Genetic consequences of colonization of a metal-polluted environment, population genetics and quantitative genetics approaches

By

Dima Souleman

Submitted to the University of Lille 1 in accordance with the requirement for the degree of Doctor of Philosophy

Speciality: Evolutionary Biology and Ecotoxicology

Supervised by:

Pr VANDENBULCKE Franck **Dr FREROT Hélène**

Jury members:

Dr NORET Nausicaa	**Université Libre de Bruxelles**
Pr BLOUIN Manuel	**AgroSup Dijon**
Dr FAUCON Michel-Pierre	**Institut Polytechnique LaSalle Beauvais**
Pr ARNAUD Jean-François	**Université de Lille 1**
Dr PAUWELS Maxime	**Université de Lille 1**

January 2017

Genetic consequences of colonization of a metal-polluted environment, population genetics and quantitative genetics approaches

By

Dima Souleman

Submitted to the University of Lille 1 in accordance with the requirement for the degree of Doctor of Philosophy

Speciality: Evolutionary Biology and Ecotoxicology

Supervised by:

Pr VANDENBULCKE Franck **Dr FREROT Hélène**

Jury members:

Dr NORET Nausicaa	**Université Libre de Bruxelles**
Pr BLOUIN Manuel	**AgroSup Dijon**
Dr FAUCON Michel-Pierre	**Institut Polytechnique LaSalle Beauvais**
Pr ARNAUD Jean-François	**Université de Lille 1**
Dr PAUWELS Maxime	**Université de Lille 1**

January 2017

Parfois notre lumière s'éteint, puis elle est rallumée par un autre être humain. Chacun de nous doit de sincères remerciements à ceux qui ont ravivé leur flamme.

Albert Schweitzer

Je dois terminer ce travail en disant un grand merci aux personnes qui ont cru en moi et qui m'ont donné la chance d'arriver au bout de cette thèse. Mais par où commencer ? par quel ordre ? c'est la décision la plus difficile, mais quand même je me lance!

Je tiens tout d'abord à remercier mon directeur de thèse, Monsieur Franck Vandenbulcke, pour m'avoir accordé sa confiance et permis d'effectuer cette thèse au sein du laboratoire LGCgE. C'est à lui que je dois toute ma reconnaissance pour mon entrée dans le monde de la recherche.

Ensuite, j'adresse mes vifs remerciements à ma co-directrice Hélène Frérot, et à Maxime Pauwels, qui ont pris le temps de m'aider à organiser et à rédiger cette thèse, et pour leur présence et disponibilité dans les moments difficiles. Je ne sais comment exprimer ma gratitude à ces deux personnes autrement qu'en leur promettant d'agir comme eux avec des étudiants dans ma situation.

J'exprime tous mes remerciements à l'ensemble des membres de mon jury : Monsieur le Professeur BLOUIN Manuel, Monsieur le Docteur FAUCON Michel-Pierre, Jean-François ARNAUD, Nausicaa NORET.

En ce qui concerne la partie bioinformatique de ce travail, je remercie Marie-joe Karam qui est devenu ma meilleure amie.
Je remercie de plus le Docteur Fabien Grumiaux pour son aide à collecter et à identifier les vers de terre.

J'adresse également mes remerciements aux personnes qui m'ont aidé avec les gros travaux à la serre (Angélique, Cédric, Eric, Nathalie), ainsi qu'au laboratoire (Cécile et Anne-Catherine).

Je remercie toutes les personnes formidables que j'ai rencontrées au LGCgE (Alain, Claude, Mathieu, Sébastien, Céline, Fabien, Sylvain).

Je remercie tous les doctorant au SN3 pour la bonne ambiance de travail mais également pour les nombreux moments passés ensemble (Hanine, Léa, Coralie, Joséphine, Fabien)

Je remercie également tous les doctorants au SN2. Particulièrement Hélène et Leslie pour leur aide concernant les analyses de génétique des populations, Julien pour ces dessins formidables et son humour.

Je tiens à remercier Agnieszka pour ses relecture et correction de mon anglais mais également pour son amitié, elle est devenue comme une sœur pour moi.

Je voudrais remercier mes amies à l'université avec qui je mange et je prends mon café chaque jour (Sabette, les nouveaux mariées Salman et Karima, Chadi, …).

Mes remerciements vont également à mes amours (Nicole, Judy, Nancy, les futures mariées Naïm et Mijo, Ali, Marc, Sami) qui m'ont permis d'oublier momentanément le travail dans des soirées, repas, sorties…

Je remercie mes parents et mes frères de m'avoir soutenu de loin.
Enfin, Je remercie mon chéri Reda qui m'a soutenu, m'a supporté, m'a encouragé … pendant toute la durée et les moments difficiles de rédaction qui n'ont pas toujours été agréables.

This thesis is dedicated to my parents.

For their endless love, support and encouragement.

Abstract

Natural habitats are more and more destructed and fragmented by urban expansion and human activities. The fragmentation of natural and agricultural areas by buildings and new infrastructures affects the size, connectivity and the quality of habitats. The populations of organisms inhabiting these anthropized territories are then more isolated. However, differentiation between populations of the same organism depends on demographic and genetic processes such as genetic drift, gene flow, mutation and natural selection. Only species that have developed special tolerance mechanisms can persist under changed environmental conditions. The introduction of contaminants such as metals in the environment may influence plants and animals evolution by modifying the evolutionary forces and thus generating differences between populations.

In this work, attention was focused on the genetic consequences of metallic pollution on two species, the earthworm *Lumbricus terrestris* and the plant model *Arabidopsis halleri*. Two different approaches have been used to study the genetic response to metallic contamination: a population genetic approach was performed in *L. terrestris* and a quantitative genetic approach was carried on in *A. halleri.*

First, it was a question of identifying and validating new microsatellite markers in *L. terrestris*. These markers were then used to characterize the neutral genetic diversity in worms collected from agricultural and urban sites. Secondly, genetic architecture of Zn tolerance and Zn hyperaccumulation was conducted investigated for the first time using an intraspecific crossing between metallicolous and non-metallicolous individuals of *A. halleri*. High density of SNP markers was used to proceed to the QTL mapping step.

Résumé

Les habitats naturels sont de plus en plus détruits et fragmentés par l'expansion urbaine et les activités humaines. La fragmentation des espaces naturels et agricoles par les bâtiments et les nouvelles infrastructures affecte la taille, la connectivité et la qualité des habitats. Les populations d'organismes vivants sur ces territoires anthropisés sont alors plus isolées. Or, la différenciation entre populations d'un même organisme dépend de processus démographiques et génétiques tels que la dérive génétique, le flux génétique, la mutation et la sélection naturelle. La persistance et le développement des populations dans des conditions environnementales modifiées dépendent de mécanismes de tolérance. Dans ce contexte, l'introduction de contaminants tels que des métaux dans l'environnement peut influencer l'évolution des plantes et des animaux en modifiant les forces évolutives et en créant des différences entre populations.

Dans ce travail, l'attention a été portée sur les conséquences génétiques de la pollution métallique sur deux espèces, le ver de terre *Lumbricus terrestris* et une plante modèle *Arabidopsis halleri*. Deux approches différentes ont été utilisées pour étudier la réponse génétique à la contamination métallique : une approche de génétique des populations chez *L. terrestris* et une approche de génétique quantitative chez *A. halleri*.

Tout d'abord, il s'est agi d'identifier et de valider de nouveaux marqueurs microsatellites chez *L. terrestris*. Ensuite, ces marqueurs ont été utilisés afin de caractériser la diversité génétique neutre chez des vers collectés sur des sites agricoles et urbanisés. Parallèlement, l'architecture génétique de la tolérance et de l'hyperaccumulation de Zn chez *A. halleri* a été explorée à l'aide d'un croisement intraspécifique entre une population métallicole et une population non métallicole. Une densité élevée de marqueurs SNP a été utilisée pour procéder à l'étape de cartographie QTL.

Table of contents

Chapter I

General Introduction

Table of contents for chapter 1

1 Anthropogenic landscape and the consequences on biodiversity

Anthropogenic activities that have increased with the expansion of cities and the industrial revolution are widely implicated in environmental changes. These changes are gathered under the term "Global Change". Drivers of Global Change are for example the production and the introduction of contaminants (pesticides, fertilizers, petroleum, metals…) in nature, the overexploitation of resources and the change of land use. About half of the land cover of the planet was subject to land use changes (Vitousek *et al.*, 1997), which consist primarily in transformation of natural habitats into agricultural or urban areas. It is predicted that the total urban area will triple between 2000 and 2030 in accordance with the explosion of human populations and their growing consumption rates of raw materials (McDonnell & Hahs, 2015). Generally land use and the other anthropogenic activities favour fine-scale changes in environmental features, and thus increase spatial environmental heterogeneity, which leads to the transformation of large homogeneous habitat into small habitat fragments isolated from each other. The fragmentation and loss of natural habitats include their replacement by either other habitats or urban structures (buildings, roads, etc.). Ecosystems will be impacted through the destruction, modification and fragmentation of habitat, degradation of the quality of soil, air and water, exploitation of native species and introduction of exotic species. These modifications can affect locally or globally the biodiversity and can lead to the extinction or decline of species (Figure 1).

The term of biodiversity means all living species and the communities formed by them in the different types of ecosystems (marine, terrestrial and aquatic). Each ecosystem is based on complex relationships between biotic and abiotic components. The ecological importance of biodiversity is reflected by the mutuality between different species, providing essential services to life, such as production of oxygen and fixing of air carbon dioxide, filtration and purification of water, pollinate plants, waste decomposition and transfer of nutrients in food chains. Biodiversity is therefore a central element to consider in the context of Global Change effect on ecosystems.

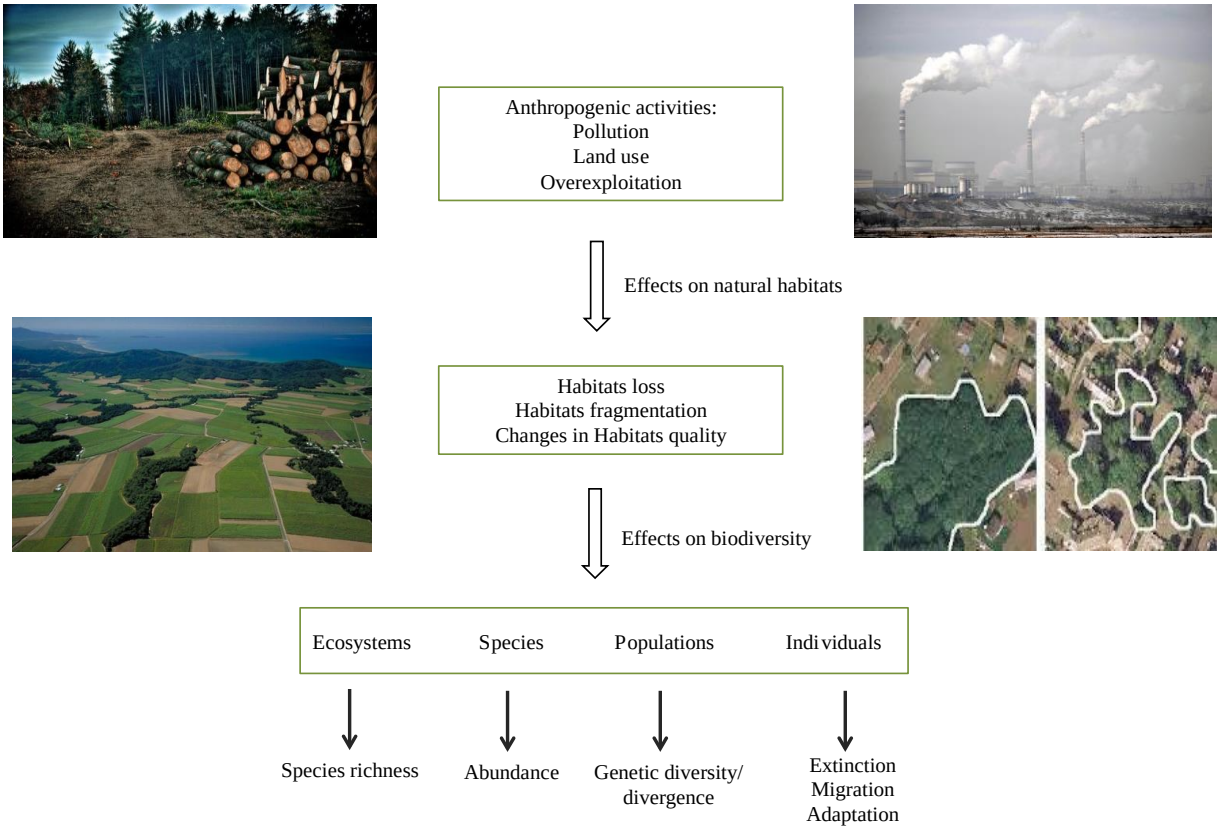


Figure 1 : Anthropogenic activities and their consequences on biodiversity

Consequences of habitat fragmentation are particularly well documented. At the ecosystem level, species richness within habitats is expected to decrease with the reduction of habitat surface (Krauss *et al.*, 2010). About 40% of species could be lost around the world because of habitat destruction and fragmentation (Seto *et al.*, 2012). For example, 85% of bird species and 86% of mammals are threatened because of their habitats destruction (Baillie *et al.*, 2004). Indeed, a lower number of species in patches of small area has been shown in several studies including plants (Cagnolo *et al.*, 2006), insect parasitoids (Kruess & Tscharntke, 2000), reptiles and mammals (Findlay & Houlahan, 1997). In fact, habitat fragmentation generally results in population fragmentation: ancestral populations are divided in a set of smaller populations separated by unsuitable environments that may act as a barrier to the movement of the individuals and thus limit migration events among populations (Fahrig, 2003). Thus, habitat fragmentation must affect the abundance of local populations (Wilcox & Murphy, 1985). Some studies were conducted to test the effect of fragmentation and isolation in shaping species abundance, for example: (Virgós, 2001) shows that the abundance of

badger populations (*Meles meles*) decreases with the diminution of the percentage of forest cover. At the genetic level, population fragmentation is expected to affect the distribution of genetic diversity among populations. Fragmented populations are expected to show lower genetic diversity (within populations) and higher genetic differentiation (among populations) (Templeton *et al.*, 1990; Young *et al.*, 1999).

Among environmental changes due to anthropogenic activities, decrease of habitat quality due to pollution is a main source of environmental heterogeneity - *i.e.* spatial and temporal variations of the environment – and stress for living organisms. However, the consequences of pollution on biodiversity have been little documented so far. Therefore, the main interest of this thesis is the evolutionary effects of pollutant in anthropogenic environment.

2 Evolutionary ecotoxicology

2.1 Background

In the context of Global Change, ecosystems are frequently and locally impacted by toxic substances (ex. pesticides, metals, polycyclic aromatic hydrocarbons). Evolutionary ecotoxicology is a discipline studying the genetic and evolutionary processes underlying organisms' reactions in response to environmental toxicants. The main goal of this discipline is to outperform the traditional parameters such as the toxicants effects at the molecular, cellular or physiological levels on individuals, species or populations by understanding the genetic basis of organisms' response to the toxicants exposure and their evolutionary impact.

2.2 Adaptation

The ways organisms can face drastic and rapid environmental changes that result from anthropogenic activities are:

- Extinction (global of local populations).
- Permanence through migration to a place with more favourable environmental conditions.
- Permanence through adaptation to the new environmental conditions.

Migration to the most suitable environments depends on organism's capacity to move and to cross the barriers between habitats. However, populations might also have locally survived to habitat loss or modification through genetic adaptation to the new environmental conditions (Aitken *et al.*, 2008). The impact of abiotic and biotic environmental changes on the reactions

and adaptations of plants and animals has always been a central issue in evolutionary biology (Møller & Nielsen, 2010). Many definitions were proposed for the word "adaptation" depending on the discipline of study (Futuyma, 1998). Adaptation for physiologists indicates the capacity of individuals to adjust their phenotype to cope with their environment. In this context, the word "adaptation" may be a source of confusion since observed changes are due to the capacity of organisms to modify their physiology or morphology in response to environmental changes. These modifications do not require genetic changes and we will prefer the terms "acclimation" or "phenotypic plasticity". In evolutionary biology, adaptation corresponds to the selection of favourable phenotypes conferring features allowing reproductive and survival success (*i.e.* adaptive trait) but requires genetic change. For biodiversity rescue, it is important to understand how organisms can adapt to their modified environmental conditions and the mechanisms underlying this adaptation.

In this document, I will use the word "adaptation" in the sense of "genetic changes leading to a higer fitness (*i.e.* the ability of an individual to reproduce and give descendants) in the new environment".

Adaptation is the result of natural selection (see the definition in the section of evolutionary forces section 3.1.4) that increases the frequency of the most favourable genotypes that have higher fitness (or selective value) in the modified environment (Hendry *et al.*, 2011). In anthropogenic environments, genetic adaptation is a process that requires prolonged and repeated exposure to stress (ex: contaminants) over several generations, which increases the magnitude of selection pressure and thus imposes a quick adaptation of organisms to persist. It can nevertheless be established quickly, in a few years and in a few generations, as observed (Klerks & Levinton, 1989) for a natural population of *Limnodrilus hoffmeisteri* (oligochaete) which has developed a genetic adaptation to a metal contamination 30 years after establishing pollution. Moreover, adaptation to altered environment is faster in organisms having a short generation time that can be measured by days, months or few years such as microorganisms and insects. However, the probability of adaptation is lower in the organisms with long generation time (ten or hundred years) such as large mammals and trees (McDonnell & Hahs, 2015).

Numerous studies revealed the selective effects of pollutants and cases of tolerance were documented, for example: tolerance to metals (Morgan *et al.*, 2007; Bourret *et al.*, 2008; Lévêque *et al.*, 2015) or pesticides (Brausch & Smith, 2009; Pelosi *et al.*, 2013) in natural populations exposed to frequent or chronic industrial sources of pollutants or agrochemicals.

Tolerance is defined as heritable changes in a population, which confer a better ability to survive and reproduce in an environmental condition which was previously lethal for the species to which this population belongs (Antonovics *et al.*, 1971). The contaminant acts as a selective force on the population by reducing the contribution of susceptible genotypes in the next generation by reducing their survival or reproduction (Klerks, 2002; Millward & Klerks, 2002). Thus, tolerant individuals will contribute more to the next generations (Dallinger & Höckner, 2013). Genetic adaptation to contamination results in a change of allele frequencies within the adapted population (Klerks, 2002). This is a microevolutionary process (Ribeiro & Lopes, 2013), since micro-evolution had been defined by Kinnison & Hendry (2001) as the evolution within and between populations.

Furthermore, two types of tolerant individuals can be selected: (1) individuals who are constitutively tolerant to pollutants, meaning that genetic mechanisms were already present before environmental pollution, and / or (2) individuals who can develop instantly and effectively defense systems against pollutants. To be selected, the latter mechanism should have genetic bases and be present in high frequency in the population adapted to the environmental pollution (Posthuma & Van Straalen, 1993).

The development of tolerance resulting from the selective effect of contaminant may affect other traits, due to genetic correlation between these traits and the trait corresponding to selection (Sgrò & Hoffmann, 2004). These effects may occur because of linkage disequilibrium between genes that control different traits or pleiotropic effects. In particular, selective effect of a contaminant can provide better tolerance toward other contaminants. This is called co-tolerance (Brausch & Smith, 2009) that may due to shared mechanisms of detoxification, such as the increased production of metallothionein in response to cadmium and copper (Morgan *et al.*, 2007).

Local adaptation is a special case of adaptation for evolutionary biology, which focuses on populations' evolution at a local scale. Local adaptation requires the action of spatially heterogeneous selective pressures that are sufficiently strong and long lasting. When populations are found in different (biotic or abiotic) environments, these environments will exert different selection pressures on different populations, then each population may evolve toward optimal local adaptation. Consequently, those populations will be differentiated owing to their different environment. In anthropogenic environments, the establishment of local adaptation depends on several factors including: 1) differentiation in survival and reproduction success among individuals (selective value for a genotype) across many

generations 2) when genetic polymorphisms exists, it provokes changes in allelic or genotype frequencies in population that are under selection pressure compared to geographically closed ones. In a spatial heterogeneous environment, the spatial variation of natural selection pressures led to changes in the structure of genetic diversity of populations connected by genes flow (Kawecki & Ebert, 2004). Local adaptation in the strict sense implies the existence of genotype × environment interaction, *i.e.* alleles and genes have different effects on the fitness of individuals in different environments. Moreover, individuals have a better selective value in their local environment than immigrants (Figure 2) (Kawecki & Ebert, 2004).

Local adaptation is necessary for species to react to changing environments. Nevertheless, this mechanism also plays essential role in species divergence, which leads to the development of new species. All those reasons make the study of local adaptation interesting for evolutionary biology. The concept of local adaptation can be studied through population genetics, describing the evolutionary forces acting on genetic diversity, or using quantitative genetic approaches, investigating the genetic determinism of quantitative traits linked to fitness.

2.3 How to study local adaptation in polluted habitats

The same approaches are currently used for the study of local adaptation and the evolutionary issue in ecotoxicology (Coutellec & Barata, 2011) such as:

1) Population and landscape genetics: is a discipline studying how alleles and genotypes frequency have changed between and within populations. Landscape genetics is an emerging discipline at the interface between landscape ecology, population genetic and the spatial statistics (Storfer *et al.*, 2006). Its objective is to describe and analyze the influence of landscape structures and environmental factors on the spatial structure of genetic variability of populations (Manel *et al.*, 2003; Storfer *et al.*, 2006; Holderegger & Wagner, 2008). The main interest of landscape genetics is to understand the interactions between landscape structures and micro-evolutionary processes such as gene flow, genetic drift or selection.
2) Quantitative genetics: the phenotypic expression of quantitative traits is the result of the influence of both genotype and environmental conditions. This relationship is the basis of quantitative genetics and their application in evolutionary ecotoxicology (Klerks *et al.*, 2011). The main goal of quantitative genetics is to explore the genetic

architecture of the quantitative traits selected in anthropogenic environment. Three approaches are available in ecotoxicology to study the variation of a quantitative trait (ex: contaminant resistance). The first approach aims at comparing the resistance levels among populations originating from polluted sites and non-polluted sites. The second approach aims at determining the relation dose-response of pollutant exposure by applying an artificial selection. The third approach analyses the genetic variation of resistance in a population, such genetic variation being important for selection to operate (Klerks *et al.*, 2011). Genetic variation also helps estimating the heritable part of quantitative traits that can influence the evolution of adaptation.

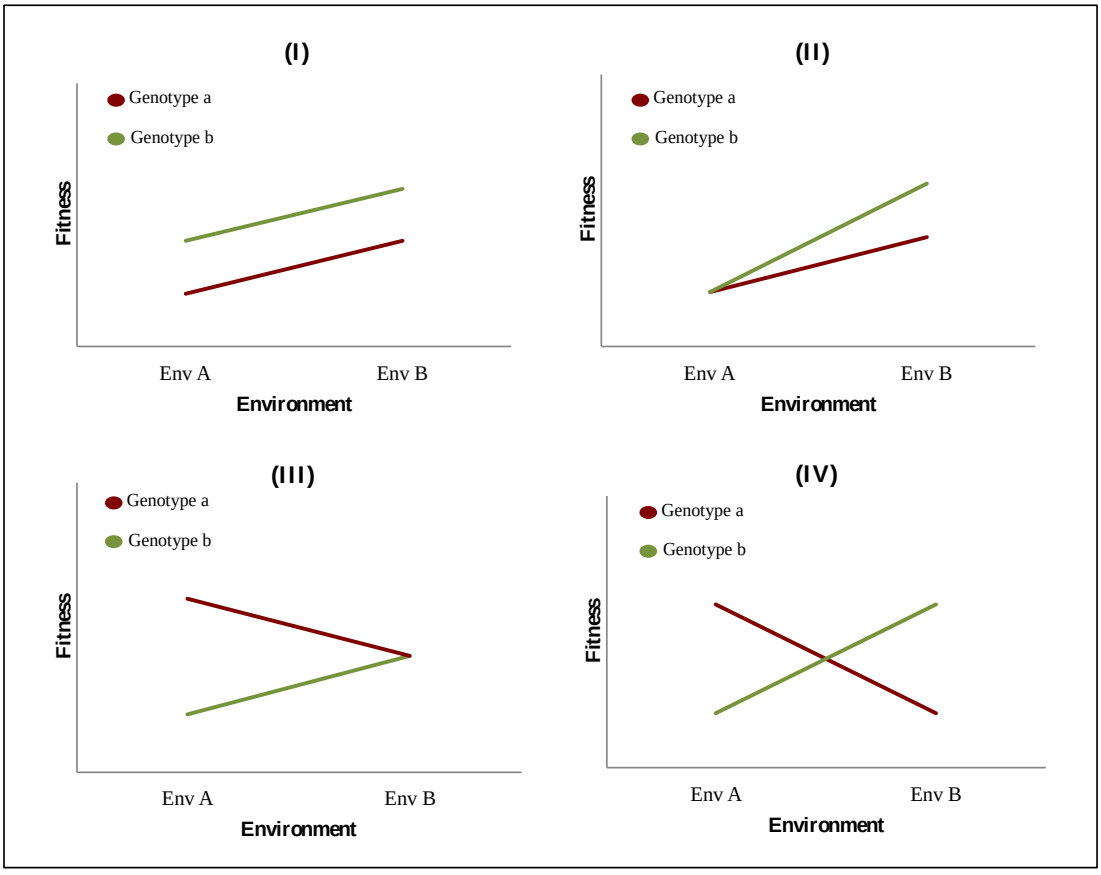


Figure 2 : Genotype and environment interaction (GEI). Four possible reaction norms are provided, the red line indicate the genotype a while the green line indicate the genotype b. in (I) the genotype b has higher fitness than genotype a in all environments, no (GEI). In (II) both genotypes have the same fitness values in the environment A and have an increased fitness in environment B, but the genotype b shows higher fitness in the environment B. In (III) genotype A shows better fitness in environment A, while both genotypes have the same fitness values in the environment B. In (IV) the reaction norms cross, the genotype a shows better fitness in environment A and the genotype b shows higher fitness in environment B (local adaptation). The three last cases of reaction norms (II, III, IV) suggest a (GEI).

3 Population genetics in anthropogenic landscape

3.1 Evolutionary forces

It is well known that landscape fragmentation affects the size, connectivity and the quality of habitat. It tends to generate small populations, genetically isolated and characterized by low genetic diversity. In other words, landscape fragmentation can have rapid evolutionary consequences (Hoffmann & Hercus, 2000). In theory, species are expected to disappear wherever environmental conditions became unsuitable and to colonize every area that might have become more suitable. Colonization success depends on the total number of introduced individuals and on both demographic (demographic and environmental stochasticity) and genetic processes (genetic drift, mutation, selection, gene flow and adaptation) (Fauvergue *et al.*, 2012).

Demographic stochasticity refers to the deviation between the observed and the expected population size estimated using growth rates such as birth rates, survival, fecundity and sex ratio. The reduction of population density is usually followed by a decrease in population growth rate, which increases the risk of extinction (Allee effect) (Allee *et al.*, 1949). Environmental stochasticity indicates the temporal variations in the habitat conditions caused by anthropic activities and environmental events such as floods, fires and droughts. Both demographic and environmental stochasticity can increase the extinction risk in a small population because of the reduction in population growth rate (Figure 3).

Changes in genetic diversity within populations and genetic differentiation among populations may result from the respective action of at least four evolutionary forces (genetic drift and inbreeding, mutation, migration, and selection) (Straalen & Timmermans, 2002) (Figure 2) which interplay with each other to change the structure of genetic diversity. In the next paragraphs, simple common definitions are given for each evolutionary force.

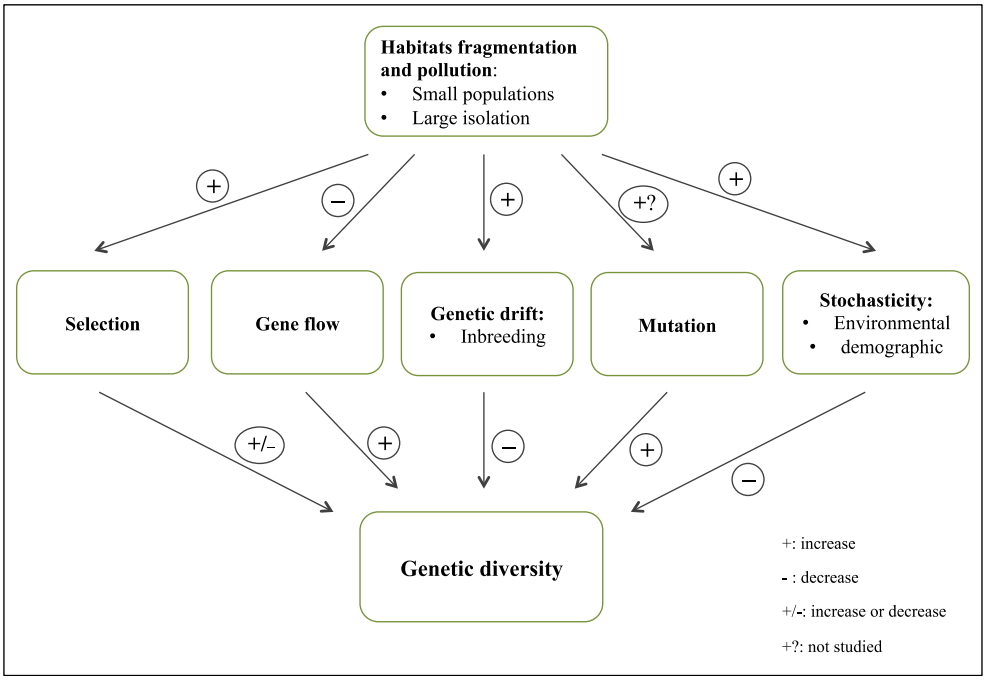


Figure 3: Summary of the factors that shape genetic diversity of populations. Note that habitat fragmentation has no expected effect on mutation.

3.1.1 Mutation

The term mutation refers to any change or alteration in the genomic material. DNA in living cells is continuously exposed to different types of pressure that could lead to the appearance of mutations. Pressure can be 'exogenous' attacks such as radiations or genotoxic compounds or 'endogenous' such as DNA replication or recombination error accidents. The chromosomal mutations may be due either to a change in chromosome number, or to a change in their structure (deletion, duplication, inversion, translocation). For example, a deletion removes some parts of the genome, while the inversion mutation causes a change in the order of certain sequences in the genome. Point mutations are sequence variation at the nucleotidic level, often of a single nucleotide leading to "Single Nucleotide Polymorphism" (SNP). Finally, the length mutation corresponds to a variation of the number of repeating units of a DNA strand, as we can observe in microsatellites (SSR for "Simple Sequence Repeat"). To study the genetic diversity in populations, molecular Microsatellites (SSR Simple Sequence Repeat) and SNP (Single Nucleotide Polymorphism) are frequently used as molecular markers because they usually show high level of variation. In this thesis,

microsatellite markers were used for population genetics study and SNPs were used for quantitative genetic part.

Some mutations are neutral even when they occur in the coding region of a gene. For example, substitution of a nucleotide may occur but without any consequence of the amino acid sequence of the coded protein. This kind of mutations has no evolutionary impact because it does not influence the fitness of individuals and thus, is not a subject for selection. However, mutations in coding sequence of a protein inducing functional changes and mutations in a non-coding sequence implied in gene expression regulation or genome organization may have an evolutionary impact because they are subject to selection pressure (Wang & Malcolm, 1999). Mutations are fundamental for evolution because they are a source of genetic diversity between individuals by generating new alleles and new haplotypes (*i.e.* the pattern of genes and alleles located on the same chromosome that segregate together from one generation to another). Mutation rate is expected to increase in metal polluted environment (Słomka *et al.*, 2011).

3.1.2 Migration

Usually, populations from the same species are geographically isolated (*i.e.* they can not exchange gametes across a landscape) by distance or geographical barriers such as roads, rivers, mountains…but they can stay connected thanks to their capacity to disperse and to move through these barriers (Hamrick *et al.*, 1993). In population genetics, the term migration refers to the exchange of genes (gene flow) due to the migration of fertile individuals or their gametes between populations. Migration tends to homogenize allele frequencies among populations. However, it can reduce the risk of inbreeding and increase genetic diversity within populations which is necessary for future adaptation (Hendry *et al.*, 2011).

Genetic studies can address the impact of fragmentation on the movements indirectly. They show that populations in fragmented landscapes are generally more genetically isolated than in unfragmented landscapes (Bergl & Vigilant, 2007; Radespiel & Bruford, 2013) (Figure 3). Increased genetic structure is partly the result of lower gene flow between populations (Coulon *et al.*, 2004).

3.1.3 Genetic drift

Genetic drift is related to demographic stochasticity and refers to a random change in allele frequencies over generations in a population of finite size (Figure 4) (Wright, 1930). In fact, in a population of finite size, as a result of chance, alleles frequencies vary from one generation to another. The effect of genetic drift increases with the reduction of population size because differences of allele frequencies are expected to be more visible in the next generations. Nei *et al* (1975) show how bottlenecks can be accompanied by strong genetic drift and cause the reduction of genetic diversity owing to i) the loss of the rare alleles ii) the decrease of the number of alleles per locus and iv) the change of the heterozygosity rate iiii) the accumulation of deleterious alleles. Drift can lead to the elimination or the fixing of an allele in the population. The chance of fixing or removal of an allele depends on its frequency in the population. Common alleles are more likely to be fixed while rare alleles have more chance to be eliminated (except rare alleles conferring a selective advantage). Furthermore, genetic drift affects both neutral and selected alleles that are subject to selection in a population. In contrast to mutation and migration, genetic drift tends to reduce the level of genetic diversity. The loss of genetic diversity by genetic drift can be recuperated by either new mutations or by the contribution of new variation by gene flow in the absence of selection.

Molecular methods that permit the estimation of genetic drift depend on the neutral genetic diversity that refers to the polymorphism in non-coding regions of the genome. These regions have *a priori* no impact on the fitness of individuals. Neutral markers are the most used in these studies considering that they are only affected by random genetic drift (ex: Microsatellite markers). Analysis of alleles and genotypes frequencies allow the estimation of genetic diversity indices and genetic parameter of populations such as the expected heterozygosity in the Hardy-Weinberg equilibrium, genetic diversity, allelic richness, the inbreeding coefficient (Fis), the effective population size (Ne) or the genetic population differentiation (Fst). Analysis of the genetic structure is typically based on the estimation of F Wright indices (Fst, Fit, Fis) from a decomposition of the variance in allele frequencies in inter-sample components, intra-sample inter-individual and intra-individual (Weir & Cockerham, 1984). These statistical analyses provide the possibility to detect fragmentation effects through the increased genetic differentiation among populations because of the effect

of genetic drift. Many studies documented the evolutionary impact of pollutants by increasing the effect of genetic drift (Barata *et al.*, 2002; Medina *et al.*, 2007).

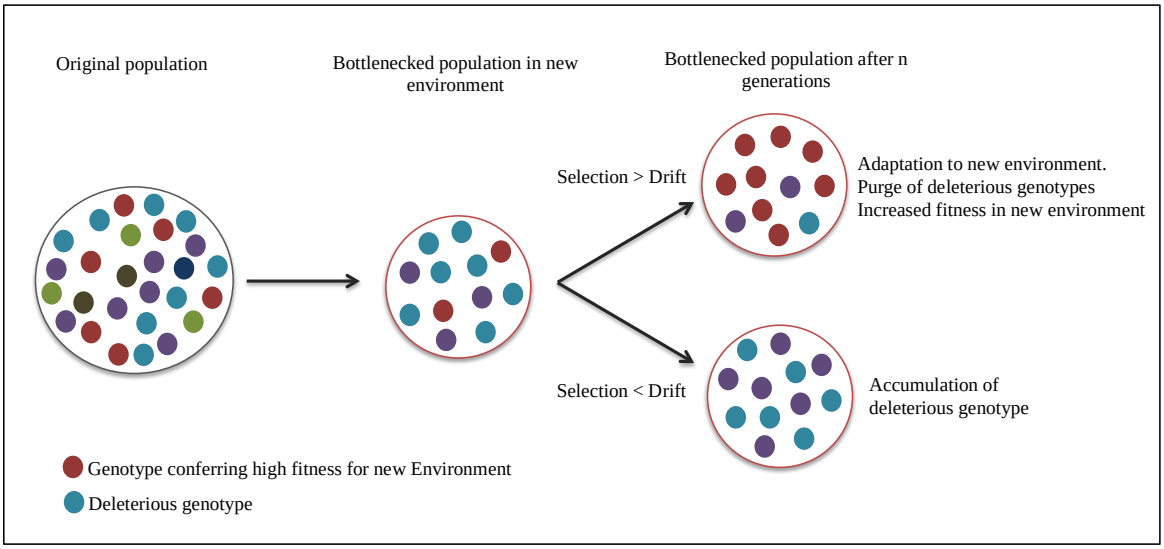


Figure 4 : Summery of the impacts of evolutionary forces in small-introduced populations. Genetic drift and selection cause modification in the structure of genetic diversity of introduced populations in new environments. Two main genotypes are provided in this figure, the red confers high fitness in the new environment and the blue is a deleterious genotype in the new environment.

3.1.4 Natural selection

Natural selection was defined by Darwin (1859) as a mechanism contributing to the evolution of species. It works by sorting the individuals within a population depending on their capacity to survive and reproduce and contribute in the shaping of genetic variation among populations within species. Process of natural selection requires four elements to take place:

- Variation: Genetic diversity within a population is very important to ensure the maintenance of a sufficient genetic resource that allows a best response to selection (Frankham, 2003, 2005).
- Inheritance: traits under natural selection should be heritable *i.e.* able to pass on from parent to offspring.
- Differential survival and reproduction: Individuals possessing the best selective values will survive and contribute more to the next generation by producing larger offspring. For example: if a population have two alleles B and b for a given gene, selection may occur if there is a difference in fitness for the three possible genotypes BB, Bb, bb.

Different modes of natural selection can be distinguished (Nielsen, 2005) (Figure 5):

- Directional selection: it favours one extreme phenotype of continuous variation and tends to eliminate variation within populations. Directional selection can be negative (purifying) or positive. Purifying selection removes deleterious mutation. While, positive selection favours advantageous mutations. Positive selection is the most interesting for evolutionary biology because of its relation with adaptation and the evolution of novel traits and functions.
- Stabilizing selection: tends to increase the intermediate phenotype of a continuous variation, which will be more represented in the next generation. This selection contributes to the decrease of genetic diversity within a population when the population stabilizes on a specific phenotype.
- Disruptive (syn. diversifying or divergent) selection: a type of selection, which happens when two or more phenotypes are favoured. This selection tends to increase the variability within populations since it favours both extremes phenotypes.

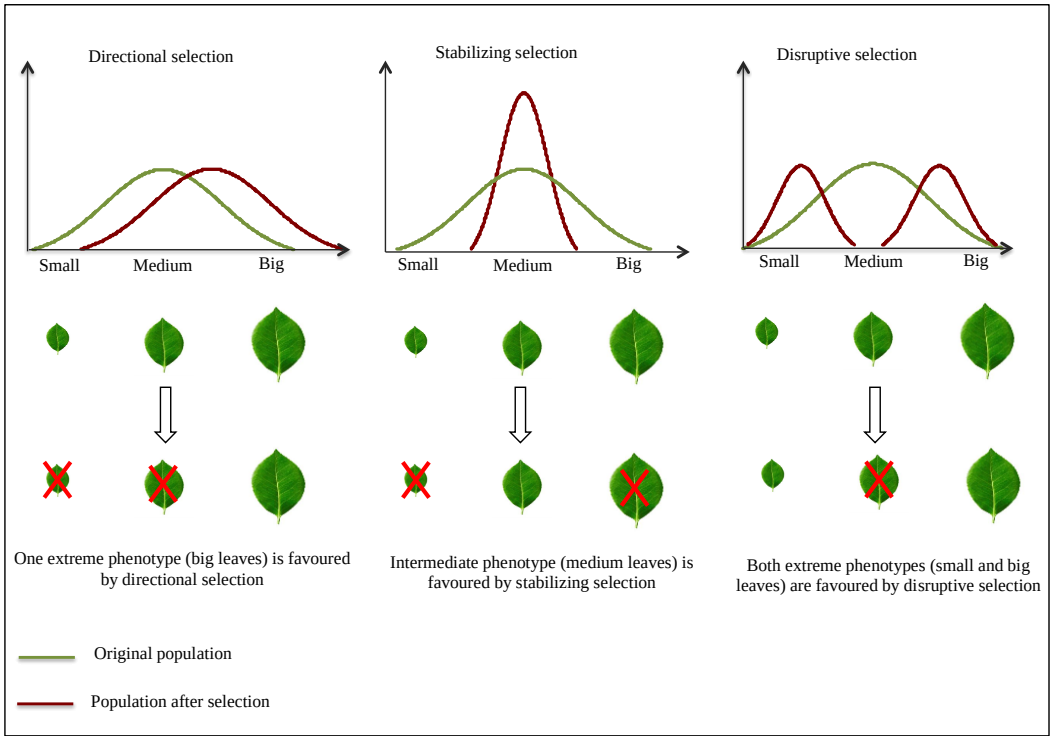


Figure 5: The three types of natural selection (Directional, Stabilizing, Disruptive) based on leaves size.

3.2 Local Adaptation to anthropogenic habitats

Local adaptation depends on the balance among the major evolutionary forces. It depends especially on the balance between gene flow and natural selection (Yeaman & Otto, 2011; Blanquart *et al.*, 2013). Selection tends to differentiate populations living in different environments, whereas gene flow contributes to homogenize genetic composition (allele frequencies) of populations (Sambatti & Rice, 2006; Savolainen *et al.*, 2007). Therefore, gene flow may represent an obstacle to local adaptation. It can hinder or slow down local adaptation by preventing genetic differentiation between stressed and unstressed populations. This effect is more visible in the case of important gene flow (Figure 6).

In particular cases, gene flow can be positive and enhance potential adaptation when it increases the fitness by bringing a new genetic variation more favourable for the new environmental conditions. However, gene flow can also be negative and prevent adaptation when it decreases fitness by bringing individuals with low fitness (outcrossing depression) (Garant *et al.*, 2007; Aitken *et al.*, 2008; Hendry *et al.*, 2011).

Inbreeding between relatives in small populations also represents another threat that can reduce genetic diversity (Charlesworth & Willis, 2009). Even though mating occurs in a panmictic way, consanguineous crosses are possible. Inbreeding increases with the relatedness in a small-colonized population (Glémin, 2003) and reduces the rate of heterozygosity through the genome faster than does genetic drift (inbreeding depression in the case of deleterious recessive alleles in homozygote states). Both genetic drift and inbreeding will reduce the genetic diversity in the recolonizing population.

In a population with high rate of inbreeding, deleterious mutations will accumulate in homozygote state after few generations. If selection pressure is strong enough, it will accelerate the elimination of these mutations *(i.e.* Individuals harbouring these mutations either die or do not reproduce) this mechanism refers to purging of the genetic load (Barrett & Schluter, 2008). However, in a very small population, the effect of genetic drift will be greater than selection and deleterious mutations will be present in high frequency and thus hinder local adaptation to occur (Figure 4).

The complex interaction between gene flow and local adaptation is changed in fragmented landscape. The existence of habitat patches with various environmental conditions favours the existence of local and divergent selection pressures (Fischer & Lindenmayer, 2007). Moreover, fragmentation increases the genetic isolation of populations, which favour local

adaptation in large populations (Lopez *et al.*, 2009). However, this isolation greatly reduces the fitness of small populations in heterogeneous landscapes. The effect of fragmentation on local adaptation is so hard to predict and empirical studies are needed to better understand this phenomenon. The study of local adaptation in fragmented landscape allows being at the center of interactions between gene flow, landscape fragmentation and local adaptation.

Adaptation to environmental modifications may occur through standing genetic variation (the presence of several alleles for one locus in a population) or new mutations. Usually, adaptation seems to be faster from the standing variation than from new mutations because Advantageous alleles are instantly accessible in high frequencies contrarily to beneficial mutations. Beneficial alleles have more chance to be fixed when they exist in multiple copies in the population (standing variation) than if they result from a single new mutation that may be lost through genetic drift. Moreover, since advantageous alleles present as standing variation are older than new mutations, they may have been submitted to previous selective pressures. These alleles might have acquired multiple advantageous genetic changes and may have reached high frequencies (pre-adaptation) (Barrett & Schluter, 2008). The high magnitude of the beneficial effect of an advantageous allele and the high effective size of populations can increase the probability of fixation (Barrett & Schluter, 2008).

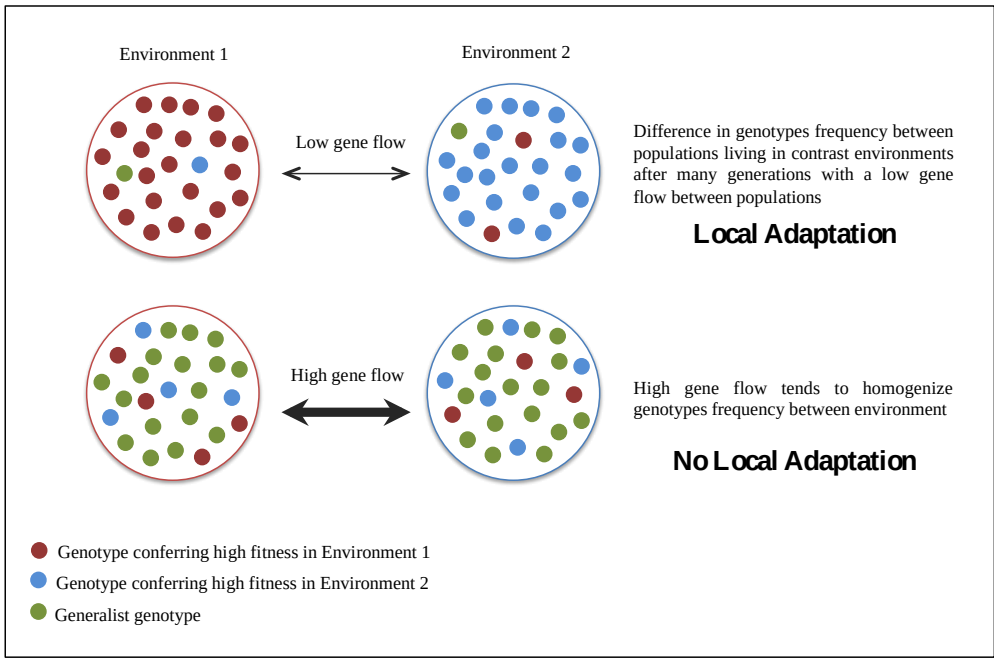


Figure 6 : How gene flow affects local adaptation. In (A), low gene flow between two contrasted environments accompanied by selection pressure facilitates local adaptation. In (B), high gene flow

tends to homogenize the genetic composition of the two contrasted environments that hinder local adaptation to happen.

4 Genetic basis of a quantitative trait

The multiple mechanisms and processes affecting local adaptation make its genetic basis complex, different in origins and remain little clear so far. Most characters involved in local adaptation may be polygenic, quantitative, associated with unknown genes or pleiotropic effects (*i.e.* a gene influences two or more unrelated phenotypic traits). Some approaches have been used to exceed the difficulty in defining some of these cases such as quantitative genetic approach. Thus, some variants having pleiotropic effects and playing adaptive role have been identified, for example, the common mutation EDAR 370A in human populations in east Asia which increases hair thickness, the morphology of the teeth and increases the number of sweat glands (Fujimoto *et al.*, 2008; Kimura *et al.*, 2009; Park *et al.*, 2012). Actually, despite the recent technology that allows the identification of candidate genes involved in adaptive mechanisms, understanding the genetics of quantitative traits requires the identification of the function linking genotypes to phenotypes, which would affect the adaptive value. This identification requires complex experiments in an attempt to identify phenotypic changes associated to a variant. Consequently, most adaptive variants previously characterized have strong and measurable phenotypic impacts. For example, genes involved in the tolerance and hyperaccumulation of heavy metals by plants (Macnair, 1993), or those involved in the tolerance of lactose in adulthood in humans (Bersaglieri *et al.*, 2004).
The study of genetic basis of local adaptation thus requires the theoretical background of quantitative genetics.

4.1 Generalities

The main goal of quantitative genetics approach is to understand the response of quantitative traits to selection pressure (environment), to study their inheritance in a segregating population and to describe genetic architecture of quantitative trait. A quantitative trait is a trait with a continuous variation within a population. Molecular basis is often complex, usually multigenic, each gene segregating in Mendelian way. Thus, the observed phenotype (P) is the result of genotype (G), the environment (E) and their interaction (I= G×E) (El-Soda *et al.*, 2014):

$$\mathbf{P = G + E + I}$$

Most studied traits and especially those involved in the adaptive response are quantitative. The main interest of quantitative genetics is to allow the decomposition of the phenotypic variability of a measured trait in a population or a group of individuals into genetic component (genetic variance VG) and environmental component (environmental variance VE), without knowing the genes involved in this trait. The genetic variance can be decomposed to additive variation (VA refers to the cumulative effect of individual loci) and dominance (VD refers to the interaction between alleles). Genetic variance is in theory directly connected to the polymorphism of the genes encoding the trait.

Broad sense heritability determine all potential sources of genetic variation (additive, dominance, epistatic (genes interaction), maternal and paternal effects): $h^2 = VG/VP$

The measurement of narrow heritability (h^2) of a trait within a population allows to determine the phenotype transmitted by additive genes effects: $h^2 = VA /VP$. This calculation is important to predict how a trait will act to face selection.

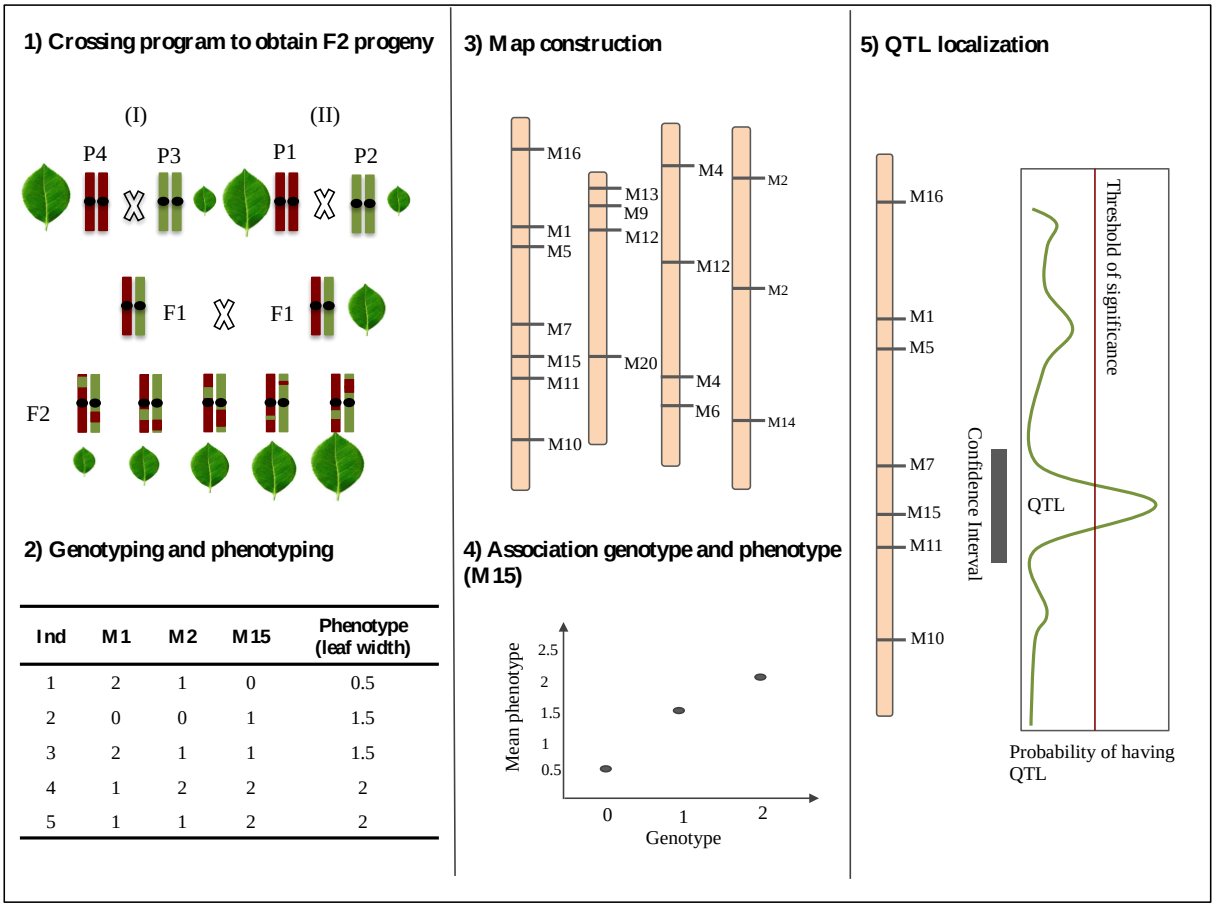
4.2 Genetic architecture for adaptive traits

A quantitative trait locus (QTL) refers to a genomic region containing one or many genes that influence the variation in a quantitative trait. This region can be determined by its linkage to polymorphic markers loci, which segregate in Mendelian way. QTL can be mapped either in natural populations (association mapping approach or linkage disequilibrium mapping) or in the segregating offspring where the relativeness among individuals is known (QTL mapping approach). Both approaches aim at uncovering genotypic variation that can explain phenotypic variations among individuals for a quantitative trait (Mackay *et al.*, 2009). Three elements are necessary for QTL Mapping analyses: i) segregating progeny, ii) molecular markers and genotyping data for the segregating progeny in order to build a genetic map iii) phenotyping data for the same progeny.

For QTL mapping approach, several types of crosses can help for construction of genetic map including F2 generation. Producing F2 progeny starts by crossing two parental lines, which present contrasted phenotypes for the studied trait to form an F1 generation. Then, it is possible to cross a F1 individual on itself or with another F1 resulting from a second parental cross to create the F2 population (Figure 7). This progeny is a necessary component to realize the genetic map through genotyping data for chosen polymorphic markers. This crossing programme provides the possibility to mix randomly by recombination all parental genes

coding for studied character, by this way, new allelic combinations will be obtained exhibiting various quantitative effects.

To construct a genetic map, a high number of molecular markers is needed to provide the opportunity to cover approximately all the genome. The distance between two markers on the linkage group (considered as a chromosome) is expressed as recombination rate between these markers. When the recombination rate is lower than 50%, the markers are related and may belong to the same linkage group.



Ind	M1	M2	M15	Phenotype (leaf width)
1	2	1	0	0.5
2	0	0	1	1.5
3	2	1	1	1.5
4	1	2	2	2
5	1	1	2	2

Figure 7: Major steps in QTL mapping approach: First step consists in obtaining a F2 progeny (1). The second step shows both genotyping data with two markers M1, M2 (that should distinguish two parental strains, 0 and 2 indicate homozygous genotypes and 1 indicate heterozygous genotype) and their phenotypes data for a quantitative trait (e.g. leaf width)(2). Phenotypes and genotypes measure are used in the third step for genetic map construction (3). The last step includes the identification of genomic region involved in the phenotypic variation for this quantitative trait. When there is a difference in the mean of phenotypic trait for the three genotypes classes for a marker, then this marker is linked to QTL (4).

The second step in the realization of a QTL analysis is the assessment for the phenotypic trait of interest. This is what is commonly called phenotyping. Like molecular markers, phenotype is only useful if it is variable among individuals of the mapping population.
Association between phenotypic and genotypic variation among F2 progeny will help to localize chromosomal regions involved in this trait (QTLs) (Mackay *et al.*, 2009) (Figure 7). When a QTL is detected, the identification of genes affecting the phenotype variation needs more detailed analysis of the chromosomal region ex: by fine mapping. QTL analyses can also help to highlight genes with minor effect.

5 Case study: adaptation to metals

Among the toxic substances released into the environment by human activities, metals are present in high quantity in the environment and thus require special attention because the short or long exposure to metallic pollution is known to cause harmful effects on living systems (terrestrial and aquatic ecosystem). These effects are difficult to quantify as they occur with varying levels of vulnerability. Both fauna and flora are affected by the presence of metals in their environment. Metals accumulate in living organisms and disrupt biological mechanisms.

5.1 Metals in the environment

Metals like cadmium (Cd), zinc (Zn), lead (Pb), arsenic (As), nickel (Ni) and copper (Cu) are inorganic substances, non-degradable by microorganisms. The term “heavy metals” has been used in the past especially for non-essential metals such as Pb and Cd, which are characterized by a high density greater than 5 g/cm^3. Therefore, actually, the term “metallic trace elements” (MTE) seems more appropriate. The most toxic are cadmium, arsenic, lead and mercury. MTEs exist naturally in all ecosystems and in all living organisms in varying concentrations following environments and organisms.
In the earth crust, heavy metals are present in the form of minerals, from which they can be mobilized by natural events such as erosion or volcanic eruptions, but also by human activities. The latter is due to physical releases related to metals and mining activities and to the releases of everyday life products such as batteries. Air emissions are also a major source of metallic pollution. While most organic molecules may be degraded, heavy metals can not and their concentration increases steadily in the soil and water.

