

**An evolutionary and functional genomics study of  
*Noccaea caerulescens*, a heavy metal hyperaccumulating  
plant species**

**Yanli Wang**

## **Thesis committee**

### **Promotor**

Prof. Dr M. Koornneef

Personal chair at the Laboratory of Genetics

Wageningen University

### **Co-promotor**

Dr M.G.M. Aarts

Associate professor, Laboratory of Genetics

Wageningen University

### **Other members**

Prof. Dr G.C. Angenent, Wageningen University

Dr M. Pauwels, Université de Lille 1 - Sciences et Technologies, Villeneuve d'Ascq,  
France

Dr C.W.B. Bachem, Wageningen University

Dr R. Geurts, Wageningen University

This research was conducted under the auspices of the Graduate School of  
Experimental Plant Sciences (EPS).

**An evolutionary and functional genomics study of *Noccaea  
caerulescens*, a heavy metal hyperaccumulating plant species**

**Yanli Wang**

**Thesis**

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Friday 9 September 2016

at 1:30 p.m. in the Aula.

Yanli Wang

An evolutionary and functional genomics study of *Noccaea caerulea*, a heavy metal hyperaccumulating plant species

PhD thesis, Wageningen University, Wageningen, NL (2016)

With references, and summary in English

ISBN 978-94-6257-856-2

DOI: 10.18174/385510

**Dedicated to my beloved family!**



## CONTENTS

### Chapter 1

<b>General Introduction</b>	9
-----------------------------	---

### Chapter 2

<b>Population genomics identifies candidate genes under selection distinguishing non-metallicolous and calamine populations of the heavy metal adapted plant species <i>Noccaea caerulea</i></b>	23
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----

### Chapter 3

<b>Construction and analysis of a <i>Noccaea caerulea</i> TILLING population</b>	65
----------------------------------------------------------------------------------	----

### Chapter 4

<b>The genetic control of flowering time in the biennial/perennial species <i>Noccaea caerulea</i></b>	97
--------------------------------------------------------------------------------------------------------	----

### Chapter 5

<b>General discussion</b>	131
---------------------------	-----

<b>References</b>	151
-------------------	-----

<b>Summary</b>	181
----------------	-----

<b>Acknowledgements</b>	184
-------------------------	-----

<b>Curriculum Vitae</b>	187
-------------------------	-----

<b>Publications</b>	188
---------------------	-----

<b>Education statement (EPS)</b>	189
----------------------------------	-----





# **Chapter 1**

## **General Introduction**

## Soil contamination

Micronutrients are elements that play important physiological and metabolic roles in plants (Ramesh *et al.*, 2004). Heavy metals such as copper (Cu), nickel (Ni) and zinc (Zn) are also micronutrients, hence they are required in small amounts to maintain normal plant growth. Micronutrients are only accessible from the soil, where they exist as a heterogeneous mixture (Jhorar *et al.*, 2004). Certain natural soils contain high levels of heavy metals, and plants have evolved homeostatic mechanisms to control the uptake and distribution of micronutrients and to achieve normal growth in the presence of variable concentrations of metals in the environment (Haydon *et al.*, 2015). Other soils are contaminated with toxic levels of heavy metals such as cadmium (Cd), Cu, Ni and Zn, reflecting human activities such as mining, smelting, energy and fuel production, and the use of sewage and municipal waste as fertiliser (Kabata-Pendias, 2011) (Sun *et al.*, 2010; Wuana and Okieimen, 2011; Gusiati and Klimiuk, 2012). The contamination of soils with heavy metals causes environmental issues because metals can leach but not degrade, so the total concentration increases and the metals persist for a long time (Adriano, 2001). Plant species differ in their sensitivity to metals, and this influences the vegetation that grows on contaminated sites.

Ni is an essential micronutrient for plant growth and development, but excess Ni is toxic to plants and inhibits normal growth. Many natural sites contain metalliferous ultramafic serpentine soils in which Ni levels can reach 2000 ppm (Assunção *et al.*, 2003). Zn is also an essential micronutrient, and is required as a cofactor for many enzymes and transcription factors in plants (Di Baccio *et al.*, 2005; Jain *et al.*, 2010; Calgaroto *et al.*, 2011). Normal soils contain ~100 ppm Zn, but this increases to more than 1000 ppm in agricultural areas due to the use of fertilisers (Broadley *et al.*, 2007). Cd is a highly toxic environmental pollutant, predominantly released as a byproduct of Zn mining (Solti *et al.*, 2008) and the manufacture of phosphate fertilisers (Panda *et al.*, 2011). Although Cd is not an essential nutrient in plants, it can be taken up by the roots, most probably due to inadvertent adsorption and translocation. High levels of Cd disrupt the uptake and distribution of essential metals such as Cu, iron (Fe) and

manganese (Mn), thus interfering with processes that depend on them, and Cd also causes oxidative stress and toxicity in plant tissues (Sridhar *et al.*, 2005; Cambroll *et al.*, 2012). Cd is also highly toxic to humans (Godt *et al.*, 2006; Bernard, 2008).

## **Strategies to cope with metals**

According to the sensitivity to the metals, plants can be divided into two groups: metal sensitive and metal tolerant species (Baker, 1987). Sensitive species are highly susceptible to metal toxicity even at concentrations a little higher than physiological levels, leading to the development of toxicity symptoms and death soon after exposure. In contrast, tolerant species can survive in the presence of metals without toxicity symptoms. Some of these tolerant species achieve high metal tolerance by restricting metal bioavailability in the soil or preventing the uptake of metals into root cells, and these are known as excluders. Others sequester metals in the root tissue, or transport the metals from the root to the shoot and store extraordinarily high concentrations of metal in the leaves, and the latter are known as hyperaccumulators (Rascio and Navari-Izzo, 2011; Lin and Aarts, 2012; van der Ent *et al.*, 2013).

## **Heavy metal hyperaccumulation**

Plant species show great variation in their ability to tolerate and accumulate different metals. The threshold that defines a Ni hyperaccumulator is a leaf Ni concentration of more than 1000 µg/g dry weight (DW) (Brooks *et al.*, 1977). The threshold for a Cd hyperaccumulator is a leaf Cd concentration of greater than 100 µg/g DW, whereas for Co, Cu, Mn, Pb and Zn, this value is 10,000 µg/g DW (Baker and Brooks, 1989). Based on these thresholds, more than 450 plant species have been identified as hyperaccumulators (Verbruggen *et al.*, 2009; Krämer, 2010). Due to the term 'hyperaccumulator' has been used with varying degrees of precision, aptness and understanding that have not always corresponded with the views of the originators of the terminology, therefore a new criterion for a hyperaccumulator has been recently proposed to clarify the circumstances in which the term 'hyperaccumulator' is appropriately and the conditions that should be met when the terms are used. According to the new threshold, plants that accumulate metal concentrations in their

leaves 2–3 orders of magnitude higher than normal when growing in standard soil, or one order of magnitude higher than normal when growing in metalliferous soil are considered as a hyperaccumulator (van der Ent *et al.*, 2013). Based on these criteria, more than 500 species have been identified as hyperaccumulators, most of which (450) are Ni hyperaccumulators. This is expected because the most abundant metalliferous soil type is serpentine, which is particularly enriched in Ni. Much smaller groups of plants are Cu, Co and Se hyperaccumulators (32, 30 and 20 species, respectively), and even smaller groups are hyperaccumulators of Pb (14), Mn (12) and Zn (12). Only few hyperaccumulators have been reported for other elements, including As, Cd and Tl, with 5, 2 and 2 species, respectively (van der Ent *et al.*, 2013). Zn hyperaccumulators often growing on calamine soils are usually Cd accumulators too, with enhanced Cd tolerance (Assunção *et al.*, 2003c).

### ***Noccaea caerulescens*, a Zn/Cd/Ni hyperaccumulator**

*Noccaea caerulescens* F.K. Meyer (formerly known as *Thlaspi caerulescens*) is the only natural Zn/Cd/Ni hyperaccumulator reported thus far. *N. caerulescens* ( $2n = 14$ ) and the Zn/Cd hyperaccumulator *Arabidopsis halleri* are prominent model plants for the study of heavy metal tolerance and hyperaccumulation. *N. caerulescens* is self-compatible and propagates largely by selfing, which makes it more likely to be homozygous in the wild than outcrossing species such as *A. halleri* (Mousset *et al.*, 2016). The *N. caerulescens* mating system facilitates fundamental genetic analysis because fewer generations are needed to produce homozygous inbred lines, which can be used to establish functional genomics tools. Furthermore, natural mutations are become fixed relatively easily and are dispersed by seeds.

*N. caerulescens* is distributed mainly in Europe (Koch *et al.*, 1998). Depending on the soil metal concentrations, three main ecotypes can be distinguished: serpentine (Ni enriched), calamine (Zn/Cd enriched) and non-metallicolous (normal soil). As an adaptation to elevated metal concentrations, metallicolous populations of *N. caerulescens* usually accumulates smaller quantities of metal but is more tolerant towards heavy metals than non-metallicolous ones under control conditions (Meerts and Van Isacker, 1997; Escarré *et al.*, 2000; Frérot *et al.*, 2003). Furthermore, even

non-metallicolous *N. caerulescens* populations have an extremely higher Zn/Cd/Ni accumulation capacity and tolerance than non-metallicolous populations of non-hyperaccumulator metallophytes such *Arabidopsis thaliana* (Lasat *et al.*, 1996; Meerts and Van Isacker, 1997; Escarré *et al.*, 2000; Assunção *et al.*, 2003). Although the serpentine ecotype is typically a Ni hyperaccumulator, it can also accumulate exogenous Zn. Zn hyperaccumulation and tolerance are therefore constitutive in *N. caerulescens* although there is heritable quantitative variation for this characteristic among populations (Zha *et al.*, 2004; Assunção *et al.*, 2006). Natural Cd hyperaccumulation is observed only in calamine populations, (Escarré *et al.* 2000) and the population from South France accumulates more Cd than other calamine varieties (Xing *et al.*, 2008). The rich phenotypic variation in Zn/Cd/Ni accumulation and tolerance makes *N. caerulescens* a highly attractive species for comparative analysis to explore the molecular and evolutionary mechanisms underlying metal tolerance and hyperaccumulation.

## **Molecular mechanisms of metal tolerance and hyperaccumulation**

Differences in the physiological processes required for the uptake, translocation and sequestration of metals in plants have been reported, commensurate with variations in heavy metal accumulation and tolerance (Lasat *et al.*, 1996; Lombi *et al.*, 2001; Zhao *et al.*, 2002; Clemens, 2001; Clemens *et al.*, 2002; Hall, 2002; DalCorso *et al.*, 2010).

The molecular mechanism underlying heavy metal hyperaccumulation has been explored by transcript profiling in the presence of excess Zn or Cd in *Arabidopsis thaliana* (Wintz *et al.*, 2003; Kovalchuk *et al.*, 2005; Herbette *et al.*, 2006) or *N. caerulescens* (Halimaa *et al.*, 2014; Lin *et al.*, 2014), and by comparative transcriptomic analysis in *A. thaliana* and *A. halleri* (Becher *et al.*, 2004; Weber *et al.*, 2004, 2006) or *A. thaliana* and *N. caerulescens* (Rigola *et al.*, 2006; Hammond *et al.*, 2006; van de Mortel *et al.*, 2006; van de Mortel *et al.*, 2008; Plessl *et al.*, 2010). These studies have revealed many genes that are overexpressed in Zn/Cd hyperaccumulator species compared to *A. thaliana*, which may therefore be related to the variation in Zn/Cd accumulation and tolerance among different species. These studies have also highlighted the

## Chapter 1

transcriptional regulation of many Zn/Cd-responsive genes that are strikingly different in *A. halleri* and *N. caerulescens*, suggesting that the molecular mechanisms of metal hyperaccumulation may have evolved independently in different species.

The modulation of gene expression is one evolutionary mechanism underlying variations in metal hyperaccumulation and tolerance. Three types of mutation can cause differential gene expression, namely the modification of cis-regulatory regions such as the promoter, copy number expansion or mutations that affect the activity of transcription factors (Stern and Orgogozo, 2008). For example, *HMA4* encodes a P-type ATPase metal transporter that loads the xylem with Zn and Cd (Hanikenne *et al.*, 2008) and its overexpression in *A. halleri* correlates with enhanced Zn/Cd translocation and accumulation (Hanikenne *et al.*, 2008; Nouet *et al.*, 2015). The overexpression of this gene in *A. halleri* is caused by mutations in the cis-regulatory and coding regions, plus the presence of three copies in *A. halleri* compared to the single copy in *A. thaliana* (Papoyan and Kochian, 2004; Hanikenne *et al.*, 2008). This P-type ATPase is also found in *N. caerulescens* as four tandem copies, although within-species variation has been reported (Lochlainn *et al.*, 2011; Craciun *et al.*, 2012). The ZIP-like transporter encoded by *ZNT1* is highly expressed under Zn/Cd stress conditions but it is single copy gene in *N. caerulescens* (Hammond *et al.*, 2006; van de Mortel *et al.*, 2006). Genetic variation in the *ZNT1* promoter corresponding to variations in Zn and Cd accumulation has been reported in *N. caerulescens* (Lin *et al.*, 2016).

Evidence for the modification of transcription factors leading to changes in the expression of several genes has been found in a tomato root iron transporter (Ling *et al.*, 2002). The highly expressed *N. caerulescens* genes identified in the previous transcript studies may be regulated by transcription factors such as *bZIP19*, a member of the basic leucine-zipper gene family (Assunção *et al.*, 2010). This transcription factor regulates the expression of a number of zinc transporter genes and the gene encoding nicotianamine synthase (NAS) under Zn deficiency, which is required for the synthesis of the metal chelator nicotianamine. The high expression level of metal

homeostasis genes in *N. caerulea* is potentially regulated by the orthologues of *bZIP19* and *bZIP23*, or by other transcription factors which remain to be identified.

As well as promoter modifications, genetic variations within introns and untranslated regions (UTRs) can also influence the gene expression level, whereas mutations in the coding region have the potential to alter protein structure and activity (Stern and Orgogozo, 2008; Pespeni *et al.*, 2012). The relevance of such mutations in relation to variations in metal accumulation and tolerance have not yet been tested for these strongly expressed genes because transcriptome analysis does not cover non-exon sequences (Halimaa *et al.*, 2014b; Lin *et al.*, 2014). These mutations can only be accessed by whole-genome sequencing.

## **Analysis of evolutionary genetics by population genome resequencing**

As discussed above, one way to investigate the molecular basis of metal tolerance and hyperaccumulation is the analysis of transcription in one or more accessions under different treatment conditions. However, advances in high-throughput sequencing make it possible to develop a comprehensive understanding of genetic and phenotypic variations shaped during the evolution of natural populations by genomic resequencing at the population level (Mitchell-Olds *et al.*, 2008).

The genetic variation underpinning the evolution of phenotypic diversity in a wild population is ecologically-dependent. Similar patterns of genetic and phenotypic variation would have arisen in those populations based on the similar niches they occupy (Antonovics, 1975). Selection signatures formed during evolution can be also detected at the genome level (Tajima, 1989; Cutter and Payseur, 2013). To identify genetic variation in ecologically-distinct populations that occupy contrasting niches, an unbiased approach involves the high-throughput sequencing of pooled genomic DNA from multiple individuals in natural populations (Turner *et al.*, 2010). By pooling DNA from natural populations with contrasting phenotypes, the allele frequency within and between populations can be estimated at the genome level and used to study the evolutionary processes. This approach has been very successful used to

identify both single basepair mutations as well as larger genomic rearrangements like the CNV and the indels (Turner *et al.*, 2010; Kofler and Schlötterer, 2014).

As stated above, *N. caerulescens* is represented by three distinguishable ecotypes: serpentine, calamine and non-metallicolous (Assunção *et al.*, 2003). For the calamine populations (Zn/Cd enriched), geographically-isolated non-metallicolous populations have been found nearby. The calamine populations usually accumulate less metal than non-metallicolous ones under control conditions but are more tolerant towards Zn/Cd (Assunção *et al.*, 2003; Verbruggen *et al.*, 2009). Common genetic factors may be responsible for phenotypic differences between the ecotypes, but this has not yet been tested. With the completion of the reference genome based on the inbred Ganges accession, a superior Cd accumulator derived from South France (Edouard Severing *et al.*, In preparation) and the rapid development of next generation sequencing technology, it is now possible to investigate genetic variation in an unbiased manner at the genome-scale by comparative genomic resequencing at the level of whole *N. caerulescens* populations. The identification of such intra-species genetic variation will provide insight into the molecular mechanisms and evolutionary basis of heavy metal hyperaccumulation.

### **Functional analysis of *Noccaea caerulescens* genes**

Following the identification of candidate genes under evolutionary selection and the corresponding mutations, which can be achieved using the population resequencing approach described above, the functional analysis of such genetic variations would help to determine their role in relation to the selective environment (Feder and Mitchell-Olds, 2003; Mitchell-Olds *et al.*, 2008; Stern and Orgogozo, 2008; Alonso-Blanco *et al.*, 2009). This is not currently possible in *N. caerulescens* due to the absence of a suitable mutant library.

Current approaches to determine gene function include the generation of specific loss-of-function mutants by reverse genetics techniques, e.g. homologous recombination /gene targeting (Bradley *et al.*, 1998; Rong and Golic, 2000), insertional mutagenesis/transposon tagging (Korswagen *et al.*, 1996), RNA interference (RNAi) (Bargmann, 2001; Vaucheret *et al.*, 2001), and the more recent CRISPR/Cas9



technique (Xie and Yang, 2013; Bortesi and Fischer, 2014; Hyun *et al.*, 2014). However, most of these approaches require genetic transformation, the efficiency of which is species and genotype dependent (Somers *et al.*, 2003; Iida and Terada, 2004; Stemple, 2004; Ko *et al.*, 2006). This requirement is a limitation in species such as *N. caerulea* with generally low transformation efficiencies (Peer *et al.*, 2003; Guan *et al.*, 2008). Alternatively, the functional analysis of candidate genes in *N. caerulea* can be achieved using the efficient *Agrobacterium rhizogenes* hairy root transformation system, although this technique is also genotype dependent, it does not introduce heritable mutations, and is only appropriate for genes with a role in the root (Limpens *et al.*, 2004; Iqbal *et al.*, 2013). These limitations mean that functional genomics tools must be generated using a non-transgenic approach in *N. caerulea*.

TILLING (targeting induced local lesions in genomes) is a non-transgenic functional genomics approach that was developed as a general strategy for reverse genetics, in which traditional mutagenesis is followed by high-throughput screening to discover point mutations (McCallum *et al.*, 2000a). Almost all TILLING populations have been produced using the chemical mutagen Ethyl Methane Sulfonate (EMS) because this introduces random point mutations with a higher frequency than other chemicals mutagens (Greene *et al.*, 2003; Henikoff and Comai, 2003; Kurowska *et al.*, 2011). An important advantage of TILLING in the context of reverse genetics is its ability to generate allelic series for most target genes, presenting a range of mutant phenotypes from near wild-type-effects caused by subtle mutations to complete loss-of-function (null) mutants produced by premature stop codons or mutations that disrupt splicing (Slade *et al.*, 2005). Allelic series corresponding to a wide range of repetitive phenotypes are desirable because such mutants provide more insight into the function of each target gene.

Several high-throughput post-PCR screening platforms are available. In the traditional approach, PCR amplification is followed by the cleavage of mismatches using endonucleases such as CEL1 or END1 that recognize single DNA strands (Triques *et al.*, 2008; Oleykowski *et al.*, 1998) and then polyacrylamide gel electrophoresis (PAGE) is used to detect the mutants on the sensitive Li-Cor gel analyser (Colbert *et al.*, 2001;

## Chapter 1

Till *et al.*, 2006). More recent non-enzymatic techniques such as high resolution melting curve analysis (HRM) are less expensive (Gady *et al.*, 2009) but next generation sequencing is the quickest and most cost effective approach (Tsai *et al.*, 2011) that have been applied to TILLING populations.

In the HRM platform, PCR is followed by the direct detection of mutations based on DNA melting curve analysis (Applied Biosystems, 2009). A fluorescent dye (LCgreen Plus+™) is added before the PCR step, saturating the double-stranded DNA. After the PCR step, mutations are detected by placing the microtitre plate on a light scanner. During thermal denaturation, fluorescence is emitted as the DNA strands are separated, and this loss of fluorescence can be visualised in real time using the light scanner. If mutations are present in a given well, heteroduplexes are formed in the post-PCR mix and these can be identified because the curve shifts at a different temperature compared to homoduplexes (Dong *et al.*, 2009b; Vossen *et al.*, 2009; Kurowska *et al.*, 2011). Mutations isolated on this basis can be confirmed by sequencing the PCR product. The combination of HRM screening and subsequent confirmation by sequencing has been applied successfully to detect mutations in TILLING populations of tomato (Gady *et al.*, 2009), wheat (Dong *et al.*, 2009), clinically relevant fungi (Mader *et al.*, 2011), and human tissue samples, the latter to identify disease genes (Poláková *et al.*, 2008; Joly *et al.*, 2011).

### **The control of flowering time**

*N. caerulescens* is a diploid, biennial or facultative perennial plant from the same Brassicaceae family as *A. thaliana*. Flowering in *N. caerulescens* is dependent on vernalisation. When grown under greenhouse conditions, ecotypes cultivated thus far require up to 32 weeks to flower, including 7–12 weeks for vernalisation (5°C, and 12 h or 8 h photoperiod) to induce flowering in plants that are 2 months old, followed by an additional 4 weeks for seed ripening (Peer *et al.*, 2003, 2006; Lochlainn *et al.*, 2011; Guimarães *et al.*, 2013). The life history of *N. caerulescens* varies among ecotypes and appears to be intermediate between the annual plant *A. thaliana* and the perennial plant *Arabis alpina* (Dechamps *et al.*, 2007). This suggests the genetic basis of flowering time in *N. caerulescens* may differ from both *A. thaliana* and *A. alpina*, but

this has not yet been tested due primarily to the absence of loss-of-function mutants in *N. caerulea* flowering repressors.

Flowering time is a well-studied subject in plant biology and has been proposed as a model trait for the integrated study of ecology, evolution and molecular biology (Kobayashi and Shimizu, 2013). The transition from vegetative to reproductive development is very important in flowering plants. This transition time is precisely controlled by many environmental conditions and their interactions with endogenous signals (Andrés and Coupland, 2012). Temperature and day length are key environmental cues that induce flowering. When flowering is only induced after a period of low temperatures, it is said to be dependent on vernalisation (Amasino, 2005). Exposure to low temperatures mimics the transition from winter to spring. Vernalisation therefore gives the plant competence for flowering when the optimal conditions are fulfilled. Day length (photoperiod) is another important environmental cue that controls flowering time, such that plants only flower when the day length is appropriate (Song *et al.*, 2013). Genetically, flowering time is strictly controlled by several pathways that converged on a small number of flowering regulators. These have been studied in detail by the analysis of genetic variation and comparisons between the model annual plant *A. thaliana* and the perennial species *A. alpina*.

In *A. thaliana*, flowering is blocked by modulating the expression level of *FLOWERING LOCUS C (FLC)*, encoding an MADS box protein that acts as a key floral repressor in the vernalisation pathway (Michaels and Amasino 1999; Sheldon *et al.* 1999). The high-level expression of *FLC* is promoted by *FRIGIDA (FRI)* but repressed by vernalisation (Clarke and Dean, 1994, Koornneef *et al.*, 1994; Lee *et al.*, 1994). The suppression of *FLC* expression is stably maintained even after transferring the plants back to warm conditions (Michaels and Amasino, 1999; Shindo *et al.*, 2006), promoting the expression of downstream floral promotion genes such as *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)* and *FT*. The FT protein is thought to be the molecular basis of the hypothesized activator of flowering originally named “florigen” (Angel *et al.*, 2011). These proteins, in turn, activate the expression of the downstream floral identity genes *LEAFY (LFY)* (Lee *et al.*, 2008) and *APETALA1 (AP1)* which induce

## Chapter 1

and maintain floral organ formation after the floral transition starts in the meristem (Komeda, 2004).

With the exception of *FLC*, flowering can also be blocked by *SHORT VEGETATIVE PHASE* (*SVP*), which encodes another negative regulator of the floral transition (Hartmann *et al.*, 2000). *SVP* represses flowering by a forming transcriptional initiation complex with *FLC* or independently of the latter (Mateos *et al.*, 2015). There are five additional *FLC*-like genes called *MADS AFFECTING FLOWERING* (*MAF*) 1–5, encoding MADS box transcription factors that also act as floral repressors in the summer annual of *A. thaliana*. The function of *FLOWERING LOCUS M* (*FLM* or *MAF1*) (Posé *et al.*, 2013) and *MAF2* is to accelerate flowering at higher growth temperatures.

The flowering habit of perennial plants is quite different to that of *A. thaliana*. Before vernalisation, perennial plants must strictly undergo a transition from the juvenile to the adult phase. This transition allows perennial species to achieve the competence to sense low temperatures during development. Perennial plants therefore only respond to vernalisation once they have reached maturity, and before this transition flowering remains blocked even if vernalisation occurs. Genetically, the inhibition of flowering in young plants is controlled by *TERMINAL FLOWER 1* (*TFL1*) (Wang *et al.*, 2011), as demonstrated in apple (Kotoda and Wada, 2005) and poplars (Mohamed *et al.*, 2010). The repetitive flowering trait in *A. alpina* depends on the abundance of *PERPETUAL FLOWERING 1* (*PEP1*) mRNA, which is orthologous to *FLC* in *A. thaliana*. The suppression of *PEP1* expression by vernalisation initiates flowering and restricts it to a short duration (Wang *et al.*, 2009). Because *PEP1* expression increases again after flowering, shoots that have not yet flowered remain vegetative and can flower in subsequent years after vernalisation (Wang *et al.*, 2009). A similar strategy has also been reported in the perennial species *A. halleri* (Franks *et al.*, 2007; Aikawa *et al.*, 2010).

The effect of vernalisation on the induction of flowering is unstable and non-heritable, which means that a vernalised plant cannot transmit its rapid-flowering trait to the next generation (Dennis and Peacock, 2007). The repression of *FLC* by vernalisation has been studied extensively at the epigenetic level and may be conserved across species (Schmitz and Amasino, 2007). Similar epigenetic mechanisms may suppress the expression of *MAF* genes although this has still been elusive (Alexandre and Hennig, 2008). In *A. thaliana*, the significant variation in requirements for and

responses to vernalisation in the context of floral induction may reflect genetic variation in the *FRI* (Werner *et al.*, 2005), *FLC* (Sheldon, 2002; Li *et al.*, 2014); In the perennial *A. alpina*, genetic variation in *PEP1* appear to contribute to variation in flowering time (Albani *et al.*, 2012). The relevance of *MAF* gene expression in the context of flowering time under ambient temperature conditions has been described for the summer annual of *A. thaliana* (Ratcliffe *et al.*, 2003; Alexandre and Hennig, 2008).

## The scope of this thesis

The work described in this thesis investigates natural genetic variation in the heavy metal hyperaccumulator *Noccaea caerulea* under conditions that select for heavy metal hyperaccumulation and tolerance, and the generation of functional genomics tools to facilitate the study of this species. **Chapter 2** describes analysis of genetic variation that has arisen during the evolution of Zn/Cd hyperaccumulation and tolerance, which is investigated by comparing the genomes of metalicolous (Zn/Cd enriched) and non-metallicolous ecotypes. Some consistent genetic differences that can explain the phenotypic distinction between these ecotypes are investigated and discussed. **Chapter 3** describes the first EMS-induced TILLING library in the Southern French accession St Felix de Pallières (SF), a Zn/Cd hyperaccumulator, suitable for functional analysis. Mutations involved in mineral homeostasis are identified by measuring the concentration for 20 mineral elements in 7000 individual plants of the M2 population, and high resolution melting curve analysis (HRM) is used as a high-throughput mutation screening platform to identify relevant allelic series. **Chapter 4** investigates variation in flowering time in response to vernalisation in different *N. caerulea* accessions, and describes floral repressors in the vernalisation pathway. The unique mechanisms controlling flowering time in the context of the biennial/perennial lifestyle of *N. caerulea* is investigated by comparing this species to the well-characterized annual plant *Arabidopsis thaliana* and the perennial plant *Arabis alpina*. **Chapter 5** discusses the main results of the experimental chapters and offers perspectives for future research.

## Chapter 2

### **Population genomics identifies candidate genes under selection distinguishing non-metallicolous and calamine populations of the heavy metal adapted plant species *Noccaea caerulescens***

Yanli Wang<sup>a†</sup> Joost van den Heuvel<sup>a,b†</sup>, Edouard Severing<sup>d</sup>, Elio Schijlen<sup>e</sup>, John Danku<sup>f</sup>, David E Salt<sup>f</sup>, Henk Schat<sup>a,c</sup>, Mark G.M. Aarts<sup>a\*</sup>

<sup>a</sup> Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

<sup>b</sup> Institute for Cell and Molecular BioSciences, Newcastle University, Campus for Ageing and Vitality, NE4 5PL Newcastle Upon Tyne, United Kingdom

<sup>c</sup> Institute of Cellular and Molecular Biology, Faculty of Earth and Life Sciences, VU University, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

<sup>d</sup> Max Planck Institute for Plant Breeding Research, Carl-von-Linné Weg 10, D-50829 Cologne, Germany

<sup>e</sup> Wageningen University and Research, Business Unit Bioscience, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands

<sup>f</sup> School of Biological Sciences, University of Aberdeen, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU, United Kingdom

<sup>†</sup> These authors contributed equally to this work

\*author for correspondence: [mark.aarts@wur.nl](mailto:mark.aarts@wur.nl)

#### Financial source:

This research was financially supported by the China Scholarship Council (CSC), the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research) and the Netherlands Genome Initiative ZonMW Horizon program through Zenith project no. 40-41009-98-11084

## Abstract

*Noccaea caerulescens* is a metal hyperaccumulator plant species found on calamine, serpentine and non-metalliferous soil, with the calamine (M) ecotype usually exhibiting higher Zn/Cd tolerance but lower accumulation than the non-metalliferous (NM) one. To identify the loci under selection we performed a whole genome sequence comparison of six *N. caerulescens* populations. For two M populations (Prayon and Pontaut), a geographically nearby NM population was available (Lellingen and Vall de Varrados, respectively), two other populations Lanestosa (M) and Navacelles (NM) are geographically distant. DNA from 30 plants per population were pooled and genomic libraries suitable for paired-end Illumina HiSeq were constructed. The loci with the highest *Fst* values between geographically close M and NM populations were considered to be under divergent selection. Validation of candidate genes in knockout mutants of *Arabidopsis thaliana* revealed that two of them, *RLK* and *GL3*, showed aberrant phenotypes for Zn/Cd tolerance or accumulation when compared to wild-type plants, corresponding with those found in the *N. caerulescens* ecotypes, indicating convergent independent local evolution. For, *RLK*, a negative peak was detected for Tajima's D, indicating directional selection. One *N. caerulescens* specific gene showed an extremely negative peak as well for Tajima's D indicating directional selection. Due to variation in Zn and/ or Cd accumulation or tolerance between the pairs PON/VDV, PR/LE, and LAN/NAV, we also compared LAN/PR vs LE/NAV, LAN/PON/PR vs LE/NAV/VDV, PON/PR vs LE/VDV and LAN/PR vs LE/NAV/VDV. Two out of 16 loci that overlapped in all the comparisons were the metal homeostasis gene *OPT3*, and the stress-responsive transcription factor *WRKY72*. The *N. caerulescens* specific gene with the most extreme Tajima's D value has been detected as well in these comparisons. The findings from this study will provide a valuable resource for future investigations to achieve a better understanding of evolutionary scenarios.



## Introduction

Only around 400 plant species, from about 45 families, have been classified as heavy metal hyper-accumulators, representing less than 0.2% of all angiosperms (Chaney *et al.* 1997; Pollard *et al.* 2002). Heavy metal hyper-accumulators can accumulate exceptionally high concentrations of certain trace elements in their aboveground parts without visible toxicity symptoms (Verbruggen *et al.*, 2009). For example, the concentrations of Zn and Cd have to be more than 10,000  $\mu\text{g g}^{-1}$  d.wt and 100  $\mu\text{g g}^{-1}$  d.wt, respectively, in leaves of a plant in order to be considered of a hyper-accumulator (Baker, 2000; van der Ent *et al.* 2013). In non-hyper-accumulators, a concentration above 300  $\mu\text{g g}^{-1}$  Zn is considered toxic (Marschner, 1995). Hyper-accumulators often combine high concentrations with extreme tolerance to the metal they accumulate. Hyper-accumulators have recently gained considerable interest because of their potential use in phytoremediation (Chaney *et al.*, 1997) and food crop biofortification (Clemens *et al.*, 2002). Next to its practical application, hyper-accumulation is a trait of interest to answer questions about the evolution, ecological relevance and genetics of novel traits (Ma *et al.*, 2001), as it has evolved multiple times in different plant lineages.

*Noccaea caerulescens* is one of the best studied hyper-accumulators, with a wide distribution in Europe (Koch *et al.*, 1998). It is one of only a few hyper-accumulators occurring on both metalliferous (rich in metal, in this case Ni/Zn/Cd) and non-metalliferous (normal) soils (Dunn, 2007). On average, the metalcolous populations growing on metalliferous soils have lower Zn and Cd accumulation capacity, but higher tolerance compared with the non-metallicolous ones growing on non-metalliferous soils (Meerts and Van Isacker, 1997; Escarré *et al.*, 2000; Frérot *et al.*, 2003). Nevertheless, even the metal tolerance of non-metallicolous *N. caerulescens* populations is much higher than normally reported for non-metallicolous populations of non-hyperaccumulating metallophytes (Lasat *et al.*, 1996; Meerts and Van Isacker, 1997; Escarré *et al.*, 2000; Assunção *et al.*, 2003).

While Zn hyper-accumulation and tolerance seem to be constitutive in *N. caerulescens*, there is heritable variation in degree between accessions (Assunção *et al.*, 2003b,

2006). Cd hyper-accumulation, on the other hand, has only been found in nature in some calamine populations (Escarré *et al.*, 2000). Heritable variation in the accumulation and translocation for both Zn and Cd has been demonstrated in *N. caerulescens* (Xing *et al.* 2008). The mechanisms, underlying this variation are incompletely understood, but attributable to variation in metal uptake, xylem loading, sequestration and detoxification in the shoot (Clemens, 2001; Hall, 2002; DalCorso *et al.*, 2010). The availability of the genome sequence of *Arabidopsis thaliana* and other *Brassicaceae* species has greatly facilitated the understanding of these physiological processes and the underlying molecular mechanisms in *N. caerulescens* (Pence *et al.*, 2000; van de Mortel *et al.*, 2008; Halimaa *et al.*, 2014b; Lin *et al.*, 2014) and *Arabidopsis halleri*, another Zn/Cd hyperaccumulating *Brassicaceae* (Clemens *et al.*, 2002; Krämer, 2010; Verbruggen *et al.*, 2013)

Progenies of crosses between *N. caerulescens* accessions with contrasting phenotypes for tolerance and accumulation have yielded more or less continuous phenotype distributions, indicating genetic control by a number of loci (Assunção *et al.* 2006; Deniau *et al.* 2006; Assunção *et al.* 2003b; Zha *et al.* 2004). (Lombi *et al.*, 2001; Zhao *et al.*, 2002). Such quantitative genetic variation has been identified in these segregation populations by quantitative trait loci (QTL) analysis (Alonso-Blanco and Koornneef, 2000). Naturally selected genetic differentiation in relation to heavy metal accumulation has been studied using the polymorphisms in a few metal transporters in Swiss *N. caerulescens* populations, but small sample size (Basic *et al.*, 2006), however, this has not yet led to the identification of target loci for natural selection that contribute to the variation in metal accumulation, translocation and tolerance in *N. caerulescens* (Koch *et al.*, 1998; Dubois *et al.*, 2003; Basic and Besnard, 2006; Jimenez-Ambriz *et al.*, 2007).

The rapid development of next generation sequencing technology, it is now possible to investigate genetic variation in an unbiased manner at the genome-scale by comparative genomic resequencing wild populations (Turner *et al.*, 2010). To identify genetic variation in ecologically-distinct populations that occupy contrasting niches, an unbiased approach involves the high-throughput sequencing of natural populations.

In pooling sequencing, multiple individuals of the populations are sampled, genomic DNA isolated and pooled per population and used to construct genomic libraries suitable for paired-end Illumina HiSeq whole genome re-sequencing. Each library is of deep sequenced. The reads will be aligned to a suitable reference sequence for comparison. Allele frequency within and between populations will be estimated and used to study the evolutionary processes. This approach has been successfully applied to detect both basepair mutations as well as large structural rearrangements e.g copy number variation and indels (Turner *et al.*, 2010).

Here, we identified target loci under selection corresponding to considerable variation in the Zn/Cd accumulation and tolerance by genome-wide resequencing of three metalicolous (Zn-Cd enriched) and three non-metallicolous (normal soil) populations in *N. caerulescens*. The potential function of strong candidate genes in relation to Zn/Cd accumulation and tolerance has been established by phenotyping *A. thaliana* T-DNA insertion mutants in orthologues of the *N. caerulescens* genes. The knockouts of two of these genes corresponded with the phenotypic variation found among the re-sequenced *N. caerulescens* populations.

## Results

### **Characterization of six natural calamine and non-metallicolous populations of *N. caerulescens***

Seeds of *N. caerulescens* (J. and C. Presl) F.K. Meyer were collected from six different populations, three growing on metalliferous calamine soil, enriched in Zn and Cd, and three growing on normal soil, not enriched in heavy metals. For two metalicolous (M) populations, Prayon (PR) and Pontaut (PON), a geographically nearby non-metallicolous (NM) population was found, i.e. Lellingen (LE) and Vall de Varrados (VDV), respectively. For the other metalicolous population, Lanestosa (LAN), a nearby non-metallicolous population has not been found. The population Navacelles (NAV) originates from non-metalliferous calcareous soil, close to the village of Ganges, South France, i.e. the region where the inbred line used to establish the reference genome (GA) originates from. The coordinates of the sites and characteristics of their soils are

listed in Table 1. As expected, the Cd and Zn concentrations were extremely high at the calamine sites LAN, PON and PR, comparable to GA, which is also a metalliferous calamine soil (Robinson *et al.*, 1998; Reeves *et al.*, 2001). Conversely, the Zn and Cd concentrations in non-metalliferous soils were low, within the ranges expected for normal soils (Schachtschabel *et al.*, 1992).

**Table 1:** Characteristics of metallicolous and non-metallicolous sites of *N. caerulescens* populations and coordinates of the sites. Soil samples were taken at a depth of 0–15 cm. The concentrations of metals are expressed as  $\mu\text{g metal g}^{-1}$  soil d. wt. Values represent means  $\pm$  SE (PR, LE, LAN); nd, not detectable. For NAV, the Zn and Cd concentration in the soil has been taken from Dubois *et al.* (2002).

Population	Site	Coordinates	Soil characteristic	Zn (ppm)	Cd (ppm)
LAN	Lanestosa, Spain	43°13'38.67"N 3°25'55.30"W	Zn/Cd mine	56300 $\pm$ 23150	21 $\pm$ 5
NAV	Navecells, France	3°30'32.8" E 43°53'15.1"N	Normal	18	1.2
PON	Pontaut, Spain	42°50'2.65"N 0°43'58.63"E	Zn/Cd mine	20000	29
VDV	Val de Varados, Spain	42°46'33.98"N 0°49'33.53"E	Normal	268	nd
PR	Prayon, France	50°34' 54.2" N 005°39' 52.0" E	Zn/Cd mine	26169 $\pm$ 56	148 $\pm$ 61
LE	Lellingen, Luxemburg	49°59' 27.6" N 006°00' 07.3" E	Normal	109 $\pm$ 15	nd

When comparing the geographically close M/NM pairs, PR was much more tolerant to both Zn and Cd than LE, which was apparent from the tolerance index, as well as the lowest exposure levels that caused leaf chlorosis (**Table 2 and 3**). Of the other pair, PON was no more Cd- or Zn-tolerant than VDV, at least according to the tolerance index. Based on leaf chlorosis, however, PON was slightly more Zn-tolerant. Although the metallicolous ecotype was on average more tolerant to both metals than the non-metallicolous one, as expected, there was some overlap, mainly because PON was barely or not more tolerant than the non-metallicolous populations, on average (**Table 2 and 3, Table S1, Figure S1**).

**Table 2:** Metal tolerance in the *N. caerulescens* populations: Pontaut (PON), Val de Varrados (VDV), Prayon (PR), Lellingen (LE), Lanestosa (LAN) and Navacelles (NAV). Tolerance was inferred from the presence or absence of chlorosis after 3 weeks of growth under metal exposure. The figures represent the lowest exposure levels at which chlorosis was observed. The series used were: control (2  $\mu$ M), 50, 100 and 1000  $\mu$ M Zn; 0.5, 5 and 50  $\mu$ M Cd. Metals were supplied as sulfates. NM: Non-Metallicolous ecotype M: Metallicolous ecotype

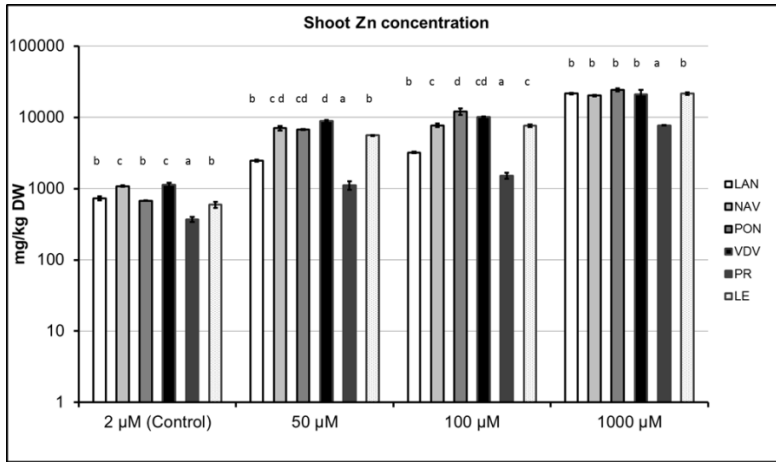
Ecotype	Population	Zn ( $\mu$ M)	Cd ( $\mu$ M)
NM	NAV	50	5
	VDV	50	5
	LE	50	5
M	LAN	1000	50
	PON	100	5
	PR	1000	50

**Table 3:** Zn and Cd tolerance index (TI) in *N. caerulescens* populations: Lanestosa (LAN) and Navacelles (NAV). Pontaut (PON), Val de Varrados (VDV), Prayon (PR), Lellingen (LE). The tolerance index (TI) for Zn and Cd was calculated as the quotient of the biomass under 1000  $\mu$ M of Zn or 50  $\mu$ M of Cd treatments and the biomass under control conditions (Baker et al 1994). Data were log-transformed prior to the statistical analysis. The biomass is the mean of 6-10 plants per treatment. Significant differences between means ( $p < 0.05$ ) are indicated by different letters.

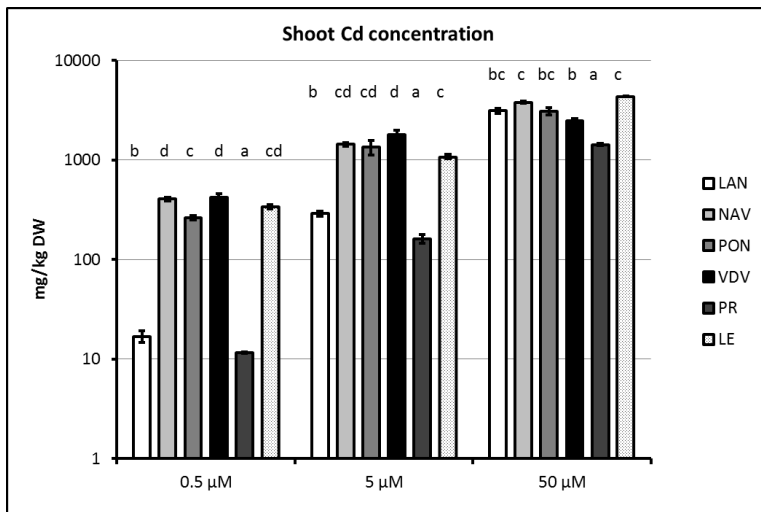
Populations	Zn (TI)		Cd (TI)	
LE	0.63 $\pm$ 0.02	b	0.27 $\pm$ 0.02	a
PON	0.42 $\pm$ 0.01	a	0.37 $\pm$ 0.02	ab
VDV	0.61 $\pm$ 0.04	b	0.39 $\pm$ 0.02	b
NAV	0.70 $\pm$ 0.02	bc	0.51 $\pm$ 0.02	c
LAN	1.09 $\pm$ 0.02	d	0.63 $\pm$ 0.02	d
PR	0.79 $\pm$ 0.07	c	0.66 $\pm$ 0.05	d

Zn and Cd concentrations in roots and shoots after three weeks of exposure were determined in hydroponically grown plants (**Figure 1 and 2, Table S2**). Both the Zn and Cd concentrations increased with increasing exposure. When exposed to 2  $\mu$ M Zn, as ZnSO<sub>4</sub>, the shoot Zn concentrations were significantly lower in PR and PON than in

their non-metallicolous counterparts LE and VDV, respectively (**Figure 1**). Also at higher exposure levels, the shoot Zn concentrations were significantly lower in PR than they were in LE, however, the difference between PON and VDV was no longer apparent. Although the shoot Zn concentrations were on average higher in the non-metallicolous ecotype than in the metallicolous one, there was clearly overlap, mainly because of relatively high concentrations, in particular, PON and, at the 2- and 1000- $\mu\text{M}$  Zn exposure levels, LAN, in comparison with, in particular, LE, at least at the lower exposure levels. Of all the populations, PR consistently showed the lowest shoot Zn concentrations, irrespective of the Zn concentration in the nutrient solution (**Figure 1**). Shoot Cd concentrations were always significantly lower in PR than in its non-metallicolous counterpart, LE, irrespective of the Cd concentration in the nutrient solution (**Figure 2**). On the other hand, there was no significant difference between PON and VDV at the 5- and 50- $\mu\text{M}$  Cd exposure levels. The shoot Cd concentration was on average lower in the metallicolous ecotype than it was in the non-metallicolous one, entirely due to the low shoot concentrations of LAN and PR across the whole range of exposure concentrations. The metallicolous population, PON, was not considerably different from the other non-metallicolous populations (**Figure 2**).



