ข้าพเจ้ากราบขอขอบคุณแม่ป้า พ่อลุง ป้านิด ลุงเถียร น้าอ้อย น้าแหล่ว น้าน้อย น้าหมู สำหรับคำสั่งสอนและการสนับสนุนทางการศึกษา ถ้าไม่มีทุกคน ข้าพเจ้าคงมาไม่ถึงจุดนี้เช่นกัน ข้าพเจ้ากราบขอขอบคุณพ่ออู๋ แม่ยุ้ยและคุณอา คุณน้า คุณป้า คุณลุงที่ผมอาจไม่ได้เอ่ยนามสำหรับการสนับสนุนให้ผมมาจนถึงวันนี้ ข้าพเจ้าขอขอบคุณพี่ ๆ น้อง ๆ ทุกคนไม่ว่าจะเป็นพี่ต้า พี่ตั้ว พี่โน้ต พี่เนย เติร์ก อัน ต้นไม้ อุ้ย ไอซ์ มะปราง ที่อยู่ด้วยกันและแบ่งปันความทรงจำดี ๆ ที่มีร่วมกัน

I would like to extend a heartfelt thank you to myself for persevering through this challenging journey. Despite facing numerous difficulties along the way, I remained resilient and refused to give up. The outbreak of Covid-19 significantly disrupted my plans and posed additional obstacles, especially in terms of conducting experiments amidst social distancing guidelines. However, I adapted to the situation and managed to navigate through these challenges with determination. I am grateful for the support and understanding extended to me by Wageningen University and the Laboratory of Genetics, who provided me with the necessary accommodations and resources to complete my PhD despite the disruptions caused by the pandemic. Writing the thesis proved to be a daunting task, even more so than conducting experiments. Despite encountering delays and unforeseen obstacles, I remained committed to the task at hand. Although the process took longer than anticipated, I am proud to have persevered and successfully completed the writing of this thesis. Reflecting on this journey, I am reaffirmed in my belief that pursuing a career as a scientist is both rewarding and fulfilling. I am grateful for the opportunity to contribute to the field of research and am excited for what the future holds.

Curriculum vitae

Jitpanu Yamjabok was born on Sunday, August 28th, 1988, in Nakhon Ratchasima, Thailand. At the age of nine, he received a set of scientific comic books covering topics about animals, plants, and astronomy. These books were both entertaining and educational, sparking his interest in becoming a scientist and astronaut. In 2004, he attended Mahidol Wittayanusorn School, Thai science high school, on a full scholarship sponsored by the Thai government.

In 2007, Jitpanu enrolled in the Department of Agriculture, major in Agricultural Biotechnology, at Kasetsart University, Thailand, with a full scholarship provided by the National Science and Technology Development Agency (NSTDA). During his bachelor's degree, he developed a strong interest in molecular marker-assisted plant breeding. His bachelor's research project focused on mapping the tomato yellow leaf curl virus resistance gene using pedigree-based gene mapping in tomato at the Plant Laboratory, Center for Agricultural Biotechnology, Kasetsart University, Thailand.

Following the completion of his bachelor's degree, he undertook an internship in 2012 at Diversity Arrays Technology Pty Ltd in Canberra, Australia. This experience exposed him to cultural diversity, advanced genotyping approaches, and English-speaking environments. Subsequently, he began working as a research assistant under the supervision of Julapark Chunwongse at the Plant Laboratory, Kasetsart University, in June 2012. His responsibilities included molecular marker genotyping (e.g., Simple Sequence Repeat and Cleaved Amplified Polymorphic Sequences) for the chili breeding program, greenhouse management, and supervision of bachelor student research projects.

After two years of work, Jitpanu received a Royal Thai scholarship to pursue master's and Ph.D. degrees focusing on molecular plant breeding using modern approaches, particularly genome-wide association and genomics prediction. In 2015, he began his master's degree in Plant Biotechnology with a specialization in molecular plant breeding and pathology at Wageningen University, The Netherlands. During his master's degree, he conducted a major thesis supervised by Robert Akkers at the Laboratory of Genetics, Wageningen University, investigating the diurnal effects on zinc deficiency response in Arabidopsis thaliana. This experience introduced him to the concept of metal homeostasis, metal hyperaccumulation, and its practical implications, inspiring him to pursue a Ph.D. in metal hyperaccumulators at the Laboratory of Genetics. For his Ph.D. thesis, he investigated the genetic architecture underlying metal hyperaccumulation in the metal hyperaccumulating species Noccaea caerulescens and developed new genetic tools for this species under the supervision of Mark Aarts, Henk Schat, and Joost van den Heuvel. The results of his Ph.D. are described in this thesis.

Following his Ph.D. graduation, he will work as a scientist for 14 years at The National Center for Genetic Engineering and Biotechnology in Bangkok, Thailand.

