Effect of metal pollution on genetic variation in natural populations of selected soil invertebrate species with different dispersal potential

Giska, I.

2016

document version
Publisher's PDF, also known as Version of record

Effect of metal pollution on genetic variation in natural populations of selected soil invertebrate species with different dispersal potential

Iwona Giska

Cover: *Magdalena Kroczak*

Photos and layout: *Iwona Giska*

Printing: AT Wydawnictwo,
ul. G. Zapolskiej 38/405, 30-126 Kraków

ISBN: 978-83-63910-58-7

Thesis 2015-12 of the Department of Ecological Science,
VU University Amsterdam, the Netherlands

To my Sister,

who would never touch a centipede

General introduction

For over a century, as the result of human population explosion and, thus, increased human activities, the environment has been changing at unprecedented rate. Most of these changes are highly negative, including environmental pollution, increasing atmospheric CO_2 concentration, habitat fragmentation, and dispersion of invasive species. Thus, natural populations are exposed to novel perturbations and face a constant need to adapt to quickly changing conditions (Bell and Collins 2008). Populations, which experience severe stress due to the high rate of environmental change, may be rescued from extinction through rapid evolution and adaptation. What allows adaptation to a changing environment, is the genetic diversity, which maintains the ability of populations to evolve.

The importance of biodiversity, the danger of its loss and the need for its conservation was recognized long ago and emphasized in the Rio Convention on Biological Diversity, 1992. It was reported in the convention that certain human activities had already led to significant reduction of biodiversity, thus, its conservation should be a common concern of humankind. Three aspects of biological diversity were mentioned in the Convention: diversity of ecosystems, diversity of species and diversity of genes. The latter results from genetic diversity of particular populations within species and the diversity among populations, and is the aspect under study in my PhD thesis.

Genetic diversity of a population is the variety of alleles and genotypes present in this population. It is generated by mutations and migrations. Normally, genetic differences between individuals in large populations are sufficiently large to enable evolution (Frankham et al. 2010). However, at certain conditions, including changing environment, genetic diversity of a population can be reduced leading to higher extinction probability. This applies particularly to isolated populations of small size (Fig. 1).

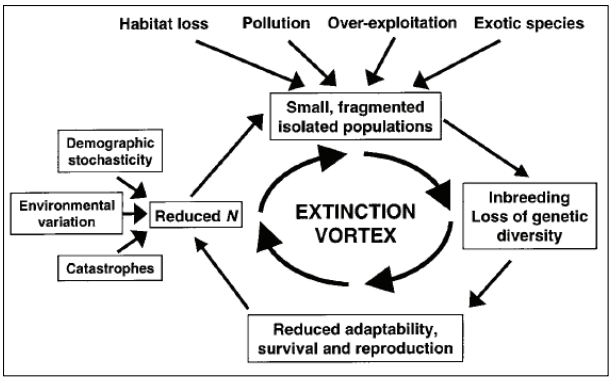


Fig. 1. Factors affecting population genetic diversity and their role in shaping the risk of extinction (from Frankham et al. 2010).

Scientists have been wondering already for decades whether metal pollution constitutes a threat for genetic diversity of natural populations (e.g., Gray 1979; Gillespie and Guttman 1989; Shugart and Theodorakis 1994). Today it is known that pollution may influence genetic diversity by the following processes: (1) the increase of mutation rate, (2) the impact on migration between contaminated and uncontaminated sites, (3) genome-wide changes (genetic drift), and (4) adaptation to the polluted environment by selection of tolerant genotypes (Bickham 2011; Van Straalen and Timmermans 2002). These processes may be identified and measured with the use of neutral genetic markers or markers under selection.

Genetic markers under selection are used to study adaptation to metal pollution. This can be done by analyzing candidate loci, e.g., genes involved in regulating the response to pollution-induced stress. Candidate genes might be found through high-density genome scans and identification of loci highly differentiated between populations from polluted and clean sites. This approach was used by e.g. by Turner et al. (2010), who found genes (F_{ST} outliers) responsible for local adaptation of *Arabidopsis lyrata* to serpentine soils characterized by high metal content. Adaptation to metal pollution might be also identified by screening for signatures of 'selective sweeps'. If a new mutation is under positive selection and increases in frequency, the neighboring region shows depleted neutral

variability due to 'selective sweep'. If local recombination rate is low and selection is strong, regions of reduced variability are long and easier to detect.

With the use of neutral genetic markers, that are not affected by selection, we can study population demographic processes and measure changes in population size. When a population is exposed to pollution, adverse effects of pollution stress on individual organisms may cause a decline of population size. If the reduction in size is deep, it might result in a decrease of the genome-wide genetic variation of populations from contaminated sites. However, scientists performing studies aiming at assessing the impact of metal pollution on the level of neutral genetic diversity of natural populations came with contradictory conclusions. They observed all possible scenarios: the decrease, the increase, and no change of population genetic diversity due to the exposure to high metal concentrations.