**Figure 1:** Zn concentration (mean  $\pm$  SE) in shoots in the *N. caerulea* metalicolous accessions LAN (white), PON (dark grey), PR (light dark), and in the non-metallicolous accessions NAV (light grey), VDV (dark) and LE (white with dots). Five to ten plants were analysed per Zn-treatment level indicated on the X-axis (concentration of ZnSO<sub>4</sub>). The Y-axis is log-transformed. Significant differences between means ( $p < 0.05$ ) are indicated by different letters.



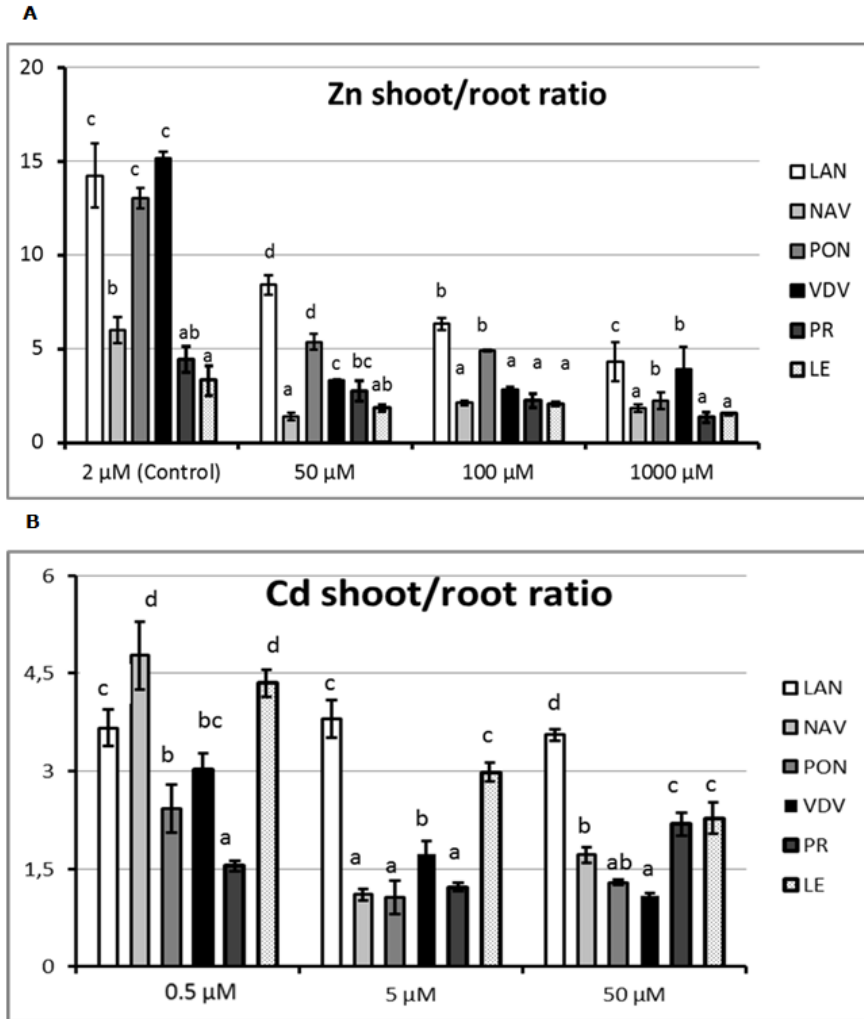
**Figure 2:** Cd concentration (mean  $\pm$  SE) in shoots in the *N. caerulea* metalicolous accessions LAN (white) PON (dark grey), PR (light dark), and in the non-metallicolous accessions NAV (light grey), VDV (dark) and LE (white with dots). Five to 10 plants were analysed per treatment level. The axis (Y) is log transformed. Significant differences between means ( $p < 0.05$ ) are indicated by different letters.

To compare the root-to-shoot translocation of Zn and Cd between ecotypes and populations, we used the shoot to root metal concentration ratio (**Figure 3**). In general, the translocation rates decreased with increasing exposure level, either gradually or abruptly, at the second treatment level (50  $\mu\text{M}$  Zn, or 5  $\mu\text{M}$  Cd), except for Cd in LAN and PR. Within the NM/M pairs, PON did not significantly differ from VDV at the lowest treatment levels (2  $\mu\text{M}$  Zn and 0.5  $\mu\text{M}$  Cd), but maintained a higher Zn translocation rate at the 50- and 100- $\mu\text{M}$  Zn treatment levels. On the other hand, VDV maintained a slightly, but significantly, higher Cd translocation at the 5- $\mu\text{M}$  Cd treatment level. Within the other pair, the Cd translocation rate was much higher in LE than in PR, except at the highest treatment level. The Zn translocation rates, however, were not significantly different, irrespective of the treatment level. Both for Zn and Cd, there were no consistent differences between ecotypes regarding the root-to-shoot translocation rates. The two most metal tolerant metallicolous populations, LAN and PR, showed almost maximally contrasting translocation phenotypes, both for Zn and Cd. The same is true for the non-metallicolous populations VDV and LE, at least at the lower Zn treatment levels (2 and 50  $\mu\text{M}$ ) and the higher Cd treatment levels (5 and 50  $\mu\text{M}$ ).

### **Genomic comparison of metallicolous and non-metallicolous populations**

Average coverages of the populations sequenced was 56.33 (LAN), 57.75 (LE), 44.19 (NAV), 74.36 (PON), 59.02 (PR) and 47.61 (VDV, see also **TableS3b**). In total 222,399,876 nucleotides were mapped and covered at least once in one of the populations. The NAV population showed a much higher number of nucleotides covered compared to the other populations. This is expected, since NAV is the population most closely situated to the 'Ganges' population from which the inbred line was reared to generate the reference genome sequence (99.04% of reference covered in NAV, compared to the next highest 92.44% in PR, **Table S4**). Because coverage differences due to variation in sequencing effort were consistent over the genome (**Figure S2**), nucleotide diversity ( $\pi$  in non-overlapping windows of 500 kb, Charlesworth and Charlesworth, 2000) was estimated from subsampled data. The

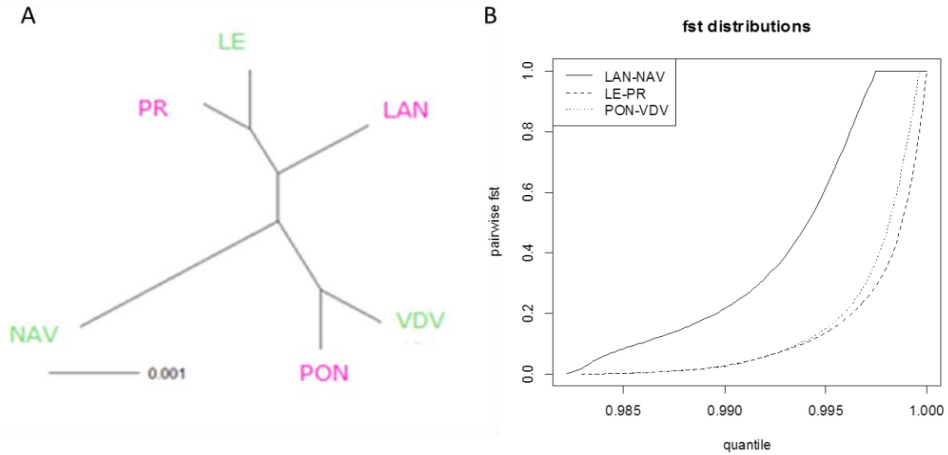




**Figure 3:** Zn (A) and Cd (B) shoot to root concentration ratios in the *N. caerulea* metalicolous accessions LAN (white) PON (dark grey), PR (light dark) and in the non-metallicolous accessions NAV (light grey), VDV (dark) and LE (white with dots). Five to ten plants were analysed per treatment level. The X-axis indicated the concentration of  $\text{ZnSO}_4$ . The Y-axis is log-transformed. Significant differences between means ( $p < 0.05$ ) are indicated by different letters.

average value of  $\pi$  varied between 0.00268 for PON to 0.00497 for NAV, while the values for LAN, LE, PR and VDV were 0.00418, 0.00443, 0.00409 and 0.00432 and therefore intermediate (**Table S3a**). Genetic variation measured as nucleotide diversity was therefore not very different between the different populations. Because PON has the highest coverage and NAV the lowest and the other populations had intermediate coverage (**Table S3b**), a negative correlation was found between coverage and  $\pi$  ( $F_{1,4}=25.91$ ,  $p<0.01$ ), potentially indicating a methodical bias in the estimation of  $\pi$ . Without the populations of PON and NAV, which have relatively high and low coverages, this relationship was not significant ( $F_{1,4}=0.24$ ,  $p=0.672$ ). Therefore, although we subsampled the data to remove a potential relationship between coverage and variation, we found a negative relationship after subsampling.

Population structure was estimated using Nei's genetic distance based on allele frequencies. This population structure resembled the geographic distances between the populations. For instance, the LE-PR and PON-VDV pairs are geographically close together, which is comparable to their genetic distances (**Figure 4A**). The genetic differences between all populations and pairwise  $F_{st}$  values between LAN-NAV, LE-PR and PON-VDV indicated that LAN-NAV populations were genetically much more different compared to the differences between the LE-PR and PON-VDV pairs (**Figure 4A**). The genetic distance between the LAN-NAV populations is similar to the distance between LAN or NAV and any other populations or exact genetic distances). A larger proportion of loci showed high  $F_{st}$  values between LAN-NAV, compared to the LE-PR or PON-VDV pairs (**Figure 4B**). For instance, while 1.78, 1.71 and 1.64% of all loci were considered variable in pairs LAN-NAV, LE-PR and PON-VDV, a much larger proportion of the  $F_{st}$  values were 1 between LAN-NAV (0.27, 0.020 and 0.058% showed an  $F_{st}$  value of 1 for LAN-NAV, PON-VDV and LE-PR resp., see also **Figure 4B**). Because we are interested in convergent allele frequency differences between ecotypes within pairs that are genetically closely related, we only tested this using the LE-PR and PON-VDV pairs.

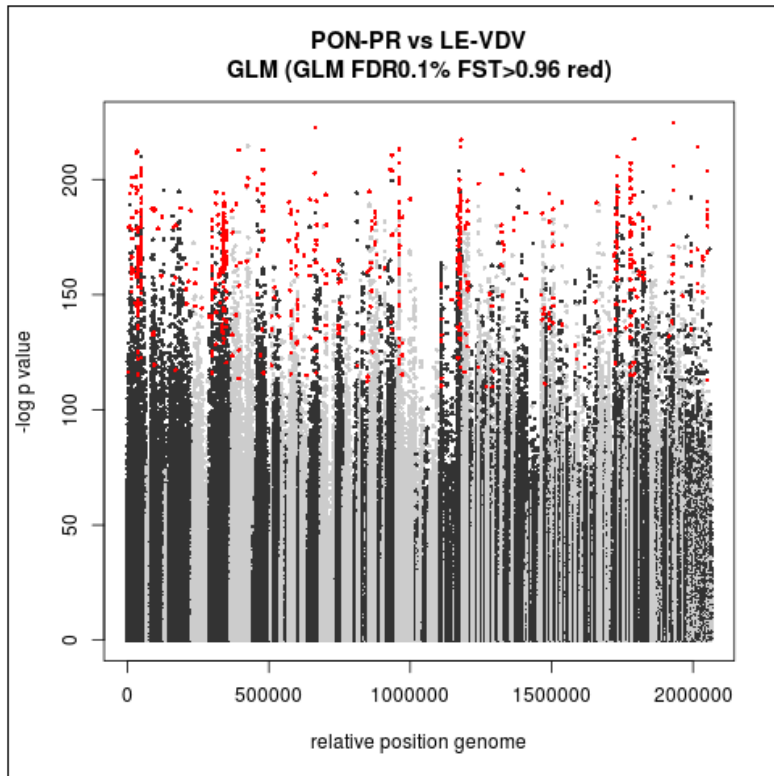


**Figure 4:** Genetic differentiation between (A) all populations using ward clustering with Nei's genetic distance and (B) pairwise  $F_{st}$  values between populations LAN-NAV, PON-VDV and PR-LE. The X-axis indicates the quantiles for  $F_{st}$  distributions taking all the positions in the genome into account. Because only variable loci are shown, and 1.78 % of the loci between LAN-NAV pair are variable, the X-axis starts at 0.9822.

Of all inspected loci, 1,985,394 (0.89% of total number of loci) were variable between any of the four populations, and were considered for Generalized Linear Model (GLM) testing contrasting LE-VDV against PR-PON. A P-value of  $e^{-108.71}$  was determined as the 0.001 FDR cut-off (contrasting LE-PON against PR-VDV, **see method**). In total 5190 loci had a lower P-value and therefore showed indication of significant (with an FDR of 0.001) convergent allele frequency differences between ecotypes LE-VDV and PR-PON.

Since we wanted to follow up a feasible number of these loci to test in an *A. thaliana* T-DNA insertion mutant experiment, we selected the significant loci with very high  $F_{st}$  values ( $F_{st} > 0.96$ ) from both LE-VDV and PR-PON- comparisons, indicating high divergence between the populations. This resulted in 452 loci with convergent and highly differentiated allele frequencies between ecotypes (as indicated in red in **Figure 5**). These loci were annotated by function and effect (i.e. non-coding vs coding, and when coding synonymous vs non-synonymous, see **TABLE S5** for all loci in coding regions). In total 117 loci were found in coding regions, of which 69 represented non-synonymous substitutions. These loci were further annotated based on published data

and for the orthologues of 13 genes we ordered *A. thaliana* T-DNA insertion knockout lines (**Table 4**).



**Figure 5:** Manhattan plot of Generalized Linear Model GLM  $-\ln$ -transformed p-values. Black and grey dots distinguish subsequent contigs, while red dots indicate loci that are significant in the GLM (FDR 0.001) and have a  $F_{ST}$  higher than 0.96 in both LE/VDV and PR/PON population comparisons. Relative genome position is calculated from ordered contigs (total number of covered loci is 222.399.876).

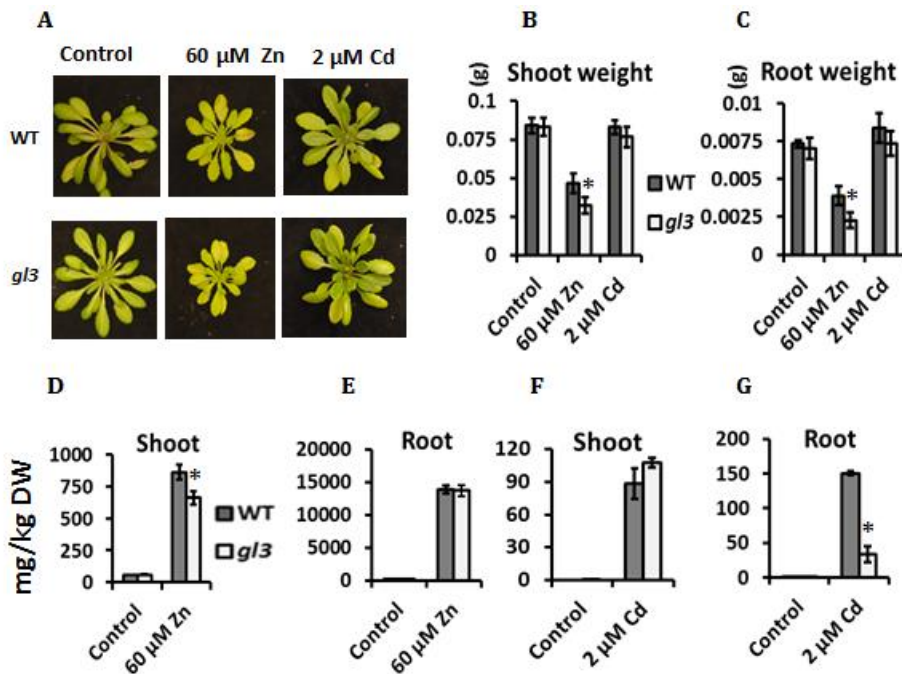
**Table 4:** Candidate polymorphisms for adaptation. The 13 candidate loci were selected from a list of 69 fully divergent loci when comparing NM with M populations (LE/VDV vs PR/PON) (nucleotide identities are indicated), causing non-synonymous differences in protein coding sequences (amino acid (AA) changes between VDV/LE and PON/PR are indicated). The orthologous *A. thaliana* gene is indicated with ID number and the predicted molecular function as listed in TAIR ([www.arabidopsis.org](http://www.arabidopsis.org)). Only the mutants indicated in **bold** showed a mutant phenotype when exposed to excess Zn or Cd (see Figs 6 and 7).

LE/VDV	PR/PON	AA change	Gene	Molecular Function
C	A	A->S	AT5G11700	protein glycosylation
G	A	S->N	AT5G14700	lignin biosynthetic process
T	G	S->R	AT1G06780	GAUT6, carbohydrate biosynthetic process, pectin biosynthetic process
C	T	A->V	AT3G21295	Tudor/PWWP/MBT superfamily protein, unknown function
A	T	I->F	AT3G21690	MATE efflux family protein, drug transmembrane transport, proline transport
A	T	V->D	<b>AT1G29740</b>	<b>Leucine-rich repeat transmembrane protein kinase, transmembrane receptor protein serine/threonine kinase activity</b>
A	G		AT4G16630	DEA(D/H)-box RNA helicase family protein
T	G		AT5G46860	VAM3, vacuole organization, amyloplast organization, Golgi vesicle transport,
G	A	H->Y	AT1G45145	ATTRX5, response to cadmium ion, cell redox homeostasis
C	A	H->G	<b>AT5G41315</b>	<b>GL3, epidermal cell fate specification, templated, trichome branching</b>
C	T	G->E	AT1G50200	ALATS, Response to cadmium ion, response to salt stress
T	C	K->E	AT5G62730	Major facilitator superfamily protein
G	A		AT5G63160	BT1, response to hydrogen peroxide, response to salicylic acid

### Validation of candidate genes under selection

To establish the potential function of these genes in metal tolerance or accumulation, the knockout mutants of *A. thaliana* for the orthologous genes were identified and hydroponically tested under excess Zn (60  $\mu$ M) and Cd (2  $\mu$ M) conditions and compared to control conditions (2  $\mu$ M Zn and no Cd). In order to evaluate potential differences in tolerance between the mutants and Col-0 wild type (WT), the shoot and root dry biomasses were measured. Only two mutants showed metal-specific differences when compared to the WT plants. The *glabrous3* mutant (*gl3*) showed a smaller rosette size and significantly reduced shoot and root dry weights, in comparison with WT, under excess Zn exposure (**Figure 6a-c**). Consistent with these results, the tolerance Index for both shoot and root was also significantly reduced in *gl3* (**data not shown**). Additionally, under excess Zn, leaf chlorosis was more prominent in this mutant (**Figure 6a**). These results indicate that the *gl3* mutant is

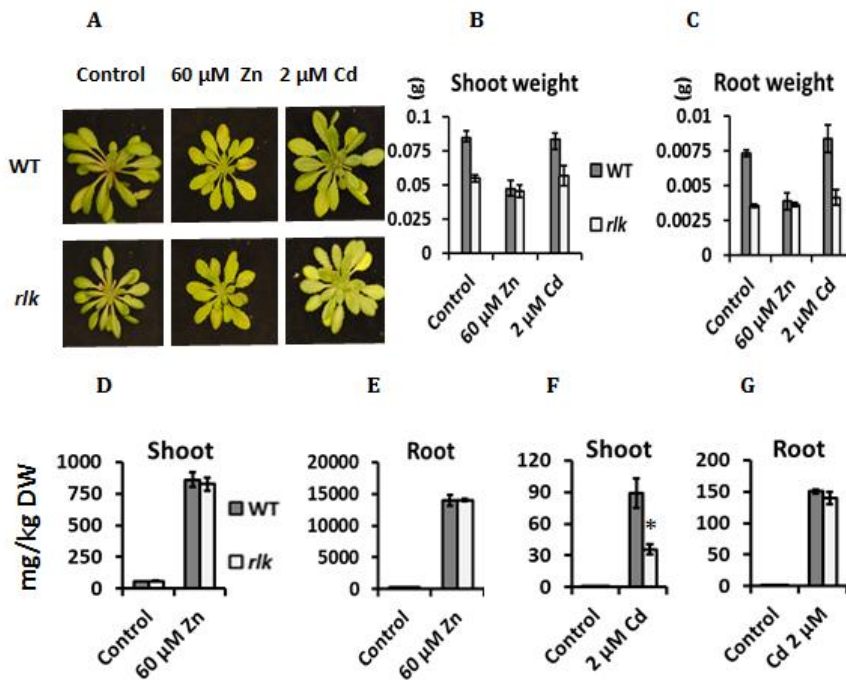
more sensitive to excess Zn than Col-0 WT plants. Remarkably, this Zn sensitivity was associated with a significantly lower Zn concentration in the shoot (**Figure 6d**). Although no differences in the shoot and root biomass were observed under Cd exposure, Cd accumulation in root was significantly reduced in *Atgl3* (**Figure 6g**), resulting in a significantly enhanced shoot-to-root Cd concentration ratio, in comparison with Col-0 (**data not shown**).



**Figure 6:** Rosette appearance and biomass of *A. thaliana gl3* mutant and Col-0 wild-type (WT) plants under control (2  $\mu$ M  $\text{ZnSO}_4$ , no Cd), excess Zn (60  $\mu$ M  $\text{ZnSO}_4$ ) and Cd (2  $\mu$ M  $\text{CdSO}_4$ ) treatments. (A) Rosettes of Plant just prior to harvest. (B) Dry biomass (mean  $\pm$  SE; in grams) of shoots and (C) roots in of WT (dark grey bars) and *gl3* mutant (light grey bars) plants. (D-G) The Zn or Cd concentrations (in mg/kg dry weight; DW) in shoot and root parts of these plants. Plants were grown for two weeks in hydroponic nutrient solutions. Six to nine plants were analysed per treatment level. Significant differences between mutant and wild types within treatments ( $p < 0.05$ ) are indicated with an asterisk (\*).

The biomass of the *A. thaliana* mutant in a Leucine-Rich Repeat receptor-like kinase gene *At1g29740 (rlk)* appeared to be hardly affected by Zn and Cd treatments, in contrast to the biomass of WT plants, entirely owing to its reduced growth under

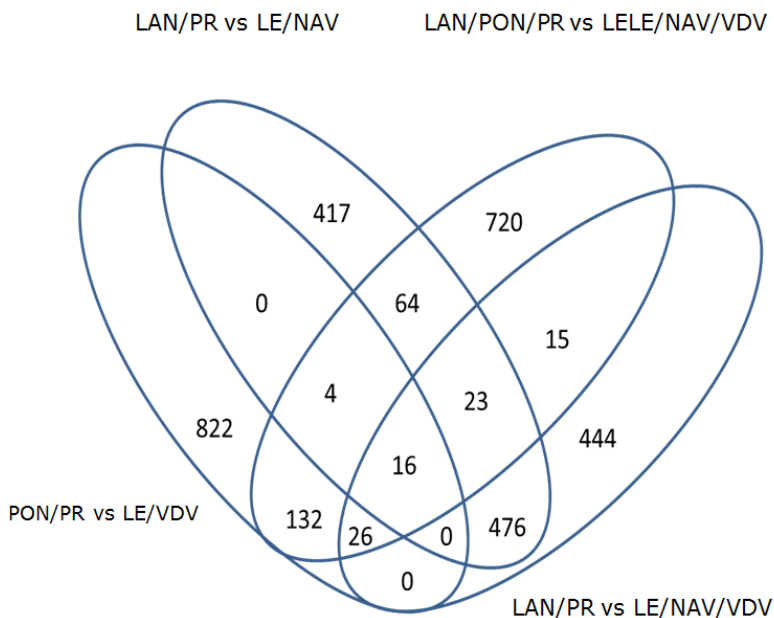
control conditions (**Figure 7b,c**). The root and shoot Zn concentrations were not different from wild type under excess Zn treatment (**Figure 7d, e**). Under Cd exposure, the Cd concentration in shoot, but not in root was significantly lower than in WT (**Figure 7f, g**). Consistent with this result, the shoot to root translocation ratio was also significantly reduced compared to WT. In addition, the mutant developed stronger leaf chlorosis than WT under Cd exposure (**Figure 7a**). The detailed phenotypic and statistical analysis results can be found in **Table S5**.



**Figure 7:** Rosette appearance and biomass of *A. thaliana* *rlk* mutant and *Col-0* wild-type (WT) plants under control (2 μM ZnSO<sub>4</sub>, no Cd), excess Zn (60 μM ZnSO<sub>4</sub>) and Cd (2 μM CdSO<sub>4</sub>) treatments. (A) Rosettes of Plant just prior to harvest. (B) Dry biomass (mean ± SE; in grams) of shoots and (C) roots in of WT (dark grey bars) and *rlk* mutant (light grey bars) plants. (D-G) The Zn or Cd concentrations (in mg/kg dry weight; DW) in shoot and root parts of these plants. Plants were grown for two weeks in hydroponic nutrient solutions. Six to nine plants were analysed per treatment level. Significant differences between mutant and wild types within treatments ( $p < 0.05$ ) are indicated with an asterisk (\*).

### Additional comparisons

As described above, the PON population did not show typical metallicolous phenotypes, and hence, other comparisons next to the LE-VDV vs. PR-PON comparison might be informative. Therefore, a GLM was performed on other population contrasts. We compared the LE-VDV against PR-PON comparison to an analysis of the complete set (all 6) of populations, using ecotype as explaining factor. Of the 1000 most significant loci of both analyses, 178 ( $132 + 4 + 26 + 16$ ) loci overlapped (**see Figure 8**). Furthermore, when we compared the significant loci from the analysis with all populations with the analysis in which PON was removed only 80 ( $26 + 16 + 23 + 15$ ) loci overlapped. This is in stark contrast with the two analyses of PR-LAN vs. LE-NAV compared to PR-LAN vs. LE-NAV-VDV, which resulted in an overlap of 515 ( $476 + 23 + 16$ ) loci. This indicates that the PON population is likely to be very different from the other metallicolous populations, both phenotypically and genetically.



**Figure 8:** Venn diagram of loci that overlap when the 1000 most significant loci were compared using different sets of populations to test for consistent allele frequency changes using a Generalized Linear Model (GLM).

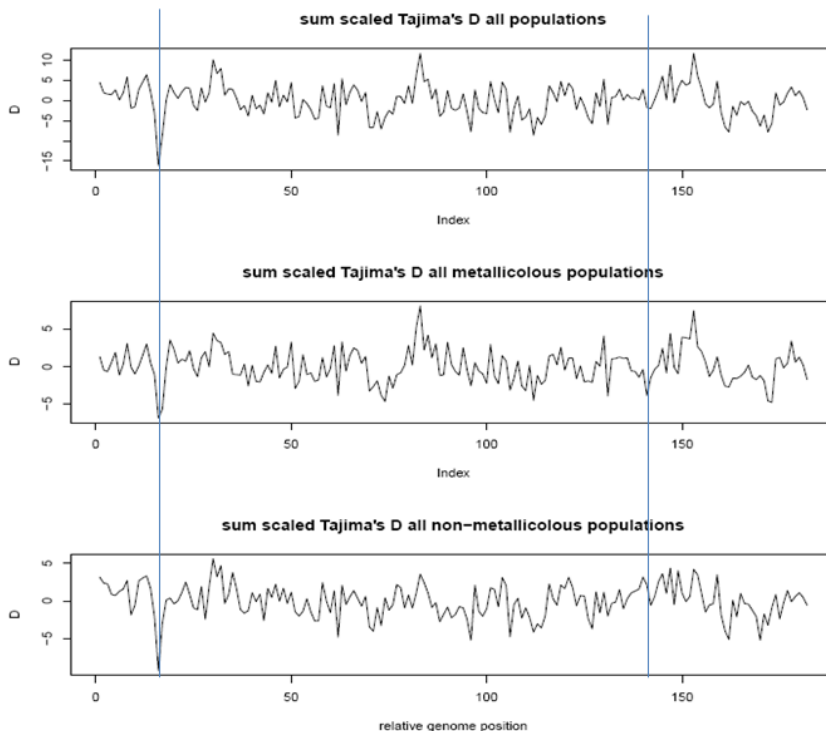


In total 16 loci overlapped between all comparisons (**Figure 8**). One of the SNPs was found in the *OLIGOPEPTIDE TRANSPORTER 3 (OPT3)* gene, which has been identified as a metal ion transporter (Zhai *et al.*, 2014). While the SNP that was highly significant in all comparisons was a non-synonymous SNP, in and around this gene (500 nucleotides up- and downstream) 50 SNPs were found between the populations. Five of these were found to result in amino acid changes in the protein (**Supplement 1**). These amino acid changes occurred in regions of the protein that are not highly conserved when compared to other plant species (**Supplement 1**). Interestingly, four of the five the amino acid substitutions found in OPT3 within *N. caerulescens* were also found to be present between different species of plants, including non-hyperaccumulators.

### Genome wide Tajima's D

We quantified Tajima's D in non-overlapping 500 kb regions. Tajima's D quantifies the difference between two measures of genetic variation, the number of segregating sites and nucleotide diversity. Under neutral evolution, these are expected to be equal, while under directional selection (and in other cases) Tajima's D is expected to become negative (Charlesworth and Charlesworth, 2000). Compared between populations over all windows in the genome, Tajima's D was not different/. This indicates that there are no differences in neutral processes causing a genome wide pattern of differentiation between the populations. To be able to compare between regions in the genome we scaled the values of Tajima's D and summed the scaled values for all populations, as well as metalicolous and non-metallicolous populations (**Figure 9**) specifically. This quantification indicated multiple low values in several places of the genome. One of the windows showed an extremely low value which corresponded with the position of the gene *RLK*. This was the only gene tested using T-DNA lines for which the amino acid change was predicted to be detrimental, using the PROVEAN protein prediction statistics. Furthermore, low values of Tajima's D were found for both metalicolous and non-metallicolous populations indicating directional selection in both directions (for completely divergent alleles). Another region in the genome (Contig1 region 1–1.5 MB) (**Figure 9**) showed by far the most

extreme negative value for Tajima's D. Since this region does not contain a locus of interest from the GLM-Fst analysis we looked at Tajima's D in smaller windows to potentially narrow down the region of interest. For smaller window sizes in this region lower Tajima's D values were found and a window of 31.25 KB between position 1218750 and 1250000 was indicated as the window with a value of -0.98137 for population LE (**Figure S3**). This window has the lowest value of Tajima's D in the larger window of 1 – 1.5 MB on contig 1 for five of the six populations. Of the four genes that are in this region three have been annotated to have homologs in *Arabidopsis* (**Figure S4**). These are AT3G20240 (mitochondrial substrate carrier), AT3G20250 (*PUMILIO5*) and AT3G20260 (unknown function) of which the *PUMILIO5* gene is known to regulate gene expression after stresses (Huh and Paek, 2014).



**Figure 9:** Variation in Tajima's D calculated from the sum of scaled Tajima's D values over the genome. The vertical line indicates the window in which RLK (one of the candidate genes from GLM – Fst analysis) was found.

Upon inspection of the pairwise  $F_{st}$  values between population pairs at all loci in this region, we found several loci with highly differentiated allele frequencies. However, only one was found to be non-synonymous and was found in a gene which has no homolog in any other species. Other SNPs with similar differentiation levels were either intergenic or intronic (**Figure S5**). We inspected the  $F_{st}$  values for the three pairs of populations, since this region was not very significant in the GLM, which might be due to a lack of divergence between either the PR-LE or PON-VDV pair. While this locus was differentiated between PR and LE and LAN and NAV, it was not in the PON-VDV pair, which suggested that the PON-VDV pair has differential allele frequency changes compared to the other pairs of populations.

Other regions in the genome show positive values of Tajima's D. Of the three regions with the highest Tajima's D values, one contained a homolog of the gene U2 small nuclear ribonucleoprotein (**Table S6**). While higher values for Tajima's D indicate balancing selection, the locus that was highly significant in the GLM and this gene shows complete divergence between ecotypes.

## Discussion

Using pool-seq (Schlötterer *et al.*, 2014) in natural populations of the metal accumulator *N. caerulescens*, we have found a large number of SNPs that show consistent allele frequency differences between geographically close metalicolous (M) and non-metallicolous (NM) populations. Among these SNPs are complete divergent positions and are predicted to be non-synonymously substituted. Furthermore, we established the function of some candidate genes under selection in relation to Zn/Cd tolerance and/or accumulation based on the analysis of knockout mutants of their *A. thaliana* orthologues. For two mutants, phenotypic changes were observed that resembled differences found between *N. caerulescens* ecotypes when grown under metal exposure conditions.

The two genes that have been studied in this respect are gene At1g29730, encoding a receptor-like protein kinase (*RLK*), and *GL3*, encoding a basic helix-loop-helix domain

transcription factor protein. Both genes seem to have some relation to root hair formation. For *GL3* this is the most obvious.

*GL3* in *A. thaliana* and many other dicot species (Bernhardt *et al.*, 2003, 2005), specifies root epidermal cell fates to form a root hair or not (Bernhardt *et al.*, 2003; Wei Yang *et al.*, 2008). Mutants show an increase in root hair formation, while over-expression reduces root hair development (Bernhardt *et al.*, 2003, 2005). In the Zn/Cd exposure experiment, *gl3* mutant plants responded very differently to excess Zn reflected in a reduced dry biomass and tolerance index, when compared with WT plants, and a lower shoot Zn concentration (**Figure. 6**). The increased hair root formation in the mutants may in turn increase the Zn or Cd uptake or mobility, or reduce the Cd sequestration in the root. The phenotypes for Zn and Cd transport seem to be very different, one explanation would be that Zn uptake may go through root hair, but Cd may not, therefore, the increased hair root formation in the mutants may in turn increase the Zn uptake. If increased root hair formation causes the sensitivity to excess Zn, then the allelic differences between the two *N. caerulescens* ecotypes might as well be causing differences in Zn sensitivity. However, when growing the two *N. caerulescens* ecotypes on agar plates, we failed to see an obvious difference in root hair patterning (**data not shown**). However, even when comparing *A. thaliana* WT and *gl3* knock-out plants, the difference was not striking and mainly visible in the upper part of the root. For *N. caerulescens* the phenotypic differences are unlikely to be equally strong, as both alleles are likely to be still functional, and differences may not be expressed in agar-grown plants.

However, the *gl3* mutant showed comparable shoot phenotypes in Zn grown *A. thaliana* plants, as we observed in the non-metallicolous *N. caerulescens* populations (NAV/VDV/LE) regarding development of chlorosis, reduced biomass and tolerance index and an increased Zn translocation ratio at excess Zn. Even for Cd exposure, such similarities were observed, although biomass of the *gl3* mutant was not affected. The Cd translocation was increased though, due to the lower accumulation of Cd in the root, and such a phenotype was also apparent in the non-metallicolous *N. caerulescens* populations at the lowest Cd concentrations tested. Between the non-metallicolous

populations (VDV/LE) and their metallicolous neighbours (PON/PR), a SNP was found (from C to A, resp.) causing a non-synonymous substitution from H to G in a conserved region in the protein. Since the metallicolous allele (A) is not found at all in the non-metallicolous populations, though there is clearly shared ancestry and probably even recruitment of metal tolerant genotypes at metal contaminated sites from local non-tolerant populations (Pauwels *et al.*, 2005; Besnard *et al.*, 2009), the metallicolous allele, conferring an altered protein function, appears to have been established after the metallicolous populations diverged from the non-metallicolous ones, which would suggest a high selection pressure on the 'new' allele. Moreover, it would have been independently evolved at least two times, suggesting that the site of mutation is important.

The leucine-rich-repeat class of receptor-like kinase (*RLK*)-encoding genes represents the largest class of putative receptor-encoding genes in the *A. thaliana* genome. The best *A. thaliana* orthologue of the *N. caerulescens* candidate gene under selection is AT1G29740. This gene resides in a cluster of four similar *RLK*-encoding genes, also including AT1G29720, AT1G29730 and AT1G29750. The only reference we could find was for AT1G29730, which shows reduced expression in roots when plants are subjected to manganese (Mn) deficiency, a condition normally leading to enhanced root hair formation (Yang *et al.*, 2008). Gene AT129720 is very poorly expressed in roots; AT1G29750 showed no response, and AT1G29740 was not included on the Affymetrix ATH1 micro-array used for the gene expression analysis. Upon Cd exposure, the shoot Cd concentration of the *at1g29740 rlk* mutant is significantly reduced in comparison with WT. This could reflect a possible link with Fe homeostasis. Mn deficiency causes down-regulation of *IRT1* (Yang *et al.*, 2008), encoding the main Fe-uptake transporter of *A. thaliana* (Vert *et al.*, 2002) but an increase in root Fe concentration, which is counterintuitive. Such an increase could be caused by alteration of root hair formation, linking reduced expression of AT1G29740 with enhanced root hair formation and reduced *IRT1* expression. If Cd needs *IRT1* to be taken up and altered root hair number has no effect, the absence of AT1G29740 function and corresponding decrease in *IRT1* expression will be sufficient to reduce

Cd uptake in roots. This is not easily measured in roots grown hydroponically in high Cd medium, but will be apparent in reduced shoot concentrations. The seemingly high Zn-tolerance ability of the mutants might be caused by the same mechanism leading to lower Zn uptake in roots, but as Zn will be needed in shoots and Cd not, the plants will make an effort to get sufficient Zn to the shoots.

In comparison with the study of Turner et al. (2010) comparing serpentine and granitic *Arabidopsis lyrata* populations by high Fst value (0.403~ 0.565), seven responsible genes for serpentine adaptation were identified, including a magnesium transporter under selection to the 3' UTR of a magnesium transporter, and a calcium channel protein with one particular amino acid substitution. In the present study, we found almost maximum Fst values (Fst > 0.96) between metallicolous and non-metallicolous populations. This is most probably because calamine soil exerts a stronger selection pressure than serpentine soil regarding metal toxicity (Brady et al., 2005) in particular because adaptation to calamine soil in *N. caerulea* is probably not older than adaptation to serpentine in *A. lyrata* species, even when considering that metallicolous *N. caerulea* populations originated after the glacials (Pauwels et al., 2012; Pollard et al., 2014), rather than due to Roman mining activities (Domergue et al., 2006). The latter is in fact unlikely, because the Romans largely exploited natural superficial ore outcrops, which probably already hosted natural metallicolous populations (van der Ent et al., 2013). The mating system and population size can also affect the population genetic structure. *A. lyrata*, as a self-incompatible, perennial species (Ross-Ibarra et al., 2008), is expected to have higher variability than the self-compatible, usually self-pollinating, biennial/perennial species *N. caerulea* (Riley, 1956; Koch et al., 1998; Dubois et al., 2003; Mousset et al., 2016). Furthermore, the outcrossing rate in *N. caerulea* was found to be higher in metallicolous populations (Koch et al., 1998; Mousset et al., 2016). Given the shorter generation time and the higher selfing rate in *N. caerulea*, it may be expected that local adaptation proceeds faster in this species in comparison with *A. lyrata*. The founder effects due to mine-to-mine seed transport by mine workers may also have promoted population differentiation, at least with regard to metal tolerance, between

metallicolous and non-metallicolous of *N. caerulea*, and decreased genetic variation in metallicolous populations.

Because the genetic distance between the geographically distant LAN and NAV populations was large, we tested for convergent evolution in the other populations (LE-PR and VDV-PON). Next to locus-specific divergence, we also quantified Tajima's D to detect mutations under natural selection (Lohmueller et al. 2011) in large non-overlapping windows, while including all populations. Whereas the LAN-NAV populations are very different from each other and all other populations, there were indications of (directional) divergent selection in the same genomic regions among all of the populations. Among the five windows with lowest values (out of 150) we found AT1G29740, which was the only gene that was predicted to carry an amino acid change predicted to cause a detrimental effect in the metallicolous populations, which resulted in phenotypes in *N. caerulea* similar to those in the corresponding T-DNA insertion mutant in *A. thaliana*. Next to this, all the populations showed a very low value of Tajima's D at the start of Contig 1 within the window of 1 – 1.5 MB region (**Figure 9**). While no initial candidate locus was located in this region, we inspected this region for additional candidate genes. Only one locus was found to be differentiated between pairs of the populations LAN-NAV and LE-PR, but not VDV-PON. While the SNP variation was found to alter the amino acid in this predicted protein, no homolog of this gene in any other species is known. However, this gene is expressed in *N. caerulea* and, therefore, an interesting candidate for further research.

While the populations LAN and NAV were genetically more different from each other compared to the geographically close pairs PON/VDV and PR/LE, the phenotypic patterns for Zn and Cd accumulation in PON/VDV were quite different from both the pair PR/LE and the geographically distant populations LAN/NAV, the difference mainly being that the metal tolerance and accumulation of PON is not very different from the non-metallicolous populations except at the lowest exposure concentrations. In addition, PON was barely or not more tolerant to Zn and Cd than LE and VDV. This may be taken to indicate that PON originated quite recently from a local non-

metallicolous population, possibly because mining started only around the 1900s at this site and lasted only for at most 50 years (Martos *et al.*, 2016). However, although the PON and VDV populations are not so different in metal accumulation and tolerance, they are genetically equally divergent from each other as LE and PR. Therefore, a likely alternative scenario is that it is not the time of divergence that is different, but rather the selection pressure.

In our experiment, NAV was more tolerant to Cd and Zn than PON, and not significantly different from PR for Zn tolerance. However, in another experiment, using a different methodology, determining the maximum metal concentration on which roots grow upon weekly transfer to increased metal concentrations, NAV was found to be more sensitive to Cd and Zn than PON and PR (H. Schat, unpublished results). Based on our biomass data, it seemed that NAV did not grow well in multiple experiments under the control conditions in our hydroponic system, thus, the Zn or Cd-tolerance of NAV might have been over-evaluated in this study. The unambiguously tolerant populations, PR and LAN, both accumulate less Cd and Zn than all the other populations, suggesting that their high levels of tolerance are in fact due to metal exclusion.

Due to the different degrees of variation in either accumulation, or tolerance, or metal root to shoot translocation between the pairs PON/VDV, PR/LE, and LAN/NAV, to identify the loci under selection responsible for a specific trait, we performed the additional comparison: LAN/PR vs LE/NAV, LAN/PON/PR vs LE/NAV/VDV, PON/PR vs LE/VDV and LAN/PR vs LE/NAV/VDV (**Figure 8**). This resulted in a clearly different overlaps of the 1000 most significant polymorphisms in those pairs. In a comparison between metallicolous and non-metallicolous populations without PON, we found more overlapping loci potentially responsible for high Zn/Cd accumulation or tolerance than with PON (583 vs 178 out of 1000, respectively), which confirming that PON is an outlier among the metallicolous populations (**Figure. 8**) However, we still identified 16 overlapping SNPs out of 1000 in all of the comparisons. One of these polymorphisms located in the metal homeostasis gene *OPT3*, which encodes a phloem-specific iron transporter that is essential for systemic iron signalling and



redistribution of iron and cadmium in *A. thaliana* (Zhai *et al.*, 2014). Large variations in the region ( between 500 nucleotides up- and downstream of this gene) were detected between populations, Five of these were found to result in amino acid changes in the conserved region of the protein (**Supplement 1**). This gene was also found to be differently expressed under metal stress in *N. caerulescens* (Halimaa *et al.*, 2014a; Lin *et al.*, 2014). In addition, we found also a non-synonymous SNP in the transcription factor *WRKY72*, which has a role in plant abiotic stresses like Pi deprivation, salt and drought (Yu *et al.*, 2010; Chen *et al.*, 2012b; Niu *et al.*, 2012). Thus, it is possible that it regulates the gene expression level under Zn/Cd stress conditions. This has to be investigated in future experiments.

## Conclusion

We tested for consistent allele frequency changes between geographically close pairs of metallicolous and non-metallicolous *N. caerulescens* populations, and found extremely divergent loci. Some of these loci were verified to affect metal-related phenotypes by examining the corresponding *A. thaliana* T-DNA knockout mutants. We have, therefore, been able to identify loci that are under selection in different populations of the same ecotype, suggesting convergent independent evolution at different localities. The stress-responsive transcription factor *WRKY72*, and the metal homeostasis gene *OPT3*, and one *N. caerulescens* specific gene under directional selection were detected in the additional analysis. However further validation of these genes would allow of a better understanding of their role in relation to Zn/Cd accumulation and tolerance in *N. caerulescens*. The SNP library created in this study will provide a valuable resource for future investigation involving *N. caerulescens* and should also be compared to related taxa to achieve a better understanding of evolutionary scenarios.

## Materials and methods

### Plant materials and growth conditions

From each natural *N. caerulea* population sampled in the field (**Table. 1**), seeds were collected from at least 30 individual plants, interspaced by at least a few meters. The seeds of Navacelles and Lanestosa were kindly offered by Thibault Sterckeman (Lab. of Soil and Environment, INRA, France) and Oihana Barrutia (Lab. of Plant Physiology, Dept. of Plant Biology and Ecology, University of the Basque Country (UPV/EHU) respectively. The seeds of Prayon and Lellingen were kindly provided by Maxime Pauwels (Lab of Genetics and Population Evolution, University of Lille, France). Hereafter, the offspring from each single plant is referred to as a family. We propagated 30 families (5 plants/family) from each population in the greenhouse under long day conditions (16/8 hrs light/dark) with temperatures set at 20/18 °C day/night and humidity set at 70%. Seeds obtained from the plants after propagation for two generations through self-fertilization, were used for phenotyping.