Metals can be divided into two categories: macro elements or essential are necessary for a normal performance of organisms like calcium (Ca), magnesium (Mg) and potassium (K), and micro elements including iron (Fe), copper (Cu), zinc (Zn), nickel (Ni)…are essential for cellular functions but in very small quantity. Some of these metals are involved in molecular processes such as the control of gene expression, protein biosynthesis, growth substances and chlorophyll. Some others metals are toxic for organisms (like Pb, Hg, Cd, ...). All metals may, from a threshold concentration, induce toxicity in organisms.

Metals have been and are still extensively used in the global economy. Thus, ore extraction and the use of raw materials such as fossil fuels are common sources of contamination. Nowadays, human activities have led to increase metallic contamination. Main polluting sources are:

- Agriculture through the massive use of fertilizers, pesticides, land application of sewage sludge.
- Industry mainly because of dust and smoke emissions, discharges of gas and of liquid effluents.
- Urbanization and its discharges of domestic waste, heavy traffic and massive use of fossil energy.

Two classes of metallic pollution can be distinguished according to origin:

- Ancient pollution: due to natural sources of these component such as geological events, rocks erosion, volcanoes and the biological decomposition that explain the presence of metals in some regions named metalliferous sites, where organisms have been exposed to high concentrations of metals for many years.
- Recent pollution: owing to anthropic activities that usually release a great amount of MTEs in aquatic and/or terrestrial ecosystems.

The presence of MTEs in aquatic or terrestrial ecosystems also generates dangerous consequences for the majority of plants and animals directly exposed to metals. However, some species develop effective means of protection against metals.

The sites that are highly polluted by MTEs (mainly Zn, Pb, Cd) resulting from anthropic activities are called calamine sites. Generally, calamine sites are characterized by drought because of the loss of vegetation (Macnair, 1987), the deficiency of some nutriments (Macnair, 1987; Becker & Dierschke, 2008), and spatial heterogeneity. These characteristics make the presence of metals in the environment as an important engine of natural selection

by reducing the contribution of susceptible genotypes to the next generation that either die or not reproduce. Thereby, natural selection of metals maximize fitness, thus its action results in increased tolerance of resident individuals to their environment (Posthuma & Van Straalen, 1993). Thus, the presence of metals is a special case, which can cause a divergence between populations in the so-called reference sites that is not affected by contamination, and populations in contaminated sites.

5.2 Adaptation to metals

Plant species able to reproduce and to grow on highly polluted sites named metallophyte species, are divided into two groups (Antonovics *et al.*, 1971):

- Absolute metallophyte: exist only in contaminated sites such as *Viola lutea* subsp. *calaminaria* (Violaceae)
- Pseudo-metallophyte: grow in both polluted (metallicolous M) and non-polluted (non-metallicolous NM) sites such as *Noccaea caerulescens* and *Arabidopsis halleri* (Brassicaceae). These species were considered as model species for local adaptation studies during past years.

5.2.1 Metal tolerance and hyperaccumulation in plants

Metal tolerance is defined as the ability of a species to grow and reproduce on soil polluted by heavy metals, on which most species could not survive because of soil toxicity (for example; negative effects on cellular processes such as inhibiting certain enzyme, reducing cellular water, inducing oxidative stress, decreasing the absorption of essential nutrients,...) (Antonovics *et al.*, 1971). From a physiological point of view, heavy metal tolerance is the set of homeostatic mechanisms that maintains cellular concentrations of MTEs in non-toxic conditions to minimize deleterious effects (Clemens *et al.*, 2002). Thus, only species that have developed MTEs tolerance mechanisms can survive on metalliferous sites such as metallophyte and pseudometallophyte species. We can distinguish between two mechanisms of heavy metal tolerance depending on the ratio of metal concentration in the aerial parts and roots (Krämer, 2010):

- The exclusion (shoot : root < 1): is a strategy in plants that refers to the case where the metal does not reach the aerial part of plant; also corresponds to an avoidance strategy (Baker, 1981; Baker & Walker, 1989). Plants using this strategy such as *Mimulus guttatus* (Scrophulariaceae, tolerant to Cu; (Macnair, 1981) and *Rumex acetosa*

(Polygonaceae, tolerant to Al; (Tolrà *et al.*, 2005) have physiological mechanisms for the storage of metals in root cells which prevent the transport of metal to aerial parts.

- The accumulation (shoot : root > 1): consists of MTEs storage in vacuoles of the aerial parts of the plants. This strategy requires the implementation of mechanisms of absorption, translocation and sequestration of metals. Some of these species accumulating MTEs have very high concentrations of MTEs in the aerial parts: these are called hyperaccumulating species of metals. *Arabidopsis halleri* (Zn and Cd; (Bert *et al.*, 2000; Kostecka, 2009) and *Noccaea caerulescens* (Zn, Cd and Ni; (Baker *et al.*, 1994), for example, are two species of Brassicaceae hyperaccumulators (Figure 8).

MTEs tolerance can be detected by measuring several morphological and physiological characters affected by the selection pressure of MTEs (Tardieu & Tuberosa, 2010). The first studies focused on root growth because the root system is the first organ in direct contact with the soil contaminated by MTEs. Three methods were proposed to measure root growth:

1) Firstly: comparing the length of roots of the same individuals (clones) developed in two different concentrations of metal. The ratio of root growth values obtained in the two concentrations is called "tolerance index" (IT). This method is used in this thesis.
2) Secondly: measuring the ability to produce new roots and the root growth (qualitative measure) in order to overcome potential biases owing to intrinsic variations of growth of the roots of the tested populations (Macnair, 1983).
3) Thirdly: determining the smallest metal concentration that inhibits root growth, called the EC100 to "effective concentration 100%" (Schat & Bookum, 1992). The plants are exposed to increasing concentrations of metals (sequential test). This method is based on a measure of presence or absence of new roots in a given concentration.

Other morphological and physiological measures (leaf width, dry biomass, photosynthetic yield) are possible to estimate the MTEs tolerance. Such measures constitute signs of toxicity in the aerial parts induced by high concentrations of metal. Recently, Meyer *et al* (2010) applied a new protocol based on (Wilkins, 1978) to study the tolerance of *Arabidopsis halleri* populations. In this test, plants are grown in hydroponic (controlled conditions) solutions of two fixed and different zinc concentrations for 4 weeks and the morphological and physiological measurements are performed twice after two weeks. The values obtained for the measured traits are generally higher for metallicolous populations. Photosynthetic yield seems to be the most informative trait related to aerial part toxicity (Assunção *et al.*, 2003).

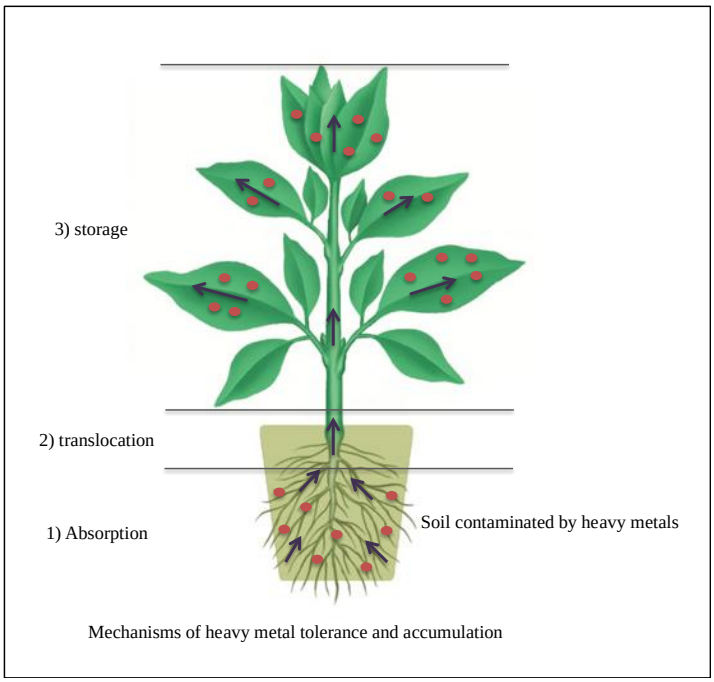


Figure 8 : Mechanisms of heavy metal tolerance and accumulation

Heavy metal hyperaccumulation refers to the ability of a plant to accumulate some metal in the aerial parts. This characteristic could be used in the phytoremediation (cleaning the soil from metal by cultivation hyperaccumulator species that extract and store MTEs) (Hassan & Aarts, 2011) and biofortification (ability to accumulate nutrition minerals by increasing their quantity in plant tissues to avoid some nutritional deficiencies) (Waters & Sankaran, 2011).

Hyperaccumulation mechanism imply three main steps (Verbruggen *et al.*, 2009; Krämer, 2010; Maestri *et al.*, 2010): i) metal absorption by roots which involve many systems, for example: Zn uptake by roots result from an overexpression of ZIP family (zinc-regulated transporter, iron-regulated transporter protein), ii) metal translocation from roots to aerial parts. P ATPase pump controlled by HMA4 gene (Heavy Metal ATP-ase 4) seems to play an essential role in the transport of Zinc and Cadmium (Willems *et al.*, 2007; Courbot *et al.*, 2007; Hanikenne *et al.*, 2008), iii) metal storage MTP (Metal Transporter Protein) allow the transport of metals from the cytoplasm into the organelles and apoplasm, Ca^{2+} appear to have a role in transporting metal to the vacuoles (Figure 9).

We can measure hyperaccumulation by estimating MTEs concentration in the aerial parts of plants *in situ* or cultivated in controlled conditions (Bert *et al.*, 2000, 2003). A plant is

considered as hyperaccumulator under natural conditions when the concentrations in aerial parts are more than 10 µg.g-1 for mercury, 100 µg.g-1for cadmium, 1000 µg.g-1for lead, cobalt, copper, arsenic, selenium, nickel and 3000 µg.g-1for manganese and zinc (Brooks *et al.*, 1977; Krämer, 2010). Moreover, we can consider a plant as hyperaccumulator in two cases:

- When the concentration of metal in shoots is higher than the metal concentration of roots (the ratio > 1) (Salt & Kraemer, 2000).
- When the concentration of metal in the shoots is higher than the total metal concentration in soil (the ratio > 1) (Baker *et al.*, 1994).

However, when an individual grows on a soil with low MTEs concentration and it contain a high concentration of MTEs in the shoots is considered as hyperaccumulator depending on Baker (1994), but not necessarily according to Salt & Kraemer, (2000). All depends on metal bioavailability (available to be absorbed by plants).

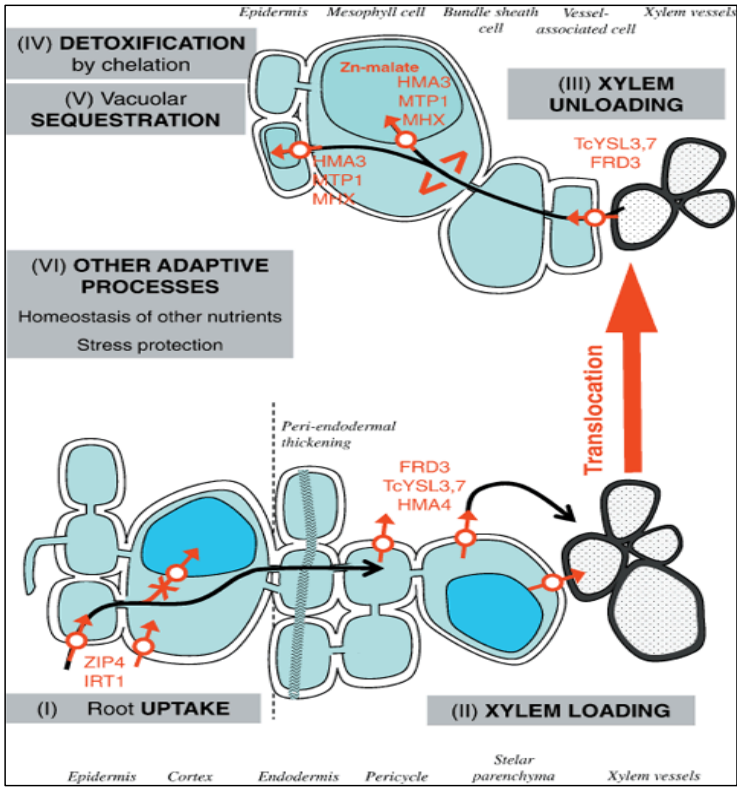


Figure 9 : The main genes involved in the hyper-accumulation of Zn **(**Verbruggen *et al.*, 2009**)**.

5.2.2 Metals tolerance and accumulation in soil macro-invertebrate (Earthworms, annélides oligochètes)

Phenomenon of local adaptation to metals have been mainly studied using plant species but local adaptation can also be considered for animals, especially macro-invertebrate species living in close contact with soil which are able to reproduce and develop in strongly contaminated sites. The annelids oligochaetes of the lumbricidae family are potentially good models to study the tolerance and accumulation of metals. Indeed, the presence and the ability of Lumbricidae to survive and reproduce in heavily contaminated soil raise the question about the mechanisms of tolerance that are implemented in these animals. For example: the earthworms *Lumbricus rubellus* collected from metal contaminated area show higher tolerance to Zn than earthworms originated from uncontaminated sites (Spurgeon & Hopkin, 1999).

Tolerance mechanisms in earthworms is related, for example, to the family of cysteine-rich proteins, metallothionein (MTs), which are involved in the detoxification process of MTEs such as Cd and homeostasis of essential metals such as zinc (Palmiter, 1998; Klaassen *et al.*, 1999) by binding essential and non-essential metals ions. Several works aimed to study the change in the expression of the genes that encode these proteins following exposure to metals. Generally, gene expression of a MT2 coding gene increases after metal exposure in earthworms such as *L. rubellus* (Morgan *et al.*, 2004) and *Eisenia fetida* (Brulle *et al.*, 2006).

Furthermore, earthworms are known for their ability to accumulate metals especially in digestive and dermal tissues higher than do other animal species do (Beyer *et al.*, 1982). However, metals have varying effects depending on the species, stage of development, lifestyle (place of life and diet) and their ability to adapt to pollution. These effects also vary according to the nature of the metal, its chemical form and soil properties.

In order to assess the impact of contaminant exposure such as metals or pesticides on the biological system of earthworms, various biological parameters can be measured in earthworms. The word ‘biomarker’ refers to the relationship between contaminants and a biological response of organisms directly exposed. Biological responses can be observed at molecular, biochemical, cellular, physiological or behavioural level. Earthworms from Lumbricidae family are considered as good models for biomonitoring (Römbke *et al.*, 2005; Sanchez-Hernandez, 2006). Therefore, they have been often used to examine the impact of anthropogenic pollutants, especially in soil contamination context. Experimental studies on

limited number of oligochaetes annelids species (*Eisenia fetida*, *Eisenia andrei*, *Lumbricus rubellus* mainly) showed that exposure to high metal concentrations can influence earthworms. For example: it can affect their growth, sexual development, cocoons production and hatching rate (Spurgeon & Hopkin, 1996; Nahmani & Lavelle, 2002). Moreover, earthworms' physiology and immunity may be affected by contaminant exposure (see Spurgeon *et al.*, 2004). Changes in enzyme activities or gene expression levels have also been reported (Brulle *et al.*, 2008, 2010).

6 Thesis objectives and studied models

6.1 Population genetics of *Lumbricus terrestris*

6.1.1 Biology of Lumbricus terrestris

Earthworms can be divided into three ecological groups depending on their locality in the soil and their feeding behaviour: (1) epigeics: that live in the litter above the soil surface, pigmented earthworms, feed on organic matter in soil surface, (2) anecics, that live in temporary vertical deep burrows and come to the surface to feed, pigmented earthworms, and (3) endogeics, that rarely come to the soil surface, make horizontal non-permanent burrows and feed on the organic matter already in the soil, this type of earthworm is un-pigmented or lightly pigmented (Römbke *et al.*, 2005). Belonging to Lumbricidae family, the species *Lumbricus terrestris* is typically an anecic earthworm. It is one of the largest species among Northern Europe earthworms. It is also one of the easiest to identify based on external characteristics (Bouché, 1972) when exhibiting a clitellum (breeding form), such as the prostomium shape, the shape and the location of the clitellum (saddle-shaped on segments 32-37), the location of the tuberculata pubertatis (segments 33-36) and the pigmented body which is darkest at the anterior part on the dorsal side. In contrast to non-permanent burrows drilled by endogeic earthworms, burrows drilled by *L. terrestris* are semi-permanent and play an important role in the transfer of water, air and nutrients through the soil. Sequencing data of the cytochrome oxidase I (COI) mitochondrial gene revealed the existence of two cryptic species (James *et al.*, 2010). Both species are supposed to live in sympatry, except in the north-west of France, where the exclusive presence of one species could be explained by ecological differentiation. *L. terrestris* is diploid n= 36, outcrossing and hermaphrodite. Their reproduction is bi-parentally by a mutual exchange of sperm between two individuals during mating. Mating behaviours of *L. terrestris* are complex including visits of prospect partner

before copulation and piercing the partner's skin by the setae of some segments, which may affect the uptake of sperm (Nuutinen & Butt, 1997; Koene *et al.*, 2005).

L. terrestris is an attractive sentinel species, which is often used (possibly in combination with other earthworms species) to assess environmental impact of soil contamination, land-use change and fragmentation of habitats. It is considered as a good candidate for biomonitoring and widely used in many studies as bio-indicator (Römbke *et al.*, 2005; Sanchez-Hernandez, 2006). Despite the importance of this species, population genetics of *L. terrestris* is poorly studied. In this study, we aimed to compare the structure of genetic diversity between populations collected from anthropogenic sites using microsatellite markers.

Figure 10 *: L. terrestris* ('Lumbricus terrestris Linnaeus, 1758 -- Discover Life'; Heidi & Hans-Juergen Koch)

6.1.2 Geographical distribution of L. terrestris

Earthworms distribution in Europe has been modified two times in the past. The first changes refer to earthworms extinction caused by the last glaciation in Europe and the recolonization of Northern Europe. The second is due to human activities that led to the transplantation of earthworms from Europe to the other continents. For example, *L. terrestris* invaded North America from introduced European populations. The successful colonization of *L. terrestris* to a new habitat indicates that this species has specific characteristics that make it a successful invader species. Current distribution of this species suggests a passive dispersal of *L. terrestris* and proves that humans have facilitated their distribution through transporting earthworms and/or their cocoons (Enckell *et al.*, 1986; Tiunov *et al.*, 2006). Besides human

activities, there are many factors affecting the distribution of *L. terrestris* such as climate conditions, life history traits and the suitability of habitat.

The present work is based on collected earthworms of *L. terrestris* from different disrupted locations in the Nord-Pas de Calais Region.

6.1.3 Approach

The most commonly used markers to study the polymorphism of DNA are microsatellite markers. They are generally dispersed throughout the genome, tandem repeats segments of DNA composed of very short motif of 1 to 6 nucleotides (for example (CA)n). Microsatellite markers are ideal for examining the phylogeographic and the genetic variation in populations and species because they are highly polymorphic, neutral to avoid the influence of selection pressure, co-dominant (homozygote and heterozygote genotypes) (Varshney *et al.*, 2005; Gao *et al.*, 2013) for many reasons:

- Generally, the maintenance of genetic variability intra-specific is very important for species conservation. Probably, there is a correlation between genetic diversity and population variability. This correlation is complicated to apprehend but we assume that the evolutionary potential of a species or a population is related to its genetic diversity. Polymorphic microsatellite markers are considered as a good estimator of genetic variability within a species.
- Several events of the demographical history of populations (for example: bottleneck, founder effect, gene flux) could be detected using microsatellite markers.

Microsatellite markers are specific and therefore they require a development for each studied species. Moreover, the number of microsatellite markers should be as large as possible in order to provide information representative of the total genetic variability of the studied species. At the beginning of this thesis, the development of new microsatellite markers was necessary given their slight availability for our studied species *L. terrestris*. Finally, we succeeded in the identification of eight polymorphic microsatellite markers for *L. terrestris* (Souleman *et al.*, 2016).

Microsatellite markers developed in the first part of this thesis have been used to study the structure of genetic diversity of *L. terrestris*. The region of study is Nord-Pas de Calais, an industrial and agricultural region in Northern France that is highly affected by anthropogenic activities. After more than a century of heavy metallurgic activities, the soil of several areas

is highly polluted by ETMs especially Zinc, Lead and Cadmium. In addition to metallic pollution, this region is highly fragmented and urbanized. We assumed that habitats fragmentation and metallic pollution could affect the structure of genetic diversity of the organisms that live in close contact with the soil. Since *L. terrestris* is an important component of terrestrial ecosystem, we aimed in this study to compare genetic diversity between disturbed habitats of *L. terrestris* earthworms collected from different sites in Northern France, Nord-Pas de Calais region, and to investigate their spatial structuration.

6.2 Genetic architecture of Zn tolerance and Zn accumulation in *Arabidopsis halleri*

6.2.1 Biology and geographical distribution of A. halleri

Arabidopsis halleri belongs to Brassicaceae family, and is a diploid (2n = 16), perennial and clonal and self-incompatible species (Van Rossum *et al.*, 2004). Leaf morphology, flower color and the level of stolons development are highly variable between three sub species (*A. halleri ovirensis*, *A. halleri halleri*, *A. halleri gemmifera*) (Al-Shehbaz & O'Kane, 2002). *A. halleri* became an important species in evolutionary biology because it shares 94% of coding genomic regions with *Arabidopsis thaliana*, a reference species in plant biology. Molecular studies showed that the separation of *A. thaliana* from the other species of Arabidopsis was about 5.8 million years ago (Koch *et al.*, 2001; Becher *et al.*, 2004). *A. halleri* is also a relative with *Arabidopsis lyrata*, they were separated 330 000 years ago (Roux *et al.*, 2011). It is a wide distributed species, from central Europe to northern Russia (Austria, Croatia, Czech Republic, Germany, Italy, Poland, Romania, Slovakia, Slovenia, Switzerland, and Ukraine). Generally, *A. halleri* grows in humid environment and mid-elevated mountain (between 600-2200 m) with acid and oligotrophic soils (Al-Shehbaz & O'Kane, 2002). *A. halleri* is a pseudometallophte species; it exists on calamine soils (metallicolous populations) and on non-polluted soils (non-metallicolous populations) (Bert *et al.*, 2000). The known metallicolous populations (M) locate in north of France, Silesia in Poland, Harz in Germany and in Northern of Italy. The current study focus on *A. halleri* collected from two sites in Northern Italy.

Figure 11 : *Arabidopsis halleri* in the natural habitat and in greenhouse.

6.2.2 Approach

All previous works aimed to study the genetic architecture of metals tolerance and hyperaccumulation using interspecific crossings between *A. halleri* and its relative *A. lyrata* (non-tolerant species). In the present work genetic architecture of Zn tolerance and Zn hyperaccumulaion was conducted for the first time on intraspecific F2 progeny of *A. halleri*. Two intraspecific crossings between M (metallicolous) and NM (non-metallicolous) parents were performed to generate two F1 generations, then one F2 generation was created by crossing two individuals randomly chosen from the two F1 progenies (Figure 7).

To evaluate Zn tolerance, we applied the protocol of Meyer *et al.*, (2010). F2 plants were grown in nutrient solution under controlled conditions (temperature, humidity, and luminosity) in two fixed concentrations of Zn (polluted and control). Morphological and physiological measures (leaf width, root length, dry biomass, photosynthetic yield) were performed 3 times during 8 weeks to estimate the variation in Zn tolerance among plants (Figure 12) (details of protocol is explained in the chapter V).

However, to evaluate Zn hyperaccumulation, F2 plants were grown on the compost in the greenhouse under controlled conditions. Two concentrations of Zn were used (control and polluted). To estimate the variation of Zn hyperaccumulation among plants, we measured the concentration of Zn in shoots and the dry biomass (Figure 13) (details of protocol is explained in the chapter V). The broad-sense heritabilities (H^2) of Zn tolerance and Zn hyperaccumulation were calculated.

Figure 12 : Hydroponic culture to evaluate Zn tolerance.

Figure 13 : Culture on compost in the greenhouse to evaluate Zn hyperaccumulation

As microsatellite markers, SNPs (single nucleotide polymorphism) markers are highly polymorphic and dispersed throughout the genome. For this reason, SNPs markers were chosen for genetic map construction because they provide the opportunity to cover almost all the genome and segregate in Mendelian way. DNA of the four parental and two F1 were

extracted purified and sequenced. Reads with more than 80 nucleotides were kept, and then paired reads were mapped against *A. halleri* assembly (not published yet). SNPs markers were chosen depending on several criteria to permit the genotyping of F2 and thus for the genetic map construction and the QTL mapping. Finally, 384 SNPs of high quality were selected **(All steps of SNPs markers identification were performed by Marie-Joe Karam).** By the mean of 384 SNPs markers, all F2 plants were genotyped. Data of SNPs genotyping were used to construct the genetic map.

After the construction of the genetic map, QTL mapping approach was used for genotype-phenotype association to localize the genomic regions involved in the variation of Zn tolerance and Zn hyperaccumulation.

7 Thesis objectives:

The objectives of the first part of this thesis are:

- Literature reviewing of knowledge about the general effects of MTEs on earthworms, highlighting the consequences of a long-term exposure to MTEs on their biology, and discussing the validity of exposure biomarkers of MTE. The fact that these sentinel species may have evolved when submitted to long-term exposure is also considered.
- Identification and characterization of new polymorphic microsatellite markers in *Lumbricus terrestris* earthworm.
- Description of the structure of genetic diversity in *L. terrestris* earthworm collected from disrupted sites in Europe, in the Nord-Pas de Calais Region.

The objectives of the second part are:

- Study the genetic architecture of heavy metals tolerance working on F2 offspring of *Arabidopsis halleri.*
- Studying the architecture of heavy metals hyperaccumulation on the same offspring of *A. halleri*.

Figure 14 provides a comprehensive plan of the realized work in each chapter of this thesis in both model species *L. terrestris* and *A. halleri*.

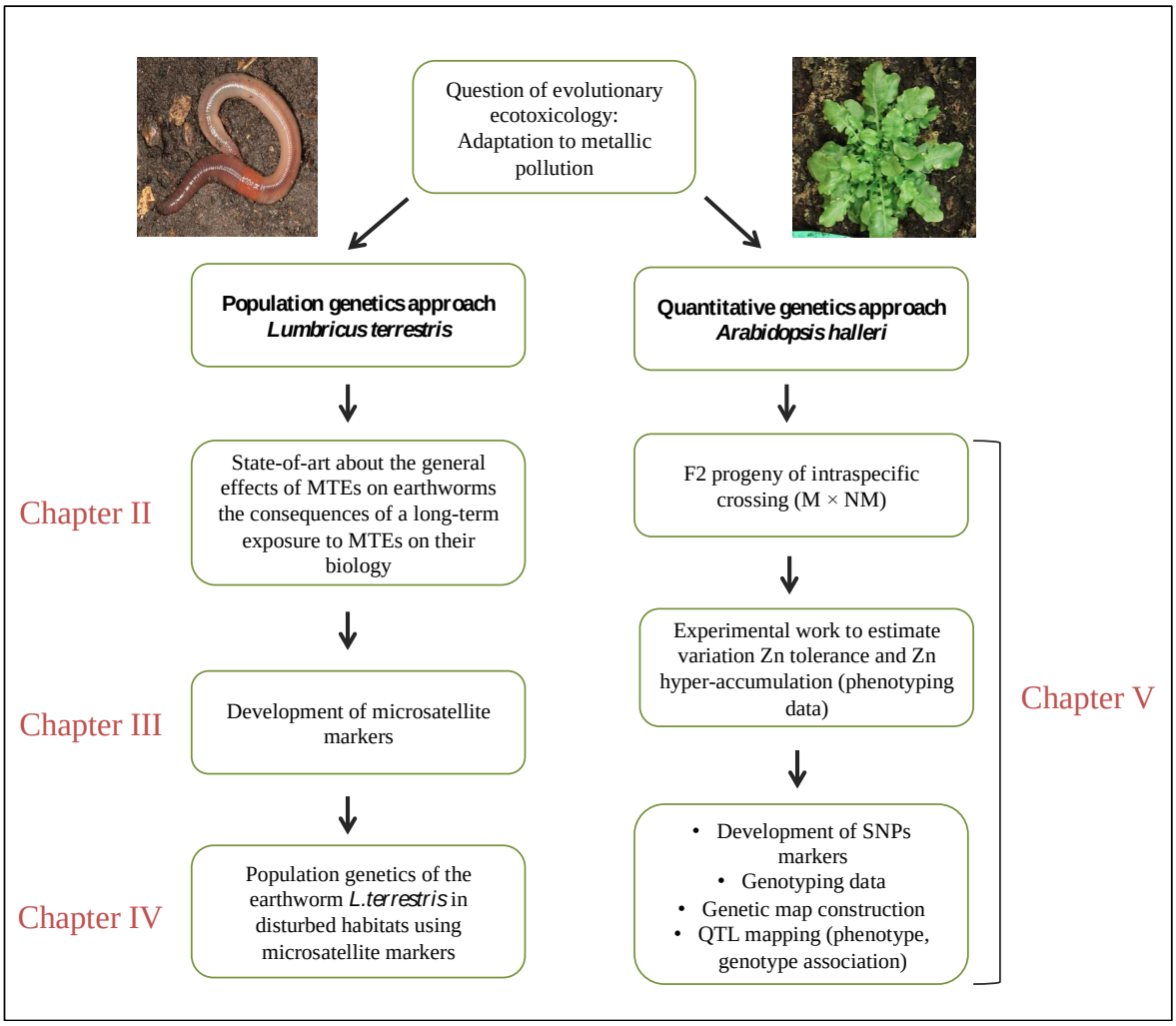


Figure 14 : Plan of thesis chapters

Chapter II

Using biomarkers in an evolutionary context: lessons from the analysis of biological responses of oligochaete annelids to metal exposure

Nowadays, the main causes of environmental pollution come primarily from the production and use of many sources of energy, then from industrial and agricultural activities. The introduction of contaminants such as metals, pesticides polycyclic aromatic hydrocarbons (PAHs) may affect the biota living in close contact with the soil such as earthworms of Lumbricidae. Biological response can vary depending on the intensity and the duration of exposure. This relalationship between biological response and exposure lead to the development of biomarkers in earthworms to pollutant.

In this chapter, we focus on the definition and validity of molecular biomarkers of metal pollution using earthworms of the Lumbricidae family. It appears that conditions that allow local adaptation to occur are present in earthworms (namely, potential strong and durable selective pressures and within-population genetically determined phenotypic variability). In addition, some infra-individual evidence of adaptation do exists. Consequently, there is a need for evolutionary studies investigating the level and distribution of genetic diversity in natural populations of earthworms (chapter 3 and 4). Therefore, it seems necessary to integrate the potential effects of local adaptation to Metallic Trace Elements (MTEs) on biological responses of organisms in the development of biomarkers (Pauwels *et al.*, 2013).

Table of contents for chapter 2

Using biomarkers in an evolutionary context: lessons from the analysis of biological responses of oligochaete annelids to metal exposure

Maxime Pauwels[1], Hélène Frérot[1], Dima Souleman[1,2] and Franck Vandenbulcke[2]

1- Université Lille Nord de France, Laboratoire de Génétique et Evolution des Populations Végétales UMR CNRS 8198, Bâtiment SN2, F-59655 Villeneuve d'Ascq cedex France

2- Université Lille Nord de France, Laboratoire de Génie Civil et géo-Environnement EA 4515 (LGCgE-Lille 1), Ecologie Numérique et Ecotoxicologie, F-59650 Villeneuve d'Ascq, France

Published in: Environmental pollution

DOI number: 10.1016/j.envpol.2013.05.005

Abstract

Anthropogenic activities may lead to the accumulation of inorganic and organic compounds in topsoils. Biota living in close contact with contaminated soils may experience stress at different levels of biological organisation throughout the continuum from molecular to community level. Biological responses observed at the individual or infra-individual level of biological organization led to the development of biomarkers. The development of biomarkers consists often in evidencing biological modifications following a contaminant stress in laboratory conditions, using naïve organisms and it is sometime proposed to use the biological state of individuals from sentinel species collected in the field to evaluate the level of environmental exposure. However, considering the possibility of local adaptation following long-term exposure, organisms' response sampled in the field may substantially differ from laboratory specimens. In this review, we discuss this point focusing on the definition and validity of molecular biomarkers of metal pollution using earthworms of the Lumbricidae family.

Keywords: metal pollution, biomarker, earthworms, local adaptation, gene expression

1 The concept of biomarker in ecotoxicology: definition and limits

Ecotoxicology is a discipline at the interface between ecology and toxicology. Concisely, ecotoxicology is concerned with toxic molecules, hereafter called contaminants, either of artificial origin (including drugs, endocrine disruptors ...) or natural agents (metals, arsenic, …), of which Human beings alters the distribution and / or cycles in the different
Compartments of the biosphere. From a fundamental point of view, ecotoxicology studies the fate and consequences of contaminants on biological systems, from ecosystems to individual organisms. From an applied point of view, it is increasingly asked to ecotoxicologists to develop tools allowing to determine the intensity and duration of contaminations and to assess associated ecological risks, through the prediction of potential effects of contaminant exposure in nature.

One approach to meet this social demand for biomonitoring methods is the development of biomarkers. This approach considers that the best method to detect the biological impact of contaminant exposure is to investigate the effects of contaminants on biological systems. Indeed, compared to traditional methods focusing on physical and chemical properties of soils, biomarkers are assumed to focus on the effects of the bioavailable (*i.e.* transmitted to living organisms) fraction of environmental chemicals and to integrate the putative interactive effects of complex mixtures of chemicals in the Ecological Risk Assessment (ERA). Theoretically, a "biomarker" can be defined from any observable and/or measurable functional response to exposure to one or several contaminants that can be characterized at the sub-individual level of biological organization (molecular, biochemical, cellular, physiological, behavioural). Importantly, the response is assumed to indicate a departure from healthy status that cannot be detected from an intact organism (Weeks, 1995; Gestel & Brummelen, 1996; Ricketts *et al.*, 2004). The concept of biomarker is thus based on the causal relationship between the contamination of environments by any chemical inducing a stress (pesticides, polycyclic aromatic hydrocarbons (PAHs), metals, ...) and biological changes induced by the contaminated environment. In practice, the development of biomarkers can be divided in two steps. It consists, first, in characterizing in laboratory conditions the effects of contaminant exposure on the biology of an organism of interest and, second, in using the characterized response to assess the level and duration of contaminant exposure of organisms sampled in the field (Weeks, 1995).

1.1 Fundamental approaches

Characterizing the effects of contaminant exposure on the biology of an organism is mostly a fundamental issue based on experimental approaches. It requires accumulating fundamental knowledge of the hazardous nature of a substance which is evaluated by toxicity studies (acute or chronic, intrinsic or cocktails ...) and establishing thresholds beyond which a substance has a toxic effect or below which it is harmless. These thresholds usually rely on the physical and chemical properties of the contaminant, the characteristics of the environment, the duration of exposure (chronic, occasional), the route of exposure (percutaneous, ingestion, inhalation ...) and the state of development of exposed individual (species, sex, age...).

In experimental conditions, the comparison of phenotypes of conspecific individuals differentially exposed to one or several contaminants is supposed to reveal the biological response to contaminant(s). In recent years, these ecotoxicological investigations have largely benefited from the emergence of molecular biology techniques, which led to a better understanding of the mechanisms of contaminants action at molecular level. In particular, the advent of transcriptomic tools (real time PCR, subtractive libraries, pyro-sequencing, DNA chips) coupled to bioinfomatic procedures and statistical data analysis, allowed a comprehensive analysis of gene expression in various ecotoxicological contexts (Brulle *et al.*, 2010). The main advantage of gene expression studies is that gene expression profiles represent the first level of integration between environmental stressors and the genome which, through the synthesis of proteins, pilots the response of the organisms to external changes (Brulle et al., 2008, 2010).

Interestingly, model organisms used in ecotoxicological studies are often "naïve", which means that they belong to model species and / or test species that have never been previously exposed to contaminant and do not descent from exposed individuals. In such cases, it seems reasonable to assume that phenotypic responses observed in contaminated conditions in contrast to control conditions may not be accounted for genetic differences among individuals, but rather are environmentally induced responses (i.e. the source of phenotypic variation is mainly environmental). Biomarkers must therefore be considered as indicators for such environmentally induced responses, revealing the ability of phenotypes to instantly adjust to environmental changes. This phenomenon is known as “phenotypic plasticity” in evolutionary biology, mostly called “acclimation” in ecotoxicology. Consequently,

biomarkers must be also considered as early markers of exposure that do not reveal long-term effects of the contaminant on the ecosystem.

1.2 Applied ecotoxicology

In contrast to fundamental approaches exposed above, the application of biomarkers for ERA purposes relies on more technical issues. Two main differences can be highlighted between the experimental accumulation of fundamental knowledge about biological responses to contaminant exposure in laboratory conditions and the practical development of biomarkers.

● Firstly, biomarkers should be used on sentinel species, *i.e.* on wild organisms sampled in natural populations from the field rather than on laboratory specimens (van der Schalie *et al.*, 1999; Beeby, 2001). Working on sentinel species implies that biomarkers may be developed on varying species corresponding to the ecosystem of interest. Considering the ERA of soil pollution in terrestrial ecosystems, it is well admitted that, because they represent important ecological functions of terrestrial ecosystems, species from the soil macrofauna should be considered as potential indicator of soil quality (Stork & Eggleton, 1992; van Straalen, 1998; Marhan & Scheu, 2005). A literature survey allows to distinguish a few species belonging to three taxa living in close contact with the soil: nematodes (Caenorhabditis elegans), Collembola and oligochaete annelids (Brulle *et al.*, 2010). In particular, in accordance with their status of soil ecosystem engineers (Darwin, 1883, cited in (Lavelle & Spain, 2001)), earthworms from the Lumbricidae family are good candidates for biomonitoring (Römbke *et al.*, 2005; Sanchez-Hernandez, 2006). Indeed, earthworms play an important role in water, nutrients and carbon cycles in terrestrial ecosystems and increase soil fertility and earthworm populations can provide information on soil structure, microclimatic conditions, nutritional status and presence of toxic elements in soils (Christensen, 1988; Edwards & Bohlen, 1996; Edwards *et al.*, 1998; Kautenburger, 2006). Therefore, they have been adopted by the international community as sentinel species for the study of the potential environmental impact (ERA) of anthropogenic contaminants such as pesticides, hydrocarbons and Metal trace elements (MTEs) (Edwards & Bohlen, 1996; Edwards *et al.*, 1998; Piearce *et al.*, 2002; Spurgeon *et al.*, 2004; Seeber *et al.*, 2005; Kautenburger, 2006).

● Secondly, using biomarkers in the field generally assumes that the relationship between the contamination of environments and induced biological changes is mutual. This means that the level of contaminant exposure in the environment could be deduced from phenotypic values of field-sampled organisms. This assumption, however, may be controversial, in particular if

characterized individuals have been sampled in natural populations exposed to contaminant over several generations. In the same way, the previous assumption that the main source of phenotypic variation is environmental may be controversial. Indeed, from an evolutionary point of view, phenotypic differences among individuals sampled in natural populations may be genetically determined (*i.e.* heritable), in particular if these populations originate from contrasted environments, and if environmental heterogeneity lasted for a relatively long period of time. In such a case, darwinian selection may have promoted the evolution of local adaptation (Kawecki & Ebert, 2004), resulting in genetically differentiated phenotypes potentially outmatching differences induced by the environment.

In this context, the aims of the present work are (1) to report a literature review of knowledge about the general effects of MTEs on earthworms as main sentinel species, (2) to particularly highlight the consequences of a long-term exposure to MTEs on earthworm biology, and (3) to debate the definition and the validity of biomarkers of MTE exposure considering that sentinel species may have evolved under such exposure.

2 Biological response of earthworms to MTEs exposure in controlled conditions

Among contaminants, MTEs, are of major environmental and human health concerns. Indeed, metals are released in large quantities by human activities (agricultural use of fertilizers, residues from metalliferous mining and smelting industries, etc., (Hopkin, 1989; Bradl, 2005) and can reach very high concentrations in soils. Unlike organic pollutants, metals remain in the environment. Consequently, they can accumulate in food chains through different trophic levels, causing toxicity to living organisms, including humans (Nawrot *et al.*, 2006; Peralta-Videa *et al.*, 2009).

Overall, MTEs soil pollution causes environmental stress leading to dysfunction of ecosystems. MTEs pollution usually reduced both species diversity and species richness (Kozlov & Zvereva, 2011). For example, high levels of MTEs affect soil microbial species causing a decrease in population size or a reduction in the species richness of soil bacterial communities (Moffett *et al.*, 2003). Studies focusing on earthworms (Annelida Oligochaeta) have also shown a decrease in population densities (Pižl & Josens, 1995) and changes in community structure (Lukkari *et al.*, 2004) in response to metal pollution. In general, loss of species richness can results in the loss of important ecosystem functions. MTE pollution is for example commonly associated with an accumulation of litter due to a reduction of the

mineralization that leads to high organic matter content observed in some heavily polluted sites (Cotrufo *et al.*, 1995). At individual and infra individual levels, many ecotoxicological studies conducted in plants and animals living in close contact with contaminated soil led to a better understanding of the biological consequences (MTEs contents in tissues, enzymes activities, genes expression levels) of the exposure of organisms to toxic levels of MTEs. For example, in plants, MTE exposure can cause a significant decrease of the growth rate of trees (Dickinson *et al.*, 1992). Photosynthesis may also be affected through the hydrolysis of chlorophyll (Manios *et al.*, 2003).

Experiments carried out under controlled conditions on a limited number of oligochaetes annelids species (*Eisenia fetida, Eisenia andrei, Lumbricus rubellus* mainly) show that exposure to MTEs may affect their physiology (see(Spurgeon *et al.*, 2004)). For example, disturbances of life history traits, enzyme activities or gene expression levels are observed. In particular, the link between changes of mRNA expression levels (transcriptomics) following exposure to chemical compounds and biochemical or physiological functions has been partially established in a limited number of soil invertebrates (Brulle *et al.*, 2010). In *E. fetida* and *L. rubellus*, the consequence of experimental contaminant exposure on the level of expression of either selected genes, using real-time PCR, or of a large set of genes, using microarrays, has been extensively studied. In addition, most work has been carried out with the aim to describe dose-response changes (i.e. to highlight a relationship between gene expression levels and contamination levels) in worms exposed to stressful concentrations of various contaminants (Cd, fluoranthene and atrazine). Particular attention was paid to genes involved in detoxification, defense, anti-oxidant, DNA repair and metabolism of metal ions (Fe2+, Zn2+, Ca2+) mechanisms. Usually, computer analyzes of the expression profiles confirm the involvement of genes controlling the mechanisms cited before. In this approach, the transcriptome is supposed to reflect the physiological state of the test species and is a relevant way to apprehend an initial stress in a dynamic manner (see (Brulle *et al.*, 2010)).

In most organisms, including earthworms, physiological tolerance to metals depends, at least partially, on the induction of genes encoding metal binding proteins. Among these molecules, metallothioneins (MTs) are cysteine-rich proteins of low molecular weight (6000-8000 Da), characterized by the absence of aromatic amino acids, and known to be involved in detoxification processes of MTEs such as Cd and homeostasis of essential metals such as zinc (Palmiter, 1998; Klaassen *et al.*, 1999). By binding essential and non-essential metal ions, MTs are involved in the homeostatic regulation and detoxification of metals (Klaassen

et al., 1999; Brulle *et al.*, 2006). For example, MTs provide detoxification function of some non-essential metals such as Cd and protection during oxidative shock (Ghoshal *et al.*, 1998; Baird *et al.*, 2006). The physiological function of MTs has led several research teams to consider that they could be involved in the biological response to environmental stress following a contamination by a toxic chemical (Beattie *et al.*, 2005). Accordingly, several research works were dedicated to the estimation of the impact of MTEs exposure based on the expression of MTs (or other proteins conferring resistance to metals) in various soil invertebrates. For instance, gene expression of a MT coding gene is induced by metals in earthworms species such as *E. fetida* (Demuynck *et al.*, 2005, 2006) and *L. rubellus* (Stürzenbaum *et al.*, 2004), but also in the nematode *C. elegans* (Liao & Freedman, 1998) and in the springtail *Orchesella cincta* (Timmermans *et al.*, 2005). Induction of MTs during exposure to organic pollutants, although much less obvious, is also suggested (Kaegi, 1993). Consequently, in *L. rubellus* and in *E. fetida*, MT (in fact MT2 form) is considered as a good biomarker of exposure because the gene showed a time-and dose-dependent increase of both the number of MT2 transcripts and the protein concentration when worms are experimentally exposed to MTEs, especially Cd (Gruber *et al.*, 2000; Brulle *et al.*, 2006; Demuynck *et al.*, 2007).

3 Interpreting the biological response of field-collected earthworms from sentinel species to MTE pollution

Experimental protocols of ecotoxicological studies interested in the characterization of the biological response of organisms to contaminant exposure are generally similar. Naïve specimens are exposed under control laboratory conditions, typically in microcosms (Fründ *et al.*, 2010), to one or several levels of contaminant concentrations, using either artificially contaminated substrates or field sampled soils. It has to be noticed that, most of the time, exposed specimens belong to a few model species that may not naturally occur on polluted soils, but offer the advantage to be model species allowing the use of a large set of molecular tools. Moreover, those studies are mostly based on the analysis of stress responses over a short period of time at most equal to individual's lifetime. Whereas the accumulation of such a fundamental knowledge is crucial for the development of biomarkers, we have to consider that may not be sufficient for two mains reasons. First, because the biology of sentinel species used for biomonitoring may differ from the biology of model species used in the laboratory. Second, because field sampled organisms may belong to natural populations that

have been exposed to contaminant over a long period of time, allowing an adaptive evolution of stress response.

- ***From model species to sentinel species*.**

The number of model species of macro-invertebrates that are commonly used in ecotoxicological experiments remains limited. Indeed, model species usually limits to species that are easy to maintain and breed in laboratory conditions and for which molecular tools are available. On the contrary, criteria for selection of sentinel species can be quite different (O'Brien *et al.*, 1993). Consequently, model and sentinel species may differ and biological and ecological differences among species may have significant consequences on their response to contaminant exposure. Considering the ecotoxicology of MTE exposure in oligochaete annelids, model species are mostly from the *Eisenia* genus. *E. fetida* and *E. Andrei*, in particular, have been used in a majority of toxicological studies (Sanchez-Hernandez, 2006), although species from the *Lumbricus* genus are increasingly studied. In particular, *E. fetida* is the reference earthworm in the international toxicity tests (Nahmani *et al.*, 2007). In contrast, candidate species for terrestrial ecosystem assessment in Central Europe belong to several genus: *Eisenia*, but also *Lumbricus*, *Dendrodrilus*, *Dendrobaena*, *Aporrectodea*, *Allolobophora*, *Proctodrilus*, *Octolasion* (Römbke *et al.*, 2005).

Table 1
Ecological characteristics of relevant Central European earthworm species.

Ecological classification	Pigmentation	Feeding behavior	Adult size	Burrows	Mobility	Ecological role	Example
Epigeics	Heavy, usually both ventrally and dorsally.	Feed on organic matter in soil surface.	Small–medium	- Do not dig. - Live in the litter above the soil surface or in the upper reaches of the mineral soil.	Rapid movement in response to disturbance.	- Decomposing litter on the soil surface, little or no soil ingested.	- *Dendrobaena octaedra* - *Dendrobaena attemsi* - *Dendrobaena illyrica* - *Eisenia andrei* - *Eisenia fetida* - *Lumbricus rubellus* - *Lumbricus castaneus*
Anecic	Medium-heavy, usually only dorsally.	- Come to the surface to feed. - Feed on the organic matter on the surface and in the soil.	Large	- Make individual burrows that are vertical and deep.	Rapid withdrawal into burrow but more sluggish than epigeics.	- Their burrows affect the transfer of water, air and nutrients through the soil.	- *Aporrectodea nocturna* - *Lumbricus terrestris* - *Nicodrilus longus*
Endogeic	Unpigmented or lightly pigmented.	- Feed on the organic matter in the soil. - Consume more soil than do others.	Medium	- Make horizontal non-permanent burrows in the soil.	Generally sluggish.	- Are responsible for the incorporation and distribution of organic matter in the soil.	- *Allolobophora chlorotica* - *Aporrectodea caliginosa* - *Aporrectodea rosea* - *Octolasion cyaneum* - *Proctodrilus oculata*

Experimental studies involving a direct or indirect comparison of the biological response of *Eisenia* species to other earthworm species (mostly from the *Lumbricus* genus) remains scarce but they suggest that putative physiological differences among species may only have a limited impact on their response to MTE exposure (Calisi *et al.*, 2009, 2011). However, recent considerations suggest that the biological response of *E. fetida* may differ from other earthworm species because of ecological differences. Indeed, earthworms can be divided into three ecological groups (Table 1): epigeics, that live in the litter above the soil surface, anecics, that live in temporary vertical deep burrows and come to the surface to feed, and endogeics, that rarely come to the soil surface, make horizontal nonpermanent burrows and feed on the organic matter already in the soil (Römbke *et al.*, 2005). Because the horizontal distribution of MTEs in soil layers may be heterogeneous, varying feeding behaviour among soil dwelling earthworms occurring on a same polluted site may determine species-specific patterns of metal exposure and accumulation in tissues (Suthar *et al.*, 2008). Accordingly, it can be shown that, at the community level, MTE pollution affects differently sympatric populations of endogeic and epigeic species (Nahmani *et al.*, 2003). At the infra individual level, uptake and bioaccumulation of MTES may vary among

earthworm ecological categories, with a general (but not systematic) tendency toward higher metal contents in tissues of epigeic and endogeic species than anecic species (Suthar *et al.*, 2008; Ernst *et al.*, 2008; Tischer, 2009). Finally, at the cellular level, the amount of Cd-induced DNA damage may also differ among ecological categories (Fourie *et al.*, 2007).

Interestingly, the relative relevance of considering biological groups for the development of generic biomarkers may also be extended at the within species level. For example, in *Allolobophora chlorotica*, two morphs have been described, depending on the presence of the biliverdin pigment (Satchell, 1967). *In situ* observations suggest that both morphs differ in ecological preferences related to soil moisture. The green form dominates in wet soils and pink form in dry soils. Laboratory experiments confirm that growth and maturation of the green form are significantly reduced in dry soil while growth of pink form is similar in dry or wet soil (Lowe & Butt, 2007). Similarly, in *Lumbricus terrestris*, sequencing data of the cytochrome oxidase I (COI) mitochondrial gene revealed the existence of two cryptic species (James *et al.*, 2010). Both species are supposed to live in sympatry, except in the north-west of France, where the exclusive presence of one cryptic species could be explained by ecological differentiation.

- ***From experimental populations to natural populations*.**

When analyzing the biological response to MTE exposure, shifting from experimentally exposed to field-sampled organisms may result in a significant difference in the history of contaminant exposure of organisms under study. Indeed, experimental protocols of toxicological studies usually extend over short periods of time, typically inferior to individual's lifetime. At this ontogenic timescale, mechanisms involved in stress response can only consist in physiological adjustments, a phenomenon known as acclimation or phenotypic plasticity (Morgan *et al.*, 2007). Thus, the overexpression of gene coding for proteins involved in detoxification processes illustrated above can be interpreted as an environmentally induced regulation of gene expression allowing single individuals to cope rapidly with a changing environment. On the contrary, organisms collected in contaminated sites may belong to natural populations that have been continuously exposed to contaminants for generations. In this case, long-term exposure, over many generations may have resulted in genetic changes conferring tolerance to toxic effects of metals in soil, leading to a genetic differentiation between exposed (or metallicolous) and non-exposed (or non-metallicolous) populations through a process called local adaptation (Kawecki & Ebert, 2004). This might

bias estimations of the level of soil pollution from the phenotype of field-sampled organisms (Medina *et al.*, 2007; Clements & Rohr, 2009; Bickham, 2011).

Local adaptation requires the action of selective pressures that are sufficiently strong and long lasting to (1) generate differential survivorship and reproductive success among individuals over several generations and therefore (2) provoke changes in allelic or genotype frequencies in metallicolous natural populations comparing to geographically close ones. Considering the toxicity of MTE polluted soils and the persistence of MTEs (that cannot be degraded) in the environment, the evolution of local adaptation in metallicolous population of any species developing in close contact with a polluted soil is highly expectable. A few studies carried on species from the soil macrofauna suggest it indeed occurs (Table 2). One of the best document example concerns the soil-living collembola *Orchesella cincta*. In this species, it has been shown that the contamination of soils by MTEs modified the genetic composition of exposed populations, resulting in the evolution of more tolerant genotypes (Janssens *et al.*, 2008; Costa *et al.*, 2012).

In earthworms, definitive evidence of local adaptation of natural population to MTE exposure does not exist yet. However, several results (Table 2) suggest that (1) MTEs can cause varying fitness among individuals and thus act as selective pressures in natural populations and that (2) genetic variation, a fundamental condition for the evolution of local adaptation, exists for genes involved in homeostatic regulation and detoxification of metals.

Table 2: Referenced studies that previously suggested local adaptation in earthworm populations.

Species	Pollutant examined	Main results	References
Dendrobaena octaedra	Cd	Although Cd exposure generally affects life history and reproductive traits, the survival of F1 offspring of worms is significantly higher when the parents originate from polluted soil.	Rozen, 2006
Dendrodrilus rubidus	Cu	When exposed to soils treated with Cu, mortality of earthworm sampled from a Cu-polluted soil is shown to be lower than mortality of earthworms from a control Cu-free soil, suggesting that local adaptation leading to higher Cu-tolerance occurs in metallicolous populations of earthworms. Interestingly, the fertility of tolerant earthworm is significantly reduced, suggesting that Cu-tolerance has a cost.	Arnold et al., 2008
Aporrectodea caliginosa *Aporrectodea longa* *Aporrectodea rosea* *Dendrodrilus rubidus* *Lumbricus rubellus* *Lumbricus terrestris* *Octolasion cyaneum* *Octolasion tyrtaeum*	Hg, Cd and Pb	The study mainly focuses on interspecific comparisons from field-collected earthworms. However, results clearly suggest that intraspecific variation exists in both metal tissue concentrations and concentration factor. Investigating the genetic component of phenotypic variation would allow discussing the occurrence of genetic resources for metal tolerance evolution.	Ernst et al., 2008
Lumbricus rubellus	As	F1, F2, F3 offspring of worms collected in As-contaminated soils show better health after As exposure than offspring of worms from uncontaminated soils, suggesting the evolution of genetically based resistance to As-toxicity.	Langdon et al., 2009
Dendrobaena octaedra	Cu	F1-generation worms obtained in laboratory conditions are used to show higher growth rate of worms from contaminated sites in both control and copper-spiked soil conditions, suggesting adaptive population differentiation. However, no statistical difference among population in survival, maturation time, cocoon production and hatchability can be shown.	Fisker et al., 2011
Eisenia fetida	Pb, Cd	The total accumulated body burden of metals for worms homogeneously exposed to a contaminated field soil is highly variable among individuals. This suggests that genetic variation may occur in the molecular pathways of metal homeostasis.	Spurgeon et al., 2011
Lumbricus rubellus	As	Although not specifically addressed, genetic variation in response to As exposure is evidenced at several life stages. Population-level effects of As exposure are also discussed. Phenotypic variation for several life-history traits is observed among phylogenetically divergent clades in controlled conditions, suggesting that genetic resources allowing adaptation may exist. However, As sensitivity does not seem to differ among clades.	Anderson et al., 2013

3.1 Reduced fitness of earthworms under metal exposure

Negative consequence of MTEs exposure on fitness-related traits has been highlighted in several studies, suggesting that local adaptation can be assumed to occur in earthworm metallicolous natural populations. For example, using a specifically designed test system, Spurgeon *et al.* revealed a exposure-dependent effects copper and cadmium toxicity on survival, growth, development time of juveniles obtained from cocoons collected in natural populations of *L. rubellus* (Spurgeon *et al.*, 2004). Considering arsenic toxicity, the results were recently extended to adult life-history traits. (Anderson *et al.*, 2013) In another study focusing on *E. fetida*, Nahmani *et al*. measured survival, body weight, cocoon production and hatching rate in earthworms exposed to metal-polluted and uncontaminated soils in control conditions. They revealed significant negative correlation among both growth and reproductive traits and metal concentration in soils, although soil characteristics other than metal concentration (organic carbon content, texture, pH) were shown to influence the level of soil toxicity (Nahmani *et al.*, 2007).

The occurrence of genetic variation of metal-related genes is highly suggested by the varying concentrations of metals in tissues of conspecific earthworms either exposed to a same level of metal exposure in laboratory conditions e.g. (Smith *et al.*, 2010) or collected at a same site (Ernst *et al.*, 2008; Uba *et al.*, 2009). However, in natural populations, varying concentrations

of metals in tissues may also be due to the heterogeneity of metal exposure among sympatric earthworms and ”common garden” designs are necessary to evidence that it can be genetically determined. In *E. fetida*, Spurgeon *et al*. specifically addressed the question of factors that could account for among individual variation for the total accumulated body burdens of Pb and Cd (Spurgeon *et al.*, 2011). By analyzing variation in individual tissue burdens of earthworms homogeneously exposed to metals, they concluded that observed differences might have a substantive biological component, suggesting that genetic variation exists for molecular mechanisms involved in uptake and/or elimination/biotransformation pathways. Interestingly, despite that, the genetic origin of phenotypic differences among metallicolous and non-metallicolous earthworms collected in the field can still be controversial. For example, Rozen showed that, in *Dendrobaena octaedra*, earthworms raised in similar laboratory conditions but collected in soils with various levels of metal pollution had different metal (Cd) body burdens, suggesting genetic differentiation (Rożen, 2006). However, the same experiment on the offspring (F1 generation) of collected earthworms revealed no difference, suggesting that observed differences in cadmium accumulation ability among parental populations might not be heritable, and thus might reveal acclimatory adjustments rather than adaptive differences.