Educational Statement of Graduate School Experimental Plant Science

Issue to: Jitpanu Yamjabok

Date: 24 April 2024

Group: Laboratory of Genetics

University: Wageningen University & Research

1. Start-up phase	*date*	*cp*
► **First presentation of your project** Genetic analysis of Zinc and Cadmium accumulation and tolerant in *Noccaea caerulescens*	May 29, 2018	1.50
► **Writing or rewriting a project proposal** Genetic analyis of Zinc and Cadmium accumulation and tolerant in *Noccaea caerulescens*		6.00
► **Writing a review or book chapter**		
► **MSc courses**		
► **Laboratory use of isotopes**		
Subtotal Start-up phase		7.50
2. Scientific Exposure (***minimum 7 ECTS credit points***)	*date*	*cp*
► **EPS PhD student days**		
EPS PhD student day	Feb 11-12, 2019	0.30
EPS PhD student day	May 3-4, 2022	0.30
► **EPS theme symposia**		
EPS theme 3 symposium, Nijmegen, NL	Oct 21 2019	0.30
EPS theme 4 symposium, Wageningen, NL	Dec 13 2019	0.30
EPS theme 3 symposium, Online	Oct 30 2020	0.3
EPS theme 3 symposium, online	Nov 5 2021	0.30
EPS theme 1 symposium, Wageningen, NL	Jun 14 2022	0.30
► **National meetings (e.g. Lunteren days) and other national platforms**		
Annual Experimental Plant Sciences Meeting, Lunteren, NL	Apr 09-10, 2018	0.60
Annual Experimental Plant Sciences Meeting, Lunteren, NL	Apr 08-09, 2019	0.60
Annual Experimental Plant Sciences Meeting, Online	Apr 12-13, 2021	0.60
Annual Experimental Plant Sciences Meeting, Lunteren, N	Apr 11-12, 2022	0.60
► **Seminars (series), workshops and symposia**		
Genetics (GSS) seminar series	Feb 2018 – Jun 2022	0.30
Symposium statistical Advances in modern plant breeding	Oct 16 2018	0.30
Statistical methods for analysis low-coverage genotyping-by-sequencing data	Jul 4 2018	0.30
Lecture: Tropical hyperaccumulator plants: a scientific travelogue (Dr. Antony van der Ent)	Feb 4 2019	0.10
The carpentries Workshop: Genomics Data	Feb 05 - 06, 2019	0.60
Seminar: Harnessing natural variation, genetic diversity and detailed, high throughput phenotyping to understand the regulation of photosynthesis (Dr. David Kramer)	Oct 10 2019	0.30
Seminar: Plant metabolism in a warming world - welcome to the dark side (Porf.Dr. Owen Atkin)	Nov 20 2019	0.30
Seminar: Mind the GxE=P; elemental content in plants (Dr. Ivan Baxter)	Dec 9 2019	0.30
Symposium: Applying Evolutionary Theory to Improve Plant Production (Dr. Jacob weiner)	Feb 25 2020	0.30
Symposium: Optimising photosynthesis for societal needs	Mar 5 2020	0.30
Open Webinar: Photosynthesis 2030+	Oct 5 2021	0.10
EPS Mendel-200 Mini-symposium	Jun 8 2022	0.10
EPS "Plants and Patents" Symposium (hybrid)	Jun 10 2022	0.15
► **Seminar plus**		

► **International symposia and congresses**		
Locomet, France	Jul 28 - 29, 2018	0.60
Locomet, France	Jul 4 - 5, 2019	0.60
TICTEC 2022, Germany	Jul 30 - 31, 2022	0.60
► **Presentations**		
Poster presentation at Lunteren meeting	Apr 08-09, 2019	1.00
Oral presentation at Locomet meeting	Jul 05 2019	1.00
Poster presentation at TICTEC 2022	Jul 30 - 31, 2022	1.00
► **IAB interview**		
► **Excursions**		
Subtotal Scientific Exposure	7	12.45
3. In-Depth Studies	*date*	*cp*
► **EPS courses or other PhD courses**		
Statistical phenotyping and GxE analysis	Jun 25 - 27 2018	0.90
Postgraduated course "Linear models"	Jun 12 - 14, 2019	0.90
Postgraduated course "Generalised linear models"	Jun 20 - 21, 2019	0.60
Postgraduated course "Mixed linear models"	Jun 27 - 28, 2019	0.60
Bioinformatic introduction course	Jul 5 - 9, 2021	1.50
CRISPR/Cas course	Sep 13 - 14, 2021	0.60
► **Journal club**		
Literature discussion Plant genetics group	Jun 2018 – Jul 2021	3.00
► **Individual research training**		
Subtotal In-Depth Studies	3	8.10
4) Personal development	*date*	*cp*
► **Skill training courses**		
English Pronunciation & Fluency	Apr 2018	0.60
Competence assessment	Nov 27, 2019	0.30
Brain training	Nov 3, 2020	0.30
Reviewing a Scientific Manuscript	Nov 12, 2020	0.30
Bridging across Cultural Differences	Apr 22 and 29, 2021	0.70
Research data and scientific publishing in a changing world	May 19, 2021	0.30
Effective and efficient verbal communication	May 26 2021	0.10
Making your science accessible	May 27 2021	0.10
Research data management	Nov 18 and 25, Dec 2 2021	0.45
Rmarkdown	Dec 6 and 13, 2021	0.60
► **Organisation of PhD students day, course or conference**		
► **Membership of Board, Committee or PhD council**		
Subtotal Personal Development	3	3.75
TOTAL NUMBER OF CREDIT POINTS*		31.80