### Metal exposure

To determine the tolerance and accumulation abilities, seeds were germinated in modified half-strength Hoagland medium (Schat *et al.*, 1996) with 0.55% agar. Two weeks after germination, the seedlings were transferred to half-strength Hoagland medium (HM) containing 2  $\mu\text{M}$   $\text{ZnSO}_4$ . After one week, seedlings were exposed to HM supplemented with different Zn (50; 100; 1000  $\mu\text{M}$   $\text{ZnSO}_4$ ) or Cd (0.5; 5; 50  $\mu\text{M}$   $\text{CdSO}_4$ ) concentrations (the latter containing 2  $\mu\text{M}$   $\text{ZnSO}_4$ ) or kept on HM (with 2  $\mu\text{M}$   $\text{ZnSO}_4$ ) as control. Solutions were refreshed once a week. Plants were grown in a climate chamber set at a 16/8 hrs light/dark cycle, a 20/18 °C day/night temperature regime and 70% relative humidity. After three weeks, the shoots and roots of each pot (three plants/pot) were harvested separately for the metal tolerance and accumulation analysis. To analyse Zn and Cd sensitivity of plants, leaf chlorosis was investigated prior to harvesting as described by Assunção *et al.*, (2003).

### Shoot and root metal accumulation assay

For biomass and mineral analysis, three plants were pooled per pot (three pots per

treatment, three plants/pot). To resorb adhering metals, the root system was washed with ice-cold 5 mM PbNO<sub>3</sub> for 30 min. Harvested shoots and roots were dried at 65 °C for 3 days. The samples were digested in a 4:1 mixture of HNO<sub>3</sub> (65%) and HCL (37%) in Teflon bombs at 140 °C overnight and analysed for Zn and Cd concentration using flame atomic absorption spectrometry (Perkin Elmer 1100B) as described by Assunção *et al.*, (2003). Metal concentrations in roots and shoots were calculated in mg/kg dry weight (DW).

### **DNA isolation and pools creation**

Two months after sowing the seeds in the greenhouse (see plant material); very young leaves from one plant per family were collected for DNA isolation. In total, the genomic DNAs from 30 plants per population were extracted using the CTAB method, adapted from Maloof Lab ([http://malooflab.openwetware.org/96well CTAB.html](http://malooflab.openwetware.org/96well%20CTAB.html)). We assessed the samples using the Nanodrop 2000 spectrophotometer (Thermo Scientific) for purity; we used the Q-bit (Invitrogen) to measure the DNA concentrations in the samples. To create a DNA pool for each population, 1 µg DNA per plant was combined, pools were created for each of the six populations and sent for library preparation and sequencing.

### **DNA library preparation**

Genomic DNA from six populations (Lellingen, Lanestosa, Navacelles, Prayon, Pontaut and Vall de Varrados each consisting of 30 individuals per pool) was used for whole genome resequencing analysis. For each sample 1 µg total DNA was sheared using a Covaris E210 device to fragment sizes of approximately 500 to 600 bp peak size and used for whole genome shotgun library preparation suitable for Illumina HiSeq paired end sequencing according to the TruSeq DNA™ Sample preparation LT protocol (Illumina Inc, San Diego CA, USA). After fragmentation, DNA was processed further for end repair, adenylation, adaptor ligation, and final library amplification following the manufacturer's protocol. Final libraries were eluted in 30 µl elution buffer. Library quality was analysed using a Bioanalyzer 2100 DNA 7500 chip (Agilent Technologies)

and quantified using Quant-iT PicoGreen (Invitrogen, Life Technologies) on an X Fluor plate reader system (Tecan).

### **Illumina whole genome re-sequencing**

Indexed libraries were equimolarly pooled and diluted to 6 pM for TruSeq Paired End Read v2 DNA clustering on five lanes divided over two flow cell lanes using a cBot (Illumina). Final sequencing was done on an Illumina HiSeq2000 platform using 101, 7, and 101 cycles for sequencing forward, index and reverse reads, respectively. All steps for clustering and subsequent sequencing were carried out according to the manufacturer's protocols. Reads were de-multiplexed per sample by corresponding index read using CASAVA 1.8 software ([http://cancan.cshl.edu/labmembers/gordon/fastq\\_illumina\\_filter/](http://cancan.cshl.edu/labmembers/gordon/fastq_illumina_filter/)).

### **Bioinformatics**

Read trimming (quality trimming, ambiguous limit 2, and quality limit 0.05) and mapping were done with CLC Genomics software Version 6.0.2. Reads were mapped (with auto detection paired distances, mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.9, similarity fraction 0.9) against a *de novo* assembled genome of *N. caeruleus* that was assembled using the sequenced library of one individual from the 'Ganges' population (noccaea.genome.v103, (Severing *et al.*, In preparation). After alignment, duplicates were removed from the bam-files using Picard tools (1.107). GATK (3.3-0 McKenna *et al.*, 2010) was used to realign around Indel regions. Then bam-files were combined in the mpileup (samtools, v1 .1.1) and synchronized format (Kofler *et al.*, 2011a). Also individual pileup files were produced to estimate  $\pi$  and Tajima's D in 500-kb windows (Kofler, *et al.*, 2011b). As coverage differences between populations were quite large, which affects these measures (Charlesworth and Charlesworth, 2010 (e.g. average coverage for PON was 61 while for NAV 34, **see Table S3b**) we subsampled the individual pileup files to a target coverage of 25 (subsample-pileup.pl, Kofler, *et al.*, 2011b). From the synchronized file allele frequencies were calculated and from these data Nei's genetic distance (Nei, 1972) between all pairs was calculated. A population cladogram was drafted using the ward

method from the `hclust` function in the `stats` package in R (v 3.2.3). Pairwise  $F_{st}$  values were calculated for the paired populations LAN-NAV, LE-PR and PON-VDV (using `fst-sliding.pl`, Kofler *et al.*, 2011a).

Allele counts taken from the synchronized files were used to perform a Generalized Linear Model (GLM) using only bi-allelic read counts. If more alleles were present, the allele with lowest count was omitted. We only used loci that had coverages higher than 25, and with a total count of the minor allele of 4 per population. By k-mer analysis the peak coverage was estimated, by which we could set a maximum coverage (see **Figure S6**). The peak of the distribution was taken as average coverage and multiplied by 2.36 ( $2 * 101 / (101 - 17 + 1) + 1$ , 101 = read length, 17 = kmer). This resulted in maximum cut off coverages for further analyses of 114, 121, 76, 173, 123 and 102 for LAN, LE, NAV, PON, PR and VDV respectively.

Because of the high genetic divergence between LAN and NAV (see results) these populations were omitted for the test of convergent allele frequency changes between ecotypes. Similar to other studies, such as in *Gasterosteus aculeatus* (Hohenlohe *et al.*, 2010) and *Arabidopsis lyrata* (Turner *et al.*, 2010), pairs of closely related populations were classified as different ecotypes. Therefore, to properly test for convergent allele frequency different it is best if populations are relatively closely related to distinguish selection from neutral processes. To test for consistent allele frequency changes we compared the metallicolous ecotype populations PR and PON against the non-metallicolous populations LE and VDV. To be able to determine a false discovery rate (FDR), we compared the p-value distributions using the ecotype as explanatory variable with one in which we estimated the effect of neutral events by contrasting LE and PON against PR and VDV. We determined the 0.001 quantile from the null hypothesis ( $H_0$ ) and called loci significant under the  $H_1$  if the p-value associated with these loci was lower than the determined value for this quantile (called a FDR of 0.001, *sensu* Tobler *et al.* 2014). The significant loci from the GLM analysis were further narrowed down by calculating the average  $F_{st}$  between the two pairs LE-PR and PON-VDV. Loci with a higher average  $F_{st}$  than 0.96 (which allows for complete divergence and +/- some sequencing error) were annotated. We determined whether variants

were within a coding region, and if so, whether nucleotide changes resulted in synonymous or non-synonymous changes. Genes with non-synonymous changes were manually annotated for biological functions of interest, which resulted in 13 genes that were followed-up using *A. thaliana* T-DNA mutant lines. The amino acid substitutions from these 13 loci were predicted and these protein sequences were used as input for PROVEAN (Choi *et al.*, 2012) to predict whether amino acid substitutions had deleterious effects on protein structure.

Next to the above described comparison between LE-VDV and PR-PON, we also performed GLM analyses using all populations (NAV, VDV, and LE against LAN, PON and PR).

### **Phenotypic analysis of *Arabidopsis thaliana* T-DNA insertion knock-out mutant lines**

To determine whether *N. caerulescens* candidate genes have functions related to heavy metal tolerance or accumulation, the corresponding knock-out lines of *A. thaliana* were phenotyped under the same conditions (see above). After 2 weeks of growth in the normal nutrient solution (see above), the plants were exposed to 60  $\mu\text{M}$  Zn or 2  $\mu\text{M}$  Cd. The nutrient solution was replaced once a week. Shoots and roots were harvested after two weeks of exposure. For the ionome analysis, the roots were washed with 1 mM EDTA (pH 8.0) and MilliQ water (Lahner *et al.*, 2003).

### **Statistical analysis**

After log transformation of the raw data, the Regression Mean Square MSR (critical significance level 5%) was used to compare means following a 2-way ANOVA (Sokal and Rohlf, 1995).

**Supplemental Information****Table S1:** Statistical analyses of plant biomass under Zn and Cd treatments in metalicolous and non-metalicolous populations of *N.caerulea*. A log transformed data were used. After two-way ANOVA, individual means were compared using MSR at  $\alpha = 0.05$ .

Variant	Source of variation	df	SS	MS	F	P	Significant
Zn	Zn treatments	3	0.271465	0.090488	16.25593	<.001	***
	populations	5	4.033556	0.806711	144.9228	<.001	***
	Zn treatments x populations	15	0.364095	0.024273	4.36056	<.001	***
	within	48	0.267191	0.005566			
Cd	Cd treatments	3	1.468	0.489333	135.731	<.001	***
	populations	5	4.088319	0.817664	226.8031	<.001	***
	Cd treatments x populations	15	0.382367	0.025491	7.070709	<.001	***
	within	48	0.173048	0.003605			

**Table S2:** Statistical analyses of Zn and Cd concentrations in shoot and root in metalicolous and non-metalicolous populations of *N.caerulea*. A log transformed data were used. After two-way ANOVA, individual means were compared using MSR at  $\alpha = 0.05$ .

Variant	Source of variation	df	SS	MS	F	P	Significant
Zn	Zn treatments	3	2.95717	0.98572	90.08	<.001	***
	populations	5	3.23043	0.64609	59.04	<.001	***
	Zn treatments x populations	15	0.69121	0.04608	4.21	<.001	***
	within	48	0.52525	0.01094			
Cd	Cd treatments	2	0.683962	0.341981	64.43	< 0.001	***
	populations	5	1.140163	0.228033	42.96	< 0.001	***
	Cd treatments x populations	10	0.643323	0.064332	12.12	< 0.001	***
	within	31	0.164546	0.005308			

## Chapter 2

**Table S3a:** Pairwise genetic distances.

Population	LAN	LE	NAV	PON	PR	VDV
LAN	-					
LE	0.00141	-				
NAV	0.00283	0.00273	-			
PON	0.00203	0.00171	0.00309	-		
PR	0.00152	0.00081	0.00271	0.00188	-	
VDV	0.00164	0.00132	0.00266	0.00095	0.00149	-

**Table S3b:** Average coverage of non-overlapping windows (100 kb)

Populations	LAN	LE	NAV	PON	PR	VDV
Coverage	44.67	44.84	34.58	61.85	46.91	39.16
	(0.22)	(0.24)	(0.10)	(0.35)	(0.25)	(0.20)



## Population genomic resequencing of *Noccaea caerulescens*

**Table S4:** Read statistics for the six populations. Mean depth was calculated as the depth in the individual bam files and hence, is the mean was calculated over the loci covered in that particular population. Since the total number of loci varies between populations (between 92.00 in LAN and 9904 in NAV) this mean coverage is inflated in LAN relative to NAV. The corrected average coverage is hence closer between LAN and NAV than the depth calculated using the individual BAM files.

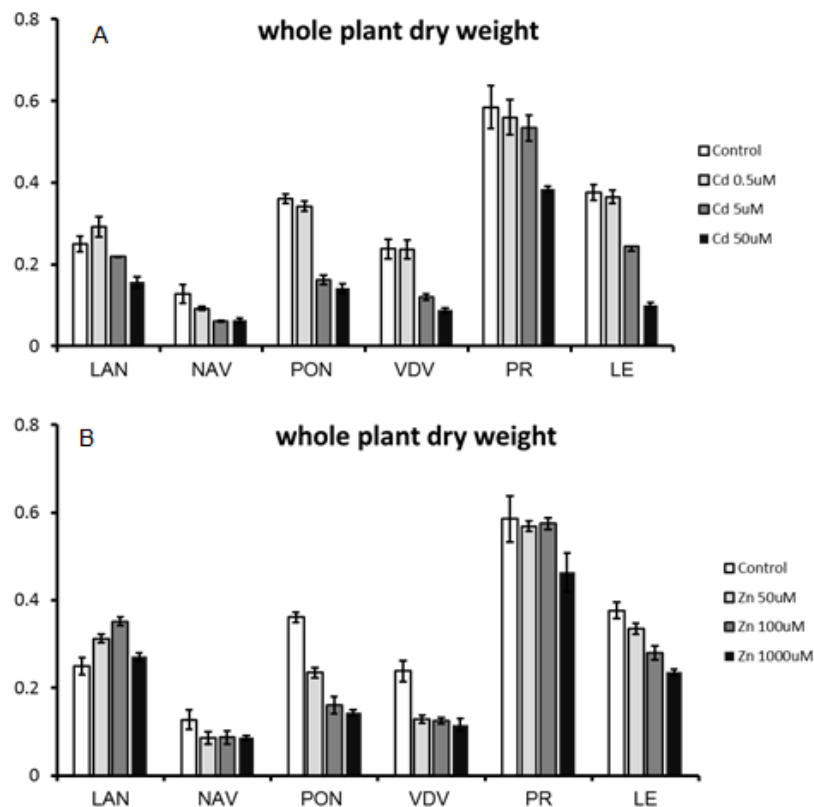
	LAN	LE	NAV	PON	PR	VDV
Raw reads sequenced	174.663.796	171.396.240	116.169.524	228.933.494	177.568.722	142.546.938
Reads aligned to reference	154.559.412	152.525.300	109.029.703	202.651.327	158.898.549	125.540.146
(percentage aligned)	(88.49)	(88.99)	(93.85)	(88.52)	(89.49)	(88.07)
Mean depth	61.2265	62.6224	44.6215	82.4983	63.8506	51.7016
Number of nucleotides from reference covered	204.618,738	205.105.796	220.256.512	200.456.104	205.589.881	204.793.300
(percentage of ref)	(92)	(92.22)	(99.04)	(90.13)	(92.44)	(92.08)
Corrected for total number of loci covered at least once in all populations	56.33	57.75	44.19	74.36	59.02	47.61

**Table S5:** Statistical analyses of shoot and root biomass of knockout mutants and *Col-0* of *Arabidopsis* under control and excess Zn treatments. A log transformed data were used. After two-way ANOVA, individual means were compared using MSR at  $\alpha = 0.05$ .

shoot					
Source	d.f.	s.s.	m.s.	v.r.	F pr.
Genotype	22	2.089179	0.094963	17.78	< 0.001
treatment	1	1.280722	1.280722	239.80	< 0.001
Genotype x treatment	22	0.290876	0.013222	2.48	0.001
Residual	91	0.486013	0.005341		
Total	136	4.146790	0.030491		

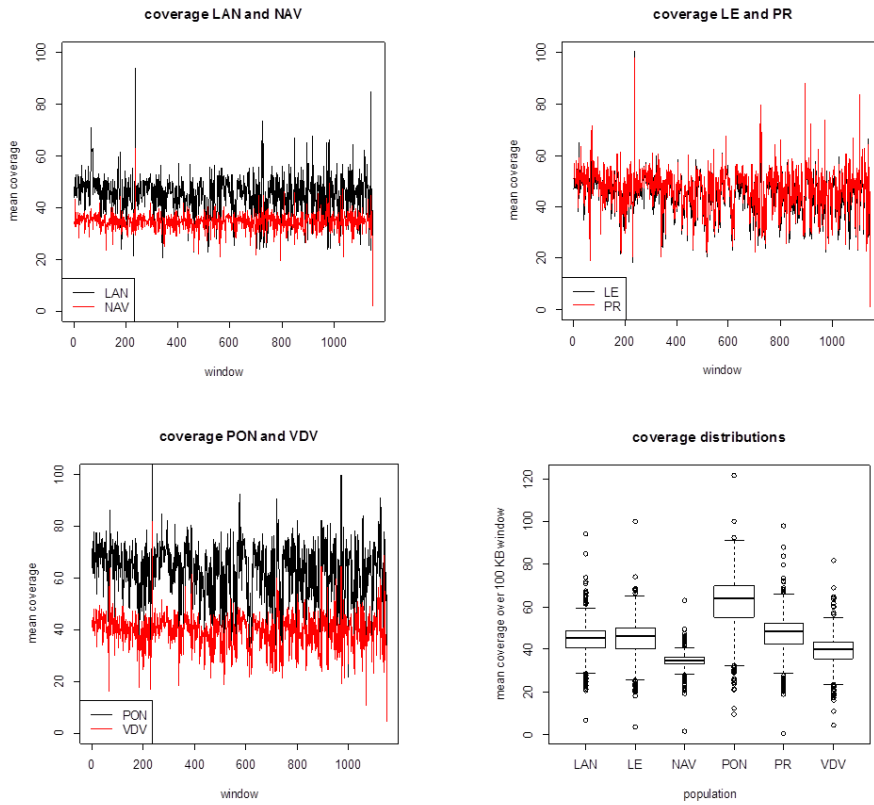
  

Root					
Source	d.f.	s.s.	m.s.	v.r.	F pr.
Genotype	22	1.60559	0.07298	5.74	< 0.001
treatments	1	1.70931	1.70931	134.43	< 0.001
Genotype x treatments	22	0.59488	0.02704	2.13	0.007
Residual	91	1.15712	0.01272		
Total	136	5.06327	0.03723		

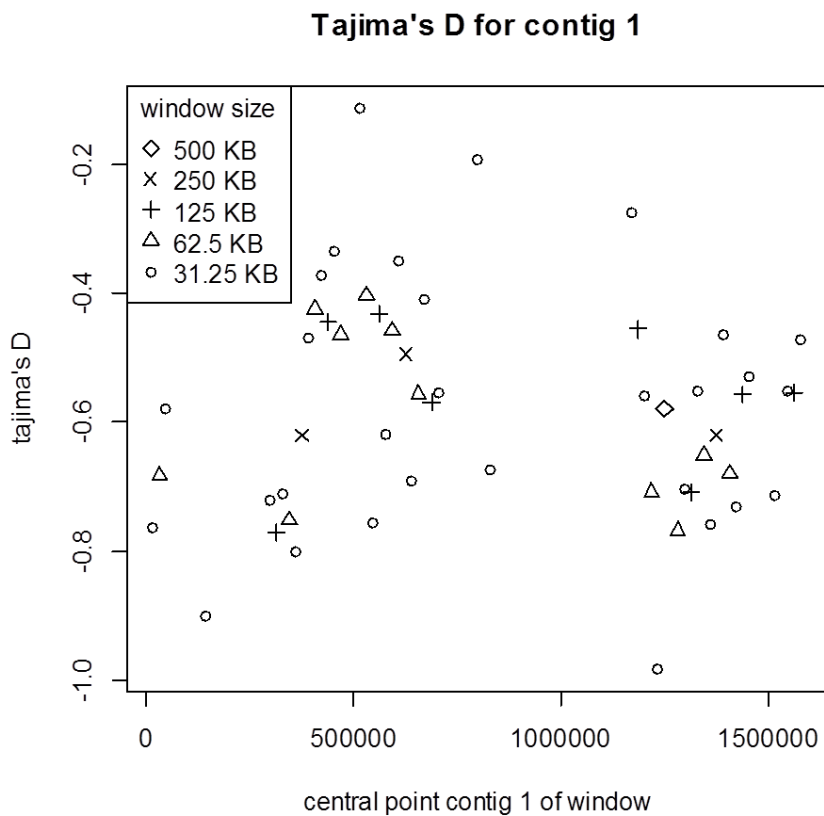


**Figure S1:** The dry biomass of the whole plants under Cd/Zn treatments. (A) Cd, (B) Zn (mean  $\pm$  SE) in roots (black bars) and shoots (grey bars) in the *Thlaspi caerulea* metalliculous accessions Lanestosa (LAN), Pontaut (PON), Prayon (PR) and in the non-metalliculous accessions NAVECELLS (NAV), Lellingen (LE) and (VDV). Plants were grown for 3 wks in nutrient solution supplemented with the metal concentrations ( $\mu$ M) indicated on the abscis. 5–10 plants were analysed per treatment level.

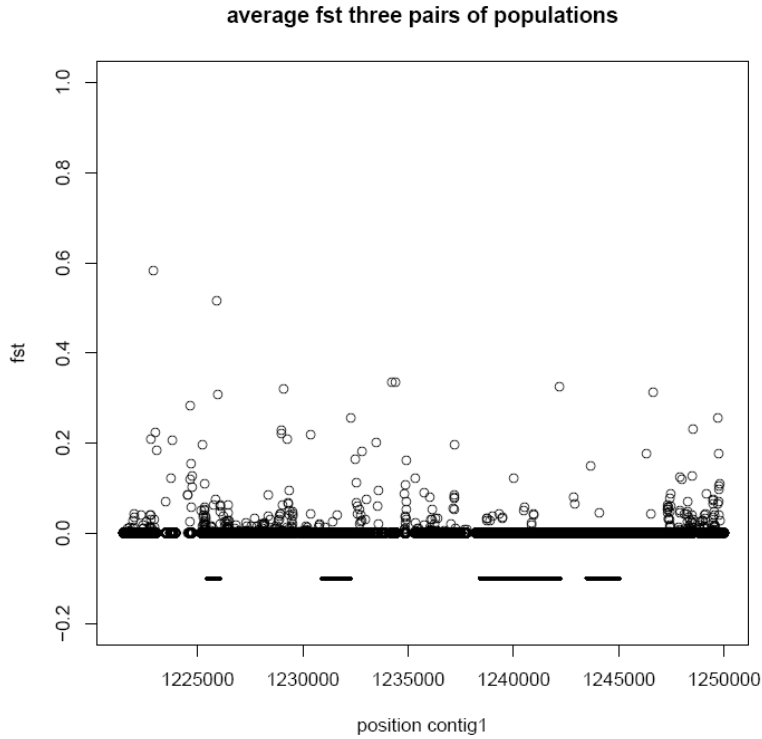
## Population genomic resequencing of *Noccaea caerulea*



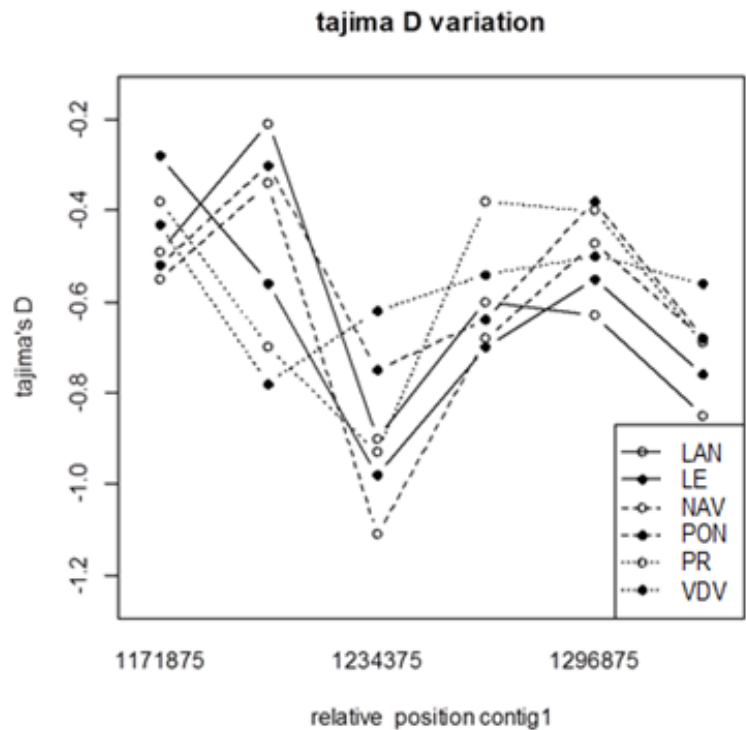
**Figure S2:** Coverage plots by window (100 kb) for the six populations per geographic pair per window (left above LAN (black) and NAV (red), right above LE (black) and PR (red), left below PON (black) and VDV (red) and the distribution of these per population. Only windows in which 95% was counted and windows for which variance between the six populations was less than 400 are shown.



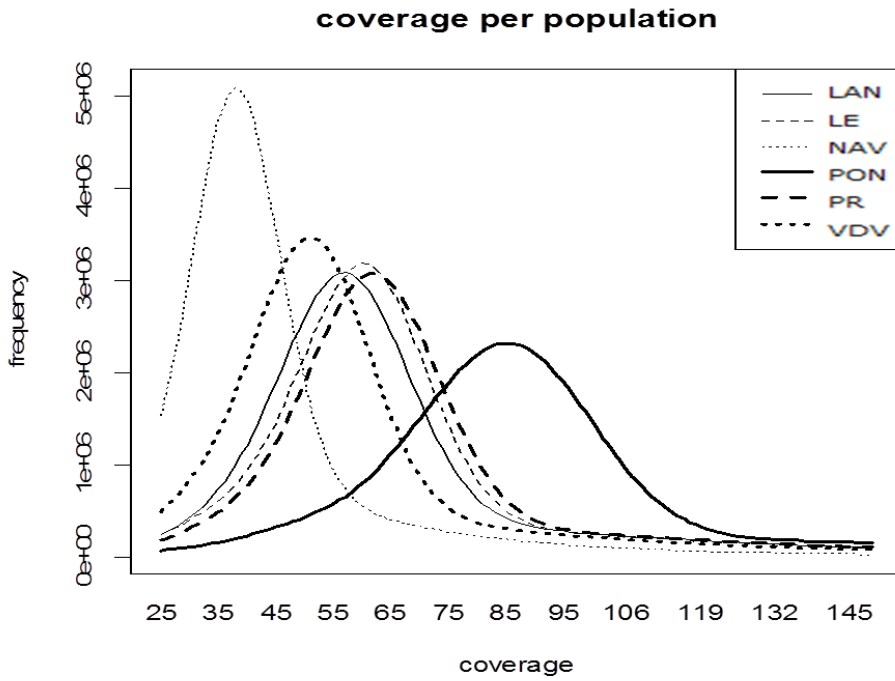
**Figure S3:** Tajima's D dependent on window size for the first part of contig 1. X-axis indicates the central position of the window. Data is for the LE population only.



**Figure S4:** Average *Fst* differentiation for the three paired *Fst* values between LAN-NAV, PON-VDV and PR-LE. The four bars indicated predicated gene positions for from left to right (unknown gene, i.e. no homolog; AT3G20240, mitochondrial substrate carrier family; AT3G20250, pumilio 5; AT3G20260, protein with unknown function (DUF1666)).



**Figure S5.** Tajima's *D* for six populations. 31250 nucleotide windows around the low position window between 1218750 and 1250000, indicated by 123475 on the x axis.



**Figure S6:** Coverage distributions per population. Coverages higher than 25 were used to estimate the distribution of coverages.

**Supplement data Table S5, Table S6 and supplement 1** can be accessed online in the Dropbox by following the link:

**<https://www.dropbox.com/home>**

**Login with username: wylfriend@163.com, password: thesis201609**

**Table S5:** Polymorphisms detected by GLM and FDR analysis between PON/PR vs PON/VDV comparison including the prediction of the amino acid substitution.

**Table S6:** candidates detected in the additional analysis based on phenotypic variation in metal accumulation and tolerance in all population.

**Supplement 1.** OPT3 sequences comparison among six re-sequenced populations of *Noccaea caerulea*

## **Acknowledgements**

This research was supported by a grant from the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research) and the Netherlands Genome Initiative ZonMW Horizon program through Zenith project no. 40-41009-98-11084 and from the Chinese Scholarship Council (CSC). The authors thank Thibault Sterckeman (Lab. of Soil and Environment, INRA, France) who provided the seeds of Navacelles and Maxime Pauwels (Lab of Genetics and Population Evolution, University of Lille, France) for providing the seeds of Prayon and Lellingen, Oihana Barrutia (Lab. of Plant Physiology, Dept. of Plant Biology and ecology, University of the Basque Country (UPV/EHU) providing the seeds of Lanestosa.



## Chapter 3

### **Construction and analysis of a *Noccaea caerulescens* TILLING population**

Yanli Wang<sup>a</sup>, YuFeng Mao<sup>a</sup>, Bastiaan Millenaar<sup>a</sup>, John Danku<sup>b</sup>, David E. Salt<sup>b</sup>, Maarten Koornneef<sup>a,c</sup> and Mark G.M. Aarts<sup>a\*</sup>

<sup>a</sup> Laboratory of Genetics, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

<sup>b</sup> School of Biological Sciences, University of Aberdeen, Cruickshank Building, St. Machar Drive, Aberdeen, AB24 3UU, United Kingdom

<sup>c</sup> Max Planck Institute for Plant Breeding Research, Carl-von-Linné Weg 10, 50829 Cologne, Germany

\*author for correspondence: mark.aarts@wur.nl

#### Financial source:

This research was financially supported by the China Scholarship Council (CSC) and the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research).

## Abstract

*Noccaea caerulescens*, an extraordinary Zn/Cd/Ni hyperaccumulating species, has been used for decades as model plant for heavy metal hyperaccumulation study. To generate the functional genomic tool in *N. caerulescens*, 3000 and 5000 seeds from two sister plants of a single-seed recurrent inbred descendant of Southern French accession St Felix de Pallières (SF) were mutagenized respectively by 0.3 and 0.4% EMS. Two subpopulations consisted of 5000 and 7000 M2 plants respectively for 0.3 and 0.4% EMS were obtained. 0.4% population had a higher mutant frequency and was used for TILLING purpose. Forward screening of 0.4% population by mineral concentration analysis (ionomics) in leaf material for each M2 plant revealed putative mutations affecting the concentration of 20 trace elements. Based on the frequency of monogenic mutants identified in the M2 population, the mutant frequency in this population was evaluated. On average 2.3 knockout mutants could be found upon screening the TILLING population of 7000 M2 plants. Every M2 plant was predicted to carry on average knock-out mutations for 10 genes. For Tilling purpose, the high resolution melting curve analysis (HRM) mutation screening platform was optimized by known point mutation and successfully applied to detected mutations for *bZIP19*, encoding a basic-region leucine-zipper (bZIP) transcription factor, which plays essential in Zn homeostasis in Plant. Of four point mutations in *ncbzip19*, two were non-synonymous substitutions, in respectively plant 73H10 and 29E8; however, these two mutations did not alter the Ionome profiles compared with WT. The EMS mutant population reported herein is the first mutant library established for *N. caerulescens*, which will be suitable for mutation identification by both forward and reverse approach.

## Introduction

In last decades, genome sequencing of many plant species (Kaul *et al.*, 2000; Matsumoto *et al.*, 2005) (review: Bolger *et al.*, 2014) has led to the availability of a large number of gene sequences in public databases which subsequently has encouraged the development of reverse genetics approaches. For this, one tries to associate the mutation of a gene with a non-wild-type phenotype in order to understand the function of the gene in wild type (review: Stemple, 2004). Especially in the model species *Arabidopsis thaliana* (Arabidopsis), random insertional mutagenesis approaches combined with sequencing of the insertion sites (Wisman *et al.*, 1998; Alonso *et al.*, 2003), RNA interference-mediated gene silencing (RNAi) (McCallum *et al.*, 2000a), screening of mutations after chemical (Till *et al.*, 2004) or physical irradiation mutagenesis (Li *et al.*, 2001), have been used as reverse genetics platforms. Insertional mutagenesis and RNAi are widely and very successfully used in Arabidopsis. However those techniques are all based on efficient *Agrobacterium tumefaciens*-mediated transformation, which limits their use because transformation efficiencies are very much genotype, or even organism, specific (Somers *et al.*, 2003; Stemple, 2004; Ko *et al.*, 2006).

*Noccaea caerulescens* (J. and C. Presl) F. K. Meyer (formerly known as *Thlaspi caerulescens*) is a diploid species from the Brassicaceae family, exhibiting extreme tolerance to zinc (Zn), nickel (Ni), cadmium (Cd) or and/or lead (Pb) exposure as well as the ability to hyperaccumulate these metals in its leaves ( Assunção *et al.*, 2003; Nascimento & Xing, 2006; Broadley *et al.*, 2007; Krämer, 2010; Mohtadi *et al.*, 2012). Together with the Zn/Cd hyperaccumulator species *Arabidopsis halleri*, *N. caerulescens* is one of the most prominent plant model systems to study the physiological and molecular basis of heavy metal hyperaccumulation and -tolerance (Krämer, 2010; Hanikenne & Nouet, 2011; Pollard *et al.*, 2014). *N. caerulescens* has received special attention because of its potential application as a source of genes that can be used to engineer phytoremediation or improve food micronutrient content through biofortification (Baker *et al.*, 1994; Assunção *et al.*, 2003c; Peer *et al.*, 2003; Rascio and Navari-Izzo, 2011). Although *N. caerulescens* is a frequently studied metal

hyper-accumulator model species, the use of *N. caerulescens* has been hampered due to the constrained reverse genetics resources.

The transformation efficiency is very low in *N. caerulescens* (Peer *et al.*, 2003; Guan *et al.*, 2008), which makes T-DNA based insertional mutagenesis cumbersome. Alternatively, the functional study of candidate genes in *N. caerulescens* could be achieved by using *Agrobacterium rhizogenes*-mediated root transformation, which has a high efficiency (about 50–90%) but is not heritable and is only limited to the analysis of gene functions in the root but not any other parts of the plant (Limpens *et al.*, 2004; Iqbal *et al.*, 2013).

A mutant population has previously been generated for *N. caerulescens* by fast neutron mutagenesis (Lochlainn *et al.*, 2011). In that experiment, the genetically diverse seeds of *N. caerulescens* collected in the field, in the vicinity of Ganges in the south of France, have been mutagenized aiming at identifying genetically stable, faster cycling lines for future genetic studies. The use of fast neutrons will often result in deletions, ranging from a single base pair to hundreds of kilobasepairs (kb). Such large deletions are likely to not only knock out the target gene but also several of the neighbouring genes, which will make it more difficult to match a mutant phenotype to the relevant gene (Li *et al.*, 2001). Furthermore, many genes are essential, and deletions in these genes will pre-dominantly cause lethality, necessitating the generation of less severe mutations to understand gene function in *N. caerulescens*.

Chemical mutagenesis, such as with the classical mutagen Ethyl Methane Sulfonate (EMS), induces predominantly point mutations randomly distributed over the genome. With EMS a high degree of mutational saturation can be achieved without excessive DNA damage (Gilchrist and Haughn, 2005). The main advantage of chemical mutagenesis is the ability to accumulate an allelic series of mutants with a range of modified functions, from wild-type to loss of function mutants. The latter may occur when EMS generates a premature stop codon or mutations that disturb intron splicing (Slade *et al.*, 2005). Such allelic series are desirable as they generate a wide repertoire of phenotypes, which provide more insight into the function of a single target gene. Additionally, chemical mutagenesis does not rely on genetic transformation or tissue

culture regeneration techniques, thus, it can be applied to any plant species, regardless of ploidy level, genome size, or genetic background (Kurowska *et al.*, 2011). An individual plant that carries a point mutation in a gene of interest can be detected in a mutagenized population through a technique called TILLING (*Targeting Induced Local Lesions IN Genomes*) (McCallum *et al.*, 2000). This technique uses post-PCR techniques to identify single bp mutations such as Li-Cor and the recently developed High Resolution Melting (HRM) curve analysis (Ishikawa *et al.*, 2010; Kurowska *et al.*, 2011; Lochlainn *et al.*, 2011) and DNA re-sequencing (Tsai *et al.*, 2011). In the traditional approach, PCR amplification is followed by the cleavage of mismatches using endonucleases such as CEL1 or END1 that recognize single DNA strands (Triques *et al.*, 2008; Oleykowski *et al.*, 1998) and detect the mutants on the sensitive Li-Cor gel analyser (Colbert *et al.*, 2001). In the HRM platform, PCR is followed by the direct detection of mutations based on DNA melting curve analysis (Applied Biosystems, 2009), thereafter subsequent confirmation by sequencing. This platform has been applied successfully to detect mutations in TILLING populations of tomato (Gady *et al.*, 2009), wheat (Dong *et al.*, 2009), clinically relevant fungi (Mader *et al.*, 2011). Since the inception of TILLING in the model plant species *A. thaliana* (McCallum *et al.*, 2000), this method has been very widely applied to agronomic crops and animals (reviews Kurowska *et al.*, 2011, (Gilchrist and Haughn, 2005; Martín *et al.*, 2009). In this study, we describe the development of a TILLING population for *N. caerulescens* based on the Saint Felix de Pallières (SF) accession, a Zn/Ni/Cd hyperaccumulator. The effectiveness of EMS on the plant growth and seeds production was evaluated in the M1 and M2 generation. Based on the element concentrations in leaf (one mature leaf) material for all M2 individuals putative mutants with disturbed ionome profiles were identified. Based on the occurrence of obvious morphological mutant phenotypes, we evaluated the mutation frequency and the number of knockout mutants. Moreover, for TILLING, the HRM curve based mutation detection platform has been optimized and applied to detect mutant alleles for the *bZIP19* gene, encoding an important transcription factor for Zn homeostasis in plants. This mutant population is the first TILLING population for *N. caerulescens*, and can be an attractive genomic tool for research in plant development as well as a

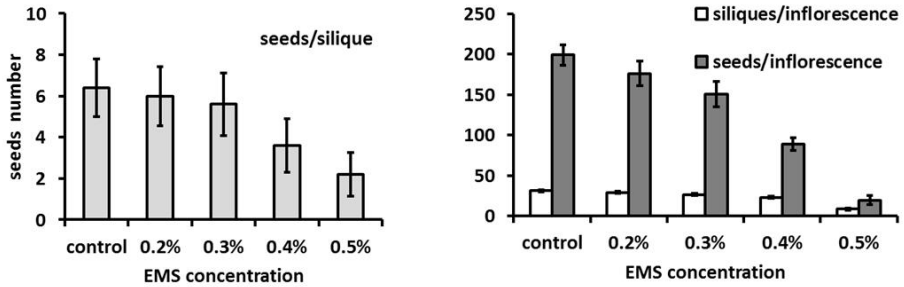
valuable resource for identifying novel genes that control ionome composition in this heavy metal hyperaccumulator species.

## Results

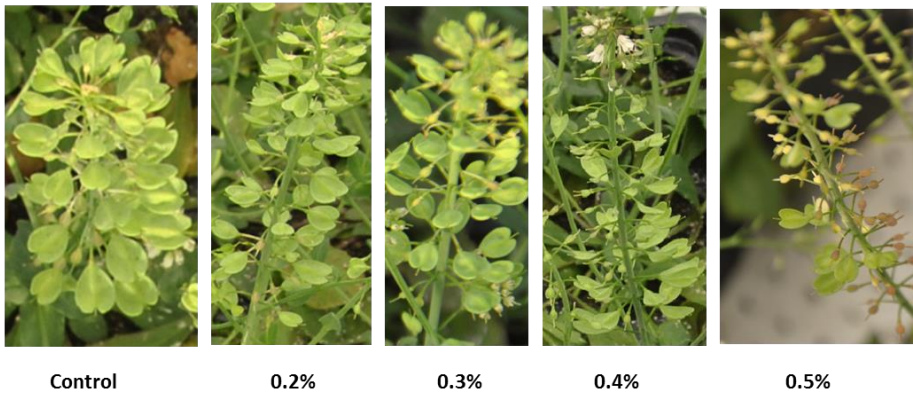
### **The effectiveness of EMS in *N. caerulescens* accession SF**

To determine a suitable EMS concentration for mutagenesis of *N. caerulescens*, 200 seeds were soaked in different EMS concentrations (0.2, 0.3, 0.4 and 0.5%). A subset of the germinated seedlings from the 0.2% and 0.5% EMS concentrations were grown to evaluate their effectiveness for mutagenesis. The number of siliques and the number of seeds/silique in the M1 plants were determined (**Figure 1**). At the higher concentrations of EMS, a stronger negative effect on fertility was observed, especially obvious at the highest concentration of 0.5% EMS. At that concentration also the number of seeds/inflorescences was negatively affected with only approximately 2-3 seeds per silique being produced. A reduction in seed number was already observed at 0.4% EMS, indicating 0.4% EMS was an effective concentration for mutagenesis in the SF accession of *N. caerulescens*. The use of 0.3% EMS was expected also to provide some effect, but probably resulting in a better fertility. Therefore the 0.3% and 0.4% EMS treatments were used to generate an M1 mutant population of *N. caerulescens*

A



B



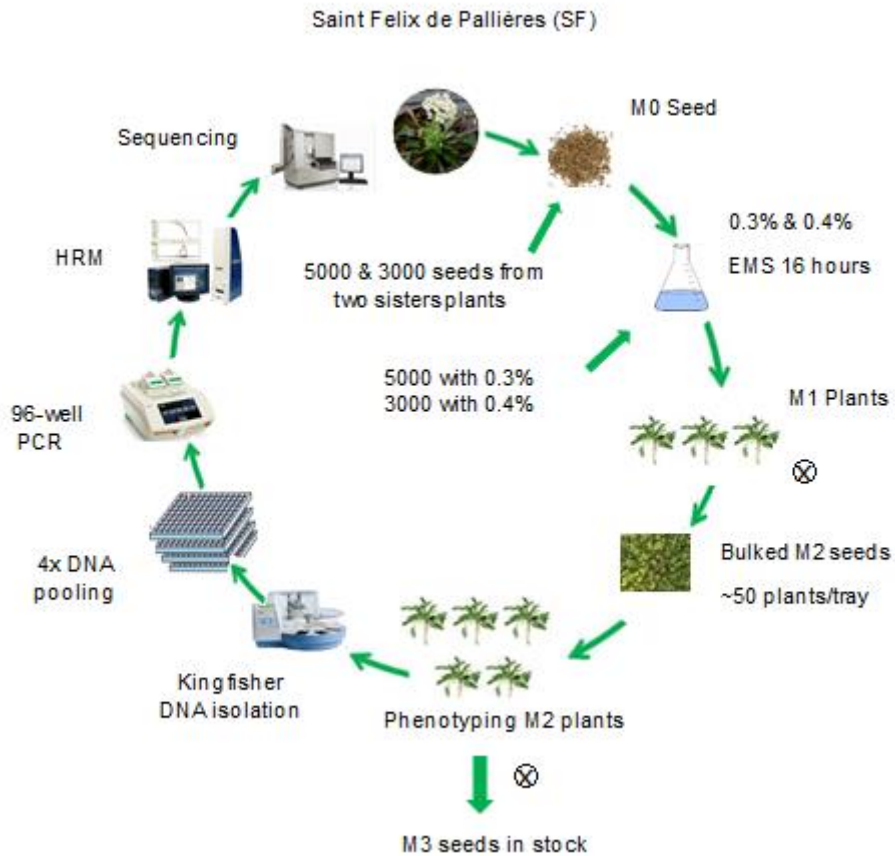
**Figure 1. Optimization of the EMS concentration for *N. caerulescens* mutagenesis.** Seeds were imbibed with four different ethyl methane sulfonate (EMS) concentrations and a control treatment with only distilled water (control) for 16 hrs in the dark. Plants were grown from these seeds, vernalized and allowed to flower upon which different traits were determined. (A) The number of seeds per silique, the number of siliques per inflorescence, and the number of seeds per inflorescence of seeds from different EMS treatments, presented in grey, white and dark bars. Data show the average of 3-5 plants  $\pm$  SE. (B) Representative examples of the inflorescences of plants treated with the indicated EMS concentration.

### **Development of an M2 population in Saint Felix (SF) background**

To create an M2 population that can be used for TILLING, seeds of an inbred line of *N. caerulea* accession 'Saint Felix de Pallières' (SF), a Zn/Cd hyperaccumulating accession, were treated with EMS (as described in MandM) to create an M1 seed batch. These M1 seeds were germinated and grown in ~150 trays in an unheated greenhouse during winter, which provided sufficient vernalisation to induce flowering in spring. The M1 plants were allowed to self-fertilize and M2 seeds were collected per tray of ~50 plants, of which 2/3 from the 0.3% and 1/3 from the 0.4% treatment. To generate the TILLING population, approximately 60 or 120 seeds per M1 progenitor tray from respectively the 0.3% and 0.4% EMS treatments were grown, and referred to as the 0.3 and 0.4 subpopulations. Since most morphological mutants were found in the 0.4 population, this subpopulation was used to test the effectiveness of TILLING and seeds from this subpopulation were harvested from individual M2 plants, resulting in M3 seeds representing approximately 7000 M2 individuals. **Figure 2** represents the work flow and procedure followed for TILLING.