It is worth to note that genetically determined variations in the level of MTEs concentration in tissues may not necessarily mean that some genotypes or populations are more adapted to metalliferous sites than others. A common garden approach comparing the sensitivity to MTE exposure of populations from both polluted and unpolluted sites would be appropriate to reveal a higher resistance of metalliferous population, *i.e.* a better fitness of those populations compared to population from uncontaminated environments (Janssens *et al.*, 2008). In *Dendrodrilus rubidus*, Arnold *et al*. (2008) compared the weight, health and mortality of earthworms from a mine spoil soil at a an abandoned copper (Cu) mine and a Cu-free control site in several treatments varying for total Cu soil concentrations (Arnold *et al.*, 2008). In highly Cu-contaminated treatments, earthworms from mine sites were shown to be healthier and to suffer significantly less change in weight and mortality than control earthworms, suggesting higher resistance. Interestingly, the acquisition of resistance was assumed to be associated with an elevated physiological cost significantly reducing cocoon production and viability. In *L. rubellus*, Langond *et al*. demonstrated that earthworms collected at an abandoned copper and arsenic (As) mine acquired genetically inherited resistance to As-toxicity (Langdon *et al.*, 2009) than control earthworms. Following offsprings of field-

sampled earthworms exposed to 2000 mg As kg^{-1} dry weight of soil over several generations, they showed that the former had lower mortality and higher cocoon viability than control earthworms. In *Dendrobaena octaedra*, Fisker *et al*. tested for the occurrence of adaptive differentiation for the level of copper among populations comparing three populations originating from copper contaminated sites to three ones originating from non-contaminated site (Fisker *et al.*, 2011). For each population, F1-generations were produced, to ensure that among population differences were heritable, and placed in uncontaminated control or copper-spiked soil. It appeared clearly from this experience that populations originating from polluted sites had a higher growth rate, an earlier acquisition of sexual maturity, a better reproduction rate and a lower mortality than populations originating from reference sites.

4 Biological responses of field collected earthworms at the transcriptome level

The evolution of local adaptation in natural earthworm populations occurring of metalliferous sites may imply that the biological response of field-collected earthworm could differ from the response of naïve specimens characterized in short-term laboratory experiments. At the transcriptome level, in particular, it may be expected than the inducible increase in the expression of gene involved in detoxification mechanisms observed in laboratory experiments (see above) cannot be confirmed in wild individuals from natural populations. Indeed, it seems that the local adaption of population exposed to stressful conditions involves the evolution of constitutive overexpression of gene associated to stress response (Roelofs *et al.*, 2008). Therefore, the level of expression of those genes may not be exclusively related to the level of exposure of individuals, as it is usually assumed in the biomarker concept focusing on transcriptomic tools.

In the specific context of MTE exposure, the constitutive overexpression of some target genes in adapted individuals of the springtail *O.cincta* comparing to sensitive ones has been demonstrated (Roelofs *et al.*, 2008). In this species, increase of resistance in natural populations has been shown to be associated with an elevated level of both constitutive and cadmium induced metallothionein expression (Sterenborg & Roelofs, 2003), whereas individuals from sensitive populations show a typical stress-induced gene expression. It has been suggested that the evolution of overexpression could be related to modifications of the genetic composition of adapted population through the increase in frequency of certain alleles for the promoter region of the MT gene. Interestingly, those alleles could be either shared

with other populations of found exclusively in Cd-resistant populations (Costa *et al.*, 2012). Similar results are not available in earthworms yet. However, it has been shown in a study testing the potential for the development of biomarkers in earthworms that the level of metallothionein expression observed in oligochaete annelids collected in a Cd-polluted site a long time ago (> 100 years) was significantly increased (Pérès *et al.*, 2011).

5 Conclusion: Integrating adaptation into the biomarker concept

Biomarkers have been mostly developed on model species in short-term laboratory experiments, and transferred as such on sentinel species from natural populations. In this paper, it has been demonstrated that, theoretically, this could lead to misinterpretation owing to adaptive mechanisms potentially occurring in natural populations that are exposed to contaminants since many generations. When especially regarding earthworms, it appears that conditions for local adaptation to occur are present (namely, potential strong and durable selective pressures and within-population phenotypic variability). In addition, some infra-individual evidence of adaptation obviously exists. Accordingly, the assumption of the evolution of constitutive overexpression of detoxification genes cannot be ruled out. Therefore, it seems reasonable to integrate the potential effects of local adaptation to MTEs on biological responses of organisms in the development of biomarkers. Practically, regarding the use of transcription levels of stress-response genes - such as MT - as biomarkers, we suggest that two additional experiments should be systematically developed to properly interpret expression levels. First, it seems necessary to analyze expression levels on offspring of field-samples individuals to ensure that overexpression is not constitutive. Second, the use of molecular genotyping markers may be useful to test the possibility of genetic changes caused by contamination exposure in natural populations. In particular, one should ensure that sampled individuals do not carry specific variant associated with constitutive overexpression. In our opinion, this second point is critical since the availability of molecular markers remains low in earthworms (Dupont, 2009).

Chapter III

Isolation and characterization of eight polymorphic microsatellite markers for the earthworm *Lumbricus terrestris*

Microsatellite are nuclear DNA markers, co-dominant, polymorphic (mutation rate approaching 10^{-3} per locus per generation), neutral (aren't subject to selection, located in non-coding regions generally able to evolve rapidly) and giving reliable results (Sunnucks, 2000). They are considered by many authors as the most powerful nuclear markers available for population genetics (structure, genetic diversity and population history) and gene flow analysis (Zhang & Hewitt, 2003), but can also be used to determine the relationship, explore a phylogeny or phylogeography. Microsatellites therefore allow both to perform phylogeographic study (i.e. spatial study) and a space-time study by understanding gene flow. Therefore, our first choice in terms of markers is naturally oriented to microsatellite markers. Microsatellite markers are specific, and therefore they require development for each studied species. Given their unavailability for the species *Lumbricus terrestris*, the development of microsatellite markers in this species of earthworm was extremely essentials for proceeding population genetic studies for this species.

Table of contents for chapter 3

Isolation and characterization of eight polymorphic microsatellite markers for the earthworm *Lumbricus terrestris*

Dima Souleman[a,b,c], Fabien Grumiaux[a,c,d], Hélène Frérot[a,b], Franck Vandenbulcke[a,c], Maxime Pauwels[a,b]

[a] Université de Lille, F-59000 LILLE, France

[b] Laboratoire Evolution, Ecologie et Paléontologie, UMR CNRS 8198, Bâtiment SN2, F-59655 Villeneuve d'Ascq cedex France

[c] LGCgE, EA4515, Univ Lille1, « Fonctionnement des Ecosystèmes Terrestres Anthropisés », SN3, Cité Scientifique, F-59655 Villeneuve d'Ascq, France.

[d] Université Lille Nord de France, Ecole Supérieure du Professorat et de l'Education (ESPE), site d'Arras, 7 bis rue Raoul François, BP 30927, F-62022 Arras Cedex, France.

Published in: European Journal of Soil Biology

DOI number: 10.1016/j.ejsobi.2016.03.009

Abstract

The earthworm *Lumbricus terrestris*, which is an ecologically important sentinel species for soils widely distributed throughout Northern Europe, has never been involved in deep population genetic studies. In order to promote future studies, we report here the isolation and characterization of 8 new polymorphic microsatellite loci isolated from sequencing of a microsatellite-enriched genomic library. Microsatellite markers were tested on 192 field-collected individuals. Allelic richness per locus varied from 4.921 to 24.876, with a mean value of 6.125 alleles per locus. The observed heterozygosity (Ho) ranged from 0.132 to 0.839, whereas the expected heterozygosity (HE) ranged from 0.407 to 0.926. Developed markers should be helpful for accurate estimation of population structure, dispersal and gene flow among populations.

Keywords: *L. terrestris*, Microsatellites, Genetic diversity, Heterozygosity.

1 Introduction

In terrestrial ecology, one of the most studied zoological groups is the Annelida, Oligochaeta group. These earthworms play key roles in most continental ecosystems, represent an important part of the soil macrofauna, and are integral to the maintenance of the structure and the fertility of soils. By drilling, earthworm can create both horizontal and vertical burrows, which can be very deep in the soil. These burrows produce pores through oxygen and water can enter and carbon dioxide can leave the soil. They increase soil fertility by mixing the soil layers and are considered as ecosystem engineers by agronomists (Edwards *et al.*, 1998; Kautenburger, 2006). Earthworms from the Lumbricidae family are also good candidates for biomonitoring (Römbke *et al.*, 2005; Sanchez-Hernandez, 2006) because they are largely distributed, easy to sample and may be used to assess the effect of soil contamination through the measure of biomarkers (Pauwels *et al.*, 2013; Pelosi *et al.*, 2013). Usually, earthworms are categorized into three ecological groups, i.e. epigeic, endogeic and anecic species. This categorization is based on the locality inside the soil and on the burrowing attitude (Capowiez *et al.*, 2010). Among Lumbricidae, the species *Lumbricus terrestris* is typically an anecic that lives in semi-permanent burrows and come to the surface to feed. It is one of the largest species among Northern Europe earthworms. It is also one of the easiest to identify based on external characteristics when exhibiting a clitellum (breeding form). As a result, *L. terrestris* is an attractive sentinel species which is often used (possibly in combination with other earthworms species) to assess environmental impact of soil contamination, land-use change and fragmentation of habitats (Capowiez *et al.*, 2010; Lapied *et al.*, 2010).

In comparison with ecological properties, the population genetic of the species has been poorly investigated and we know almost nothing about the distribution of genetic diversity within the species. This situation is noticeably illustrated by the recent discovery, through sequencing of the Cytochrome Oxidase I gene (COI) that, in Europe, *L. terrestris* likely includes a cryptic species named *L. herculeus* (James *et al.*, 2010). This situation may result from a lack of appropriate molecular tools. In this context, the development of molecular genotyping markers, such as microsatellites, would be very useful for accurate estimation of population genetic structure, levels of within-population genetic diversity or among-population gene flow. Microsatellites are neutral, codominant, highly polymorphic markers. They have been shown as highly suitable markers for population genetics (Guichoux *et al.*, 2011). Some microsatellite markers have been previously developed for other earthworm

species: eight markers in *Lumbricus rubellus* (Harper *et al.*, 2006), sixteen markers in *Eisenia fetida* (Somers *et al.*, 2011), eight markers in *Allolobophora chlorotica* (Dupont *et al.*, 2011) and eleven markers in *Aporrectodea longa* (Strunk *et al.*, 2012). Ten markers have also been proposed for *L. terrestris* (Velavan *et al.*, 2007). Surprisingly, however, developers used only three of them in a subsequent study (Velavan *et al.*, 2009). We ourselves failed in obtaining satisfying results testing these markers on individuals we investigated (data not shown).

In this context, the aim of our study was thus to use a recently developed high-throughput method for isolating microsatellite markers (Malausa *et al.*, 2011) to increase the number of microsatellite markers available for population genetic studies of *L. terrestris*. We also performed multiplexing of selected markers to make high-throughput genotyping easier. We tested the efficiency and level of polymorphism of proposed markers in a high number of individuals collected from the Nord-Pas de Calais Region in France.

2 Materials and Methods

2.1 Sampling

Twelve European individuals of *L. terrestris* have been used for the development of a microsatellite-enriched genomic library. They were collected in northern (Nord-Pas de Calais Region) and southern France (Bordeaux and Avignon) and in southern Poland (Krakow). Allyl isothiocyanate (AITC) was used to extract earthworms from their subterranean burrows (Zaborski, 2003). Collecting individuals in different parts of the species distribution in Europe was supposed to favour the selection of microsatellite loci whose mutational dynamics does not depend on any reference genotype. This may help to develop markers whose levels of polymorphisms are sufficient at various geographic scales.

Selection of microsatellites and validation of the selected microsatellites were performed on 192 individuals collected from 44 sampling sites Nord-Pas de Calais Region, France (maximum distance among sites: 3 kilometres). Each sample site was made of 1m² quadrat. AITC was used to extract earthworms. Worms have been identified with the taxonomic key of Bouché, (1972) ('Phylogenetic species delimitation of the earthworms *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* Bouché, 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences') using characteristics such as the prostomium shape, the shape and the location of the clitellum (saddle-shaped on segments 32-37) and the location of the tuberculata pubertatis (segments 33-36).

2.2 DNA extraction

Earthworms were cleaned and stored in ethanol 96% until use. Total genomic DNA was isolated using NucleoSpin® 96 Tissue (Macherey Nagel) kits that allow for the purification of multiples of 96 samples. Small part of the end of each earthworm tail was used to extract genomic DNA. Then, DNA extracts were 1:20 diluted before use.

2.3 COI barcoding and analysis

Given that two cryptic species were identified in the morphospecies *L. terrestris* (James *et al.*, 2010), the taxonomical identity of our samples was checked using COI sequencing. COI was amplified using universal DNA primers LCO1490 and HCO2198 (Folmer *et al.*, 1994). However, in order to ensure that amplification failure was not due to primer-template mismatches, an additional amplification was performed using redesigned amplification primers (LCO1490_Lt_F: ACTCAACTAATCACAAAGATATTGG; HCO2198_Lt_R: TATACTTCTGGGTGACCAAAGAATCA) from the complete *L. terrestris* mitochondrial genome sequence (GenBank NC001673.1). Standard protocols were used for amplification and sequencing reactions (James *et al.*, 2010). The Millipore-Multiscreen purification kit was used for amplicon purification and PCR fragments were sequenced using the BigDye TerminatorKit 3.1 (Applied Biosystems) and run on an ABI-3130 capillary sequencer (Applied Biosystems). All sequences were checked manually with CodonCode Aligner Version 5.1.4. Sequence analysis was performed as suggested in (James *et al.*, 2010).

2.4 Microsatellite library

A stoichiometric mixture of the 12 European DNAs was performed to produce a microsatellite enriched genomic library. Library was constructed by the Genoscreen Company (Lille) by coupling multiplex microsatellite enrichment isolation techniques with the 454 GS-FLX Titanium pyrosequencing (Malausa *et al.*, 2011). Enrichment was performed using probes containing the following microsatellite motifs: TG, TC, AAC, AAG, AGG, ACG, ACAT, ACTC. Overall, 5931 sequences containing a single microsatellite motif were found. This allows the identification of 244 independent microsatellite loci candidates. Amplification primer pairs were designed for each *in silico*.

2.5 Microsatellite amplification

For 20 microsatellite loci, one primer pair was chosen for *in vitro* Polymerase Chain Reaction (PCR) amplification trials. Microsatellite loci and primer pairs were chosen depending on the repeated motif (only di-nucleotides and tri-nucleotides motifs were selected to avoid very large allele size range), maximizing the number of repeats of the motif observed in the library (considering that loci with higher number may also show higher allelic richness in natural populations) and selecting for expected amplicon size so as to allow multiplexing. Amplification trials were first performed on a subset of individuals sampled in the Nord – Pas de Calais region (n = 16) so as to ensure that selected microsatellite markers show the highest rate of positive results (to limit the possibility of null alleles), high PCR amplification yields (estimated through peak levels), low allele drop-out, and no multiple peak profiles (stutter artifacts). We used a four-step procedure for validation. Firstly, PCR amplifications were conducted in 15µl reaction including 0.3µl of both forward and reverse primers (10mM), 1.5µl of 10X DreamTaq Buffer (containing 20 mM Tris-HCl (pH 8.0),1 mM DTT, 0.1 mM EDTA, 100 mM KCl, stabilizing agent and 50% (v/v) glycerol), 1.2µl of dNTPs (2.5mM), 0.3µl of BSA, 0.075µl DreamTaq polymerase and 3µl of diluted DNA. PCR were designed with initial denaturation at 94 °C for 5 min, followed by 35 cycles of 1 min at 94 °C denaturation, 1 min at 60°C annealing, 1 min at 72 °C extension, followed by a final extension of 5 min at 72 °C. PCR products (yield, amplicon size) were checked using gel electrophoresis in 1% agarose gels and ethidium bromide staining. GeneRulerTM 100bp DNA ladder was used to determine fragment size. Secondly, the 20 markers were tested separately using labelled forward primers. Then, 1.5µl of each PCR product was mixed with 9.75µl of formamide (Applied Biosystems) and 0.25 µl of Genescan-500 LIZ size standard, and analyzed through electrophoresis on an ABI3130 Genetic Analyzer (Applied Biosystem). Only eight microsatellite markers show both consistent amplification results and genetic variation among individuals. During this step, five microsatellite markers available in the litterature for *L. terrestris* (Velavan *et al.*, 2007, 2009) and that could be multiplexed were also tested. Only one of them (LT163) gave satisfying results. Thirdly, the eight microsatellite markers were multiplexed and co-amplified in two subsets of loci called Mplx1, Mplx2. Each multiplex contains four markers, which are directly labeled on the forward primer with different fluorescent loading dye (6-FAM, NED, PET, or VIC, Applied Biosystems (Table 1). PCRs were performed for multiplex adjust under the following

conditions: denaturation step at 95°C, for 15 min, 5 cycles including (45s at 95°C, 5 min at 68 °C, and 60s at 72 °C), 5 cycles including (45s at 95°C, 1 min at 58 °C, and 60s at 72 °C), 20 cycles including (45s at 95°C, 30s at 47 °C, and 60s at 72 °C), and extra extension step of 30 min at 60 °C, using QIAGEN Multiplex PCR Kit. After, 1.5µl of PCR product was mixed with 9.75µl of formamide (Applied Biosystems) and 0.25 µl of Genescan-500 LIZ size standard, and analysed by DNA sequencer (Applied Biosystem). Fourthly, the two developed multiplexes were tested using the same PCR conditions on a set of 192 samples coming from 44 locations in northern France (Nord-Pas de Calais).

In order to discuss the possibility of cross-species transfer of developed microsatellite markers, the latter were also tested on a set of 16 genotypes from each of the *L. castaneus* and *L. rubellus* species. DNA extraction and amplification procedures were the same as above described.

2.6 Microsatellite data analysis

Fragment analysis and allele size determination were carried out using GENEMAPPER 3.7 software (Applied biosystem). Depending on marker peaks' position and allowing an offset of up to 0.8 bases among peaks at a same position, a set of detected alleles was determined for each locus. The association between genotypes at pairs of loci was tested using the log-likelihood ratio G-statistic in FSTAT version 2.9.3. FSTAT version 2.9.3 was also used to determine genetic diversity parameters such as, the number of alleles and the FIS coefficients per locus. Then null allele frequency for each locus was estimated using the EM algorithm implemented in the FreeNA software (Chapuis & Estoup, 2007). The observed heterozygosity (H_O) and expected heterozygosity (H_E) were established for each locus using GeneClasse2 (Piry *et al.*, 2004).

Table1: Identification of 8 polymorphic microsatellite loci for *L. terrestris*. A: alleles number; NA: estimated null allele frequency; H_O: observed heterozygosity; H_T: expected heterozygosity; F_{IT}: fixing index. *** : Significant value.

Locus	PCR product size	Primers	Repeat motif	Loading dye	Percentage of positive results	Allele size range	Most frequent allele size	A	NA	H_O	H_T	F_{IT}
Lt3	127	F: CAAAGATGCGAACAAACGTC R: CGCTTGATACATAATAGCACGG	(ca)8	PET	66%	115-175	129	11	0.25	0.305	0.753	0.596***
Lt8	134	F: TCATCGTCATCGTCCTCATC R: CATCATTATTATCGTGAAAGCGA	(tca)8	VIC	73%	125-189	143	5	0.25	0.187	0.575	0.675***
Lt16	187	F: ACTTGTGCGACACGATTGAA R: TTTCCTCGAGTCTTCTTCGC	(gag)10	6-FAM	86%	169-226	171-177	6	0.28	0.256	0.742	0.655***
Lt19	129	F: CTAAGCAGAACACCTTGGGC R: ACGTTGTCTGGCGTTTTAGC	(gt)7	NED	79%	170-243	204	6	0.22	0.132	0.407	0.674***
Lt17	210	F: TCAAGCGCTCAAAGATGTCA R: TGGAAACTGGTATCTCGGATG	(ca)7	VIC	77%	197-264	211	21	0.04	0.810	0.906	0.107
Lt9	141	F: ACTTGACACAGCGCCAAAG R: CGCTTGATACATAATAGCACGG	(ca)8	NED	68%	110-193	141	25	0.24	0.385	0.846	0.546***
Lt12	212	F: TGCATTGTGAGAGCAATTGTG R: TGAGATATGATAGCGGAGTGAA	(aca)11	6-FAM	79%	190-282	215	19	0.33	0.205	0.803	0.744***
Lt 163	171	F:GCCGGAGCGTTAGGAGCGATAG R:GGATACGCCCGACTCACCACTAA	(tgc)12	PET	81%	138–216	150	23	0.05	0.839	0.926	0.094

3 Results and Discussion

Over the 192 tested individuals, 145 gave positive COI amplification results and 177 gave positive amplification results for at least one microsatellite marker (Table S1). Individuals that gave negative results for 8 (n = 15), 7 (n = 5) or 6 (n = 7) microsatellite markers were generally not amplified by COI primers (n = 13, n = 5 and n = 3, respectively, see Table S1). However, all COI sequences we obtained, including sequences of individuals with not optimal genotyping results, felt in the *L. terrestris* cluster identified in (James *et al.*, 2010) (Figure 1). This suggests that the cryptic species *L. herculeus* does not exist in the geographic area we sampled and that, whereas developed microsatellite markers appeared to be highly specific (see below), amplification failing is not due to incorrect species identification. As a result, we believe that the absence of amplification for resulted from insufficient quality of DNA after an extended storage period in ethanol 96%.

Statistical analyses using Fstat version 2.9.3 indicate that the eight validated loci are highly polymorphic, especially loci Lt17, Lt9, Lt12, Lt163, Lt3 which exhibited a high level of allelic diversity (Table 1). By means of 192 individuals, a total of 116 alleles were identified overall. The number of alleles ranged from 5 at the locus Lt8 to 25 at the locus Lt9. The tests for genotypic disequilibrium were non-significant for all pairs of loci.

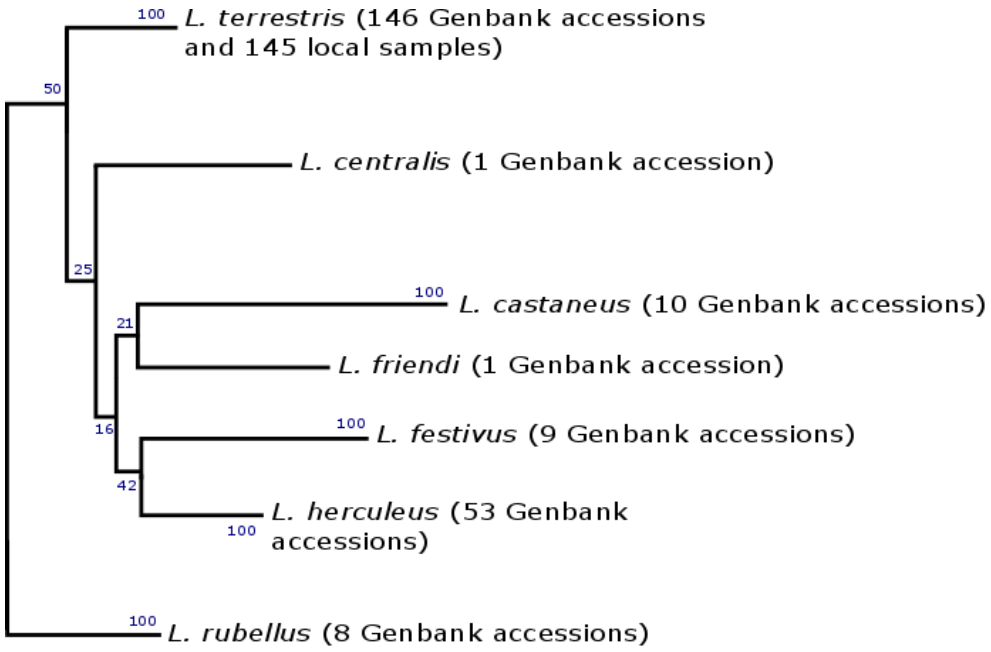


Figure 1: Neighbor joining tree (Kimura 2-parameter) for 6 species of genus *Lumbricus*, based on the COI 5' barcoding fragments, bootstrap value is recorded at the end of each cluster branch.

Among the eight loci, two (locus Lt17 and locus Lt163) showed relatively low estimated null allele frequencies. On the other hand, independently of allelic richness, the estimated null allele frequencies were higher at the other six loci. Estimated null allele frequencies may be higher than true frequencies for several reasons related to the population genetic model used for the estimation. Indeed, null allele frequencies are estimated assuming that individuals belong to a panmictic population and that heterozygote deficiencies from expected proportions are only due to the presence of null alleles. However, given the current absence of knowledge about *L. terrestris* populations' genetics, we cannot ensure that sampled individuals belonged to a unique population. Moreover, it is very likely that mating systems of earthworms showing hermaphrodism and assortative mating over short geographic distances differ from panmixia (Darío J. Díaz Cosín, 2010). In general, earthworms are considered as a cross-fertilization hermaphrodite and they are rarely self-fertilizing. However, some species make an exception of this reproduction strategy. For example, spermatozoids of *E. fetida* were able to fertilize ova from the same individual (Domínguez *et al.*, 2003). Besides, genetic diversity may be affected by the presence of parthenogenesis in earthworms. Accordingly, in our study, levels of overall observed heterozygosity were always lower than levels of overall expected heterozygosity (Table 1) and FIT values varied among markers from 0.094 (Lt163) to 0.744 (Lt12) and were significantly positive for 6 of 8 loci, suggesting a non-panmictic mating system. However, such result may also result from population genetic structure in our sampling area, such as a Wahlund effect (Garnier-Géré & Chikhi, 2001).

Tests for cross-species amplification of microsatellite markers revealed that they were highly specific to *L. terrestris*. Indeed, except for Lt163, that successfully amplified from 15 *L. castaneus* individuals and Lt3, that successfully amplified from 9 *L. rubellus* individuals, amplification success for other markers concerned less than half of the 16 tested individuals from each species. This specificity interrogates the capacity of developed markers to amplify other *Lumbricus* species, including *L. herculeus*. Indeed, COI barcodes revealed that *L. herculeus* was not included in our study.

The high variability of the microsatellite loci suggests that the developed markers will have a sufficient degree of polymorphism for population genetic studies. The study of population genetic structure can be useful to investigate effective population sizes as well as the respective effects of genetic drift, migration and selection in the evolution of populations. For example, population genetic data allow discussing how genotypes from distinct

populations with contrasted ecology are genetically connected. High levels of neutral genetic differentiation would suggest genetic divergence among populations, whereas low levels could reveal a favourable background for the detection of outlier loci involved in local adaptation (Hendry *et al.*, 2011). Population genetic surveys should be particularly important for species that, like *L. terrestris*, are used as sentinel species in ecotoxicological studies. Indeed, *L. terrestris* naturally occurs in anthropized habitats, such as metal-polluted sites. The evolution of adaptive divergence among metallicolous and non-metallicolous populations cannot be excluded. Such divergence may affect biological parameters that are used as biomarkers to assess the impact of a contaminant on physiology, since divergent genotypes may behave differentially in front of a specific contaminant (Pauwels *et al.*, 2013). This possibility is hardly taken into account in ecotoxicological studies. Currently, in most biomarker trials, various *L. terrestris* genotypes are generally used without any knowledge about their respective evolutionary history, and, *a fortiori*, the possibility of genetic differences in their phenotypic response. Such knowledge should be made available from population genetic data. In addition, from a more fundamental point of view, the availability of microsatellite markers may also be appropriate to discuss the demographic relationships among the two cryptic species that have been recently detected in *L. terrestris* (James *et al.*, 2010).

In the present study, we report the isolation and characterization of eight polymorphic microsatellites markers for the earthworm *Lumbricus terrestris*, which can be used to investigate the population sizes as well as the respective effects of genetic drift, migration and selection in the evolution of populations.

Chapter IV

Population genetics of the earthworm *Lumbricus terrestris* in disturbed habitats using microsatellite markers

The explosion of human populations led to increasing exploitation of land surface. About 3-6% of land cover was transformed to urban area to accommodate human population bulge during the period of 1980-2000 (Johnson *et al.*, 2015). Urbanisation and anthropogenic activities threaten the local and global ecosystems through dramatic modifications of landscape. Urbanization and its related consequences such as habitats loss, fragmentation and the change of habitats quality may affect the evolution of organisms. One may think, that invertebrates living in close contact with soil (like earthworms) and which are not very mobile, are more sensitive to changing environment. Studying the evolution of these organisms can help to understand the effects of habitats modification on the evolution of species.

The modification of abiotic (water, temperature, pollution) and biotic (parasite, pollinators, fragmentation) components of an environment can modify natural selection and the related adaptive response. Moreover, these modifications can also affect non-adaptive evolution owing to the alteration of evolutionary forces such as gene flow, genetic drift or non-random mating (Donihue & Lambert, 2014). Urbanization is expected to generate differentiation in genetic diversity between populations through urban area re-colonization, which may induce bottleneck in population size (Bartlewicz *et al.*, 2015). Pollutions may also represent one of the abiotic factors that can alter environmental conditions and affect biodiversity.

Nord Pas-de-Calais region, situated in Northern France, is one of the most urbanized region in France. In addition, this region has been producing coal in the past and therefore sheltered numerous metallurgical industries that have spawned a metallic pollution in the upper layers of soil. According to (DREAL, 2013), formerly mining areas like Auby stand by a high proportion of urbanized landscape and by a surface metallic pollution. Two smelters located in this area were the main sources responsible for the production of metallic contamination. Fragmentation related to urbanization and the change of environmental quality related to metallic pollution are supposed to generate environmental stress, which may affect organisms living in close contact with soil.

Among invertebrates living in soil, earthworms are considered as good candidates for biomonitoring (Lavelle & Spain, 2001). In terrestrial ecosystem, earthworms play an essential role in water, carbon and nutrients cycles and they increase soil fertility. *L. terrestris* is an annelid oligochaete which is widely distributed even in disturbed environments. It can survive in

both metallic contaminated and uncontaminated soil. Thus, it provides a good model to investigate the effect of urbanization and metallic pollution on genetic diversity.

We aimed in this work to investigate the genetic variability of *L. terrestris* collected from metallic contaminated and un-contaminated sites in the Nord-Pas de Calais Region. We assumed that changes in the environmental conditions due to the presence of heavy metals and urbanization can induce a difference in genetic diversity among *L. terrestris* populations.

Table of contents for chapter 4

Population genetics of the earthworm *Lumbricus terrestris* in disturbed habitats using microsatellite markers

Dima Souleman[a,b,c], Fabien Grumiauxa,[c,d], Hélène Frérot,[a,b], Maxime Pauwels, [a,b] Franck Vandenbulcke,[a,c],

[a] Université de Lille, F-59000 LILLE, France

[b] Laboratoire Evolution, Ecologie et Paléontologie, UMR CNRS 8198, Bâtiment SN2, F-59655 Villeneuve d'Ascq cedex France

[c] LGCgE, EA4515, Univ Lille1, « Fonctionnement des Ecosystèmes Terrestres Anthropisés », SN3, Cité Scientifique, F-59655 Villeneuve d'Ascq, France.

[d] Université Lille Nord de France, Ecole Supérieure du Professorat et de l'Education (ESPE), site d'Arras, 7 bis rue Raoul François, BP 30927, F-62022 Arras Cedex, France.

Abstract

Transformation of natural habitats into agricultural or urban areas is the main threats to species survival. Terrestrial ecosystems are widely impacted by the destruction, pollution and fragmentation of habitats that represent the primary causes of global diversity decline. To understand the genetic consequence of habitats fragmentation, we compared genetic diversity of five locations of terrestrial species; *Lumbricus terrestris* that are subject to the degrading effects of habitats fragmentation using eight microsatellite markers. We assume that habitats alteration may be affect the structure of genetic diversity in this important species. Microsatellite data showed no differences in the level of genetic variation among locations. However, significant level of genetic differentiation was detected among localities that may due to isolation by geographical distance. However, genetic structure seems to be correlated with the type of fragmentation. No genetic structure was detected in the agriculture location, while genetic structure at fine-scale was revealed in the highly urbanized and metals polluted locations were physical barriers are present. Our results suggest that fragmentation have been modified the structure of genetic diversity of *L. terrestris* without cause any reduction in the level of genetic diversity.

Keywords: Anthropogenic landscape, *L. terrestris*, Microsatellite markers, genetic structure

1 Introduction

In anthropogenic landscapes, human activities have generally strongly affected the environment. For example, anthropic activities often favoured spatial environmental heterogeneity. Environmental changes progressively have fragmented the spatial distribution of species habitats (Merriam, 1984; Saunders et al., 1991; Ellis et al., 2006, 2010; Ellis & Ramankutty, 2008). Habitats could have been either strongly modified, e.g. following pollution, or partially or totally replaced by other habitats or (almost) inhabited urban structures (roads, parking lots, buildings, etc.). *In fine*, habitat fragmentation is usually expected to negatively affect levels of species diversity in the landscape (Polus *et al.*, 2006; Krauss *et al.*, 2010).

Local species extinction results from a negative impact of habitat fragmentation on the survival of populations, through a reduction of mean fitness of populations (Lienert, 2004). However, and prior to species extinction, the intensity of major micro-evolutionary forces driving the evolution of local populations (random drift, selection, migration) should also be affected. As a result, habitat fragmentation should have consequences on population dynamics (*i.e.* population sizes) and, at the molecular level, on the structure of genetic diversity among populations. First, habitat fragmentation should result in population fragmentation. In such a case, large ancestral populations are divided in a set of smaller populations experiencing higher level of genetic drift and separated by unsuitable environments that may act as barriers limiting migration events among populations (Fahrig, 2003). Fragmented, isolated, populations are generally assumed to show reduced genetic diversity and higher levels of inbreeding within populations, and higher genetic differentiation among populations (Templeton *et al.*, 1990; Young *et al.*, 1999). Population fragmentation may also be associated with migration towards newly suitable habitats, resulting in founder effects and reduced genetic diversity in derived populations. Meanwhile, remaining populations may experience the evolution of local adaptation, through the selection of individuals that are tolerant to new environmental conditions. This should provoke strong genetic bottlenecks reducing genetic diversity within populations. It should also promote adaptive divergence among populations that are ecologically differentiated. Adaptive divergence should occur firstly at selected loci. Therefore, divergence could also be genome-wide if local adaptation favoured the evolution of genetic mechanisms of reproductive isolation, limiting among population gene flow.

With 17.2 % and 67% of urban and agricultural areas, respectively, the Nord-Pas de Calais region is among the most anthropized regions in France (in comparison, national means are 9% and 51.5%, respectively) (DREAL, 2013). It is also one of the most densely populated, with 324 inhabitants per square kilometre (national mean is 118). Anthropogenic activities have significantly impacted landscape and biodiversity. Considering plant biodiversity, for example, it has been reported that regional habitats have experienced major changes in species composition over the last century (Van Calster *et al.*, 2008; Hautekèete *et al.*, 2015). In addition, from the XVIIIe century, the region had a long industrial history, including intense mining and metallurgic activities. In particular, the region hosted three important metal (mostly zinc and lead) smelters, including, from 1893 to 2003, one of the biggest lead smelter in Europe, called "Metaleurop-Nord". Consequently, soils have been locally highly polluted by metallic trace element, especially Zinc (Zn), Lead (Pb) and Cadmium (Cd) (Sterckeman *et al.*, 2000; Douay *et al.*, 2007). Metal pollution has so strongly affected the landscape that it generated a specific habitat, called calaminarian grassland that is protected by the European Habitat Directive (Baumbach, 2012). The habitat is characterized by strong selective pressures exerted by the toxicity of high metal concentrations in soils. Accordingly, calaminarian grasslands of Nord-Pas-de-Calais also host very specific plant and animal communities (Nahmani & Lavelle, 2002; Lemoine, 2016). Little is known, however, on the consequences of anthropization on the population genetics of resident species.

In terrestrial ecosystem, earthworms, so-called “engineers of soil”, are key elements of terrestrial species communities because of their strong interactions with many soil abiotic and biotic soil components. They can change the structure of the soil, through their vertical and horizontal burrows (Edwards & Bohlen, 1996). They can increase soil fertility, mixing soil layers and contributing to the humification of organic matter. They also increase the bioavailability of nutrients for plants, or provide aeration and humidity proper for microbial activity (Ndegwa & Thompson, 2001). As a result, earthworm communities can be shown to be sensitive to any anthropogenic activities that may impact soil characteristics, such as land use (Decaëns, 2010; Rutgers *et al.*, 2016), agricultural practices (Chan, 2001), soil pollution by, for example, metals or pesticides (Pelosi *et al.*, 2013; Lévêque *et al.*, 2015) or urbanization (Amossé *et al.*, 2016). Modifications of earthworm communities suggest that, at the population level, population dynamics and genetics should be affected by habitat anthropization. Habitat quality may also modify the earthworm dispersal behaviour, and therefore the level of connectivity among populations (Mathieu *et al.*, 2010; Eijsackers, 2011;

Caro *et al.*, 2013). Finally, habitat change may also favor the evolution of local adaptation in earthworm populations (Pauwels *et al.*, 2013). Altogether, this suggests that the intensity of major micro-evolutionary forces may be affected in earthworm populations occurring in anthropogenic landscape, and that the genetic structure of these populations may be altered. Despite of earthworms' importance and the theoretical effects of landscape anthropization on populations' genetic structure, population genetics of those invertebrates is poorly studied (Costa *et al.*, 2013). Genetic characterization of earthworms' populations is an essential component of species conservation, particularly in small and highly fragmented populations. Few works were focused on the population genetics of earthworms of *L. terrestris* (Kautenburger, 2006), *Allolobophora chlorotica*, *Aporrectodea icterica* and hormogastrid earthworms (Novo *et al.*, 2010; Torres-Leguizamon *et al.*, 2014; Dupont *et al.*, 2015). The objective of this work was to examine if fragmentation and the change of habitat quality have modified the structure of genetic diversity of *L. terrestris*. Thereby, this study was conducted on *L. terrestris* samples collected from five fragmented sites in Northern France, Nord-Pas de Calais region. Genetic differentiation between locations and genetic structure were compared using mitochondrial (COI) sequences and genotyping data obtained using microsatellite markers developed by Souleman *et al.*, (2016).

2 Materials and methods

2.1 Sampling and DNA extraction

This study was carried out in Northern France in the Nord-Pas de Calais region. Earthworms of *L. terrestris* were collected from five different locations. Each location includes several sampling sites that correspond to 1 m^2 quadrat. The first location called AGB (50.848528/1.864583- 50.8683311/1.855883), contained 43 sites and corresponded to organic agriculture soils. The second location called MEU (50.4111139/3.030917 – 50.451572/3.049203), contained 60 sites. It is a highly urbanized area located 100 km east of (AGB). This location hosted the "Metaleurop-Nord" lead smelter from 1893 to 2003, when the smelter was destroyed. Concentric circles of decreasing concentrations of Cd, Pb and Zn in soils were highlighted around the former smelter (Fritsch *et al.*, 2011). The three other locations (Coc (8 sites), Gen (12 sites) and Tour (7 sites) group together in a non-polluted area at short geographic distance north east of MEU. They correspond to agricultural areas showing a moderate level or urbanization (Figure 1).

In total, 828 earthworms were extracted from their subterranean burrows using Allyl isothiocyanate (AITC) (Zaborski, 2003). The number of individuals varied among locations (AGB=253, Meu=412, Coc=25, Gen=86, Tour=52) (Table 1). Worms have been identified with the taxonomic key of Bouche (1972) using characteristics such as the prostomium shape, the shape and the location of the clitellum (saddle-shaped on segments 32-37) and the location of the tuberculata pubertatis (segments 33-36). In addition to this identification based on morphological characters, 321 randomly taken samples were chosen to ensure their taxonomic status using COI barcoding. After identification, earthworms were rinsed and conserved in ethanol 96%. Using NucleoSpin® 96 Tissue (Macherey Nagel) kits, which permit the purification of multiples of 96 samples, total genomic DNA was extracted from a small section of earthworms' tails (Souleman *et al.*, 2016). This DNA was used for 1:20 dilution.

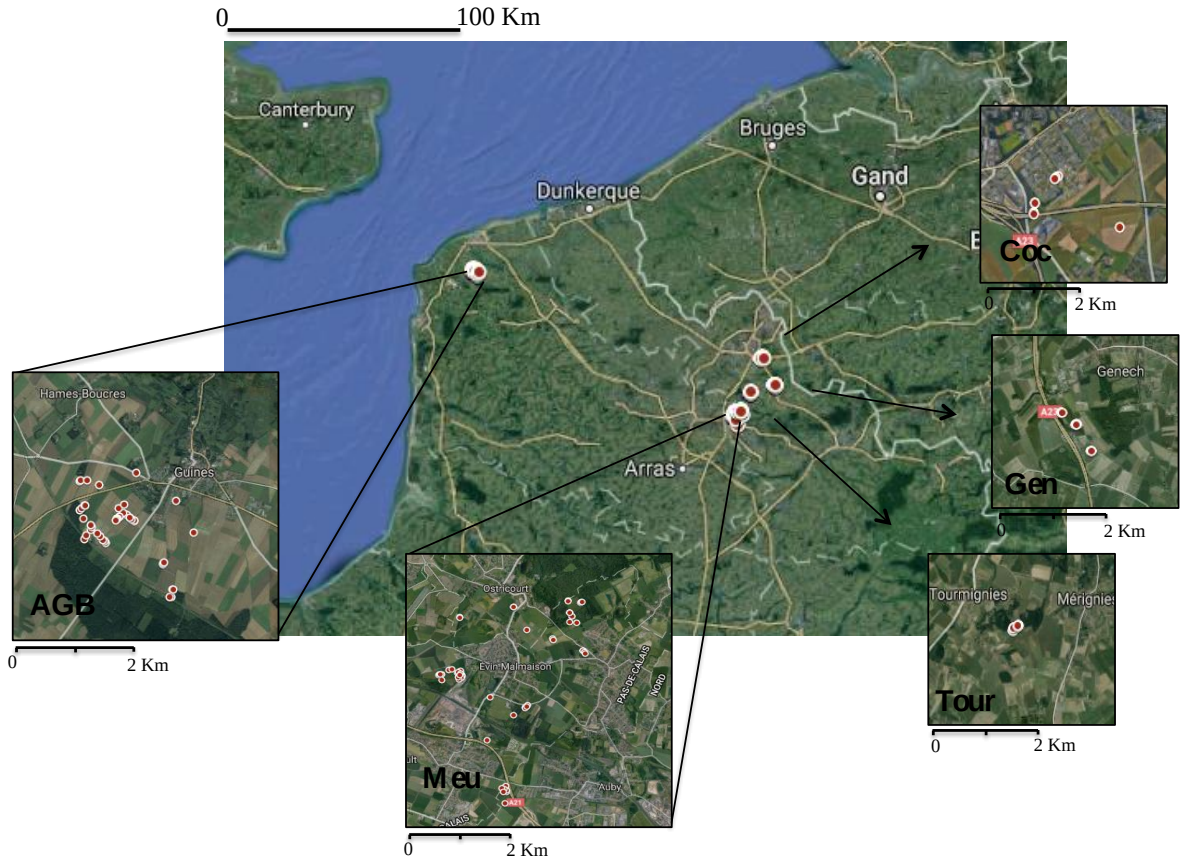


Figure 1 : Geographical distribution of sampling locations of *L. terrestris*. AGB (organic agriculture soils), Meu (highly urbanized area), Coc , Gen and Tour: three locations in non-polluted areas.

2.2 COI sequencing

COI sequencing was used to barcode 321 samples randomly chosen from the *L. terrestris*, 145 sequences were cited in (Souleman *et al.*, 2016). Briefly, COI was amplified using universal DNA primers LCO1490 and HCO2198 (Folmer *et al.*, 1994). However, in order to ensure that amplification failure was not due to primer-template mismatches, an additional amplification was performed using redesigned amplification primers (Souleman *et al.*, 2016) from the complete *L. terrestris* mitochondrial genome sequence (GenBank NC001673.1). Standard protocols were used for amplification and sequencing reactions (James *et al.*, 2010). The Millipore-Multiscreen purification kit was used for amplicon purification and PCR fragments were sequenced using the BigDye TerminatorKit 3.1 (Applied Biosystems) and run on an ABI-3130 capillary sequencer (Applied Biosystems).

2.3 COI sequences analysis

The number of COI sequences varied among locations 111, 9, 29, 18, 145 sequences for Meu, Coc, Gen, Tour and AGB respectively. Then, sequences were checked manually with CodonCode Aligner Version 5.1.4. Sequences analysis was performed as suggested in (James *et al.*, 2010). Then, sequences were aligned with MEGA 4 (Tamura *et al.*, 2007) software using ClustalW. MEGA 4 was also used to examine bases composition intra location and to calculate pairwise genetic distances for the total number of sequences employing the Kimura status. To verify the correspondence of the obtained sequences with *L. terrestris* species, 228 COI sequences were added from Genbank of Lumbricus genus identified by James et al. (2010). The phylogenetic tree was displayed using a Neighboring Joining (NJ) method, and with 1000 bootstraps. By the mean of DnaSP 4.0 (Librado & Rozas, 2009), supplemented descriptions of population genetic were performed for each location such as the number of polymorphic sites (S), number of haplotype, haplotype diversity (Hd), nucleotide diversity (π) and the average number of nucleotide difference (K).

2.4 Microsatellite Genotyping

DNA samples of *L. terrestris* were genotyped using eight microsatellite markers developed by Souleman *et al.*, (2016). For two multiplex of four microsatellite markers, PCRs were realized using QIAGEN Multiplex PCR Kit under the following conditions: denaturation step at 95°C, for 15 min, 5 cycles including (45s at 95°C, 5 min at 68 °C, and 60s at 72 °C), 5 cycles including (45s at 95°C, 1 min at 58 °C, and 60s at 72 °C), 20 cycles including (45s at 95°C,

30s at 47 °C, and 60s at 72 °C), and extra extension step of 30 min at 60 °C. Then, a mixture of 1.5µl of PCR product with 9.75µl of formamide (Applied Biosystems) and 0.25 µl of Genescan-500 LIZ size standard was done and analysed using ABI3130 Genetic Analyzer (Applied Biosystem). GENEMAPPER 3.7 software (Applied Biosystem) was used to analyse the fragments obtained and to determine genotype at each locus for each individual. Allele determination was performed according to the marker peaks' position, allowing an offset up to 0.8 bases among peaks assigned to same allele.

2.5 Genetic analysis of microsatellite data

2.5.1 Genetic diversity analysis

For our five locations and by the mean of microsatellite data for each individual, the number of alleles per loci (N_{all}), the expected heterozygosity (H_E), observed heterozygosity (H_O), and the intra-locations fixation index (F_{is}) were estimated (per locus and over locations) and (per location and over loci) using GENETIX (Version 4.05.2) (Goudet, 1995). In parallel, the average of allelic richness A_r was calculated for each location employing FSTAT (Version 1.2) (Goudet, 1995). A_r refers to the unbiased estimate of the expected allele numbers independent of the sample size. Differences in means for N_{all}, H_E, H_O, A_r and F_{is} were tested using Kruskal-Wallis H-test. Furthermore, genotypic disequilibrium and departure from Hardy-Weinberg were estimated over loci and locations using FSTAT. The number of private alleles was calculated in each location using GeneClass v2.0.f software (Piry *et al.*, 2004).
The function GENHET in R program was used to estimate five of individual heterozygosity estimates: 1) the ratio of heterozygous loci in an individual (PHt= number of heterozygous loci/number of genotyped loci); 2) standardized heterozygosity depending on expected heterozygosity's average (Hs-exp= PHt/mean expected heterozygosity of typed loci); 3) standardized heterozygosity depending on observed heterozygosity's average (Hs-obs= PHt/mean observed heterozygosity of typed loci). 4) Internal relatedness (IR). 5) Homozygosity by locus (HL) (Coulon, 2010).

2.5.2 Bayesian clustering

To infer the presence of genetically distinct clusters or gene pools, we used microsatellite data to perform Bayesian clustering analysis (BCA) using STRUCTURE 2.1 software (Pritchard *et al.*, 2000) under the admixture model. The STRUCTURE procedure uses a Markov Chain Monte Carlo (MCMC) algorithm to detect the number of genetic clusters (K) that are

consistent with some population genetic assumptions (linkage and Hardy-Weinberg equilibrium) in a group of individuals. The number of possible K was tested from K=1 to K=40, with 30 analytical replica for each K value. Burn-in duration was defined for $2x10^5$ followed by 10^6 MCMC repetitions. This analysis was performed two times, firstly with no prior information about individuals' origin and secondly taking into account the information about the sites where individuals were collected (LocPRIOR option). Moreover, BCA was designed in a hierarchical way starting with the totality of sampled individuals to determine an upper level of population structure, and afterwards a sub-analysis was carried out on resulting clusters (Figure 2). At each step, the optimal K was chosen depending on the ΔK statistics (Evanno *et al.*, 2005). CLUMPP v1.1 was used after the identification of the most likely K to calculate the mean of individuals' adhesion coefficient over replication to each cluster. Finally, DISTRUCT (Rosenberg, 2004) software was used to visualise the clustering results.

2.5.3 Population genetic structure

The package R (HierFstat) was used to estimate hierarchical F-statistic among locations and sites within locations (de Meeûs & Goudet, 2007). HIERFSTAT allows calculating the components of the variance of allele frequencies for any number of hierarchical levels, estimates F-statistics at each specified level and can test the significance of F-statistics values using permutation tests. The effects of two hierarchical levels (location, sites within location) on genetic structure were estimated (giving F_{CT} and F_{SC} indices, respectively), using 1000 permutations.

To analyse the fine-scale genetic structure within localities, a spatial autocorrelation analysis were proceed as explained in (Hardy & Vekemans, 1999) using SPAGeDi software version 1.0 (Hardy & Vekemans, 2002). The association between genetic and spatial distances allow the investigation of isolation by distance patterns. Pairwise genetic distance matrix was estimated using kinship coefficient using 10 distance classes with 10000 permutations that give 95% of confidence interval to ensure the presence of spatial genetic structure (Hardy & Vekemans, 2002). A plot of the average of pairwise genetic distance of each set of distance class (a) was displayed against demographical distance between individual (r) to visualise the spatial genetic structure. To assess the possibility of IBD partial Mantel test was executed by a permutation of the data in PASSaGE 2 software. This test is usually used to estimate the correlation between two matrixes of variable (Rosenberg & Anderson, 2011).

3 Results

3.1 COI sequences diversity

145 COI gene sequences belong to AGB location were mentioned in a previous work (Souleman *et al.*, 2016) with GenBank accession numbers (from KU888473 to KU888617). In this study, 167 sequences were added to represent the four other locations and to check the taxonomic status of sampled earthworms. 167 obtained sequences felt in the *L. terrestris* cluster identified by James *et al.*, (2010). Totally, 312 sequences belong to *L. terrestris* suggest that *L. herculeus,* a close relative and cryptic species, does not exist in the sampling area. Between locations, haplotype number ranged from 7 to 24 in Coc and Meu respectively. While, haplotype diversity showed a slight difference between locations and varied from 0.87 to 0.94 (Table 1).

3.2 Genetic diversity per locus and per location

The eight microsatellite loci used in this study were highly polymorphic (Souleman *et al.*, 2016). The level of polymorphism for 828 individuals varied between 10 and 56 alleles per locus. Averaged observed heterozygosity (H_O) ranged from 0.221 (locus Lt8) to 0.805 (locus Lt163), while gene diversity index (H_E) varied between 0.583 and 0.908 respectively (Table 2). Allelic richness estimates over loci varied among locations between (6.51 and 9.5), observed heterozygosity (H_O) ranged from 0.394 to 0.503, expected heterozygosity (H_E) from 0.69 to 0.81 and inbreeding coefficient (F_{is}) spread between 0.35 and 0.43 (Table 1). Means of genetic diversity indices H_O, H_E, A_r, N_{all} and F_{is} show slight and not significant differences among locations (P-value > 0.05 by Kruskal-Wallis test for all). The number of private alleles among locations ranged between 2 and 27 in Coc and AGB respectively. The frequencies of private allele varied between 0.001 and 0.154.

Table 1 : Over loci genetic diversity distribution for five locations of *L. terrestris* (±SD), N°Ind individuals number within location, N°site sites number within location, Ar allelic richness, Ap private alleles number, Nall alleles number, HE gene diversity, HO observed heterozygosity, Fis the fixation index, N°seq number of sequences, N° (S) number of segregating sites, N°haplo number of haplotyps, Hd haplotype diversity, Pi nucleotide diversity.

location	Microsatellite diversity								COI sequence diversity				
	N°ind	**N°sites**	**Ar**	**Nall**	**Ap**	**HE**	**HO**	**Fis**	**N°seq**	**N° (S)**	**N°haplo**	**Hd**	**Pi(π)**
Coc	25	8	6.51 ± 2.86	57	2	0.666 ± 0.17	0.394 ± 0.27	0.43 ± 0.19	9	40	7	0.94	0.025
Gen	86	12	7.38 ± 2.94	83	7	0.69 ± 0.17	0.458 ± 0.23	0.35 ± 0.11	29	42	8	0.87	0.023
Tour	52	6	9.264 ± 2.84	102	26	0.808 ± 0.08	0.503 ± 0.16	0.38 ± 0.14	18	29	6	0.84	0.014
AGB	253	43	8.21 ± 3.5	132	20	0.754 ± 0.12	0.432 ± 0.26	0.43 ± 0.05	145	62	16	0.837	0.023
Meu	412	60	9.55 ± 3.84	154	27	0.81 ± 0.1	0.465 ± 0.2	0.43 ± 0.04	107	96	24	0.087	0.02

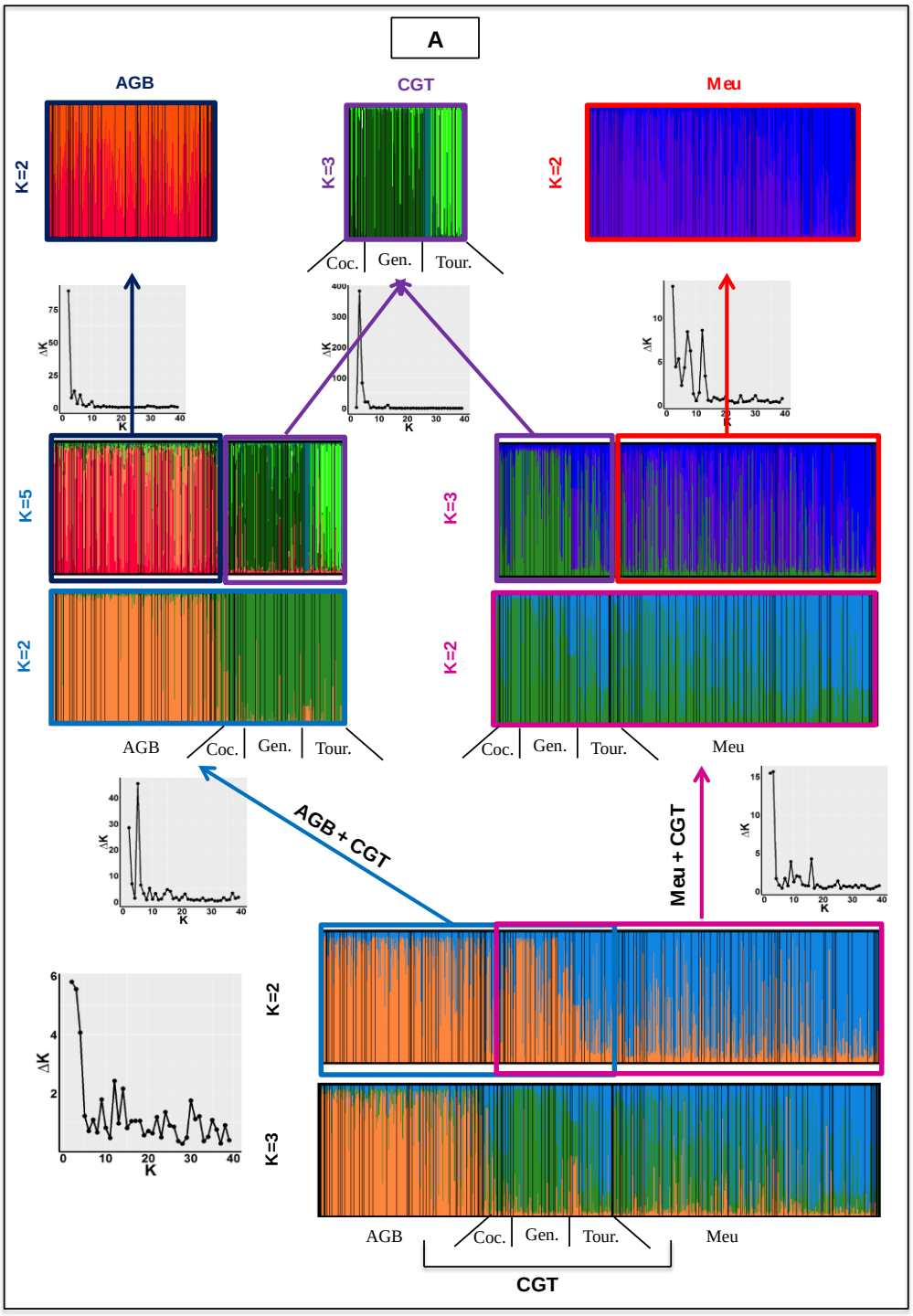
Table 2: Genetic diversity for eight microsatellite loci overall samples and alleles, (HE) gene diversity, (HO) observed heterozygosity, (Ar) allelic richness, (Nall) total number of alleles.