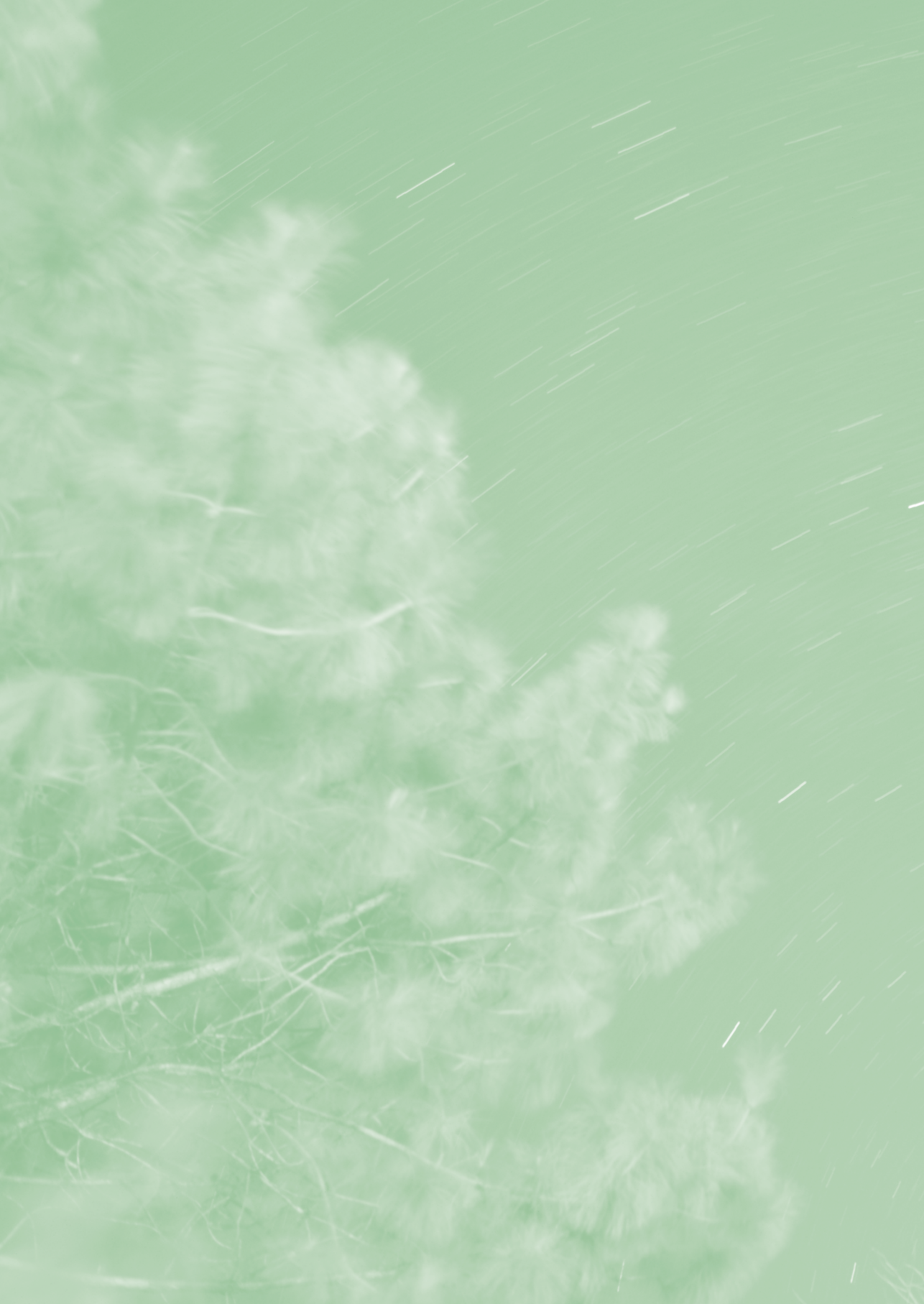
Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS

* A credit represents a normative study load of 28 hours of study

Sponsor

The project was financed by Royal Thai government and the Laboratory of Genetics, Wageningen University, The Netherlands.

Financial support from the Laboratory of Genetics, Wageningen University for printing this thesis is gratefully acknowledgement.

Cover design by Jitpanu Yamjabok

Printed by ProefschriftMaken