## Construction a Tilling Population of *Noccaea caerulescens*

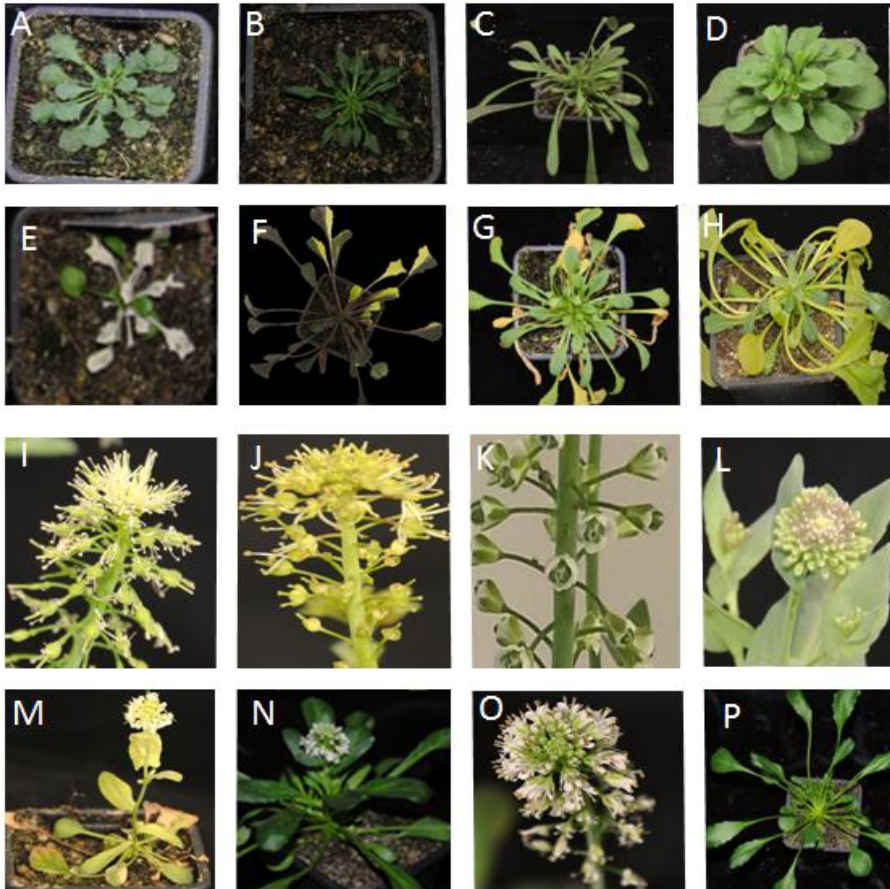


**Figure 2. A schematic overview of the mutagenesis and TILLING procedure.** M0 seeds from two sister plants of a single-seed recurrent inbred descendant of *N. caerulescens* accession Saint Felix de Pallières were treated with 0.3 or 0.4% EMS to generate M1 seeds. These were sown in trays with 54 plants each. M1 plants were grown and allowed to self-fertilize to generate M2 seed, which was bulk-harvested per tray. A subset of M2 seeds / tray were sown to grow M2 plants in the greenhouse. In total 5520 and 7000 M2 plants were grown to obtain the 0.3 and 0.4% populations, respectively, the latter of which was used for TILLING. The M2 generation was phenotyped for plant morphological traits and their ionome profile. M3 seeds were harvested from individual M2 plants of the TILLING population. For TILLING by HRM, the genomic DNA of individual plants was isolated using a Kingfisher DNA isolation robot. DNA of four plants was pooled in single wells of 96-well microtiter plates. Each pool was used for PCR amplification. PCR products per pool were subjected to High Resolution Melting curve analysis (HRM). The wells that contained mutations were selected and subjected to a second HRM screen. Positive samples were again PCR amplified and confirmed by DNA-sequencing the PCR product.

### **Morphological phenotypes in the mutagenized population**

To evaluate the effectiveness of mutagenesis, a broad range of visual morphological phenotypes were monitored in both populations, during the vegetative and reproductive stages. The most striking phenotypes are summarized in **Table 1**. As expected, the 0.4 population yielded a higher percentage of mutant phenotypes than 0.3 population in the M2 mutants for plant architectural traits, such as leaf shape, leaf colour, inflorescence number, flower development and flowering time were observed in the M2 population. The most commonly observed phenotypes in both populations were related to anthocyanin pigmentation. The second largest number of the mutants is related to the leaf pigmentation ranging from narrow yellow spots, necrotic spots and yellow sectors (variegation), or generally lighter or darker green plants. On average, the frequency of the pigmentation mutations was 1.2% and 2.1% in the 0.3 and 0.4 populations, respectively. Flower morphology mutants were investigated upon flowering of the plants. 1.9% of the plants in the 0.4 population showed altered flower morphology phenotypes. Eceriferum mutants with a bright green appearance due to an aberrant cuticular wax layer, and plants without secondary shoots account for 0.11 and 0.04%, respectively, in the 0.4 population. Flowers with smaller petals and stamens and with multiple petals (double flowering) were exclusively observed in the 0.4 population, accounting for 0.18% in total.

A selection of morphological phenotypes observed in the M2 population is presented in **Figure 3**. Many more other abnormal phenotypes were observed at different development stages, as listed in **Figure S1**. The frequency and the wide range of the phenotypic mutations observed in these two populations implied that especially the 0.4 population, would be a valuable resource for screening desired *N. caerulea* mutations in forward and reverse genetics screens.



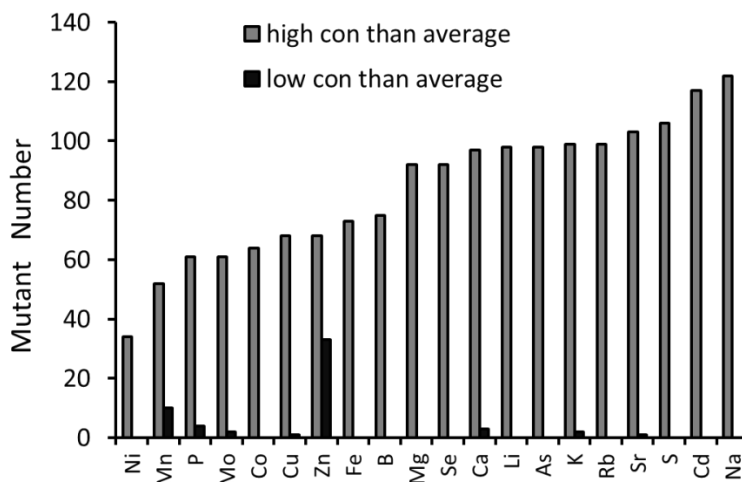
**Figure 3. Examples of mutant phenotypes in the M2 population.** Examples of morphological mutants illustrating the mutation spectrum observed in the mutagenized *N. caerulescens* M2 population. A-D: abnormal leaf shape and organization; E-H: chlorophyll aberration; I-J: apetala-like phenotype; K: agamous-like “double flower” phenotype; L: cauliflower-like phenotype; M: no side shoots, N: non-vernalisation required early flowering mutant, O: wild type SF inflorescence, P: wild type SF rosette (2 months old). Additional mutant phenotypes are described in the **Figure s1**.

**Table 1** Frequency of indicated mutant phenotypes observed in the 0.3% and 0.4% EMS M2 populations (pop). The frequency was calculated by dividing the number of mutant plants by the total number of plants in each population and expressed in %.

	Phenotypes	0.3% EMS	0.4% EMS	Frequency 0.3% pop (%)	Frequency 0.4% pop (%)
leave colour	Anthocyanin	105	158	1.91%	2.26%
	Chlorophyll	76	144	1.20%	2.06%
Flower morphology	Late flowering	15	50	0.27%	0.71%
	Fasciation	17	40	0.31%	0.57%
	Loose flower	3	10	0.05%	0.14%
	<i>terminal flower-like</i>	2	8	0.04%	0.11%
	<i>apetala-like</i>	2	8	0.04%	0.11%
	<i>cauliflower-like</i>	1	6	0.02%	0.09%
	Early flowering	2	5	0.04%	0.07%
	No petals and stamens	0	3	0	0.04%
	Fused petals	1	2	0.02%	0.03%
	<i>agamous-like</i>	0	2	0	0.03%
Stem colour	<i>eceriferum-like</i>	0	8	0	0.11%
Shoot number	No side shoots	1	3	0.02%	0.04%

### Identification of mutants with a disturbed ionome profile

Considering that *N. caerulescens* is a Zn/Cd/Ni hyperaccumulator, it will be interesting to identify mutants in which mineral uptake and distribution is disturbed. Therefore all plants were sampled to determine their leaf ionome by Coupled Plasma Mass Spectrometry (ICP-MS). Samples in which one of the 20 trace elements that were analysed (Li, B, Na, Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd) was found to have a concentration that was three or more standard deviations different from the concentration in the WT (Z-score analysis) were considered to be putative high or low element mutants (Lahner *et al.*, 2003) (**Figure 4**). Given the Zn/Cd/Ni hyperaccumulating characteristics of the SF background, we initially focused on the identification of low Zn mutants in the 0.4 population. In total, 34 putative low Zn mutants were identified in this M2 population. These mutants were found in different seed batches, suggesting they arose from independent mutation events (**Figure S4**). The putative mutants for each element with the altered elemental profiles are listed in the **Supplement excel file 1**.

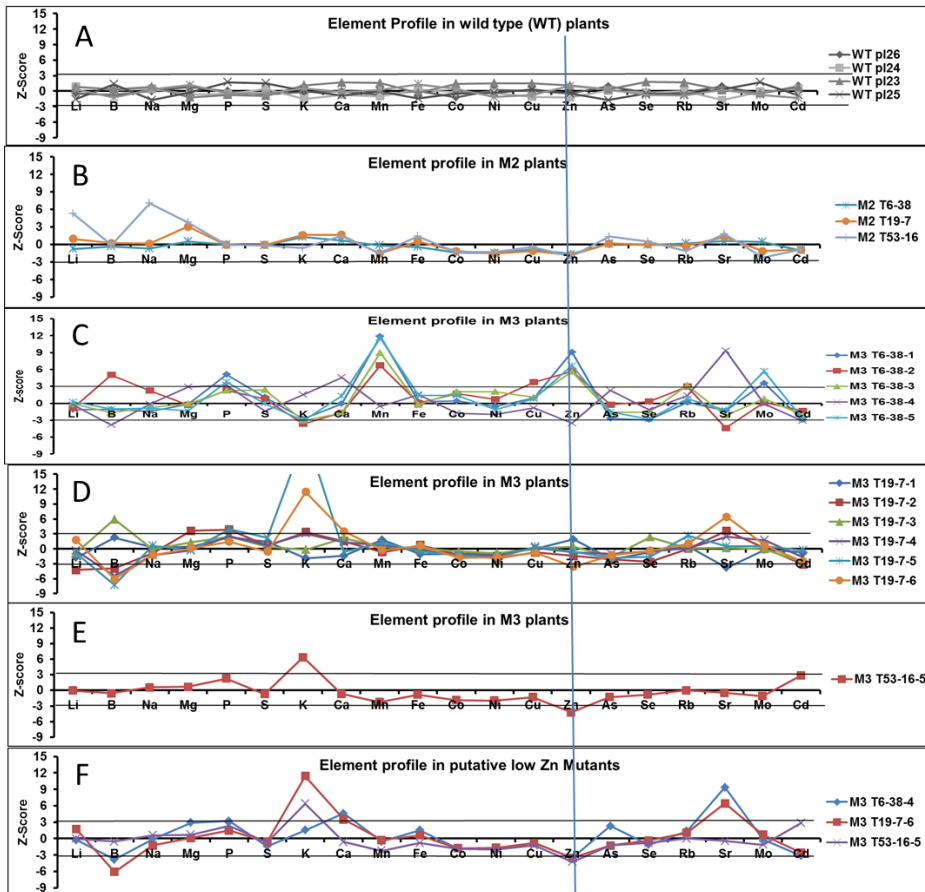


**Figure 4. Number of putative ionome mutants in the *N. caerulescens* M2 population.** Putative ionome mutant numbers identified based on Z-score values to have either a higher (grey bars) or lower concentration (con) (black bars) than average for at least the indicated element.

To further confirm the element profile of the putative mutants, representative putative low Zn mutants T6-38, T19-7 and T53-16 were selected and 10-15 progeny from those mutants were grown in the greenhouse. These profiles were not unambiguous, i.e. the progenies showed heterogeneous phenotypes instead of homogeneous phenotypes expected for progenies of recessive mutants (**Figure 5**). Among the progeny of T6-38, plant T6-38-4 accumulated very little Zn, but showed high Sr and Ca concentrations. The four sister plants, however, contained much higher Zn and Mn concentrations than the wild type plants. Plant T6-48-4 was much reduced in size compared to the other plants and displayed strong leaf chlorosis. However, none of the other putative mutants, neither in the same line nor in others, displayed any obvious visible phenotypes (**data not shown**). As for the progeny of T19-7, plant T19-7-6 showed a very low Zn concentration, but a higher concentration of K and Sr. Of the progeny of T54-16, only one plant was alive when sampling. It confirmed the low Zn concentration of the M2 parent, and also showed a high K concentration. The element profiles of the progenies compared to their mother plant (M2) indicated the

## Chapter 3

element profiles of the M3 plants T6-38-4, T19-7-6 and T53-16-5 are similar to that being found in the M2 generation

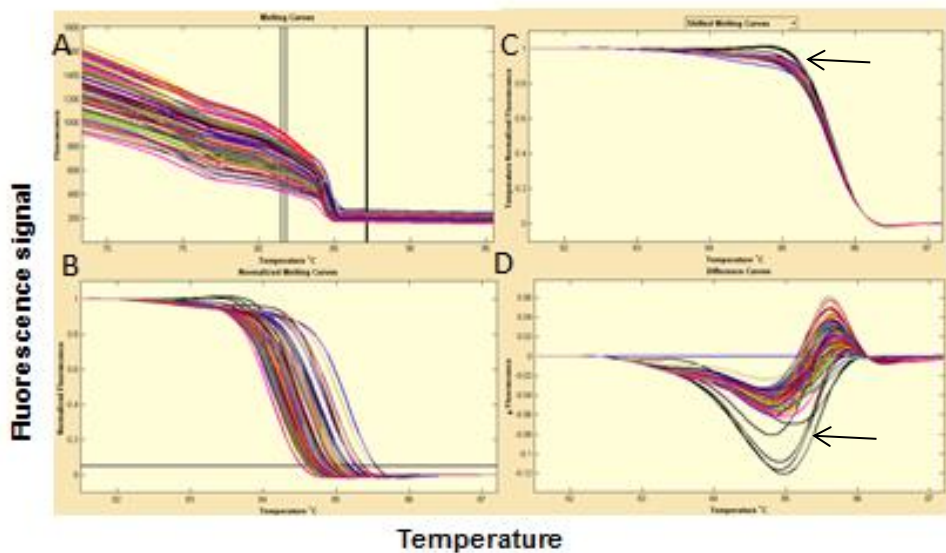


**Figure 5. Leaf ionome profiles of representative putative *N. caerulescens* low Zn mutants** Ionome profiles based on ICP-MS analysis of one mature, expanded leaf per M2 or M3 plant, indicating the Z-scores for each element of indicated plants. The Z-scores for M2 plants were normalized to the average of 6500 individuals. The ionome profiles are shown for (A) *SF* wild type (WT) plants; (B) low Zn M2 plants; (C) M3 progeny of M2 T6-38; (D) M3 progeny of M2 T19-7; (E) M3 progeny of M2 T53-16; (F) putative low Zn mutants selected from each M3 progeny.

### Optimization of the High Resolution Melting Curve analysis

DNA samples were prepared from 7000 fertile M2 individuals of the 0.4 % population. Samples from four M2 individuals were combined (4x pooling). To reduce cost and

work load for screening the mutants in an efficient way, the High Resolution Melting Curve (HRM) analysis was optimized based on the analysis of a known EMS-induced point mutation in the *Flowering Locus C (FLC)* gene of *N. caerulea* (**Chapter 4**). For this test, DNA samples from the *flc-1* mutant, carrying a G->A mutation (483bps after the ATG) were spiked to five wells in a 96-well plate, which contained the 4x pooled DNA samples from wild-type plants. The initial *FLC* HRM assay of this plate yielded seven positive samples that displayed aberrant melting curves indicating PCR fragments carrying point mutations in these wells (**Figure 6**). After the second round of screening, four wells, all containing the *flc-1* mutant DNA, were confirmed as positive, meaning only one spiked sample got undetected, but no false positives were found.



**Figure 6. High Resolution Melting analysis of second round PCR products of 96 4x-fold pooled samples.** The figure shows the analysis of the amplicons containing known point mutation (G-A) in *flc-1*. (A). the original total fluorescence for 96-samples captured by the light scanner. (B). normalising curves by changing the black bars in figure A. (C). Normalized temperature-shifted curves, the black curves pointed by arrow containing *flc-1* mutants (G->A) in corresponding wells and (D) difference curves with G->A point mutations highlighted in black.

### Identification of mutants in the *NcbZIP19* gene

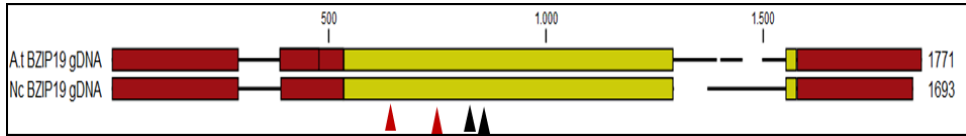
The *A. thaliana* *bZIP19* (*AtbZIP19*) gene encodes a basic-region leucine-zipper (bZIP) transcription factor, which acts together with transcription factor bZIP23 in the control of Zn deficiency responsive gene expression of, mainly, Zn transporters. These genes act largely redundantly, with the *bzip19bzip23* mutant being unresponsive to Zn deficiency and Zn deficiency hypersensitive (Assunção *et al.*, 2010). Orthologues of these genes are found in *N. caerulea*, with only the *NcbZIP19* gene being expressed. In order to study if mutation of *NcbZIP19* affects expression of the *N. caerulea* orthologues of the Zn uptake transporters found to be controlled by *AtbZIP19*, we screened the M2 population for mutants in *NcbZIP19*.

After screening 7000 M2 plants by applying the HRM assay, 28 samples were identified as positives in the first round of screening. After the second round of screening, 14 of these were selected as putative mutants. Four mutant allele serials in *NcbZIP19* were identified after subsequence confirmation by sequencing of HRM-positive PCR amplicons (**Figure 7B**). Of four mutations in exons, two were non-synonymous, leading to an alteration in the amino acid sequence, while the other two caused synonymous changes. Of these two nonsynonymous mutations, plant 73H10 carried a G>A mutation at position 78 of the coding region, leading to a Serine (Ser) to Asparagine (Asn) amino acid substitution. In plant 29E8, a G ->A mutation at position 115 of the coding region caused a substitution from Aspartic acid (Asp) to Asparagine (Asn). The two synonymous mutations were located at positions 196 and 211, respectively in the coding region (**Table 2**). **Figure 7** shows a schematic representation of the *NcbZIP19* gene, marking the location of the induced mutations.

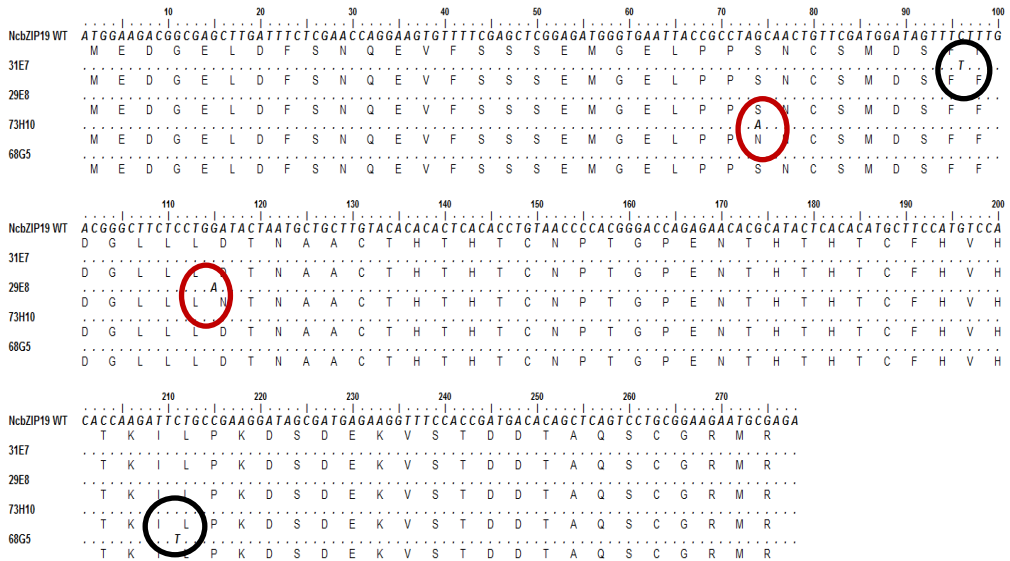


## Construction a Tilling Population of *Nocca caerulea*

A



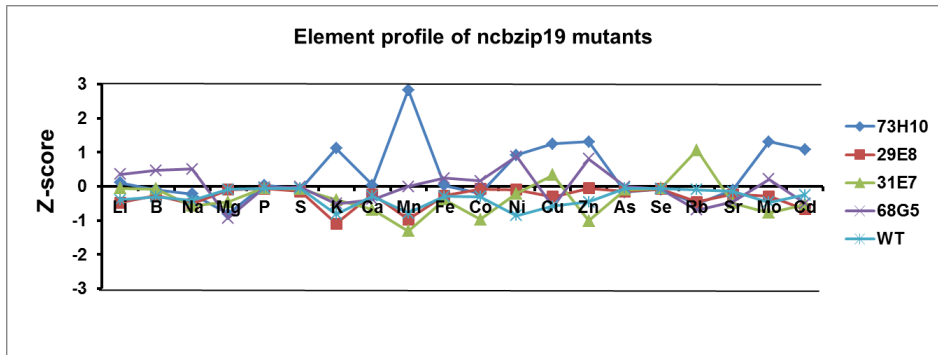
B



**Figure 7. Sequencing results of four *ncbzip19* mutants (A).** Genome organization of the *bZIP19* gene from *A. thaliana* (*At*) and *N. caerulea* (*Nc*). Untranslated regions are indicated with red boxes, exons are indicated with yellow boxes, and introns are indicated with black vertical lines. The EMS-induced mutations are indicated by red (non-synonymous) and black (synonymous) arrows below the first exon. (B). Point mutations confirmed by sequencing. Two non-synonymous substitutions are shown in the red circles. Two synonymous substitutions are shown in black circles

The availability of the element profiles for the whole M2 population allows us to check the ionic profiles of the two non-synonymous *ncbzip19* mutants, 73H10 and 29E8. According to the Z-score analysis, the element profiles of these mutants were not

significantly different from the wild type (**Figure 8**), although the concentrations of K, Ni, Cu, Zn, Mo, Cd and particularly Mn, were higher than in WT plant. Unfortunately neither of these two mutants produced viable offspring (no seeds or no germinating seeds), which made further analysis of them impossible.



**Figure 8. The Ionome profiles of *ncbzip19* mutants and WT plants.** In plants 73H10 and 29E8, a mutation leading to a non-synonymous amino acid substitution was found. For plants 31E7 and 68G5, two synonymous substitutions were found. Profiles of single M2 individual for *ncbzip19* mutants are shown. Profile data of WT is the average of six plants. Z-score: standard deviation from means of WT.

### Genome-scale mutant frequency

The frequency of visible phenotypes in a mutant population indirectly reflects the overall, genome-wide mutation frequency. To determine the chance of obtaining a target mutation with an obvious phenotype in the TILLING population (0.4% population), we evaluated the mutation frequency based on the observation of visually mutant phenotypes that correspond to mutations in supposedly or confirmed single genes. A total of five early flowering and 2 double flower mutants were identified (**Table 1**). Of the early flowering mutants, we identified three mutations in the *FLC* gene (**Chapter 4**), causing aberrant splicing leading to truncated proteins. The two other early flowering mutants were due to the mutations in the *SVP* gene, producing respectively a premature stop codon and a non-synonymous substitution. The double flower phenotype is most likely due to mutations in the *AGAMOUS* (*AG*) gene (Yanofsky *et al.*, 1990), which is a single copy gene in *A. thaliana* and *N. caerulea* (Edouard Severing *et al.*, In preparation) Two to three mutants for each

of the three monogenic phenotypes suggests that full mutant saturation was almost complete in the 0.4 % population of 7000 individuals which was used for TILLING, with a chance of 13.5 % to find no mutants for a gene of comparable size (~700bps in ORF) in this population based on a Poisson distribution. Two-three mutants obtained for these three genes from the M2 population (respectively *FLC*, *SVP* and *AGAMOUS*), we estimated the knockout mutations/plant in this 0.4% population (7000 individuals). Assuming one mutant/3000 M2 plants, with around 30,000 genes (29,712) being predicted in the *N. caerulescens* genome (Lin *et al.*, 2014), we expect every M2 plant to carry on average knock-out mutations for 10 genes.

## Discussion

*Noccaea caerulescens* is a Zn/Cd/Ni hyperaccumulator. This species has been used for decades as prominent model to study the molecular mechanism underlying of heavy metal accumulation hyperaccumulation (Assunção *et al.* 2003). However, the use of *N. caerulescens* has been hampered due to this species is recalcitrant to transformation, therefore currently reverse genetics resources are not applicable to this species. The present study describes a mutant population in *N. caerulescens* generated by EMS-mutagenesis, a non-transgenic approach. The variety of visible phenotypes observed in the M2 generation, and the leaf ionome data demonstrated that this population has been effectively mutagenized and is likely to be a valuable resource to isolate mutants disturbed in elemental composition as well as plant morphology or any other relevant trait, including metal tolerance and hyperaccumulation. For reverse genetics purposes, the HRM analysis platform was optimized and successfully applied to screen this TILLING population for an allelic mutant series of the *bZIP19* transcription factor gene. The obtained results demonstrated that the population serves an adequate resource for both forward and reverse genetics purposes.

The mutant frequencies in the *N. caerulescens* mutant population were evaluated based on the frequency of single gene morphological mutants identified in the M2 population. We expect that for every gene on average 2.3 knockout mutants can be found upon screening the TILLING population of 7000 M2 plants. A similar mutation frequency has been found in the early research in *A. thaliana* to provide suitable

mutagenized populations (Koornneef *et al.*, 1982). In order to obtain the required number of the mutations allowance for genetic analysis, double size of the current population would have been screened due to the chance of getting a knockout mutation also depends on nucleotides composition and the size of gene. To improve the accuracy of the estimated mutant frequency, more single genes for which clear phenotypes have been described should be tested.

An alternative procedure is to estimate the mutant frequency at the DNA level using targeted approaches also used for TILLING, such as Li-COR and HRM technology, or genome wide screening by Next generation Sequencing (NGS) as reviewed by Schneeberger *et al.*, (2014). The genome-wide mutant frequency was evaluated by whole genome screening of mutations in a few representative targeted genes in e.g. *Arabidopsis thaliana* (Lai *et al.*, 2012), *Brachypodium distachyon* (Dalmais *et al.*, 2013), in rice (*Oryza sativa*), wheat (*Triticum aestivum* L) and tomato (*Solanum lycopersicum* L.) (Slade *et al.*, 2005; Till *et al.*, 2007; Cooper *et al.*, 2008; Stephenson *et al.*, 2010; Chen *et al.*, 2012a; Gady *et al.*, 2012). The mutant frequencies for those species are 1/300 kb in *Arabidopsis* (Greene *et al.*, 2003), 1/396 kb in *B. distachyon*, 1/451 kb in rice, 1/500 kb in *Brassica. rapa* (Stephenson *et al.*, 2010) and 1/737 kb in tomato (Gady *et al.*, 2009). Based on the mutant frequency, the number of the point mutants one plant carries was estimated  $\{(\text{genome size}/(\text{mutant rate}))\}$ , varying from a few to several hundred per plant e.g. around 1000/plant for *Medicago truncatula* (Le Signor *et al.*, 2009), 10.000/plant in *B. rapa* (Stephenson *et al.*, 2010) and 430/ plant in rice (Wu *et al.*, 2005) (Šimek and Novoselovic, 2006; Barkley and Wang, 2008; Stephenson *et al.*, 2010; Kurowska *et al.*, 2011). The evaluation of mutant frequency in these studies also include basepair mutations that cannot cause obvious phenotypes.

We could make a rough calculation of the knockout mutant frequency at genome level based on the loss-of-function mutations in the *FLC*, *SVP* and *AG* genes. The genome frequency of identifying the knockout mutations of a monogenic phenotype from our TILLING population was approximately one knockout mutant per 2000kbs (The gene length screened x the total number of Tilling individuals / The total number of morphological mutants). This would suggest that 7000 M2 plants are sufficient to

identify at least one mutant per 2000 kbs. The mutation frequency and the mutations one plant carrying that we observed are not comparable to the mutant frequency reported in previous TILLING screened populations (Martín *et al.*, 2009; review: Šimek and Novoselovic, 2006; Kurowska *et al.*, 2011) due to intragenic region and the nonsynonymous without showing obvious phenotypes were not included in present study.

The mutation density in a TILLING population depend upon various factors such as species, genotype, mutagen dosages and polyploidy (Talamè *et al.*, 2008). Of those factors, mutagen concentration would be a very crucial one. Higher mutant density would be achieved with the higher EMS concentration applied, but the fertility will be also negatively affected. For the polyploidy, high dosages of mutagen might be applied in order to obtain knockout mutants with obvious phenotype for redundant genes like wheat (1/ 20-50kb)(Dong *et al.*, 2009a). However diploid cannot withstand such high mutant frequency due to the severer sterility that will cause., For instance, in Arabidopsis, the mutant frequency is around 1/300kb has been reported (Greene *et al.*, 2003). Thus, these factors need to be in balance. Taking these factors into consideration, we concluded that 0.4% EMS would be the maximum concentration that could be applied in this *N. caerulea* SF accession and possibly to identify two to three loss-of-function mutations for the monogenic traits in the 0.4 % population. With the availability of the optimized TILLING Platform (HRM) in this *N. caerulea* SF mutant population, the genome mutant frequency would be evaluated accurately after more targeted genes are screened. Such could be comparable to the reported species

The HRM platform was used to identify potential mutation events based on a visual detection of aberrant HRM curves. In order to detect mutations in *N. caerulea* in an efficient and effective way, we adapted a 4x pooling strategy. This seemed adequate, as when we tested a known point mutation in the *FLC* gene, we did not find false positives and only failed to detect the mutation in one pool (one false negative). To minimize false negatives, a large number of false positives were accepted in the first screen for the *BZIP19* gene. After the final appraisal of the mutations by sequencing of

the amplicons, four *BZIP19* mutants were identified. *NcbZIP19* controls the expression of *ZIP* and *NAS* genes (Ya-Fen Lin *et al.*, *In preparation*), as has been found in *A. thaliana* (Assunção *et al.*, 2010) and these genes would be less expressed if the *NcbZIP19* would have been mutated, and leading to a lower Zn hyperaccumulation. The two *ncbzip19* mutants with non-synonymous substitutions did not display an altered ionome profile when compared to WT, suggesting that the consequence of these two nonsynonymous substitutions was neutral. The altered element profiles caused by the mutation in *NcbZIP19* are not identified in the Ionome data suggests other genes are involved in those putative low Zn mutants. To identify knockout mutants for target elemental composition, forward screening of the elemental putative mutants with altered ionome profiles by the next generation sequencing would be very much fruitful (review: Schneeberger, 2014).

We also used the TILLING population for a forward genetic screen for ionome mutants. Identification of variation in elemental accumulation among putative *N. caerulea* mutants provides an excellent starting point for identifying genes important for regulation of the plant ionome, and with regards to Zn and Cd, for regulation of Zn/Cd hyperaccumulation. We selected a few putative low Zn mutants as representatives, and partially confirmed their phenotype in the M3 generation. However, not all progeny showed the expected low Zn phenotype. This was unexpected assuming the mutant phenotypes would be recessive, as is normally the case for EMS-induced point mutations (Comai and Henikoff, 2006). One explanation for the observed phenotypic variation would be that outcrossing has occurred between the identified mutants and neighbouring plants. Although *N. caerulea* is predominantly self-fertilizing with a cross-fertilization rate in the field ranging from 5–10% under normal growth conditions (Riley, 1956; Mousset *et al.*, 2016), this could be higher if either the growth conditions or the mutagenesis process caused some (male) sterility, as was described for sorghum (Xin *et al.*, 2008). Alternatively, the mutant itself could be partially self-sterile. Not unlikely assuming that zinc is a key constituent of many enzymes and proteins, and Zn deficient plants are known to be partially sterile (Talukdar and Aarts, 2008). In any case, we were not able to recover many seeds of these low Zn mutants,

which support the notion that partial sterility may be the consequence of the mutation. In addition, in comparison to the sister plants and the WT, T-38-4 had a less Zn accumulation capacity, and looked very poorly grown and developed severe leaf chlorosis. The limited seed production complicates genotyping and phenotyping in the M3 generation. This could be a pleiotropic effect of the mutation, but it could also reflect the high mutation rate in the 0.4% population. To avoid losing valuable materials, more care should be taken when propagating the mutants identified from this population for functional analysis. The 0.3% population showed a relatively higher fertility, unfortunately, with a lower mutant frequency in comparison to 0.4% population. 0.3% population would allow of an additional screening for mutants, but a relatively larger population size would be screened.

Most of the *N. caerulescens* mutants profiled in this study had at least one element that was significantly different (Z-score) from the wild type, indicating the mineral homeostasis genes might be involved. Or alternatively, more than one gene controlling mineral compositions might be affected in those mutants. The large number of mutants collected by ionome and forward screening of morphological traits indicating this mutant population is effectively mutated. This would allow the identification of mutant series affecting virtually any gene of interest by TILLING. The identification of the novel genes for targeted traits is promising in this population, via forward techniques like map-based cloning (Li *et al.*, 2010; Peng *et al.*, 2012), or mapping by sequencing (James *et al.*, 2013; Schneeberger, 2014). This is the first functional genomics tool in *N. caerulescens*, an extraordinary Zn/Cd/ Ni hyperaccumulating species. This mutant library will facilitate the application of functional genomics in *N. caerulescens* as well as comparative functional analysis with related species.

## Material and Methods

### Plant growth conditions

For mutagenesis, an inbred line of the *N. caerulea* accession 'Saint Felix de Pallières' (SF) has been used, which was obtained from Dr. Henk Schat (Free University, Amsterdam) and collected in the southern Cevennes, France, from calamine soil at a former Zn mine in the vicinity of St. Felix de Pallières (44° 2'40.03" N, 3°56'18.05" E). To ensure the homogeneity of the seeds used for mutagenesis, the plants were self-fertilized for 5 generations by single seed descent (SSD). After EMS treatment (see below), all M1 plants were multiplied under greenhouse conditions (from October 2012 to May 2013) in trays with 54 pots each, filled with a mixture of peat and sand. Plants were watered twice a day. The temperature was above 18°C during the night and around 23°C during the day; complementary artificial lighting was provided to attain a 16-h day. Two months after sowing, the plants were naturally vernalized (16-h day) in an unheated greenhouse for two and half months (day temperature was between 5-10°C; the night temperature did varied between 4-6 °C). After vernalisation, the plants were transferred to a heated greenhouse for flowering, under long day conditions (16-h day/8-h night) with (day/night) temperatures around 20-23 °C. In these greenhouse conditions, the duration of the growth cycle per generation is around 7 months. For the M2 generation, the plants were grown from October 2013 until June 2014 under the same culturing regime. All the plants used for this experiment were growing in the same greenhouse and under same conditions.

### EMS mutagenesis

The mutagenesis scheme is outlined in **Figure 2**. 5000, respectively 3000 seeds obtained from self-fertilization of two sister plants, were treated with a 0.3% and 0.4% Ethyl Methane Sulfonate (EMS, Sigma-Aldrich) concentration, respectively, by following the EMS protocol for higher plant mutagenesis (Koornneef, 2002). The seeds were imbibed in water for 16 hrs before being treated with 0.3% (~5000 seeds) and 0.4% (~3000 seeds) (v/v) of a freshly prepared EMS solution. After treatment, the seeds were washed at least 10 times with demi-water. After washing, the seeds



were suspended in 0.15–0.2% [w/v] water agar solution and sown directly on soil in 148 trays (54 wells / tray) using a pipette. Six months after sowing (including two months of vernalisation), the M2 seeds from ~50 plants per tray were bulk-harvested after removing the plants that had not produced seeds. From the obviously sterile plants within each tray, the seeds were harvested separately. The M2 generation was planted in October 2013. To maximize the M2 plants derived from as many M1 plants as possible within each tray, 120 bulked seeds / tray were sown in separate pots (7x7x8 cm) and were individually labelled. The plants were transferred to a “cold” greenhouse (only heated to keep minimum temperature above 4 °C) for vernalisation between end of December 2013 until mid of March 2014. Morphological and pigmentation characters were investigated at different development stages. After transfer to a fully heated greenhouse (~20 °C) the non-flowering plants were removed (~300) from the M2 population, resulting in ~7000 individuals in the 0.4% population for ionome analysis and TILLING. The M3 seeds from these plants were collected and are kept in long-term storage for future analysis. The 0.3% population consisted of around 5000 individuals, which was used as back-up for TILLING.

### **Element profiling by ICP-MS**

To determine the shoot ionome profile, one mature leaf was harvested from each M2 individual when the 5-months-old plants were flowering. Samples were collected in 96-deep-well plates (QIAGEN) and dried for 72 h at 60 °C. Elemental analyses were performed using inductively coupled plasma mass spectrometry (ICP-MS; Elan DRC II; PerkinElmer, <http://www.perkinelmer.com> ) to determine the concentrations of Li, B, Na, Mg, P, S, K, Ca, Mn, Fe, Co, Ni, Cu, Zn, As, Se, Rb, Sr, Mo and Cd as described by Lahner *et al.*, (2003). Putative mutants for any of the analysed elements were selected based on the Z-score value (three times higher or lower than WT) (Lahner *et al.*, 2003).

### **DNA extraction and quantification**

Genomic DNA was extracted from the inflorescence of each plant in the 0.4% population. For mutation screening purposes, the DNA samples were normalized and 4 xs pooled. For each DNA isolation, around 50 mg of inflorescence tips were collected

from M2 plants in 96-deep-well plates (QIAGEN) with two 4-mm glass beads in each tube and stored at -80 °C until further use. Prior to the DNA isolation, the samples were frozen in liquid nitrogen for one minute and homogenized twice using a Deep well shaker (Vaskon 96 grinder, Belgium; <http://www.vaskon.com>) for 2 minutes at 16.8 Hz, thereafter following the Nucleo Mag Plant isolation protocol (<http://www.mn-net.com/>), the lysis buffer (300 µl) with (RNase) was added and samples were incubated in a water bath (65°C) for 30 mins. After centrifuge for 10 mins at maximum speed, 300 µl supernatant were transferred to the new 96-deep well, mixed with binding buffer (150 µl) and magnetic beads (30 µl). After this step, the prepared samples were ready for DNA isolation by KingFisher™ Flex Magnetic Particle Processors (Thermo) by following the protocol of the kit (<http://www.mn-net.com/>)

For the final step, the DNA was eluted by 100 µl Elution buffer instead of 150 µl recommended by the protocol to reduce contaminations. A few plates were randomly selected for evaluation of the DNA quality by Nanodrop and gel electrophoresis. The DNA samples were all highly homogenous and the DNA concentrations were comparable, at around 30-40 ng/µl, for the whole population. The OD260/280 ratios were around ~1.8 to 2.0, indicating good quality DNA. After 5 times dilution, the 4x pooling in 96-well format was obtained by mixing four DNA samples/well. 2 µl of pooled genomic DNA was used in multiplex PCRs for mutation detection analysis.

### PCR primer design and amplification

All PCRs were performed using FramStar 96-wells plates (4titude, UK, <http://www.4ti.co.uk>). Specific primer sets were designed for this study based on preliminary *N. caerulea* whole genome sequences (Ganges accession) (manuscript in preparation) or on published cDNA sequences (Halimaa *et al.*, 2014a; Lin *et al.*, 2014) using Primer3plus ( <http://primer3plus.com/>) software. The primer sequences are listed in **Table S1**. PCR fragments (~400bp) obtained from a characterized EMS-induced SF early flowering *flc-1* mutant (**Chapter 4**) were used to setup the HRM mutation detection platform. The PCR were performed on 4x flat pools in FramStar 96-wells plates (4titude, UK, <http://www.4ti.co.uk>). HRM primer amplification

efficiencies and specificities were determined by PCR amplification of wild-type SF DNA. The reactions mix and PCR conditions were listed in the supplement **Table S2 and S3**.

### **High Resolution Melting (HRM) curve analysis**

Positive pools were selected by analysing the melting temperature profiles. If the pool contains a mutation it showed a lower melting temperature than homogeneous DNA fragments. The positive pools were screened again by using a single DNA sample in each tube. Positive samples were confirmed by sequencing (Ishikawa *et al.*, 2010).

### **Sequencing**

Mutations were validated by DNA fragment sequence analysis. The PCR product was purified using the NucleoSpin® Gel and PCR Clean-up kit and purified fragments were sent for sequencing to Eurofin Genome Sequencing Company (Ebersberg, Germany) (<https://www.eurofinsgenomics.eu/>). The sequences and amino-acid sequences of *NcbZIP19* were aligned by using the Bioedit program (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). Identical amino acids are indicated by dots, the deletion of the amino acid was indicated by dashes line.

## Supplemental information

**Supplement table 1.** *Primer sequences used to optimize PCR and HRM System*

Primer name	Gene		Primer
YISP1F	<i>NcFLC</i>	Forward	AACATGCCGATGATCTTAAAGC
YISP1R		Reverse	GAACGAGGGAATCGACACTTAC
bzip19P1_F	<i>NcbZIP19</i>	Forward	AAATCGTTCTCTACGTCCATGTCT
bzip19P1_R		Reverse	TCAAAATCGACAATTTCTAAGAAGC
bzip19BP2_F		Forward	GAGTACGTTACTTTGTTGGTGGT
bzip19P2_R3		Reverse	ATACTTTCGAACCGCTTCTCGATTTC

**Supplement table 2:** *Optimized PCR reaction system*

Reagents	Volume/reaction
Onetaq polymerase	0.05 µl
5x Onetaq PCR buffer	3.0 µl
10 µM Forward /Reverse Primers	0.2 µl
LC Green fluorescent dye	1.0 µl
5 mM dNTPs	0.4 µl
DNA template	2.0 µl
MQ water	3.35 µl
Total	10 µl

**Supplement table 3:** *PCR running conditions for flc-1 and bzip19 genes*

PCR step	Temp	Time
Initial Denaturation	94°C	30 seconds
30 Cycles	94°C	15-30 seconds
	45-68°C	15-60 seconds
	68°C	30 seconds
Final Extension	68°C	5 minutes
Hold	4°C	storage

## Construction a Tilling Population of *Noccaea caerulescens*

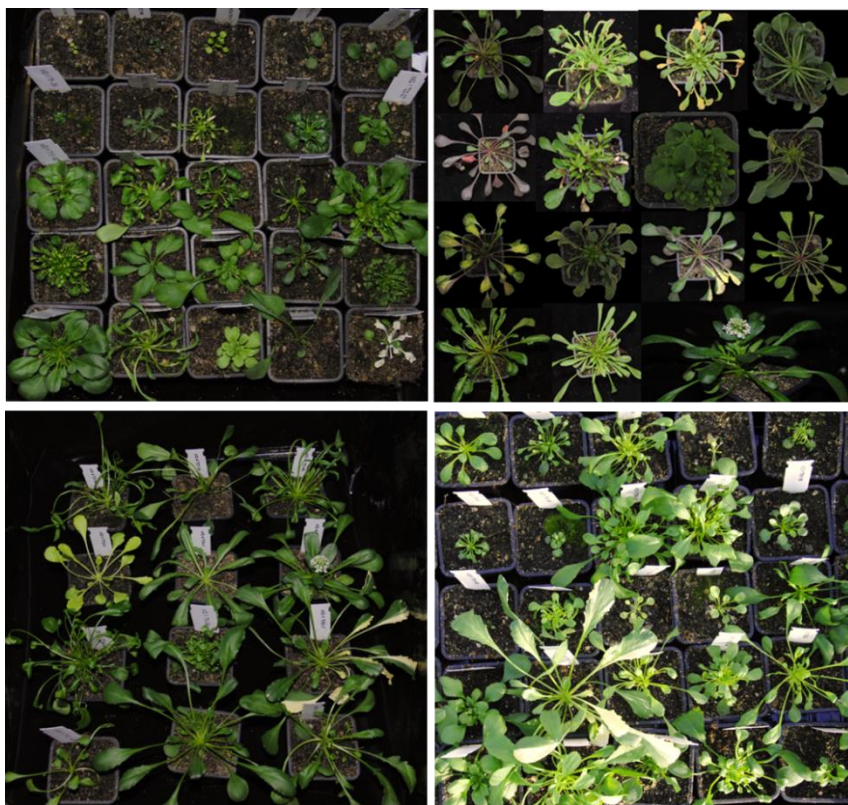
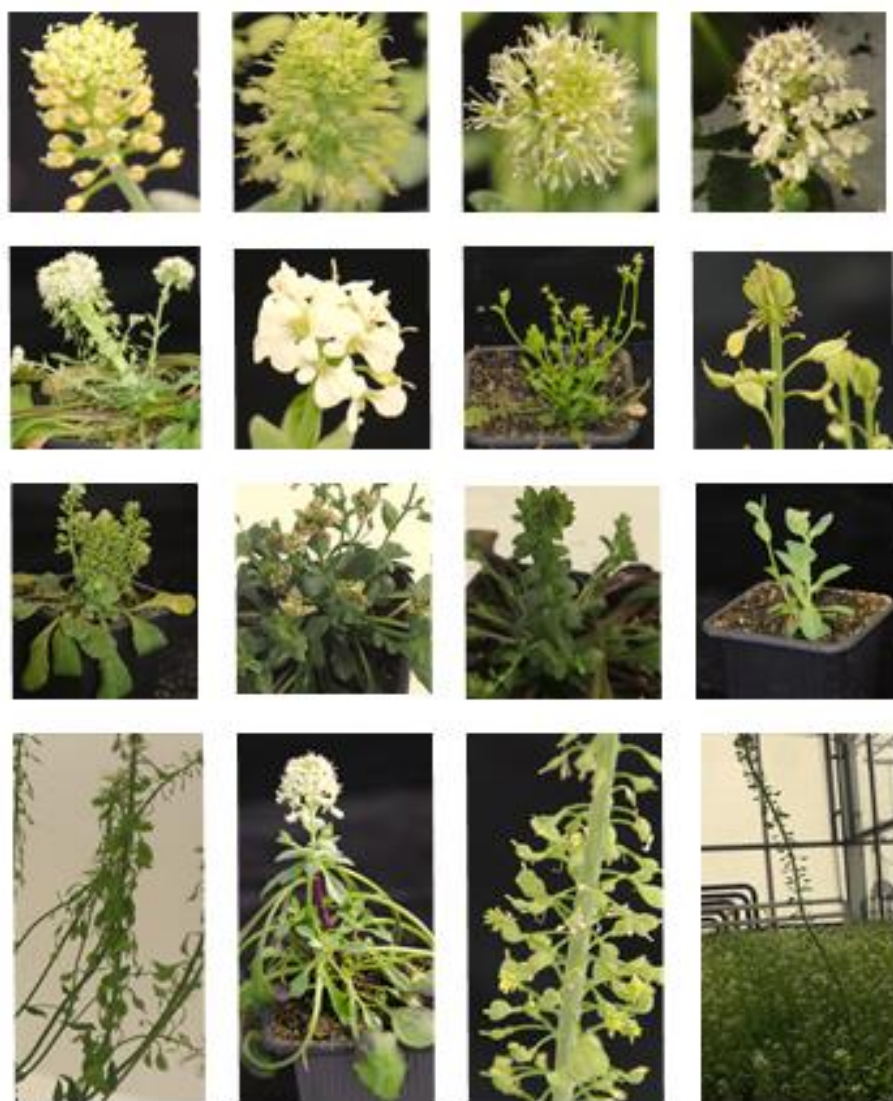
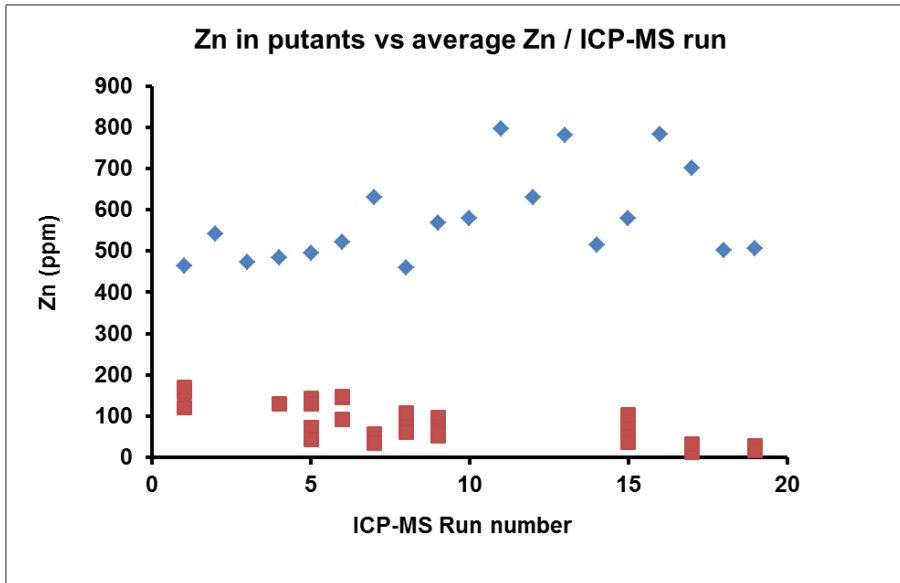


Figure S1. Additional morphological mutations identified in the M2 generation in vegetative growth stage.