Locus	HE	HO	Ar	Nall	Fis	Fit	Fst
Lt12	0.791	0.453	9.8	43	0.41 ± 0.09	0.435 ± 0.09	0.027 ± 0.02
Lt163	0.908	0.805	14.35	56	0.14 ± 0.04	0.164 ± 0.041	0.023 ± 0.01
Lt17	0.875	0.686	11.12	33	0.218 ± 0.06	0.264 ± 0.06	0.06 ± 0.006
Lt9	0.729	0.264	6.96	22	0.65 ± 0.026	0.669 ± 0.018	0.058 ± 0.047
Lt16	0.707	0.373	5.5	15	0.47 ± 0.07	0.508 ± 0.06	0.078 ± 0.017
Lt19	0.583	0.526	5	10	0.23 ± 0.067	0.245 ± 0.06	0.021 ± 0.01
Lt3	0.703	0.278	6.67	19	0.69 ± 0.038	0.69 ± 0.024	0.039 ± 0.05
Lt8	0.672	0.221	6.06	22	0.69 ± 0.035	0.734 ± 0.008	0.128 ± 0.08
Over all loci					0.416 ±0.082	0.45 ± 0.082	0.063 ± 0.012

3.3 Bayesian clustering

For the total number of individuals, Bayesian clustering analysis taking or not in consideration the origin of individuals (*i.e.* with and without the locPRIOR option) both revealed that K=2 and K=3 as the most likely number of clusters (Figure 2A and 2B). At K=3, both analyses accordingly delimited clusters corresponding to three groups of samples that we named AGB, containing AGB samples, Meu, containing Meu samples, and CGT, (containing all samples from Coc, Gen, Tour localities. Nevertheless, at K=2 (highest ΔK values), results were somehow incongruent. The analysis ignoring the origin of individuals gathered AGB and CGT samples in a unique cluster, whereas taking into account the origin of individuals gathered CGT samples with Meu ones (Figure 2). This difference led us to test the genetic structure in a hierarchical way, analysing separately subsets of samples including either AGB with CGT or CGT with Meu. For the AGB+CGT subset, K=2 and K=5 (when the LOCPRIOR option is off) were the most likely numbers of clusters. K=2 separated AGB and CGT samples, whereas K=5 corresponded to a lower of genetic structure within localities, as confirmed by a clustering analysis considering AGB and CGT samples separately. For the CGT+Meu subset, K=2 and K=3 (when the LOCPRIOR option is off) were the most likely numbers of clusters. Here again, K=2 separated CGT samples from others. K=3 revealed a

lower level of genetic structure in Meu, as confirmed by a clustering analysis considering Meu samples separately.



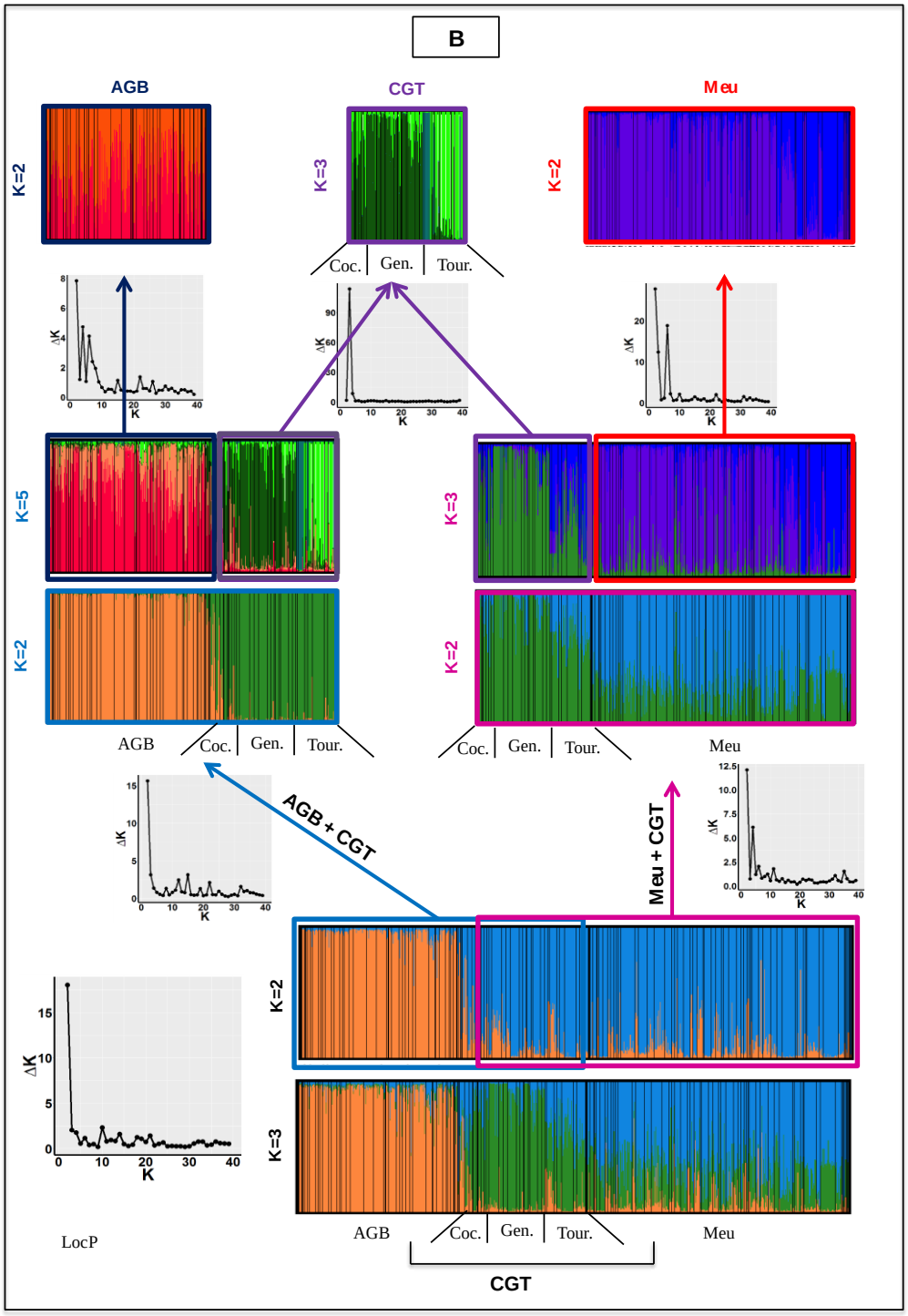


Figure 2: Genetic clustering of *L. terrestris* groups based on microsatellite data represented from the bottom up to the top, from the total number of individuals until locations, AGB (organic agriculture soils), Meu (highly urbanized area), CGT = Coc + Gen + Tour (three locations in non-polluted area). Each individual is representing by a vertical bar showing admixture degree. Each graph shows the number of cluster, which has the highest ΔK. In A) Bayesian clustering results without locPRIOR option while in B) Bayesian clustering results with locPRIOR option

3.4 Levels of genetic differentiation

Low but significant level of genetic variation was found overall samples and locus (F_{ST} = 0.063 ± 0.012, P-value <0.001). The mean of a multilocus Fit-estimate was 0.45 ± 0.082. HierFstat analysis of the partition of the genetic variance among sites within location (F_{SC}) and among locations (F_{CT}) revealed that samples are significantly structured at 2 levels, level 1 correspond to locations (AGB, Meu, Coc, Gen, Tour), where genetic variance among locations was significant (F_{CT} = 0.075, P-value = 0.001), while level 2 corresponds to sites within each location, where (F_{SC} = 0.043, P-value = 0.001). HierFstat results were confirmed by performing the same analysis using Fstat v.2.9.3.2 (pairwise comparisons among locations and among sites) and results were identical (Figure 3).

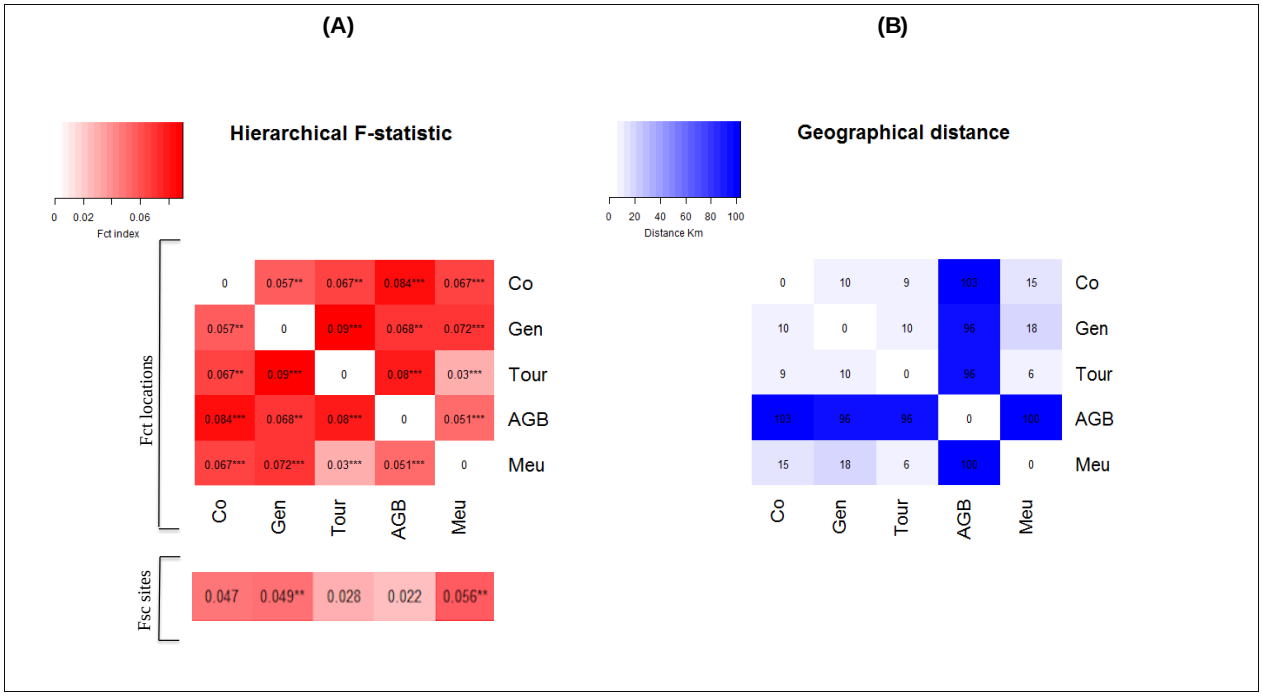
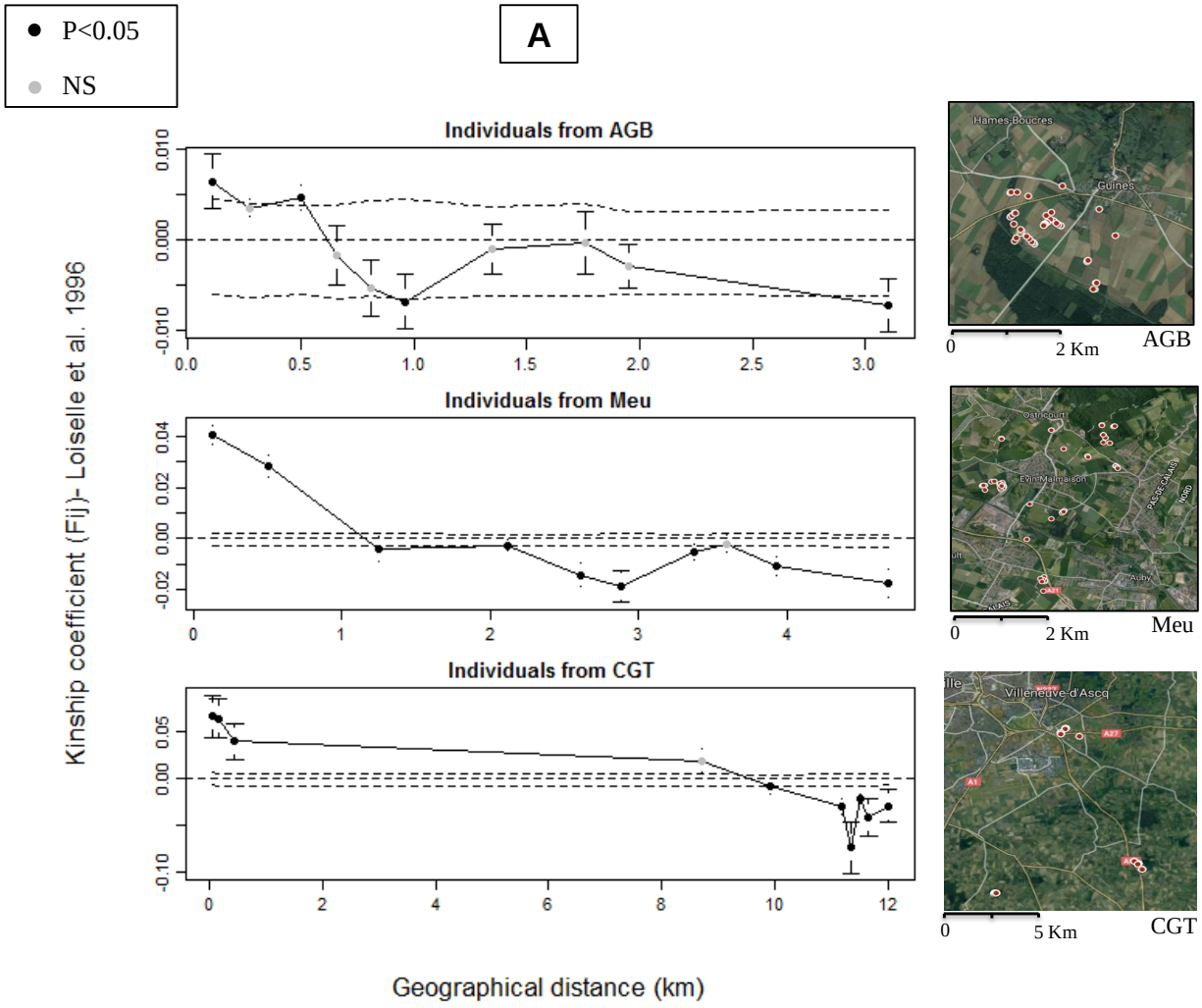


Figure 3 : Hierarchical F-statistic among and within locations (Co, Gen, Tour, AGB, Meu); AGB (organic agriculture soils), Meu (highly urbanized area), Coc + Gen + Tour (three locations in non-polluted area) presented as heatmap, In red: a matrix of Fct allelic differentiation among locations within the total of individuals, Fcs allelic differentiation among sites within location, in blue: a matrix of geographical distance among locations

3.5 Spatial genetic structure

Results of partial test of Mantel suggest a significant correlation between geographical distances and genetic structure at the scale of two groups (IBD) (r=0.077, P-value=0.001 for

Meu), (r=0.22, P-value=0.001 for CGT), but no significant correlation was detected in AGB group (r=0.024, P-value=0.2). This IBD was confirmed by an autocorrelation analysis using SPAGeDi considering 10 distance classes; all over samples SPAGeDi correlogram show significant spatial structuring at a short distance (Figure 4). The same result of fine-scale genetic structure was observed for Meu location, while no significant genetic structure was detected for the other locations.



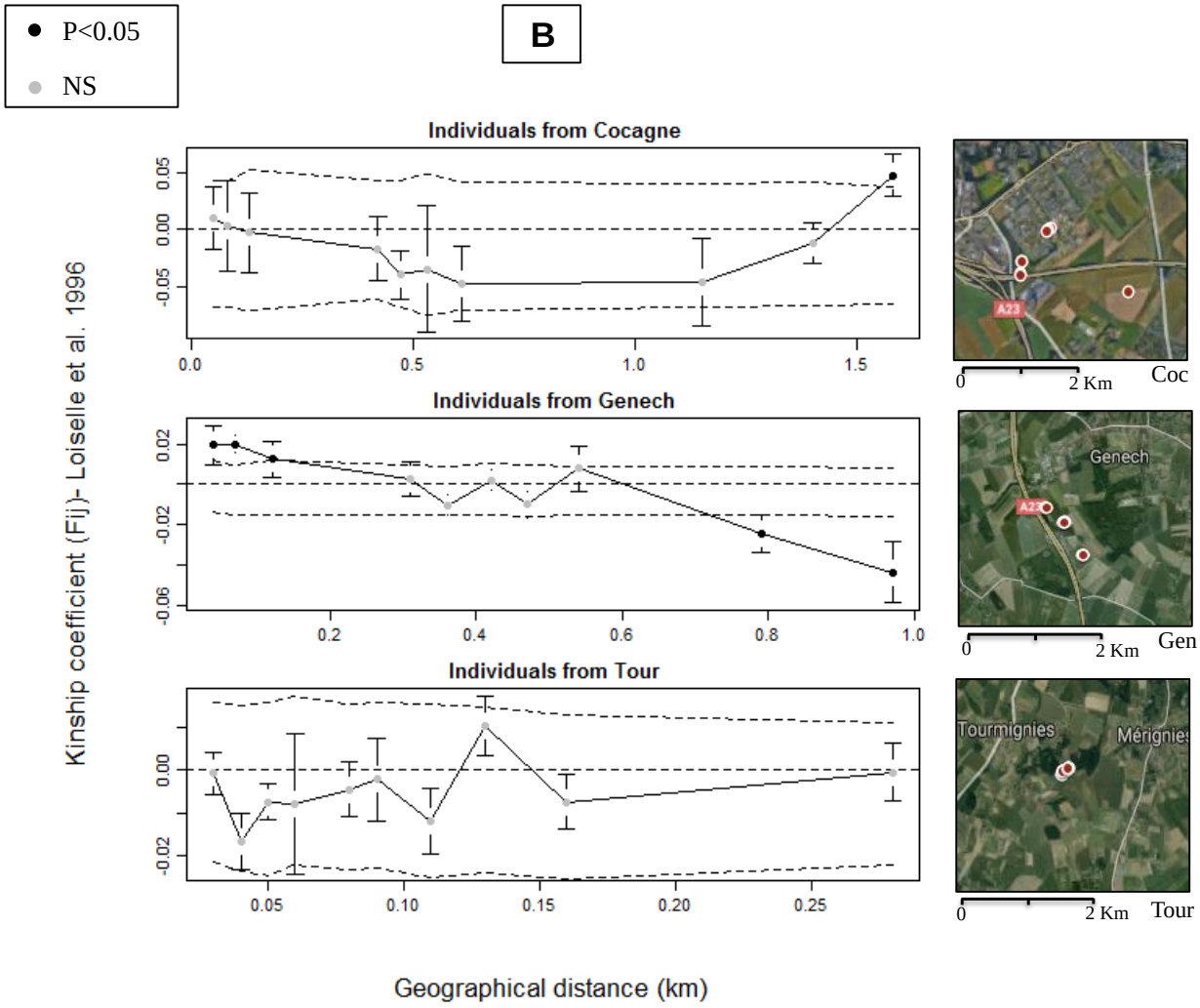


Figure 4: Correlograms of kindship coefficient (Fij)- Loiselle et al. 1996, in **A**: for AGB (organic agriculture soils), Meu (highly urbanized area), CGT = Coc + Gen + Tour (three locations in non-polluted area) in **B**: the three groups (Coc, Gen, Tour).

4 Discussion

Anthropogenic activities and urbanization can alter the quality and the connectivity of terrestrial habitats either by changing the abiotic and biotic conditions or by creating physical barriers. As a terrestrial species, *L. terrestris* can be among the species that are highly impacted by habitats modifications. The influence of habitats fragmentation can be at the level of earthworm's communities and may reach the genetic level and affect the genetic diversity of terrestrial species. The main objective of this study was to characterize the structure of genetic diversity and to determine how fragmentation can shape the structure of genetic diversity of *L. terrestris* collected from different fragmented sites. For this aim, earthworms were collected from 5 locations (CO, Gen, Tour, AGB, Meu), which represent a

gradient of fragmented and anthropogenic habitats from the most fragmented location (Meu) to the less fragmented one (AGB).

4.1 A hierarchical structure of genetic diversity

Overall, the level of polymorphism detected at nuclear microsatellite loci was high (Table 1). The number of alleles ranged from 10 to 56 alleles per locus, and was thus significantly higher than previously reported by Souleman *et al*. 2016 (between 5 and 25 alleles per locus). Not that the previous survey was spatially limited and only included individuals from the AGB location. This difference illustrates the benefit of broadening the scope of earthworm population genetic survey to get access to higher level of genetic polymorphisms. It also suggests that, at the regional scale, alleles may not be shared (private alleles), and therefore that genetic diversity is structured among populations.

Accordingly, BCA revealed that three genetic pools could be delimited from our sampling set. The signal was stronger (*i.e.* K=3 corresponded to the modal value for ΔK) when the origin of individuals was taken into account in the analysis. However, a second BCA on clusters detected at K=2 in both analyses (AGB+CGT using the default model and CGT+Meu using the LOCPRIOR model) revealed that CGT cluster was actually differentiated from AGB and Meu. This liability in the results may be explained by the level of genetic differentiation among populations at the regional scale. Indeed, although we detected significant levels of genetic differentiation among localities, F_{CT} values were small (from 0.03 to 0.09), suggesting that population structure is moderate and that private allele, likely at low frequency, marginally participate to population differentiation.

BCA performed on the three detected clusters revealed that they could actually be divided in several subclusters. For AGB, a modal, but low, value of ΔK was obtained for K=2. However, all individuals show elevated membership coefficient to both subclusters, suggesting that AGB is actually no structured. This absence of structuration was supported by the near absence of isolation by distance pattern in the analysis of the evolution of genetic correlation among individuals with geographic distance in this locality (Figure 2). For CGT, subclusters more or less corresponded to the Coc, Gen and Tour localities, although Coc appear to be intermediate. Tour was the only locality that possesses a specific subcluster that was actually almost restricted to a single sampling site. The structure of genetic diversity within CGT was supported by significant levels of genetic differentiation among localities (F_{SC}). Surprisingly, genetic differentiation among CGT localities was even as high as the levels of genetic differentiation among clusters (AGB/CGT/Meu) whereas geographic distances may be ten

times lower. For example, whereas Tour and Coc are separated from one another by about 9 kilometers and Gen and AGB are separated by about 90 kilometers, pairwise F_{CT} are similar (0.67 and 0.68, respectively, Figure 3). For Meu, two subclusters were detected, showing a rather clear geographic pattern. Indeed, the spatial distribution of membership coefficient appeared to show a North/South gradient. The assumption of spatial genetic structure within Meu was supported by the regular decreasing of genetic correlation among individuals with geographic distance in this locality.

Overall, a high F_{IS} value was noticed. This might be due to mating behaviour of the species. *L. terrestris* is a hermaphrodite earthworm but their reproduction is bi-parental by a mutual exchange of sperm between two individuals during mating. This problematic may be different between earthworm species, as some species have been suggested to be parthenogenesis (Edwards & Bohlen, 1996).

4.2 Determinants of genetic structure at the regional scale

Overall, a significant genetic differentiation was detected among sites (F_{ST} = 0.063). However, this level of genetic differentiation is rather low if we consider the geographical scale at which our study has been carried out (Max.distance = 100 Km). It was greater than observed in *Aporrectodea icterica* and *Allolobophora chlorotica* (F_{ST}= 0.018, F_{ST} = 0.014 respectively), but at smaller geographical scale (about 500m) (Dupont *et al.*, 2015). The level of population differentiation we detected was however far lower than observed in *L. terrestris* at a larger spatial scale (about 300 Km, Φ_{ST} = 0.4243) using 61 RAPD markers (Kautenburger, 2006). This moderate genetic differentiation may be due to recent evolutionary history and a common ancestry for locations that colonized new habitats by the mean of human activities or by a historical gene flow.

Overall analyses suggested that genetic structure is related to geographical distances among the three clusters (AGB, Meu, CGT). Nevertheless, this does not explain why differentiation was has high between AGB and CGT or Meu and between CGT and Meu whereas distances between AGB and CGT /Meu were higher (*ca*. 100 Km) than distances between CGT and Meu (*ca*. 25 Km). We suggest that these results could be explained by levels of urbanization, which is more important in the centre of the region than in the West. Higher urbanization could enhance genetic isolation among populations and result in higher levels of genetic structure at smaller spatial scales.

Our results show no significant differences in genetic diversity indices among the five locations where H_E ranged between 0.666 and 0.81 and H_O ranged between 0.394 and 0.503.

These results suggest that adaptation to metallic pollution within Meu, if it exists, does not cause neither loss in genetic diversity nor genetic differentiation between populations originating from contaminated and un-contaminated sites. In this case, many hypotheses are possible i) local adaptation does not exist. However, previous studies on another earthworm species such as *Lumbricus rubellus* and *Eisenia fetida* support the possibility of genetic adaptation to metal exposure (Spurgeon *et al.*, 2011; Anderson *et al.*, 2013) ii) fragmentation of habitat and the change of habitat quality due to metallic pollution does not provoke bottleneck in population sizes during the colonisation of new habitats iii) human activities through earthworms' and their cocoons transplantation at the moment of Nord-Pas de Calais smelter destruction did not affect genetic diversity of earthworms.

4.3 Determinant of genetic structure at the local scale

Analysis of spatial autocorrelations revealed a genetic structure at fine-scale for Meu location, but not AGB. Indeed, it is possible that this structure is the result of a special adaptation to the local context, which is amplified in Meu location by intensive human activities, including the presence of a water channel and the construction of the highways that led to strong fragmentation, which probably highly affect *L. terrestris* connectivity and preventing gene flow. However, in AGB that is a bio-agriculture area, anthropogenic impact is limited to agriculture activities that can less modify earthworms habitats. Apparently, this structure has no relationship with the presence of metals in the soil of Meu location. However, metal trace elements MTEs have been shown to cause a decrease of earthworm population densities (Pižl & Josens, 1995) and cause modifications in the structure of their communities (Lukkari *et al.*, 2004) that could have an effect on the structure of genetic diversity. Moreover, some studies revealed that some genes which are involved in detoxification processes of MTEs such as Cd-metallothionein gene show high gene expression in earthworms exposed experimentally to ETMs, especially Cd (Gruber *et al.*, 2000; Brulle *et al.*, 2006; Demuynck *et al.*, 2007; Mustonen *et al.*, 2014). Induction of Cd-metallothionein coding gene has also been observed in earthworms collected in metal-contaminated fields, thus exposed *in situ* (Pérès *et al.*, 2011)

Bayesian structure analysis detect sub-structure in Meu location that revealed two clusters which apparently may be due to the presence of the channel of water that hinder the movement of *L. terrestris* for reproduction.

In contrast to Meu location, no relationship between geographic and genetic distance was detected for AGB location that may be explained by heavy agricultural activities in this

region. This kind of human activities, as a passive dispersal, can help *L. terrestris* earthworms to expand their ranges. Similar results were reported for *Aporrectodea icterica* (Torres-Leguizamon *et al.*, 2014)

This genetic structure may due to the mating among relatives because of the limited dispersal of *L. terrestris. L. terrestris* is not able to colonize new habitats by itself owing to: i) a slow dispersal 4 m/year in grazed grassland (Hoogerkamp *et al.*, 1983) and 1.5 m/year in grassland (Ligthart & Peek, 1997) compared to the fastest colonizer earthworm *Aporrectodea longa* (between 5 and 8 m/year) (Eijsackers, 2011), ii) dispersal of *L. terrestris* increases with a high earthworms density and may be affected by the presence of another earthworm species such as *L. rubellus* as a competitor on food resources (Lowe & Butt, 2002), iii) as an anecic, *L. terrestris* live in deep vertical burrows. Thus, they disperse less than earthworms from another ecological categories (epigeics and endogeics). Finally, the heterogeneity of food sources distribution that affect *L.terrestis* dispersal (Grigoropoulou & Butt, 2010).

Chapter V

Genetic architecture of local adaptation to anthropogenic habitat: evidence from an F2 progeny from an inter-ecotypic cross in *Arabidopsis halleri*

The quantitative variation of traits results from the combined action of genes and environment. The approach of QTL (Quantitative Trait Loci) is commonly used to understand the relationship between the genotype-phenotype and to dissect the architecture of complex traits. Metallic pollution is a problem associated with the area of intensive industry. Only some taxa are preferentially observed on sites highly contaminated by metals, suggesting that in addition to the large phenotypic plasticity of some opportunistic taxa, evolutionary processes are involved in this adaptation to metalliferous environments. Metals tolerance and hyperaccumulation are quantitative and adaptive traits. Tolerance to metals contaminations refers to the ability of organisms to survive and reproduce on highly polluted sites without showing any sign of metals toxicity, while hyperaccumulation can be considered as a mechanism of tolerance through the accumulation of metals in plants shoots to protect both: roots from toxic effects of metals and shoots from herbivorous. The genetic basis of tolerance and hyperaccumulation have been extensively studied (Antonovics *et al.*, 1971; Baker, 1987; Macnair, 1993). All such quantitative traits, metals tolerance and hyperaccumulation are expected to be controlled by multiple genes and their interactions.

Pseudometallophyte plants such as *Arabidopsis halleri*, which survive on soil contaminated and uncontaminated by metals, provide a good model to study the genetic architecture of metals tolerance and hyperaccumulation. Pseudometallophyte plants provide the opportunity to distinguish between contrasted ecotypes (metallicolous grow on contaminated soil) and (non-metallicolous grow on uncontaminated soil). Studying the genetic architecture of traits require a crossing program between two contrasting phenotypes for these traits.

Previous studies that aimed on the identification of genetic basis of metal tolerance and hyperaccumulation were performed on interspecific crosses between *A. halleri* and its non-tolerant, non-hyperaccumulator relative *A. lyrata*. In this study and for the first time, we aimed to investigate the genetic architecture for these traits using intraspecific cross between metallicolous and non-metallocolous individuals of *A. halleri* collected from Italy to generate F2 progeny.

Table of contents for chapter 5

Genetic architecture of local adaptation to anthropogenic habitat: evidence from an F2 progeny from an inter-ecotypic cross in *Arabidopsis halleri*

Dima SOULEMAN[1*], Marie-Joe KARAM[1*], Sophie GALLINA[1], Marc HANIKENNE[2], Sol SCHVARTZMAN[2], Maxime PAUWELS[1], Hélène FRÉROT[1]

[1]Unité Evo-Eco-Paléo - UMR 8198, CNRS/Université de Lille - Sciences et Technologies, Bâtiment SN2, 59655 Villeneuve d'Ascq Cedex, France

[2]Functional Genomics and Plant Molecular Imaging, Center for Protein Engineering (CIP), Department of Life Sciences, University of Liège, Liège, Belgium

*These authors contributed equally to this work

Abstract

Arabidopsis halleri is a well-studied species for its ability to tolerate and hyperaccumulate metals, particularly zinc and cadmium. In previous studies, these quantitative traits displayed genetic differentiation between populations from polluted and non-polluted soils, suggesting that they are involved in local adaptation to metallic pollution. Deciphering the genetic architecture of these quantitative traits has only been done at the inter-specific level using *A. lyrata* ssp. *petraea* x *A. halleri* backcross and F2 progenies. The major QTL regions identified included candidate genes implied in the traits such as genes from the *hma4* (heavy-metal transporting ATPase 4) and the *mtp1* (metal tolerance protein 1) families. So far, no QTL mapping study has been done at the intra-specific level. We aimed at studying the genetic architecture of zinc tolerance and hyperaccumulation in an *A. halleri* F2 progeny (175 individuals) by crossing individuals coming from polluted and non-polluted soils. 384 single nucleotide polymorphism (SNP) markers issued from high-throughput sequencing of the genomes of parental plants have been developed and used for the genetic map construction. Growth parameters and photosynthetic yield were performed under control and polluted conditions in hydroponics (10 *vs.* 2000 µM of $ZnSO_4$). One major QTL was identified for photosynthetic yield that explained about 30% of the variation for the trait. This QTL region does not co-localize with the known candidate genes for metal tolerance that have been identified so far at the inter-specific level. This suggests that, in *A. halleri*, molecular mechanisms involved in local adaptation to metal-polluted soils at the intra-specific level differ from those that distinguish the level of metal tolerance observed in the species from those observed in its non-tolerant close relatives.

Keywords: *Arabidopsis halleri*; zinc tolerance; QTL mapping; intraspecific F2 cross; Single Nucleotide Polymorphism

1 Introduction

Anthropogenic activities are one of the main sources of Global Change. Locally, they may cause habitat fragmentation, modification or even destruction. Several factors can be responsible of habitat modification at a local scale, thus generating spatially heterogeneous environments. These factors can be biological (*e.g.* invasive species), physical (*e.g.* the increase of freshwater temperature due to industrial activity), or chemical (*e.g.* soil pollution by organic or inorganic compounds). Consequently, populations are differentially affected according to their geographical location. Persistence of populations facing new environmental conditions relies on the capacity of individuals to migrate or rapidly adjust their phenotype. Such rapid changes may result from non-genetic and genetic processes. Although some non-genetic determinants can be transmitted to the next generations, like epigenetic marks, genetic processes stay the main drivers of adaptation in the long term (Danchin & Pocheville, 2014). As a result, populations are locally adapted when resident populations outperform foreign populations in their local environment in field reciprocal transplants as well as in common garden experiments, thus revealing genotype × environment interactions (Kawecki & Ebert, 2004). However, the fact that evolution through genetic adaptation could be as fast as human activities' development is not a trivial issue (Schoener, 2011). According to these authors, rapid adaptation may depend on how ecology can affect evolution, *i.e.*, on the evolutionary dynamics of ecologically relevant traits. These traits can be relatively simple (e.g. tolerance to herbicides, insecticides or pollutants) or more complex (life history traits like, e.g. body size, fecundity, or mating behavior) (Reznick & Ghalambor, 2001), which particularly raises the question of their genetic bases. Studying genetic bases of rapid and local adaptations actually refers to the study of genetic bases of local adaptation for traits related to a specific human disturbance (if we omit the rare cases of rapid and natural environmental changes). Interestingly, several authors indicated that genetic bases of local adaptation are still largely uncovered, and also reported appropriate genomic approaches to investigate them (Savolainen *et al.*, 2007; Bergelson & Roux, 2010). According to these authors, high-throughput molecular tools have become particularly helpful for phenotype-genotype association studies requiring moderate (*e.g.* for Quantitative Trait Locus Mapping) to high (*e.g.* for Genome-Wide Association Mapping) density of markers along the genome.

In this study, we aim at exploring the genetic bases of rapid and local adaptation to a particular case of anthropogenic impact: metallic pollution. Mining and industrial activities are therefore responsible for the expansion of areas highly contaminated with zinc (Zn),

cadmium (Cd) and lead, also called calamine areas. For this purpose, we will focus on two adaptive traits that have evolved in some plant species growing on this type of soil, also known as metallophyte species. Metal tolerance is defined as the capacity of a plant to survive and reproduce on a metal-contaminated soil without showing any toxicity symptom like do other non-tolerant plants (Macnair, 1993). Hyperaccumulation refers to the capacity of plants to concentrate high quantities of metals in their shoots (Baker, 1981; AlanJM Baker *et al.*, 1999). Pseudometallophyte species are particularly interesting models for local adaptation studies because they are able to grow on both contaminated (metallicolous populations, M) and non-contaminated (non-metallicolous populations, NM) soils, often geographically close to each other (Antonovics *et al.*, 1971; Macnair, 1983). *Arabidopsis halleri* is a pseudometallophyte, which has been extensively studied for its Zn and Cd tolerance and hyperaccumulation properties (Bert *et al.*, 2000, 2002; Macnair, 2002; Pauwels *et al.*, 2006; Roosens *et al.*, 2008; Meyer *et al.*, 2015; Stein *et al.*, 2016). According to all these studies, Zn tolerance and hyperaccumulation are quantitative and species-wide properties, while Cd tolerance and hyperaccumulation are not. It is a diploid species with 2n = 16, with a genome size of approximately 250 Mbp/1C (Briskine *et al.*, 2016). It shares about 94% DNA sequence identity into coding part with its relative *Arabidopsis thaliana* (a non-hyperaccumulator and sensitive species to Cd and Zn, with a genome size of 135 Mbp) (Becher *et al.*, 2004), from which it diverged 5 Myr ago (Bechsgaard *et al.*, 2006). It is even closer to *Arabidopsis lyrata* a non-hyperaccumulator and sensitive species to Cd and Zn, with a close genome size (210 Mbp/1C) and the same number of chromosomes (Hohmann *et al.*, 2014). *A. halleri* and *A. lyrata* may have diverged 337 Kyr ago (Roux *et al.*, 2011), and can be completely intercrossed. Consequently, to ensure significant segregation of metal tolerance and hyperaccumulation, the genetic architecture of these traits in *A. halleri* was first investigated by using interspecific crosses between *A. halleri* and *A. lyrata*. *A. halleri* parental origins were either metallicolous (Willems *et al.*, 2007; Courbot *et al.*, 2007; Frérot *et al.*, 2010; Willems *et al.*, 2010) or non-metallicolous (Meyer *et al.*, 2016). These QTL Mapping analyses revealed that phenotypic variation in Zn, Cd tolerance and Zn hyperaccumulation is controlled by a restricted number of QTLs. Furthermore, candidate genes occurring in some of these QTL regions were identified by synteny with the *A. thaliana* genome and positional cloning (Roosens *et al.*, 2007, 2008) and were also functionally validated: the *HMA4* (Heavy Metals ATPase 4) and the *MTP1* (Metal Transport Protein 1) gene families (Dräger *et al.*, 2004; Hanikenne *et al.*, 2008). Remarkably, the region of *AhHMA4* tandem copies was colocalizing

with a shared QTL, occurring in all studies. This gene, which encodes a membrane protein activating root-to-shoot translocation of Zn and Cd, was thus supposed to be involved in any ancient acquisition conferring basic levels of, at least, Zn tolerance to the whole *A. halleri* species (Meyer *et al.*, 2016). These ancient acquisition may be contemporary of the speciation between *A. halleri* and *A. lyrata* (Roux *et al.*, 2011). On the contrary, the region of *AhMTP1-A* and *AhMTP1-B* genes were only revealed when using a M parental origin of *A. halleri* (Willems *et al.*, 2007), suggesting a more recent acquisition. Whether these genes, which are involved in sequestration of overtaking metals from the cytoplasm into vacuoles to maintain Zn homeostasis (Zaal et al., 1999), were selected during colonization of metal-polluted soils is probable but not clearly evidenced. When confronted, all these results suggest that the species-wide Zn tolerance basic level in *A. halleri* and the enhanced Zn tolerance level secondarily acquired in M populations would have different genetic architectures. Therefore, a direct approach, *i.e.*, an intra-specific cross between a NM and M populations of *A. halleri*, appeared timely. That is why we performed a F2 cross between a NM and M *A. halleri* geographically close populations collected from the Lombardy region in Italy. The linkage map was constructed using Next Generation Sequencing and the mapping of Single Nucleotide Polymorphisms (SNPs). The questions addressed by this study are: (1) Can we detect any significant QTL region involved in local adaptation of *A. halleri* to metal pollution through the assessment of Zn tolerance and hyperaccumulation in a controlled experiment? (2) Are these regions overlapping a QTL already highlighted in previous studies? (3) What can we infer about the process of local adaptation to anthropogenic habitats?

2 Materials and methods

2.1 Sampling sites

A. halleri is a montane species distributed throughout Europe (Northern France, Silesia in Poland, Harz in Germany and Northern Italy) (O'Kane & Al-Shehbaz, 1997; Al-Shehbaz & O'Kane, 2002). This study was carried out using genetic accessions of *A. halleri* from two different sites in Bergamo province of Italy. The first one, called I35, is located in the "Val del Riso" valley from southern Alps (N 45°55'01.6, S 9°47'40.1). In this valley, some old mines wastes and ruins of mining industries can be observed since mining activities occurred between 1850's and 1980's. I35 is situated in a mine exit (Figure 2). Consequently, the soil of this site presents high concentrations of metal trace elements. For example, the mean concentration of ammonium-EDTA extractable Zn is 11417 ± 4566 $\mu g.g^{-1}$ (from 5 soil

samples). The second site, called I30, is located in another valley about 40 km away from the first valley near the "Sommaprada" locality (N 45°59'28.03 S 10°16'19.38) (Figure 1). This site is covered by hayfields and forests and is not contaminated by metal trace elements, since the mean concentration of ammonium-EDTA extractable Zn is 22.81 ± 13.00 $\mu g.g^{-1}$ (from 5 soil samples) (Figure 2) (Decombeix, 2011).

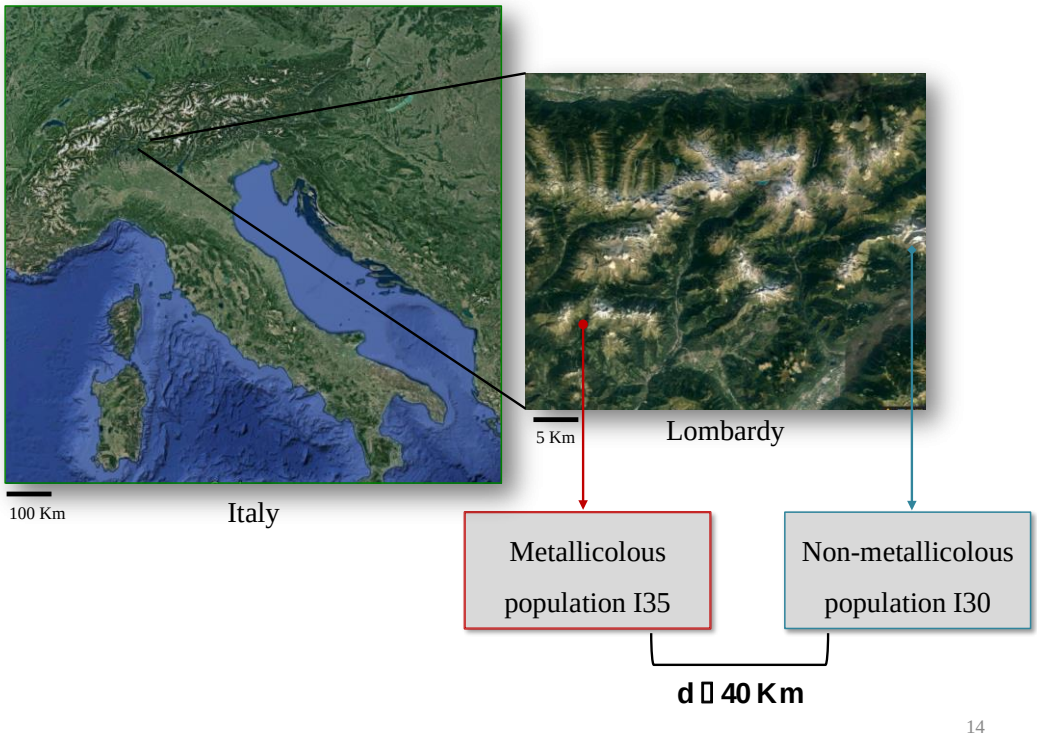


Figure1: locality of populations used to performed F2 progeny.

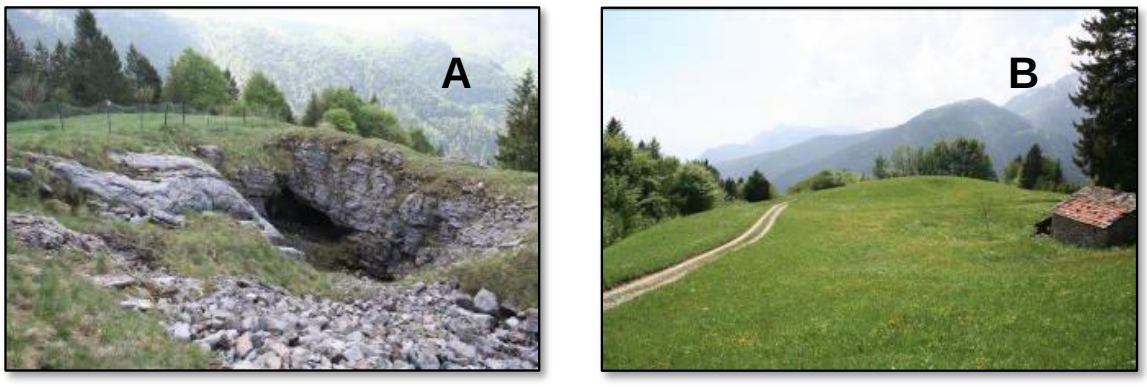


Figure 2: A: Sampling site of metallicolous population (I35), **B**: Sampling site of non-metallicolous population (I30) (Decombeix thesis 2011)

2.2 Plant material and crossing program

We aimed at producing a F2 progeny. As *A. halleri* is a self-incompatible species, two F1 progenies were first produced from two different inter-ecotypic crosses, *i.e.*, involving one M and one NM parent. The two NM parents were originating from the non-contaminated site described above (I30 site, individuals I30.13 and I30.16, used as females) and the two M parents were originating from the contaminated one (I35 site, individuals I35.6 and I35.12, respectively, used as males). Two F1 individuals, one per cross, were randomly selected and crossed to generate the F2 progeny (Figure 3). All crosses were made by hand. Finally, 177 F2 individuals were positively assigned to the right crossing parents using a multiplex of eight microsatellite markers (Godé *et al.*, 2012).

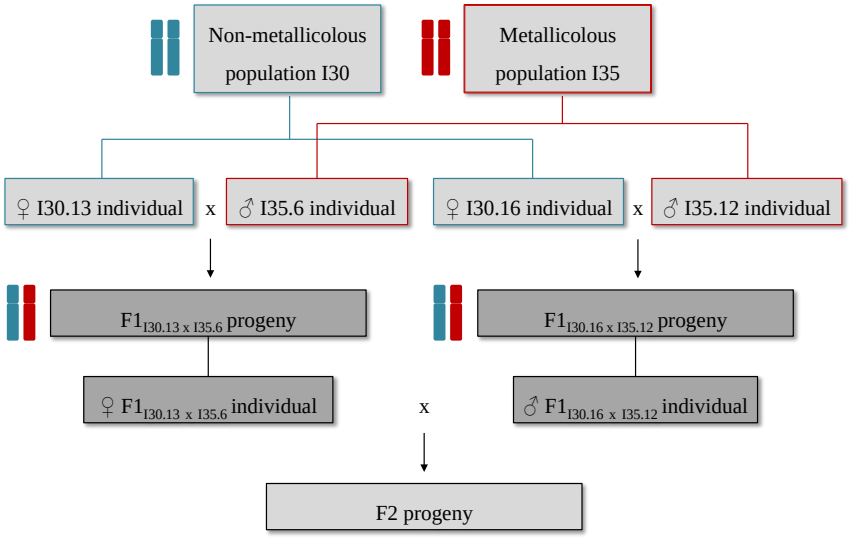


Figure 3 : *A. halleri* F2 progeny production

2.3 Assessment of Zn tolerance

Six replicates of each parental, F1 and F2 genotypes were obtained by cutting, generating a total of 1062 plants (Meyer *et al.*, 2010). Then, the replicates were grown in the greenhouse on sand during eight weeks so as to obtain rooted plants. After that, they were transferred to pots containing one liter of nutrient solution for eight weeks in a growth control chamber (temperature: 20°C day and 15°C night; light: 14 h day and 10 h night; hygrometry: 80% for the first two weeks of acclimation and then 65%). Plants were randomly distributed among pots (three plants/pot) and pots were randomly placed on three rotated tables to homogenize culture conditions (1062 plants in 354 pots in total) (Figure 4).

The nutrient solution contained essential micro and macro-elements for plants to grow including 0.02 mM FeEDDHA, 0.1 µM $(NH_4)_6Mo_7O_{24}.H_2O$, 0.1 µM $CuSO_4.5H_2O$, 0.025 mM H_3BO_3, 2 µM $MnSO_4.H_2O$, 1 µM KCl, 0.1 µM NaCl, 0.5 mM $MgSO_4.7H_2O$, 1 mM $NH_4H_2PO_4$, 2 mM $Ca(NO_3)_2.4H_2O$, 3 mM KNO_3 and 10 µM of Zn added as $ZnSO4.7H_2O$. Metal bioavailability was insured using a buffer solution of 2 mM of MES (2-(N-morpholino)ethanesulfonic acid) whose pH was adjusted to 5 using a KOH solution. To provide these elements regularly for plants, the nutrient solution was changed each week for all pots. For the acclimation phase, replicates were grown in the same dose of Zn (10 µM) for two weeks. Then, the Zn tolerance test started by assigning three replicates per genotype to either of two conditions of Zn concentrations: 10 µM as a non-polluted condition and 2000 µM as a polluted condition (Meyer *et al.*, 2010).

The Zn tolerance test lasted six weeks (week 0 to week 5). Zn tolerance was estimated by measuring five phenotypic traits for all individuals: biomass-related traits that are root length (three measures at weeks 0, 2 and 4, represented by t_0, t_1 and t_2, respectively), mean leaf width (three measures at weeks 0, 2 and 4, represented by t_0, t_1 and t_2, respectively), root dry biomass and shoot dry biomass, and one physiological trait that is mean photosystem II yield ("Φ_{PSII}", three measures at weeks 2 and 4, represented by t_1 and t_2, respectively). Plant shoots and roots were harvested at the end of the experiment, cleaned using osmosis water, oven-dried at 60°C for 48 h and weighed.

For each replicate, tolerance indices were computed by dividing the phenotypic value in the polluted condition by the median of the phenotypic values of the three replicates in the non-polluted condition. Then, for each genotype, the phenotypic measures in polluted and non-polluted conditions and tolerance indices were averaged across the replicates (Meyer *et al.*, 2010).

The broad-sense heritability (H^2) of Zn tolerance that is the ratio of the total genetic variance to the total phenotypic variance was estimated using the mean square values (*MS*) from the ANOVA following the formula for cloning data based on three replicates (Lynch & Walsh, 1998):

$H^2 = \frac{(\mathrm{MSgenotype} - \mathrm{MSerror})/n}{(\mathrm{MSgenotype} - \mathrm{MSerror})/n + \mathrm{MSerror}}$; where "*n*" is the number of replicates, here $n = 3$.

Thus, only individuals with available phenotypic data for the three clones were considered (*i.e.*, from 136 to 161 individuals according to the trait and condition considered, 153 individuals on average).

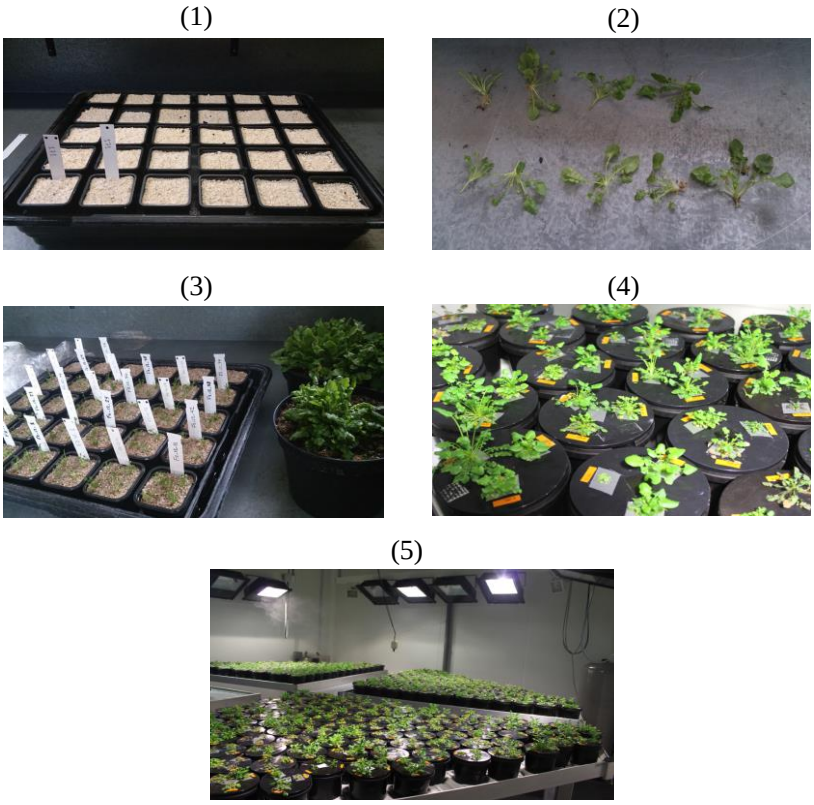


Figure 4: Experimental design to evaluate Zn tolerance.

2.4 Assessment of Zn hyperaccumulation

We performed the same cutting program as for Zn tolerance, for each parental, F1 and F2 genotypes to obtain six replicates per genotype. Replicates were grown on compost in the greenhouse for eight weeks so as to obtain rooted plants. Then, the plants were transferred in 1L-pots with 650 g of fresh compost. They were subjected to two Zn concentrations, with three replicates per genotype per concentration. The first concentration was set to 10 $mg.kg^{-1}$ of Zn (non-polluted condition). The second concentration was set to 1500 $mg.kg^{-1}$ of $ZnSO4.7H_2O$ (polluted condition) (Frérot *et al.*, 2010). As we wanted to assess the hyperaccumulation plant capacities in healthy conditions, the concentrations in the non-polluted and polluted conditions were chosen to be moderate. Pots were distributed randomly into three balanced blocks. After six weeks, shoots of each replicate were harvested and rinsed with demineralized water and dried at 55°C for three days, and then shoot dry biomass was weighed. For each replicate, shoot Zn concentration was measured using the zincon method developed for *A. halleri* by (Macnair & Smirnoff, 1999). It is a UV-visible spectrophotometry-based method that uses zincon as a colored Zn-chelating agent. For that,

25 mg of ground leaves were digested in 750 µl of a 2% sulfosalicylic acid solution during 24 h. Four microliters of the resulting solution were mixed with 40 µl of 0.03% zincon solution and 156 µl of a buffer at pH 9.6. Absorbance values were measured at 606 nm on a microplate absorbance reader (SUNRISE Tecan V 3.17, Grödig, Austria). Shoot Zn concentration was expressed in $mg.kg^{-1}$ of shoot dry weight. The broad-sense heritability (H^2) of leaf Zn concentration and shoot biomass was computed as described above.

2.5 High throughput sequencing and reads processing

DNA was extracted from the four parental and two F1 individuals of the cross described above, followed by RNAse treatment and ethanol precipitation. The six individuals were paired-end sequenced (2 x 100 bp) on three lanes of an Illumina HiSeq 2000 using one tag per individual. Library preparation and sequencing were done at the GET PLAGE platform (http://get.genotoul.fr/).

Reads were de-multiplexed according to the six tags. The *PRINSEQ-lite* 0.20.4 tool was used for reads processing: poly-N, poly-A and poly-T tails were trimmed at the 5'- and 3'-ends of the reads with a minimum length of 1 for poly-N tails and 5 for poly-A tails and poly-T tails. Reads were trimmed by quality score from 3'-end with a quality threshold of 20, and they were filtered out if their average score was below 25 or if they included characters other than A, C, G, T or N. Low complexity sequences were filtered out with the *entropy* method with a minimum allowed value of 70. Reads with less than 80 nucleotides were filtered out. Adaptors were removed using *cutadapt* (Martin, 2011). Pre and post processed *FATSQ* files (read files) were quality controlled using *FastQC* (v0.11.4, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

2.6 SNP and genotype calling

After read processing, only paired reads of the four P1 and two F1 individuals were mapped against the *A. halleri* assembly (version 1 PL22 polish individual) constructed by Sol Schvartzman and Marc Hanikenne at the University of Liège in Belgium. Mapping was done using the *Bowtie short-read aligner* version 2-2.2.4 (Langmead *et al.*, 2009) with the *sensitive* mode, the *end-to-end* option and a fragment length for valid paired-end alignment comprised between 0 and 1000. *Bowtie* alignments were piped to *SAMtools* version 1.2 (Li *et al.*, 2009) for reformatting into BAM files, sorting, indexing and PCR duplicates removal. We used

QualiMap (v.2.0.2, (García-Alcalde *et al.*, 2012) to obtain statistics on the BAM files (*i.e.*, number of mapped reads, depth of coverage, mapping quality and genome coverage).
We aimed at selecting 384 high-quality SNPs suitable for genotyping the F2 generation and thus for the genetic map construction. Variant and genotype calling across all the four P1 and two F1 individuals were done using *GATK* version 3.3-0 (McKenna *et al.*, 2010). Both the UnifiedGenotyper (UG) and HaplotypeCaller (HC) probabilistic algorithms were tested with a minimum base quality score of 20 (to consider a base for calling). UG considers each locus independently for variant calling and detects SNPs and INDELs separately whereas HC considers haplotypes for variant calling (it does not suppose independence between loci) and allows a simultaneous detection of SNPs and INDELs *via* a local *de novo* assembly. We used UG exclusively for SNP calling.

2.7 Assessing Mendelian violations

SNPs with two or more alleles common to UG and HC were used to evaluate the Mendelian violation between the parents and F1 individuals (2,970,216 SNPs). This was done using the *SelectVariants* function from *GATK* version 3.3-0 (McKenna *et al.*, 2010) by providing a pedigree and indicating a minimum genotype quality threshold of 100 (risk of false genotype $\leq 10^{-10}$) that all members of a trio must have in order for a SNP to be accepted as a Mendelian violation.

2.8 SNP filtering

We applied filtering steps on the SNP issued from the UG and HC algorithms, based on different quality scores, genotyping and mapping requirements. Only bi-allelic SNPs with long enough, polymorphism free flanking regions (50 bp from each side with no N-nucleotides) and no missing genotypes among the six individuals were kept. SNPs were also selected according to their suitability for genetic map construction, *i.e.*, SNPs that showed at least one ecotype-specific allele (metallicolous *vs.* non-metallicolous) in parents, clearly transmitted p to the F1 individuals. This implies that selected SNPs had a Mendelian segregation. Finally, only SNPs that were common between the UG and HC algorithms were kept (Figure 5).
According to the HC output quality scores, we applied the following filters: MQ $\geq$ 40.33 (first quantile), QUAL $\geq$ 3142.27 (first quantile), the individual variant position depth comprised between 28 (first quantile) and 100 and genotype quality $\geq$ 78 (first quantile). We applied

additional filters that take into consideration the allele bias: BaseQRankSum comprised between -1.96 and 1.96, FS ≤ 13.01 (error rate of 0.05), ClippingRankSum comprised between -1.96 and 1.96, FS ≤ 13.01 (error rate ≤ 0.05), MQRankSum comprised between -1.96 and 1.96, QD ≥ 26.83 (third quantile) and ReadPosRankSum comprised between -1.96 and 1.96 (error rate ≤ 0.05).

In our final 384 SNP dataset that we chose for genotyping the F2 progeny, we wanted to include SNPs located within or close to known metal tolerance and hyperaccumulation candidate genes. Sequences of the *Ah-mpt1* gene family (*mtp1-A1*, *mtp1-A2*, *mtp1-B*, *mtp1-C* and *mtp1-D;* GenBank accessions: FN428855, FN386317, FN386316 and FN386315: (Shahzad *et al.*, 2010) and *Ah-hma4* gene family (*hma4-1*, *hma4-2*, *hma4-3*; GenBank accessions: EU382073.1 and EU382072.1, from: (Hanikenne *et al.*, 2008) were aligned to the *A. halleri* genomic assembly using the *blastn* function from the *BLAST 2.2.28+* (Camacho *et al.*, 2009) software. Blast hits with alignment length equal to or greater than the query length and with an identity percentage equal to or greater than 95% were kept. The SNPs within or close to these matching regions were forced to be in the final 384 SNP. The rest of the SNPs were randomly chosen among those that passed all the filtration criteria.

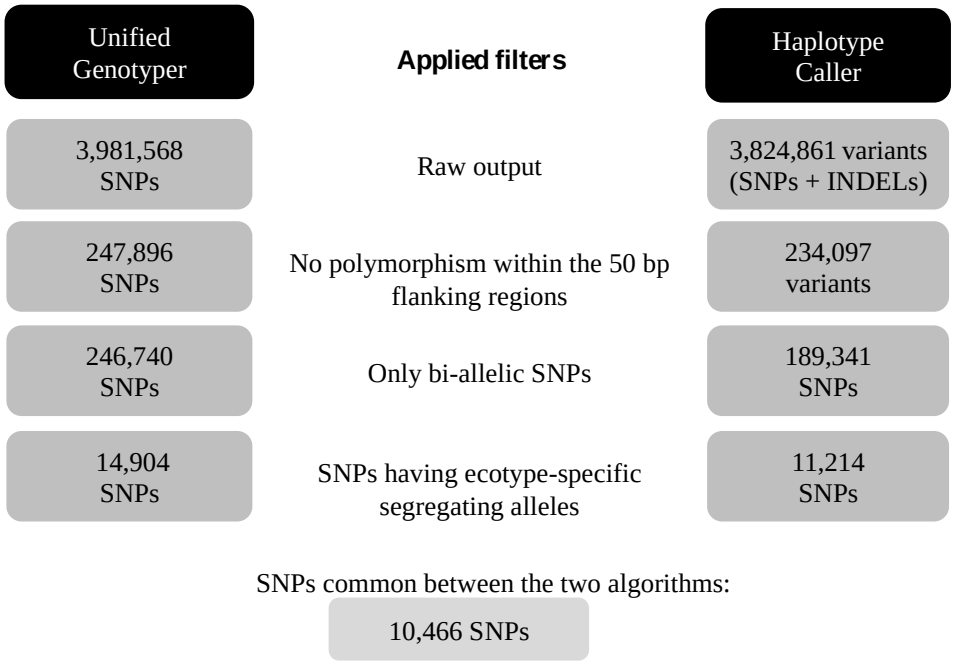


Figure 5: First step of SNP filtration

2.9 SNP genotyping

The DNA of 176 F2 individuals was extracted at the GENTYANE platform (INRA, Clermont-Ferrand, France, http://www.ibisa.net/plateformes/detail.php?tri=&srch=&q=101) from oven dried leaves using the sbeadex^{TM} maxi plant kit (LGC Genomics, Teddington Middlesex, UK) on the oKtopure^{TM} robot. The mean concentration was 8.57 ± 4.82 ng.µL^{-1} and the mean A260/A280 ratio was 1.84 ± 0.06. For genotyping, DNA was normalized to 5 ng.µL$^{-1.}$

Genotyping was performed with the KASP chemistry (LGC Genomics) using the Fluidigm dynamic array platform (Fluidigm, South San Francisco, CA, USA) at the GENTYANE platform. Plant material replicates (independent DNA extractions) and DNA replicates (a single DNA extraction) were included in the experiment for assessing the genotyping error rate which was computed as [the number of SNPs x number of genotypes that have discordant information over the replicates] divided by [the total number of individuals x number of SNPs] x 100. As individuals were grouped in two 96 well plates, the error rate was computed at the intra-plate and inter-plate levels. Genotypes were visually checked using the FLUIDIGM SNP GENOTYPING ANALYSIS software.

2.10 Construction of the genetic map

The genetic map construction was done using the *R/qtl* package (version 1.39-5, (Broman *et al.*, 2003). Before map construction, six individuals and 27 SNPs with more than 25% of missing data were discarded from the analysis. Three pairs of individuals shared more than 99% of genotype identity. We discarded one individual from each pair as they could be sample duplications. Additionally, we discarded five markers showing significant segregation distortion (Chi-square test, $\alpha_{Bonf} \simeq$ 14e-05)), *i.e.*, differing from the expected 1:2:1 ratio for an F2 progeny. Finally, 167 individuals and 352 SNP markers were included in the genetic map construction.

For linkage groups construction, the logarithm-of-odds (LOD) score minimum threshold was set to 10 and the maximum recombination fraction threshold to 0.5. Ordering markers along each linkage group was done with the greedy algorithm by fixing the positions of previously ordered markers and adding one marker at a time in the position giving the minimum number of obligate crossovers. The genotyping error rate that is taken into consideration in the genetic map construction was estimated to 0.07 by quantifying the degree of repeatability observed when comparing duplicated individuals and SNPs included in the genotyping experiment. We

used the Kosambi's mapping function to obtain the genetic distances from the recombination rates (Kosambi, 1943).

2.11 Shared synteny of *A. halleri* genomic sequences with *A. lyrata* and *A. thaliana*

Synteny is defined as the degree to which the genomic loci remain on corresponding chromosomes over time (Tang *et al.*, 2008). In order to study the shared synteny of our genetic map with the genomes of *A. lyrata* and *A. thaliana*, we extracted the scaffolds including SNPs in the genetic map of *A. halleri* from the *A. halleri* genome assembly (version 1) using *SAMtools* version 1.2 (Li *et al.*, 2009) and grouped them by linkage groups. Then, we used the *blastn* function from the *BLAST 2.2.28+* (Camacho *et al.*, 2009) software to compare the eight *A. lyrata* chromosome sequences (Hu *et al.*, 2011), Phytozome V9.0, version 1.0 (107)) and the five *A. thaliana* chromosome sequences ((The *Arabidopsis* Genome Initiative (AGI), 2000), TAIR10 genome release) to these linkage group scaffolds. The *blastn* results were filtered according to an *alignment length* minimum threshold of 5000.

2.12 QTL mapping and annotation

The QTL mapping was done using the *R/qtl* package (version 1.39-5, (Broman *et al.*, 2003) on the tolerance indices of the traits. We started by the Interval Mapping method (IM) implemented in *R/*qtl, particularly the single QTL genome scan. The LOD score representing the likelihood of a QTL has been computed every centiMorgan (cM) along the linkage groups using the non-parametric model and the expectation maximization (EM) method with 4,000 iterations. The genome-wide significance threshold for the LOD score was set for each trait using a permutation test (1,000 permutations) and corresponding to a 0.05 significance threshold (Churchill & Doerge, 1994). For the detected QTLs, an approximate Bayesian credible interval was computed, with a probability coverage of the interval of 0.95. We computed the QTL additive "*a*" and dominance "*d*" effects of the QTL using the *effectscan* function implemented in *R/qtl*. The additive effect is calculated as the half difference between the means of the phenotype values for the homozygotes for a given trait. The dominance effect is calculated as the difference between the mean phenotypic value of the heterozygotes and the midpoint between the mean phenotypic values for the homozygotes for a given trait. The degree of dominance for a QTL was computed as the absolute value of the ratio of dominance effect to the additive effect ($|d/a|$). A QTL is considered additive if $|d/a|$ is lower

than 0.2, partially dominant if |*d*/*a*| is between 0.2 and 0.8, dominant if |*d*/*a*| is between 0.8 and 1.2 and over-dominant if |*d*/*a*| is greater than 1.2 (Frérot *et al.*, 2010).

3 Results

3.1 Segregating and heritability of the traits

Eight traits were finally retained for QTL mapping: (1) leaf width at t1 - leaf width at t0, (2) leaf width at t2 - leaf width at t0, (3) root length at t1 - root length at t0, (4) root length at t2 - root length at t0, (5) shoot biomass, (6) root biomass, (7) photosystem II yield (ΦPSII at t1) and (8) photosystem II yield (ΦPSII at t2). Leaf and root growth were estimated through the non-destructive traits (1) and (2) in order to focus on the dynamics of plant response to Zn. To these tolerance variables were added two traits measured in the hyperaccumulation experiment: (1) leaf Zn concentration and (2) shoot biomass.

To illustrate the trait responses to Zn, we plotted the traits distribution between polluted and non-polluted conditions (Figure 6). Our results show a significant increase in phenotypic values for all biomass traits in response to Zn exposure. Conversely, a significant decrease in the phenotypic values of the photosystem II yield traits was observed in response to Zn exposure (Wilcoxon statistical test for paired data using R 3.1.2). Remarkably, a positive correlation between the photosystem II yield trait and biomass traits appeared in the non-polluted condition, namely for root growth, shoot biomass and root biomass. This correlation disappeared in the polluted condition (Spearman coefficient, Figure 7). A more concrete representation of tolerance indices distribution is shown in Figure 8 and Figure 9 by segregation histograms. The parental and F1 individuals died before this experiment, this is why they were not represented on the histograms. Tolerance indices varied between 0.86 +/- 5.58 and 3.42 +/- 4.22, with no significant difference from one for all of them (Table 1, Wilcoxon statistical test using R 3.1.2).

The broad-sense heritabilities (H^2) of the traits in the polluted and non-polluted conditions and of the tolerance indices were estimated and summarized in Table 1. For all traits except for the root length growth between t2 and t0, we observed an increase in the heritability values in polluted condition in comparison to non-polluted one (3.8 fold as a mean). Moreover, we also observed for all traits higher heritability values based on tolerance indices in comparison with values based on the polluted condition (4.8 fold as a mean).

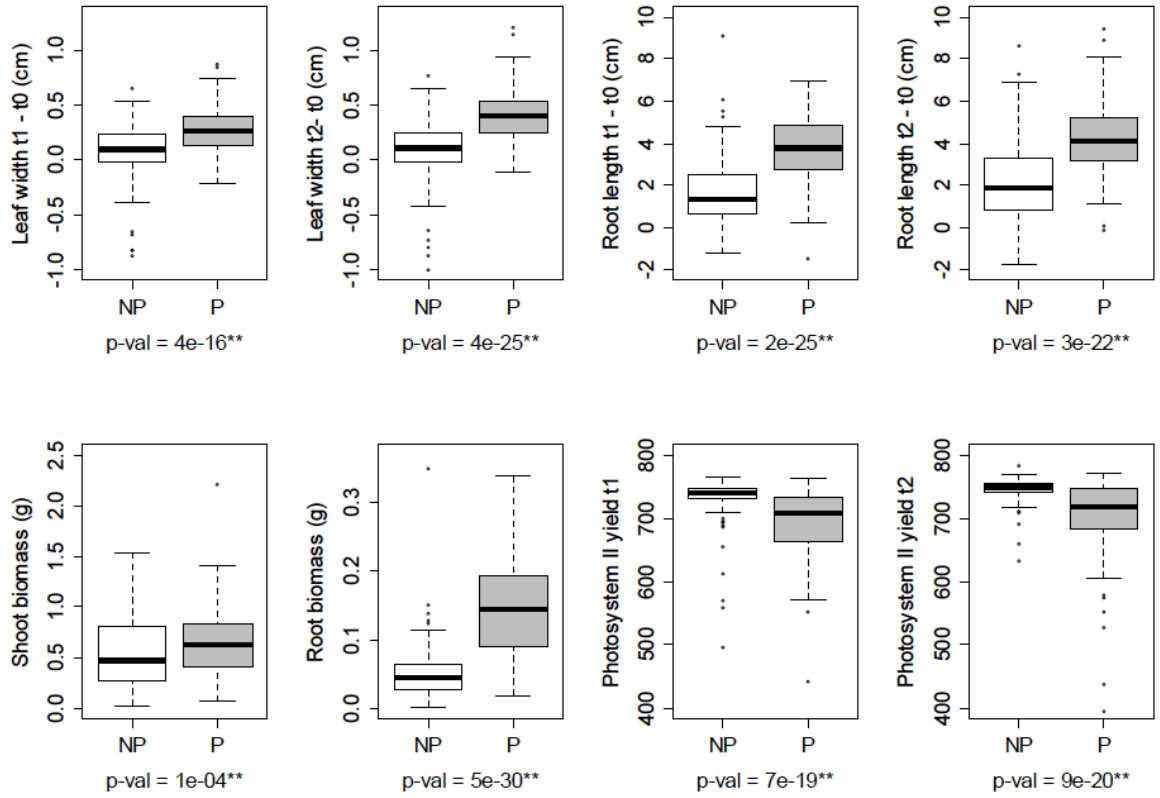
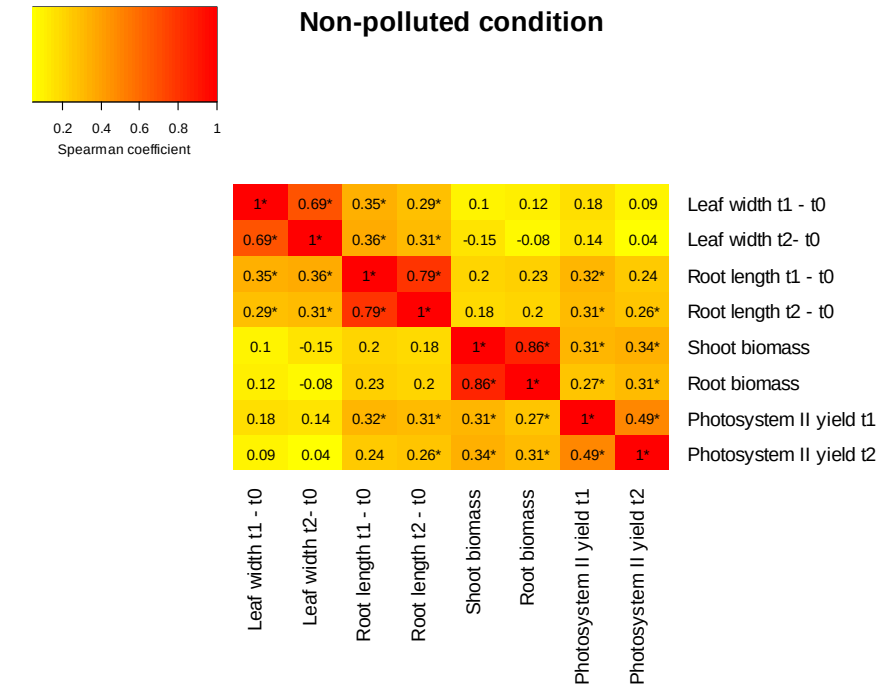


Figure 6: Traits distributions in the non-polluted and polluted conditions for all genotypes

NP = non-polluted condition; P = Zn-polluted condition.
P-values are the results of the Wilcoxon statistical test for paired data.



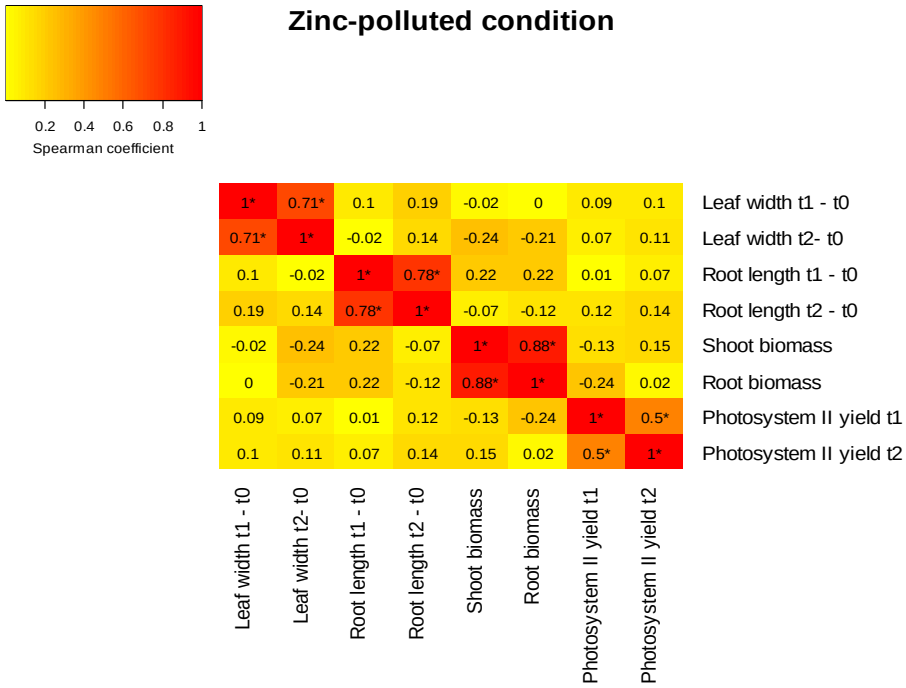


Figure 7: Correlation rates between the traits in the non-polluted and Zn-polluted conditions (α_{BONF} = 7.8e-04)

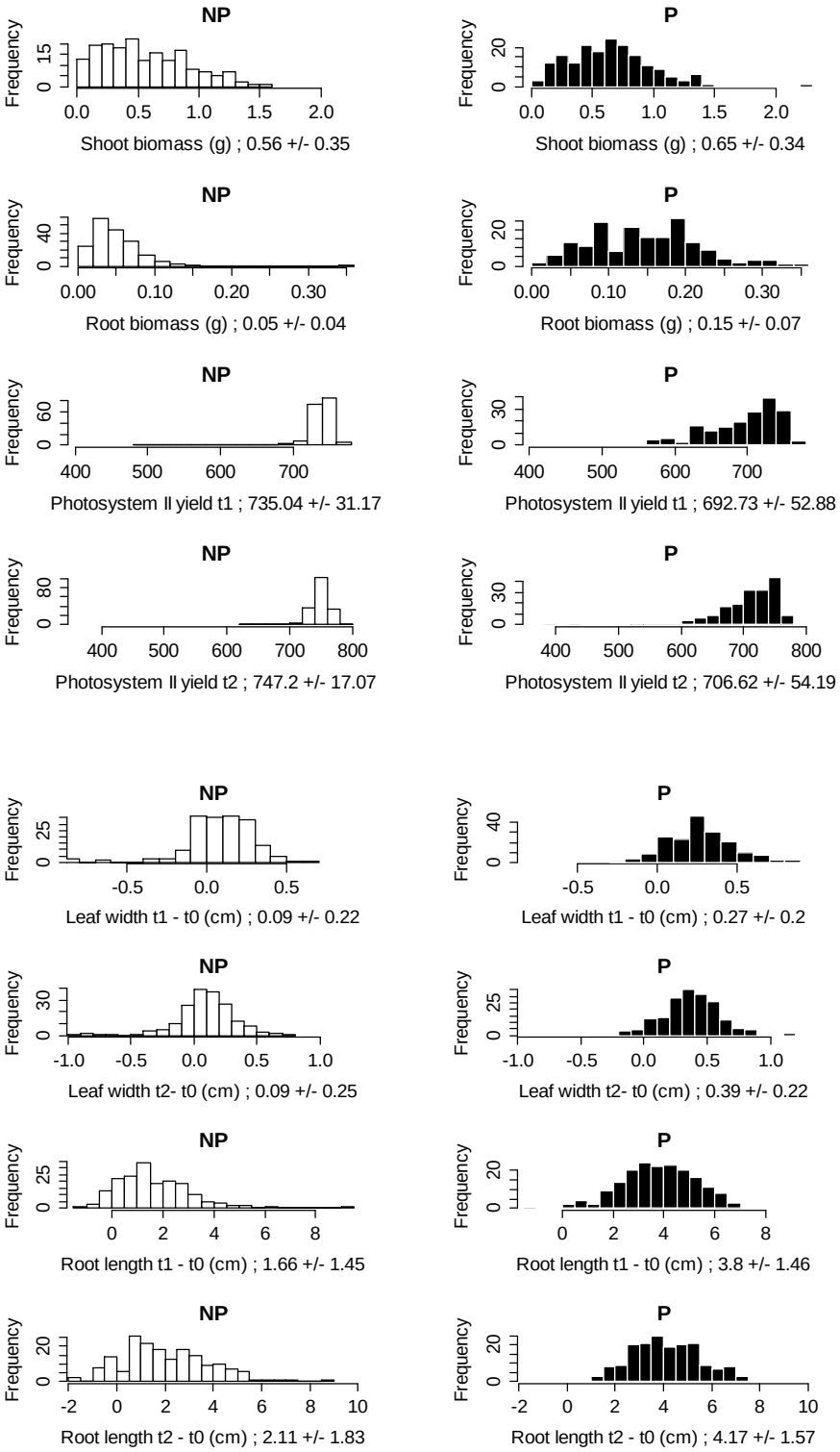


Figure 8: Traits segregation histograms in the non-polluted and polluted conditions for all genotypes. NP = non-polluted condition; P = Zn-polluted condition. Below each histogram, the trait mean +/- the standard deviations are written.

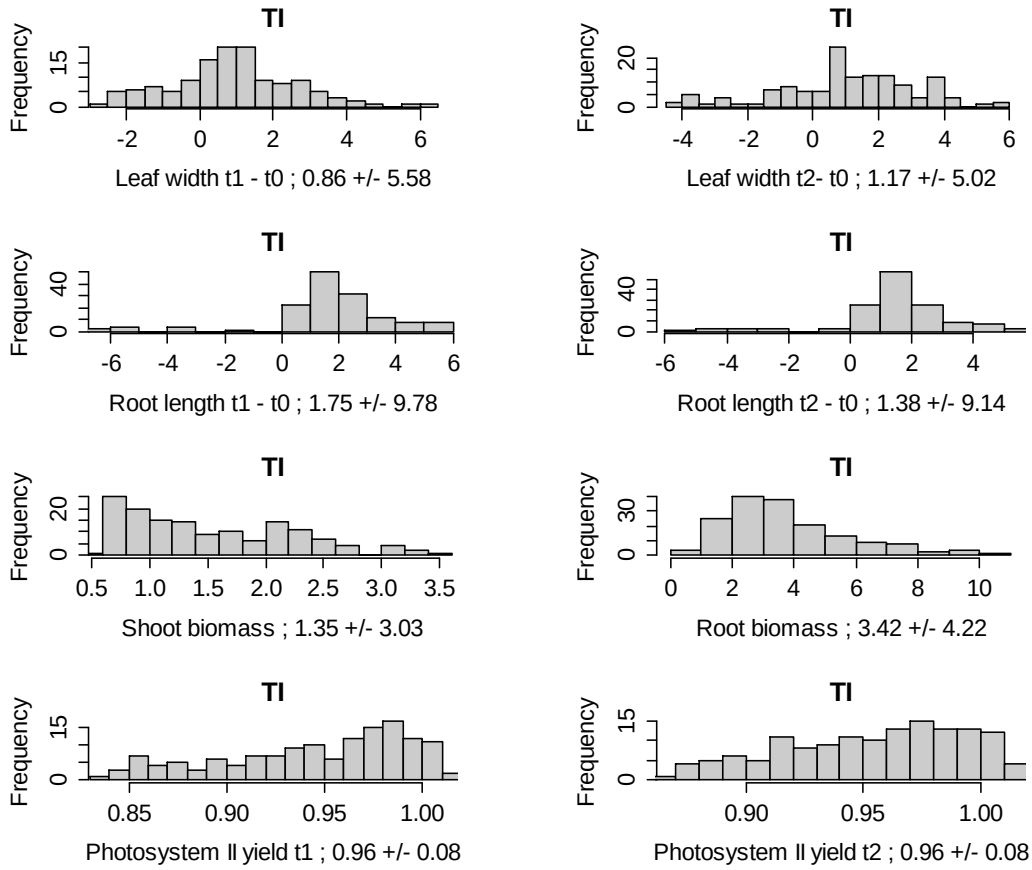


Figure 9: Segregation histograms of the tolerance traits indices from the first to the ninth decile for all genotypes

TI = tolerance index.
Below each histogram, the traits TI median +/- the standard deviation are written

Table 1: Traits, tolerance indices and tolerance indices broad-sense heritabilities (H^2)

Trait	**Tolerance index (mediane ± sd)**	***H²* in the non-polluted condition**	***H²* in the polluted condition**	***H²* of the train tolerance index**
Leaf width at t_1 - leaf width at t_0 (cm)	0.86 ± 5.58	-0.04	0.15	0.65
Leaf width at t_2 - leaf width at t_0 (cm)	1.17± 5.02	0.06	0.12	0.50
Root length at t_1 - root length at t_0 (cm)	1.75 ± 9.78	0.09	0.15	0.65
Root length at t_2 - root length at t_0 (cm)	1.38 ± 9.14	0.15	0.04	0.81
Shoot biomass (g)	1.35 ± 3.03	0.26	0.41	0.62
Root biomass (g)	3.42 ± 4.22	0.06	0.31	0.62
Photosystem II yield ΦPSII at t_1	0.96 ± 0.08	0.03	0.40	0.42
Photosystem II yield ΦPSII at t_2	0.96 ± 0.08	0.14	0.33	0.37

3.2 From reads to SNPs

SNPs were detected by mapping the paired processed reads (841,110,582 reads, representing 90.46% of the raw reads) against the *A. halleri* assembly (version 1). The average number of paired processed reads across the four parental and two F1 individuals was 140,185,097 ± 38,924,730 reads. The average number of mapped reads was 66,088,413 ± 9,071,361, depth of coverage 40.37 ± 5.54 X and genome coverage 81.10 ± 0.59% across these individuals.

We examined Mendelian violations between the parental and F1 individuals in the SNPs detected with *GATK*. 95% of the examined SNPs have a QUAL equal to or above 100 (risk of false polymorphism $\leq 10^{-10}$). None of these SNPs showed Mendelian violation confirming that the P1 individuals are the true parents of the F1 individuals.

According to the filters described in the materials and methods section, 705 SNPs were selected. From these high-quality 705 SNPs, 377 SNPs were randomly chosen and five forced to be in the dataset because they were included in a scaffold matching significantly on Ah-mtp1-C and Ah-mtp1-D genes. In addition to these SNPs, two SNPs located within a scaffold matching significantly on Ah-hma4-2 were added.

The genotyping of the F2 progeny (384 SNPs x 175 individuals) was manually checked and validated. The results showed a conversion rate (number of polymorphic SNPs over the total number of SNPs) of 93% approximately. The percentage of missing data per individual had a mean of 22.5% ± 20.5. The percentage of missing data per SNP had a mean of 22.8% ± 23.1. Plant material replicates (independent DNA extractions) gave a mean error rate at the intra-plate level of 5.2% ± 3.1 and a mean error rate at the inter-plate level of 10.6% ± 3.5. DNA replicates (a single DNA extraction) gave a mean error rate at the inter-plate level of 4.7% ± 0.3.

Table 2: Summary of the read number after processing and the mapping

**A. halleri* genome assembly (version 1)

Individual	**Number of raw reads**	**Number of paired processed**	**Number of mapped reads**	**Mean depth of coverage**	**Mean mapping**	***A. halleri* genome coverage (%)***
I30.13	140,127,836	127,847,806 (91.24%)	66,951,489 (47.78%)	40.90	23.27	80.45%
I30.16	238,894,204	212,734,468 (89.05%)	58,441,253 (24.46%)	35.70	23.48	80.53%
I35.6	109,757,778	100,200,986 (91.29%)	54,072,668 (49.27%)	33.03	23.34	81.05%
I35.12	144,921,280	132,552,426 (91.47%)	78,288,260 (54.02%)	47.82	23.21	81.00%
$F1_{I30.13 \times I35.6}$	130,752,396	119,331,600 (91.27%)	65,068,657 (49.76%)	39.75	23.45	81.70%
$F1_{I30.16x\ I35.12}$	165,403,764	148,443,296 (89.75%)	73,708,153 (44.56%)	45.01	23.65	81.89%
Mean ± SD	**154,976,210 ± 44,953,396**	**140,185,097 ± 38,924,730**	**66,088,413 ± 9,071,361**	**40.37 ± 5.54**	**23.4 ± 0.16**	**81.10 ± 0.59**

3.3 Genetic Map

With a maximum authorized missing data threshold of 25%, the map construction was conducted with 167 individuals and 352 SNPs. With a LOD score threshold of 10, 17 linkage groups were formed, from which the eight first groups carried 97% of the markers, approximately, which corresponds to the expected number of chromosomes in *A. halleri*. In these first eight linkage groups, the number of markers per linkage group varied from 29 to 61 markers. The scaffolds including the SNPs in the map has a minimum size of 1,003 bp, a maximum size of 485,200 bp, and a median size of 82,380 bp. Markers reordering and QTL mapping were carried out with these first eight linkage groups. Their size varied from 57.7 to 124.8 cM. The average spacing and maximum spacing per linkage group varied between 1.3 and 3.0 and between 9.9 and 61.4 (Figure 10).

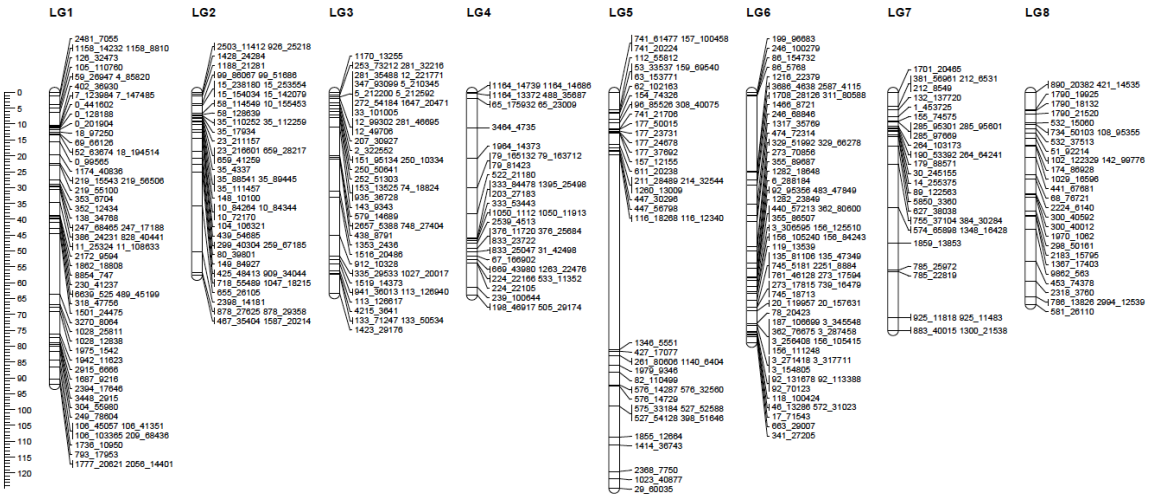


Figure 10: Linkage map of the *A. halleri* F2 cross constructed with R/qtl (version 1.39-5)

LG = linkage group.
Distances on the genetic map are expressed in cM on the left bar.
The 352 SNPs included in the map are represented by the scaffold number and the position on the scaffold separated by an underscore. The scaffolds belong to the *A. halleri* genome assembly (version 1).

The comparison of the *A. lyrata* and *A. thaliana* genomes, respectively, to our *A. halleri* genetic map showed a strong signal of shared synteny between *A. halleri* and its close relative *A. lyrata*,

where each linkage group in *A. halleri* corresponded mostly to one chromosome in *A. lyrata* (Figure 11). Shared synteny was more partial between *A. halleri* and *A. thaliana*, where each *A. thaliana* chromosome shared considerable syntenic blocks with at least two *A. halleri* linkage groups (Figure 11) as if the AtChr1 is a fusion of the AhLG1 and AhLG2, the AtChr2 a fusion of the AhLG3, AhLG4 and AhLG5, the AtChr3 a fusion of the AhLG3 and AhLG5, the AtChr4 a fusion of the AhLG6, AhLG7 and AhLG8 and the AtChr5 a fusion of the AhLG6, AhLG7 and AhLG8.

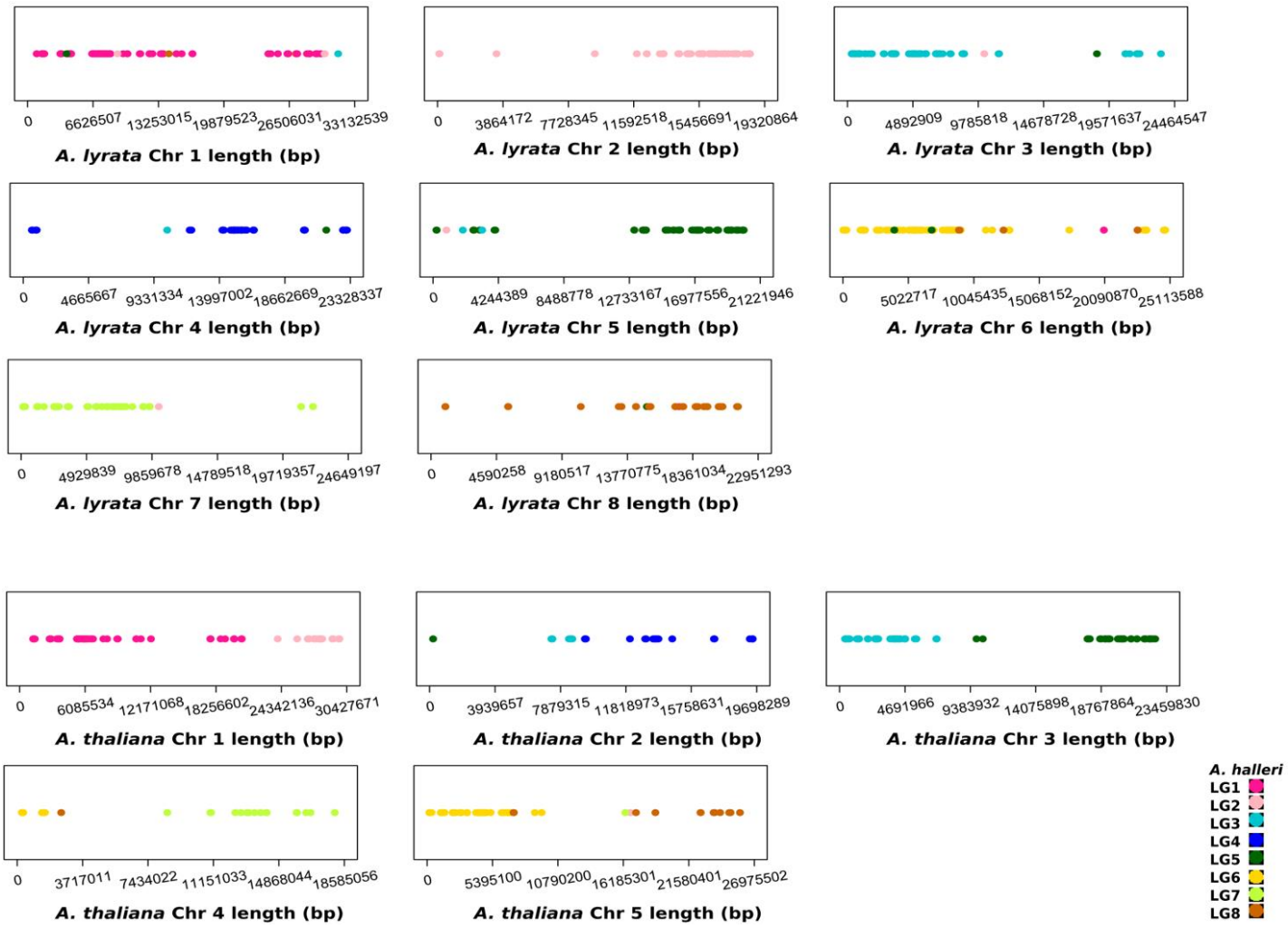


Figure 11: Shared synteny study between the *A. halleri* linkage group constituting scaffolds and the *A. lyrata* and *A. thaliana* chromosomes, respectively. LG = linkage group; Chr = chromosome.

3.4 QTL detection and annotation of the QTL region

The Interval Mapping approach for QTL detection was done with the tolerance indices of our traits. It gave one significant QTL region associated with the Photosystem II yield. We note that this QTL signal is significant in the polluted condition but not in the non-polluted one. The QTL region is located on the linkage group 4. The LOD score significance thresholds was set at the genome wide scale to 3.22 for t1 and 3.04 for t2 with $\alpha = 0.05$. The association was significant at both t1 and t2 and more pronounced at t2. At t1 (resp. t2), the maximum LOD score was of 5.96 (resp. 8.19), located at the position of 5 cM (resp. 4 cM) at a pseudo-marker position. The credible interval with a probability coverage of the interval of 0.95 was between 0 and 10 cM for t1 and 0 and 8 cM for t2) (Figure 12). The markers included in the credible interval of the QTL are: "1164_14739", "1164_14686", "1164_13372", "488_35687", "65_175932" and "65_23009". The contribution of the QTL to the phenotypic variance was 18.02% for t1 and 27.19% for t2. For this QTL, the metallicolous allele increased the photosynthetic yield (Figure 13). The additive effect of the QTL was estimated to 0.0428 ± 0.0076 at t1 and 0.0463 ± 0.0072 at t2. The dominance effect of the QTL was estimated to 0.0221 ± 0.0109 at t1 and 0.0458 ± 0.0103 at t2. The degree of dominance of the QTL was estimated to 0.52 at t1 and 0.99 at t2, so the QTL is partially dominant to dominant, as reflected by the position of the phenotypic distribution of the heterozygotes that is closer to that of the metallicolous homozygotes in Figure 13 .

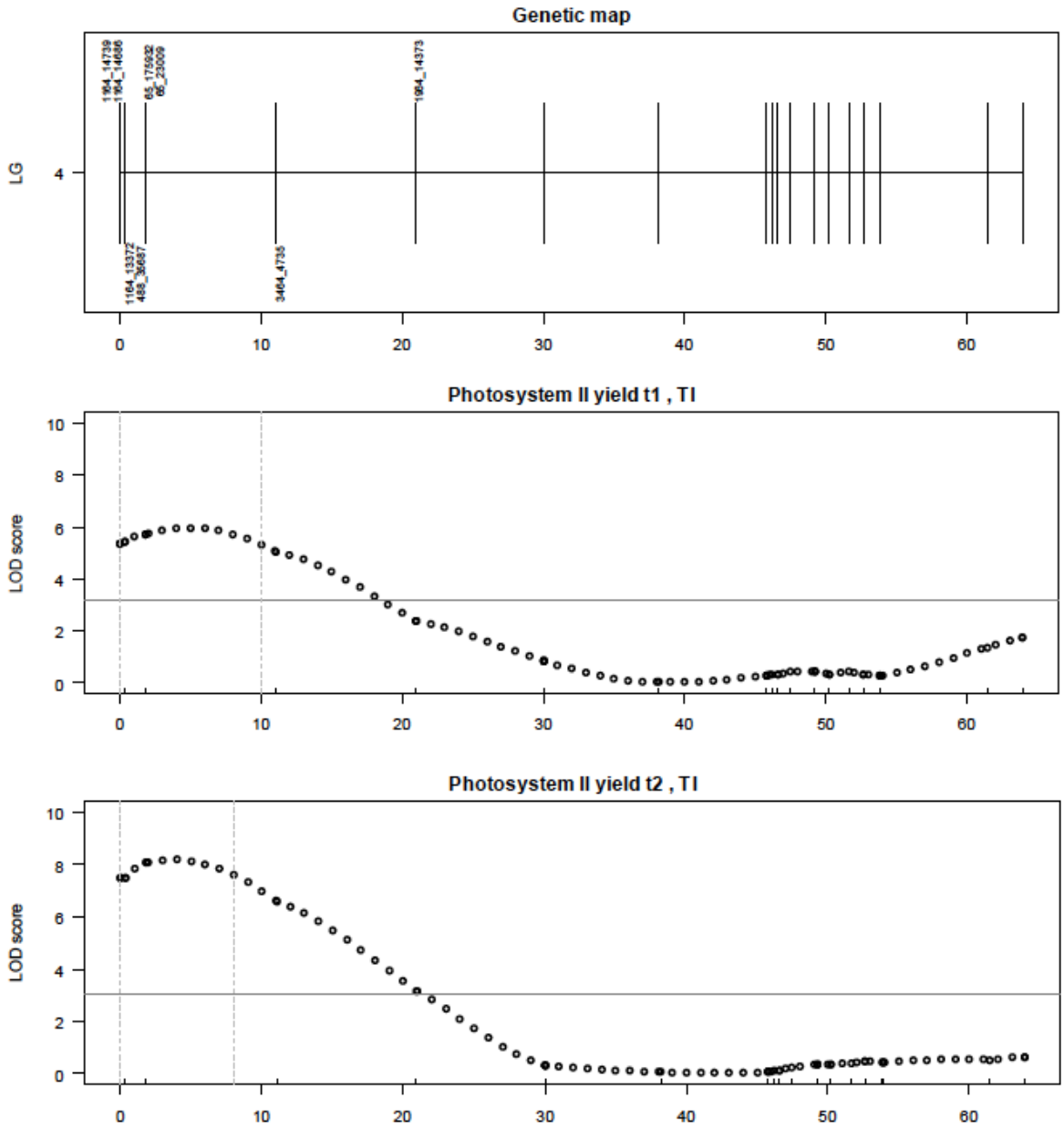


Figure 12 : QTL mapping results for the tolerance indices of the photosystem II yield at t1 and t2 by the interval mapping method

The linkage group 4 is represented in the upper part of the figure, and only the names of the first eight markers are displayed.
The horizontal lines represent the LOD score significance thresholds (3.22 for t1 and 3.04 for t2) with $\alpha = 0.05$. The vertical lines represent the limits of the credible interval with a probability coverage of the interval of 0.95 (positions 0-10 for t1 and positions 0-8 for t2).

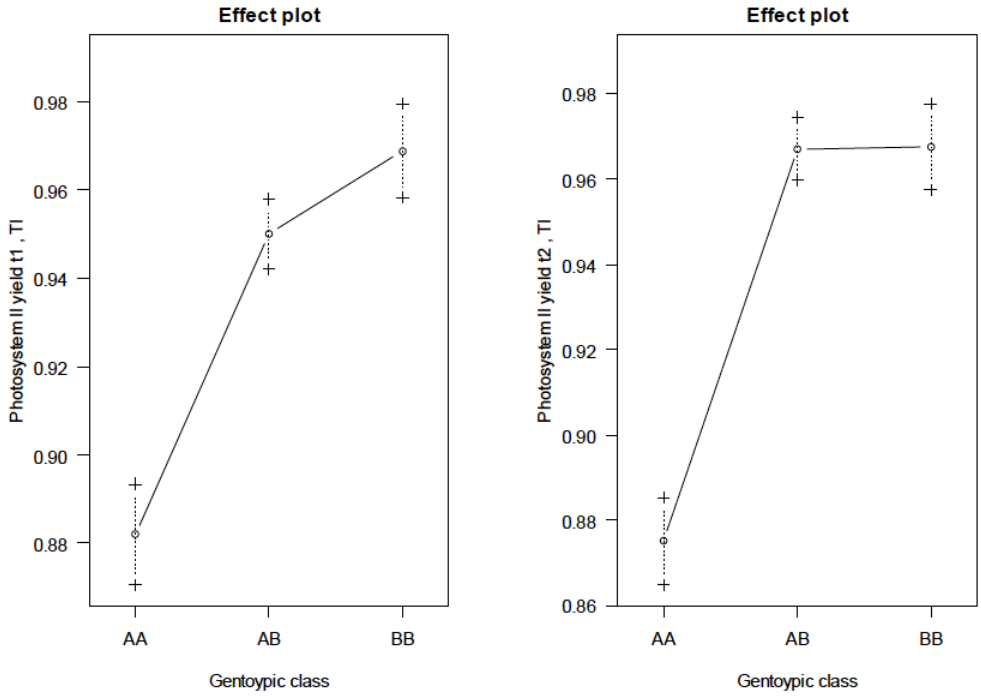


Figure 13 : Effect plots of the QTL associated with the tolerance indices of the Photosystem II yield at t_1 and t_2

The allele « *A* » is exclusively transmitted by the non-metallicolous parents and the allele « *B* » by the metallicolous parents.

4 Discussion

4.1 Response of the intra-specific F2 progeny to Zn exposure

Rapid local adaptation is a major challenge conditioning the survival of species in front of the fast pace of the environmental changes due to anthropogenic activities. In this study, we aimed at adding knowledge about the little known on the genetic architecture of local adaptation. For that, we used a QTL mapping approach on an intra-specific F2 progeny, generated by crossing parents issued from contrasted geographically close environments. We were particularly interested in the genetic architecture of local adaptation of a pseudometallophyte species *A. halleri*, to recent Zn pollution, acting since less than 200 years. We worked on two isolated Italian populations (I30 (NM) and I35 (M)) located in two environments highly contrasted by the levels of metal pollution in the soil. Previous results on M and NM *A. halleri* populations all over the range

distribution of the species (Pauwels *et al.*, 2006) and on Polish and Slovakian populations (Meyer *et al.*, 2010), have shown that in spite of the continuous distribution of the tolerance traits, we observe extreme phenotypes between the M and NM populations whose genetic bases are interesting to study. More particularly in Italian populations, (Decombeix, 2011), unpublished data) measured Zn tolerance traits in a NM population "I31" that has very similar geographic and edaphic characteristics to I30, notably the heavy metal soil concentrations such as Zn, cadmium and lead, and in I35 in common garden experiment. As expected, they showed that the I31 was less fit than the I35 population in Zn contaminated environments: they found lower leaf width, leaf length and photosynthetic yield in I31 relatively to all the studied metallicolous populations and particularly the I35 in two polluted conditions (1,000 $\mu g.g^{-1}$ and 8,000 $\mu g.g^{-1}$ of Zn added as ZnO). The geographical and edaphic similarities between I30 and I31 made us assume that there are exploitable Zn tolerance genetic differences between I30 and I35 for QTL mapping. We note that the I30 and I35 populations didn't show significant differences for Zn hyperaccumulation in common garden experiment ((Decombeix, 2011), unpublished data).

As *A. halleri* is known to be self-incompatible, two inter-ecotype crosses were done using individuals from and I30 and I35 populations to generate a F2 progeny, subsequently phenotyped and genotyped for QTL mapping. Zn tolerance was estimated based on variation in morphological traits that are leaf growth and root growth that may be directly affected by metal. Root growth was considered because this organ is in direct contact with the soil pollution. Shoot growth was considered because shoots are the storing organ after metal translocation. Zn tolerance was also estimated based on a physiological trait that is the Photosystem II yield that reflects the effect of metal on the efficiency of the photosynthesis (Meyer *et al.*, 2010). Zn hyperaccumulation was estimated by measuring Zn content in leaves. Heritability values were increased in the polluted condition relatively to the non-polluted one for the majority of the tolerance traits. We also observed a heritability increase for tolerance indices for all traits compared to the heritabilities in the polluted condition. This suggests that for an *A. halleri* genotype, variation in tolerance trait phenotypic values under Zn stress is smaller than under a non-polluted environment and that the variation of the amplitude of phenotypic response to Zn stress is even smaller. This confirms the strong genetic basis of Zn tolerance in *A. halleri*. Zn exposure caused a general decrease in the photosynthetic yield of the F2 plants while,

unexpectedly, biomass was increased with Zn exposure. These findings differ from what was observed by (Meyer *et al.*, 2010) in the same conditions as our experiment (same medium and same Zn concentrations) in Polish and Slovakian *A. halleri* populations, where the biomass-related traits (shoot and root biomass, root length and leaf width) had lower values in the polluted condition independently on the ecotype. This contradictory behavior could be related to different genetic backgrounds among the compared populations: the *A. halleri* populations from the northern range of the species including the Polish and Slovakian populations seems to belong to a phylo-geographical group that is different from that of the populations from the southern range including the Italian populations (Pauwels *et al.*, 2008, 2012). The increase in the biomass-related traits we observed could be mainly due to a biomass that is not affected by the Zn pollution in some genotypes (as observed in (Zhao *et al.*, 2000) or slightly hindered by it, namely the non-metallicolous ones, and a biomass that is favored in the presence of the metal, namely in the metallicolous genotypes. This scenario was observed in hydroponic solution in another tolerant species *Noccaea caerulescens* where with increasing Zn concentration, biomass was relatively constant in non-metallicolous populations and was increased in metallicolous ones (Escarré *et al.*, 2013). When we looked at the correlations between the traits we measured, Zn exposure seemed to break positive correlations existing between biomass and photosynthetic yield in non-polluted conditions. This suggests that Zn may differently affect the biomass-related traits and the photosynthetic yield, and that genetic bases underlying these traits are potentially unlinked.

4.2 Construction and validation of the *A. halleri* genetic map

Next Generation Sequencing (NGS) technologies are nowadays offering a fast and low-cost tool for developing a large number of DNA markers that can be used in population and quantitative genetics (Metzker, 2009; Nielsen *et al.*, 2011). In this paper, the NGS of the parental and F1 individuals allowed us to detect millions of SNPs widely distributed across the genome. We were particularly interested in SNPs because they are easy to detect, reproducible, segregate strongly among populations and give the ability to examine both neutral and under selection variations in the genome (Helyar *et al.*, 2011). The high density of SNPs we obtained allowed us to apply very stringent quality filters and mapping filters and select 352 SNPs that we used to genotype the F2 progeny and build a robust *A. halleri* genetic map with a LOD score threshold of 10. As,

expected, it consisted in eight linkage groups, which probably correspond to the eight *A. halleri* chromosomes in *A. halleri*. Our map was larger in size (622.8 *vs.* 567 and 526 cM), more dense in markers (384 *vs.* 85 and 70 markers, 2.0 *vs.* 6.6 and 8.5 cM average spacing between two adjacent markers) relatively to the *A. halleri* * *A. lyrata* genetic maps constructed by (Willems *et al.,* 2007) and (Frérot *et al.,* 2010). QTL mapping would be ideal with molecular markers homogenously distributed on the linkage groups. Nevertheless, we observed non-random marker distribution in our map reflected by large gaps between marker groups. This could reflect the bias due to the NGS read filtration and mapping and the SNP quality filtration that should prevent detecting SNPs in highly repeated or low complexity genomic regions. To assess the biological relevance of the linkage groups, we analyzed the shared synteny between the scaffolds in the linkage groups and the chromosomes of *A. lyrata* and *A. thaliana.* A strong shared synteny was observed between the *A. halleri* linkage groups and the *A. lyrata* chromosomes, with a one-to-one chromosome correspondence. Nevertheless, some punctual genomic blocks escaped this shared synteny profile. These blocks could be due to local genomic rearrangements that happened after speciation. They could also reveal errors in the genomic assemblies such as the presence of chimeric scaffolds. The shared synteny profile obtained between the *A. halleri* linkage groups and the *A. thaliana* chromosomes revealed the same chromosomic fusions in *A. thaliana* relatively to both *A. lyrata* and *A. thaliana* (Willems *et al.*, 2007; Hu *et al.*, 2011).

The *A. halleri* genetic map we constructed was made of eight linkage groups that are biologically relevant, and has a high density of SNP markers that cover both neutral and non-neutral loci. Consequently it offered a good global representation of the genome and constituted a powerful tool to proceed to the QTL mapping step.

4.3 Detection of a major QTL for Zn tolerance

Among the eight tolerance and two hyperaccumulation traits we measured, we detected one single QTL associated with the Photosystem II yield. This QTL was located on the linkage group 4 at the position of 4 cM approximately. It was detected after two weeks and after four weeks of Zn exposition, and interestingly, the QTL signal got stronger with time, confirming the robustness of the association. It appeared that it is a major tolerance QTL because it explained up to 27% of the total phenotypic variance. Given that the corresponding estimated broad-sense heritability is equal to 37%, this QTL would explain 73% approximately of the genetic variance

of the trait. However, we should be careful in interpreting the effect size of the QTL, because it may be inflated because of the limited number of markers and individuals in the experiment. In fact, it has been shown that increasing sample size and marker density increases the number of QTL detected and decreases the effect size of each QTL (Mackay *et al.*, 2009). The Photosystem II yield QTL that we detected showed a partially dominant to dominant heredity, with the metallicolous allele conferring the increased photosynthetic yield to the heterozygotes and the metallicolous homozygotes. The absence of QTLs for the other traits could be due to multiple reasons. In our case, we can exclude the absence of segregation of the traits in the F2 progeny because they showed variability in that generation. However, in spite of the segregation, the traits would not have differential distributions across the genotypic because of similar means of large variances. Moreover, we could imagine that the genetic bases of these biomass traits is much complex (integrative traits) and that they are governed by multiple QTLs with small effects that require a detection power that we didn't reach in our experiment. For Zn hyperaccumulation, the absence of QTL was expected as the Zn concentrations measured in the parent populations in (Decombeix, 2011) were similar.

Interestingly, our study corroborates the results of (Meyer *et al.*, 2010) that found a mean Q_{ST} value for Zn tolerance indices calculated from Photosystem II yield significantly higher than the mean F_{ST} value in Polish and Slovakian populations. This suggests that divergent selection may act at least on the photosynthetic yield in several regions of the species range.

4.4 QTL specificity at the intra-specific and local scale

Until recently, the genetic architecture of Zn tolerance and hyperaccumulation in *A. halleri* was studied at the inter-specific level, where crossing *A. halleri* with its non-tolerant relative *A. lyrata* was a guaranty to observe segregation in the progeny given the constitutive aspect of the trait. Deciphering the genetic architecture of the increased Zn tolerance observed in the metallicolous populations was only possible by comparing an inter-specific crosse between a metallicolous accession of *A. halleri* and *A. lyrata* (Willems *et al.*, 2007) to an inter-specific crosse between a non-metallicolous accession of *A. halleri* and *A. lyrata* (Meyer *et al.*, 2016). This comparison showed that the *AhMTP1-A* and *AhMTP1-B* genes would be implied in that increased tolerance. This work is the first to directly study the genetic architecture of increased Zn tolerance in *A. halleri* at a local scale. Our detected QTL is not located in a genomic overlapping with

previously identified QTL in *A. halleri*. This could be due to the differences in the genetic backgrounds between the previously studied populations (Auby France) and the Italian one. It would be necessary to identify the genes that are included in our QTL region that would be related to Zn tolerance and sequence these genes in addition to the *AhMTP1-A* and *AhMTP1-B* genes in metallicolous and non-metallicolous natural populations and see if we can detect signals of selection that differ according to the genetic background.

Our study showed that rapid local adaptation to anthropogenic environmental stress could occur in natural populations. Whether the dominant heredity of our detected QTL is related to this fast adaptation has to be dug out.