**Continuous Figure S1.** Additional morphological mutations identified in the M2 generation in reproductive growth stage.



**Supplement Fig 2:** Distribution of putative low Zn mutants in the whole M2 population. 7000 M2 plants were measured in 20 different runs by inductively coupled plasma mass spectrometry (ICP-MS). Blue boxes matched with run number on x-axis stand for the average of Zn concentration in the M2 plants measured in the matched runs. Red boxes matched with x-axis stand for the Zn concentration of putative low Zn mutants in that run.

**Supplement file:** Ionome profile data for putative mutants of 20 trace elements can be accessed online in the Dropbox page by following the link:

<https://www.dropbox.com/home>

Login with username: wylfriend@163.com, password: thesis201609



## Acknowledgements

We thank Dr. Christian Bachem (Laboratory of Plant Breeding, Wageningen University and Research centre) for sharing the instruments for DNA isolation and mutation screening, Rinie Verwoert (PPO/PRI Unifarm of Wageningen University) for help with management of the greenhouse and taking care of plants, Corrie Hanhart (Lab of genetics, Wageningen University) for help harvesting the seeds. The help from my colleagues for collecting the leaf materials for ICP-MS is gratefully acknowledged. The Chinese Scholarship Council (CSC) and the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research) are acknowledged for the financial support.



**Sincere thanks to my colleagues for great help!**



## Chapter 4

### **The genetic control of flowering time in the biennial/perennial species *Noccaea caerulescens***

Yanli Wang,<sup>a</sup> Maarten Koornneef,<sup>a, b</sup> Mark GM Aarts<sup>a</sup>

<sup>a</sup> Laboratory of Genetics, Wageningen University, 6708 PB Wageningen, The Netherlands

<sup>b</sup> Max Planck Institute for Plant Breeding Research Carl-von-Linné Weg 10, 50829 Cologne, Germany

#### Financial source:

This research was financially supported by the China Scholarship Council (CSC) and Consortium for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research).

## Abstract

The appropriate timing of flowering is crucial for the reproductive success of plants. The molecular mechanism of flower induction has been well-studied in the model Brassicaceae species *Arabidopsis thaliana*, with long days and vernalisation as major environmental promotive factors. The vernalisation requirement, however, varies greatly among species. The related Brassicaceae species *Noccaea caerulea* has an obligate vernalisation requirement that has not yet been studied at the molecular level. Here we characterize the variation in vernalisation response of four geographically diverse biennial/perennial *N. caerulea* accessions, Ganges (GA), Lellingen (LE), La Calamine (LC) and St. Felix-de-Pallières (SF). Differences in vernalisation responsiveness among accessions indicate there is natural genetic variation for vernalisation response within *N. caerulea*. Mutants for the *FLOWERING LOCUS C* (*FLC*) and *SHORT VEGETATIVE PHASE* (*SVP*) genes are identified that no longer need vernalisation to flower, indicating that these genes encode key repressors of the *N. caerulea* vernalisation response pathway. However, the reversion of the strong, non-vernalisation requiring *flc1-1* mutant to vegetative growth at elevated temperature (30 °C), accompanied with reduced expression of *LEAFY* (*LFY*) and *APETALA1* (*AP1*) indicates that under higher temperatures, such reversion might be advantageous as it will suppress flowering in dry and hot summer conditions, allowing plants to return to a vegetative state and survive summer as a perennial plant. Therefore the “crosstalk” between vernalisation and ambient temperature might be a strategy to cope with the fluctuation temperature in *Noccaea* species. This strategy might facilitate a flexible evolutionary response to changing environment across a species range.

## Introduction

The transition from vegetative to reproductive growth is an important event in a plant's life cycle. In many plant species it is determined by an interaction between developmental programs and pathways that respond to environmental cues such as day length and temperature (Andrés and Coupland, 2012). In most temperate biennial and perennial species, including important crops, the transition to reproductive growth is accelerated by vernalisation (Oliver *et al.*, 2009). The genetic of the vernalisation response has been well-studied in the plant model species *Arabidopsis thaliana*. In winter-annual accessions, the promotion of flowering by vernalisation is genetically controlled by the interaction between *FLOWERING LOCUS C* (*FLC*) and *FRIGIDA* (*FRI*) (Clarke and Dean, 1994, Koornneef *et al.*, 1994; Lee *et al.*, 1994 ). *FLC* encodes for a MADS domain protein that acts as a repressor of flowering (Michaels and Amasino 1999; Sheldon *et al.* 1999). Its expression can be activated by *FRI*, which acts as part of a transcription initiation complex that binds to the *FLC* promoter (Choi *et al.*, 2011). *VERNALISATION INSENSITIVE 3* (*VIN3*) (Bond *et al.*, 2009; Kim *et al.*, 2010) is the most upstream gene in the vernalisation pathway (Kim *et al.*, 2010). The *VIN3* protein acts as a partner in a protein complex including also LIKE - HETEROCHROMATIN PROTEIN 1 (*LHP1*) and POLYCOMB REPRESSION COMPLEX 2 (*PRC2*), which controls expression of *FLC* through locus specific methylation modification (Sung and Amasino, 2004). Repression of *FLC* expression in the cold is maintained even when cold-treated plants are transferred to warm conditions (Michaels and Amasino, 1999; Shindo *et al.*, 2006), thereby releasing the repression of *FLOWERING LOCUS T* (*FT*), a potent activator of flowering and considered to be the main floral integrator in *A. thaliana* (Angel *et al.*, 2011).

*SHORT VEGETATIVE PHASE* (*SVP*) is another negative regulator of the floral transition (Hartmann *et al.*, 2000). It also encodes a MADS box transcription factor repressing flowering either as partner in a transcription initiation complex with *FLC* or independent from the latter (Mateos *et al.*, 2015). In the perennial species *Arabis alpina*, the vernalisation requirement and response is controlled by an *FLC* orthologue called *PERPATRUL FLOWERING 1* (*PEP1*). *PEP1* has a complex duplicated gene

structure, which differs from the simple structure of *FLC/PEP1* orthologues in related annual species such as *A. thaliana*. Furthermore, *PEP1* expression is upregulated again when the plants are transferred to warmer temperatures after the vernalisation treatment, with the consequence that meristems that have not been induced to flower remain vegetative (Wang *et al.*, 2009; Albani *et al.*, 2012a).

*Noccaea caerulescens* (J. and C. Presl) F. K. Meyer (formerly known as *Thlaspi caerulescens*), is a diploid ( $2n = 14$ ), biennial or facultative perennial plant from the same *Brassicaceae* family as *A. thaliana*. *N. caerulescens* is an extremophile, adapted to grow on soil with high concentrations of Ni, Zn, Pb or Cd. Next to displaying extreme heavy metal tolerance, it is also a heavy metal hyperaccumulator, with genotypes able to accumulate Ni, Zn and Cd to over 1% of their dry weight in shoots (Assunção *et al.*, 2003; Nascimento and Xing, 2006; Broadley *et al.*, 2007; Krämer, 2010). Together with the Zn/Cd hyperaccumulator species *Arabidopsis halleri*, *N. caerulescens* is among the most prominent plant model systems to study heavy metal hyperaccumulation and associated hypertolerance (Krämer, 2010; Hanikenne and Nouet, 2011; Pollard *et al.*, 2014), however, molecular genetic studies are challenging in this species because of its biennial life cycle of 7-9 months, including a 2-3 months vernalisation period to induce flowering (Peer *et al.*, 2003, 2006). This disadvantage limits the efficiency of genetic studies and breeding efforts to enhance its application in metal phytoremediation. To overcome this, one non-vernalisation requiring, rapid- cycling mutant has been generated from the *N. caerulescens* Ganges background through fast neutron mutagenesis. The genetic locus (or loci) affected in this mutant, to cause early flowering has not yet been elucidated (Lochlainn *et al.*, 2011).

Naturally collected *A. thaliana* accessions show extensive genetic variation in their vernalisation requirement and the efficiently to prompt a vernalisation response (Nordborg and Bergelson, 1999; Kim *et al.*, 2010; Dean, 2015; Duncan *et al.*, 2015). The same is likely to be the case for other temperate species which are less investigated. Recently, variation in vernalisation requirements and flowering time among *N. caerulescens* accessions has been reported (Guimarães *et al.*, 2013), but the

genetic basis and molecular mechanisms of flowering time regulation is still unresolved.

A better understanding of the genetic control of flowering time in *N. caerulescens* will help to understand its phenotypic diversity as an adaptation to climate conditions. Day lengths of 8 or 12 hours did not affect the flowering time when plants were exposed to a 4 °C cold treatment to induce flowering, indicating that only the temperature seems important to induce flowering in *N. caerulescens* (Guimarães *et al.*, 2013). In addition to the natural occurring variation to uncover the key genes involved in induction of flowering, it will also be important to identify genes through analysis of early flowering time mutations, such as those reported previously (Lochlainn *et al.*, 2011) to confirm the function of orthologues of known flowering time genes and to find new functions not found in e.g. *A. thaliana*. The present study investigates the variation in the vernalisation requirement and response of four representative *N. caerulescens* accessions collected from environmentally diverse locations, to understand the role of the major floral repressor genes *FLC* and *SVP* at the transcriptional level. It confirms the essential role of these two genes upon the identification and analysis of EMS-induced mutants for either of them, obtained by screening of a mutagenized M2 population for non-vernalisation requiring plants. A model for the regulation of flowering time in this biennial/perennial species is proposed and compared with the current models established for the annual *A. thaliana* and the perennial *A. alpina*.

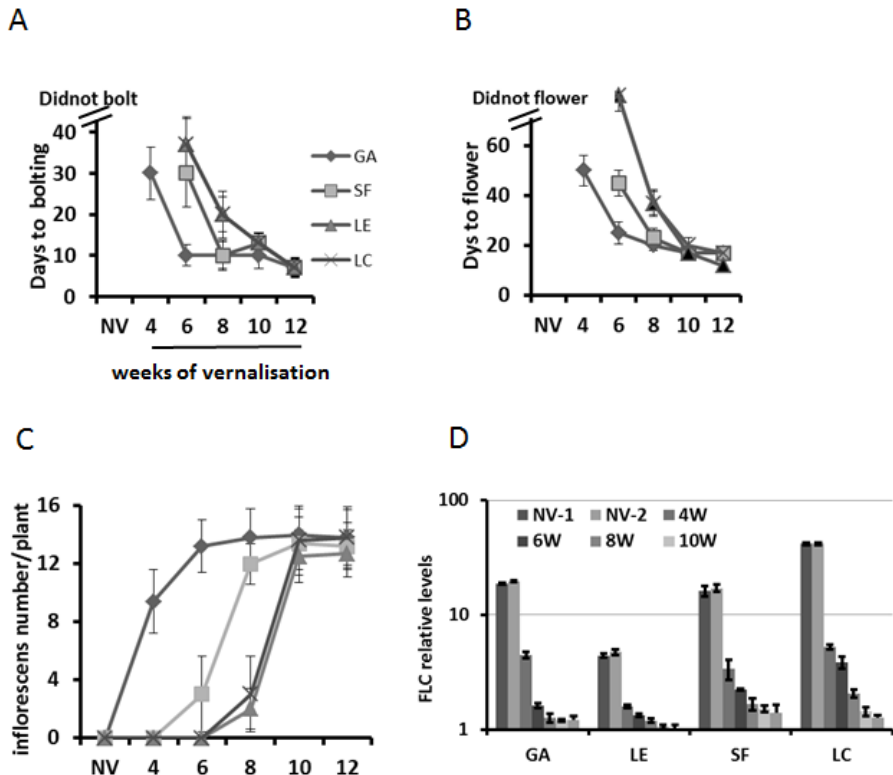
## Results

### **Different *N. caerulea* accessions require different vernalisation times**

Four *N. caerulea* accessions originating from different environments, both in terms of local soil metal concentrations and climate conditions were examined to determine their vernalisation requirement. Sixty days after sowing, the plants were vernalised by keeping them for 4, 6, 8, 10 or 12 weeks of cold (4 °C) and then returned to warm conditions. The days to bolting (first flower buds visible) and the days to flowering (first petals visible) for each accession were counted (Fig.1A and 1B). As the length of the cold period increased, the response to the cold treatment varies greatly among the accessions. In general, the prolonged cold accelerated the time to bolting and flowering in all accessions after returning them to warm greenhouse conditions. The largest difference in flowering time among accessions was observed after 4 weeks of cold treatment. Upon this vernalisation, flowering of GA was induced completely, with all plants (12) flowering 57 days after 4 weeks of vernalisation, whereas for SF, LC and LE flowering did not occur until the end of the experiment (180 days after sowing). After 6 weeks of cold treatment GA and SF flowered after respectively 25 and 45 days after vernalisation, whereas LE and LC still did not flower after the whole 180-day experiment.

The number of inflorescences was also higher upon the increase of the cold period. Eight weeks of vernalisation shortened the days to bolting in all accessions, and increased the inflorescence numbers compared with six weeks vernalisation (Fig.1C). For the GA and SF accessions more inflorescences developed after 8 weeks of vernalisation whereas in LC and LE, only a few inflorescences developed in this cold treatment but the time to flowering was slower in LE and LC and some of the inflorescences that had bolted did not flower at all, indicating that not in all meristems the transition to flowering was completed. After 10 weeks of cold treatment the number of days to bolting and to flowering was reduced to on average respectively 10, and 20 days for all accessions, indicating that the vernalisation was complete. After the 10 and 12 weeks of treatments, LE flowered the earliest among all accessions after the transfer to warm conditions. It also had stopped flowering, with all siliques being

well-developed, at 30 days after 10 and 12 weeks of cold, whereas the other three accessions were still flowering.



**Fig 1. Variation in vernalisation response and FLC expression of selected *N. caerulea* accessions.** (A) The transition to flowering as determined based on the time to bolting (i.e. first bud visible) (NB means no bolting); and (B) the time to flowering for non-vernalized plants (NV) or for plants vernalised for the indicated number of weeks at 4 °C (NF means no flowering). (C) The number of inflorescences that developed per plant when vernalised for the indicated number of weeks. (D) Relative expression levels of FLC upon vernalisation for the indicated number of weeks (W). The Y-axis is log-transformed. The housekeeping gene of Tubulin is used as reference. The rosette leaves are used for the expression level analysis. Data is mean of three to five plants with standard error. NV-1: 2-months-old non-vernalised plants; NV-2: 5-months-old non-vernalised plants.

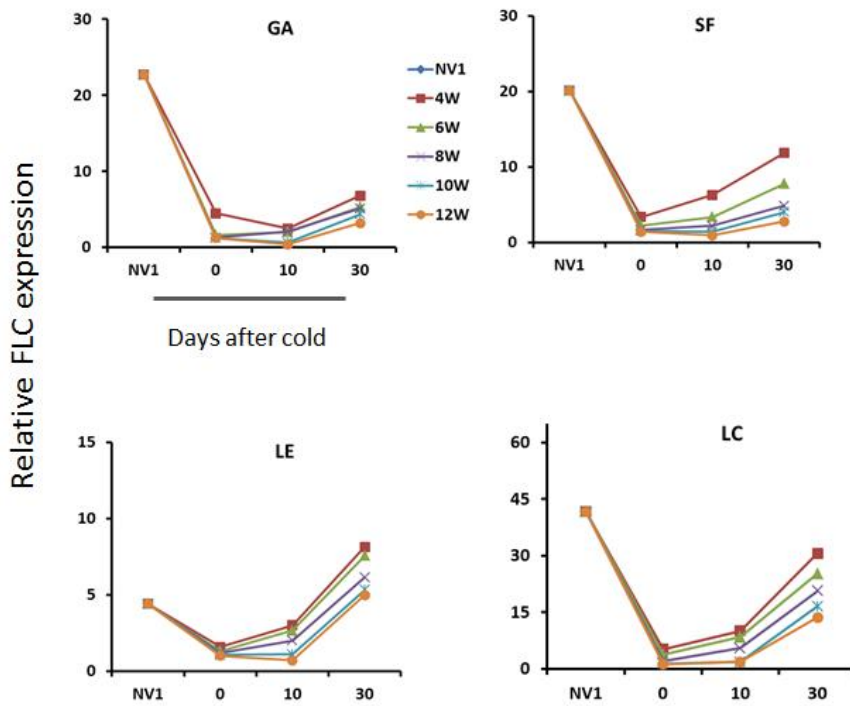
### Vernalisation of all four accessions repressed the *FLC* expression

To explore the role of flowering repressors in *N. caerulea* we determined the *FLC* and *SVP* expression levels in the four examined accessions after different vernalisation

periods. Prolonged cold treatment resulted in a downregulation of *FLC* transcript levels in all accessions (Fig.1D), but *SVP* transcript levels were not changed with prolonged cold treatment (**data not shown**). *FLC* expression in all accessions is down-regulated after 4 weeks of cold treatment, and continues decreasing the longer the cold period (Fig.1D). However, the sensitivity to vernalisation among accessions did not correlate with the initial expression level of *FLC* in *N. caerulea*. LE, with the lowest initial *FLC* mRNA level among all accessions, flowers only after 10 weeks of cold treatment, and, more importantly, only minor changes in *FLC* expression level were observed between 8 and 10 weeks of cold treatment, which were very different in their effect on inducing flowering (Fig.1A and Fig.1B) in this accession. Therefore, the *FLC* mRNA threshold to repress flowering seems much higher in the GA, SF and LC accessions, than in LE, with GA only requiring 4 weeks of vernalisation to induce flowers at a higher *FLC* expression level than LE.

One reason why the four accessions required different weeks of cold for the full acceleration of flowering could be that *FLC* expression recovers differently between accessions after the transfer to warmer conditions. Therefore, we also determined the maintenance of *FLC* gene repression in the accessions at 10 and 30 days after the vernalisation treatment (Fig. 2). After vernalisation for 4 to 8 weeks, the *FLC* expression indeed increased 10 days after return to the warm greenhouse in LC and LE, while in other two accessions, especially GA *FLC* expression stayed more or less constant. In all accessions, the *FLC* expression increased again after 30 days in the warm greenhouse, but very mild in GA and SF accessions, than in LC and LE. In general, the *FLC* expression remained more repressed the longer the cold period. In LE and LC, however, *FLC* mRNA levels are relatively more stable repressed when the cold treatment lasted up to 10 weeks, when flowering is induced in all accessions. This suggests that the recovery of *FLC* mRNA immediately after transfer to warm greenhouse correlates with the absence of flowering after a short period of vernalisation.





**Fig 2. Variation in the maintenance of repression of FLC expression.** FLC expression in *N. caerulescens* rosette leaves of accessions GA (A), SF (B), LE (C) and LC (D) of plants either not vernalised (NV1) or vernalised for 4, 6, 8, 10 or 12 weeks (#W) at 4°C after two months growth at 23°C, and subsequently transferred back to the warm greenhouse. Rosette leaves were collected at 0, 10 and 30 days following the vernalisation. Each data point represents at three to five plants. SE bars are omitted for clarity.

### Non-vernalisation requiring early flowering mutants in *N. caerulescens*

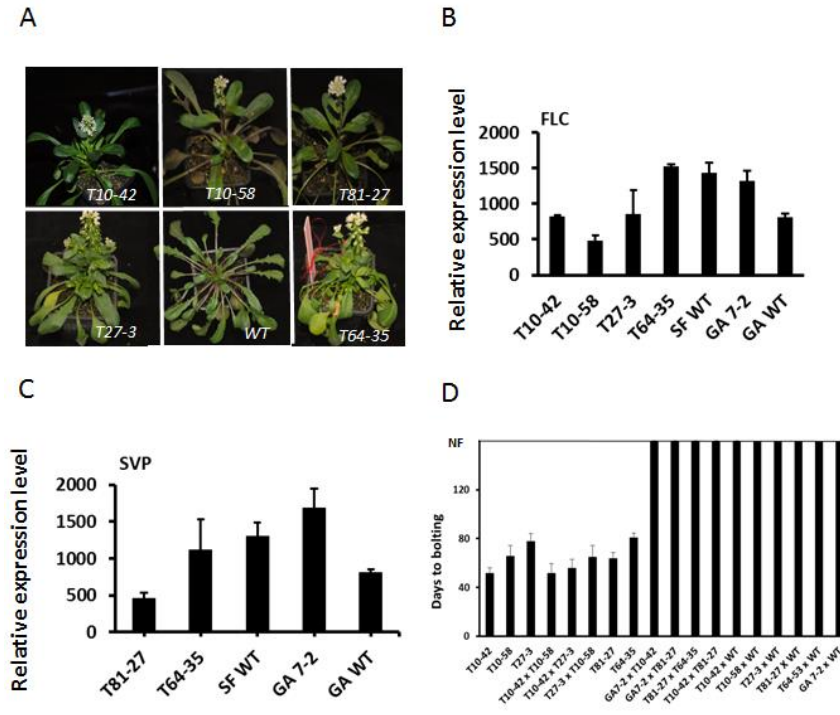
To study the molecular mechanisms controlling flowering of *N. caerulescens*, mutants showing an impaired vernalisation response were identified upon screening a total of 8000 EMS-mutagenized M2 seedlings (from 3500 M1 SF-plants). During the screen, the day temperature was kept at 20°C and long day conditions were applied (16h day/8h night), which did not induce flowering on control SF plants unless vernalized for two months. The screen revealed five early flowering mutants that lacked the obligate vernalisation requirement (Fig.3A). The *T10-42* mutant is the most extreme early flowering plant, flowering only 51 days after sowing without vernalisation. The

*T10-58* and *T27-3* mutants started flowering respectively at 66 and 78 days after sowing. The *T81-27* mutant and *T64-35* started flowering respectively 64 and 84 days, after sowing. No morphological differences were found between the mutants and WT in their inflorescence and flower morphology.

The expression of *FLC* and *SVP* were determined in all these *N. caerulescens* mutants. Three mutants (*T10-42*, *T10-58* and *T27-3*) exhibited *FLC* mRNA levels significant lower than that of WT suggesting that flowering without vernalisation obviously correlate with reduced *FLC* mRNA level (Fig.3B). However, two other mutants (*T81-27* and *T64-35*) exhibited *FLC* mRNA levels at least as high as those found in the WT, but *SVP* mRNA level was significantly lower in one mutant (*T81-27*) compared than that of in WT (Fig.3C). For the mutant *T64-35* as well as for the GA mutants (GA 7-2, GA 2-1) derived from the same mutant event (Lochlainn *et al.*, 2011), the genetic defects did not correlate with reduced expression of *FLC* and/or *SVP* (Fig.3B and 3C).

### Genetic analysis

To test which mutations are allelic, the non-vernalisation requiring mutants were inter-crossed. In addition, the mutants were also backcrossed (BC) to WT. Mutant *T64-35* was very sterile and no hybrid seeds could be obtained. The flowering time of all F1 plants were assessed as the days to bolting after sowing in long days in the greenhouse under non-vernalising conditions (Fig.3D). All F1 plants derived from the inter-crosses between the three mutants with low *FLC* mRNA (*T10-42*, *T10-58* and *T27-3*) flowered without vernalisation, indicating that these mutants carry inactive allelic recessive alleles at the same locus (Fig.3D). All F1 progeny from the backcrosses between the mutants (respectively *T10-42*, *T10-58*, *T27-3* and *T81-27*) and WT and inter-crosses between (*T10-42*, *T10-58* and *T27-3*) and *T81-27* mutants did not flower in these non-vernalising conditions (Fig 3D), indicating *T81-27* carry mutation in a different locus. Analysis of the BC1S1 progeny of the mutant x WT backcrosses under non-vernalising conditions, revealed all of these to be in line with a Mendelian 3:1 segregation of non-flowering : flowering plants.



**Fig 3. Identification and analysis of early flowering *N. caerulescens* mutants.** (A) Identified early flowering mutants (T#-#) and SF wild-type (WT) after growing without vernalisation. The pictures were taken 54, 69, 81, 74 and 94 days after the plant flowered respectively for T10-42, T10-58, T27-3, T81-27 and T64-35. (B) The relative repression of FLC in rosette leaves of WT and *flc-1* mutant plants. (C) The relative expression of SVP in the putative WT and *svp* mutants. (D) Days to flowering of self-fertilized and F1 progeny of the early flowering mutants and inter-mutant crosses when grown under non-vernalising conditions in a warm greenhouse, GA7-2 is the Ganges early flowering plant obtained through fast neutron mutagenesis. For gene expression, rosette leaves were collected when the first flower had opened. Each data point represents at least three plants  $\pm$  SE. NF: did not flower until end of the experiment.

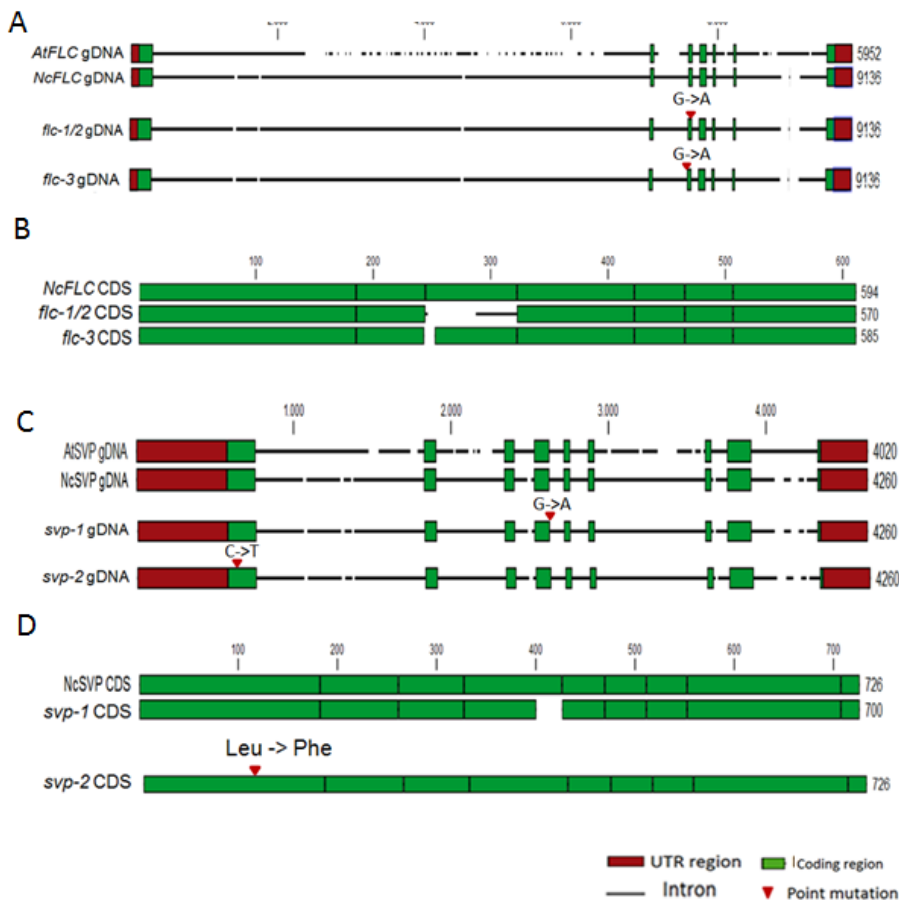
To explore whether the previously identified non-vernalisation requiring mutant identified in a fast-neutron generated M2 population in a ‘Ganges’-like background (Lochlainn *et al.*, 2011) was allelic with SF mutants, both Ganges mutant lines (derived from the same mutant event) were crossed to these four mutants in SF background. All F1 hybrids showed a late WT phenotype, indicating that an additional flower repressor is mutated in the Ganges mutant (Figure 3D).

### Splicing-site mutations in *FLC* and *SVP* result in early flowering mutants

Based on the information of orthologues in *Arabidopsis thaliana*, we deduced the genetic information of *FLC* and *SVP* in *N. caerulea* by using the transcriptome data (Lin *et al.*, 2014). Given the lower expression level of *FLC* and *SVP* in the early flowering mutants compared to WT, we cloned and sequenced *NcFLC* and *NcSVP* coding sequences (CDs) and genomic DNA (gDNA) from all mutants and WT. After alignment of the gDNA and CD sequences, four G- to A point mutations were found, with three located at different splicing junction sites of *FLC* in respectively T10-48, T10-52 and T27-3 and one in T81-27. Consistent with this, the CDs of *FLC* or *SVP* were alternatively spliced resulting in shorter CDs variants. T10-48 and T10-52 carrying mutations at same positions in *FLC* gene were identified within one tray in M2 generation, indicating these two mutant lines might derive from one mutant event. Based on these results, we designated the mutants T10-42, T10-58, T27-3 and T81-27 respectively *flc-1/2*, *flc-1/2*, *flc-3* and *svp-1* as mutant of respectively the *FLC* and *SVP* genes.

The genetic DNA organisation and the coding sequences of *FLC* and *SVP* in the wild plant and the *flc* mutants were presented in (Fig 4A,B) and (Fig 4C, D). For *flc-1/2* mutants, the point mutation is located exactly at the splicing acceptor site (483bps after ATG start codon) at the 3' end of the third exon resulting in a complete loss of the exon. For the *flc-3* mutant, 9 bps at the very beginning of the third exon was missing from the coding sequence because of another point mutation (421bps after ATG), this time at the 5' splicing site (Fig 4A,B). The short deletions within the *FLC* protein coding sequence were predicted to impair its activity (supplementary Fig.S1). For *svp-1*, a G to A point mutation (400bps after ATG) at the exon-intron boundary at the 3' end of the fourth exon was identified, resulting in a 26 bps deletion (Fig.4C, D). This deletion results in an open reading frame shift and the introduction of a premature stop codon just N-terminal to the highly conserved MADS box domain in the SVP protein (supplementary Fig.S2). No *FLC* and *SVP* sequences difference were found in the Ganges mutant (**data not shown**).

The genetic complementation together with the expression analysis and the sequencing results suggests that the identified point mutations explain the loss of *FLC*, respectively *SVP* functionality in at least four of the identified mutants, which altered their control of flowering. For the T64-35 mutant, coding sequencing indicated no mutations in the *FLC* gene, but did reveal a C to T mutation at the position 103 in the first exon of the *SVP* gene (Fig 4C,D). This gives rise to a nonsynonymous amino acid substitution in the ORF, altering Leu to Phe (35 from N-terminal) in a conserved region of the predicted protein sequence (Supplement Fig 2). As this is likely to alter SVP function in this mutant, we designated this mutant as *svp-2*. Unfortunately, this could not be confirmed by allelism tests because of the complete sterility of this mutant.

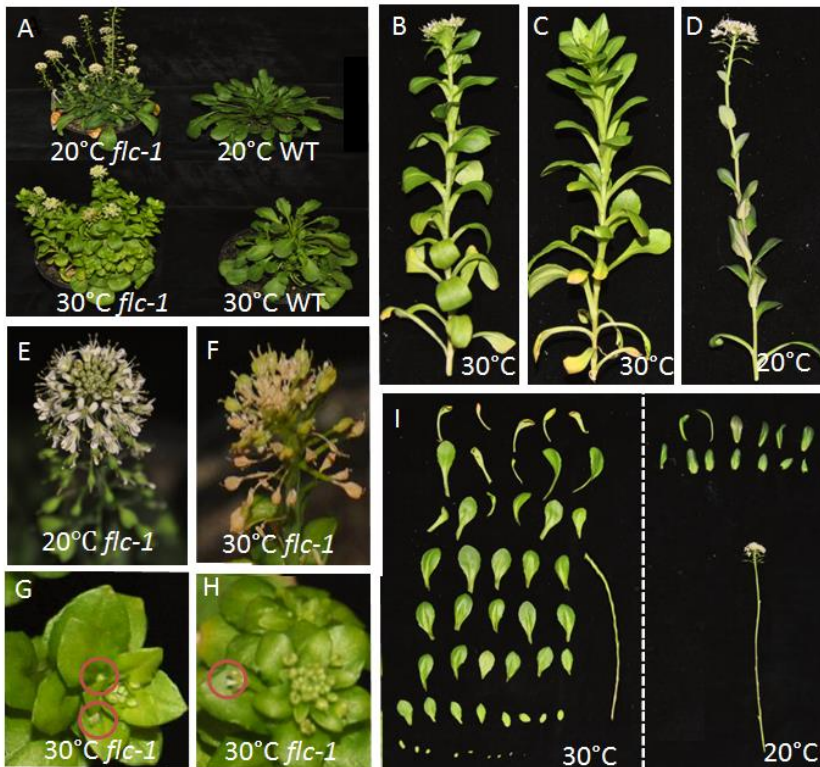


**Fig 4. Analysis of the *N. caerulea* FLC and SVP genomic DNA and derived coding sequences (ORF) in wild type (WT) and *flc* or *svp* mutant plants.** (A). Genomic DNA sequencing revealed point mutations in *NcFLC* in 3 out of 5 mutants, (B) CDS comparison among the *FLC* mutants and WT, for CDS of *flc1/2*, firm line stands from the transcribed intron after splicing. For both *AtFLC* and *NcFLC*, the 7 green boxes are exons, firm lines in between exons are intron regions, the red boxes are 5'UTR (left) and 3'UTR (right), (C). Genomic DNA sequencing revealed point mutations in *NcSVP* in 2 out of 5 mutants. (D). *SVP* CDS (ORF) comparison among the *SVP* mutants and WT. The organisation of *SVP* is presenting in the same way of *FLC*.

### The effect of ambient temperature on *N. caerulea* flowering

So far, the plants used in this study had been selected and grown during winter, in a heated greenhouse. When growing the mutants and wild type plants in late spring or summer, we noticed that these plants did not flower properly (Supplement Fig.3).

During these periods the greenhouse day temperature often exceeded 30°C, sometimes even 35°C. This prompted us to examine whether higher ambient temperatures affect flowering time in *N. caerulescens*. WT and *flc-1* mutant plants were therefore grown with the day temperature set at either 20°C or 30°C and the night temperature set at 18°C under long day conditions (16/8hs). The flowering time was determined as the number of days between sowing and bolting of each plant.



**Fig.5 The phenotype of wild type (WT) and *flc-1* mutant plants under different day temperatures.** (A) 2-month-old flowering plants of WT and *flc-1* mutant plants grown at 20 or 30°C. (B) The primary inflorescence, and (C) a secondary inflorescence of *flc-1* grown at 30°C. (D) The primary inflorescence of *flc-1* grown at 20°C (E) The flowering head of *flc-1* grown at 20 or (F) 30°C. (G-H) Poorly developed flowers and flower buds indicative of floral reversion to a vegetative phase in *flc-1* grown at 30°C. Single reverted flowers were in red circles. (I) The leaves on side inflorescence of *flc-1* mutant under 30/20°C. The side shoot under 20 similar to the main shoot, it is not shown here.

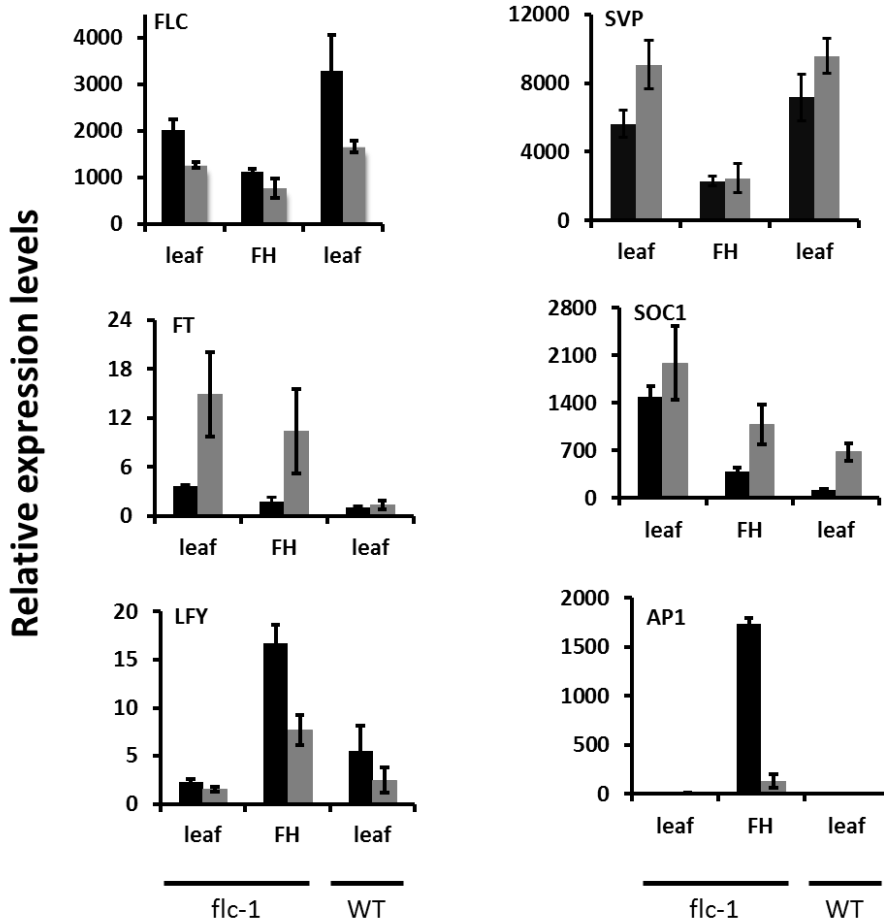
Various morphological differences were observed when comparing the *flc-1* mutants grown at these two temperatures (Fig.5.-*svp* and Ganges mutants in supplement Fig.3). The plants started bolting 50 days after sowing, in both temperature regimes. However the plants at 30°C took one additional week longer until the first flower opened. At 20°C the primary and secondary inflorescences of *flc-1* flowered properly after 10-13 cauline leaves had developed (Fig.5D and I). However, for the *flc-1* plants at 30°C, more than 30 leaves were formed before the primary inflorescence flowered (Fig.5B). Although secondary inflorescences were formed and had started to elongate, most of these inflorescences developed more than 50 leaves without any visible flowers (Fig.5C and I). If some flowers appeared on the primary inflorescences at 30°C, the siliques dried out (Fig.5F) and no seeds appeared to be formed. However, at 20°C siliques developed very well (Fig 5E) and more than 300 seeds were produced per plant. This indicates that the fitness of the plants is fine at an ambient day temperature of around 20°C, but strongly reduced when grown at 30°C.

### **Flowering time *N. caerulescens* gene expressions are affected by an elevated day temperature**

To further investigate the effect of ambient temperature differences on the flowering time of *N. caerulescens*, the expression of flowering time regulating genes was determined in rosette leaves and flower heads (FH: the top part of an inflorescence including flower buds, but no open flowers) of *flc-1* plants growing at 20 and 30°C day temperature regimes (Fig. 6). The analysis included the floral repressor genes *FLC* and *SVP*, the floral integrator genes *FLOWER LOCUS T (FT)* and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 (SOC1)*, as well as the floral organ identify genes *LEAFY (LFY)* and *APETALA1 (AP1)*. Higher *FLC* transcripts were detected in the rosette leaf of WT at 20°C than at 30°C, indicating the temperature has an effect at the expression level of *FLC*. *SVP* is down-regulated at 20°C in both the *flc-1* mutant and WT compared to 30°C with slightly lower expression being observed in the *flc-1* mutant. Both flower promotion genes, *FT* and *SOC1* are more up-regulated in leaves of the *flc-1* mutant compared to WT under both temperatures and their transcripts level were up-regulated under high temperature. In contrast, the expression level of the



floral organ identify genes *LFY* and *AP1*, especially *AP1*, were 14 times down-regulated in the flower heads under 30°C compared to that of in the 20°C.



**Fig 6. Gene expression of flowering time regulating *N. caerulea* genes in different organs and temperatures.** The relative expression levels of *FLC*, *SVP*, *FT*, *SOC1*, *LFY* and *AP1* are determined for rosette leaves (leaf) and flowering heads (FH) of SF wild-type (WT) and *flc-1* mutant plants grown at 20°C (black bars) or 30°C (grey bars). Expression levels are expressed relative to the expression of reference gene *NcTubulin*. Data is mean of three to five plants with standard error.

## Discussion

### **Natural variation for vernalisation requirement and response among *N. caerulea* accessions**

Plants ensure their reproduction at the most appropriate time and correct stage of development by monitoring their environment and internal signals. (Balasubramanian, Sureshkumar, Lempe and Weigel, 2006). Therefore the accurate regulation of the transition between vegetative and reproductive growth is critical for propagation and survival (Komeda, 2004). The promotion of flowering in response to prolonged exposure to cold temperatures (vernalisation) is an adaptation to prevent plants from flowering in the fall, prior to winter, and to enable them to flower in spring. Natural genetic variation in vernalisation requirement, together with the temperature regimes, defines when the plant begins to flower and is critical for adaptation to different environments. However the critical environmental conditions, such as vernalisation temperatures or length of cold exposure, required for the optimal vernalisation response vary among species (Kim *et al.*, 2010).

In order to investigate natural variation for vernalisation requirement in *N. caerulea* we selected four accessions from diverse environments. Our results provide evidence that vernalisation requirement and response among accessions varies within this *Noccea* species (Fig. 1). The flowering time differs among *N. caerulea* accessions depending on the length of the vernalisation treatment. Both GA and SF are biennials from a relative dry region in the south of France with a short winter (Dubois, 2005)(Mousset *et al.*, 2016). Consistent with this, a faster vernalisation response is predicted to confer a selective advantage to plants that flower earlier and thus avoid early summer droughts and high temperatures that result in poor seed set. LE and LC originate from Luxembourg and Belgium respectively and are biennial/perennial accessions, dealing with long winter and spring with milder temperature (8.5–9°C), but very dry summers (Dechamps *et al.*, 2007). This implies that these accessions have more opportunities to satisfy the longer vernalisation period they require and allow them flower later than South

France, but have to restrict the flowering time within short period before hot summer to avoid dry out.

### **The identification of floral repressors affected by vernalisation in *N. caerulescens***

Plants that require vernalisation to flower encode repressors that block flowering during summer or autumn, and this block is relieved by reducing the expression of the repressor when the plants are exposed to low, vernalising, temperatures (Andrés and Coupland, 2012). The floral repressors and regulatory framework for the vernalisation response varies greatly among different species. In the related model species *A. thaliana*, expression of the repressor *FLC* drops during vernalisation, upon which the transcription of *FT* is induced, which promotes floral initiation (Bastow *et al.*, 2004; Helliwell *et al.*, 2006; Kim *et al.*, 2010; Deng *et al.*, 2011). Similarly, *SVP* acts parallel to *FLC* to repress flowering in *A. thaliana* (Lee *et al.*, 2007; Mateos *et al.*, 2015). In the preliminary genome sequences of *N. caerulescens* we found one expressed *FLC* and one *SVP* orthologue (Lin *et al.*, 2014), as well as at least two expressed orthologues of *MAF* genes but no *FLM* and *FRI* orthologues were found at the corresponding co-linear sites where they reside in *A. thaliana* (Severing *et al.*, in preparation). Our finding that loss-of-function mutations of either *NcFLC* or *NcSVP* genes abolish the vernalisation requirement and cause early flowering, shows that *NcFLC* and *NcSVP* are the floral repressors in the control of flowering by the vernalisation pathway in *N. caerulescens*. The ability of the *flc-1* and *svp-1* mutants, we identified in the SF background, to complement the locus mutated in the two early flowering mutant lines in the GA background (Lochlainn *et al.*, 2011), indicates there is a third locus in *N. caerulescens* that controls flowering time. These allelic GA mutants were not affected in expression of *FLC*, *SVP*, or any of the *FLM/MAF*-like genes we identified, and did not show any mutations in the cDNA sequences of these genes (**data not shown**). Since none of the above mentioned candidate genes are affected, map based cloning and genome sequencing will be needed to identify the mutation. The observation that single mutants of all three genes result in early flowering indicates that they operate in mutual dependency.