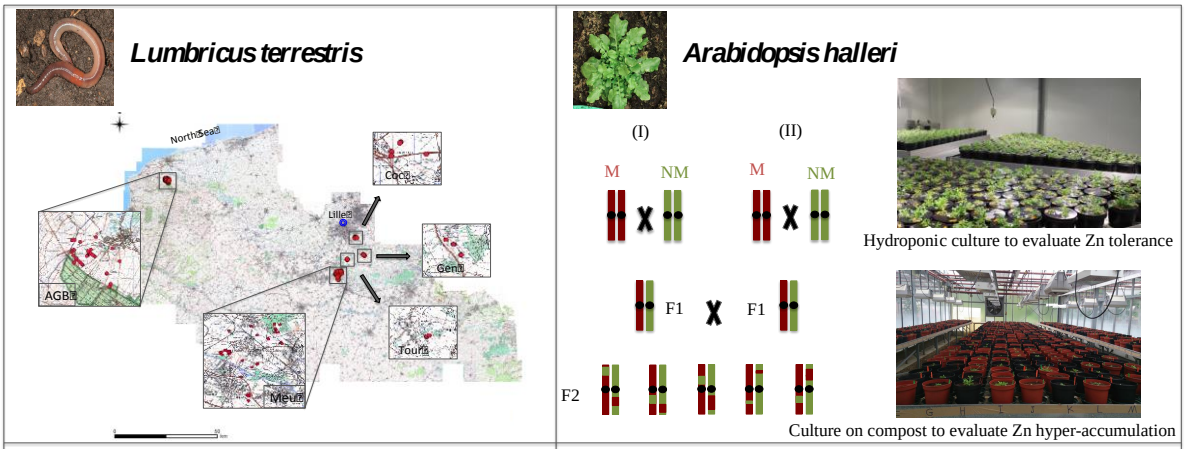
Chapter VI

General discussion and perspectives

1 General discussion and perspectives

Habitat fragmentation through urbanization and change in habitat quality due to pollution (metals, pesticides…) can affect the resident species. Since, habitats modifications are usually spatially heterogeneous and populations are differentially affected depending on their geographical locality. Permanence of populations is related to their capacity to migrate to a most favourable environment or to adjust their phenotype through adaptation or phenotypic plasticity. Adaptation requires genetic modifications while phenotypic plasticity implicates non-genetic modifications (mainly physiological plasticity). The objective of this thesis was to assess the genetic consequences of organisms' response to habitat fragmentation and metallic exposure using two approaches on two tolerant model species: a population genetics approach in the *Lumbricus terrestris* earthworm species and a quantitative genetics approach in the *Arabidopsis halleri* plant species. Population genetics approach, using microsatellite markers allowed us to evaluate the effect of the fragmentation and MTEs pollution on the genetic variability and the genetic structure of *L. terrestris* earthworm populations, while the quantitative genetics approach was used to understand the genetic architecture of Zn tolerance and Zn hyperaccumulation in *A. halleri*.

Both selected models live in close contact with the soil matrix and are highly exposed to soil modifications due to anthropogenic activities. *L. terrestris* was selected as sentinel species often used in ecotoxicology because it is found in hostile environments. Nevertheless, earthworms communities can be affected by the modifications of biotic and abiotic parameters such as agricultural practice (Chan, 2001), soil pollution by metals or pesticides (Pelosi *et al.*, 2013; Lévêque *et al.*, 2015) and urbanization (Amossé *et al.*, 2016). These modifications in earthworm's communities can affect populations' dynamics and may alter the genetic structure of these populations. The first part of this thesis was conducted to study the effect of fragmentation and the change of habitats quality on the structure of genetic diversity in *L. terrestris*. In parallel, *A. halleri*, a metal tolerant and hyperaccumulator plant, was selected because it is considered as a great model to study local adaptation in metal polluted areas given its physiological characteristic and its proximity to the model species *A. thaliana* (Figure 1).



Aims: • Isolation and characterization of polymorphic microsatellite markers • studying the structure of genetic diversity of *L.terrestris* earthworms in disturbed habitats in Northern France, Nord-Pas de Calais region	**Aims:** • deciphering the genetic architecture of Zn tolerance and Zn hyperaccumulation in *A. halleri*.
Materials and methods: • Sampling in five different locations • COI sequencing for earthworms identification • Genotyping with eight microsatellite data • Statistical analysis	**Materials and methods:** • Performing F2 cross between NM (non-metallicolous) and M (metallicolous) of *A.halleri* geographically close populations. • Experimental work to evaluate Zn tolerance and hyperaccumulation (phenotyping) • Developing of SNPs markers and genotyping • Genetic map construction and QTL mapping (phenotype-genotype association)
Main results: • Identification of eight microsatellite markers • No difference was reported in genetic diversity between earthworms collected from polluted and non-polluted area. • Hierarchical structure in two levels • Genetic structure at fine-scale in polluted and highly urbanised location comparing to less urbanised one.	**Main results:** • 177 F2 individuals, five morphological and physiological traits to estimate Zn tolerance, two traits to estimate Zn hyperaccumulation • 384 SNPs markers + Genetic map • One major QTL involved in Zn tolerance that explains 30% of phenotypic variation

Figure 1 : summary of the works presented in this thesis**.**

2 A population genetics approach to study neutral genetic diversity

Isolated populations in fragmented habitats have generally less genetic diversity than their ancestral populations because of founder effects or bottlenecks in populations' size that have lead to genetic drift. When fragmented populations are founded by a small number of individuals, those individuals will bring a sample of alleles from the set of alleles of the ancestral population. By this way, derived populations will differ by genotype, allele and phenotype frequencies from each other and from the initial population. Moreover, geographical distances and physical barriers can hinder gene flow among populations. Indeed, populations surrounded

by roads, water channels and others barriers may receive fewer migrants than other populations, which may lead to an increase in inbreeding rates (Fauvergue *et al.*, 2012). In addition to the genetic drift and reduced gene flow, natural selection due to the change of environmental conditions and the presence of contaminants such as metals can explain the loss of a part of the genetic diversity (only for adaptive genes).

In this work, we assumed that habitat fragmentation resulting from anthropogenic activities has modified the genetic structure of *L. terrestris* populations and has provoked a loss of genetic diversity. To understand how habitats fragmentation can shape the structure of genetic diversity in *L. terrestris*, we compared the genetic diversity between different groups of *L. terrestris* earthworms collected from differently fragmented sites i.e. either by urbanization and metallic pollution or by agriculture activities, using microsatellite markers. To our knowledge, our study has pioneered the exploration of the genetic diversity and its structuration in *L. terrestris* and the possible impacts of habitats fragmentation on it.

Microsatellite markers are ideal to study the genetic variation within and among populations. These markers are highly used to study several events of the demographical history of populations such as bottlenecks, founder effects and gene flows. Since those markers are not available for the studied species, this work started by identifying eight polymorphic microsatellite markers in *L. terrestris.*

On the local scale, earthworms collected from the urbanized area showed a genetic structure, which is mostly related to populations connectivity modified by habitats fragmentation but there is no evidence in favour of an effect of metallic pollution in this structure (Figure 2). This figure shows no relationship between genetic clusters obtained in samples collected from polluted area and the concentration of Cd in the soil. No genetic structure was revealed in earthworms collected from the agriculture area where fragmentation is only related to agricultural activities.

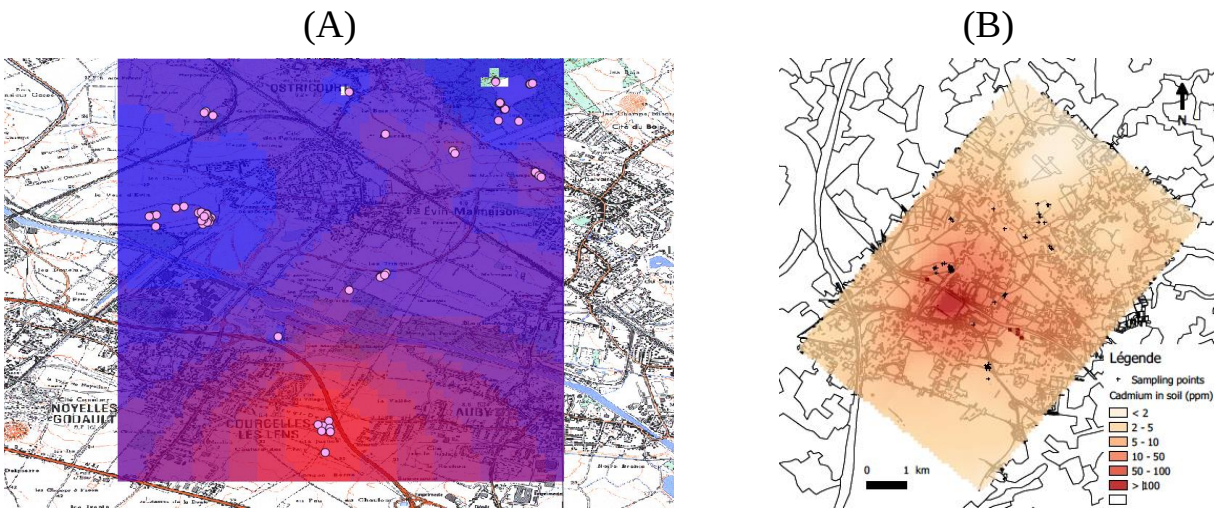


Figure 2 : In A) the distribution of two clusters obtained from samples collected in metals polluted area. The first group is in red in the south and the second one is in blue in the north. In B) cadmium concentration in the soil where samples were collected.

Our results show no difference in the level of genetic diversity between the metallicolous and non-metallicolous populations suggesting that metallic pollution did not affect the neutral genetic diversity in the *L. terrestris* ecotypes. The absence of metallic pollution effects on the neutral genetic diversity may be explained by the ecological category of the studied earthworms' species. Indeed, as an anecic earthworm, *L. terrestris* may be less exposed to metallic pollution than epigeics and endogeic earthworms, which live in closer contact with surface pollution in soil. Another hypothesis can be evoked since behavioural studies showed that earthworms could be able to detect the pollution and thus to avoid it and escape it (Capowiez *et al.*, 2003; Capowiez & Bérard, 2006).

In perspective, identification of more microsatellite markers will help to ensure the obtained results. Furthermore, large sampling of *L. terrestris* populations in different sites at large scale for example: in Europe will help to study the demographical history of this species, to characterize the genetic diversity in wide-range and could help to determine the environmental factors that influence the genetic diversity of this widely distributed and ecologically important species.

In *A. halleri*, population genetics studies were previously done to investigate the structure of neutral genetic diversity among populations collected in anthropogenic metal polluted and non-polluted sites. Genetic diversity in *A. halleri* has been studied using microsatellite markers (Van

Rossum *et al.*, 2004), RFLPs in cpDNA (Pauwels *et al.*, 2005) and using sequence level variation (Kawabe & Miyashita, 2003). Results from neutral genetic studies were similar to those obtained for *L. terrestris*, no difference in genetic diversity was detected between ecotypes (Pauwels *et al.*, 2005). Moreover, no evidence of neutral genetic divergence was reported among populations of *A. halleri* collected from a range of heterogeneous metallic polluted sites (Van Rossum *et al.*, 2004).

Even though, metallic pollution was expected to become a structuring landscape factor influencing the spatial structure of genetic diversity. The results obtained on *L. terrestris* (In Chapter 4 of this thesis) and *A. halleri* (Pauwels *et al.*, 2005)There is no evidence that metallic pollution present a barrier that hinder gene flow between metallicolous (M) and non-metallicolous (NM) populations. Moreover, habitats fragmentation related to metal contamination did not generate a bottleneck in population size that implies a reduction in neutral genetic diversity or founder effect on recolonized soils. Therefore, on metal contaminated soils, sufficient genetic diversity should be available for natural selection to act and thus for local adaptation to occur.

3 A quantitative genetics approach to study the genetic architecture of local adaptation

Particular gene or several genes normally control a specific adaptation. Identifying the genes implicated in such adaptation is a major challenge of quantitative genetics studies. The presence of *A. halleri* on calamine soils that are heavily contaminated by Zn, Cd and Pb makes it one of the best species to illustrate adaptation to metallic pollution. Therefore, many studies aimed at the identification of genomic region implicated in Zn and Cd hyperaccumulation and tolerance in *A. halleri*.

These two traits have been demonstrated to be quantitative. In particular, the distribution of Zn tolerance in M and NM populations of *A. halleri* was determined by (Pauwels *et al.*, 2006). The results confirm that Zn tolerance is a quantitative trait since populations from both polluted and non-polluted sites showed a continuous variation in Zn tolerance from the less tolerant population till the most tolerant ones, with M populations revealing higher average tolerance than NM populations (Figure 2). Zn tolerance is also a constitutive trait, since all the *A. halleri* populations showed higher tolerance levels than the relative non-metallophyte species *A. thaliana* and *A. lyrata*.

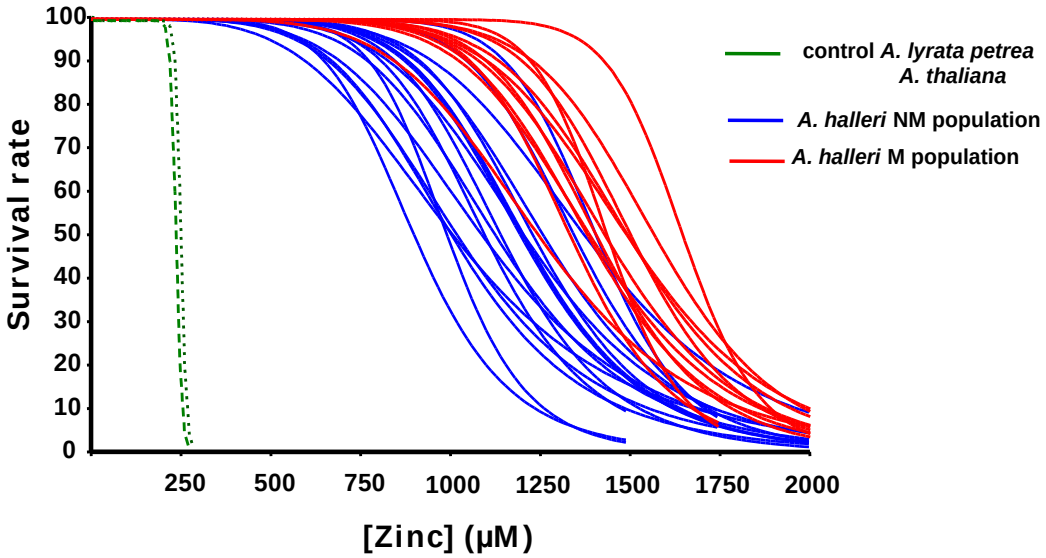


Figure 3 : population survival curves obtained from fitting to a sigmoidal dose-response model *Arabidopsis halleri* **(Pauwels *et al.*, 2006)**.

To ensure sufficient segregation in progenies, these traits were usually studied through interspecific crosses between *A. halleri* (metallicolous and non-metallicolous) and *A. lyrata* (non-tolerant, non-hyperaccumulator relative species) (Willems *et al.*, 2007; Meyer *et al.*, 2016). In this thesis and for the first time, using QTL mapping approach, we examined the architecture of Zn tolerance and hyperacccumuation using intraspecific F2 progeny resulting from a cross between individuals of metallicolous population (I35) and non-metallicolous population (I30) of *A. halleri* collected in Italy. Common garden experiment showed differences in Zn tolerance between the two populations used to realize this cross (Decombeix, 2011). We thus assumed that major Zn hyperaccumulation and Zn tolerance loci that control intraspecific variation are different from those involved in interspecific variation.

Variation of Zn tolerance was estimated through morphological traits such as root length, leaf width and shoot and root dry biomass, and a physiological trait that is photosystem II yield "Φ_{PSII}". The linkage map was constructed using genotyping data of 384 SNPs (single nucleotide polymorphisms) and then the association of genotype and phenotypes was done using QTL mapping approach. As a result, one major QTL involved in Zn tolerance was detected. The later explains 30% of phenotypic variation in photosystem yield II. This QTL has been never highlighted previously in studies performed on inter-specific crosses between *A. halleri* and *A. lyrata*. Consequently, we supposed that Zn tolerance mechanism at intraspecific level could be

controlled by different loci from those at interspecific level. One candidate gene that belongs to NRAMP family is located in the QTL region we detected. This family of genes is implicated in metal ions transport across the phagosomal and plasma membrane of cells. On the contrary, no QTL were detected for Zn hyperaccumulation (not showed results). This may be due to the similar capacity of Zn accumulation in both M and NM populations of *A. halleri* which were used to generate F2 progeny (Decombeix, 2011), or the small effect in trait variation of the loci which control Zn hyperaccumulation making them hard to detect.

The next step of this work will be the densification of SNPs makers at QTL region in order to identify the genes involved in Zn tolerance variation at intra-specific level. Since we have a candidate gene (NRAMP), we can study the relative gene expression and genomic copy numbers of this gene using some individuals of M population (I35) and some individuals of NM population (I30) cultivated on polluted soil in control conditions. In the case of *L. terrestris*, the best candidate gene widely used as a biomarker in the annelide oligochaete is the gene that encodes the metallothionein proteins (MTs) (Brulle *et al.*, 2010). These proteins are able to fix metallic ions and play an important role in metal detoxification, notably cadmium and zinc. Gene expression of Cd-methallothionein usually show differences between polluted and non-polluted conditions depending on the duration and the intensity of the exposure (Bernard *et al.*, 2010; Brulle *et al.*, 2010; Spurgeon & Hopkin).

Finally, genetic mechanisms of local adaptation to metallic pollution exist in the two species. In both cases, the role of these candidate genes needs to be validated in natural population. We assume that the colonization of metal contaminated sites by those two species may led to a difference in MT coding gene expression. Since MT2 expression phenotype is contrasted between worms living in metals contaminated sites versus worms living in non-contaminated sites, studies comparing gene expression between individuals from different ecotypes are required.

4 Phenotypic plasticity and local adaptation

The effect of phenotypic plasticity and local adaptation are usually confounded when comparing any adaptive phenotype between two contrasting populations *in situ.* It is effectively impossible to distinguish between local adaptation and phenotypic plasticity without an experimental

approach comparing adaptive traits between populations in controlled conditions so as to eliminate the environmental effect.

Common garden approaches are widely used to compensate this problem by placing individuals (ideally cocoon or juveniles) from different populations in the same environment. If there is phenotypic plasticity in the strict sense, it is assumed that all individuals will response in the same way. On the contrary, any observed difference between populations for the studied phenotype would be of genetic origin. This point is valid even in the presence of G*E interaction, since all individuals are in the same environment. In this case, we will simply limit the scope of the analysis in a particular environment (for example, metals polluted environment).

In perspective, local adaptation could be detected in *L. terrestris*, using common garden to detect genetic differences in metal tolerance between individuals collected from polluted and non-polluted sites. This approach would allow eliminating most of variations related to other environmental conditions. Genetic differences between metallicolous and non-metallicolous populations could be revealed by measurement of life history traits such as growth rate, sexual maturity, cocoon number and mortality rate as well as measurement of gene expressions. Common garden experiment had already been investigated in metallicolous populations and non-metallicolous populations of *A. halleri.* These experiments showed differences in metal tolerance between two ecotypes (Meyer *et al.*, 2010; Decombeix, 2011).

Nevertheless, the best approach to detect local adaptation for any phenotypic trait is called reciprocal transplantation (Kawecki & Ebert, 2004). The idea of this experiment is to grow local individuals and individuals from other populations (immigrants) in several meta-population environments. If there is local adaptation, it is expected that, at least in some environments, local individuals will have a better selective value than immigrants.

Annexes

New Phytologist

Research

Potential preadaptation to anthropogenic pollution: evidence from a common quantitative trait locus for zinc and cadmium tolerance in metallicolous and nonmetallicolous accessions of Arabidopsis halleri

Claire-Lise Meyer[1], Maxime Pauwels[2], Loic Briset[1], Cecile Gode[2], Pietro Salis[1], Angelique Bourceaux[2], Dima Souleman[2], Helene Frerot[2]* and Nathalie Verbruggen[1]*

[1]Laboratoire Physiologie et Genetique Moleculaire des Plantes, Universite Libre de Bruxelles, Boulevard du Triomphe, Campus Plaine, B-1050 Brussels, Belgium; [2]Universite de Lille – Sciences et Technologies, CNRS, UMR 8198 Evo-Eco-Paleo, F-59655 Villeneuve d'Ascq, France

Author for correspondence:
Nathalie Verbruggen
Tel: +32 2 650 21 28
Email: nverbru@ulb.ac.be

Received: 6 April 2016
Accepted: 31 May 2016

New Phytologist (2016)
doi: 10.1111/nph.14093

Key words: constitutive tolerance, genetic architecture, metal pollution, preadaptation, pseudometallophytes, rapid evolution.

Summary

As a drastic environmental change, metal pollution may promote the rapid evolution of genetic adaptations contributing to metal tolerance. In Arabidopsis halleri, genetic bases of zinc (Zn) and cadmium (Cd) tolerance have been uncovered only in a metallicolous accession, although tolerance is species-wide. The genetic determinants of Zn and Cd tolerance in a non-metallicolous accession were thus investigated for the first time.

The genetic architecture of tolerance was investigated in a nonmetallicolous population (SK2) by using first backcross progeny obtained from crosses between SK2 and Arabidopsis lyrata petraea, a nonmetallophyte species.

Only one significant and common quantitative trait locus (QTL) region was identified explaining 22.6% and 31.2% of the phenotypic variation for Zn and Cd tolerance, respec-tively. This QTL co-localized with HEAVY METAL ATPASE 4 (AhHMA4), which was previ-ously validated as a determinant of Zn and Cd tolerance in a metallicolous accession. Triplication and high expression of HMA4 were confirmed in SK2. In contrast, gene duplica-tion and high expression of METAL TOLERANT PROTEIN 1A (MTP1A), which was previously associated with Zn tolerance in a metallicolous accession, were not observed in SK2.

Overall, the results support the role of HMA4 in tolerance capacities of A. halleri that may have pre-existed in nonmetallicolous populations before colonization of metal-polluted habi-tats. Preadaptation to metal-contaminated sites is thus discussed.

Introduction

Global change, namely the impact of human activities on the Earth, has become a major concern. In particular, the way in which habitat disturbances affect species diversity is now increas-ingly studied. Anthropogenic disturbances resulting in abrupt changes towards hostile environmental conditions (e.g. use of insecticides or herbicides, or soil, air and water pollution) are expected to locally reduce species diversity because they challenge the maintenance of species populations in local communities. Theoretically, population persistence may rely on the levels of phenotypic plasticity, dispersal or genetic adaptations. However, environmental changes are often so drastic that phenotypic plas-ticity and migration are not sufficient to avoid extinction. Accordingly, many examples suggest that population persistence, following human-induced environmental disturbances, implies the rapid evolution of genetic adaptations (Reznick & Ghalambor, 2001). Thus, whereas strong selective pressures resulting from habitat modifications should cause population decline, adaptive evolutionary changes must occur quickly enough to restore population growth before extinction, a phenomenon known as evolutionary rescue (Gomulkiewicz & Holt, 1995; Gonzalez et al., 2013; Carlson et al., 2014). Whether populations can rapidly adapt or undergo evolutionary rescue actually depends on several extrinsic and intrinsic factors (Carlson et al., 2014). Among them, intrinsic factors, such as population size, the initial levels of standing variation, and the initial degree of maladaptation, seem essential (Gomulkiewicz & Holt, 1995; Bell & Gonzalez, 2009; Carlson et al., 2014).

Metal pollution is one of the major sources of anthropogenic disturbances inducing toxic environmental conditions for many plant populations, and is mainly caused by the use of agricultural fertilizers or the accumulation of residues from mining or smelting industries. Before the molecular biology era, several studies

*These authors contributed equally to this work.

revealed that the maintenance of plant populations facing metal pollution and the associated selective pressures implied the rapid evolution of an adaptive trait called 'metal tolerance' (Wu et al., 1975; Al-Hiyaly et al., 1988). It was also shown that the evolu-tion of metal tolerance in metallicolous (M) populations has occurred in pseudometallophyte species which displayed low fre-quencies of metal-tolerant individuals in nonmetallicolous (NM) populations growing in nonpolluted areas (Ingram, 1988, cited in Baker & Proctor, 1990). Therefore, the maintenance of M populations in those species suggested that they have undergone evolutionary rescue thanks to the rapid selection of favorable alle-les from initial standing genetic variation (Bradshaw, 1991).

Such rapid evolution of metal tolerance supported the assump-tion that it may be controlled by a few major genes, although minor modifier genes could be secondarily selected (Macnair, 1983, 1997; Schat & Ten Bookum, 1992). Using quantitative trait locus (QTL) mapping, genetic studies have investigated zinc (Zn) and cadmium (Cd) tolerance in Arabidopsis halleri, a pseu-dometallophyte model species that develops on metal-polluted sites where pollution is only of anthropogenic origin. From an interspecific cross (BC1) between a Zn/Cd-tolerant M accession (Auby, France) and its nontolerant close relative Arabidopsis lyrata petraea, QTL analyses revealed that the genetic architecture of Zn and Cd tolerance traits involved a limited number of genomic regions accounting for a substantial part of phenotypic variance (Courbot et al., 2007; Willems et al., 2007). Two candidate genes co-localizing with these major QTLs were mainly detected. First, HEAVY METAL ATPASE 4 (HMA4), encoding a plasma mem-brane metal pump involved in root-to-shoot metal translocation and cellular metal detoxification (Hanikenne et al., 2008), co-localized with major QTLs for Zn and Cd tolerance. Second, paralogs of the METAL TOLERANT PROTEIN 1 gene (AhMTP1A1 and AhMTP1A2 in tandem repeat and AhMTP1B), encoding Zn transporters involved in vacuolar sequestration (Dr€ager et al., 2004; Shahzad et al., 2010), co-localized with QTLs for Zn tolerance. More recently, CATION EXCHANGER 1 (CAX1), a gene co-localizing with the second major QTL for Cd tolerance, was identified after fine-mapping. CAX1 encodes a vac-uolar Ca^{++}/H^{+} exchanger that was proposed to limit Cd-induced reactive oxygen species (ROS) accumulation under low-Ca condi-tions (Baliardini et al., 2015). To date, only the contribution of AhHMA4 has been validated by RNA interference (RNAi)-mediated silencing (Hanikenne et al., 2008).

As QTLs were identified in an M accession, they might have been selected during the rapid adaptation of A. halleri to anthropogenic metal-polluted sites. However, it was demon-strated that NM populations of A. halleri exhibited (1) signifi-cant standing genetic variation (Meyer et al., 2009) as well as (2) a high degree of Zn and Cd tolerance, so that metal tol-erance in this pseudometallophyte species can be considered as species-wide or 'constitutive' (Pauwels et al., 2006; Meyer et al., 2010, 2015). Considering that M populations were probably founded from NM populations nearby (Pauwels et al., 2005), it can be reasonably assumed that (1) initial (i.e. present in NM populations) levels of standing variation were sufficient and (2) the initial degree of maladaptation of A. halleri to anthropogenic metal-polluted habitats was limited enough to promote the maintenance of M populations on metal-polluted soils. Nevertheless, significant differences in average tolerance levels between M and NM populations have been observed (Pauwels et al., 2006; Meyer et al., 2010, 2015). This means that the maintenance of M populations also has required specific adaptive changes, as expected in an evolutionary rescue scenario. However, it remains unclear whether AhHMA4, AhMTP1A1/A2/B and/or AhCAX1 would participate in the local adaptation of M populations to the impacted environment or whether they would be involved in constitutive tolerance capacities shared by M and NM popu-lations.

To examine whether the genetic determinants of Zn and Cd tolerance that have been identified in the M accession of A. halleri were selected during adaptation to anthropogenic metal-polluted habitats or pre-existed in NM populations, the genetic architecture of Zn and Cd tolerance was explored for the first time using an NM A. halleri accession. For this purpose, QTL mapping of a BC1 progeny (called BC1 SK2) generated from a cross between one A. halleri individual from an NM Slo-vakian population (SK2) and A. lyrata petraea was performed. To allow comparison with the previous QTL mapping of the BC1 generated with the Auby accession (BC1 AU), the same pheno-typing protocol as in Willems et al. (2007) and Courbot et al. (2007) was used. In addition, the genomic copy number and transcript levels of the candidate genes HMA4 and MTP1A that were identified in BC1 AU were determined in BC1 SK2 and in the parental populations. By comparison with previous studies only involving M individuals, we discuss the possibility that the identified mechanisms are putative preadaptations to anthropogenic metalliferous soils.

Materials and Methods

Plant material

The NM SK2 population is a Slovakian population located in the Tatras Mountains that was initially referred to as Sl2 (Bert et al., 2002). It was chosen for this study because of its particu-larly low Zn and Cd tolerance levels compared with other studied Arabidopsis halleri (L.) O'Kane & Al-Shehbaz popula-tions (Pauwels et al., 2006; Meyer et al., 2010, 2015). A single cross was performed between one individual from the SK2 population and one individual from the nontolerant species Arabidopsis lyrata petraea (L.) O'Kane & Al-Shehbaz (from Unhost', Central Bohemia: Macnair et al., 1999). Both species are self-incompatible and usually outcrossing. Therefore, to avoid any inbreeding depression effect, one randomly selected F1 individual was used as the pollen donor to fertilize a second A. lyrata petraea genotype, generating the interspecific backcross progeny (BC1). The BC1 population used for linkage map construction consisted of 335 individuals of which 129 and 70 were phenotyped for Zn and Cd tolerance, respectively. The different sample sizes resulted from difficulties in maintaining some genotypes by cutting.

New
Phytologist

Research 3

Evaluation of Zn and Cd tolerance

Zn and Cd tolerance were measured in two separate sequential growth tests as described in Willems et al. (2007) and Courbot et al. (2007). Genotypes were propagated by cutting and after 5 wk of growth on sand, three cuttings per BC1 genotype were transferred to vessels filled with nutrient solution. The Zn tolerance test was performed at the University of Lille (France) in a growth chamber with the following parameters: 13 h light d ^{1}, 80 lmol photons m ^{2} s ^{1} irradiance, 20°C : 18°C, day : night, and hygrometry of 65%. To minimize microenvironmental effects, vessels were randomly distributed in the chamber on a turntable (Rotoplan system; Strader, Saint Sylvain d'Anjou, France). Cd tolerance was measured under controlled glasshouse conditions (13 h light d ^{1}, 100 lmol photons m ^{2} s ^{1} irradi-ance, and 20°C : 17°C, day : night) at the Free University of Brussels. Vessels were randomly distributed in the chamber and moved around once a week during the change of nutrient solu-tion. After 3 wk in nutrient solution, plants were sequentially transferred to increasing concentrations of Zn (10–3000 lM $ZnSO_4$) or Cd (10–250 lM $CdSO_4$). Individual tolerance levels were determined as the lowest concentration at which no increase in root length (for Zn tolerance) or fresh biomass (for Cd toler-ance) was observed. This concentration is currently known as the effective concentration for 100% growth inhibition (EC_{100}).

Statistical analysis

For Zn tolerance and Cd tolerance, a one-way analysis of variance (ANOVA) was performed in SAS (GLM procedure; SAS Institute, Cary, NC) considering 'genotype' as a random effect. The nor-mality of residuals was tested using the Kolmogorov–Smirnov D test and a normal quantile-quantile (Q-Q) plot. Departure from normality was low for both traits (D = 0.15 and D = 0.14, respec-tively). For these reasons, and because normality was also difficult to improve, data sets were not transformed before parametric tests. The broad-sense heritability (H^2) of Zn and Cd tolerance was then calculated by dividing the genetic variance (which includes additive genetic variance and other sources of genetic variance) by the total phenotypic variance. Variance components were calculated using the REML method of the VARCOMP pro-cedure of SAS. The arithmetic mean of Zn and Cd tolerance for each genotype was then calculated for the three replicates. The correlation between Zn and Cd tolerances was estimated from genotype values using the Spearman nonparametric coefficient.

Marker analysis

The genomic DNA of the four parental genotypes and 335 individuals of the BC1 progeny was extracted using a Kit NucleoSpin 96 Plant (Macherey-Nagel, Hoerdt, France). Sixty single nucleotide polymorphism (SNP) or microsatellite markers were selected for genotyping (Supporting Information Table S1). Thirty-nine new microsatellite markers were developed as described in Frerot et al. (2010). Briefly, the markers were selected from a microsatellite-enriched genomic library developed from an enrichment procedure with Dynabeads (Thermo Fisher Scientific, Waltham, MA, USA) (Glenn & Schable, 2005). Primer sequences were designed in flanking regions of A. halleri microsatellites using PRIMER3 software (http://frodo.wi.mit.edu/). In order to allow multiplexing of markers, primer combinations were selected according to a 60°C (5°C) melting temperature and compatibility of PCR product sizes (100–150, 150–250 and 250–350 bp). The microsatellite markers were combined in 12 multiplexes labeled with FAM, PET, NED or VIC fluorescent dyes. Multiplex PCR was carried out in 10-ll reactions containing 19 PCR buffer II (Applied Biosystems, Foster City, CA, USA), 2.5 mM $MgCl_2$, 150 nM dNTP (Euromedex, Souffelweyersheim, France), 0.075 lM of the forward-M13 primer, 0.375 lM of the reverse-PIG primer, 1.5 lM of fluorescent dye-labeled M13 (Applied Biosystems), 0.5 U of Qiagen Multiplex PCR kit, and 5 ll of DNA (20–60 ng). The cycling conditions consisted of one initial denaturation step of 15 min at 95°C, followed by two touchdowns of five cycles each: denaturation for 45 s at 95°C, annealing at 68°C (2°C/cycle) for 5 min for the first touchdown and at 58°C (2°C/cycle) for 1 min for the second, and extension for 60 s at 72°C; then 27 cycles of 45 s at 95°C, 30 s at 47°C, and 60 s at 72°C; and a final extension of 10 min at 72°C. Amplification products were separated on an ABI Prism 3130 DNA sequencer (Applied Biosystems). Alleles at microsatellite loci were scored using GENEMAPPER software version 3.7 (Applied Biosystems).

In addition, re-sequencing data sets were used to design seven SNP markers. Illumina paired-end sequencing reads from six A. halleri populations (P. Saumitou-Laprade, unpublished data) were aligned to the Araly1 assembly of the A. lyrata lyrata genome (http://genome.jgi.doe.gov/Araly1/Araly1.info.html). Within regions of interest, the criteria used for the selection of SNPs were: interspecific polymorphism; fixed polymorphism within A. halleri or A. lyrata; 50-base flanking regions with low polymorphism within A. halleri and between A. halleri and A. lyrata; read depth > 20. The SNPs were genotyped by KBioscience competitive allele-specific polymerase chain reaction (KASP) (LGC Genomics, Teddington, UK). KASP assays were performed in a final reaction volume of 8 ll containing 4 ll of KASP master mix V2 low ROX (LGC Genomics), 0.125 ll of KASP mix assay (LGC Genomics) and c. 100 ng of genomic DNA. The PikoReal real-time PCR system (Thermo Scientific, Breda, the Netherlands) was used with the following cycling conditions: 15 min at 94°C; 10 touchdown cycles of 20 s at 94°C and 60 s at 61–55°C (the annealing temperature for each cycle being reduced by 0.6°C per cycle); and 26 cycles of 20 s at 94°C and 60 s at 55°C. Fluorescence detection was performed at the end of each cycle and the data sets were analyzed using the allelic discrimination RFU-based method of the PIKOREAL software 2.1.

Linkage map construction

The A. halleri SK2 9 A. lyrata petraea linkage map was constructed with the JOINMAP 3.0 program (Van Ooijen & Voorrips, 2001). The grouping of loci is based on a test for independence translated into a logarithm of odds (LOD) score. Linkage groups

were obtained at an LOD of 4. Markers along each linkage group were then ordered using the sequential method implemented in JOINMAP 3.0 in which loci are added one by one starting from the two most strongly linked loci. For each added locus, the best position is determined by comparing the goodness of fit of the resulting maps for each tested position. Kosambi's mapping function was used to translate recombination frequencies into map distances (Kosambi, 1944). When dealing with an interspecific cross, segregation distortion frequently occurs. Deviations from Mendelian ratios were assessed using a v^2 test implemented in JOINMAP 3.0 at a locus-by-locus significance level of a = 0.05.

QTL mapping

Detection of QTLs was performed using the MAPQTL 4.0 software (Van Ooijen et al., 2002). A Kruskall–Wallis rank test was first performed on each locus separately to find potential regions of QTLs. Interval mapping (IM) analysis then allowed finer detection by determining whether a QTL occurred and computing an LOD score for every centiMorgan (cM) along the linkage groups. The LOD score represents the 10-base logarithm of the quotient of two likelihoods: the likelihood of the presence of a segregating QTL (alternative hypothesis) divided by the likelihood of no segregating QTL (null hypothesis). To establish the occurrence of a QTL, the calculated LOD scores were compared with an LOD score threshold obtained by a permutation test (1000 permutations), which corresponds to a genome-wide empirical significance threshold at the 5% level (Churchill & Doerge, 1994). A multiple-QTL model (MQM) analysis was finally performed every cM, in which markers close to detected QTLs (by IM mapping) were selected as cofactors to take over the role of the nearby QTLs in the approximate multiple-QTL models used in the subsequent MQM analysis. This method reduces the residual variance and enhances the power of searching for other segregating QTLs. It also improves the precision of QTL positions. After manual selection of cofactors, an automatic selection of cofactors was executed to keep a restricted set of sig-nificant cofactors. The LOD score profiles showing QTLs with their one- and two-LOD support intervals were obtained using MAPCHART 2.1 (Voorrips, 2002).

Additionally, the power of analysis to detect the QTLs identified by Courbot et al. (2007) and Willems et al. (2007) was calculated using the R/QTLDESIGN software (Sen et al., 2007).

HMA4 and MTP1A relative transcript levels and gene copy numbers

Eighteen BC1 progenies with different allelic combinations (A. lyrata/A. lyrata or A. lyrata/A. halleri) in the genomic region of HMA4 (markers Chr2-06046 to Chr2-08800) and MTP1A (markers Chr2-17890 to Chr2-19598) were grown on soil in a controlled growth chamber (16 h light d^{1}, 100 lmol photons m^{2} s^{1} irradiance, 20°C : 18°C, day : night and 70% humidity). Nine leaves of each genotype were harvested, immediately frozen in liquid nitrogen, homogenized and stored at 80°C. Total RNA was extracted with TRI Reagent (Sigma, St Louis, MO, USA) and cDNA was synthesized from 1 lg of DNaseI-treated (Promega, Fitchburg, WI, USA) total RNA using oligo-dT and the Goscript reverse transcription system (Promega). Real-time quantitative PCRs were conducted in 96-well plates with the PikoReal real-time PCR system using Takion no ROX SYBR Mastermix (Eurogentec, Cologne, Germany). Volumes of 2.5 ll of cDNA were used for PCR in a 10-ll mix containing 5 ll of SYBR mastermix, 2 ll of H_2O and 0.5 lM of each primer. A total of three technical repeats were run per cDNA and primer pair combination. Relative transcript levels of HMA4 and MTP1A were calculated by normalization to ELONGATION FACTOR 1 ALPHA (EF1a) as a constitutively expressed reference gene (Talke et al. 2006). HMA4, MTP1A and EF1a primers were described previously (Talke et al., 2006; Shazhad et al., 2010; Hanikenne et al., 2013). In the study by Shazhad et al. (2010), the MTP1A primers were designed for amplification on A. halleri DNA. By sequencing, we checked the specificity of these primers on A. lyrata DNA and then the PCR efficiency was evaluated from the analysis of 1 : 1, 1 : 3, 1 : 12 and 1 : 48 dilution series. Genomic copy numbers of HMA4 and MTP1A were evaluated on genomic DNA of the BC1 SK2 parents. The single-copy gene SHORT ROOT (SHR) was used as a reference gene and a total of 15 technical repeats were run per DNA and primer pair combination according to the protocol previously described. All primers have been described previously (Baliardini et al., 2015).

Results

Zn and Cd tolerance in BC1 SK2

For both Zn and Cd tolerance, the broad-sense heritability (H^2) was moderate: 0.54 and 0.50, respectively. The phenotypic distribution of these traits was examined in the BC1 population (Fig. 1). A slight departure from normality was observed for both traits (D = 0.15 and D = 0.14, respectively; P-value < 0.01), with skewness (0.13 and 0.15, respectively) and kurtosis (3.07 and 3.07, respectively) values barely different from those of a normal distribution (0 and 3, respectively). Seven individuals showed higher Cd tolerance values (EC_{100} from 150 to 225 lM $CdSO_4$) than the average tolerance in the SK2 population (EC_{100} = 133 lM $CdSO_4$; Meyer et al., 2015). This result has to be treated with caution because parent representatives are not the original parents and some individuals from this population were able to survive up to 250 lM $CdSO_4$ in the study of Meyer et al. (2015). For both Zn and Cd tolerance, the phenotypic distribu-tions were rather different between the BC1 SK2 and the BC1 AU progenies, suggesting different genetic architectures (Fig. 1). The correlation between Zn and Cd was positive and significant (N = 50; r = 0.29; P = 0.039).

Linkage map

The 60 markers were assigned to the eight linkage groups of the A. halleri 9 A. lyrata petraea genetic map (Willems et al., 2007; Frerot et al., 2010). The length of each linkage group varied from 52.1 to 85.2 cM, while the marker number varied from five to

New
Phytologist

Research 5

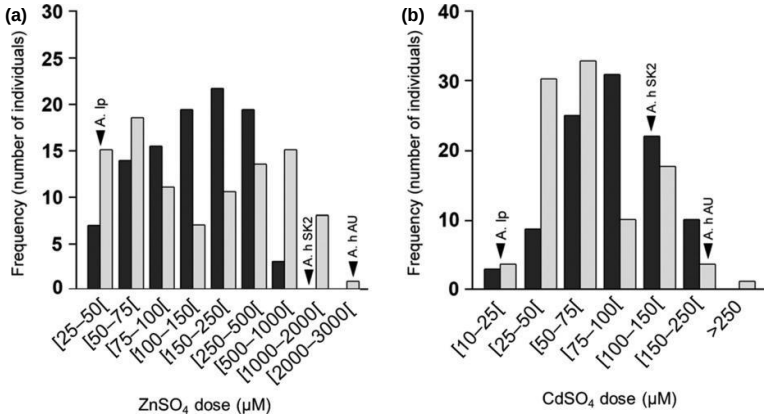


Fig. 1 Segregation profile for Zn (a) and Cd (b) tolerance in backcross progenies generated either using one SK2 individual (dark gray bars) or one Auby individual (light gray bars; from Courbot et al. (2007) and Willems et al. (2007)). Parents' representatives are noted above the arrows. A.lp, Arabidopsis lyrata ssp petraea; A.h SK2, A. halleri of population SK2; A.h AU, A. halleri from population AU (Auby).

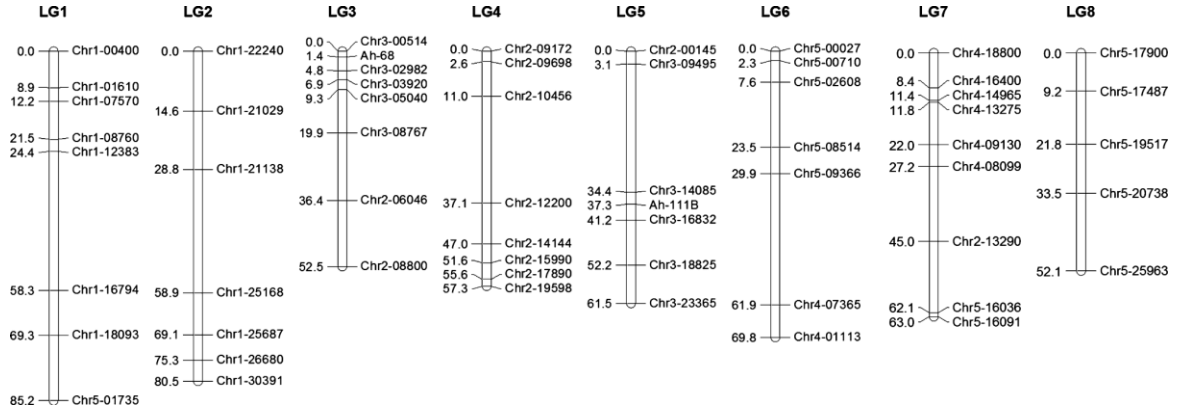


Fig. 2 Linkage map of the Arabidopsis halleri SK2 9 A. lyrata petraea BC1 progeny. Markers are labeled to the right of linkage groups by their position in the A. thaliana genome (chromosome number – position in kb) except for some markers with unknown positions. Distances on the genetic map are expressed in cM. Markers in segregation distortion are underlined. LG, linkage group.

nine by linkage group (Fig. 2). The total length of the map (522 cM) was very similar to the previous estimations (567 and 526 cM; Willems et al., 2007; Frerot et al., 2010). The order of the markers in the present genetic map was very similar to that in A. thaliana and A. lyrata physical maps (Fig. 2; Table S1). Dis-crepancies among these three maps were only observed for one marker in linkage group 7 (Chr2-13290) and three markers at the beginning of linkage group 2 (markers Chr1-22240, Chr1-21029 and Chr1-21138). The order in this group of three mark-ers was similar between the A. lyrata physical map and the present genetic map but the localization was different (end of scaffold 1 for A. lyrata; data not shown). At a significance threshold of 5%, 32 markers (i.e. 53% of markers) showed significant departure from the expected Mendelian ratio (Fig. 2). These markers were located mainly in LG1, LG2 and LG4 and showed distortion in the same direction (excess of the homospecific A. l. petraea/A. l. petraea allelic combination) as was observed in the BC1 AU progeny (Willems et al., 2007).

QTLs of Zn and Cd tolerance

Only one significant QTL was detected for Zn tolerance and one for Cd tolerance (called SK2_ZnTol1 and SK2_CdTol1, respec-tively) with LOD scores higher than the threshold obtained by permutation (i.e. 2.3). These two QTLs overlapped and were located at the end of linkage group 3 between markers Chr2-08800 and Chr2-06046. They explained 22.6% and 31.2% of the total phenotypic variance for Zn and Cd tolerance, respec-tively (Table 1; Fig. 3). Negative additive effects were measured for both traits, meaning that the A. halleri alleles increased toler-ance compared with the A. lyrata petraea alleles (Table 1). Inter-estingly, confidence intervals of SK2_ZnTol1 and SK2_CdTol1

Table 1 Summary characteristics of quantitative trait loci (QTLs) detected for zinc (Zn) and cadmium (Cd) tolerance in the Arabidopsis halleri SK2 9 A. lyrata petraea BC1 progeny

Tolerance to	QTL[a]	LG[b]	Marker[c]	LOD score[d]	R^{2} [e]	a[f]
Zn	SK2_ZnTol1	3	Chr2-08800/ Chr2-06046	6.85	22.6	178.50
Zn	SK2_ZnTol2	6	Chr5-08514	1.65	4.6	79.61
Zn	SK2_ZnTol3	8	Chr5-17487	1.29	3.9	75.00
Cd	SK2_CdTol1	3	Chr2-08800/ Chr2-06046	4.33	31.2	48.74
Cd	SK2_CdTol2	1	Chr5-01735	1.55	10.5	27.71
Cd	SK2_CdTol3	5	Chr3-14085	1.39	7.3	23.30

[a]QTLs are named according to the trait and ordered according to their significance level. Suggestive QTLs are indicated in italics.
[b]Linkage groups (LGs) where the QTLs were detected.
[c]Marker closest to the higher logarithm of odds (LOD) score.
[d]Maximum LOD score for the linkage group obtained by the multiple-QTL model (MQM) mapping method.
[e]Percentage of variance explained by the QTL. [f]Additive effect of the QTL.

overlapped with the position of the HMA4 gene (Chr2-08280 in the Arabidopsis thaliana genome). Additionally, two peaks near the LOD threshold were observed in linkage groups 6 and 8, and 1 and 5 for Zn and Cd tolerance, respectively (Fig. 3; Table 1) which may indicate minor-effect QTLs. These putative QTLs did not co-localize with the QTLs previously identified in BC1 AU. Together, they explained 8.5% and 17.8% of the Zn and Cd phenotypic variance, respectively.

Using 129 BC1 individuals, markers spaced c. 7 cM apart, broad-sense heritability of 0.89 and an LOD threshold of 2.3 for declaring significance, 99% power was found to detect QTLs with additive effects similar to those of ZnTol2 and ZnTol3 (Willems et al., 2007). With the same parameters, 70 BC1 indi-viduals and a broad-sense heritability of 0.79, 75% power to detect a QTL similar to CdTol2 was obtained and 50% power for a QTL similar to CdTol3 (Courbot et al., 2007). These results indicated that the absence of QTLs similar to those of BC1 AU in the present study is not attributable to a lack of statis-tical power.

Transcript levels and genomic copy numbers of HMA4 and MTP1A

In order to confirm HMA4 as a candidate gene for the overlapping QTLs SK2_ZnTol1 and SK2_CdTol1, relative expression and genomic copy numbers were examined in some BC1 SK2

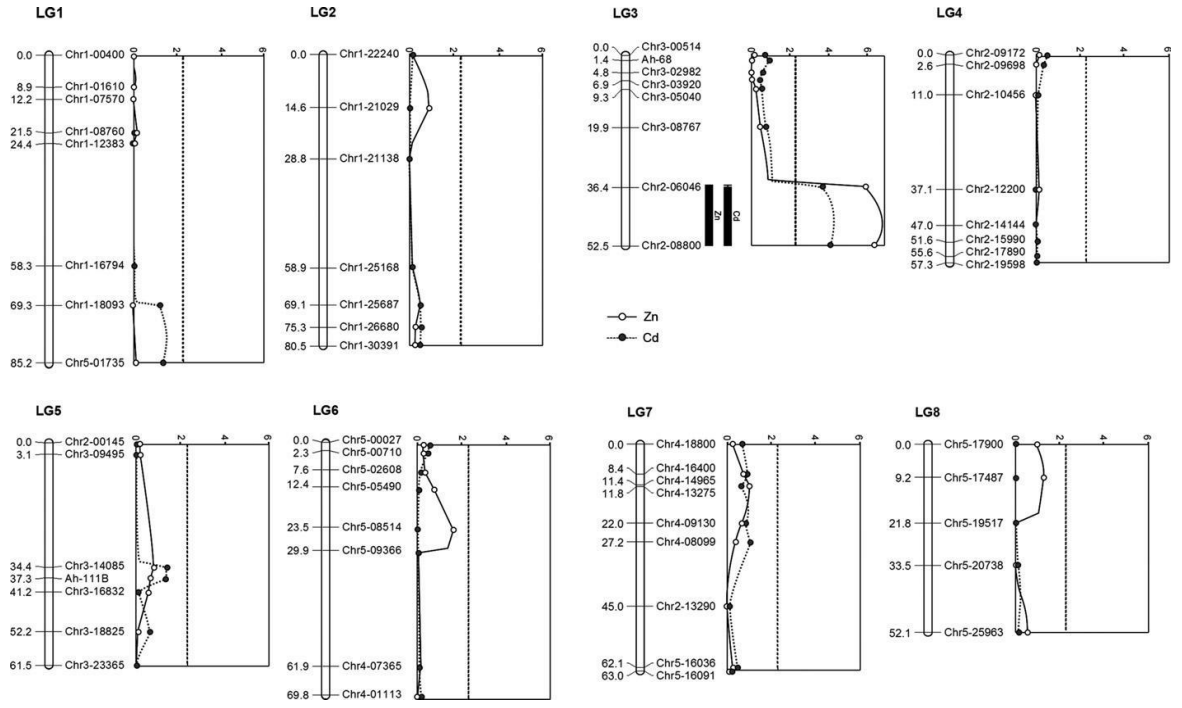


Fig. 3 Quantitative trait locus (QTL) mapping for zinc (Zn) and cadmium (Cd) tolerance in the Arabidopsis halleri SK2 9 A. lyrata petraea progeny obtained by the multiple QTL model (MQM) mapping method. Marker names are designated by the position on A. thaliana chromosomes except for some markers with unknown positions. Logarithm of odds (LOD) scores are indicated on the horizontal axes. The vertical dashed lines represent the LOD score threshold (2.3) at a 5% error level for QTL detection. The positions of QTLs are indicated by bars representing the one-LOD support intervals (one LOD score unit on either side of the QTL peak) and whiskers representing the two-LOD support intervals (two LOD score unit on either side of the QTL peak). Black line, LOD scores for Zn tolerance; dotted line, LOD scores for Cd tolerance.

New
Phytologist

Research 7

individuals and in the parents of the cross. The same analysis was performed on MTP1A to further confirm that MTP1 was not a candidate gene for Zn tolerance in SK2. Relative expression of HMA4 and MTP1A was studied in the shoots of 18 BC1 individuals representing the two genotypic classes (A. l. petraea/A. l. petraea and A. l. petraea/A. halleri) in each genomic region. On average, 10-fold higher HMA4 transcript levels were observed in the BC1 genotypes harboring the A. halleri allele at the QTLs SK2_ZnTol1 and SK2_CdTol1 than in the individuals with two A. lyrata petraea alleles in this interval (Fig. 4a). By contrast, MTP1A was similarly expressed in the different BC1 individuals independently of the presence of the A. halleri allele in the QTL region (Fig. 4b). Using quantitative PCR on genomic DNA, the copy number in the parents of the BC1 SK2 was estimated. As expected, three HMA4 gene copies were found for A. halleri and two for A. lyrata petraea (Fig. 5a). It should be noted that one of the two copies detected in A. lyrata is a truncated pseudogene (Hanikenne et al., 2013). Only one copy of MTP1A was detected in the parent A. halleri SK2 and in A. lyrata petraea (Fig. 5b).

Discussion

Genetic architecture of Zn and Cd tolerance in an NM A. halleri accession

The genetic architecture of Zn and Cd tolerance was investigated for the first time in progeny from an interspecific cross between an NM accession of A. halleri (SK2) and A. lyrata petraea, a non-metallophyte close relative species. Previous QTL analyses suggested a genetic architecture involving six major genetic determinants for Zn and Cd tolerance, called Zntol-1, Zntol-2, Zntol-3, Cdtol-1, Cdtol-2 and Cdtol-3 (Courbot et al., 2007; Willems et al., 2007). Associated percentages of explained variance were moderate for Zn tolerance (12.2%, 11.2% and 5.6%, respectively) and higher for Cd tolerance (42.9%, 23.7% and 15.9%, respectively). By contrast, the genetic architecture detected here was apparently simpler, as only one significant QTL region was detected for each trait (SK2_ZnTol1 and SK2_CdTol1, respectively). The associated explained variances were reasonably high (22.6% and 31.2%, respectively), even though they may be overestimated as a result of sample size effects (Beavis, 1994). Interestingly, however, SK2_ZnTol1 co-localized with Zntol-1 in Willems et al. (2007) and SK2_CdTol1 co-localized with Cdtol-1 in Courbot et al. (2007). Moreover, SK2_ZnTol1 and SK2_CdTol1 co-localized with each other and represented the only QTL region that was shared between Zn and Cd tolerance in previous studies (Fig. 3). Additionally, two peaks near the LOD threshold which may indicate minor-effect QTLs explained 8.5% and 17.8% of the Zn and Cd phenotypic variance, respectively. These putative QTLs did not co-localize with the QTLs previously identified with BC1 AU.

The phenotypic distributions in BC1 SK2 were unimodal with an intermediate mode between the values of the parents' representatives (Fig. 1). By contrast, bimodal distributions of phenotypic values were observed for Zn and Cd tolerance of the BC1 AU progeny (Courbot et al., 2007; Willems et al., 2007). A bimodal distribution is commonly interpreted as evidence for one major effect locus in association with a few minor effect loci. Nevertheless, a unimodal – even normal – distribution is not evidence of genetic control by more than a single major gene (Lynch & Walsh, 1998). The phenotypic distributions in BC1 SK2 thus remained in accordance with the detection of one major-effect QTL, even though a contribution of additional weak-effect genes for Zn and Cd tolerance in A. halleri is very likely.

Putative role of AhHMA4 in constitutive tolerance in A. halleri

SK2 showed among the lowest levels of Zn and Cd tolerance compared with other NM and M A. halleri accessions (Pauwels

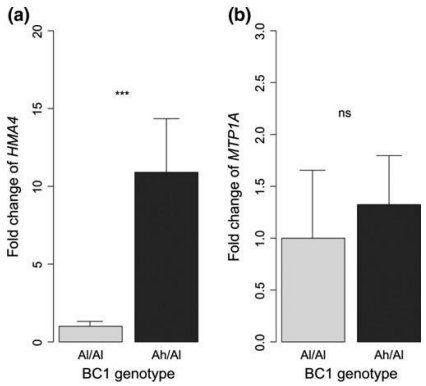


Fig. 4 Leaf transcript levels of with HEAVY METAL ATPASE 4 (HMA4) and METAL TOLERANT PROTEIN 1A (MTP1A) in BC1 SK2 individuals displaying the Arabidopsis lyrata petraea or A. halleri alleles (Al and Ah, respectively). Values (mean + SE; n = 9) are given relative to transcript levels of the allelic class Al/Al. Asterisks indicate significant differences: ***, P < 0.001; ns, nonsignificant.

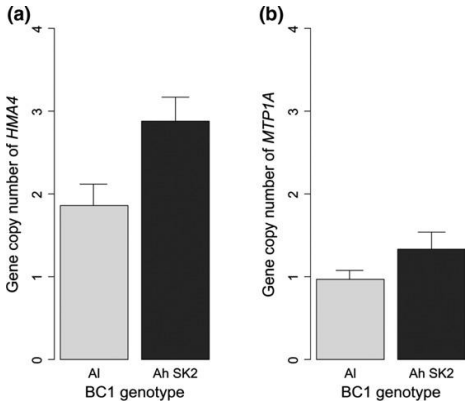


Fig. 5 Gene copy number in the parents of the Arabidopsis halleri SK2 9 A. lyrata petraea BC1 progeny for (a) HEAVY METAL ATPASE 4 (HMA4) and (b) METAL TOLERANT PROTEIN 1A (MTP1A). Values (mean + SE; n = 15 technical replicates) are given relative to the single-copy gene SHORT ROOT (SHR). Arabidopsis lyrata petraea (Al) genomic DNA served as a calibrator. Note that, among the AlHMA4 copies, one is a truncated pseudogene (Hanikenne et al., 2013).

et al., 2006; Meyer et al., 2010, 2015). This suggests that the tolerance capacities of SK2 are representative of the constitutive capacities that are shared by the whole species and must have evolved very early after the species emergence (Bert et al., 2000; Pauwels et al., 2006). Accordingly, we assume that genetic bases of constitutive background tolerance are species-wide and, a fortiori, shared among M and NM populations. As SK2_ZnTol1 and SK2_CdTol1 (the overlapping QTLs) corresponded to the common QTL region identified for BC1 AU and BC1 SK2, this region may be involved in constitutive tolerance in A. halleri.