***NcFLC* expression pattern comparison of *A. thaliana* and *A. alpina***

The expression pattern of *FLC* and *PEP1* differs in Annual and perennials with the *FLC* expressed stably repressed after the plants transferred to warm greenhouse for flowering in annual plant *A. thaliana*. However, in perennial *A. alpina* the expression of *PEP1*, the orthologue of *A. thaliana FLC*, rises again after the plants had flowered. By doing this, the subsequent shoot which had not yet flowered are blocked, which allows of these shoots flowering next year after vernalisation. Analysis of intra- and inter-species variation demonstrated that the structure of *PEP1* is more complex than has been found in *FLC*. *A. alpina* contains a tandem duplication of exon1 in *PEP1* which created two transcriptional start sites and two overlapping transcripts respectively *PEP1a* and *PEP1b*. Both transcripts show very similar patterns of transcriptional regulation before, during and immediately after vernalisation, but *PEP1b* expression persists more strongly in flowers and leaves, later after vernalisation than *PEP1a* transcript (Albani *et al.*, 2012). The organisation of *NcFLC* is more like the structure of *PEP1*, however, only one transcript has been detected in *N. caerulea* (Lin *et al.*, 2014).

The analysis of *FLC* expression after returning plants from cold to warm conditions suggested the repression of *NcFLC* is partial stable especially upon prolonged cold (Fig. 2). However, its expression increased gradually when transferred back to the warm greenhouse after a short period of 4 weeks of cold, especially in the case of SF. Such increase indicates incomplete vernalisation, reactivating *FLC* expression to perform the repression on flowering after transfer back to the warm greenhouse. A similar phenomenon was also observed after relatively short vernalisation times in some *Arabidopsis* accessions, especially those requiring very long vernalisation periods (Shindo *et al.*, 2006; Angel *et al.*, 2011). With the increase of the cold treatment, the *NcFLC* mRNA is much more stably repressed. Up to 10 weeks of cold treatment, upon which vernalisation the flowering was completely induced in all accessions, the slight increase of *FLC* mRNA detected was perhaps due to the increased turnover of the mRNA as plant growth accelerated (Shindo *et al.*, 2006). The most likely role of the recovery of *NcFLC* expression after flowering relates to the maintenance of later

formed meristems in a vegetative state, which is essential for perennial species such as *A. alpina* (Wang *et al.*, 2009; Albani *et al.*, 2012). Since both biennial and perennial plants were commonly found in the field for especially LC and LE accessions, the perpetual flowering habit might be part of the life style in some *N. caerulescens* accessions. It seems much rarer in the SF and GA accessions, but this most likely relates to summer conditions that do not support growth of *N. caerulescens* in the field. At those sites, only plant growing in the shade, with a little moisture, will stand a chance to survive. The partial stable expression of *NcFLC* seems to be intermediate between those of *FLC/PEP1* in *A. thaliana* and *A. alpina*, corresponding to a somewhat intermediate life style. Further research, including field observations, will be interesting to fully understand this complex trait.

### **The higher expression of *FT* and *SOC1* did not promote *LFY* and *AP1* expression under high temperature**

During the vernalisation, with the decrease of *FLC* expression, *FT* and *SOC1* was upregulated indicating the repression of *FLC* on flowering via binding to downstream gene *FT* and *SOC1* (**data not shown**). Both *FT* and *SOC1* were higher expressed in the non-functional *flc* *N. caerulescens* mutant than in WT, given the repression effects of *FLC* were eliminated. Here, we observed that plants display a different flowering depending on the growth season (Supplement Fig.4). When the temperature was higher than 30°C the inflorescences reverted back to vegetative growth (Fig.5 G-H) and the flowers became sterile, observations which promoted us to check the effect of day-time temperature on the flowering time control in the mutant under different temperature.

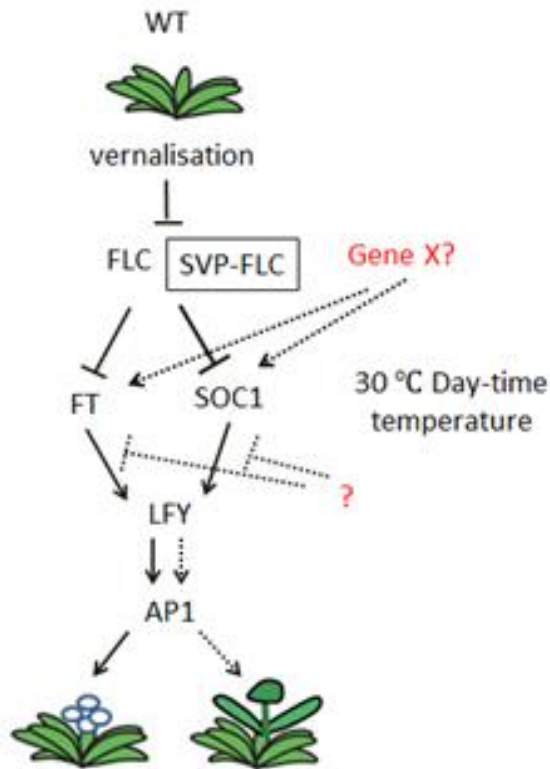
The up-regulation of *FT* in the *flc-1* mutant compared to WT under the same condition suggested that also in *N. caerulescens* repression on flowering occurs via the repression on the *FT* gene (Fig. 6). In the *flc-1* mutant *FT* and *SOC1* are highly expressed at 30°C. However, their downstream genes, respectively, *LFY* and *AP1* were notably lower expressed than under 20°C. We infer that the high expression of *FT* and *SOC1* are the direct effect of elevated temperature (30°C). In *A. thaliana* accession (Col), alternatively spliced isoforms of *FLM* and *MAF2* were up-regulated under high

temperature, which encode proteins that cannot combine with SVP to repress flowering (Posé *et al.*, 2013; Airoidi *et al.*, 2015). In *N. caerulea*, however, this scenario is unlikely given the fact that promotion of flower initiation in *flc-1* (*svp* and *Gange* mutants in supplement Fig.4) has not been observed at elevated temperature (30°C). Alternatively it was reported that in *A. thaliana* a repression complex formed by LIKE HETEROCHROMATIN PROTEIN 1 (LHP1) and POLYCOMB REPRESSIVE COMPLEX 2 (PRC2) is required to maintain repression of *FT* expression (Jiang *et al.*, 2008). In *N. caerulea* the repression of *FT* was not maintained under high temperature indicating that such complex might exist in this species. Future research could test if maintenance of such complex in *N. caerulea* would be sensitive to high temperature and thus explain loss of *FT* repression.

When *SOC1* is expressed in the meristem, it interacts with AGL24, another MADS box transcription factor, and this complex promotes the activation of transcription of *LEAFY* (*LFY*), a meristem identity gene that is involved in the initiation of flower development (Lee *et al.*, 2008). *APETALA1* (*AP1*), activated mainly by *FT*, is also necessary to establish and maintain flower meristem identity. Both *LFY* and *AP1* are pivotal for the specification of floral organ identities after the meristems have been converted towards floral identity (Komeda, 2004). The high expression of *FT* and *SOC1* in the *flc-1* mutant, however, did not induce the expression of *LFY* and *AP1* at 30°C, suggesting that the promotion of *LFY* and *AP1* by respectively *SOC1* and *FT* in *N. caerulea* was repressed, either directly by the high temperature or indirectly by other components involved in this process. Thus, retarded inflorescence development or reversion from floral identity meristems to vegetative meristems at 30°C (Fig. 5G-H) is very likely due to the low *LFY* and *AP1* expression level. The findings in this research indicated that under higher temperatures in the field, such reversion might be advantageous as it will suppress flowering in dry and hot summer conditions, which are not likely to be successful for reproduction.

### **An integrated flowering regulation model in *N. caerulea***

In annual plant *A. thaliana*, the *FLC* represses flowering via the repression of *FLC* on *SOC1* and *FT* in both leaf and meristem (Deng *et al.*, 2011). In perennial *A. alpina*, the



**Fig 7. Model of the flowering time regulation pathway in *N. caerulea*.** In general, the vernalisation pathway is more similar to *A. thaliana*. Vernalisation represses *FLC* expression, which removes the repression of *FLC* on downstream floral promotion genes *FT* and *SOC1*, resulting in the flowering initiation. In contrast to *A. thaliana*, under high temperatures (30°C), *FT* and *SOC1* are highly expressed, but they are no longer able to induce expression of the downstream floral identity genes *LFY* and *AP1* in *N. caerulea*, therefore, the floral initiation is delayed under high temperature. The dotted line stands for the possible effects imposed under 30°C. Gene *x* stands for the third unidentified floral repressor, which functions in parallel of *FLC* and *SVP* to repress flower in *N. caerulea*. The Question Mark (?) stands for the components that induce the expression of *LFY* or *AP1* might be sensitive to high temperature, which will result in low expression of *LFY* and *AP1*, consistently disrupting the floral organ development and maintenance.

flowering is induced after the repression of *TFL1* on *LFY*, along with the expression of *PEP1* were removed by vernalisation in the adult plant (Andrés and Coupland, 2012). The identification of genes that control flowering in *N. caerulea* allows us to compare the molecular pathways controlling seasonal flowering-responses in *N.*

*caerulescens* with those in *A. thaliana* and *A. alpina*. Based on the transcriptional analysis of the key vernalisation genes in the non-vernalisation requiring mutants, we propose a flowering time regulation model for *N. caerulescens* (Fig. 7). In general, this model is similar to that in *A. thaliana*. However, the vernalisation pathway in *N. caerulescens* is affected by down regulation of the floral identity genes *LFY* and *AP1* under high temperatures, similar as what we expect will happen in summer, despite the fact that under these conditions the expression of floral integrators *FT* and *SOC1* is upregulated. This temperature sensitivity of the regulators inhibiting the promotion of *LFY* and *AP1* expression by upstream genes will be an interesting topic for further research in this biennial/perennial species.



## Material and Methods

### Plant materials and growth conditions

*N. caerulescens* accessions Lellingen (LE), La Calamine (LC), Ganges (GA) are obtained by single seed descent propagation of lines described by Assunção *et al.* (2003). They originate from non-metallicolous soil in Wilwerwiltz, close to Lellingen in Luxemburg (49°59'1.83" N, 5°59'39.00" E); from calamine soil at the entry to a former Zn mine in La Calamine/Kelmis in Belgium (50°42'38.78" N, 6° 0'37.42" E); and most likely collected from calamine soil at a former Zn smelter in Les Avinières, close to St. Laurent le Minier in the south of France (43° 56' 11.2" N, 3° 40' 17.2" E). The accession Saint Felix de Pallières (SF) has also been collected in the south of France, from calamine soil at a former Zn mine (44° 2'40.03" N, 3°56'18.05" E) and was obtained from Dr. Henk Schat (Free University, Amsterdam). Before this experiment these accessions have been propagated by self-pollination, for at least 4-5 generations (SF) or more than 8 (LE, LC, GA) since collection in the field. Non-vernalisation requiring early flowering mutants were identified in an EMS-mutagenized M2 population (~8000 plants) generated in the inbred SF accession of (Wang *et al.*, manuscript in preparation). Seeds of two early flowering lines (A2 and A7) originating from one mutant event were obtained from Dr. Martin Broadley (Univ. of Nottingham, UK). The mutants were selected out from a fast-neutron mutagenized M2 population, the seeds used for mutagenesis were collected from wild growing plant *N. caerulescens* Gange (Lochlainn *et al.*, 2011)

For all the plants used in this experiment, the seeds were imbibed at 4°C for 4 days before sowing in pots containing mix soil (peat, sand plus nutrients) under standard greenhouse conditions (16-h light/8-h dark cycle) at 23°C and with 65% humidity. For the flowering time variation study, the plants were vernalized in cold room (at 4 °C, 16-h light/8-h dark cycle) for 4, 6,8,10 and 12 weeks after 60 days of vegetative growth, and subsequently transferred back to the same greenhouse (above) for flowering. To examine the temperature effect on the flowering time, seeds were sown under 20°C and 30°C chamber (16-h light/8-h dark cycle). The rosette leaves and

flowering heads (FHs) were collected for RNA extraction when the first flower had opened. The FH samples contained a few small, leaf-like bracts.

### **Gene expression analysis**

RNA was extracted from rosette leaves following the Trizol protocol (Oñate-Sánchez and Vicente-Carbajosa, 2008) and treated with DNA-free DNase (Promega, [www.promega.com](http://www.promega.com)). RNA quality and quantity were determined using agarose gel electrophoresis and a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). cDNA was synthesized from 1 µg of total RNA using a cDNA reverse transcription kit (iScript™cDNA Synthesis Kit, Bio-Rad) following the protocol recommended by the manufacturer.

Gene expression was determined upon quantitative reverse transcription-polymerase chain reaction (qRT-PCR) analysis. 10x dilution of the cDNA was used for each reaction by mixing 4 µl cDNA, 1 µl forward and reverse primers (3 µM) and 5 µl SYBR Green quantitative PCR buffer (Bio-Rad, Cat. no.18080-044). The primer sequences of the different genes are presented in Supplementary table 1. The following PCR protocol was used: denaturation at 95°C for 10 min, followed by 45 cycles of denaturation at 95°C for 10 s, annealing at the primer-specific annealing temperature for 30 s and extension at 70°C for 30 s. Following the last cycle, a melting curve was determined in the temperature range 57–95°C. A last step of cooling was performed at 40°C for 10s. The relative expression was determined based on delta Ct values. The expression levels were normalized to those of the housekeeping gene *NcTubulin*, the same gene as used by Nguyen (2010), but different primer sequences. Three to five biological replicates for each treatment were used for each analysis. Before performing the qRT-PCR, the primer efficiency and specificity of each gene was evaluated.

### **Analysis of candidate mutant alleles and amino acid sequences**

The coding regions of *NcFLC*, *NcSVP*, *NcMAF-like-1* and *NcMAF-like-2* were PCR-amplified from WT and mutant plants from by using 2 µl 10x diluted cDNA (same cDNA used for q-PCR) as PCR template. The PCR was performed by following the tag-

enzyme manual structuration (run for 35 cycles). The PCR products were gel-purified using the NucleoSpin® Gel and PCR Clean-up kit (<http://www.mn-net.com/>) and purified fragments were send for sequencing (Eurofin Genome Sequencing Company, Ebersberg, Germany). *FLC* and *SVP* genomic fragments (~600 bp) were amplified and sequenced in Eurofin Company. The cDNA and DNA sequences were aligned using Bioedit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). The primers were designed using Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>), based on preliminary whole genome *N. caerulescens* reference sequence of the Ganges accession or on published cDNA sequences (Halimaa *et al.*, 2014a; Lin *et al.*, 2014). The primer sequences are available in Supplemental Table1.

The amino-acid sequences of *FLC* from *A. thaliana*, *B. napus*, *B. rapa*, and the splice variants of *FLC* and *SVP* were aligned by using the Bioedit program. Identical amino acids are indicated by dots, the deletion of the amino acid was indicated by dashes line.

### **Genetic complementation analysis**

The three mutants carrying *flc* mutant alleles were inter-crossed and crossed with *svp* mutants (*svp-1* and putative *svp-2*) and WT. Both early flowering lines obtained from the Broadley lab (Lochlainn *et al.*, 2011) were crossed to *flc-1* and *svp-1*. The F1 plants were grown under control conditions (same as above). The flowering time was determined as the number of days after the first flower is open after seed sowing. The flowering time of backcrossed F2 derived from F1 crosses between early flowering plant and wild type (SF) were evaluated under the same condition.

### **The determination of inflorescence numbers**

The effect of a cold period on inflorescence number was evaluated by counting the number of inflorescence 30 days after transfer back to warm greenhouse, the subsequence inflorescence on the side shoots were not included. Ten plants per accession were scored for each treatment.

## Supplemental information

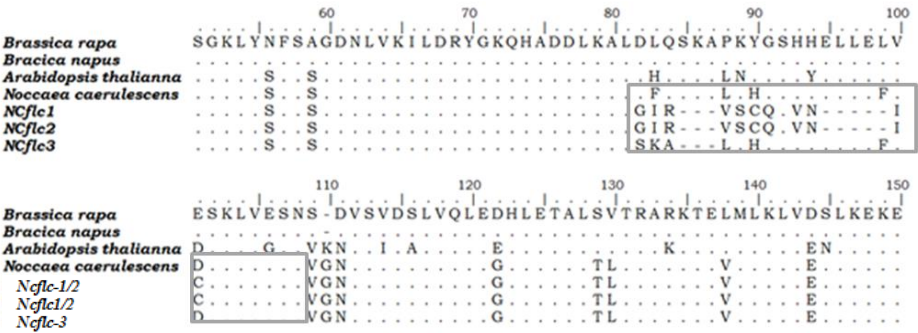
**Supplementary table 1A.** Primers sequences used for real-time quantitative Reverse Transcriptase-PCR.

Gene		Primer
<i>NcFLC</i>	Forward	GTGGAATCAAATGTCGGTAATGTA
	Reverse	CTCTAGTCAGAGTGAGGGCTGTCT
<i>NcSVP</i>	Forward	GCAACCGCGAGACAAGTAAC
	Reverse	TGACTGCAAGTTGTGCCTCT
<i>NcTubulin</i>	Forward	ACTTGGTCCCTTACCCGAGAATCC
	Reverse	CATGGAAGCTGGCTCGAAAGC
<i>NcSOC1</i>	Forward	GAGGCATACCAAGGATCGAA
	Reverse	TCTCCAAGAGTTTGCGTTT
<i>NcFT</i>	Forward	CATCGTATCGTGCTGGTATTGT
	Reverse	CTCACGAGTGTTGAAGTTTGA
<i>NcLFY</i>	Forward	ATCTTCCGTTTGAGCTTCTC
	Reverse	GGCGTCTAGAAGATTCCTCT
<i>NcAP1</i>	Forward	TTAGGGCACACAAGAGCAAT
	Reverse	CATGTAAGGGTGCTGGATTG

**Supplementary Table 1B.** Primers sequences used for normal PCR

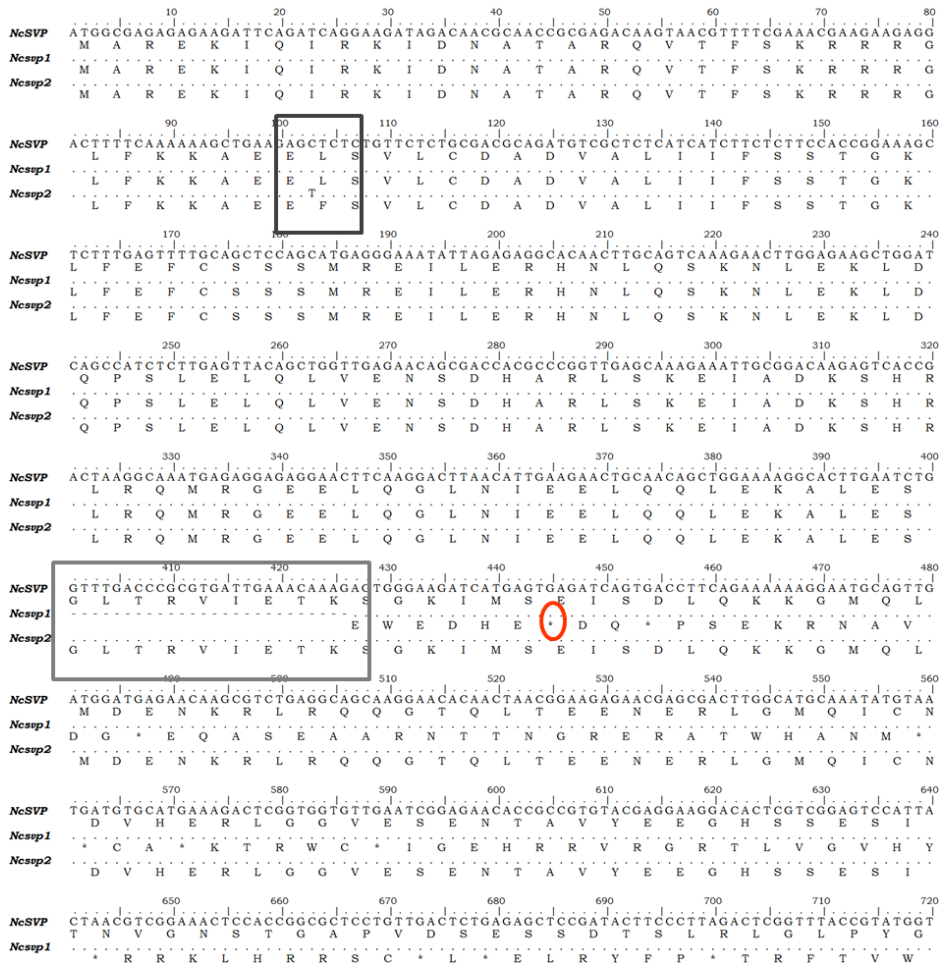
Gene		Primer
<i>NcMAF-like-1</i>	Forward	AGAGTCTCCGGACACCAAAA
	Reverse	GTTGATGACGGtGGTTACTTGA
<i>NcMAF-like-2</i>	Forward	CCGGACATCagACCcAAAT
	Reverse	GATGACGGTGGTTACTTGCG
<i>NcSVP</i>	Forward	CAAGAGTCACCGACTAAGGTACAC
	Reverse	CATCAACTGCATTCCCTAAACAC
<i>NcFLC</i>	Forward	AACATGCCGATGATCTTAAAGC
	Reverse	GAACGAGGGAATCGACACTTAC

Genetic control of flowering time in *Noccaea caerulea*

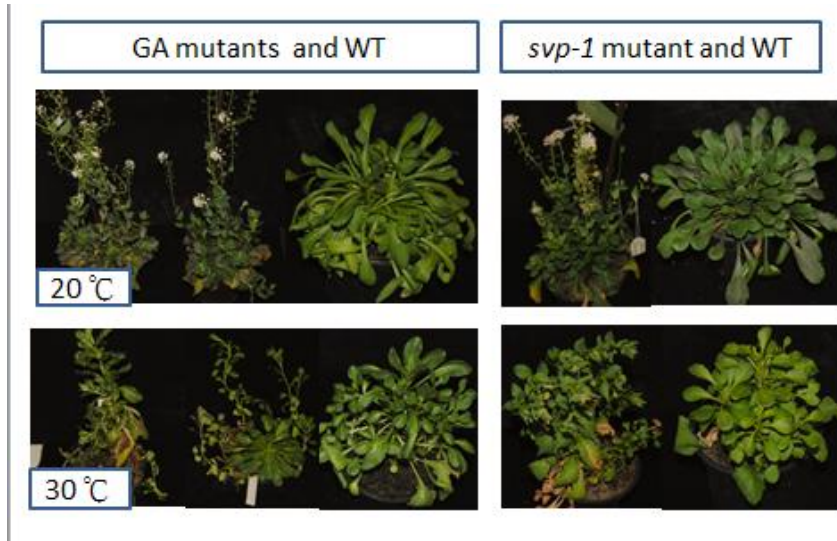


**Supplementary Fig 1. Sequence comparison of MADS domains of FLC from four different species.** Amino acid is expressed in single letter code, deletions that are part of MAD-Box K conserved domain are in grey box, and a dashed line stands for deletion regions as well as the new amino acid produced after the mutation events.

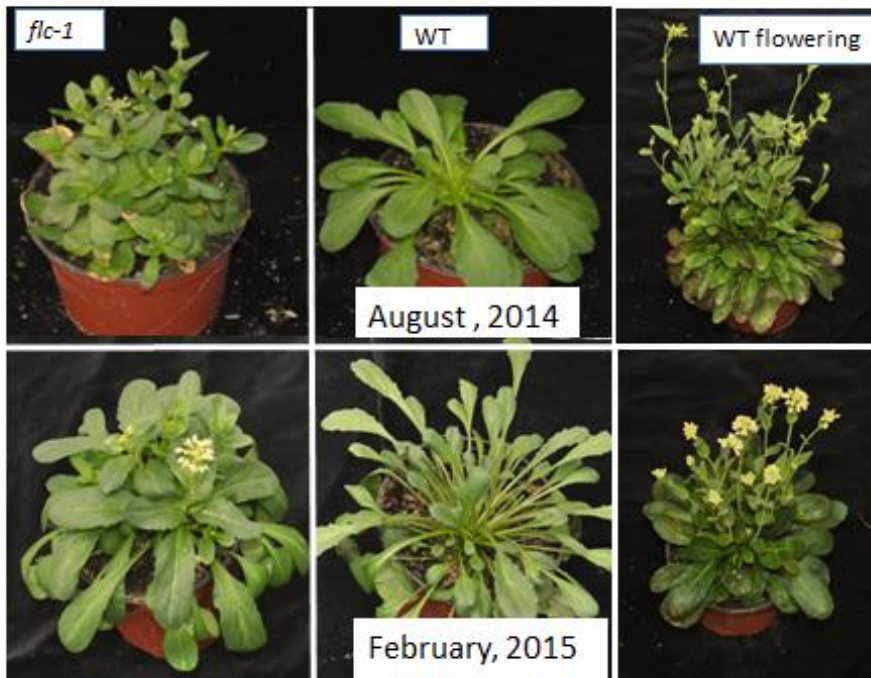
## Chapter 4



**Supplementary Fig 2. cDNA sequences and amino acid sequences comparisons.** The deletion in the svp-1 mutant is expressed in the grey box, and the red circle indicates one of the premature stop codons produced after shifting of the Open reading frame. The Amino acid substitution from Leu to Phe in svp-2 is shown in a black box.



**Supplementary Fig 3. The phenotype of wild type (WT) and *svp-1* and *Ganges* mutant plants under different day temperatures.** 2-month-old flowering plants of WT and *Ganges* and *svp-1* mutant plants grown at 20 or 30°C. The primary and secondary inflorescence of both mutant plants are flowering properly under 20°C, but poorly developed under 30°C.



**Supplementary Fig 4. The flowering phenotype of *flc-1* and wild type (WT) plants under different season.** Both the *flc-1* mutant and wild plant developed flowers properly in cooler season (February). In August, the temperature was above 30 °C when performing this experiment in 2014.



## **Acknowledgements**

This research was financially supported by the China Scholarship Council (CSC) and the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research).



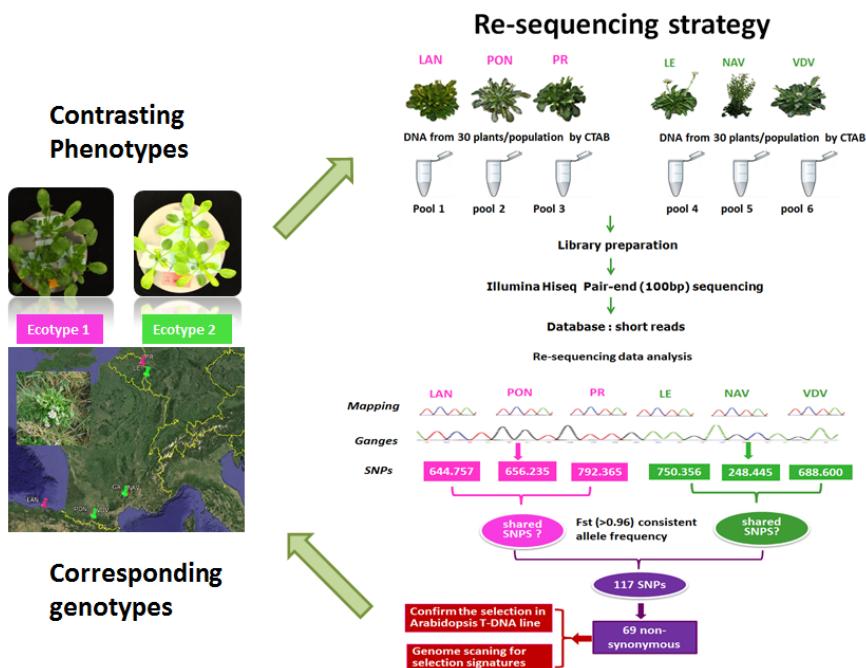
## **Chapter 5**

### **General discussion**

Plant species that are heavy metal hyperaccumulators have a higher capacity to accumulate metals as well as increased metal tolerance. *Noccaea caerulescens* is a Zn/Cd/Ni hyperaccumulator, whereas *Arabidopsis halleri* is a Zn/Cd hyperaccumulator. Researchers have worked on both species for decades as prominent models to study the evolution of the hyperaccumulation (Assunção *et al.*, 2003c). Their morphological and physiological differences, corresponding to variations in accumulation capacity and tolerance, have been well characterized (Clemens *et al.*, 2002; Lin and Aarts, 2012). The potential evolutionary changes at the transcriptional level have been investigated by transcriptomic profiling in *A. halleri* (Becher *et al.*, 2004; Weber *et al.*, 2004, 2006) and *N. caerulescens* (Rigola *et al.*, 2006; Hammond *et al.*, 2006; van de Mortel *et al.*, 2006; van de Mortel *et al.*, 2008; Plessl *et al.*, 2010; Halimaa *et al.*, 2014a; Lin *et al.*, 2014) predominantly under Zn or Cd treatment conditions. *N. caerulescens* is one of only a few species that grows on both metalliferous (Zn/Cd-enriched) and non-metalliferous (normal) soils, with metallicolous accessions typically accumulating less metal but showing greater metal tolerance than non-metallicolous accessions. The availability of the complete *N. caerulescens* genome sequence and the efficiency of next generation sequencing technology now make it easier and less expensive to inspect genome-scale variations responsible for the evolution of phenotypic diversity and tolerance in Zn/Cd hyperaccumulators, particularly wild populations growing in metalliferous and non-metalliferous soils.

This thesis describes a strategy to identify genetic variations in *N. caerulescens* under selection for Zn/Cd hyperaccumulation and tolerance by whole genome resequencing (**Figure 1**). The resulting data revealed the independent divergent selection of a number of genes in metallicolous and non-metallicolous populations. In addition a functional genomics tool for *N. caerulescens* was developed and tested on a population mutagenized using ethyl methanesulfonate (EMS) which will facilitate functional analysis in this species and functional comparisons with relatives. A genetic study of the early flowering mutants selected from the EMS-M2 population confirmed that *FLC* and *SVP* are the two major floral repressors in the *N. caerulescens* vernalisation

pathway. The transcriptional analysis of the floral repressors and key floral identity genes under different temperature regimes demonstrated that *N. caerulescens* has a unique response mechanism to vernalisation and ambient temperature compared to the annual species *Arabidopsis thaliana* and the perennial species *Arabis alpina*.



**Figure 1. Overview of the whole-genome resequencing strategy.** To identify *N. caerulescens* genetic variations under selection for Zn/Cd hyperaccumulation and tolerance, calamine (Zn/Cd-enriched) and non-metalliferous (normal) ecotypes with contrasting Zn/Cd tolerance and accumulation profiles were subject to whole-genome resequencing. The extremely divergent loci under selection in the coding regions were selected. Their function in relation to Zn/Cd accumulation and/or tolerance was established by studying *A. thaliana* T-DNA insertion mutations. Corresponding genotypes responsible for phenotypic variation between ecotypes were identified and confirmed.

### Molecular basis of heavy metal hyperaccumulation in *N. caerulescens*

The natural selection of mutations that improve fitness leads to reproductive advantages. Individuals with such mutations therefore become more prevalent in a population where there is selection pressure, leading to a change in the allele frequency within a population (Turner *et al.*, 2010). The combination of mutations in

different genes under long-term selection contributes to phenotypic evolution. At the molecular level, three types of mutations can be distinguished: (1) mutations in cis-regulatory regions such as the promoter, UTR or intron, or in genes that control transcription factors, which alter gene expression levels; (2) mutations in the coding region of a gene, which alter the function of the encoded protein; (3) mutations that change the copy number of a gene, which can lead to changes in expression levels and protein function due to divergence between copies. Specifically, copy number variation (CNV) reflects structural rearrangements such as gene duplication events, which affect the expression level, whereas mutations in the coding regions of different copies can lead to changes in protein function resulting in functionally distinct isoforms (Wray, 2007; Stern and Orgogozo, 2008).

Advances in molecular and developmental biology have revealed that mutations affecting gene expression are more likely to contribute to phenotypic evolution than mutations that modify the amino acid sequence of a protein (Stern and Orgogozo, 2008; Alonso-Blanco *et al.*, 2009). The modulation of gene expression levels due to natural genetic variation in cis-regulatory regions, or genes encoding transcription factors, is probably more flexible than structural mutations due to the nature of temporal and spatial expression profiles. Conversely, altered protein functions may induce more severe effects on plant development (Feder and Mitchell-Olds, 2003; Alonso-Blanco *et al.*, 2009).

The work described in this thesis focused on the analysis of nonsynonymous mutations because these alter the amino acid sequence and are thus more likely to affect protein structure, stability, activity and binding properties (Stern and Orgogozo, 2008). In comparison to mutations in cis-regulatory regions, mutations in coding regions are also easier to detect and predict. The functional validation of selected candidate loci in corresponding *A. thaliana* T-DNA insertion mutations for the impact on Zn/Cd tolerance and accumulation revealed two coding mutations that may be responsible for phenotypic variation in *N. caerulea* (**Chapter 2**). The nonsynonymous mutation in *RLK* was predicted to be detrimental to the protein, whereas the *GL3* polymorphism may affect protein function by controlling its stability

or activity. These genes did not show significant differential expression in the GA accession under different Zn concentrations (Lin *et al.*, 2014) nor between metallicolous (LC) and non-metallicolous (LE) accessions (Halimaa *et al.*, 2014a). However the most conclusive evidence of differing protein function requires the exchange of nucleotides and the analysis of the resulting phenotypes, which can be achieved using gene editing technology.

Among the major candidate genes obtained from transcript analysis were those involved in physiological processes that contribute to differences in metal accumulation and tolerance (Clemens *et al.*, 2002). The candidate genes from the population resequencing did not overlap with those detected in previous profiling studies. One potential explanation is that this study focused on extremely divergent loci ( $F_{st} > 0.96$ ) and metal homeostasis genes may not present as much diversity as those genes. Alternatively, because this study focussed on the allele frequency with a consistent pattern among geographically close pairs, the differences may reflect the non-comparable methodology and experiment design. Given that metal hyperaccumulation and tolerance is constitutive in *N. caerulescens* and *A. halleri* (Krämer, 2010), various selection pressures are likely to have been present during evolution rather than appearing solely due to mining/smelting and other human activities. There is substantial evidence that hyperaccumulation in plants probably evolved as a defence against the herbivores (Boyd and Martens, 1994; Pollard and Baker, 1997; Boyd, 2007; Krämer, 2010) (Lin *et al.*, unpublished data). Therefore, the non-overlapping loci detected in this and previous studies may reflect the various selection pressures that existed before the populations diverged within species. Accordingly, the loci selected at that time are older and must differ from those which have adapted more recently.

In the earlier transcript and transcriptome studies mentioned above, candidate genes with enhanced expression levels could also originate as gene duplicates derived from whole-genome duplication events, or single-copy genes duplicated double-stranded DNA break repair involving unequal crossover events (Jelesko *et al.*, 1999; Blanc *et al.*, 2003; Soltis and Soltis, 2003; Kramer *et al.*, 2004; Narayanan *et al.*, 2006; Yandea-

Nelson *et al.*, 2006). The comparative genome sequencing of six *N. caerulescens* populations revealed intra-species CNV, including the well-known examples of *HMA4*, which plays an essential role in Cd uptake and transport from the root to the shoot (Hanikenne *et al.*, 2008; Halimaa *et al.*, 2014a; Lin *et al.*, 2014, 2016), and Metallothionein 3 (*MT3*), which is involved in metal homeostasis (Milner and Kochian 2008). *MT3* and *HMA4* are very strongly expressed, indeed *HMA4* is the most abundantly transcribed metal transporter gene in *N. caerulescens* roots discovered thus far (Halimaa *et al.*, 2014b). Accordingly, three to five copies of these two genes were detected in the six populations. The intra-species CNV within *N. caerulescens* may have evolved as a response to selection pressure caused by exposure to metals. *HMA4* is one of many examples of a correlation between phenotypic variation and the expression of multiple gene copies (Hanikenne *et al.*, 2008; Craciun *et al.*, 2012; Iqbal *et al.*, 2013; Nouet *et al.*, 2015). Due to functional redundancy, the selection pressure on multi-copy genes is relaxed compared to single-copy loci, allowing them to diversify and acquire new protein functions (Stern and Orgogozo, 2008).

Transcription factor families are expressed and activated in different spatiotemporal patterns to regulate plant development and control the expression of downstream genes (Tomancak *et al.*, 2002). Therefore, genetic variations affecting transcription factors may have broad effects on gene expression profiles in plants (Wray *et al.*, 2003; Davidson, 2006). In the population resequencing study, *WRKY72* was detected in all the additional comparisons. This stress response gene may therefore be under selection via the regulation of its downstream genes, which could include metal transporters. The functional validation of this gene in corresponding *A. thaliana* T-DNA knockout mutants would allow functional annotation in relation to Zn/Cd accumulation or tolerance in *N. caerulescens*. Mutations in the cis-regulatory region, introns or splicing sites of genes revealed by previous transcript studies under excess Zn/Cd conditions may also be under selection, but this is still under investigation. If detected, the essential cis-regulatory region of the promoter should be identified by promoter deletion techniques as described for the metal homeostasis gene *ZNT1* (Lin *et al.* 2016). The sequence database from this study will provide a valuable resource



for future investigations involving *N. caerulea* and should also be compared to related taxa to achieve a better understanding of evolutionary scenarios.

Tajima's D is a statistical test that distinguishes mutations under natural selection from those with a neutral effect. The former would be allocated a negative Tajima's D value representing directional selection (Lohmueller *et al.*, 2011). Mutations under selection shift towards beneficial phenotypes during evolution. The *N. caerulea* gene with the most extreme Tajima's D value is not homologous to known genes involved in metal accumulation or tolerance. The species-dependent function of the genes could be identified by screening loss-of-function mutants in the TILLING population, along with *RLK*, *GL3* and other genes assumed to be involved in metal homeostasis.

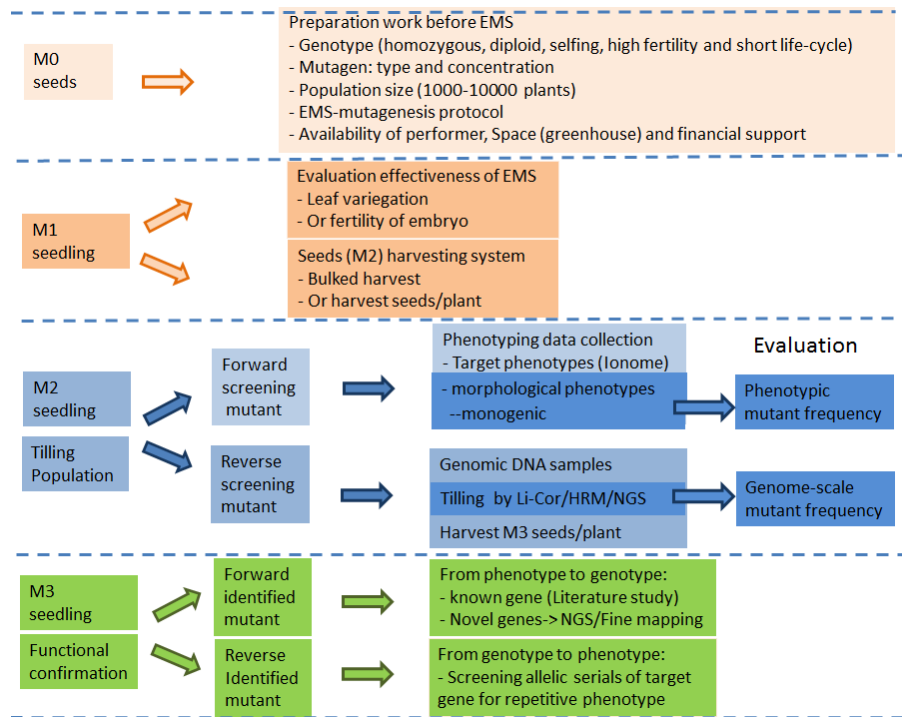
### **Development of a functional genomics tool in *N. caerulea***

Next generation sequencing is rapid and cost-efficient, allowing large-scale DNA sequencing projects to be carried out in both model and non-model organisms (Kaul *et al.*, 2000; Matsumoto *et al.*, 2005) (review: Bolger *et al.*, 2014). However, the functional analysis of genes is always more laborious, so many genes are currently characterized at the sequence level but their functions are unknown. Similarly, the impact of sequence variation on phenotypic diversity is also unclear. The traditional way of determine the relationship between genotype and phenotype (forward genetics, i.e. starting with a phenotype and finding the corresponding mutation) has given way to reverse genetics, where an uncharacterized sequence is mutated and tested to determine the impact on phenotype. Reverse genetics may be facilitated by genome-wide mutagenesis followed by screening to identify mutations causing interesting phenotypes, e.g. the Salk mutant collections generated by insertional mutagenesis in *A. thaliana* (Alonso *et al.*, 2003). Alternatively, specific genes can be targeted by gene knockout or knock-down, using techniques such as RNAi, amiRNA or CRISPR/Cas9 (Xie and Yang, 2013; Bortesi and Fischer, 2014). Whereas genome-wide mutagenesis is applicable in all species, specific gene targeting methods are species-dependent because they rely on the efficiency of gene transfer and expression in different species (Somers *et al.*, 2003; Stemple, 2004; Ko *et al.*, 2006).

The species-dependent limitations of gene targeting mean that general mutagenesis approaches are often favoured in non-model organisms that are recalcitrant to standard transformation methods. TILLING (targeting induced local lesions in genomes) is one such approach that was developed as a general strategy for reverse genetics (McCallum *et al.*, 2000a). Traditional mutagenesis (chemical or physical mutagenesis) is followed by high-throughput screening to discover point mutations in the targeted sequence. Almost all TILLING populations have been developed using chemical mutagens, primarily the alkylating agent ethyl methanesulfonate (EMS) but occasionally other chemicals such as sodium azide and/or methyl nitrosourea (MNU). EMS is preferred because it induces mutations at a higher frequency than the others allowing smaller populations to be screened (Greene *et al.*, 2003; Henikoff and Comai, 2003; Kurowska *et al.*, 2011). These benefits increase the likelihood of finding mutations even in small genes or cis-regulatory elements (Henikoff and Comai, 2003). An important advantage of TILLING is the ability to accumulate an allelic series of mutants with various modified protein functions ranging from mild impairment to complete loss of function, the latter generally produced by mutations that introduce premature stop codons or disrupt RNA splicing (Slade *et al.*, 2005). Such allelic series are desirable because they generate a range of phenotypes that provide more insight into the function of a single target gene. Several high-throughput screening platforms are available to identify mutations at the desired target gene, often combining PCR and traditional enzymatic digestion. In the standard approach, PCR amplification is followed by the specific digestion of mismatches using endonucleases such as CEL1 or END1 that nick single strands (Triques *et al.*, 2008, Oleykowski *et al.*, 1998), followed by polyacrylamide electrophoresis and visualization in the sensitive Li-Cor gel analyser system (Colbert *et al.*, 2001; Till *et al.*, 2006). More recently, non-enzymatic techniques such as high resolution melting curve analysis (HRM) have become popular because they are less expensive (Gady *et al.*, 2009), as well as next generation sequencing methods that also save time and reduce costs (Tsai *et al.*, 2011).

TILLING was first applied in the model plant *A. thaliana* (McCallum *et al.*, 2000a), but has been successful in many different species (reviews Kurowska *et al.*, 2011

(Gilchrist and Haughn, 2005; Martín *et al.*, 2009). Over the last few decades, TILLING has become a recognized tool in the field of functional genomics, with the largest databank for different species collected in the Seattle TILLING project (<http://comailab.genomecenter.ucdavis.edu/index.php/TILLING>). Thus far, TILLING has been widely used in *A. thaliana* and some agronomic crops, but has not been applied in metal hyperaccumulators such as *N. caerulescens*. Nevertheless, physiological differences between hyperaccumulators like *N. caerulescens* and non-hyperaccumulator species have been collected, revealing a number of genes that may explain these differences (Clemens *et al.*, 2002; Rigola *et al.*, 2006; Hammond *et al.*, 2006; van de Mortel *et al.*, 2006; van de Mortel *et al.*, 2008; Plessl *et al.*, 2010; Halimaa *et al.*, 2014a; Lin *et al.*, 2014). The functional analysis of such genes in *N. caerulescens* has not been possible thus far due to the absence of a reproducible transformation method. However, a *N. caerulescens* TILLING population has been created which is the first population generated specifically for functional genomics in this species (**Chapter 3**).



**Figure 2:** Overview of steps and crucial factors required for generating a Tilling platform. Corresponding to three generations for Tilling, the crucial decisions and data collection were listed respectively in each generation.

### **Crucial steps for each generation in a high-quality TILLING population**

Before EMS treatment (M0) certain factors must be taken into consideration, including the genotype to be used for mutagenesis. A fully homozygous diploid line is an ideal genotype configuration to initiate TILLING because it avoids any confusion between new mutations and existing polymorphisms, and a diploid species is better than a polyploid one because the genome is less complex and has fewer redundant genes (Talamè *et al.*, 2008). A high dosage of EMS is needed to achieve a mutation frequency sufficient to generate obvious phenotypes in a polyploidy species, e.g. one hit per 20–50 kb in wheat (Chen *et al.*, 2012a). However, such a high mutation frequency would be not tolerated in diploid species because it would result in extensive sterility, e.g. one hit per 300 kb in preferable in *A. thaliana* (Greene *et al.*, 2003). In diploid species, redundant genes exist due to ancient gene duplications events. The functional analysis of such genes therefore requires the availability of mutants affecting the individual paralogs – these can then be crossed to generate double mutants.

A self-compatible genotype that produces abundant seeds is beneficial because this helps to avoid contamination caused by out-crossing. Self-pollination also facilitates genetic analysis in the M3 generation. Furthermore, primer design and sequence analysis is facilitated if the mutagenized population has the same genotype as the population used to define the reference genome. The *N. caerulea* accession Saint Felix de Pallières (SF) used herein is geographically close (25 km) to the origin of the Ganges accession used for the reference genome. SF was mutagenized instead of Ganges because the latter is less efficient for self-pollination and produces fewer seeds.

After EMS mutagenesis (M1), two further issues are the evaluation of the effective EMS dosage and the optional harvesting system for M2 seeds. The seeds can be bulk harvested or harvested individually. In the standard TILLING procedure, seeds are harvested individually from M1 plants (Krysan, 2004) so that the same batch can be used for future forward screens and as a backup for the isolation of the same mutant allele known to segregate in a particular family. This is important when the phenotypically interesting M2 mutants are sterile and one relies on heterozygous

sister plants. However, considerable time is needed to harvest and clean individual plants (Gady *et al.*, 2009; Gottwald *et al.*, 2009; Porch *et al.*, 2009; Reddy *et al.*, 2012). Instead, seeds can be harvested in bulk from a limited number of plants (modified TILLING), which saves considerable time avoiding the need for labelling, harvesting and cleaning of the M1 plants. However, bulk harvesting causes problems if the mutant plant is sterile. This is not an issue when plants are harvested individually because it is easier to trace heterozygous and fertile relatives.

Once established, a TILLING population (M2) is a valuable source of mutations for both forward and reverse genetics. In many cases, the M2 generation is used to record phenotypic observations (Talamé *et al.*, 2008; Himmelblau *et al.*, 2009; Minoia *et al.*, 2010). In the standard TILLING approach, an informative phenotypic database for a saturated mutagenized M2 population comprising 5000–13,000 individuals would be required with a description of target phenotypes in every plant, as achieved in tomato (Minoia *et al.*, 2010) and sorghum (Xin *et al.*, 2008). Corresponding genomic DNA samples are collected from each plant for molecular and genetic research.

Forward screening in large M2 populations can be based on many different types of phenotypic data. The *N. caerulea* TILLING population was created to facilitate the identification of loss-of-function mutations involved in mineral uptake and distribution to provide a better understanding of the molecular mechanisms. However, the visible phenotypes caused by such loss-of-function mutations affecting homeostasis network remain unclear, so alternative readouts such as the concentration of selected minerals in the M2 population (the “ionome”) would be more straightforward (Salt *et al.*, 2008). The *N. caerulea* M2 population was screened for variations in the distribution of 20 different elements, producing a smaller collection of mutants that can be used to identify novel genes involved in different physiological processes, focusing on genes involved in Zn/Cd/Ni accumulation and tolerance (Clemens *et al.*, 2002).

Another possibility is to screen for easily scoreable traits in the M2 population, e.g. morphological phenotypes in the vegetative and reproductive stages (**Chapter 3**). The frequency of monogenic phenotypes could be used to estimate the overall frequency

of mutation, combined with allelism tests for mutants with similar unambiguous phenotypes (Koornneef et al. 1982), but accurate mutation frequencies require the analysis of sequences at specific loci in a TILLING experiment (Slade *et al.*, 2005; Till *et al.*, 2007). Gathering detailed morphological descriptions for each M2 individual is optional and was not applied to the *N. caerulea* TILLING population due to the amount of time required. However, such information has been collected in public-access platforms such as the Settle project for tomato, rice and wheat.

For reverse screening, the HRM platform was used to detect a mutant series in the *bZIP19* gene confirming that individuals carrying single mutations could be genotyped by HRM in this TILLING library. Although HRM is much less expensive than traditional approaches, one drawback is that sensitivity depends on the length of the DNA fragments that are screened. The optimal fragment length is 300–400 bp (Applied Biosystems, 2009; Gady *et al.*, 2009) so more time is required for primer design, specificity testing and PCR optimization. The advent of next generation sequencing allows thousands of pooled individuals from mutagenized populations to be sequenced and analysed without pre-screening steps (Deschamps and Campbell 2010). Whole genome re-sequencing of mutants with altered ionic profiles would therefore pinpoint novel genes involved in metal homeostasis. However, the detection of single-nucleotide polymorphisms requires deep coverage (usually 20–30 fold for a diploid species) and is still costly, although cost-per-base prices continue to fall (Tsai *et al.*, 2011).

*N. caerulea* is a biennial or facultative perennial plant from the same Brassicaceae family as *A. thaliana* and it has an obligate vernalisation requirement. When grown under greenhouse conditions, the accessions cultivated thus far require up to 32 weeks to flower, including a 7–12 week period of short-day vernalisation (5°C and 12 h or 8 h photoperiod) for young plants (2 months old) to induce flowering, with an additional 4 weeks for seed ripening, and a few extra weeks of dry storage to break seed dormancy (Assunção *et al.*, 2003c; Peer *et al.*, 2003, 2006; Lochlainn *et al.*, 2011; Guimarães *et al.*, 2013). These limitations, particularly the requirement for greenhouse space to achieve vernalisation during the winter, mean that only one

generation per year can be produced when developing the TILLING population. Practically, this makes it impossible to generate a reverse genetics population in less than 2 years, even without considering seed harvest and cleaning. A second round of TILLING using seeds from the early flowering mutant *flc-1* identified in this study therefore offers a promising strategy to create a parallel functional genomics tool in *N. caerulea*. In such a TILLING population, the challenge posed by the long period of vernalisation is eliminated. Furthermore, this mutant has an average life cycle of 3–4 months per generation, which speeds up the functional analysis and genetic complementation analysis. However, one disadvantage of the *flc-1* mutant as a basis for TILLING is that one the temperature must be maintained at ~20°C to avoid reversion (**see below**).

The EMS mutant population reported herein is the first functional genomics tool established for *N. caerulea*. A fully saturated population would allow the identification of mutant series affecting virtually any gene of interest, but genome-wide reverse genetics combined with targeted reverse genetics allows the comprehensive functional analysis of redundant genes, including those affected by CNV. The current reverse genetics technique of targeted genome editing using CRISPR/Cas9 technology would help to optimise both gene discovery and trait development because genomes can be modified rapidly and in a precise and predictable manner (Bortesi and Fischer, 2014; Woo *et al.*, 2015). However, this technique requires an efficient transformation method and is therefore likely to be applicable in a restricted range of genotypes.

### **Flowering time, a novel trait for ecological and evolutionary investigations**

*FLC* is the key floral repressor in *A. thaliana*, and *SVP* represses flowering by forming a transcriptional initiation complex either with or without *FLC* (Mateos *et al.*, 2015). There are five *FLC* paralogues in *A. thaliana* with *MAF2-5* span ~24 kb at the distal region of chromosome V, with interspaces of 1.5 kb (*MAF2-5*) (Ratcliffe *et al.*, 2001). The study of five early-flowering plants confirmed that the *NcFLC* and *NcSVP* genes encode the floral repressors of the vernalisation pathway in *N. caerulea*. However, in the Ganges early-flowering mutants (Lochlainn *et al.*, 2011), the mutation in the



third locus responsible for these early-flowering events has not yet been detected (**Chapter 4**).

A comparison of the annual species *A. thaliana* and its perennial relative *A. alpina* revealed that the expression of *PEP1/FLC* is distinct, and is more complex in perennial plants corresponding to the complex duplicated structure of the *PEP1* gene (Albani *et al.*, 2012). The structure of the single-copy *FLC* gene has been described in detail in *A. thaliana* (Michaels and Amasino, 1999). Comparative sequences revealed that *A. lyrata* harbours two full-length *FLC* genes, whereas *A. arenosa* contains two full copies and a partial copy (Nah and Jeffrey Chen, 2010). Similarly, *Brassica oleracea* contains a tandem duplication of *BoFLC1* (Razi *et al.*, 2008). In *A. alpina*, *PEP1* is partially duplicated (Wang *et al.*, 2009; Albani *et al.*, 2012). The structure of the *FLC* gene in *N. caerulea* resembles *PEP1*, with partial duplication of the first *FLC* exon (Lin *et al.*, 2014). The partial duplication of *FLC* found in *N. caerulea* and of *PEP1* in *A. alpina* may originate from a similar type of rearrangement, but the single copy of *FLC* in *A. thaliana* may also have originated from an ancestral *FLC* locus that contained two or three tandem copies. *N. caerulea* shares the structure of *PEP1* but the transcription profile in response to vernalisation is intermediate. *N. caerulea* is a biennial or facultative perennial species, so the intermediate expression pattern may correspond to the intermediate life history between the annual *A. thaliana* and perennial *A. alpina*. A further investigation of this trait in the field will provide more information about the correlation between this unique expression pattern and flowering time regulation in *N. caerulea*.

There are five *FLC* paralogues in *A. thaliana*, with four arranged as tandem duplications at the distal region of chromosome V (Ratcliffe *et al.*, 2001). The tandem duplications of the *MAF2-5* genes have been well-characterized in *A. thaliana* accessions. Mutations in the *MAF* genes give rise to an early-flowering phenotype confirming they function as floral repressors (Ratcliffe *et al.*, 2003). All the *MAF* genes respond to vernalisation. In response to prolonged cold, *MAF2-4* expression is downregulated but to a markedly lesser extent than *FLC*. *MAF2* acts downstream of *FLC* and controls an *FLC*-independent pathway to prevent premature vernalisation in

response to short periods of cold (Ratcliffe *et al.*, 2003). *MAF3* and *MAF4* probably maintain the vernalisation response. In contrast to the others, *MAF5* is upregulated by vernalisation indicating that it opposes *FLC* in the vernalisation response (Ratcliffe *et al.*, 2003). Conversely, *FLOWERING LOCUS M* (*FLM* or *MAF1*) is a repressor of flowering which together with *MAF2* accelerates flowering at higher growth temperatures (Posé *et al.*, 2013; Airoldi *et al.*, 2015).

Interestingly, *N. caerulea* has fewer *MAF* genes than *A. thaliana*. *N. caerulea* appears to have only one expressed *MAF* gene, which appears closely related to *MAF4* and *MAF5*. The second *MAF*-like gene is located between the *MAF2/MAF3/FLM* and *MAF4/MAF5* clusters (Edouard Severing *et al.*, in preparation). Comparison of *MAF2* and *MAF3* in *A. thaliana* accessions revealed high levels of nonsynonymous single-nucleotide polymorphisms, indels and rearrangements in these genes, including novel gene fusions involving *MAF2* and portions of *MAF3* (Caicedo *et al.*, 2009). The *MAF*-like genes present in *N. caerulea* may have undergone similar fusion events. *FLC* is a true homologue because the synteny around the *FLC* locus is perfectly conserved. In contrast, *FLM* appears to be missing in *N. caerulea*. Similar results have been reported in *Brassica rapa* by the Bonnema group (Edouard, *et al.*, in preparation). These data together suggest that these flowering repressor genes have undergone independent deletion events in *N. caerulea*.

Splicing variants in MADS box transcription factors play an important role in the control of flowering. Alternative splicing of the single copy of *FLC* in *A. thaliana* has also been described (Caicedo *et al.*, 2004). A missense mutation affecting the threonine at position 895 in *BRR2a* causes defects in *FLC* splicing and greatly reduces *FLC* transcript levels (Mahrez *et al.*, 2016). Natural mutations in *PEP1* generate two distinct transcripts with different expression profiles, matching the perennial life cycle of *A. alpina* (Albani *et al.*, 2012). The current *N. caerulea* EST database contains only one *FLC* transcript instead of two, although the structure of *NcFLC* looks more like *PEP1*. This suggests that only one functional transcript is present in *N. caerulea* or that the second transcript has not been detected due to the stage or tissue used for sampling (Lin *et al.*, 2014)

The clade members *FLM* (*MAF1*) and *MAF2–5* also undergo alternative splicing, and in the case of *FLM* and *MAF2* this may affect flowering time in response to higher ambient temperatures (Posé *et al.*, 2013; Airoidi *et al.*, 2015). In the *N. caerulea* Ganges mutant, two or three different transcripts have been identified for the *MAF*-like genes, although no mutations were detected in the genomic DNA. Such splicing differences may influence the flowering time in different *N. caerulea* accessions or modify their sensitivity to vernalisation, which will further experimental investigation.

Natural genetic variation in flowering time is well-characterized in *A. thaliana*, including variations in the response to vernalisation which have been traced to genetic variation in *FRI* (Werner *et al.*, 2005) and *FLC* (Sheldon, 2002; Li *et al.*, 2014). In contrast, the major source of variation in flowering time in response to vernalisation and seasonal perennial flowering in *A. alpina* is *PEP1* (Albani *et al.*, 2012). Natural genetic variation and alternative splicing in relation to flowering time under ambient temperature conditions has been reported for *MAF2* genes in *Arabidopsis* spp. (Ratcliffe *et al.*, 2003; Alexandre and Hennig, 2008; Rosloski *et al.*, 2013). The relevance of *MAF* gene variants in the context of vernalisation remains elusive. Consistent with the different vernalisation requirements in *N. caerulea* accessions, the initial *NcFLC* mRNA level required to repress flowering varies greatly in an accession-dependent manner (**Chapter 4**). Resequencing revealed single nucleotide polymorphisms in the first intron and promoter region of *NcFLC*. However, the extent to which this *FLC* sequence variation accounts for natural variation in the vernalisation response can only be determined by collecting more flowering time phenotypes and genetic data in *N. caerulea*.

The mutation causing early flowering in the Ganges mutant (Lochlainn *et al.*, 2011) has not yet been identified (**Chapter 4**). This phenotype may reflect a mutation in a *MAF*-like gene, although the evidence has not yet been detected possibly due to the complexity of the structural rearrangements of the *MAF*-like genes in *N. caerulea*. If *MAF*-like genes are not responsible, then novel genes involved in this early flowering mutant must be identified by map-based cloning and genome sequencing (James *et al.*, 2013; Schneeberger, 2014). Further investigation is required to correlate

alternative splicing in *MAF*-like genes and vernalisation requirements, which would provide more insight into the regulation of flowering time regulation in *N. caerulea*.

Although the control of flowering time in the vernalisation pathway is understood in detail, less is known about the effect of ambient temperature on flowering time (Posé *et al.*, 2013; Airoidi *et al.*, 2015) and almost nothing is known about the interaction between these pathways (**Chapter 5**). A successful flowering event requires the completion of two defined stages: floral induction and floral organ formation/maintenance. The former begins with the expression of the floral promotion genes *FT* and *SOC1* following the removal of *FLC* by sufficient vernalisation (Angel *et al.*, 2011). Thereafter, *FT* and *SOC1* activate the expression of the downstream floral identity genes *LFY* (Lee *et al.*, 2008) and *SOC1* (Komeda, 2004), respectively, to initiate floral organ development. The *FT* and *SOC1* genes in *N. caerulea* are more strongly expressed in the flowering heads at 30°C than 20°C, but *LFY* and *AP1* are expressed at minimal levels. Accordingly, if the development of existing inflorescences (flowering heads) is disturbed, the organs can revert to vegetative growth (**Chapter 5**). This may represent a novel evolutionary strategy to protect the plant against desiccation at high temperatures, thus permitting flowers to develop only when optimal conditions are available. Alternatively, this mechanism may repress subsequent meristems and allow flowering the following year as part of the perennial life cycle. Similar phenomena have not yet been reported in the model species *A. thaliana* and perennial *A. alpina*, but may be adapted across the species range as a strategy to cope with complex environmental conditions.

### ***Noccaea caerulea*, a model plant for evolutionary studies**

The discipline of ecological and evolutionary functional genomics (EEFG) has for years aimed to understand the function of genetic variation shaped by ecological forces in relation to the fitness of individuals in a population (Feder and Mitchell-Olds, 2003; Mitchell-Olds *et al.*, 2008). However, little progress has been made in the analysis of plants, compared to microbes and animals mainly due to the absence of a model plant suitable for EEFG (Wray, 2007).

Genetic variation in plants has been under long-term selection due to the presence of biotic and abiotic stress (e.g. herbivory, pathogens, heat, drought and cold) so phylogenetically related adapted and non-adapted species pairs may no longer exist. The absence of comparative analysis between adapted and non-adapted relatives makes it difficult to understand molecular mechanisms that have evolved as part of specialisation. In contrast, diverged ecotypes with contrasting hyperaccumulation and/or tolerance traits may have evolved recently, due to mining, smelting and other human activities. The increase in the number of hyperaccumulators indicates that selection pressure caused by heavy metals is stronger, and rapid adaptation would be necessary to survive (Antonovics, 1975; Wu *et al.*, 1975). The tolerance that has evolved in a few generations suggested that the effect of strong selection pressure on phenotype may be predictable over short periods of time (Wu *et al.*, 1975). These data suggest that heavy metal hyperaccumulation could be used for ecological and evolutionary analysis.

*N. caerulescens*, a Zn/Cd/Ni hyperaccumulator, is one of the two most prominent model species for heavy metal hyperaccumulation (Assunção *et al.*, 2003c; Krämer, 2010). It is widely distributed in Europe, with three distinguishable ecotypes and many more small populations, providing sufficient genetic variation for adaptive evolution in the presence of toxic metal concentrations. The reference genome sequence of the inbred line Ganges has a size of ~319 Mb (Edouard Severing *et al.*, in preparation). A comprehensive transcriptome database has been assembled with 29,712 predicted genes representing 15,874 gene families (Lin *et al.*, 2014). *N. caerulescens* is a member of the Brassicaceae and thus facilitates comparative genome analysis with other brassicas to identify speciation events (Edouard Severing *et al.*, In preparation), as well as providing a basis for population genome resequencing comparisons to identify genetic variations responsible for phenotypic diversity among ecotypes (**Chapter 2**).

As well as identifying genetic variation that can help to understand ecological and evolutionary forces, the functional analysis of genetic variation in *N. caerulescens* has been made possible through the whole-genome functional analysis tool (**Chapter 3**)

and targeted reverse genetics using *Agrobacterium rhizogenes* (Limpens *et al.*, 2004; Iqbal *et al.*, 2013). These tools initially developed for the model plant *A. thaliana* could facilitate the functional analysis of *N. caerulea* despite the absence of facile transformation methods (Peer *et al.*, 2003; Guan *et al.*, 2008).

The resources now available for the heavy metal hyperaccumulator *N. caerulea*, along with data from this population resequencing project (**Chapter 2**) and additional evidence of genetic variation obtained by studying the classical ecological and evolutionary trait flowering time (**Chapter 4**), have identified multiple evolutionary events that underpin the adaption of *N. caerulea* to selective environments. This species is not only a model for heavy metal hyperaccumulation, but also provides a comprehensive system that will help to explain the molecular mechanisms underlying phenotypic variation for other traits shaped by diverse ecological and evolutionary forces.

## References

- Adriano. 2001.** Trace elements in terrestrial environments. Biogeochemistry, bioavailability and risks of metals. *Springer-Verlag, New York* **32**: 374.
- Aikawa S, Kobayashi MJ, Satake A, Shimizu KK, Kudoh H. 2010.** Robust control of the seasonal expression of the Arabidopsis *FLC* gene in a fluctuating environment. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 11632–11637.
- Airoidi CA, McKay M, Davies B. 2015.** *MAF2* Is Regulated by Temperature-Dependent Splicing and Represses Flowering at Low Temperatures in Parallel with *FLM*. *Plos One* **10**: e0126516.
- Albani MC, Castaings L, Wötzel S, Mateos JL, Wunder J, Wang R, Reymond M, Coupland G. 2012.** *PEP1* of Arabis alpina Is Encoded by Two Overlapping Genes That Contribute to Natural Genetic Variation in Perennial Flowering. *PLoS Genetics* **8**.
- Alexandre CM, Hennig L. 2008.** *FLC* or not *FLC*: the other side of vernalization. *Journal of experimental botany* **59**: 1127–35.
- Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al. 2003.** Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science (New York, N.Y.)* **301**: 653–657.
- Alonso-Blanco C, Aarts MGM, Bentsink L, Keurentjes JJB, Reymond M, Vreugdenhil D, Koornneef M. 2009.** What has natural variation taught us about plant development, physiology, and adaptation? *The Plant cell* **21**: 1877–96.
- Alonso-Blanco C, Koornneef M. 2000.** Naturally occurring variation in Arabidopsis: An underexploited resource for plant genetics. *Trends in Plant Science* **5**: 22–29.
- Amasino RM. 2005.** Vernalization and flowering time. *Current Opinion in Biotechnology* **16**: 154–158.
- Andrés F, Coupland G. 2012.** The genetic basis of flowering responses to seasonal cues. *Nature reviews. Genetics* **13**: 627–39.
- Angel A, Song J, Dean C, Howard M. 2011.** A Polycomb-based switch underlying



quantitative epigenetic memory. *Nature* **476**: 105–8.

**Antonovics J. 1975.** Metal tolerance in plants perfecting an evolutionary paradigm. *International Conference on Heavy Metals in the Environment*: 169–186.

**Assunção AGL, Bookum WM, Nelissen HJM, Vooijs R, Schat H, Ernst WHO. 2003.** Differential metal-specific tolerance and accumulation patterns among *Thlaspi caerulescens* populations originating from different soil types. *New Phytologist* **159**: 411–419.

**Assunção AGL, Ten Bookum WM, Nelissen HJM, Vooijs R, Schat H, Ernst WHO. 2003b.** A cosegregation analysis of zinc (Zn) accumulation and Zn tolerance in the Zn hyperaccumulator *Thlaspi caerulescens*. *New Phytologist* **159**: 383–390.

**Assunção AGL, Herrero E, Lin YF, Huettel B, Talukdar S, Smaczniak C, Immink RG, van Eldik M, Fiers M, Schat H, et al. 2010.** *Arabidopsis thaliana* transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc Natl Acad Sci U S A* **107**: 10296–10301.

**Assunção AGL, Pieper B, Vromans J, Lindhout P, Aarts MGM, Schat H. 2006.** Construction of a genetic linkage map of *Thlaspi caerulescens* and quantitative trait loci analysis of zinc accumulation. *New phytologist* **170**: 21–32.

**Assunção AGL, Schat H, Aarts MGM. 2003c.** *Thlaspi caerulescens*, an attractive model species to study heavy metal hyperaccumulation in plants. *New Phytologist* **159**: 351–360.

**Di Baccio D, Kopriva S, Sebastiani L, Rennenberg H. 2005.** Does glutathione metabolism have a role in the defence of poplar against zinc excess? *New Phytologist* **167**: 73–80.

**Baker AJM. 1987.** Metal tolerance. *New Phytologist* **106**: 93–111.

**Baker AJM, Brooks RR. 1989.** Terrestrial higher plants which hyperaccumulate metallic elements - a review of their distribution, ecology and phytochemistry. *Biorecovery* **1**: 81–126.

- Baker AJM, McGrath SP, Sidoli CMD, Reeves RD. 1994.** The possibility of in situ heavy metal decontamination of polluted soils using crops of metal-accumulating plants. *Resources, Conservation and Recycling* **11**: 41–49.
- Balasubramanian S, Sureshkumar S, Lempe J, Weigel D. 2006.** Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. *PLoS Genetics* **2**: 0980–0989.
- Bargmann CI. 2001.** High-throughput reverse genetics: RNAi screens in *Caenorhabditis elegans*. *Genome biology* **2**: REVIEWS1005.
- Barkley NA, Wang ML. 2008.** Application of TILLING and EcoTILLING as Reverse genetics Approaches to Elucidate the Function of Genes in Plants and Animals. *Current genomics* **9**: 212–226.
- Basic N, Besnard G. 2006.** Gene polymorphisms for elucidating the genetic structure of the heavy-metal hyperaccumulating trait in *Thlaspi caerulescens* and their cross-genera amplification in Brassicaceae. *Journal of Plant Research* **119**: 479–487.
- Basic N, Salamin N, Keller C, Galland N, Besnard G. 2006.** Cadmium hyperaccumulation and genetic differentiation of *Thlaspi caerulescens* populations. *Biochemical Systematics and Ecology* **34**: 667–677.
- Bastow R, Mylne JS, Lister C, Lippman Z, Martienssen RA, Dean C. 2004.** Vernalization requires epigenetic silencing of *FLC* by histone methylation. *Nature* **427**: 164–167.
- Becher M, Talke IN, Krall L, Krämer U. 2004.** Cross-species microarray transcript profiling reveals high constitutive expression of metal homeostasis genes in shoots of the zinc hyperaccumulator *Arabidopsis halleri*. *The Plant journal: for cell and molecular biology* **37**: 251–268.
- Bernard A. 2008.** Cadmium & its adverse effects on human health. *The Indian journal of medical research* **128**: 557–564.
- Bernhardt C, Lee MM, Gonzalez A, Zhang F, Lloyd A, Schiefelbein J. 2003.** The bHLH genes *GLABRA3* (*GL3*) and *ENHANCER OF GLABRA3* (*EGL3*) specify epidermal

cell fate in the Arabidopsis root. *Development (Cambridge, England)* **130**: 6431–6439.

**Bernhardt C, Zhao M, Gonzalez A, Lloyd A, Schiefelbein J. 2005.** The bHLH genes *GL3* and *EGL3* participate in an intercellular regulatory circuit that controls cell patterning in the Arabidopsis root epidermis. *Development* **132**: 291–298.

**Besnard G, Basic N, Christin PA, Savova-Bianchi D, Galland N. 2009.** *Thlaspi caerulescens* (Brassicaceae) population genetics in western Switzerland: Is the genetic structure affected by natural variation of soil heavy metal concentrations? *New Phytologist* **181**: 974–984.

**Blanc G, Hokamp K, Wolfe KH. 2003.** A recent polyploidy superimposed on older large-scale duplications in the Arabidopsis genome. *Genome Research* **13**: 137–144.

**Bolger ME, Weisshaar B, Scholz U, Stein N, Usadel B, Mayer KFX. 2014.** Plant genome sequencing - applications for crop improvement. *Current Opinion in Biotechnology* **26**: 31–37.

**Bond DM, Wilson IW, Dennis ES, Pogson BJ, Jean Finnegan E. 2009.** *Vernalization insensitive 3 (VIN3)* is required for the response of *Arabidopsis thaliana* seedlings exposed to low oxygen conditions. *Plant Journal* **59**: 576–587.

**Bortesi L, Fischer R. 2014.** The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnology advances* **33**: 41–52.

**Boyd RS. 2007.** The defense hypothesis of elemental hyperaccumulation: Status, challenges and new directions. *Plant and Soil*. 153–176.

**Boyd RS, Martens SN. 1994.** Nickel hyperaccumulated by *thlaspi montanum* var. *montanum* is acutely toxic to an insect herbivore. *Oikos* **70**: 21–25.

**Bradley A, Zheng B, Liu P. 1998.** Thirteen years of manipulating the mouse genome: A personal history. *International Journal of Developmental Biology* **42**: 943–950.

**Brady KU, Kruckeberg AR, Bradshaw Jr. HD. 2005.** Evolutionary Ecology of Plant Adaptation To Serpentine Soils. *Annual Review of Ecology, Evolution, and Systematics* **36**: 243–266.

- Broadley MR, White PJ, Hammond JP, Zelko I, Lux A. 2007.** Zinc in plants: Tansley review. *New Phytologist* **173**: 677–702.
- Brooks RR, Lee J, Reeves RD, Jaffre T. 1977.** Detection of nickeliferous rocks by analysis of herbarium specimens of indicator plants. *Journal of Geochemical Exploration* **7**: 49–57.
- Burn JE, Smyth DR, Peacock WJ, Dennis ES. 1993.** Genes conferring late flowering in *Arabidopsis thaliana*. *Genetica* **90**: 147–155.
- Caicedo AL, Richards C, Ehrenreich IM, Purugganan MD. 2009.** Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana* *MAF2-5* flowering time gene cluster. *Molecular Biology and Evolution* **26**: 699–711.
- Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD. 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 of the United States of America* **101**: 15670–5.
- Calgaroto NS, Cargnelutti D, Rossato LV, Farias JG, Nunes ST, Tabaldi LA, Antes FG, Flores EMM, Schetinger MRC, Nicoloso FT. 2011.** Zinc alleviates mercury-induced oxidative stress in *Pfaffia glomerata* (Spreng.) Pedersen. *BioMetals* **24**: 959–971.
- Cambrollé J, Mancilla-Leytón JM, Muñoz-Vallés S, Luque T, Figueroa ME. 2012.** Tolerance and accumulation of copper in the salt-marsh shrub *Halimione portulacoides*. *Marine Pollution Bulletin* **64**: 721–728.
- Cempel M, Nikel G. 2006.** Nickel: A review of its sources and environmental toxicology. *Polish Journal of Environmental Studies* **15**: 375–382.
- Chaney RL, Malik M, Li YM, Brown SL, Brewer EP, Angle JS, Baker AJM. 1997.** Phytoremediation of soil metals. *Current Opinion in Biotechnology* **8**: 279–284.
- Chen L, Huang L, Min D, Phillips A, Wang S, Madgwick PJ, Parry M a J, Hu YG. 2012a.** Development and characterization of a new TILLING population of common

bread wheat (*Triticum aestivum* L). *PLoS ONE* **7**.

**Chen L, Song Y, Li S, Zhang L, Zou C, Yu D. 2012b.** The role of WRKY transcription factors in plant abiotic stresses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1819**: 120–128.

**Choi K, Kim J, Hwang H-J, Kim S, Park C, Kim SY, Lee I. 2011.** The *FRIGIDA* complex activates transcription of *FLC*, a strong flowering repressor in Arabidopsis, by recruiting chromatin modification factors. *The Plant cell* **23**: 289–303.

**Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. 2012.** Predicting the functional effect of amino acid substitutions and indels. *PloS one* **7**: e46688.

**Clarke JH, Dean C. 1994.** Mapping *FRI*, a locus controlling flowering time and vernalization response in *Arabidopsis thaliana*. *MGG Molecular & General Genetics* **242**: 81–89.

**Clemens S. 2001.** Molecular mechanisms of plant metal tolerance and homeostasis. *Planta* **212**: 475–486.

**Clemens S, Palmgren MG, Krämer U. 2002.** A long way ahead: Understanding and engineering plant metal accumulation. *Trends in Plant Science* **7**: 309–315.

**Colbert T, Till BJ, Tompa R, Reynolds S, Steine MN, Yeung a T, McCallum CM, Comai L, Henikoff S. 2001.** High-throughput screening for induced point mutations. *Plant physiology* **126**: 480–484.

**Comai L, Henikoff S. 2006.** TILLING: Practical single-nucleotide mutation discovery. *Plant Journal* **45**: 684–694.

**Cooper JL, Till BJ, Laport RG, Darlow MC, Kleffner JM, Jamai A, El-Mellouki T, Liu S, Ritchie R, Nielsen N, et al. 2008.** TILLING to detect induced mutations in soybean. *BMC plant biology* **8**: 9.

**Coustham V, Li P, Strange A, Lister C, Song J, Dean C. 2012.** Quantitative modulation of polycomb silencing underlies natural variation in vernalization. *Science (New York, N.Y.)* **337**: 584–7.

- Craciun AR, Meyer CL, Chen J, Roosens N, De Groodt R, Hilson P, Verbruggen N. 2012.** Variation in HMA4 gene copy number and expression among *Noccaea caerulea* populations presenting different levels of Cd tolerance and accumulation. *Journal of Experimental Botany* **63**: 4179–4189.
- Cutter AD, Payseur BA. 2013.** Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature reviews. Genetics* **14**: 262–74.
- DalCorso G, Farinati S, Furini A. 2010.** Regulatory networks of cadmium stress in plants. *Plant signaling & behavior* **5**: 663–667.
- Dalmaï M, Antelme S, Ho-Yue-Kuang S, Wang Y, Darracq O, d'Yvoire MB, Cézard L, Légée F, Blondet E, Oria N, et al. 2013.** A TILLING Platform for Functional Genomics in *Brachypodium distachyon*. *PLoS ONE* **8**.
- Davidson EH. 2006.** The regulatory genome: gene regulatory networks in development and evolution. *Developmental Biology* **310**: 304.
- Dechamps C, Lefèbvre C, Noret N, Meerts P. 2007.** Reaction norms of life history traits in response to zinc in *Thlaspi caerulescens* from metalliferous and nonmetalliferous sites. *New Phytologist* **173**: 191–198.
- Deng W, Ying H, Helliwell CA, Taylor JM, Peacock WJ, Dennis ES. 2011.** *FLOWERING LOCUS C (FLC)* regulates development pathways throughout the life cycle of Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 6680–6685.
- Deniau AX, Pieper B, Ten Bookum WM, Lindhout P, Aarts MGM, Schat H. 2006.** QTL analysis of cadmium and zinc accumulation in the heavy metal hyperaccumulator *Thlaspi caerulescens*. *Theoretical and applied genetics*. **113**: 907–20.
- Dennis ES, Peacock WJ. 2007.** Epigenetic regulation of flowering. *Current Opinion in Plant Biology* **10**: 520–527.
- Domergue C, Serneels V, Cauuet B, Pailler J, Orzechowsk S. 2006.** Mines et métallurgies en Gaule à la fin de l'âge du Fer et à l'époque romaine., *Glux-en-Glenne*: 131–162.

- Dong C, Dalton-Morgan J, Vincent K, Sharp P. 2009a.** A Modified TILLING Method for Wheat Breeding. *The Plant Genome Journal* **2**: 39.
- Dong C, Vincent K, Sharp P. 2009b.** Simultaneous mutation detection of three homoeologous genes in wheat by High Resolution Melting analysis and Mutation Surveyor. *BMC plant biology* **9**: 143.
- Dubois S, Cheptou PO, Petit C, Meerts P, Poncelet M, Vekemans X, Lefèbvre C, Escarré J. 2003.** Genetic structure and mating systems of metallicolous and nonmetallicolous populations of *Thlaspi caerulescens*. *New Phytologist* **157**: 633–641.
- Duncan S, Holm S, Questa J, Irwin J, Grant A, Dean C. 2015.** Seasonal shift in timing of vernalization as an adaptation to extreme winter. *eLife* **4**.
- Dunn CE. 2007.** New Perspectives on Biogeochemical Exploration. *Proceeding of Exploration 07: Fifth Decennial International Conference on Mineral Exploration*: 249–261.
- van der Ent A, Baker AJM, Reeves RD, Pollard AJ, Schat H. 2013.** Hyperaccumulators of metal and metalloid trace elements: Facts and fiction. *Plant and Soil* **362**: 319–334.
- 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**: 429–437.
- Feder ME, Mitchell-Olds T. 2003.** Evolutionary and ecological functional genomics. *Nature Reviews Genetics* **4**: 651–657.
- Franks SJ, Sim S, Weis AE. 2007.** Rapid evolution of flowering time by an annual plant in response to a climate fluctuation. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 1278–1282.
- 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: 643–648.

**Gady AL, Hermans FW, Van de Wal MH, van Loo EN, Visser RG, Bachem CWB. 2009.** Implementation of two high through-put techniques in a novel application: detecting point mutations in large EMS mutated plant populations. *Plant methods* **5**: 13.

**Gady AL, Vriezen WH, Van de Wal MHB, Huang PP, Bovy AG, Visser RG, Bachem CWB. 2012.** Induced point mutations in the phytoene synthase 1 gene cause differences in carotenoid content during tomato fruit ripening. *Molecular Breeding* **29**: 801–812.

**Gilchrist EJ, Haughn GW. 2005.** TILLING without a plough: A new method with applications for reverse genetics. *Current Opinion in Plant Biology* **8**: 211–215.

**Godt J, Scheidig F, Grosse-Siestrup C, Esche V, Brandenburg P, Reich A, Groneberg DA. 2006.** The toxicity of cadmium and resulting hazards for human health. *Journal of occupational medicine and toxicology (London, England)* **1**: 22.

**Gottwald S, Bauer P, Komatsuda T, Lundqvist U, Stein N. 2009.** TILLING in the two-rowed barley cultivar ‘Barke’ reveals preferred sites of functional diversity in the gene *HvHox1*. *BMC research notes* **2**: 258.

**Greene EA, Codomo CA, Taylor NE, Henikoff JG, Till BJ, Reynolds SH, Enns LC, Burtner C, Johnson JE, Odden AR. 2003.** Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* **164**: 731–740.

**Guan ZQ, Chai TY, Zhang YX, 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**: 77–86.

**Guimarães MA, Loureiro ME, Salt DE. 2013.** Inducing flowering in *Noccaea caerulescens* (J. & C. Presl) F. K. Mey (Brassicaceae), a species having high heavy-metal accumulation. *Revista Ciência Agronômica* **44**: 834–841.

**Gusiatin ZM, Klimiuk E. 2012.** Metal (Cu, Cd and Zn) removal and stabilization during multiple soil washing by saponin. *Chemosphere* **86**: 383–391.



- Halimaa P, Blande D, Aarts MGM, Tuomainen M, Tervahauta A, Kärenlampi S. 2014a.** Comparative transcriptome analysis of the metal hyperaccumulator *Noccaea caerulescens*. *Frontiers in plant science* **5**: 213.
- Halimaa P, Lin YF, Ahonen VH, Blande D, Clemens S, Gyenesei A, Häikiö E, Kärenlampi SO, Laiho A, Aarts MGM, et al. 2014b.** Gene expression differences between *Noccaea caerulescens* ecotypes help to identify candidate genes for metal phytoremediation. *Environmental Science and Technology* **48**: 3344–3353.
- Hall JL. 2002.** Cellular mechanisms for heavy metal detoxification and tolerance. *Journal of experimental botany* **53**: 1–11.
- Hammond JP, Bowen HC, White PJ, Mills V, Pyke K a, Baker AJM, Whiting SN, May ST, Broadley MR. 2006.** A comparison of the *Thlaspi caerulescens* and *Thlaspi arvense* shoot transcriptomes. *New phytologist* **170**: 239–60.
- Hanikenne M, Nouet C. 2011.** Metal hyperaccumulation and hypertolerance: A model for plant evolutionary genomics. *Current Opinion in Plant Biology* **14**: 252–259.
- Hanikenne M, Talke IN, Haydon MJ, 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**: 391–395.
- Hartmann U, Höhmann S, Nettesheim K, Wisman E, Saedler H, Huijser P. 2000.** Molecular cloning of *SVP*: A negative regulator of the floral transition in Arabidopsis. *Plant Journal* **21**: 351–360.
- Haydon MJ, Román Á, Arshad W. 2015.** Nutrient homeostasis within the plant circadian network. *Frontiers in plant science* **6**: 299.
- Helliwell CA, Wood CC, Robertson M, James Peacock W, Dennis ES. 2006.** The Arabidopsis *FLC* protein interacts directly in vivo with *SOC1* and *FT* chromatin and is part of a high-molecular-weight protein complex. *Plant Journal* **46**: 183–192.
- Henikoff S, Comai L. 2003.** Single-nucleotide mutations for plant functional genomics. *Annual Review of Plant Biology* **54**: 375–401.

- Herbette S, Tacconat L, Hugouvieux V, Piette L, Magniette MLM, Cuine S, Auroy P, Richaud P, Forestier C, Bourguignon J, et al. 2006.** Genome-wide transcriptome profiling of the early cadmium response of *Arabidopsis* roots and shoots. *Biochimie* **88**: 1751–1765.
- Himelblau E, Gilchrist EJ, Buono K, Bizzell C, Mentzer L, Vogelzang R, Osborn T, Amasino RM, Parkin I a P, Haughn GW. 2009.** Forward and reverse genetics of rapid-cycling *Brassica oleracea*. *TAG. Theoretical and applied genetics*. **118**: 953–61.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson E a, Cresko W a. 2010.** Population genomics of parallel adaptation in *threespine stickleback* using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.
- Hyun Y, Kim J, Cho SW, Choi Y, Kim JS, Coupland G. 2014.** Site-directed mutagenesis in *Arabidopsis thaliana* using dividing tissue-targeted RGEN of the CRISPR/Cas system to generate heritable null alleles. *Planta* **241**: 271–284.
- Iida S, Terada R. 2004.** A tale of two integrations, transgene and T-DNA: Gene targeting by homologous recombination in rice. *Current Opinion in Biotechnology* **15**: 132–138.
- Lochlainn ÓS, Bowen HC, Fray RG, Hammond JP, King GJ, White PJ, Graham NS, Broadley MR. 2011.** Tandem quadruplication of *HMA4* in the zinc (Zn) and cadmium (Cd) hyperaccumulator *Noccaea caerulescens*. *PLoS ONE* **6**.
- Iqbal M, Nawaz I, Hassan Z, Hakvoort HWJ, Blik M, Aarts MGM, Schat H. 2013.** Expression of *HMA4* cDNAs of the zinc hyperaccumulator *Noccaea caerulescens* from endogenous *NcHMA4* promoters does not complement the zinc-deficiency phenotype of the *Arabidopsis thaliana* *hma2hma4* double mutant. *Frontiers in plant science* **4**: 404.
- Ishikawa T, Kamei Y, Otozai S, Kim J, Sato A, Kuwahara Y, Tanaka M, Deguchi T, Inohara H, Tsujimura T, et al. 2010.** High-resolution melting curve analysis for rapid detection of mutations in a Medaka TILLING library. *BMC molecular biology* **11**: 70.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al. 2007.** The grapevine genome sequence suggests

ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463–7.

**Jain R, Srivastava S, Solomon S, Shrivastava AK, Chandra A. 2010.** Impact of excess zinc on growth parameters, cell division, nutrient accumulation, photosynthetic pigments and oxidative stress of sugarcane (*saccharum* spp.). *Acta Physiologiae Plantarum* **32**: 979–986.

**James GV, Patel V, Nordström KJ V, Klasen JR, Salomé P a, Weigel D, Schneeberger K. 2013.** User guide for mapping-by-sequencing in Arabidopsis. *Genome biology* **14**: R61.

**Jelesko JG, Harper R, Furuya M, Gruissem W. 1999.** Rare germinal unequal crossing-over leading to recombinant gene formation and gene duplication in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 10302–7.

**Jhorar RK, Van Dam JC, Bastiaanssen WGM, Feddes RA. 2004.** Calibration of effective soil hydraulic parameters of heterogeneous soil profiles. *Journal of Hydrology* **285**: 233–247.

**Jiang D, Wang Y, Wang Y, He Y. 2008.** Repression of *FLOWERING LOCUS C* and *FLOWERING LOCUS T* by the Arabidopsis Polycomb Repressive Complex 2 Components. *PLoS ONE* **3**.

**Jiménez-Ambriz G1, 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**: 199–215.

**Kabata-Pendias A. 2011.** Trace elements in soils and plants. *CRC Press*: 1–534.

**Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin XY, et al. 2000.** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

**Kim D-H, Zografos BR, Sung S. 2010.** Mechanisms underlying vernalization-mediated *VIN3* induction in Arabidopsis. *Plant signaling & behavior* **5**: 1457–1459.

- Ko T-S, Korban SS, Somers D a. 2006.** Soybean (*Glycine max*) transformation using immature cotyledon explants. *Methods in molecular biology (Clifton, N.J.)* **343**: 397–405.
- Kobayashi MJ, Shimizu KK. 2013.** Challenges in studies on flowering time: Interfaces between phenological research and the molecular network of flowering genes. *Ecological Research* **28**: 161–172.
- 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. *Biochemical Systematics and Ecology* **26**: 823–838.
- Kofler R, Orozco-terWengel P, De Maio N, Pandey R V, Nolte V, Futschik A, Kosiol C, Schlotterer C. 2011a.** PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE* **6**: 9.
- Kofler R, Orozco-terWengel P, de Maio N, Pandey RV, Nolte V, Futschik A, Kosiol C, Schlötterer C. 2011b.** Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* **6**.
- Kofler R, Pandey RV, Schlötterer C. 2011c.** PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**: 3435–3436.
- Kofler R, Schlötterer C. 2014.** A guide for the design of evolve and resequencing studies. *Molecular Biology and Evolution* **31**: 474–483.
- Komeda Y. 2004.** Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annual review of plant biology* **55**: 521–535.
- Koornneef M. 2002.** Classical mutagenesis in higher plants. *Molecular plant biology*: 1–11.
- Koornneef M, Dellaert LWM, van der Veen JH. 1982.** EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L.) Heynh. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **93**: 109–123.

- Korswagen HC, Durbin RM, Smits MT, Plasterk RH. 1996.** Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 14680–5.
- Kotoda N, Wada M. 2005.** *MdTFL1*, a *TFL1*-like gene of apple, retards the transition from the vegetative to reproductive phase in transgenic Arabidopsis. *Plant Science* **168**: 95–104.
- Kovalchuk I, Titov V, Hohn B, Kovalchuk O. 2005.** Transcriptome profiling reveals similarities and differences in plant responses to cadmium and lead. *Mutation research* **570**: 149–161.
- Krämer U. 2010.** Metal hyperaccumulation in plants. *Annual review of plant biology* **61**: 517–534.
- Kramer EM, Jaramillo MA, Di Stilio VS. 2004.** Patterns of Gene Duplication and Functional Evolution during the Diversification of the *AGAMOUS* Subfamily of MADS Box Genes in Angiosperms. *Genetics* **166**: 1011–1023.
- Krysan P. 2004.** Ice-Cap. A High-Throughput Method for Capturing Plant Tissue Samples for Genotype Analysis. *Plant Physiology* **135**: 1162–1169.
- Kurowska M, Daszkowska-Golec A, Gruszka D, Marzec M, Szurman M, Szarejko I, Maluszynski M. 2011.** TILLING - a shortcut in functional genomics. *Journal of Applied Genetics* **52**: 371–390.
- Lahner B, Gong J, Mahmoudian M, Smith EL, Abid KB, Rogers EE, Guerinot ML, Harper JF, Ward JM, McIntyre L, et al. 2003.** Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. *Nature Biotechnology* **21**: 1215–1221.
- Lai K-S, Kaothien-Nakayama P, Iwano M, Takayama S. 2012.** A TILLING resource for functional genomics in *Arabidopsis thaliana* accession C24. *Genes & genetic systems* **87**: 291–7.
- Lasat MM, Baker AJM, Kochian L V. 1996.** Physiological Characterization of Root Zn<sup>2+</sup> Absorption and Translocation to Shoots in Zn Hyperaccumulator and

Nonaccumulator Species of Thlaspi. *Plant physiology* **112**: 1715–1722.