Thus, considering the co-localization of SK2_ZnTol1 and SK2_CdTol1 with Zntol-1 (Willems et al., 2007) and Cdtol-1 (Courbot et al., 2007), AhHMA4 can be considered as a relevant candidate gene for constitutive tolerance. Although the wide confidence intervals of the QTL regions mean that we cannot rule out the possibility that one or more genes in linkage disequilib-rium may also be valid candidates, strong evidence supports AhHMA4 as a candidate gene for constitutive tolerance. First, AhHMA4 was already identified as a candidate gene underlying Zntol-1 and Cdtol-1 in the BC1 AU studied by Willems et al. (2007) and Courbot et al. (2007). A contribution of HMA4 to Zn and Cd tolerance traits was further demonstrated by RNAi-mediated silencing in the German M accession of Langelsheim (Hanikenne et al., 2008). In addition, high HMA4 expression was associated with tandem triplication of AhHMA4, whereas the gene is present as a single copy in nontolerant relatives (Hanikenne et al., 2008). A combination of demographic and molecular evolution approaches suggested that AhHMA4 triplica-tion occurred well before the establishment and colonization of anthropogenic metal-polluted habitats (Roux et al., 2011). Actu-ally, the first duplication event of AhHMA4 could have occurred very close to the time of the split between A. lyrata and A. halleri, and was subsequently followed by the second AhHMA4 duplica-tion (Roux et al., 2011). More recently, a population survey suggested that triplication was shared among M and NM accessions from Germany (Hanikenne et al., 2013). Our results confirmed that AhHMA4 was also triplicated in SK2, a population that can be considered to be genetically isolated from M accessions than have been characterized so far (Pauwels et al., 2012). Taken together, previous findings and the results of the present study strongly support the assumption that the tandem triplication and the consecutive overexpression of AhHMA4 may be a major determinant of constitutive Zn and Cd tolerance in A. halleri.

Role of standing genetic variation in adaptive evolution to anthropogenic metal-polluted habitats

The subtractive comparison of QTL regions detected for BC1 AU and BC1 SK2 should be informative about the genetic mechanisms involved in the local adaptation to anthropogenic metal-polluted habitats. It is remarkable that Zntol-2 and Zntol-3, which are commonly associated with AhMTP1A and AhMTP1B, respectively (Willems et al., 2007), Cdtol-2, which is associated with AhCAX1 (Baliardini et al., 2015), and Cdtol-3 were not detected in BC1 SK2. It is also remarkable that AhMTP1A was not overexpressed in SK2 compared with control sensitive species (Fig. 4b). This might indicate that those candidate genes may rather contribute to enhanced metal tolerance observed in M populations instead of constitutive tolerance. We suggest that the maintenance of M populations in anthropogenic polluted sites (i.e. their evolutionary rescue) might have involved selection of particular variants of those genes, either in the coding or in the promoting region. Considering AhMTP1, this would suggest that an improvement of Zn detoxification mechanisms, through enhanced sequestration in leaf cell vacuoles, would have been advantageous for plants exposed to higher concentrations of Zn in soils. In contrast, effective Zn detoxification is not necessary on nonmetalliferous soils, and basic activity of AhMTP1 should be sufficient to ensure Zn homeostasis.

More generally, most adaptive genetic changes are expected to result from standing variation, that is, the selection of neutral or mildly deleterious alleles present at variable frequency in populations (Matuszewski et al., 2015). Based on investigations of within-population genetic variation, it has already been suggested that M populations of A. halleri may have evolved from large standing variation existing in NM populations (including SK2) rather than from new mutations (Pauwels et al., 2006; Meyer et al., 2010). Several authors came to the same conclusion regard-ing tolerance polymorphism within NM populations of Noccaea caerulescens (Meerts and Van Isacker, 1997; Escarre et al., 2000).

Preadaptation of A. halleri to anthropogenic metal-polluted habitats

Apart from selection of standing variation, local adaptation may also benefit from the existence of advantageous alleles that are already at high frequency or even fixed in populations. In this case, individuals can be considered as 'preadapted'. In some pseu-dometallophytes, it has already been suggested that adaptation to metal-polluted soils may have benefited from preadapting pro-cesses. For example, preadaptation to high soil Zn concentrations was reported for Canadian populations of Deschampsia cespitosa as all NM individuals tested revealed quite high Zn tolerance levels (Cox & Hutchinson, 1981). Similarly, the widespread tol-erance to high magnesium : calcium ratios and nickel concentra-tions covered by a granite outcrop population of Phacelia dubia suggested that this species showed some degree of preadaptation to serpentine habitats (Taylor & Levy, 2002). In A. halleri, the pre-existence of constitutive tolerance and of AhHMA4 copy number expansion (Roux et al., 2011; Hanikenne et al., 2013) suggests that the species was somehow preadapted to anthro-pogenic metal-polluted habitats. It also interrogates the evolu-tionary 'raison d'^etre' of constitutive metal tolerance in a pseudometallophyte. Evidence of positive selection in the AhHMA4 genomic region in M as well as NM populations implies that AhHMA4 paralogs may have played a role in a non-metalliferous environment (Hanikenne et al., 2013). However, the selective pressure that may have favored the selection of AhHMA4 overexpression remains unclear.

A preadaptation may occur when a trait that is currently associated with a particular adaptation was previously selected for another function in response to a different selective pressure

(Gould & Vbra, 1982). The current trait is thus considered as an 'exaptation', in comparison with an 'adaptation' which is an original trait that evolved for its present function. A preadaptation is thus a potential exaptation if it becomes selected for a different function. Considering AhHMA4 paralogs, available data indicate that they may not only be involved in tolerance of toxic metals in soils. In particular, AhHMA4 gene copies have been shown to participate to Zn and Cd hyperaccumulation (Hanikenne et al., 2008; Frerot et al., 2010; Willems et al., 2010), that is, the abnormal concentrations of these trace elements in aerial plant parts (for recent reviews, see Kr€amer, 2010; van der Ent et al., 2013; Verbruggen et al., 2009). Indeed, HMA4 is a plasma membrane pump that plays a role in Zn and Cd xylem loading and cellular detoxification of Zn and Cd excess (Hanikenne et al., 2008). The selective pressures that promote the evolution of hyperaccumulation have been, and are still, a subject of debate (Boyd & Martens, 1992; Pollard et al., 2014). Interestingly, some of these putative selective pressures could act in a nonpolluted environ-ment. For example, elemental defense against herbivores and pathogens or elemental allelopathy have been suggested as poten-tial roles for metal hyperaccumulation (Boyd & Martens, 1992, 1998).

Finally, our study provides consistent arguments in favor of some degree of preadaptation to Zn and Cd tolerance in A. halleri. These arguments will be strengthened when the occur-rence of AhHMA4 copies, associated with significant Zn and Cd tolerance capacities, is demonstrated in several NM populations. Undoubtedly, the more integrated NM accessions are in future molecular studies, the greater will be the potential to elucidate the evolutionary origin of tolerance capacities in A. halleri. Other important implications of our work relate to investigations into mechanisms of local adaptation to metal-polluted soils. In this regard, it should be enlightening (1) to pursue the identification of candidate genes associated with the QTL regions specific to BC1 AU, (2) to explore the association between their expression and quantitative variations for Zn and Cd tolerance in natural populations of A. halleri, and (3) to examine the consequence of local adaptation to metal-polluted sites for the population structure of genetic diversity at those genes.

Acknowledgements

This project was financially supported by the FNRS (PDR T.0206.13; grants to C-L.M. and L.B.) and the French National Agency (ELOCANTH project; ANR-AAP JCJC–SVSE 7 call of proposal). We are grateful to Anne Creach for help with BC1 SK2 follow-up in the glasshouse at the University of Lille, to Sophie Gallina for making available sequencing data to define SNP markers, and Sebastien Osterreith for help with Cd tolerance phenotyping of BC1 SK2 at the University of Brussels. We also thank Ophelie Ronce for fruitful scientific discussions. The two laboratories involved (Laboratoire de Physiologie et Genetique Moleculaire des Plantes and Unite Evo-Eco-Paleo) are members of the International Research Network LOCOMET funded by the French National Centre of Scientific Research (CNRS).

Author contributions

C-L.M., H.F. and N.V. planned and designed the research, analyzed data and wrote the manuscript; C-L.M., L.B. and C.G. performed experiments; M.P. helped with the MTP1 genotyping and revised the manuscript; A.B. and P.S. maintained plant collections; D.S. performed preliminary experiments.

References

Al-Hiyaly SA, McNeilly T, Bradshaw AD. 1988. The effect of zinc contamination from electricity pylons – evolution in a replicated situation. New Phytologist 110: 571–580.

Baker AJM, Proctor J. 1990. The influence of cadmium, copper, lead, and zinc on the distribution and evolution of metallophytes in the British Isles. Plant Systematics and Evolution 173: 91–108.

Baliardini C, Meyer C-L, Salis P, Saumitou-Laprade P, Verbruggen N. 2015. CAX1 co-segregates with Cd tolerance in the metal hyperaccumulator Arabidopsis halleri and plays a role in limiting oxidative stress in Arabidopsis. Plant Physiology 169: 549–559.

Beavis WD. 1994. The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the 49th Annual Corn & Sorghum Industry Res. Conf. Am Seed Trade Assoc., Chicago, IL, USA, 250–266.

Bell G, Gonzalez A. 2009. Evolutionary rescue can prevent extinction following environmental change. Ecology Letters 12: 942–948.

Bert V, Bonnin I, Saumitou-Laprade P, de Laguerie P, Petit D. 2002. Do Arabidopsis halleri from nonmetalicolous populations accumulate zinc and cadmium more effectively than those from metallicolous populations? New Phytologist 155: 47–57.

Bert V, Macnair MR, De Laguerie P, Saumitou-Laprade P, Petit D. 2000. Zinc tolerance and accumulation in metallicolous and nonmetallicolous populations of Arabidopsis halleri (Brassicaceae). New Phytologist 146: 225–233.

Boyd RS, Martens SN. 1992. The raison d'^etre for metal hyperaccumulation by plants. In: Baker AJM, Proctor J, Reeves RD, eds. The ecology of ultramafic (Serpentine) soils. Andover, UK: Intercept, 279–289.

Boyd RS, Martens SN. 1998. Nickel hyperaccumulation by Thlaspi montanum var. montanum (Brassicaceae): a constitutive trait. American Journal of Botany 85: 259–265.

Bradshaw AD. 1991. Genostasis and the limit to evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 333: 289–305.

Carlson SM, Cunningham CJ, Westley PAH. 2014. Evolutionary rescue in a changing world. Trends in Ecology & Evolution 29: 521–530.

Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971.

Courbot M, Willems G, Motte P, Arvidsson S, Roosens N, Saumitou-Laprade P, Verbruggen N. 2007. A major quantitative trait locus for cadmium tolerance in Arabidopsis halleri colocalizes with HMA4, a gene encoding a heavy metal ATPase. Plant Physiology 144: 1052–1065.

Cox RM, Hutchinson TC. 1981. Multiple and co-tolerance to metal in the grass Deschampsia cespitosa: adaptation, preadaptation and the cost. Journal of Plant Nutrition 3: 731–741.

Dr€ager DB, Desbrosses-Fonrouge A-G, Krach C, Chardonnens AN, Meyer RC, Saumitou-Laprade P, Kr€amer U. 2004. Two genes encoding Arabidopsis halleri MTP1 metal transport proteins co-segregate with zinc tolerance and account for high MTP1 transcript levels. Plant Journal 39: 425–439.

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.

Frerot H, Faucon M-P, Willems G, Gode C, Courseaux A, Darracq A, Verbruggen N, Saumitou-Laprade P. 2010. Genetic architecture of zinc hyperaccumulation in Arabidopsis halleri: the essential role of QTL 9 environment interactions. New Phytologist 187: 355–367.

Glenn TC, Schable NA. 2005. Isolating microsatellite DNA loci. Methods in Enzymology 395: 202–222.

Gomulkiewicz R, Holt RD. 1995. When does evolution by natural selection prevent extinction? Evolution 49: 201–207.

Gonzalez A, Ronce O, Ferriere R, Hochberg ME. 2013. Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. Philosophical Transactions of the Royal Society B: Biological Sciences 368: 20120404.

Gould SJ, Vbra ES. 1982. Exaptation – a missing term in the science of form. Paleobiology 8: 4–15.

Hanikenne M, Kroymann J, Trampczynska A, Bernal MP, Motte P, Clemens S, Kr€amer U. 2013. Hard selective sweep and ectopic gene conversion in a gene cluster affording environmental adaptation. PLoS Genetics 9: e1003707.

Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Kr€amer U. 2008. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of HMA4. Nature 453: 391–396.

Kosambi D. 1944. The estimation of map distances from the recombination values. Annals of Eugenics 12: 172–175.

Kr€amer U. 2010. Metal hyperaccumulation in plants. Annual Review of Plant Biology 61: 517–534.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland, MA, USA: Sinauer Associates.

Macnair MR. 1983. The genetic control of copper tolerance in the yellow monkey flower, Mimulus guttatus. Heredity 50: 283–293.

Macnair MR. 1997. The evolution of plants in metal-contaminated environments. In: Bijlsma R, Loeschcke V, eds. Environmental stress, adaptation and evolution. Basel, Switzerland: Birkh€auser-Verlag, 3–24.

Macnair MR, Bert V, Huitson SB, Saumitou-Laprade P, Petit D. 1999. Zinc tolerance and hyperaccumulation are genetically independant characters. Proceedings of the Royal Society of London B: Biological Sciences 266: 2175–2179.

Matuszewski S, Hermisson J, Kopp M. 2015. Catch me if you can: adaptation from standing genetic variation to a moving phenotypic optimum. Genetics 200: 1255–1274.

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

Meyer C-L, Kostecka AA, Saumitou-Laprade P, Creach A, Castric V, Pauwels M, Frerot H. 2010. Variability of zinc tolerance among and within populations of the pseudometallophyte Arabidopsis halleri and possible role of directional selection. New Phytologist 185: 130–142.

Meyer C-L, Vitalis R, Saumitou-Laprade P, Castric V. 2009. Genomic pattern of adaptive divergence in Arabidopsis halleri, a model species for tolerance to heavy metal. Molecular Ecology 18: 2050–2062.

Pauwels M, Frerot H, Bonnin I, Saumitou-Laprade P. 2006. A broad-scale study of population differentiation for Zn-tolerance in an emerging model species for tolerance study: Arabidopsis halleri (Brassicaceae). Journal of Evolutionary Biology 19: 1838–1850.

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, Gode C, Frerot 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.

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

Reznick DN, Ghalambor CK. 2001. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica 112–113: 183–198.

Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. 2011. Does speciation between Arabidopsis halleri and Arabidopsis lyrata coincide with major changes in a molecular target of adaptation? PLoS ONE 6: e26872.

Schat H, Ten Bookum WM. 1992. Genetic control of copper tolerance in Silene vulgaris. Heredity 68: 219–229.

Sen S, Satagopan JM, Broman KW, Churchill GA. 2007. R/qtlDesign: inbred line cross experimental design. Mammalian Genome 18: 87–93.

Shahzad Z, Gosti F, Frerot H, Lacombe E, Roosens N, Saumitou-Laprade P, Berthomieu P. 2010. The five AhMTP1 zinc transporters undergo different evolutionary fates towards adaptive evolution to zinc tolerance in Arabidopsis halleri. PLoS Genetics 6: e1000911.

Taylor SI, Levy F. 2002. Responses to soils and a test for preadaptation to serpentine in Phacelia dubia (Hydrophyllaceae). New Phytologist 155: 437–447.

Talke IN, Hanikenne M, Kra¨mer U. 2006. Zinc-dependent global transcriptional control, transcriptional deregulation, and higher gene copy number for genes in metal homeostasis of the hyperaccumulator Arabidopsis halleri. Plant Physiology 142: 148–167.

Van Ooijen JW, Boer MP, Jansen RC, Maliepaard C. 2002. MapQTL 4.0: software for the calculation of QTL positions on genetic maps. Wageningen, the Netherlands: Plant Research International.

Van Ooijen JW, Voorrips RE. 2001. Joinmap 3.0: software for the calculation of genetic linkage maps. Wageningen, the Netherlands: Plant Research International.

Verbruggen N, Hermans C, Schat H. 2009. Molecular mechanisms of metal hyperaccumulation in plants. New Phytologist 181: 759–776.

Voorrips RE. 2002. Mapchart: software for the graphical presentation of linkage maps and QTLs. Heredity 93: 77–78.

Willems G, Dr€ager D, Courbot M, Gode C, Verbruggen N, Saumitou-Laprade P. 2007. The genetic basis of zinc tolerance in the metallophyte Arabidopsis halleri ssp. halleri (Brassicaceae): an analysis of quantitative trait loci. Genetics 176: 659–674.

Willems G, Frerot H, Gennen J, Salis P, Saumitou-Laprade P, Verbruggen N. 2010. Quantitative trait loci analysis of mineral element concentrations in an Arabidopsis halleri 9 Arabidopsis lyrata petraea F_2 progeny grown on cadmium-contaminated soil. New Phytologist 187: 368–379.

Wu L, Bradshaw AD, Thurman DA. 1975. The potential for evolution of heavy metal tolerance in plants. III. The rapid evolution of copper tolerance in Agrostis stolonifera. Heredity 34: 165–187.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Table S1 List of markers used in linkage map construction

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the New Phytologist Central Office.

Annex 2: Posters realized during this thesis

36ème réunion du Groupe d'Etude de Biologie et Génétique des Populations (Petit Pois Déridé Orsay 2014)

Nouvelles avancées sur l'origine évolutive de de la tolérance aux métaux chez l'espèce pseudométallophyte *Arabidopsis halleri* : que nous apprennent les populations non métallicoles ?

Dima Souleman[1], Claire-Lise Meyer[2], Loïc Briset[2], Nathalie Verbruggen[2], Hélène Frérot[1]

[1]Laboratoire Génétique et Evolution des Populations Végétales, UMR CNR 8198, Bâtiment SN2, Université Lille 1, 59655 Villeneuve d'Ascq Cedex, France

[2] Laboratoire de Physiologie et Génétique Moléculaire des plantes, Bâtiment BC, Campus Plaine, Université de Bruxelles, 1050 Bruxelles, Belgique

Tel : 0622657573

Email : demasouleman@yahoo.com

Résumé

Les métaux lourds peuvent être émis dans l'air sous forme de très fines particules qui finissent par retomber sur le sol, mais ils peuvent aussi être présents dans les déchets miniers et industriels déposés à la surface des sols. Certaines plantes sont adaptées à cette contamination, et sont dites tolérantes aux métaux. Parmi ces espèces, certaines sont également hyperaccumulatrices de métaux dans leurs parties aériennes, c'est le cas d'*Arabidopsis halleri* (Brassicacée), qui hyperaccumule le zinc et le cadmium. L'identification des gènes impliqués dans la tolérance et l'hyperaccumulation permettrait de bien comprendre les mécanismes évolutifs impliqués dans l'adaptation aux milieux anthropisés. Un croisement BC1 entre *A. halleri* et une proche parente non tolérante (*A. lyrata petraea*) a permis d'identifier plusieurs QTL pour la tolérance et l'accumulation du Zn et du Cd (Courbot et al. 2007, Willems et al. 2007, Frérot et al. 2010, Willems et al. 2010). Ces études ont cependant utilisé un seul individu d'*A. halleri* provenant

d'une population métallicole. Or des études récentes de phylogéographie (Pauwels et al. 2005, Pauwels et al. 2012) et de phénotypage en milieux controllés (Pauwels et al. 2005, Meyer et al. 2010, Meyer et al. submitted) ont montré que plusieurs unités génétiques existaient au sein de l'aire de distribution de cette espèce, que les populations métallicoles étaient fondées de manière indépendantes à partie des populations non-métallicoles les plus proches géographiquement et qu'il existait une grande variabilité de tolérance entre les populations. Nous avons donc voulu tester si l'architecture génétique de la tolérance au Zn et Cd étaient la même dans une population non-métallicole. Pour cela nous avons produit un nouveau croisement BC1 entre un individus *A. halleri* d'une population non-métallicole et *A lyrata petraea*. Nous avons ainsi identifié plusieurs QTL de tolérance au Zn et Cd dont certains co-localisent avec ceux obtenu dans les études précédents et d'autres non. Ces résultats suggèrent différents mécanismes de tolérance au sein de l'espèce et une évolution des bases génétiques dans les milieux métallifères. L'acquisition secondaire de la tolérance et de l'hyperaccumulation lors de la colonisation des sites pollués est en cours d'étude grâce à un croisement intraspécifique *A. halleri* métallicole x *A. halleri* non métallicole.

Nouvelles avancées sur l'origine évolutive de de la tolérance aux métaux chez l'espèce pseudométallophyte *Arabidopsis halleri* : que nous apprennent les populations non métallicoles ?

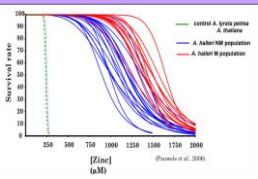
Dima Souleman1, Claire-Lise Meyer2, Loïc Briset2, Nathalie Verbruggen2, Hélène Frérot1

1-Laboratoire Génétique et Evolution des Populations Végétales, UMR CNRS 8198, Bâtiment SN2, Université Lille 1, 59655 Villeneuve d'Ascq Cedex, France
2- Laboratoire de Physiologie et Génétique Moléculaire des plantes, Bâtiment BC, Campus Plaine, Université de Bruxelles, 1050 Bruxelles, Belgique

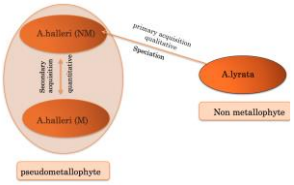
Introduction

De multiples activités humaines sont responsable de la **pollution métallique** (Zn, Pb, Cd, Hg...). Certaines plantes sont adaptées à cette contamination: elles développent des mécanismes de **tolérance**. C'est le cas de l'espèce modèle pseudométallophyte ***Arabidopsis halleri*** (Brassicacée).

La tolérance au Zn est un **trait constitutif** de l'espèce, par opposition à *A. lyrata* ou *A. thaliana*. C'est aussi un trait présentant des **variations quantitatives** entre populations métallicoles et non métallicoles, et au sein de ces mêmes populations.

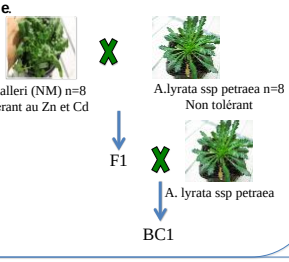


Les mécanismes de tolérance et d'hyperaccumulation ont probablement été **acquis au moment de la spéciation** entre *A. halleri* et *A. lyrata* (variation qualitative entre espèces). L'acquisition secondaire de la surtolérance des populations métallicoles (variation quantitative au sein de l'espèce) correspond probablement à la **colonisation récente** des sites pollués.



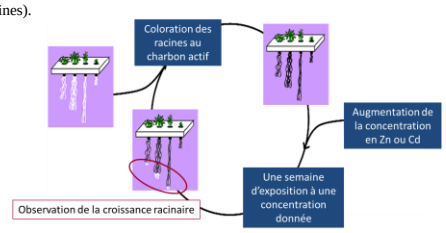
L'objectif:

Il s'agit de détecter les bases génétiques responsables de l'acquisition primaire de la tolérance au Zn et au Cd chez *A.halleri* par QTL Mapping en utilisant un croisement interspécifique **entre *A. lyrata* et *A. halleri* issu de population non métallicole**.

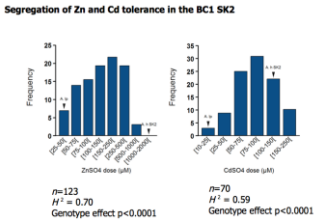


Matériels et Méthodes:

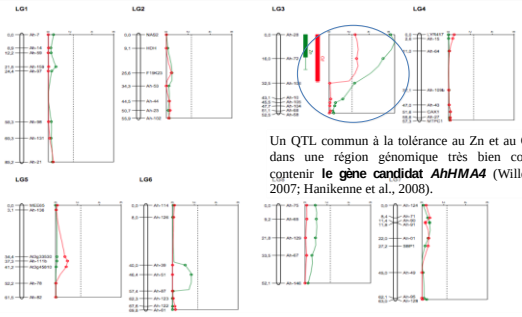
Pour les tests séquentiels de tolérance, 123 et 70 génotypes de la descendance BC1 ont été bouturés de manière à obtenir 3 répliquas par génotype, et ont été cultivés en solution nutritive avec des concentrations croissantes en Zn et en Cd, respectivement, selon les protocoles de Willems et al. (2007) et Courbot et al. (2007). Les valeurs de EC100 sont relevées à chaque pas de concentration, il s'agit de la concentration en Zn ou en Cd à laquelle on n'observe aucune nouvelle racine (100% d'inhibition de la croissance des racines).



Résultats, discussion et perspectives:



QTL Mapping



Un QTL commun à la tolérance au Zn et au Cd apparaît dans une région génomique très bien connue pour contenir **le gène candidat *AhHMA4*** (Willems et al., 2007; Hanikenne et al., 2008).

Le gène *AhHMA4* code pour un transporteur de cations, et a pour fonction de détoxiquer les racines en chargeant le xylème, faisant ainsi monter le métal dans les feuilles (conséquence sur l'hyperaccumulation, voir Frérot et al., 2010). Si ce gène est bien celui qui est révélé ici par l'analyse QTL, il pourrait donc être responsable de l'acquisition primaire de la tolérance chez *A. halleri*. Il reste maintenant à savoir quelles sont les forces évolutives qui sont à l'origine d'une telle modification de l'homéostasie des métaux chez *A. halleri* par rapport aux espèces-sœurs non métallophytes.

Bibliographie

(1) Willems G, Drager D.B, Courbot M, Godé C, Verbruggen N, Saumitou-Laprade P (2007). The genetic basis of zinc tolerance in the metallophyte Arabidopsis halleri ssp. Halleri (Brassicaceae): An analysis of quantitative trait loci. Genetics 176, 659-674. (2) Courbot M, Willems G, Motte P, Arvidsson S, Roosens N, Saumitou-Laprade P, Verbruggen N (2007) A Major Quantitative Trait Locus for Cadmium Tolerance in Arabidopsis halleri Colocalizes with HMA4, a Gene Encoding a Heavy Metal ATPase1. Plant Physiology 144, 1052–1065 **(3)**Fréot H, Faucon M, Willes G, Godé C, Courseaux A, Darracq A, Verbruggen N, Saumitou-Laprade P (2010). Genetic architecture of zinc hyperaccumulation in *Arabidopsis halleri*: The essential role of QTL · environment interactions. *New Phytol*ogist 187, 355-367 (4)Hanikenne M, N. Talke I, J. Haydon M, 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

Society of Environmental Toxicology and Chemistry (SETAC Europe 25th annual meeting – Barcelona 2015)

Development of new microsatellites markers to investigate genetic diversity between metallicolous and non-metallicolous earthworm populations

Dima SOULEMAN[1,2,3], Maxime PAUWELS[1,2], Fabien GRUMIAUX[1,3], Hélène FREROT[1,2], Franck VANDENBULCKE[1,3]

[1] Université de Lille, F-59000 LILLE, France

[2] Laboratoire de Génétique et Evolution des Populations Végétales UMR CNRS 8198, Bâtiment SN2, F-59655 Villeneuve d'Ascq cedex France

[3] LGCgE, EA4515, Univ Lille1, « Fonctionnement des Ecosystèmes Terrestres Anthropisés », SN3, Cité Scientifique, F-59655 Villeneuve d'Ascq, France

E-mail contact : demasouleman@yahoo.com

1 Introduction

Mining, smelting and many anthropic activities can result in heavy metal contamination of urban and agricultural soils. Pollution may induce physiological and genetic changes in organisms living in close contact with soil. Biological responses observed at the individual or infra-individual level of biological organization led to the development of biomarkers. The development of biomarkers consists often in evidencing biological modifications following a contaminant stress in laboratory conditions, using naïve organisms and it is sometimes proposed to use the biological state of individuals from sentinel species collected in the field to evaluate the level of environmental exposure. However, considering the possibility of local adaptation following long-term exposure, response of organisms sampled in the field may substantially differ from laboratory specimens.

In a recent review, we investigated this point focusing on the definition and validity of molecular

biomarkers of metal pollution using earthworms of the Lumbricidae family [1]. It appears that conditions that allow local adaptation to occur are present in earthworms (namely, potential strong and durable selective pressures and within-population genetically determined phenotypic variability). In addition, some infra-individual evidence of adaptation obviously exists. Consequently, there is a need for evolutionary studies investigating the level and distribution of genetic diversity in natural populations of earthworms. Therefore, it seems necessary to integrate the potential effects of local adaptation to Metallic Trace Elements (MTEs) on biological responses of organisms in the development of biomarkers. [1]

Three species of Lumbricidae family belonging to two ecological groups were choosen for this study: *Lumbricus terrestris,* an anecic earthworm which lives in deep vertical burrows and come to surface to feed and *Lumbricus rubellus* and *Lumbricus castaneus*, epigeic earthworms which live in the litter above the soil surface. We assume that exposure of the two groups is different due to their respective ecology. In this work, we decided to investigate a possible local adaptation, owing to pollution, in earthworms from Lumbricidae family populations using microsatellite markers approach. Microsatellite markers may provide datato investigate the genetic differences between earthworm populations issued from polluted and reference sites.

2 Materials and methods

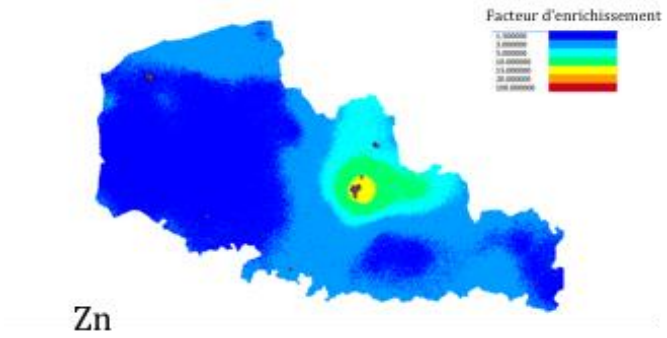
Individuals of Lumbricidae family (*L. terrestris*, *L. rubellus* and *L. castaneus*) used for developping microsatellites were collected in northern and southern France and in southern Poland to build a microsatellite enriched genomic library. Collecting indiviuals in different parts of the species distribution in Europe was supposed to be helpful to develop molecular markers whose efficacity does not depend on any reference genotype. Earthworms were stored in ethanol 96% until use. Total genomic DNA was isolated using NucleoSpin®Tissue (Macherey Nagel). For each species, a mixture of collected DNA was performed. The microsatellite enriched genomic library was built and sequenced by high throughput pyrosequencing (GS FLX®, Roche Diagnostics®) by Genoscreen company. This allows the identification of hundreds of microsatellites loci.Those sequence were used to define molecular markers *in silico*. First, microsatellites identification was performed for *L. terrestris* because of its abundance in the field and because of the relatively easy morphology-based identification of the species. Among 255 microsatellites, twenty markers were selected for a test using PCR reaction kit (QIAGEN) and

DNA sequencer (Applied Biosystem). Finally, eight microsattelite markers were validated using QIAGEN Multiplex PCR Kit and cappillary sequencinq (3730XL®, Applied Biosystems®). A broader validation of the eight markers was then performed on two hundred worms collected in northern France.

3 Results and discussion

Eight new microsatellite markers have been identified in *L. terrestris*. These microsatellite markers are highly polymorphic and can be used in population genetic analysis. Microsatellites were co-amplified using QIAGEN Multiplex PCR Kit and divided into two multiplex PCR reactions. The size of the different alleles was determined using GENEMAPPER 3.7 software (Applied biosystem) (Figure 1A). Microsatellite markers identification is actually ready for *L. rubellus* and a large genotyping (800 individual) of *L. terrestris* is actually ongoing. Large sampling was performed in a workshop area located in Nord-Pas de Calais Region, France. The workshop area exhibits historic contamination by heavy metals. Populations were collected from many sites around Metaleurop factory (a lead smelter) and from reference sites. Metal contamination (Pb, Cd and Zn mainly) is illustrated below by Zn content in soil at the regional scale. (Figure 1B).

marker	size of PCR product	repeat number	repeat motif
Lt3	127	8	ca
Lt8	134	8	tca
Lt16	187	10	gag
Lt19	129	7	gt
Lt10	204	7	gt
Lt17	210	7	ca
Lt9	141	8	ca
Lt12	212	11	aca



(A) **(B)**

Figure 3: A. microsatellite markers identified in this study. B. Sampling sites for earthworm populations collected in northern France

4 Conclusion

Eight microsatellite markers were identified in *L. terrestris*. The same approach of microsatellite identification is ongoing in *L. rubellus* and *L. castaneus*. Microsatellite markers are used to investigate the differences of microsatellite alleles size between metallicolous and non-metallicolous of *L. terrestris* populations. These microsatellite markers are essential tools to understand the evolutionary dynamics of populations and direct and indirect selective forces that shape genetics variation in natural populations. We hypothesized that probable lose of genetic variation may occur due to selection pressure and genetic drift in population belong to polluted area.

5 References

[1] Pauwels M, Frérot H, Souleman D, Vandenbulcke F. 2013. using biomarkers in an evolutionary context : lessons from the analysis of biological responses of oligochaete annelids to metal exposure. Enivronmental pollution 179 :343-350.
[2] Bourennane H, Douay F, Sterckeman T. 2010 Mapping of anthropogenic trace elements inputs in agricultural topsoil from Northern France using enrichment factors. Geoderma 157:165-174.

[3] Velavan T.P, Schulenburg H, Michiels N.K. 2007 Development and characterization of novel microsatellite markers for the common earthworm (Lumbricus terrestris). Molecular Ecology Notes 7:1060-1062.

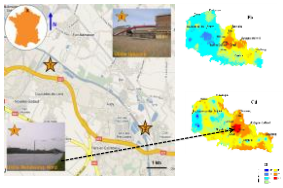
Development of new microsatellites markers to investigate genetic diversity between metallicolous and non-metallicolous earthworm populations

Dima **SOULEMAN**[1,2,3], Maxime **PAUWELS**[1,2], Fabien **GRUMIAUX**[1,3], Hélène **FREROT**[1,2], Franck **VANDENBULCKE**[1,3]

[1] **Université de Lille,** F-59000 LILLE, France
[2] **Laboratoire de Génétique et Evolution des Populations Végétales** UMR CNRS 8198, Bâtiment SN2, F- 59655 Villeneuve d'Ascq cedex France
[3] **LGCgE,** EA4515, Univ Lille1, « Fonctionnement des Ecosystèmes Terrestres Anthropisés », SN3, Cité Scientifique, F-59655 Villeneuve d'Ascq, France

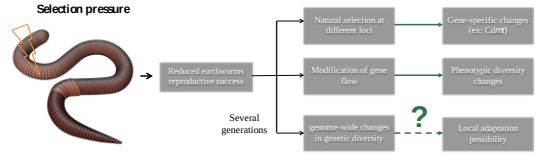
demasouleman@yahoo.com

Introduction



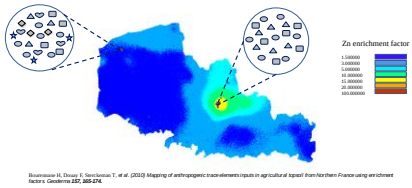
Heavy metal pollution is often encountered in areas of intensive industries like smelting and mining. Organisms living in close contact with contaminants are often affected. Usually, shortly after exposure, organisms exhibit acclimation responses (physiological adaptation) which may consist in changes observed at the infra-individual level (mainly molecular, biochemical and cellular parameters). Many parameters showing variation following exposure to a contaminant are pollution biomarker candidates. Therefore, a long-term exposure may not only cause environmental phenotypic variation (acclimation) among individuals but also adaptative evolution unlike short-term exposure of experimental approaches (Hendry et al., 2011; Pauwels et al., 2013).

At the field, contaminants may not usually result in either populations decline because of toxic effects or in the loss of population size due to the reproductive success reducing and this process can lead to genetic variability diminution through genetic drift, natural selection and gene flow alteration. Several methods are evoked to explain the loss of genetic diversity. One of them consists in using neutral genetic markers at the genome-wide scale such as microsatellite markers which are widely used in population genetic studies and may provide data to investigate the genetic diversity within and between populations. These microsatellite markers are essential tools to understand the evolutionary dynamics of populations and to understand direct and indirect selective forces that shape genetics variation in natural populations.

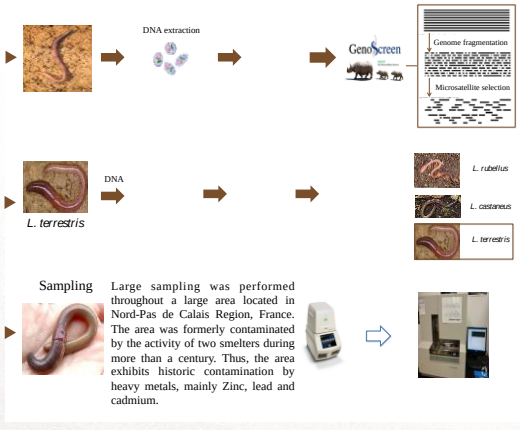


Earthworms from the Lumbricidae represent good candidates for soil biomonitoring (Rombke et al., 2005). We chose the anecic species *Lumbricus terrestris* to investigate the possible local adaptation due to metals using microsatellite markers.

Objectives

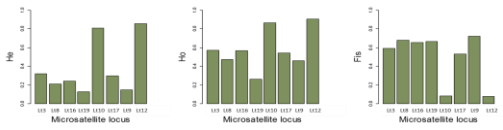


Bourennane H, Douay F, Sterckeman T, et al. (2010) Mapping of anthropogenic trace elements inputs in agricultural topsoil from Northern France using enrichment factors. Geoderma 157, 165-174.



Large sampling was performed throughout a large area located in Nord-Pas de Calais Region, France. The area was formerly contaminated by the activity of two smelters during more than a century. Thus, the area exhibits historic contamination by heavy metals, mainly Zinc, lead and cadmium.

In this study, we succeed to identify and to validate eight polymorphic microsatellite loci which can be use to explore the genetic diversity of field collected populations of *L.terrestris*. Microsatellite loci do not depend on any reference genotype and have been designed to be used throughout the entire distribution area of the species. Then, allelic diversity was assessed using 192 individuals. A total of 116 alleles were determined. The number of different alleles per locus ranged from 5 alleles at locus Lt8 to 25 alleles at locus Lt9 with an average of 6.125 for all loci. The observed heterozygosity (H_o) ranged from 0.128 to 0.855, whereas the expected heterozygosity (H_E) ranged from 0.262 to 0.904.



marker	size of PCR product	repeat number	repeat motif
Lt3	127	8	ca
Lt8	134	8	tca
Lt16	187	10	gag
Lt19	129	7	gt
Lt10	204	7	gt
Lt17	210	7	ca
Lt9	141	8	ca
Lt12	212	11	aca

microsatellite markers identified in this study

F_{IS} values obtained by Fstat version 2.9.3 analysis revealed that 6 of 8 loci could have been affected by one or more null alleles. F_{IS} values varied among markers from 0,082 (Lt163) to 0,722 (Lt12). Hight F_{IS} values for loci (Lt3, Lt8, Lt16, Lt19, Lt9, Lt12) may related to the possibility to appearance of null alleles .

Perspectives and conclusion

- Eight microsatellite markers have been identified in L. terrestris.
- We are actually performing the genetic diversity analysis using these 8 markers on populations which have been collected from many sites around Metaleurop factory (a lead smelter) and from reference sites.
- The same approach of microsatellite identification is ongoing in L. rubellus and L. castaneus.

References

Pauwels M, Frérot H, Souleman D, Vandenbulcke F. 2013. using biomarkers in an evolutionary context : lessons from the analysis of biological responses of oligochaete annelids to metal exposure. Environmental pollution 179 :343-350.
Rombke J, Jansch S, Didden W (2005) The use of earthworms in ecological soil classification and assessment concepts. Ecotoxicology and Environmental Safety 62 (2):24 doi:10.1016/j.ecoenv.2005.03.027.
Bourennane H, Douay F, Sterckeman T, et al. (2010) Mapping of anthropogenic trace elements inputs in agricultural topsoil from Northern France using enrichment factors. Geoderma 157, 165-174.

International Conference of Ecological Sciences, Marseille- 2016

QTL mapping for zinc tolerance at the intra-specific level in Arabidopsis halleri using a F2 progeny

Marie-Joe KARAM, Dima SOULEMAN, Sophie GALLINA, Marc HANIKENNE, Sol SCHVARTZMAN, Maxime PAUWELS, Hélène FRÉROT

Arabidopsis halleri is a well-studied species for its ability to tolerate and hyperaccumulate metals, particularly zinc and cadmium. Deciphering the genetic architecture of these quantitative traits has only been done at the inter-specific level using *A. lyrata* ssp. *petraea* x *A. halleri* backcross and F2 progenies. The major QTL regions identified included candidate genes implied in the traits such as genes from the *hma4* (heavy-metal transporting ATPase 4) and the *mtp1* (metal tolerance protein 1) families. So far, no QTL mapping study has been done at the intra-specific level. We aimed at studying the genetic architecture of zinc tolerance and hyperaccumulation in an *A. halleri* F2 progeny (175 individuals) by crossing individuals coming from polluted and non-polluted soils. 384 single nucleotide polymorphism (SNP) markers issued from high-throughput sequencing of the genomes of parental plants have been developed and used for the genetic map construction. Growth parameters and photosynthetic yield were performed under control and polluted conditions in hydroponics (10 *vs.* 2000 µM of $ZnSO_4$). One major QTL was identified for photosynthetic yield that explained about 30% of the variation for the trait. This QTL region does not co-localize with the known candidate genes for metal tolerance that have been identified so far at the inter-specific level. This suggests that, in *A. halleri*, molecular mechanisms involved in local adaptation to metal-polluted soils at the intra-specific level differ from those that distinguish the level of metal tolerance observed in the species from those observed in its non-tolerant close relatives.

Keywords: *Arabidopsis halleri*; zinc tolerance; QTL mapping; intraspecific F2 cross; Single Nucleotide Polymorphism

Annex chapter 5:

Assessment of Zn hyperaccumulation

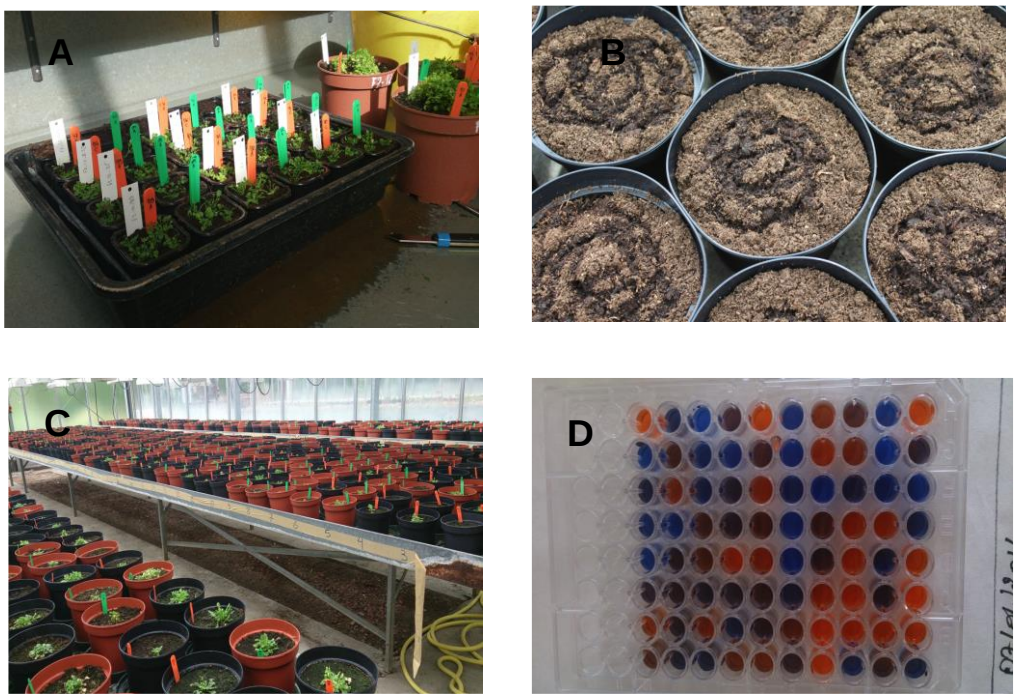


Figure 1: Experimental design to assess Zn hyperaccumulation. A) Replicates on compost. B) Adding Zn to polluted soil. C) Culture on compost to evaluate Zn hyper-accumulation. D) Zn dosage in shoots using Zincon method.

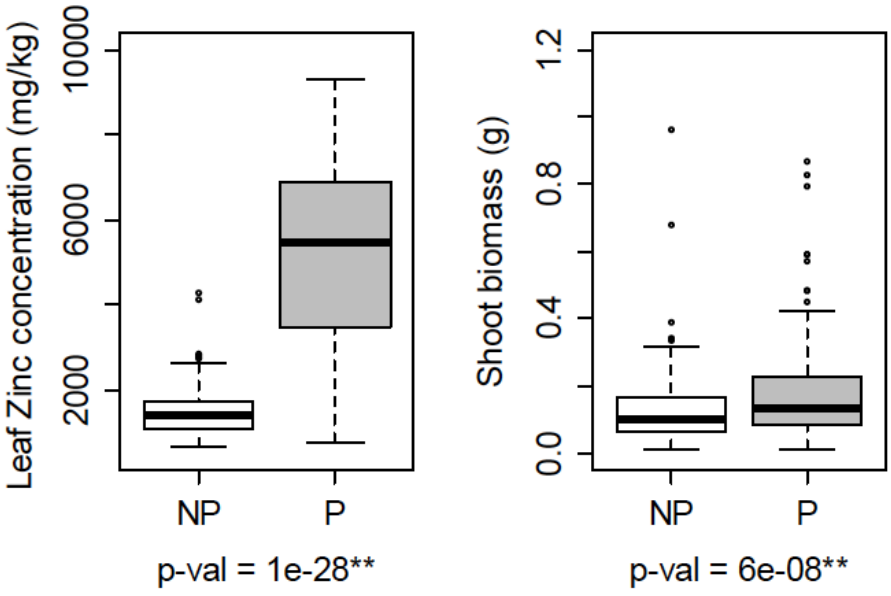


Figure 2: Traits distributions in the non-polluted and polluted conditions for all genotypes in the hyper-accumulation experiment

NP = non-polluted condition; P = Zn-polluted condition.

P-values are the results of the Wilcoxon statistical test for paired data.

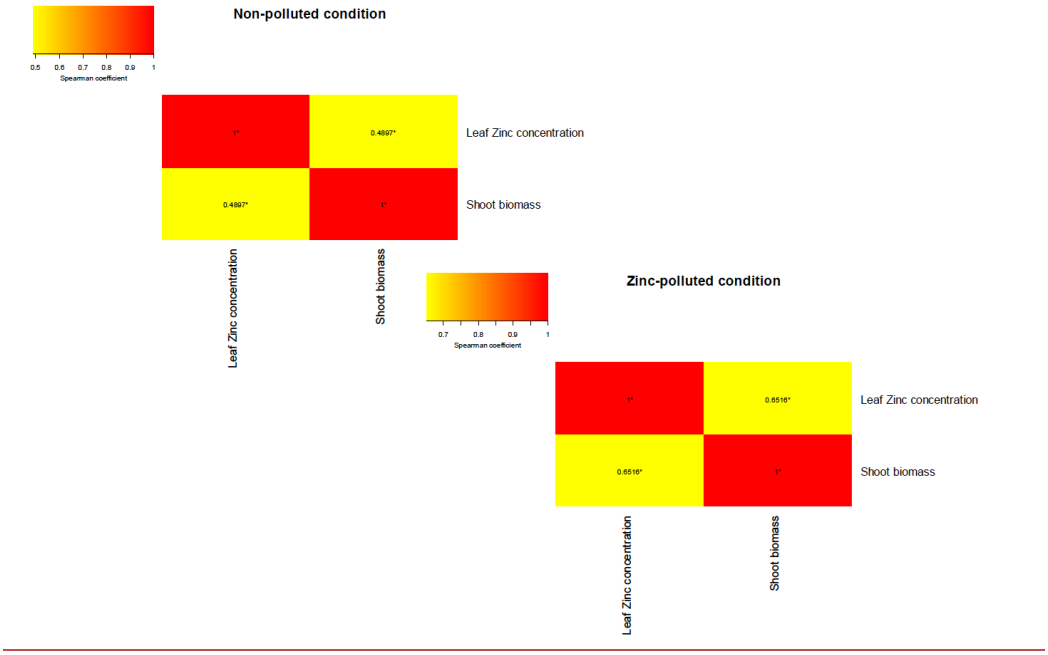


Figure 3: Correlation rates between the traits in the non-polluted and Zn-polluted conditions in the hyper-accumulation experiment (α = 0.05)

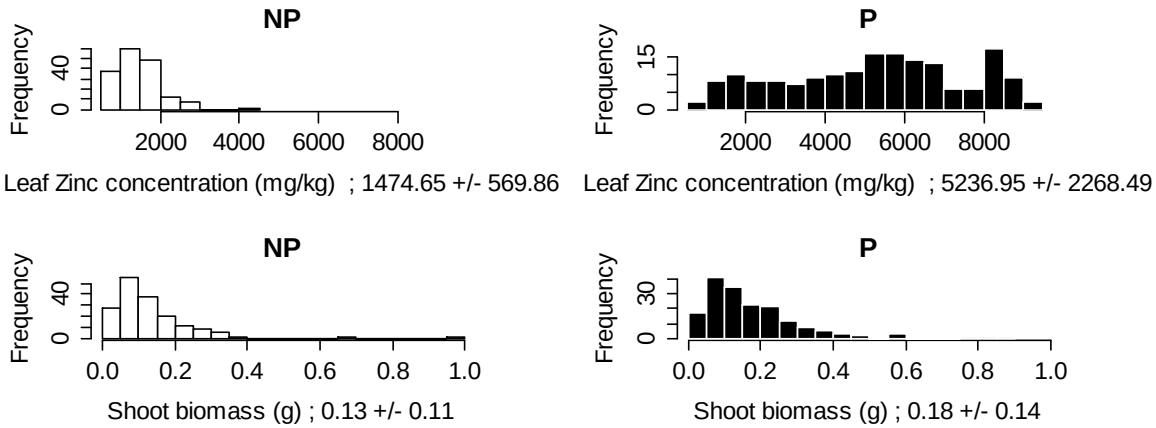


Figure 4: Traits segregation histograms in the non-polluted and polluted conditions in the hyper-accumulation experiment for all genotypes

NP = non-polluted condition; P = Zn-polluted condition.
Below each histogram, the trait mean +/- the standard deviation is written.

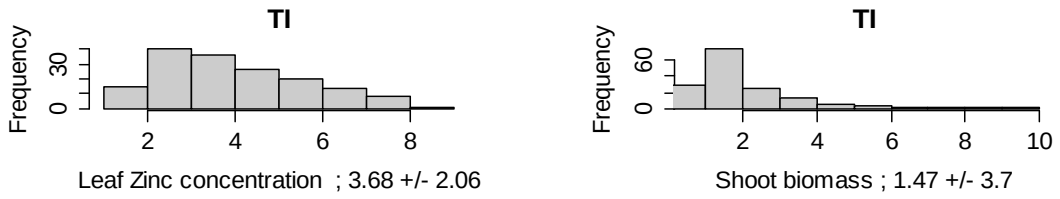


Figure 5: Segregation histograms of the traits tolerance indices from the first to the ninth decile for all genotypes in the hyper-accumulation experiment

TI = tolerance index.
Below each histogram, the traits TI median +/- the standard deviation are written.

Trait	Block effect	H^2 in the non-polluted condition	H^2 in the polluted condition
Leaf Zn concentration ($mg.Kg^{-1}$)	No	0.08	0.22
Shoot biomass (g)	No	0.32	0.39

Table 1 : Traits broad-sense heritabilities (H^2) in the hyperaccumulation experiment

REFERENCES

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. **2008**. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

AlanJM Baker, SP McGrath, RogerD Reeves, JAC Smith. **1999**. Metal Hyperaccumulator Plants. Phytoremediation of Contaminated Soil and Water. CRC Press, .

Allee W., Park, Emerson A., Park T, Schmidt K. 1949. Principles of animal ecology.

Al-Shehbaz IA, O'Kane SL. **2002**. Taxonomy and Phylogeny of Arabidopsis (Brassicaceae). *The Arabidopsis Book*: e0001.

Amossé J, Dózsa-Farkas K, Boros G, Rochat G, Sandoz G, Fournier B, Mitchell EAD, Le Bayon R-C. **2016**. Patterns of earthworm, enchytraeid and nematode diversity and community structure in urban soils of different ages. *European Journal of Soil Biology* **73**: 46–58.

Anderson CJ, Kille P, Lawlor AJ, Spurgeon DJ. **2013**. Life-history effects of arsenic toxicity in clades of the earthworm Lumbricus rubellus. *Environmental Pollution* **172**: 200–207.

Antonovics J, Bradshaw AD, Turner RG. **1971**. Heavy Metal Tolerance in Plants. In: Cragg JB, ed. Advances in Ecological Research. Academic Press, 1–85.

Arnold RE, Hodson ME, Langdon CJ. **2008**. A Cu tolerant population of the earthworm *Dendrodrilus rubidus* (Savigny, 1862) at Coniston Copper Mines, Cumbria, UK. *Environmental Pollution* **152**: 713–722.

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.

Baillie J, Hilton-Taylor C, Stuart SN. **2004**. *2004 IUCN Red List of Threatened Species: A Global Species Assessment*. IUCN.

Baird SK, Kurz T, Brunk UT. **2006**. Metallothionein protects against oxidative stress-induced lysosomal destabilization. *Biochemical Journal* **394**: 275–283.

Baker AJM. **1981**. Accumulators and excluders - strategies in the response of plants to heavy metals. *Journal of Plant Nutrition* **3**: 643–654.

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

Baker AJM, McGrath SP, Sidoli CMD, Reeves RD. **1994**. Environmental biotechnology in waste treatment and recycling The possibility of in situ heavy metal decontamination of polluted soils using crops of metal-accumulating plants. *Resources, Conservation and Recycling* **11**: 41–49.

Baker AJM, Walker P. **1989**. *Ecophsiology Of Metal Uptake By Tolerant Plants*. CRC Press.

Barata C, Baird DJ, Mitchell SE, Soares AMVM. **2002**. Among- and within-population variability in tolerance to cadmium stress in natural populations of *Daphnia magna*: Implications for ecological risk assessment. *Environmental Toxicology and Chemistry* **21**: 1058–1064.

Barrett RDH, Schluter D. **2008**. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Bartlewicz J, Vandepitte K, Jacquemyn H, Honnay O. **2015**. Population genetic diversity of the clonal self-incompatible herbaceous plant *Linaria vulgaris* along an urbanization gradient. *Biological Journal of the Linnean Society* **116**: 603–613.

Baumbach H. **2012**. *Metallophytes and Metallicolous Vegetation: Evolutionary Aspects, Taxonomic Changes and Conservational Status in Central Europe*.

Beattie JH, Owen HLH, Wallace SM, Arthur JR, Kwun I-S, Hawksworth GM, Wallace

HM. **2005**. Metallothionein overexpression and resistance to toxic stress. *Toxicology Letters* **157**: 69–78.
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* **37**: 251–268.
Bechsgaard JS, Castric V, Charlesworth D, Vekemans X, Schierup MH. **2006**. The Transition to Self-Compatibility in *Arabidopsis thaliana* and Evolution within S-Haplotypes over 10 Myr. *Molecular Biology and Evolution* **23**: 1741–1750.
Becker T, Dierschke H. **2008**. Vegetation response to high concentrations of heavy metals in the Harz Mountains, Germany. *Phytocoenologia* **38**: 255–265.
Beeby A. **2001**. What do sentinels stand for? *Environmental Pollution* **112**: 285–298.
Bergelson J, Roux F. **2010**. Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. *Nature Reviews Genetics* **11**: 867–879.
Bergl RA, Vigilant L. **2007**. Genetic analysis reveals population structure and recent migration within the highly fragmented range of the Cross River gorilla (Gorilla gorilla diehli). *Molecular Ecology* **16**: 501–516.
Bernard F, Brulle F, Douay F, Lemière S, Demuynck S, Vandenbulcke F. **2010**. Metallic trace element body burdens and gene expression analysis of biomarker candidates in *Eisenia fetida*, using an 'exposure/depuration' experimental scheme with field soils. *Ecotoxicology and Environmental Safety* **73**: 1034–1045.
Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN. **2004**. Genetic Signatures of Strong Recent Positive Selection at the Lactase Gene. *The American Journal of Human Genetics* **74**: 1111–1120.
Bert V, Bonnin I, Saumitou-Laprade P, De Laguérie P, Petit D. **2002**. Do *Arabidopsis halleri* from nonmetallicolous populations accumulate zinc and cadmium more effectively than those from metallicolous populations? *New Phytologist* **155**: 47–57.
Bert V, Macnair MR, de Laguerie P, Saumitou-Laprade P, Petit D. **2000**. Zinc Tolerance and Accumulation in Metallicolous and Nonmetallicolous Populations of *Arabidopsis halleri* (Brassicaceae). *The New Phytologist* **146**: 225–233.
Bert V, Meerts P, Saumitou-Laprade P, Salis P, Gruber W, Verbruggen N. **2003**. Genetic basis of Cd tolerance and hyperaccumulation in *Arabidopsis halleri*. *Plant and Soil* **249**: 9–18.
Beyer WN, Chaney RL, Mulhern BM. **1982**. Heavy Metal Concentrations in Earthworms From Soil Amended with Sewage Sludge1. *Journal of Environment Quality* **11**: 381.
Bickham JW. **2011**. The four cornerstones of Evolutionary Toxicology. *Ecotoxicology* **20**: 497–502.
Blanquart F, Kaltz O, Nuismer SL, Gandon S. **2013**. A practical guide to measuring local adaptation. *Ecology Letters* **16**: 1195–1205.
Bouché M. 1972. Lombriciens de France écologie et systématique.
Bourret V, Couture P, Campbell PGC, Bernatchez L. **2008**. Evolutionary ecotoxicology of wild yellow perch (*Perca flavescens*) populations chronically exposed to a polymetallic gradient. *Aquatic Toxicology* **86**: 76–90.
Bradl H. **2005**. *Heavy Metals in the Environment: Origin, Interaction and Remediation*. Academic Press.
Brausch JM, Smith PN. **2009**. Pesticide resistance from historical agricultural chemical exposure in *Thamnocephalus platyurus* (Crustacea: Anostraca). *Environmental Pollution* **157**: 481–487.