**Lee J, Oh M, Park H, Lee I. 2008.** *SOC1* translocated to the nucleus by interaction with *AGL24* directly regulates *LEAFY*. *Plant Journal* **55**: 832–843.

Jeong HL, Seong JY, Soo HP, Hwang I, Jong SL, Ji HA. 2007. Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. *Genes and Development* **21**: 397–402.

**Li P, Filiault D, Box MS, Kerdaffrec E, van Oosterhout C, Wilczek AM, Schmitt J, McMullan M, Bergelson J, Nordborg M. 2014.** Multiple *FLC* haplotypes defined by independent cisregulatory variation underpin life history diversity in *Arabidopsis thaliana*. *Genes and Development* **28**: 1635–1640.

**Li X, Song Y, Century K, Straight S, Ronald P, Dong X, Lassner M, Zhang Y. 2001.** A fast neutron deletion mutagenesis-based reverse genetics system for plants. *Plant Journal* **27**: 235–242.

**Li W, Wu J, Weng S, Zhang D, Zhang Y, Shi C. 2010.** Characterization and fine mapping of the glabrous leaf and hull mutants (*gl1*) in rice (*Oryza sativa* L.). *Plant Cell Reports* **29**: 617–627.

**Limpens E, Ramos J, Franken C, Raz V, Compaan B, Franssen H, Bisseling T, Geurts R. 2004.** RNA interference in *Agrobacterium rhizogenes*-transformed roots of *Arabidopsis* and *Medicago truncatula*. *Journal of Experimental Botany* **55**: 983–992.

**Lin Y-F, Aarts MGM. 2012.** The molecular mechanism of zinc and cadmium stress response in plants. *Cellular and Molecular Life Sciences* **69**: 3187–3206.

**Lin Y-F, Hassan Z, Talukdar S, Schat H, Aarts MGM. 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**: e0149750.

**Lin Y-F, Severing EI, Te Lintel Hekkert B, Schijlen E, Aarts MGM. 2014.** A comprehensive set of transcript sequences of the heavy metal hyperaccumulator *Noccaea caerulescens*. *Frontiers in plant science* **5**: 261.

- Ling H-Q, Bauer P, Bereczky Z, Keller B, Ganai M. 2002.** The tomato fer gene encoding a bHLH protein controls iron-uptake responses in roots. *Proceedings of the National Academy of Sciences of the United States of America* **99**: 13938–13943.
- Lochlainn SÓ, Amoah S, Graham NS, Alamer K, Rios JJ, Kurup S, Stoute A, Hammond JP, Østergaard L, King GJ, et al. 2011.** High Resolution Melt (HRM) analysis is an efficient tool to genotype EMS mutants in complex crop genomes. *Plant Methods* **7**: 43.
- Lohmueller KE, Bustamante CD, Clark AG. 2011.** Detecting directional selection in the presence of recent admixture in African-Americans. *Genetics* **187**: 823–835.
- Lombi E, Zhao FJ, McGrath SP, Young SD, Sacchi GA. 2001.** Physiological evidence for a high-affinity cadmium transporter highly expressed in a *Thlaspi caerulescens* ecotype. *New Phytologist* **149**: 53–60.
- Ma LQ, Komar KM, Tu C, Zhang W, Cai Y, Kennelley ED. 2001.** A fern that hyperaccumulates arsenic. *Nature* **409**: 579.
- Mahrez W, Shin J, Muñoz-Viana R, Figueiredo DD, Trejo-Arellano MS, Exner V, Siretskiy A, Gruissem W, Köhler C, Hennig L. 2016.** *BRR2a* Affects Flowering Time via *FLC* Splicing. *PLOS Genetics* **12**: e1005924.
- Marschner H. 1995.** *Mineral nutrition of higher plants*. Book  
ISBN: 978-0-12-473542-2
- Martín B, Ramiro M, Martínez-Zapater JM, Alonso-Blanco C. 2009.** A high-density collection of EMS-induced mutations for TILLING in *Landsberg erecta* genetic background of Arabidopsis. *BMC plant biology* **9**: 147.
- Martos S, Gallego B, Sáez L, López-Alvarado J, Cabot C, Poschenrieder C. 2016.** Characterization of Zinc and Cadmium Hyperaccumulation in Three *Noccaea* (Brassicaceae) Populations from Non-metalliferous Sites in the Eastern Pyrenees. *Frontiers in Plant Science* **7**: 1–13.
- Mateos JL, Madrigal P, Tsuda K, Rawat V, Richter R, Romera-Branchat M, Fornara**

- F, Schneeberger K, Krajewski P, Coupland G. 2015.** Combinatorial activities of *SHORT VEGETATIVE PHASE* and *FLOWERING LOCUS C* define distinct modes of flowering regulation in Arabidopsis. *Genome Biology* **16**: 1–23.
- Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, et al. 2005.** The map-based sequence of the rice genome. *Nature* **436**: 793–800.
- McCallum CM, Comai L, Greene E a, Henikoff S. 2000a.** Targeting induced local lesions IN genomes (TILLING) for plant functional genomics. *Plant physiology* **123**: 439–442.
- McCallum CM, Comai L, Greene E a, Henikoff S. 2000b.** Targeted screening for induced mutations. *Nature biotechnology* **18**: 455–457.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010.** The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* **20**: 1297–303.
- Meerts P, Van Isacker N. 1997.** Heavy metal tolerance and accumulation in metalicolous and non-metallicolous populations of *Thlaspi caerulescens* from continental Europe. *Plant Ecology* **133**: 221–231.
- Michaels SD, Amasino RM. 1999.** *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *The Plant cell* **11**: 949–956.
- Minoia S, Petrozza A, D’Onofrio O, Piron F, Mosca G, Sozio G, Cellini F, Bendahmane A, Carriero F. 2010.** A new mutant genetic resource for tomato crop improvement by TILLING technology. *BMC research notes* **3**: 69.
- Mitchell-Olds T, Feder M, Wray G. 2008.** Evolutionary and ecological functional genomics. *Heredity* **100**: 101–2.
- Mohamed R, Wang CT, Ma C, Shevchenko O, Dye SJ, Puzey JR, Etherington E, Sheng X, Meilan R, Strauss SH, et al. 2010.** *Populus CEN/TFL1* regulates first onset of flowering, axillary meristem identity and dormancy release in *Populus*. *Plant*



*Journal* **62**: 674–688.

**Mohtadi A, Ghaderian SM, Schat H. 2012.** A comparison of lead accumulation and tolerance among heavy metal hyperaccumulating and non-hyperaccumulating metallophytes. *Plant and Soil* **352**: 267–276.

**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*: mcv191.

**Nah G, Jeffrey Chen Z. 2010.** Tandem duplication of the *FLC* locus and the origin of a new gene in Arabidopsis related species and their functional implications in allopolyploids. *New phytologist* **186**: 228–238.

**Narayanan V, Mieczkowski PA, Kim HM, Petes TD, Lobachev KS. 2006.** The Pattern of Gene Amplification Is Determined by the Chromosomal Location of Hairpin-Capped Breaks. *Cell* **125**: 1283–1296.

**Nascimento CWA do, Xing B. 2006.** Phytoextraction: a review on enhanced metal availability and plant accumulation. *Scientia Agricola* **63**: 299–311.

**Nei M. 1972.** Genetic distance between populations. *The American Naturalist* **106**: 283–292.

**Niu CF, Wei W, Zhou QY, Tian AG, Hao YJ, Zhang WK, Ma B, Lin Q, Zhang ZB, Zhang JS, et al. 2012.** Wheat *WRKY* genes *TaWRKY2* and *TaWRKY19* regulate abiotic stress tolerance in transgenic Arabidopsis plants. *Plant, Cell and Environment* **35**: 1156–1170.

**Nordborg M, Bergelson JOY. 1999.** The effect of seed and rosette cold treatment on germination and flowering time in some *Arabidopsis thaliana* (Brassicaceae) ecotypes. *American Journal of Botany* **86**: 470–475.

**Nouet C, Charlier JB, Carnol M, Bosman B, Farnir F, Motte P, Hanikenne M. 2015.** Functional analysis of the three *HMA4* copies of the metal hyperaccumulator *Arabidopsis halleri*. *Journal of Experimental Botany* **66**: 5783–5795.

- Ó Lochlainn S, Fray RG, Hammond JP, King GJ, White PJ, Young SD, Broadley MR. 2011.** Generation of nonvernal-obligate, faster-cycling *Noccaea caerulescens* lines through fast neutron mutagenesis. *New Phytologist* **189**: 409–414.
- Oliver SN, Finnegan EJ, Dennis ES, Peacock WJ, Trevaskis B. 2009.** Vernalization-induced flowering in cereals is associated with changes in histone methylation at the *VERNALIZATION1* gene. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 8386–8391.
- Oñate-Sánchez L, Vicente-Carbajosa J. 2008.** DNA-free RNA isolation protocols for *Arabidopsis thaliana*, including seeds and siliques. *BMC research notes* **1**: 93.
- Panda P, Nath S, Chanu TT, Sharma GD, Panda SK. 2011.** Cadmium stress-induced oxidative stress and role of nitric oxide in rice (*Oryza sativa* L.). *Acta Physiologiae Plantarum* **33**: 1737–1747.
- 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 1. *Plant physiology* **136**: 3814–3823.
- Pauwels M, Saumitou-Laprade P, Holl AC, 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**: 4403–4414.
- 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**: 916–928.
- Peer WA, Mahmoudian M, Freeman JL, Lahner B, Richards EL, Reeves RD, Murphy AS, Salt DE. 2006.** Assessment of plants from the Brassicaceae family as genetic models for the study of nickel and zinc hyperaccumulation. *New Phytologist* **172**: 248–260.
- Peer WA, Mamoudian M, Lahner B, Reeves RD, Murphy AS, Salt DE. 2003.**

Identifying model metal hyperaccumulating plants: Germplasm analysis of 20 brassicaceae accessions from a wide geographical area. *New Phytologist* **159**: 421–430.

**Pence NS, Larsen PB, Ebbs SD, Letham DL, Lasat MM, Garvin DF, Eide D, Kochian L V. 2000.** The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 4956–4960.

**Peng Y, Zhang Y, Lv J, Zhang J, Li P, Shi X, Wang Y, Zhang H, He Z, Teng S. 2012.** Characterization and Fine Mapping of a Novel Rice Albino Mutant low temperature albino 1. *Journal of Genetics and Genomics* **39**: 385–396.

**Pespeni MH, Garfield DA, Manier MK, Palumbi SR. 2012.** Genome-wide polymorphisms show unexpected targets of natural selection. *Proceedings. Biological sciences / The Royal Society* **279**: 1412–20.

**Plessl M, Rigola D, Hassinen VH, Tervahauta A, Kärenlampi S, Schat H, Aarts MGM, Ernst D. 2010.** Comparison of two ecotypes of the metal hyperaccumulator *Thlaspi caerulescens* (J. & C. PRESL) at the transcriptional level. *Protoplasma* **239**: 81–93.

**Pollard AJ, Baker AJM. 1997.** Deterrence of herbivory by zinc hyperaccumulation in *Thlaspi caerulescens* (Brassicaceae). *New Phytologist* **135**: 655–658.

**Pollard AJ, Powell KD, Harper FA, Smith JAC. 2002.** The Genetic Basis of Metal Hyperaccumulation in Plants. *Critical Reviews in Plant Sciences* **21**: 539–566.

**Pollard AJ, Reeves RD, Baker AJM. 2014.** Facultative hyperaccumulation of heavy metals and metalloids. *Plant Science* **217-218**: 8–17.

**Porch TG, Blair MW, Lariguet P, Galeano C, Pankhurst CE, Broughton WJ. 2009.** Generation of a Mutant Population for TILLING Common Bean Genotype BAT 93. *Journal of the American Society for Horticultural Science* **134**: 348–355.

**Posé D, Verhage L, Ott F, Yant L, Mathieu J, Angenent GC, Immink RGH, Schmid M. 2013.** Temperature-dependent regulation of flowering by antagonistic *FLM* variants. *Nature* **503**: 414–7.

- Ramesh SA, Choimes S, Schachtman DP. 2004.** Over-expression of an Arabidopsis zinc transporter in *Hordeum vulgare* increases short-term zinc uptake after zinc deprivation and seed zinc content. *Plant Molecular Biology* **54**: 373–385.
- 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**: 169–181.
- Ratcliffe OJ, Kumimoto RW, Wong BJ, Riechmann JL. 2003.** Analysis of the Arabidopsis *MADS AFFECTING FLOWERING* gene family: *MAF2* prevents vernalization by short periods of cold. *The Plant cell* **15**: 1159–1169.
- Ratcliffe OJ, Nadzan GC, Reuber TL, Riechmann JL. 2001.** Regulation of flowering in Arabidopsis by an *FLC* homologue. *Plant physiology* **126**: 122–132.
- Razi H, Howell EC, Newbury HJ, Kearsey MJ. 2008.** Does sequence polymorphism of *FLC* paralogues underlie flowering time QTL in *Brassica oleracea*? *Theoretical and Applied Genetics* **116**: 179–192.
- Reddy TV, Dwivedi S, Sharma NK. 2012.** Development of TILLING by sequencing platform towards enhanced leaf yield in tobacco. *Industrial Crops and Products* **40**: 324–335.
- Reeves RD, Schwartz C, Morel JL, Edmondson J. 2001.** Distribution and Metal-Accumulating Behavior of *Thlaspi caerulescens* and Associated Metallophytes in France. *International Journal of Phytoremediation* **3**: 145–172.
- International Rice Genome Sequencing Project. 2005.** The map-based sequence of the rice genome. *Nature* **436**: 793–800.
- Rigola D, Fiers M, Vurro E, Aarts MGM. 2006.** The heavy metal hyperaccumulator *Thlaspi caerulescens* expresses many species-specific genes, as identified by comparative expressed sequence tag analysis. *New Phytologist* **170**: 753–766.
- Riley R. 1956.** The Influence of the Breeding System on the Genecology of *Thlaspi alpestre* L. *New Phytologist* **55**: 319–330.
- Robinson BH, Leblanc M, Petit D, Brooks RR, Kirkman JH, Gregg PEH. 1998.** The

potential of *Thlaspi caerulescens* for phytoremediation of contaminated soils. *Plant and Soil* **203**: 47–56.

**Rong YS, Golic KG. 2000.** Gene targeting by homologous recombination in *Drosophila*. *Science* **288**: 2013–2018.

**Rosloski SM, Singh A, Jali SS, Balasubramanian S, Weigel D, Grbic V. 2013.** Functional analysis of splice variant expression of *MADS AFFECTING FLOWERING 2* of *Arabidopsis thaliana*. *Plant Molecular Biology* **81**: 57–69.

**Ross-Ibarra J, Wright SI, Foxe JP, Kawabe A, DeRose-Wilson L, Gos G, Charlesworth D, Gaut BS. 2008.** Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE* **3**.

**Salt DE, Baxter I, Lahner B. 2008.** Ionomics and the study of the plant ionome. *Annual review of plant biology* **59**: 709–733.

**Schat H, Vooijs R, Kuiper E. 1996.** Identical major gene loci for heavy metal tolerances that have independently evolved in different local populations and subspecies of *Silene vulgaris*. *Evolution* **50**: 1888–1895.

**Schlötterer C, Tobler R, Kofler R, Nolte V. 2014.** Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* **15**: 749–763.

**Schmitz RJ, Amasino RM. 2007.** Vernalization: A model for investigating epigenetics and eukaryotic gene regulation in plants. *Biochimica et Biophysica Acta - Gene Structure and Expression* **1769**: 269–275.

**Schneeberger K. 2014.** Using next-generation sequencing to isolate mutant genes from forward genetic screens. *Nature reviews. Genetics* **15**: 662–676.

**Schubert D, Primavesi L, Bishopp A, Roberts G, Doonan J, Jenuwein T, Goodrich J. 2006.** Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *The EMBO journal* **25**: 4638–4649.

**Sheldon CC. 2002.** Different Regulatory Regions Are Required for the Vernalization-

Induced Repression of *FLOWERING LOCUS C* and for the Epigenetic Maintenance of Repression. *THE PLANT CELL ONLINE* **14**: 2527–2537.

**Sheldon CC, Burn JE, Perez PP, Metzger J, Edwards JA, Peacock WJ, Dennis ES. 1999.** The *FLF* MADS box gene: a repressor of flowering in Arabidopsis regulated by vernalization and methylation. *The Plant cell* **11**: 445–458.

**Shindo C, Lister C, Crevillen P, Nordborg M, Dean C. 2006.** Variation in the epigenetic silencing of *FLC* contributes to natural variation in Arabidopsis vernalization response. *Genes and Development* **20**: 3079–3083.

**Le Signor C, Savoie V, Aubert G, Verdier J, Nicolas M, Pagny G, Moussy F, Sanchez M, Baker D, Clarke J, et al. 2009.** Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnology Journal* **7**: 430–441.

**Šimek R, Novoselovic D. 2012.** The Use of Reverse genetics Approach in Plant Genomics. *Poljoprivreda* **18**:14-18

**Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D. 2005.** A reverse genetics, nontransgenic approach to wheat crop improvement by TILLING. *Nature biotechnology* **23**: 75–81.

**Sokal RR, Rohlf FJ. 1995.** *Biometry: The Principles and Practices of Statistics in Biological Research*. Book ISBN: 0-7167-2411-1

**Solti A, Gáspár L, Mészáros I, Szigeti Z, Lévai L, Sárvári E. 2008.** Impact of iron supply on the kinetics of recovery of photosynthesis in Cd-stressed poplar (*Populus glauca*). *Annals of Botany* **102**: 771–782.

**Soltis DE, Soltis PS. 2003.** The role of phylogenetics in comparative genetics. *Plant physiology* **132**: 1790–1800.

**Somers D a, Samac D a, Olhoft PM. 2003.** Recent advances in legume transformation. *Plant physiology* **131**: 892–899.

**Song YH, Ito S, Imaizumi T. 2013.** Flowering time regulation: Photoperiod- and temperature-sensing in leaves. *Trends in Plant Science* **18**: 575–583.

- Sridhar BBM, Diehl SV, Han FX, Monts DL, Su Y. 2005.** Anatomical changes due to uptake and accumulation of Zn and Cd in Indian mustard (*Brassica juncea*). *Environmental and Experimental Botany* **54**: 131–141.
- Stemple DL. 2004.** TILLING--a high-throughput harvest for functional genomics. *Nature reviews. Genetics* **5**: 145–150.
- Stephenson P, Baker D, Girin T, Perez A, Amoah S, King GJ, Østergaard L. 2010.** A rich TILLING resource for studying gene function in *Brassica rapa*. *BMC plant biology* **10**: 62.
- Stern DL, Orgogozo V. 2008.** The loci of evolution: How predictable is genetic evolution? *Evolution* **62**: 2155–2177.
- Sun Y, Zhou Q, Xie X, Liu R. 2010.** Spatial, sources and risk assessment of heavy metal contamination of urban soils in typical regions of Shenyang, China. *Journal of Hazardous Materials* **174**: 455–462.
- Sung S, Amasino RM. 2004.** Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3. *Nature* **427**: 159–164.
- Tajima F. 1989.** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Talame V, Bovina R, Sanguineti MC, Tuberosa R, Lundqvist U, Salvi S. 2008.** TILLMore, a resource for the discovery of chemically induced mutants in barley. *Plant Biotechnology Journal* **6**: 477–485.
- Talukdar S, Aarts MGM. 2008.** *Arabidopsis thaliana* and *Thlaspi caerulescens* respond comparably to low zinc supply. *Plant and Soil*.85–94.
- Till BJ, Cooper J, Tai TH, Colowit P, Greene E a, Henikoff S, Comai L. 2007.** Discovery of chemically induced mutations in rice by TILLING. *BMC plant biology* **7**: 19.
- Till BJ, Reynolds SH, Weil C, Springer N, Burtner C, Young K, Bowers E, Codomo C a, Enns LC, Odden AR 2004.** Discovery of induced point mutations in maize genes by TILLING. *BMC plant biology* **4**: 12.

- Tobler R, Franssen SU, Kofler R, Orozco-Terwengel P, Nolte V, Hermisson J, Schlötterer C. 2014.** Massive habitat-specific genomic response in *D. melanogaster* populations during experimental evolution in hot and cold environments. *Molecular Biology and Evolution* **31**: 364–375.
- Tomancak P, Beaton A, Weiszmänn R, Kwan E, Shu S, Lewis SE, Richards S, Ashburner M, Hartenstein V, Celniker SE. 2002.** Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome biology* **3**: RESEARCH0088.
- Tsai H, Howell T, Nitcher R, Missirian V, Watson B, Ngo KJ, Lieberman M, Fass J, Uauy C, Tran RK, et al. 2011.** Discovery of rare mutations in populations: TILLING by sequencing. *Plant physiology* **156**: 1257–1268.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin S V. 2010.** Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature genetics* **42**: 260–3.
- Vaucheret H, Béclin C, Fagard M. 2001.** Post-transcriptional gene silencing in plants. *Journal of Cell Science* **114**: 3083.
- Verbruggen N, Hermans C, Schat H. 2009.** Molecular mechanisms of metal hyperaccumulation in plants. *The New phytologist* **181**: 759–76.
- Verbruggen N, Juraniec M, Meyer CBC. 2013.** Tolerance to cadmium in plants : the special case of hyperaccumulators. : 633–638.
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Guerinot M Lou, Briat J-F, Curie C. 2002.** *IRT1*, an *Arabidopsis* transporter essential for iron uptake from the soil and for plant growth. *The Plant cell* **14**: 1223–33.
- van de Mortel JE, Almar Villanueva L, Schat H, Kwekkeboom J, Coughlan S, Moerland PD, Ver Loren van Themaat E, Koornneef M, Aarts MGM. 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**: 1127–47



- van de Mortel JE, Schat H, Moerland PD, Ver Loren van Themaat, van der Ent S, Blankestijn H, Ghandilyan A, Tsiatsiani S, Aarts MGM. 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**: 301–324.
- Vossen RHAM, Aten E, Roos A, Den Dunnen JT. 2009.** High-resolution melting analysis (HRMA) - More than just sequence variant screening. *Human Mutation* **30**: 860–866.
- Wang R, Albani MC, Vincent C, Bergonzi S, Luan M, Bai Y, Kiefer C, Castillo R, Coupland G. 2011.** *AaTFL1* confers an age-dependent response to vernalization in perennial *Arabis alpina*. *The Plant cell* **23**: 1307–1321.
- Wang R, Farrona S, Vincent C, Joecker A, Schoof H, Turck F, Alonso-Blanco C, Coupland G, Albani MC. 2009.** *PEP1* regulates perennial flowering in *Arabis alpina*. *Nature* **459**: 423–7.
- Weber M, Harada E, Vess C, Roepenack-Lahaye E V, Clemens S. 2004.** Comparative microarray analysis of *Arabidopsis thaliana* and *Arabidopsis halleri* roots identifies nicotianamine synthase, a ZIP transporter and other genes as potential metal hyperaccumulation factors. *The Plant journal : for cell and molecular biology* **37**: 269–281.
- Weber M, Trampczynska A, Clemens S. 2006.** Comparative transcriptome analysis of toxic metal responses in *Arabidopsis thaliana* and the Cd<sup>2+</sup>-hypertolerant facultative metallophyte *Arabidopsis halleri*. *Plant, Cell and Environment* **29**: 950–963.
- Wei Yang TJ, Perry PJ, Ciani S, Pandian S, Schmidt W. 2008.** Manganese deficiency alters the patterning and development of root hairs in *Arabidopsis*. *Journal of Experimental Botany* **59**: 3453–3464.
- Werner JD, Borevitz JO, Henriette Uhlenhaut N, Ecker JR, Chory J, Weigel D. 2005.** *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. *Genetics* **170**: 1197–1207.

- Wintz H, Fox T, Wu YY, Feng V, Chen W, Chang HS, Zhu T, Vulpe C. 2003.** Expression Profiles of *Arabidopsis thaliana* in Mineral Deficiencies Reveal Novel Transporters Involved in Metal Homeostasis. *Journal of Biological Chemistry* **278**: 47644–47653.
- Wisman E, Hartmann U, Sagasser M, Baumann E, Palme K, Hahlbrock K, Saedler H, Weisshaar B. 1998.** Knock-out mutants from an *En-1* mutagenized *Arabidopsis thaliana* population generate phenylpropanoid biosynthesis phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 12432–12437.
- Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, Kim S-G, Kim S-T, Choe S, Kim J-S. 2015.** DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology* **33**: 1162–1164.
- Wray G a. 2007.** The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics* **8**: 206–16.
- Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman M V., Romano LA. 2003.** The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution* **20**: 1377–1419.
- Wu L, Bradshaw AD, Thurman DA. 1975.** The potential for evolution of heavy metal tolerance in plants. *Heredity* **34**: 165–187.
- Wu JL, Wu C, Lei C, Baraoidan M, Bordeos A, Madamba MRS, Ramos-Pamplona M, Mauleon R, Portugal A, Ulat VJ, et al. 2005.** Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Molecular Biology* **59**: 85–97.
- Wuana RA, Okieimen FE. 2011.** Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. *ISRN Ecology* **2011**: 1–20.
- Xie K, Yang Y. 2013.** RNA-Guided genome editing in plants using a CRISPR-Cas system. *Molecular Plant* **6**: 1975–1983.
- Xin Z, Wang ML, Barkley N a, Burow G, Franks C, Pederson G, Burke J. 2008.**

Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC plant biology* **8**: 103.

**Xing JP, Jiang RF, Ueno D, Ma JF, Schat H, McGrath SP, Zhao FJ. 2008.** Variation in root-to-shoot translocation of cadmium and zinc among different accessions of the hyperaccumulators *Thlaspi caerulescens* and *Thlaspi praecox*. *New Phytologist* **178**: 315–325.

**Yandeau-Nelson MD, Xia Y, Li J, Neuffer MG, Schnable PS. 2006.** Unequal sister chromatid and homolog recombination at a tandem duplication of the *a1* locus in maize. *Genetics* **173**: 2211–2226.

**Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann K a, Meyerowitz EM. 1990.** The protein encoded by the Arabidopsis homeotic gene *agamous* resembles transcription factors. *Nature* **346**: 35–39.

**Yu S, Ligang C, Liping Z, Diqiu Y. 2010.** Overexpression of *OsWRKY72* gene interferes in the abscisic acid signal and auxin transport pathway of Arabidopsis. *Journal of Biosciences* **35**: 459–471.

**Zha HG, Jiang RF, Zhao FJ, Vooijs R, Schat H, Barker JHA, McGrath SP. 2004.** Co-segregation analysis of cadmium and zinc accumulation in *Thlaspi caerulescens* interecotypic crosses. *New Phytologist* **163**: 299–312.

**Zhai Z, Gayomba SR, Jung H-I, Vimalakumari NK, Piñeros M, Craft E, Rutzke M a, Danku J, Lahner B, Punshon T, et al. 2014.** *OPT3* Is a Phloem-Specific Iron Transporter That Is Essential for Systemic Iron Signaling and Redistribution of Iron and Cadmium in Arabidopsis. *The Plant cell* **26**: 2249–2264.

**Zhao F-J, Hamon RE, Lombi E, McLaughlin MJ, McGrath SP. 2002.** Characteristics of cadmium uptake in two contrasting ecotypes of the hyperaccumulator *Thlaspi caerulescens*. *Journal of experimental botany* **53**: 535–543.



## Summary

### **An evolutionary and functional genomics analysis of the heavy metal hyperaccumulator *Noccaea caerulescens***

*Noccaea caerulescens* is the only known Zn/Cd/Ni hyperaccumulator. The Ganges accession (2n = 14) has an, yet unpublished, genome size of ~319 Mb, with 29,712 predicted genes representing 15,874 gene families. This species is distributed mainly in Europe. Three ecotypes can be distinguished: two metalicolous ecotypes, resident to serpentine soil (Ni enriched) and calamine soil (Zn/Cd enriched), and a non-metallicolous ecotype, growing on regular, non-metalliferous soils. The physiological differences that underlie variation in heavy metal accumulation and tolerance are well-understood, and the molecular basis of hyperaccumulation and tolerance has been explored by transcript profiling in the presence of metals and by comparative transcriptome analysis using *N. caerulescens* and non-hyperaccumulators such as *Arabidopsis thaliana*. The genetic variation which emerged during the evolution of metal hyperaccumulation has not yet been investigated. The work described in this thesis considers the identification of genetic variation under selection for Zn/Cd hyperaccumulation and tolerance by next generation resequencing of the wild metalicolous (calamine) and non-metallicolous populations and the generation of a mutant *N. caerulescens* library for functional analysis. The regulation of flowering time was also investigated, using early flowering mutants selected from the mutant library.

**Chapter 2** describes the identification of genetic variation under selection for Zn/Cd accumulation or tolerance using a comparative whole-genome resequencing approach featuring the metalicolous (M) calamine (Zn/Cd enriched) and non-metallicolous (NM; normal) ecotypes. Two M populations (Prayon and Pontaut) and nearby NM populations (Lellingen and Vall de Varrados, respectively) were used for this purpose, as well as the geographically remote populations Lanestosa (M) and Navacelles (NM). Mutations in coding regions that are consistently and extremely divergent ( $F_{st} > 0.96$ ) between the two geographically-close pairs were investigated and nonsynonymous substitutions were identified. The functional validation of strong candidates in the

corresponding orthologous genes from *A. thaliana* T-DNA knockout mutants revealed that mutants for *GL3* and a *LRR-RLK* gene showed phenotypes for Zn/Cd tolerance and accumulation, indicating convergent independent local evolution in *N. caerulescens*. Many further genes were identified in an additional comparison based on phenotypic variation in either accumulation, or tolerance or translocation, including the stress-responsive transcription factor *WRKY72*, and the metal homeostasis gene *OPT3*. One gene under directional selection was found to be unique to *N. caerulescens*. The SNP library created herein will provide a valuable resource for future investigations. Evidence for further genetic variation was detected, indicating that *N. caerulescens* is a comprehensive system for the analysis of molecular mechanisms underlying the evolution of metal hyperaccumulation and tolerance.

**Chapter 3** describes the generation of a mutant library based on the Southern French accession St Felix de Pallières (SF), a Zn/Cd hyperaccumulator. The frequency of mutants with observable phenotypes in the M2 population was 0.4%, suggesting that saturation had been achieved and the library could be used to identify mutants for most genes of interest. The forward screening of 7000 individuals from the M2 population by mineral concentration analysis (ionomics) in leaf material revealed putative mutants affecting the concentration of 20 trace elements. This mutant collection could be used to identify novel genes involved in mineral homeostasis. The high resolution melting curve analysis (HRM) was optimized and successfully applied to identify mutant allelic series in the TILLING population. This is the first reported TILLING population for *N. caerulescens*, and will facilitate the application of functional genomics in this species as well as comparative functional analysis with its relatives.

**Chapter 4** investigated flowering time variations reflecting the requirement for and response to vernalisation in four *N. caerulescens* accessions from distinct environments. The role of *FLC* and *SVP* as major floral repressors was confirmed by analysing EMS-induced mutants in both genes, obtained by screening the mutagenized M2 population for vernalisation-independent plants. *N. caerulescens* is a biannual/perennial species with an intermediate lifestyle between the annual plant *Arabidopsis thaliana* and the perennial plant *Arabis alpina*, and *FLC* expression in this

species is partially repressed by vernalisation. The flowering plant reverted from reproductive to vegetative growth under high temperature conditions, correlating with the low-level expression of floral organ genes, suggesting this species has evolved a unique strategy to cope with changing climates. Based on those findings, a model has been developed to explain the regulation of flowering time by the vernalisation pathway in this biannual/perennial species and more evidence is provided in comparison to *A. thaliana* and *A. alpina*.

In **Chapter 5** the main results of the experimental chapters are discussed and a perspective on future research is provided.

## Acknowledgements

Finally my PhD thesis is done, and it is time to show my thanks to many people I have met especially in department of Plant genetics of Wageningen University, because without the help and accomplish of those people it is impossible for me to finish the whole process.

I would like to show my deepest thanks to my dear promoter Maarten Koornneef, thank you so much for your patience guidance, motivation and huge contribution to the thesis. It is really my immense honour to be part of your group during my PhD as you have always been source of inspiration, motivation and support whenever needed. Every time after the discussion with you, my thoughts and logic become clear and I am fully motivated to perform the experiment. No words could be good enough to express how great it was to work with you. I am very much impressed and enjoyable of the way you interpret science!

Dear daily supervisor Mark, thank you very much for offering me the opportunity to do this PhD project and for your huge contribution to the thesis. Having you as my daily supervisor has been very inspiring and constructive. Thank you so much for your suggestion every time when I asked for courses, meetings, and for offering innovative ideas for the experiments. I really appreciated your great help and impressed by your professional experience and critical scientific thinking, which influenced me very much when I performed the experiments. Apart from science, many thanks for picking me up at the train station in Ede-Wageningen in the deep evening of 6th of Oct, 2011, the first day I arrived in the Netherlands. I will never forget that you showed me the windmill in Wageningen when I asked why I had not seen any windmill in the Netherlands yet. After that you guided me to the campus as well! Also I am very much appreciated for your hospital invitation to celebrate the Sinterklaas at your places; by the way the soup you cooked is so tasty!

Dear external supervisor Henk, thank you so much for your great help for mineral concentration measurement and statistical analysis. I was surprised that you have



always done the statistical analysis manually. I am very happy that I have learned this skill from you : ) I know I could do it with computer easily, but I really enjoyed the process of doing the statistical manually together with you and checked the raw data when something went wrong unexpectedly. I was very much enjoyable to correct the chapter together with you as well.

Dear Joost van den Heuvel, many thanks for doing the bioinformatics analysis for the population resequencing project. You are a very kind person, you always explain the bioinformatics and figures in a very patient way, which makes me feel very comfortable and very interested to learn new things. Every time when I have questions, you are very patient to answer and try to explain things in an easier way. You are a very nice and humours colleague who is very easy to collaborate with. With your great help my PhD life has been much more relaxed!

I would like to show my thanks to the students who have been involved in part of the work described in this thesis. I am very happy to work together with my students Bastiaan Millenaar, Yufeng Mao and Hui Su. Having you around contributed not only to my research but also to my understanding of how to be a supervisor.

I also would like to express my thanks to my lab fellows: Valeria, Tania, Mina, Aina, Pingping, Roxanne, Edouard, Ya-Fen, Robert, Mohammed, Ross, Zeshan, Diana, Nilma, Ana Carolina, Ana Paula, Charles, Nihal, Padraic, Frank, Michele, Sam, Rene, Kaige, Xingzheng, Jianhua and Eric, Cris, Ramon, Xianwen, Alex, Claudio, Tina, Andy, Sabine, Margo, Jelle, Florian, Anneloes, Kim, Marjon for all the help you offered and nice moments you created. My experiment would not have been possible without the technical support, so I am very grateful to Corrie, Frank, Jose and Bertha for the help in the lab whenever I need, and to Wytske for your very kind management of official formalities. I also would like to thank Technician Fien in Plant breeding for assisting me when I needed to isolate the very high quality of DNA samples from *Noccaea caerulescens* for Pacbio and many thanks for assisting me when I screened the mutations by the Light scanner. Finally, I would like to express my gratitude to Bas Zwaan for always keeping the nice and friendly atmosphere inside of the genetics group.

To all my Chinese friends in Wageningen, thank you very much for your accompany during my PhD, Xuan, Zhao Tao, Yanxia, Li Tao, Wei Zhen, Tinting, Yuanyuan, Bai Bing, Li Tao, Li Hui, Du Yu, Wang Yan, Hanzi, Junwei, Yunmeng, Song Wei, He Jun and Xiaohua, Huchen, Cheng Xu, Zhu Feng, Song Yin, Jinling, Guiling, Jinbing, Cheng Xi, Shuhang, Lin xiao, Kaile, Xiaoxue, Yan Ting, Chunxu and Qin Wei, Minggang and Fengjiao, Yang Ting and Liu Qing, Huicui, Mengli, Xiao Dong, Xiaomei, Deng Ying, Shuxin Xuan, Shougang Ren and all other Chinese friends. I am very happy to know all of you in Wageningen, your friendship make my life in the Netherlands very colourful.

To my friends in China 学长尚力青，超超，明玉，郭伟，立伟，久树，大学四位舍友，郑同学，文生同学，非常感谢你们的陪伴支持和鼓励！同时也感谢所有初中，高中，大学，硕士研究生的所有同学们，你们的友谊是我人生中最宝贵的财富，非常感谢在不同的人生阶段能有你们的陪伴！

Last but not least, to my family members: 爷爷奶奶，爸爸妈妈，叔叔姑姑舅舅阿姨，哥哥弟弟姐妹妹，小外甥（女）小侄子们，Thank you all so much for your understanding and encouragement. I know you all are waiting for me, finally I will be back soon to you to enjoy the happiness of our big family!

## Curriculum Vitae



Yanli Wang was born on the 20<sup>th</sup> of May 1983 in Weichang, Hebei, China. In 2004 she studied at the department of Horticulture in Hebei Normal University of Science and Technology in Qinhuangdao, China. After she received her Bachelor degree of Horticulture in June 2008, in the same year she moved to the Northwest A&F University in Shaanxi, China to perform her master's study in the Department of Horticulture with a specialization in pomology. Where she studied apple breeding and the thinning effect of organic chemicals on the apple flowers. She received a Master degree of horticulture in 2011 and in the same year she obtained a scholarship from the China Scholarship Council and started her PhD study at Wageningen University under the supervision of Dr. Mark Aarts and Prof. Maarten Koornneef. For the PhD degree, she did research about evolutionary and functional analysis of a heavy metal hyperaccumulating plant species *Noccaea caerulea* which resulted in this thesis.

## Publications

**Yanli Wang, Joost van den Heuvel, Edouard Severing, Elio Schijlen, John Danku, David E Salt, Henk Schat, Mark G.M. Aarts.** Population genomics identifies candidate genes under selection distinguishing non-metallicolous and calamine populations of the heavy metal adapted plant species *Noccaea caerulescens*. In preparation

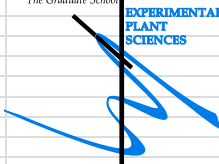
**Yanli Wang, YuFeng Mao, Bastiaan Millenaar, John Danku, David E. Salt, Maarten Koornneef and Mark G.M. Aarts.** Construction and analysis of a *Noccaea caerulescens* TILLING population. In preparation.

**Yanli Wang, Maarten Koornneef, Mark GM Aarts.** The genetic control of flowering time in the biennial/perennial species *Noccaea caerulescens*. In preparation.

**Chuankui Song, Yanli Wang, Zhengcao Xiao, Bin Xiao** (2015) Protective effects of Green Tea Polyphenols against Benzo[a]pyrene-induced reproductive and developmental toxic effects in Japanese Medaka (*Oryzias latipes*). **Journal of functional foods**,14:354–362.

**Graaf RM, Krosse S, Swolfs AE, Brinke E, Prill N, Leimu R, van Galen PM, Wang Y, Aarts MG, van Dam NM** (2015) Isolation and identification of 4- $\alpha$ -rhamnosyloxy benzyl glucosinolate in *Noccaea caerulescens* showing intraspecific variation. **Phytochemistry**,110:166–171.

## Education statement (EPS)

Education Statement of the Graduate School		The Graduate School
Experimental Plant Sciences		EXPERIMENTAL PLANT SCIENCES
Issued to:	Yanli Wang	
Date:	9 September 2016	
Group:	Laboratory of Genetics	
University:	Wageningen University & Research	
<b>1) Start-up phase</b>		<u>date</u>
▶ <b>First presentation of your project</b>		
Title: 'Evolutionary genetics of plant metal hyperaccumulation for improved plant nutrition and quality'		Apr 24, 2012
▶ <b>Writing or rewriting a project proposal</b>		
Title: 'Evolutionary genetics of plant metal hyperaccumulation for improved plant nutrition and quality'		Mar 10, 2012
▶ <b>Writing a review or book chapter</b>		
▶ <b>MSc courses</b>		
▶ GEN20806 Plant Biotechnology		Nov-Dec 2011
GEN30306 Genetic Analysis, Tools and Concepts (GATC)		Sep-Oct 2012
▶ <b>Laboratory use of isotopes</b>		
Subtotal Start-up Phase		9.0 credits *
<b>2) Scientific Exposure</b>		<u>date</u>
▶ <b>EPS PhD student days</b>		
4th joint PhD retreat, Norwich UK		Aug 14-17, 2012
5th joint PhD retreat, Gent Belgium		Jul 22-25, 2013
EPS PhD Student days 'Get2Gether', Soest, NL		Jan 29-30, 2015
▶ <b>EPS theme symposia</b>		
EPS Theme 4 Symposium 'Genome Biology', Wageningen University		Dec 09, 2011
EPS Theme 1 Symposium 'Developmental Biology of Plants', Wageningen University		Jan 19, 2012
EPS Theme 2 Symposium 'Interactions between Plants and Biotic Agents', Wageningen University		Feb 10, 2012
EPS Theme 3 Symposium 'Metabolism and Adaptation', Utrecht University		Apr 26, 2012
EPS Theme 4 Symposium 'Genome Biology', Radboud University Nijmegen		Dec 07, 2012
EPS Theme 3 Symposium 'Metabolism and Adaptation', University of Amsterdam		Mar 22, 2013
EPS Theme 4 Symposium 'Genome Biology', Wageningen University		Mar 11, 2014
EPS Theme 3 Symposium 'Metabolism and Adaptation', Wageningen University		Feb 10, 2015
EPS Theme 3 Symposium 'Metabolism and Adaptation', University of Amsterdam		Feb 23, 2016
▶ <b>Lunteren days and other National Platforms</b>		
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL		Apr 02-03, 2012
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL		Apr 22-23, 2013
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL		Apr 22-23, 2014
Annual Meeting 'Experimental Plant Sciences', Lunteren, NL		Apr 11, 2016
▶ <b>Seminars (series), workshops and symposia</b>		
Invited seminar : 'Bananas for the 21st Century: pushing back the threat of extinction'		Sep 20, 2012
Plant Genome Mining (CBSG Bioinformatics Workshop)		Dec 13, 2012
Mini-symposium 'How to write a World-class Paper'		Oct 17, 2013
Mini-Symposium: Developments in Arabidopsis Research		Apr 11, 2013
Symposium: All-inclusive Breeding: Integrating high-throughput science		Oct 16, 2014
Invited seminar 'Ionomics: Genomics of mineral nutrient and trace elements in plants'		Oct 23, 2014
Invited seminar: 'Genetics and epigenetics : a complex relationship'		Nov 19, 2014
Invited seminar: 'Seasonal flowering in annual and perennial plants'		Jan 19, 2015
Invited seminar: 'The evolutionary significance of gene and genome duplications'		Feb 03, 2015
Symposium 'Future for cytogenetics in plant genomics and breeding'		Oct 01, 2015
Invited seminar: 'Analysis of qPCR data. The use and usefulness of amplification curve analysis'		Mar 14, 2016
▶ <b>Seminar plus</b>		
▶ <b>International symposia and congresses</b>		
LOCOME, France		Jul 02-03, 2013
Evolutionary genomics and genetics, Suzhou, China		Oct 07-12, 2014
LOCOME, France		Jul 02-03, 2014

## Education statement (EPS)

► <b>Presentations</b>	
Poster Summer School MISTRAL	July 02-13, 2012
Talk: LOCOME, France	Jul 02-03, 2013
Poster: 5th joint PhD retreat, Gent Belgium	Jul 22-25, 2013
Talk: LOCOME, France	Jul 02-03, 2014
Talk: Evolutionary genomics and genetics, SuZhou, China	Oct 07-12, 2014
Poster: Evolutionary genomics and genetics, SuZhou, China	Oct 07-12, 2014
► <b>IAB interview</b>	
► <b>Excursions</b>	
<i>Subtotal Scientific Exposure</i>	<i>18.3 credits *</i>
<b>3) In-Depth Studies</b>	<u><i>date</i></u>
► <b>EPS courses or other PhD courses</b>	
Summer school MISTRAL IRNA, France	Jul 02-13, 2012
Summer School of Natural Variation in Plants, Wageningen	Aug 21-24 2012
Current Trends in Phylogenetics	Oct 22-26, 2012
Bioinformatics-A User's Approach	Aug 27-31, 2012
Mixed model based QTL mapping in GenStat	May 14-16 2012
Transcription factors and transcriptional regulation	Dec 17-19, 2013
The Power of RNA-seq	Jun 05-07, 2013
De Novo Assembly from NGS data	Jan 08-09, 2013
► <b>Journal club</b>	
Literature Discussion (Lab of Genetics)	2012-2016
► <b>Individual research training</b>	
<i>Subtotal In-Depth Studies</i>	<i>12.1 credits *</i>
<b>4) Personal development</b>	<u><i>date</i></u>
► <b>Skill training courses</b>	
Competence Assessment (CA)	Mar 20, 2012
Course: English Academic writing II	Mar-July 2012
Course: English IV Speaking and listening	Mar-July 2012
Course: Project and Time Management (P&TM)	Nov 06, 20 & Dec 18, 2012
Interpersonal Communication for PhD students	Apr 10-11, 2014
► <b>Organisation of PhD students day, course or conference</b>	
► <b>Membership of Board, Committee or PhD council</b>	
<i>Subtotal Personal Development</i>	<i>6.0 credits *</i>
<b>TOTAL NUMBER OF CREDIT POINTS*</b>	<b>45.5</b>
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.	

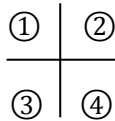
## Sponsors

The project was financed by the China Scholarship Council (CSC) and the Laboratory of Genetics, Wageningen University, The Netherlands.

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

## Colophon

Cover photography:



① *Noccaea caerulea* plants grown under hydroponic system (Chapter 2) ② *Arabidopsis thaliana* plants grown under hydroponic system (Chapter 2); ③ Non-vernalisation required early flowering mutant selected from EMS mutagenesis experiment(Chapter 3 and Chapter 4); ④ Leaves from Early flowering plants grown under two different temperature regimes (Chapter 4).

Thesis layout and cover design: Yanli Wang

Printed by Proefschriftmaken.nl// Uitgeverij BOXPress