Briskine RV, Paape T, Shimizu-Inatsugi R, Nishiyama T, Akama S, Sese J, Shimizu KK. **2016**. Genome assembly and annotation of Arabidopsis halleri, a model for heavy metal hyperaccumulation and evolutionary ecology. *Molecular Ecology Resources*: n/a-n/a.
Broman KW, Wu H, Sen Ś, Churchill GA. **2003**. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**: 889–890.
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.
Brulle F, Cocquerelle C, Mitta G, Castric V, Douay F, Leprêtre A, Vandenbulcke F. **2008**. Identification and expression profile of gene transcripts differentially expressed during metallic exposure in *Eisenia fetida* coelomocytes. *Developmental & Comparative Immunology* **32**: 1441–1453.
Brulle F, Mitta G, Cocquerelle C, Vieau D, Lemière S, Leprêtre A, Vandenbulcke F. **2006**. Cloning and Real-Time PCR Testing of 14 Potential Biomarkers in *Eisenia fetida* Following Cadmium Exposure. *Environmental Science & Technology* **40**: 2844–2850.
Brulle F, Morgan AJ, Cocquerelle C, Vandenbulcke F. **2010**. Transcriptomic underpinning of toxicant-mediated physiological function alterations in three terrestrial invertebrate taxa: A review. *Environmental Pollution* **158**: 2793–2808.
Cagnolo L, Cabido M, Valladares G. **2006**. Plant species richness in the Chaco Serrano Woodland from central Argentina: Ecological traits and habitat fragmentation effects. *Biological Conservation* **132**: 510–519.
Calisi A, Lionetto MG, Schettino T. **2009**. Pollutant-induced alterations of granulocyte morphology in the earthworm *Eisenia foetida*. *Ecotoxicology and Environmental Safety* **72**: 1369–1377.
Calisi A, Lionetto MG, Schettino T. **2011**. Biomarker response in the earthworm *Lumbricus terrestris* exposed to chemical pollutants. *Science of The Total Environment* **409**: 4456–4464.
Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden T. **2009**. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
Capowiez Y, Bérard A. **2006**. Assessment of the effects of imidacloprid on the behavior of two earthworm species (*Aporrectodea nocturna* and *Allolobophora icterica*) using 2D terraria. *Ecotoxicology and Environmental Safety* **64**: 198–206.
Capowiez Y, Dittbrenner N, Rault M, Triebskorn R, Hedde M, Mazzia C. **2010**. Earthworm cast production as a new behavioural biomarker for toxicity testing. *Environmental Pollution* **158**: 388–393.
Capowiez Y, Rault M, Mazzia C, Belzunces L. **2003**. Earthworm behaviour as a biomarker – a case study using imidacloprid. *Pedobiologia* **47**: 542–547.
Caro G, Decaëns T, Lecarpentier C, Mathieu J. **2013**. Are dispersal behaviours of earthworms related to their functional group? *Soil Biology and Biochemistry* **58**: 181–187.
Chan KY. **2001**. An overview of some tillage impacts on earthworm population abundance and diversity — implications for functioning in soils. *Soil and Tillage Research* **57**: 179–191.
Chapuis M-P, Estoup A. **2007**. Microsatellite Null Alleles and Estimation of Population Differentiation. *Molecular Biology and Evolution* **24**: 621–631.
Charlesworth D, Willis JH. **2009**. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**: 783–796.
Christensen O. **1988**. Lumbricid earthworms as bio-indicators relative to soil factors in different agro-ecosystems. : 839–849.
Churchill GA, Doerge RW. **1994**. Empirical Threshold Values for Quantitative Trait

Mapping. *Genetics Society of Americ* **Genetics**: 963–971.
Clemens S, Palmgren MG, Krämer U. **2002**. A long way ahead: understanding and engineering plant metal accumulation. *Trends in Plant Science* **7**: 309–315.
Clements WH, Rohr JR. **2009**. Community responses to contaminants: Using basic ecological principles to predict ecotoxicological effects. *Environmental Toxicology and Chemistry* **28**: 1789–1800.
Costa D, Mariën J, Janssens TKS, Gestel CAM van, Driessen G, Sousa JP, Straalen NM van, Roelofs D. **2012**. Influence of adaptive evolution of cadmium tolerance on neutral and functional genetic variation in *Orchesella cincta*. *Ecotoxicology* **21**: 2078–2087.
Costa D, Timmermans MJTN, Sousa JP, Ribeiro R, Roelofs D, Van Straalen NM. **2013**. Genetic structure of soil invertebrate populations: Collembolans, earthworms and isopods. *Applied Soil Ecology* **68**: 61–66.
Cotrufo MF, De Santo AV, Alfani A, Bartoli G, De Cristofaro A. **1995**. Effects of urban heavy metal pollution on organic matter decomposition in Quercus ilex L. Woods. *Environmental Pollution* **89**: 81–87.
Coulon A. **2010**. Genhet: an easy-to-use R function to estimate individual heterozygosity. *Molecular Ecology Resources* **10**: 167–169.
Coulon A, Cosson JF, Angibault JM, Cargnelutti B, Galan M, Morellet N, Petit E, Aulagnier S, Hewison AJM. **2004**. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual–based approach. *Molecular Ecology* **13**: 2841–2850.
Courbot M, Willems G, Motte P, Arvidsson S, Roosens N, Saumitou-Laprade P, Verbruggen N. **2007**. A Major Quantitative Trait Locus for Cadmium Tolerance in Arabidopsis halleri Colocalizes with HMA4, a Gene Encoding a Heavy Metal ATPase. *Plant Physiology* **144**: 1052–1065.
Coutellec M-A, Barata C. **2011**. An introduction to evolutionary processes in ecotoxicology. *Ecotoxicology* **20**: 493–496.
Dallinger R, Höckner M. **2013**. Evolutionary concepts in ecotoxicology: tracing the genetic background of differential cadmium sensitivities in invertebrate lineages. *Ecotoxicology* **22**: 767–778.
Danchin É, Pocheville A. **2014**. Inheritance is where physiology meets evolution. *The Journal of Physiology* **592**: 2307–2317.
Darío J. Díaz Cosín MN. **2010**. Reproduction of Earthworms: Sexual Selection and Parthenogenesis. : 69–86.
Darwin C. **1859**. darwin - Google Scholar. *On the origin ofspecies by means of natural selection, or the preservation of favoured races in the struggle for life*.
Decaëns T. **2010**. Macroecological patterns in soil communities. *Global Ecology and Biogeography* **19**: 287–302.
Decombeix I. **2011**. Etude de l'adaptation aux milieux calaminaires chez *Arabidopsis halleri* :approche écologique, génétique et phénotypique.
Demuynck S, Grumiaux F, Descamps M, Leprêtre A. **2005**. Differential expression of stress protein in response to heat and cadmium in two oligochaete species *Eisenia fetida* and *E.andrei.*
Demuynck S, Grumiaux F, Mottier V, Schikorski D, Lemière S, Leprêtre A. **2006**. Metallothionein response following cadmium exposure in the oligochaete *Eisenia fetida. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **144**: 34–46.
Demuynck S, Grumiaux F, Mottier V, Schikorski D, Lemière S, Leprêtre A. **2007**. Cd/Zn

exposure interactions on metallothionein response in *Eisenia fetida* (Annelida, Oligochaeta). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* **145**: 658–668.
Dickinson NM, Turner AP, Watmough SA, Lepp NW. **1992**. Acclimation of Trees to Pollution Stress: Cellular Metal Tolerance Traits. *Annals of Botany* **70**: 569–572.
Domínguez J, Velando A, Aira M, Monroy F. **2003**. Uniparental reproduction of *Eisenia fetida* and E. andrei (Oligochaeta: Lumbricidae): Evidence of self-insemination. *Pedobiologia* **47**: 530–534.
Donihue CM, Lambert MR. **2014**. Adaptive evolution in urban ecosystems. *AMBIO* **44**: 194–203.
Douay F, Pruvot C, Roussel H, Ciesielski H, Fourrier H, Proix N, Waterlot C. **2007**. Contamination of Urban Soils in an Area of Northern France Polluted by Dust Emissions of Two Smelters. *Water, Air, and Soil Pollution* **188**: 247–260.
Dräger DB, Desbrosses-Fonrouge A-G, Krach C, Chardonnens AN, Meyer RC, Saumitou-Laprade P, Krämer U. **2004**. Two genes encoding *Arabidopsis halleri* MTP1 metal transport proteins co-segregate with zinc tolerance and account for high MTP1 transcript levels. *The Plant Journal* **39**: 425–439.
DREAL. **2013**. L'occupation du sol en Nord - Pas-de-Calais - DREAL (Direction régionale de l'environnement de l'aménagement et du logement) HAUTS-DE-FRANCE.
Dupont L. **2009**. Perspectives on the application of molecular genetics to earthworm ecology. *Pedobiologia* **52**: 191–205.
Dupont L, Grésille Y, Richard B, Decaëns T, Mathieu J. **2015**. Dispersal constraints and fine-scale spatial genetic structure in two earthworm species. *Biological Journal of the Linnean Society* **114**: 335–347.
Dupont L, Lazrek F, Porco D, King RA, Rougerie R, Symondson WOC, Livet A, Richard B, Decaëns T, Butt KR, *et al.* 2011. New insight into the genetic structure of the *Allolobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data. *Pedobiologia* **54**: 217–224.
Edwards CA, Bohlen PJ. **1996**. *Biology and Ecology of Earthworms*. Springer Science & Business Media.
Edwards CA, Dominguez J, Neuhauser EF. **1998**. Growth and reproduction of Perionyx excavatus (Perr.) (Megascolecidae) as factors in organic waste management. *Biology and Fertility of Soils* **27**: 155–161.
Eijsackers H. **2011**. Earthworms as colonizers of natural and cultivated soil environments. *Applied Soil Ecology* **50**: 1–13.
Ellis EC, Klein Goldewijk K, Siebert S, Lightman D, Ramankutty N. **2010**. Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography* **19**: 589–606.
Ellis EC, Ramankutty N. **2008**. Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment* **6**: 439–447.
Ellis EC, Wang H, Xiao HS, Peng K, Liu XP, Li SC, Ouyang H, Cheng X, Yang LZ. **2006**. Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. *Remote Sensing of Environment* **100**: 457–473.
El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM. **2014**. Genotype × environment interaction QTL mapping in plants: lessons from Arabidopsis. *Trends in Plant Science* **19**: 390–398.
Enckell PH, Bengtson S-A, Douwes P, Niklasson M, Stille B, Wiman B. **1986**. The dispersal pattern of an anthropochorous species: Genetic variation in populations of *Lumbricus terrestris*

L. (Lumbricidae) in the Faroe Islands. *Hereditas* **104**: 253–261.
Ernst G, Zimmermann S, Christie P, Frey B. **2008**. Mercury, cadmium and lead concentrations in different ecophysiological groups of earthworms in forest soils. *Environmental Pollution* **156**: 1304–1313.
Escarré J, Lefèbvre C, Frérot H, Mahieu S, Noret N. **2013**. Metal concentration and metal mass of metallicolous, non metallicolous and serpentine *Noccaea caerulescens* populations, cultivated in different growth media. *Plant and Soil* **370**: 197–221.
Evanno G, Regnaut S, Goudet J. **2005**. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14**: 2611–2620.
Fahrig L. **2003**. Effects of Habitat Fragmentation on Biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **34**: 487–515.
Fauvergue X, Vercken E, Malausa T, Hufbauer RA. **2012**. The biology of small, introduced populations, with special reference to biological control. *Evolutionary Applications* **5**: 424–443.
Findlay CS, Houlahan J. **1997**. Anthropogenic Correlates of Species Richness in Southeastern Ontario Wetlands. *Conservation Biology* **11**: 1000–1009.
Fischer J, Lindenmayer DB. **2007**. Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography* **16**: 265–280.
Fisker KV, Sørensen JG, Damgaard C, Pedersen KL, Holmstrup M. **2011**. Genetic adaptation of earthworms to copper pollution: is adaptation associated with fitness costs in *Dendrobaena octaedra*? *Ecotoxicology* **20**: 563–573.
Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. **1994**. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294–299.
Fourie F, Reinecke SA, Reinecke AJ. **2007**. The determination of earthworm species sensitivity differences to cadmium genotoxicity using the comet assay. *Ecotoxicology and Environmental Safety* **67**: 361–368.
Frankham R. **2003**. Genetics and conservation biology. *Comptes Rendus Biologies* **326**, **Supplement 1**: 22–29.
Frankham R. **2005**. Genetics and extinction. *Biological Conservation* **126**: 131–140.
Frérot H, Faucon M-P, Willems G, Godé C, Courseaux A, Darracq A, Verbruggen N, Saumitou-Laprade P. **2010**. Genetic architecture of zinc hyperaccumulation in Arabidopsis halleri: the essential role of QTL × environment interactions. *New Phytologist* **187**: 355–367.
Fritsch C, Cœurdassier M, Giraudoux P, Raoul F, Douay F, Rieffel D, Vaufleury A de, Scheifler R. **2011**. Spatially Explicit Analysis of Metal Transfer to Biota: Influence of Soil Contamination and Landscape. *PLOS ONE* **6**: e20682.
Fründ H-C, Butt K, Capowiez Y, Eisenhauer N, Emmerling C, Ernst G, Potthoff M, Schädler M, Schrader S. **2010**. Using earthworms as model organisms in the laboratory: Recommendations for experimental implementations. *Pedobiologia* **53**: 119–125.
Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, Mustofa MS, Samakkarn U, Settheetham-Ishida W, Ishida T, *et al.* 2008. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Human Molecular Genetics* **17**: 835–843.
Futuyma. **1998**. Futuyma(1998). Evolutionary Biology - Google Scholar.
Gao C, Ren X, Mason AS, Li J, Wang W, Xiao M, Fu D. **2013**. Revisiting an important component of plant genomes: microsatellites. *Functional Plant Biology* **40**: 645–661.
Garant D, Forde SE, Hendry AP. **2007**. The multifarious effects of dispersal and gene flow on

contemporary adaptation. *Functional Ecology* **21**: 434–443.
García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. **2012**. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* **28**: 2678–2679.
Garnier-Géré P, Chikhi L. **2001**. Population Subdivision, Hardy–Weinberg Equilibrium and the Wahlund Effect. eLS. John Wiley & Sons, Ltd, .
Gestel CAMV, Brummelen TCV. **1996**. Incorporation of the biomarker concept in ecotoxicology calls for a redefinition of terms. *Ecotoxicology* **5**: 217–225.
Ghoshal K, Wang Y, Sheridan JF, Jacob ST. **1998**. Metallothionein Induction in Response to Restraint Stress TRANSCRIPTIONAL CONTROL, ADAPTATION TO STRESS, AND ROLE OF GLUCOCORTICOID. *Journal of Biological Chemistry* **273**: 27904–27910.
Glémin S. **2003**. How Are Deleterious Mutations Purged? Drift Versus Nonrandom Mating. *Evolution* **57**: 2678–2687.
Godé C, Decombeix I, Kostecka A, Wasowicz P, Pauwels M, Courseaux A, Saumitou-Laprade P. **2012**. Nuclear microsatellite loci for *Arabidopsis halleri* (Brassicaceae), a model species to study plant adaptation to heavy metals. *American Journal of Botany* **99**: e49–e52.
Goudet J. **1995**. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity* **86**: 485–486.
Grigoropoulou N, Butt KR. **2010**. Field investigations of *Lumbricus terrestris* spatial distribution and dispersal through monitoring of manipulated, enclosed plots. *Soil Biology and Biochemistry* **42**: 40–47.
Gruber C, Stürzenbaum S, Gehrig P, Sack R, Hunziker P, Berger B, Dallinger R. **2000**. Isolation and characterization of a self-sufficient one-domain protein. *European Journal of Biochemistry* **267**: 573–582.
Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F, *et al.* 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11**: 591–611.
Hamrick JL, Murawski DA, Nason JD. **1993**. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. In: Fleming TH,, In: Estrada A, eds. Advances in vegetation science. Frugivory and seed dispersal: ecological and evolutionary aspects. Springer Netherlands, 281–297.
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.
Hardy OJ, Vekemans X. **1999**. Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* **83**: 145–154.
Hardy OJ, Vekemans X. **2002**. spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**: 618–620.
Harper GL, Cesarini S, Casey SP, Morgan AJ, Kille P, Bruford MW. **2006**. Microsatellite markers for the earthworm *Lumbricus rubellus*. *Molecular Ecology Notes* **6**: 325–327.
Hassan Z, Aarts MGM. **2011**. Opportunities and feasibilities for biotechnological improvement of Zn, Cd or Ni tolerance and accumulation in plants. *Environmental and Experimental Botany* **72**: 53–63.
Hautekèete N-C, Frachon L, Luczak C, Toussaint B, Van Landuyt W, Van Rossum F, Piquot Y. **2015**. Habitat type shapes long-term plant biodiversity budgets in two densely populated regions in north-western Europe. *Diversity and Distributions* **21**: 631–642.

Heidi & Hans-Juergen Koch. Common earthworm, nightcrawler, Lumbricus terrestris | Heidi & Hans-Juergen Koch.

Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, *et al.* 2011. Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology Resources* **11**: 123–136.

Hendry AP, Kinnison MT, Heino M, Day T, Smith TB, Fitt G, Bergstrom CT, Oakeshott J, Jørgensen PS, Zalucki MP, *et al.* 2011. Evolutionary principles and their practical application. *Evolutionary Applications* **4**: 159–183.

Hoffmann AA, Hercus MJ. **2000**. Environmental Stress as an Evolutionary Force. *BioScience* **50**: 217–226.

Hohmann N, Schmickl R, Chiang T-Y, Lučanová M, Kolář F, Marhold K, Koch MA. **2014**. Taming the wild: resolving the gene pools of non-model Arabidopsis lineages. *BMC Evolutionary Biology* **14**.

Holderegger R, Wagner HH. **2008**. Landscape Genetics. *BioScience* **58**: 199–207.

Hoogerkamp M, Rogaar H, Eijsackers HJP. **1983**. Effect of earthworms on grassland on recently reclaimed polder soils in the Netherlands. In: Satchell JE, ed. Earthworm Ecology. Springer Netherlands, 85–105.

Hopkin. **1989**. Ecophysiology of Metals in Terrestrial Invertebrates.

Hu TT, Pattyn P, Bakker EG, Cao J, Cheng J-F, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H, *et al.* 2011. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nature Genetics* **43**: 476–481.

James SW, Porco D, Decaëns T, Richard B, Rougerie R, Erséus C. **2010**. DNA Barcoding Reveals Cryptic Diversity in *Lumbricus terrestris* L., 1758 (Clitellata): Resurrection of L. herculeus (Savigny, 1826). *PLOS ONE* **5**: e15629.

Janssens TKS, Lopéz R del R, Mariën J, Timmermans MJTN, Montagne-Wajer K, van Straalen NM, Roelofs D. **2008**. Comparative Population Analysis of Metallothionein Promoter Alleles Suggests Stress-induced Microevolution in the Field. *Environmental Science & Technology* **42**: 3873–3878.

Johnson MTJ, Thompson KA, Saini HS. **2015**. Plant evolution in the urban jungle. *American Journal of Botany* **102**: 1951–1953.

Kaegi JH. 1993. Evolution, structure and chemical activity of class I metallothioneins: an overview.

Kautenburger R. **2006**. Genetic structure among earthworms (*Lumbricus terrestris L.*) from different sampling sites in western Germany based on random amplified polymorphic DNA. *Pedobiologia* **50**: 257–266.

Kawabe A, Miyashita NT. **2003**. DNA Polymorphism in Active Gene and Pseudogene of the Cytosolic Phosphoglucose Isomerase (PgiC) Loci in *Arabidopsis halleri ssp. gemmifera*. *Molecular Biology and Evolution* **20**: 1043–1050.

Kawecki TJ, Ebert D. **2004**. Conceptual issues in local adaptation. *Ecology Letters* **7**: 1225–1241.

Kimura R, Yamaguchi T, Takeda M, Kondo O, Toma T, Haneji K, Hanihara T, Matsukusa H, Kawamura S, Maki K, *et al.* 2009. A Common Variation in EDAR Is a Genetic Determinant of Shovel-Shaped Incisors. *The American Journal of Human Genetics* **85**: 528–535.

Kinnison MT, Hendry AP. **2001**. The pace of modern life II: From rates of contemporary microevolution to pattern and process. In: Hendry AP,, In: Kinnison MT, eds. Contemporary

Issues in Genetics and Evolution. Microevolution Rate, Pattern, Process. Springer Netherlands, 145–164.

Klaassen CD, Jie Liu, Choudhuri and S. **1999**. METALLOTHIONEIN: An Intracellular Protein to Protect Against Cadmium Toxicity. *Annual Review of Pharmacology and Toxicology* **39**: 267–294.

Klerks PL. **2002**. Adaptation, Ecological Impacts, and Risk Assessment: Insights from Research at Foundry Cove, Bayou Trepagnier, and Pass Fourchon. *Human and Ecological Risk Assessment: An International Journal* **8**: 971–982.

Klerks PL, Levinton JS. **1989**. Rapid Evolution of Metal Resistance in a Benthic Oligochaete Inhabiting a Metal-polluted Site. *The Biological Bulletin* **176**: 135–141.

Klerks PL, Xie L, Levinton JS. **2011**. Quantitative genetics approaches to study evolutionary processes in ecotoxicology; a perspective from research on the evolution of resistance. *Ecotoxicology* **20**: 513–523.

Koch M, Haubold B, Mitchell-Olds T. **2001**. Molecular systematics of the Brassicaceae: evidence from coding plastidic matK and nuclear Chs sequences. *American Journal of Botany* **88**: 534–544.

Koene JM, Pförtner T, Michiels NK. **2005**. Piercing the partner's skin influences sperm uptake in the earthworm *Lumbricus terrestris*. *Behavioral Ecology and Sociobiology* **59**: 243.

Kosambi DD. **1943**. The Estimation of Map Distances from Recombination Values. *Annals of Eugenics* **12**: 172–175.

Kostecka AA. **2009**. *Adaptations of* Arabidopsis halleri *to habitats rich in heavy metals in southern Poland*. Lille 1.

Kozlov MV, Zvereva EL. **2011**. A second life for old data: Global patterns in pollution ecology revealed from published observational studies. *Environmental Pollution* **159**: 1067–1075.

Krämer U. **2010**. Metal Hyperaccumulation in Plants. *Annual Review of Plant Biology* **61**: 517–534.

Krauss J, Bommarco R, Guardiola M, Heikkinen RK, Helm A, Kuussaari M, Lindborg R, Öckinger E, Pärtel M, Pino J, *et al.* 2010. Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. *Ecology Letters* **13**: 597–605.

Kruess A, Tscharntke T. **2000**. Species richness and parasitism in a fragmented landscape: experiments and field studies with insects on *Vicia sepium*. *Oecologia* **122**: 129–137.

Langdon CJ, Morgan AJ, Charnock JM, Semple KT, Lowe CN. **2009**. As-resistance in laboratory-reared F1, F2 and F3 generation offspring of the earthworm *Lumbricus rubellus* inhabiting an As-contaminated mine soil. *Environmental Pollution* **157**: 3114–3119.

Langmead B, Trapnell C, Pop M, Salzberg S. **2009**. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: R25.

Lapied E, Moudilou E, Exbrayat J-M, Oughton DH, Joner EJ. **2010**. Silver nanoparticle exposure causes apoptotic response in the earthworm *Lumbricus terrestris* (Oligochaeta). *Nanomedicine* **5**: 975–984.

Lavelle P, Spain A. **2001**. *Soil Ecology*. Springer Science & Business Media.

Lemoine G. **2016**. Chapter 15 - Brownfield Restoration as a Smart Economic Growth Option for Promoting Ecotourism, biodiversity, and Leisure: Two Case Studies in Nord-Pas De Calais A2 - Prasad, M.N.V. Bioremediation and Bioeconomy. Elsevier, 361–388.

Lévêque T, Capowiez Y, Schreck E, Mombo S, Mazzia C, Foucault Y, Dumat C. **2015**. Effects of historic metal(loid) pollution on earthworm communities. *Science of The Total Environment* **511**: 738–746.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup 1000 Genome Project Data Processing. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.
Liao VH-C, Freedman JH. **1998**. Cadmium-regulated Genes from the NematodeCaenorhabditis elegans IDENTIFICATION AND CLONING OF NEW CADMIUM-RESPONSIVE GENES BY DIFFERENTIAL DISPLAY. *Journal of Biological Chemistry* **273**: 31962–31970.
Librado P, Rozas J. **2009**. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
Lienert J. **2004**. Habitat fragmentation effects of fitness of plant populations - A review. **12**: 53–72.
Ligthart TN, Peek GJCW. **1997**. Evolution of earthworm burrow systems after inoculation of lumbricid earthworms in a pasture in the Netherlands. *Soil Biology and Biochemistry* **29**: 453–462.
Lopez S, Rousset F, Shaw FH, Shaw RG, Ronce O. **2009**. Joint Effects of Inbreeding and Local Adaptation on the Evolution of Genetic Load after Fragmentation. *Conservation Biology* **23**: 1618–1627.
Lowe CN, Butt KR. **2002**. Influence of organic matter on earthworm production and behaviour: a laboratory-based approach with applications for soil restoration. *European Journal of Soil Biology* **38**: 173–176.
Lowe CN, Butt KR. **2007**. Life-cycle traits of the dimorphic earthworm species *Allolobophora chlorotica* (Savigny, 1826) under controlled laboratory conditions. *Biology and Fertility of Soils* **43**: 495–499.
Lukkari T, Taavitsainen M, Väisänen A, Haimi J. **2004**. Effects of heavy metals on earthworms along contamination gradients in organic rich soils. *Ecotoxicology and Environmental Safety* **59**: 340–348.
Lumbricus terrestris Linnaeus, 1758 -- Discover Life.
Lynch M, Walsh B. **1998**. *Genetics and Analysis of Quantitative Traits*. Sinauer.
Mackay TFC, Stone EA, Ayroles JF. **2009**. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* **10**: 565–577.
Macnair MR. **1981**. The Uptake of Copper by Plants of Mimulus Guttatus Differing in Genotype Primarily at a Single Major Copper Tolerance Locus. *New Phytologist* **88**: 723–730.
MacNair M. **1983**. Heredity - Abstract of article: The genetic control of copper tolerance in the yellow monkey flower, *Mimulus guttatus*. *Heredity* **50**: 283–293.
Macnair MR. **1987**. Heavy metal tolerance in plants: A model evolutionary system. *Trends in Ecology & Evolution* **2**: 354–359.
Macnair MR. **1993**. The genetics of metal tolerance in vascular plants. *New Phytologist* **124**: 541–559.
Macnair MR. **2002**. Within and between population genetic variation for zinc accumulation in *Arabidopsis halleri*. *New Phytologist* **155**: 59–66.
Macnair MR, Smirnoff N. **1999**. Use of zincon to study uptake and accumulation of zinc by zinc tolerant and hyperaccumulating plants. *Communications in Soil Science and Plant Analysis* **30**: 1127–1136.
Maestri E, Marmiroli M, Visioli G, Marmiroli N. **2010**. Metal tolerance and hyperaccumulation: Costs and trade-offs between traits and environment. *Environmental and Experimental Botany* **68**: 1–13.
Malausa T, Gilles A, Meglécz E, Blanquart H, Duthoy S, Costedoat C, Dubut V, Pech N,

Castagnone-Sereno P, Délye C, *et al.* 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* **11**: 638–644.
Manel S, Schwartz MK, Luikart G, Taberlet P. **2003**. Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**: 189–197.
Manios T, Stentiford EI, Millner PA. **2003**. The effect of heavy metals accumulation on the chlorophyll concentration of Typha latifolia plants, growing in a substrate containing sewage sludge compost and watered with metaliferus water. *Ecological Engineering* **20**: 65–74.
Marhan S, Scheu S. **2005**. The influence of mineral and organic fertilisers on the growth of the endogeic earthworm *Octolasion tyrtaeum* (Savigny). *Pedobiologia* **49**: 239–249.
Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.
Mathieu J, Barot S, Blouin M, Caro G, Decaëns T, Dubs F, Dupont L, Jouquet P, Nai P. **2010**. Habitat quality, conspecific density, and habitat pre-use affect the dispersal behaviour of two earthworm species, *Aporrectodea icterica* and *Dendrobaena veneta*, in a mesocosm experiment. *Soil Biology and Biochemistry* **42**: 203–209.
McDonnell MJ, Hahs AK. **2015**. Adaptation and Adaptedness of Organisms to Urban Environments. *Annual Review of Ecology, Evolution, and Systematics* **46**: 261–280.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky 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**: 1297–1303.
Medina MH, Correa JA, Barata C. **2007**. Micro-evolution due to pollution: Possible consequences for ecosystem responses to toxic stress. *Chemosphere* **67**: 2105–2114.
de Meeûs T, Goudet J. **2007**. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. *Infection, Genetics and Evolution* **7**: 731–735.
Merriam G. **1984**. Connectivity: a fundamental ecological characteristic of landscape pattern. *Methodology in landscape ecological research and planning : proceedings, 1st seminar, International Association of Landscape Ecology, Roskilde, Denmark, Oct 15-19, 1984 / eds. J. Brandt, P. Agger*.
Metzker ML. **2009**. Sequencing technologies - the next generation. *Nature Reviews Genetics* **11**: 31–46.
Meyer C-L, Juraniec M, Huguet S, Chaves-Rodriguez E, Salis P, Isaure M-P, Goormaghtigh E, Verbruggen N. **2015**. Intraspecific variability of cadmium tolerance and accumulation, and cadmium-induced cell wall modifications in the metal hyperaccumulator Arabidopsis halleri. *Journal of Experimental Botany*: erv144.
Meyer C-L, Kostecka AA, Saumitou-Laprade P, Créach A, Castric V, Pauwels M, Frérot H. **2010**. Variability of zinc tolerance among and within populations of the pseudometallophyte species *Arabidopsis halleri* and possible role of directional selection. *New Phytologist* **185**: 130–142.
Meyer C-L, Pauwels M, Briset L, Godé C, Salis P, Bourceaux A, Souleman D, Frérot H, Verbruggen N. **2016**. Potential preadaptation to anthropogenic pollution: evidence from a common quantitative trait locus for zinc and cadmium tolerance in metallicolous and nonmetallicolous accessions of *Arabidopsis halleri*. *New Phytologist*: n/a-n/a.
Millward RN, Klerks PL. **2002**. Contaminant-Adaptation and Community Tolerance in Ecological Risk Assessment: Introduction. *Human and Ecological Risk Assessment: An*

International Journal **8**: 921–932.
Moffett BF, Nicholson FA, Uwakwe NC, Chambers BJ, Harris JA, Hill TCJ. **2003**. Zinc contamination decreases the bacterial diversity of agricultural soil. *FEMS Microbiology Ecology* **43**: 13–19.
Møller AP, Nielsen JT. **2010**. Fear screams and adaptation to avoid imminent death: effects of genetic variation and predation. *Ethology Ecology & Evolution* **22**: 183–202.
Morgan AJ, Kille P, Stürzenbaum SR. **2007**. Microevolution and Ecotoxicology of Metals in Invertebrates. *Environmental Science & Technology* **41**: 1085–1096.
Morgan AJ, Stürzenbaum SR, Winters C, Grime GW, Aziz NAA, Kille P. **2004**. Differential metallothionein expression in earthworm (*Lumbricus rubellus)* tissues. *Ecotoxicology and Environmental Safety* **57**: 11–19.
Mustonen M, Haimi J, Väisänen A, Knott KE. **2014**. Metallothionein gene expression differs in earthworm populations with different exposure history. *Ecotoxicology* **23**: 1732–1743.
Nahmani J, Hodson ME, Black S. **2007**. Effects of metals on life cycle parameters of the earthworm Eisenia fetida exposed to field-contaminated, metal-polluted soils. *Environmental Pollution* **149**: 44–58.
Nahmani J, Lavelle P. **2002**. Effects of heavy metal pollution on soil macrofauna in a grassland of Northern France. *European Journal of Soil Biology* **38**: 297–300.
Nahmani J, Lavelle P, Lapied E, van Oort F. **2003**. Effects of heavy metal soil pollution on earthworm communities in the north of France. *Pedobiologia* **47**: 663–669.
Nawrot T, Plusquin M, Hogervorst J, Roels HA, Celis H, Thijs L, Vangronsveld J, Van Hecke E, Staessen JA. **2006**. Environmental exposure to cadmium and risk of cancer: a prospective population-based study. *The Lancet Oncology* **7**: 119–126.
Ndegwa PM, Thompson SA. **2001**. Integrating composting and vermicomposting in the treatment and bioconversion of biosolids. *Bioresource Technology* **76**: 107–112.
Nei M, Maruyama T, Chakraborty R. **1975**. The Bottleneck Effect and Genetic Variability in Populations. *Evolution* **29**: 1–10.
Nielsen R. **2005**. Molecular Signatures of Natural Selection. *Annual Review of Genetics* **39**: 197–218.
Nielsen R, Paul JS, Albrechtsen A, Song YS. **2011**. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**: 443–451.
Novo M, Almodóvar A, Fernández R, Trigo D, Díaz Cosín DJ. **2010**. Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Molecular Phylogenetics and Evolution* **56**: 507–512.
Nuutinen V, Butt KR. **1997**. The mating behaviour of the earthworm *Lumbricus terrestris* (Oligochaeta: Lumbricidae). *Journal of Zoology* **242**: 783–798.
O'Brien DJ, Kaneene JB, Poppenga RH. **1993**. The use of mammals as sentinels for human exposure to toxic contaminants in the environment. *Environmental Health Perspectives* **99**: 351–368.
O'Kane SL, Al-Shehbaz IA. **1997**. A Synopsis of Arabidopsis (Brassicaceae). *Novon* **7**: 323–327.
Palmiter RD. **1998**. The elusive function of metallothioneins. *Proceedings of the National Academy of Sciences* **95**: 8428–8430.
Park J-H, Yamaguchi T, Watanabe C, Kawaguchi A, Haneji K, Takeda M, Kim Y-I, Tomoyasu Y, Watanabe M, Oota H, *et al.* 2012. Effects of an Asian-specific nonsynonymous EDAR variant on multiple dental traits. *Journal of Human Genetics* **57**: 508–514.

Pauwels M, Frérot H, Bonnin I, Saumitou-Laprade P. **2006**. A broad-scale analysis of population differentiation for Zn tolerance in an emerging model species for tolerance study: *Arabidopsis halleri* (Brassicaceae). *Journal of Evolutionary Biology* **19**: 1838–1850.

Pauwels M, Frérot H, Souleman D, Vandenbulcke F. **2013**. Using biomarkers in an evolutionary context: Lessons from the analysis of biological responses of oligochaete annelids to metal exposure. *Environmental Pollution* **179**: 343–350.

Pauwels M, Roosens N, Frérot H, Saumitou-Laprade P. **2008**. When population genetics serves genomics: putting adaptation back in a spatial and historical context. *Current Opinion in Plant Biology* **11**: 129–134.

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.

Pelosi C, Barot S, Capowiez Y, Hedde M, Vandenbulcke F. **2013**. Pesticides and earthworms. A review. *Agronomy for Sustainable Development* **34**: 199–228.

Peralta-Videa JR, Lopez ML, Narayan M, Saupe G, Gardea-Torresdey J. **2009**. The biochemistry of environmental heavy metal uptake by plants: Implications for the food chain. *The International Journal of Biochemistry & Cell Biology* **41**: 1665–1677.

Pérès G, Vandenbulcke F, Guernion M, Hedde M, Beguiristain T, Douay F, Houot S, Piron D, Richard A, Bispo A, *et al.* 2011. Earthworm indicators as tools for soil monitoring, characterization and risk assessment. An example from the national Bioindicator programme (France). *Pedobiologia* **54**, **Supplement**: S77–S87.

Phylogenetic species delimitation of the earthworms Eisenia fetida (Savigny, 1826) and Eisenia andrei Bouché, 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences.

Piearce TG, Langdon CJ, Meharg AA, Semple KT. **2002**. Yellow earthworms: distinctive pigmentation associated with arsenic- and copper-tolerance in *Lumbricus rubellus*. *Soil Biology and Biochemistry* **34**: 1833–1838.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A. **2004**. GENECLASS2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity* **95**: 536–539.

Pižl V, Josens G. **1995**. Earthworm communities along a gradient of urbanization. *Environmental Pollution* **90**: 7–14.

Polus E, Vandewoestijne S, Choutt J, Baguette M. **2006**. Tracking the effects of one century of habitat loss and fragmentation on calcareous grassland butterfly communities. *Biodiversity and Conservation* **16**: 3423–3436.

Posthuma L, Van Straalen NM. **1993**. Heavy-metal adaptation in terrestrial invertebrates: A review of occurrence, genetics, physiology and ecological consequences. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology* **106**: 11–38.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**: 945–959.

Radespiel U, Bruford MW. **2013**. Fragmentation genetics of rainforest animals: insights from

recent studies. *Conservation Genetics* **15**: 245–260.
Reznick DN, Ghalambor CK. **2001**. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* **112–113**: 183–198.
Ribeiro R, Lopes I. **2013**. Contaminant driven genetic erosion and associated hypotheses on alleles loss, reduced population growth rate and increased susceptibility to future stressors: an essay. *Ecotoxicology* **22**: 889–899.
Ricketts HJ, Morgan AJ, Spurgeon DJ, Kille P. **2004**. Measurement of annetocin gene expression: a new reproductive biomarker in earthworm ecotoxicology. *Ecotoxicology and Environmental Safety* **57**: 4–10.
Roelofs D, Aarts MGM, Schat H, Van Straalen NM. **2008**. Functional ecological genomics to demonstrate general and specific responses to abiotic stress. *Functional Ecology* **22**: 8–18.
Römbke J, Jänsch S, Didden W. **2005**. The use of earthworms in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety* **62**: 249–265.
Roosens NHCJ, Willems G, Godé C, Courseaux A, Saumitou-Laprade P. **2007**. The use of comparative genome analysis and syntenic relationships allows extrapolating the position of Zn tolerance QTL regions from *Arabidopsis halleri* into *Arabidopsis thaliana*. *Plant and Soil* **306**: 105–116.
Roosens NHCJ, Willems G, Saumitou-Laprade P. **2008**. Using Arabidopsis to explore zinc tolerance and hyperaccumulation. *Trends in Plant Science* **13**: 208–215.
Rosenberg NA. **2004**. distruct: a program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137–138.
Rosenberg MS, Anderson CD. **2011**. PASSaGE: Pattern Analysis, Spatial Statistics and Geographic Exegesis. Version 2. *Methods in Ecology and Evolution* **2**: 229–232.
Roux C, Castric V, Pauwels M, Wright SI, Saumitou-Laprade P, Vekemans X. **2011**. Does Speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* Coincide with Major Changes in a Molecular Target of Adaptation? *PLoS ONE* **6**: e26872.
Rożen A. **2006**. Effect of cadmium on life-history parameters in *Dendrobaena octaedra* (Lumbricidae: Oligochaeta) populations originating from forests differently polluted with heavy metals. *Soil Biology and Biochemistry* **38**: 489–503.
Rutgers M, Orgiazzi A, Gardi C, Römbke J, Jänsch S, Keith AM, Neilson R, Boag B, Schmidt O, Murchie AK, *et al.* 2016. Mapping earthworm communities in Europe. *Applied Soil Ecology* **97**: 98–111.
Salt DE, Kraemer U. **2000**. Mechanisms of metal hyperaccumulation in plants. In: Raskin I,, In: Ensley BD, eds. Phytoremediation of Toxic Metals: Using Plants to Clean UP the Environment. New York: John Wiley & Sons, 231–245.
Sambatti JBM, Rice KJ. **2006**. Local Adaptation, Patterns of Selection, and Gene Flow in the Californian Serpentine Sunflower (helianthus Exilis). *Evolution* **60**: 696–710.
Sanchez-Hernandez JC. **2006**. Earthworm Biomarkers in Ecological Risk Assessment. In: In: Ware DGW,, In: Whitacre DM,, In: Albert LA,, In: Voogt P de,, In: Gerba CP,, In: Hutzinger O,, In: Knaak JB,, In: Mayer FL,, In: Morgan DP,, In: Park DL, et al., eds. Reviews of Environmental Contamination and Toxicology. Reviews of Environmental Contamination and Toxicology. Springer New York, 85–126.
Satchell JE. **1967**. Lumbricidae. *Soil Biology*: 259–322.
Saunders DA, Hobbs RJ, Margules CR. **1991**. Biological Consequences of Ecosystem Fragmentation: A Review. *Conservation Biology* **5**: 18–32.

Savolainen O, Pyhäjärvi T, Knürr T. **2007**. Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics* **38**: 595–619.
van der Schalie WH, Gardner HS, Bantle JA, De Rosa CT, Finch RA, Reif JS, Reuter RH, Backer LC, Burger J, Folmar LC, *et al.* 1999. Animals as sentinels of human health hazards of environmental chemicals. *Environmental Health Perspectives* **107**: 309–315.
Schat H, Bookum WT. **1992**. Heredity - Abstract of article: Genetic control of copper tolerance in Silene vulgaris. *Heredity* **68**: 219–229.
Schoener TW. **2011**. The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics. *Science* **331**: 426–429.
Seeber J, Seeber GUH, Kössler W, Langel R, Scheu S, Meyer E. **2005**. Abundance and trophic structure of macro-decomposers on alpine pastureland (Central Alps, Tyrol): effects of abandonment of pasturing. *Pedobiologia* **49**: 221–228.
Seto KC, Güneralp B, Hutyra LR. **2012**. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences* **109**: 16083–16088.
Sgrò CM, Hoffmann AA. **2004**. Genetic correlations, tradeoffs and environmental variation. *Heredity* **93**: 241–248.
Shahzad Z, Gosti F, Frérot H, Lacombe E, Roosens N, Saumitou-Laprade P, Berthomieu P. **2010**. The Five AhMTP1 Zinc Transporters Undergo Different Evolutionary Fates towards Adaptive Evolution to Zinc Tolerance in Arabidopsis halleri. *PLoS Genet* **6**: e1000911.
Słomka A, Sutkowska A, Szczepaniak M, Malec P, Mitka J, Kuta E. **2011**. Increased genetic diversity of Viola tricolor L. (Violaceae) in metal-polluted environments. *Chemosphere* **83**: 435–442.
Smith BA, Egeler P, Gilberg D, Hendershot W, Stephenson GL. **2010**. Uptake and Elimination of Cadmium and Zinc by *Eisenia andrei* During Exposure to Low Concentrations in Artificial Soil. *Archives of Environmental Contamination and Toxicology* **59**: 264–273.
Somers CM, Neudorf K, Jones KL, Lance SL. **2011**. Novel microsatellite loci for the compost earthworm *Eisenia fetida*: A genetic comparison of three North American vermiculture stocks. *Pedobiologia* **54**: 111–117.
Souleman D, Grumiaux F, Frérot H, Vandenbulcke F, Pauwels M. **2016**. Isolation and characterization of eight polymorphic microsatellites markers for the earthworm *Lumbricus terrestris*. *European Journal of Soil Biology* **74**: 76–80.
Spurgeon DJ, Hopkin SP. **1996**. Effects of Metal-Contaminated Soils on the Growth, Sexual Development, and Early Cocoon Production of the Earthworm *Eisenia fetida*,with Particular Reference to Zinc. *Ecotoxicology and Environmental Safety* **35**: 86–95.
Spurgeon DJ, Hopkin SP. **1999**. Tolerance to Zinc in Populations of the Earthworm Lumbricus rubellus from Uncontaminated and Metal-Contaminated Ecosystems. *Archives of Environmental Contamination and Toxicology* **37**: 332–337.
Spurgeon DJ, Lister L, Kille P, Pereira MG, Wright J, Svendsen C. **2011**. Toxicokinetic studies reveal variability in earthworm pollutant handling. *Pedobiologia* **54**, **Supplement**: S217–S222.
Spurgeon DJ, Svendsen C, Kille P, Morgan AJ, Weeks JM. **2004**. Responses of earthworms (*Lumbricus rubellus*) to copper and cadmium as determined by measurement of juvenile traits in a specifically designed test system. *Ecotoxicology and Environmental Safety* **57**: 54–64.
Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA. **2007**. Simple sequence repeats in mycobacterial genomes. *Journal of Biosciences* **32**: 3–15.

Stein RJ, Höreth S, de Melo JRF, Syllwasschy L, Lee G, Garbin ML, Clemens S, Krämer U. **2016**. Relationships between soil and leaf mineral composition are element-specific, environment-dependent and geographically structured in the emerging model *Arabidopsis halleri*. *New Phytologist*: n/a-n/a.

Sterckeman T, Douay F, Proix N, Fourrier H. **2000**. Vertical distribution of Cd, Pb and Zn in soils near smelters in the North of France. *Environmental Pollution* **107**: 377–389.

Sterckeman T, Douay F, Proix N, Fourrier H, Perdrix E. Assessment of the Contamination of Cultivated Soils by Eighteen Trace Elements Around Smelters in the North of France. *Water, Air, and Soil Pollution* **135**: 173–194.

Sterenborg I, Roelofs D. **2003**. Field-selected cadmium tolerance in the springtail *Orchesella cincta* is correlated with increased metallothionein mRNA expression. *Insect Biochemistry and Molecular Biology* **33**: 741–747.

Storfer A, Murphy MA, Evans JS, Goldberg CS, Robinson S, Spear SF, Dezzani R, Delmelle E, Vierling L, Waits LP. **2006**. Putting the 'landscape' in landscape genetics. *Heredity* **98**: 128–142.

Stork NE, Eggleton P. **1992**. Invertebrates as determinants and indicators of soil quality. *American Journal of Alternative Agriculture* **7**: 38–47.

Straalen NM van, Timmermans MJTN. **2002**. Genetic Variation in Toxicant-Stressed Populations: An Evaluation of the 'Genetic Erosion' Hypothesis. *Human and Ecological Risk Assessment: An International Journal* **8**: 983–1002.

van Straalen NM. **1998**. Evaluation of bioindicator systems derived from soil arthropod communities. *Applied Soil Ecology* **9**: 429–437.

Strunk H, Hochkirch A, Veith M, Hankeln T, Emmerling C. **2012**. Isolation and characterization of eleven polymorphic microsatellite markers for the earthworm *Aporrectodea longa* (Ude). *European Journal of Soil Biology* **48**: 56–58.

Stürzenbaum SR, Georgiev O, Morgan AJ, Kille P. **2004**. Cadmium Detoxification in Earthworms: From Genes to Cells. *Environmental Science & Technology* **38**: 6283–6289.

Sunnucks P. **2000**. Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**: 199–203.

Suthar S, Singh S, Dhawan S. **2008**. Earthworms as bioindicator of metals (Zn, Fe, Mn, Cu, Pb and Cd) in soils: Is metal bioaccumulation affected by their ecological category? *Ecological Engineering* **32**: 99–107.

Tamura K, Dudley J, Nei M, Kumar S. **2007**. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* **24**: 1596–1599.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. **2008**. Synteny and Collinearity in Plant Genomes. *Science* **320**: 486–488.

Tardieu F, Tuberosa R. **2010**. Dissection and modelling of abiotic stress tolerance in plants. *Current Opinion in Plant Biology* **13**: 206–212.

Templeton AR, Shaw K, Routman E, Davis SK. **1990**. The Genetic Consequences of Habitat Fragmentation. *Annals of the Missouri Botanical Garden* **77**: 13–27.

The *Arabidopsis* Genome Initiative (AGI). **2000**. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Timmermans MJTN, Ellers J, Roelofs D, Straalen NM van. **2005**. Metallothionein mRNA Expression and Cadmium Tolerance in Metal-stressed and Reference Populations of the Springtail *Orchesella cincta*. *Ecotoxicology* **14**: 727–739.

Tischer, S. **2009**. Earthworms (Lumbricidae) as bioindicators: the relationship between in-soil

and in.
Tiunov AV, Hale CM, Holdsworth AR, Vsevolodova-Perel TS. **2006**. Invasion Patterns of Lumbricidae Into the Previously Earthworm-free Areas of Northeastern Europe and the Western Great Lakes Region of North America. *Biological Invasions* **8**: 1223–1234.
Tolrà RP, Poschenrieder C, Luppi B, Barceló J. **2005**. Aluminium-induced changes in the profiles of both organic acids and phenolic substances underlie Al tolerance in *Rumex acetosa L*. *Environmental and Experimental Botany* **54**: 231–238.
Torres-Leguizamon M, Mathieu J, Decaëns T, Dupont L. **2014**. Genetic Structure of Earthworm Populations at a Regional Scale: Inferences from Mitochondrial and Microsatellite Molecular Markers in *Aporrectodea icterica* (Savigny 1826). *PLOS ONE* **9**: e101597.
Uba S, Uzairu A, Okunola OJ. **2009**. Content of Heavy Metals in *Lumbricus Terrestris* and Associated Soils in Dump Sites.
Van Calster H, Vandenberghe R, Ruysen M, Verheyen K, Hermy M, Decocq G. **2008**. Unexpectedly high 20th century floristic losses in a rural landscape in northern France. *Journal of Ecology* **96**: 927–936.
Van Rossum F, Bonnin I, Fénart S, Pauwels M, Petit D, Saumitou-Laprade P. **2004**. Spatial genetic structure within a metallicolous population of *Arabidopsis halleri,* a clonal, self-incompatible and heavy-metal-tolerant species. *Molecular Ecology* **13**: 2959–2967.
Varshney RK, Graner A, Sorrells ME. **2005**. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48–55.
Velavan TP, Schulenburg H, Michiels NK. **2007**. Development and characterization of novel microsatellite markers for the common earthworm (*Lumbricus terrestris L.*). *Molecular Ecology Notes* **7**: 1060–1062.
Velavan TP, Weller S, Schulenburg H, Michiels NK. **2009**. High genetic diversity and heterogeneous parasite load in the earthworm Lumbricus terrestris on a German meadow. *Soil Biology and Biochemistry* **41**: 1591–1595.
Verbruggen N, Hermans C, Schat H. **2009**. Molecular mechanisms of metal hyperaccumulation in plants. *New Phytologist* **181**: 759–776.
Virgós E. **2001**. Role of isolation and habitat quality in shaping species abundance: a test with badgers (*Meles meles L.)* in a gradient of forest fragmentation. *Journal of Biogeography* **28**: 381–389.
Vitousek PM, Mooney HA, Lubchenco J, Melillo JM. **1997**. Human Domination of Earth's Ecosystems. *Science* **277**: 494–499.
Wang W, Malcolm BA. **1999**. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. *BioTechniques* **26**: 680–682.
Waters BM, Sankaran RP. **2011**. Moving micronutrients from the soil to the seeds: Genes and physiological processes from a biofortification perspective. *Plant Science* **180**: 562–574.
Weeks JM. **1995**. The value of biomarkers for ecological risk assessment: academic toys or legislative tools? Applied Soil Ecology 2, 215e216.
Weir BS, Cockerham CC. **1984**. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**: 1358–1370.
Wilcox BA, Murphy DD. **1985**. Conservation Strategy: The Effects of Fragmentation on Extinction. *The American Naturalist* **125**: 879–887.
Wilkins DA. **1978**. The Measurement of Tolerance to Edaphic Factors by Means of Root Growth. *New Phytologist* **80**: 623–633.

Willems G, Dräger DB, Courbot M, Godé C, Verbruggen N, Saumitou-Laprade P. **2007**. The Genetic Basis of Zinc Tolerance in the Metallophyte *Arabidopsis halleri* ssp. halleri (Brassicaceae): An Analysis of Quantitative Trait Loci. *Genetics* **176**: 659–674.
Willems G, Frérot H, Gennen J, Salis P, Saumitou-Laprade P, Verbruggen N. **2010**. Quantitative trait loci analysis of mineral element concentrations in an *Arabidopsis halleri* × *Arabidopsis lyrata petraea* F2 progeny grown on cadmium-contaminated soil. *New Phytologist* **187**: 368–379.
Wright S. **1930**. *EVOLUTION IN MENDELIAN POPULATIONS*.
Yeaman S, Otto SP. **2011**. Establishment and Maintenance of Adaptive Genetic Divergence Under Migration, Selection, and Drift. *Evolution* **65**: 2123–2129.
Young AG, Brown AHD, Zich FA. **1999**. Genetic Structure of Fragmented Populations of the Endangered Daisy Rutidosis leptorrhynchoides. *Conservation Biology* **13**: 256–265.
Zaal BJ van der, Neuteboom LW, Pinas JE, Chardonnens AN, Schat H, Verkleij JAC, Hooykaas PJJ. **1999**. Overexpression of a Novel Arabidopsis Gene Related to Putative Zinc-Transporter Genes from Animals Can Lead to Enhanced Zinc Resistance and Accumulation. *Plant Physiology* **119**: 1047–1056.
Zaborski ER. **2003**. Allyl isothiocyanate: an alternative chemical expellant for sampling earthworms. *Applied Soil Ecology* **22**: 87–95.
Zhang D-X, Hewitt GM. **2003**. Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**: 563–584.
Zhao FJ, Lombi E, Breedon T, M SP. **2000**. Zinc hyperaccumulation and cellular distribution in *Arabidopsis halleri*. *Plant, Cell & Environment* **23**: 507–514.

Abstract

Natural habitats are more and more destructed and fragmented by urban expansion and human activities. The fragmentation of natural and agricultural areas by buildings and new infrastructures affects the size, connectivity and the quality of habitats. The populations of organisms inhabiting these anthropized territories are then more isolated. However, differentiation between populations of the same organism depends on demographic and genetic processes such as genetic drift, gene flow, mutation and natural selection. Only species that have developed special tolerance mechanisms can persist under changed environmental conditions. The introduction of contaminants such as metals in the environment may influence plants and animals evolution by modifying the evolutionary forces and thus generating differences between populations. In this work, attention was focused on the genetic consequences of metallic pollution on two species, the earthworm *Lumbricus terrestris* and the plant model *Arabidopsis halleri*. Two different approaches have been used to study the genetic response to metallic contamination: a population genetic approach was performed in *L. terrestris* and a quantitative genetic approach was carried on in *A. halleri.* First, it was a question of identifying and validating new microsatellite markers in *L. terrestris*. These markers were then used to characterize the neutral genetic diversity in worms collected from agricultural and urban sites. Secondly, genetic architecture of Zn tolerance and Zn hyperaccumulation was conducted investigated for the first time using an intraspecific crossing between metallicolous and non-metallicolous individuals of *A. halleri*. High density of SNP markers was used to proceed to the QTL mapping step.

Résumé

Les habitats naturels sont de plus en plus détruits et fragmentés par l'expansion urbaine et les activités humaines. La fragmentation des espaces naturels et agricoles par les bâtiments et les nouvelles infrastructures affecte la taille, la connectivité et la qualité des habitats. Les populations d'organismes vivants sur ces territoires anthropisés sont alors plus isolées. Or, la différenciation entre populations d'un même organisme dépend de processus démographiques et génétiques tels que la dérive génétique, le flux génétique, la mutation et la sélection naturelle. La persistance et le développement des populations dans des conditions environnementales modifiées dépendent de mécanismes de tolérance. Dans ce contexte, l'introduction de contaminants tels que des métaux dans l'environnement peut influencer l'évolution des plantes et des animaux en modifiant les forces évolutives et en créant des différences entre populations.

Dans ce travail, l'attention a été portée sur les conséquences génétiques de la pollution métallique sur deux espèces, le ver de terre *Lumbricus terrestris* et une plante modèle *Arabidopsis halleri*. Deux approches différentes ont été utilisées pour étudier la réponse génétique à la contamination métallique : une approche de génétique des populations chez *L. terrestris* et une approche de génétique quantitative chez *A. halleri*.

Tout d'abord, il s'est agi d'identifier et de valider de nouveaux marqueurs microsatellites chez *L. terrestris*. Ensuite, ces marqueurs ont été utilisés afin de caractériser la diversité génétique neutre chez des vers collectés sur des sites agricoles et urbanisés. Parallèlement, l'architecture génétique de la tolérance et de l'hyperaccumulation de Zn chez *A. halleri* a été explorée à l'aide d'un croisement intraspécifique entre une population métallicole et une population non métallicole. Une densité élevée de marqueurs SNP a été utilisée pour procéder à l'étape de cartographie QTL.