Decreased genetic diversity of populations inhabiting polluted environments was reported, for instance, in populations of the marsh frog (*Rana ridibunda*) from wetlands of Sumgayit, Azerbaijan (Matson et al. 2006) or the sandhopper (*Talitrus saltator*) from metal-polluted beaches of Tyrrhenian coast, Italy. The genetic erosion hypothesis of Van Straalen and Timmermans (2002) was supported by the studies of Andre et al. (2010), who suggested that the level of population genetic diversity of the earthworm *Lumbricus rubellus* at sites highly polluted with Pb was reduced due to the loss of distinct genetic lineages. On the contrary, Eeva et al. (2006) reported increased mitochondrial nucleotide diversity in populations of the great tit (*Parus major*) from sites with metal and nuclear contamination in Finland and Russia and suggested that it was due to an increase in mutation rate. The increased mutation rate due to contamination was suggested also by Rinner et al. (2011), who studied the mosquitofish (*Gambusia holbrookii*) from Sumgayit, Azerbaijan, where Matson et al. (2006) studied the abovementioned marsh frog. Although in populations of the marsh frog reduction of genetic diversity was reported, in case of both species heteroplasmy was observed, indicating an increased mutation rate at highly polluted sites. Štambuk et al. (2013), who studied populations of the mussel (*Mytilus galloprovincialis*) in the eastern Adriatic, considered enhanced mutation load as a possible explanation of higher microsatellite genetic diversity. However, finally they concluded that the increase of population genetic diversity resulted rather from high gene inflow and higher fitness of

more heterozygous populations. Intensive gene flow was suggested as a reason of no changes of genetic diversity in populations of the wood mouse (*Apodemus sylvaticus*) inhabiting a gradient of metal pollution in Belgium (Berckmoes et al. 2005).

The brief overview presented above shows that results of the studies aiming at evaluating effects of metal pollution on population genetic diversity are inconclusive. Thus, the question to what extent metal pollution may alter genetic diversity of natural populations remains open. It is also worth noting that the mentioned studies were of a rather small scale, in terms of both the number of investigated species and the molecular markers applied. So far there are no larger scale studies aiming at comparing several species sampled from the same area. Such studies could reveal whether the effect of pollution is species specific. In addition, studies involving species with different dispersal abilities could show the role of gene flow in maintaining genetic diversity of populations from contaminated sites.

When considering exposure of animals to metals in soil, we need to be aware that the exposure cannot be fully described by total soil concentrations. What determines toxicity of pollutants is their bioavailability because organisms respond only to that fraction of the metal that enters the body. This fraction must be high enough to cause adverse effects in individual organisms and consequently impact the fitness of natural populations. Thus, when assessing pollution effects on any population parameter, including genetic diversity, we should first check if the pollutants are available for uptake from the environment. This can be done with chemical methods, e.g., extraction with $CaCl_2$. However, more reliable are biological methods, such as toxicokinetics studies (Peijnenburg et al. 1999), which are based on measuring the amount of a pollutant that actually enters the body. The toxicokinetics approach allows for predicting the physiological fate of metals as it provides information about bioaccumulation and time to reach steady-state tissue concentration.

What in recent years appeared to be of a great importance in studies on soil invertebrates, is the cryptic speciation exhibited by many species (Van der Wurff et al. 2003; Spelda et al. 2011; Donnelly et al. 2013). The presence of cryptic species within a study area may hamper the comparison of genetic diversity between populations if they represent separate cryptic species. An interesting issue to test in such case is whether pollution itself could lead to cryptic speciation or cause cryptic species

to disappear from a highly polluted areas due to species-specific tolerance. Such phenomenon was suggested by Andre et al. (2010) in case of the genetic lineages of the earthworm *L. rubellus* inhabiting sites characterized by different levels of Pb pollution.

Hypotheses

Considering the existing knowledge and the above-mentioned questions, the following hypotheses were tested in the thesis:

1) Long term metal pollution impacts the level of genetic diversity of natural populations of soil invertebrates.
2) The scale of the impact of pollution on population genetic diversity depends on the species dispersal abilities.
3) In case of low-migratory species showing cryptic speciation pollution controls the distribution of the genetic lineages.

Animals

Three soil and litter dwelling invertebrate species were selected for the study: the earthworm *Lumbricus rubellus* (Fig. 2), the centipede *Lithobius forficatus* (Fig. 3), and the rove beetle *Staphylinus erythropterus* (Fig. 4). These species were selected to represent different dispersal abilities. The earthworm *L. rubellus* represents species with low dispersal abilities. The migration rate of earthworms in the field was shown to be < 20 m/year (Marinissen and Van den Bosch 1992). The centipede *L. forficatus* is considered a species with medium dispersal abilities, while the rove beetle, which is a flying insect, represents species with high dispersal abilities.

Fig. 2. The earthworm *Lumbricus rubellus.*

Fig. 3. The centipede *Lithobius forficatus.*

Fig. 4. The rove beetle *Staphylinus erythropterus*.

In soil ecotoxicology, earthworms are considered to be sentinel species. In the field, the shape of earthworm populations reflects soil conditions. They are commonly used in standard ecotoxicological laboratory tests. Thus, we considered it appropriate to choose *L. rubellus* for our study, especially because it is an earthworm species widespread in Europe. Earthworms are known to accumulate heavy metals. They are sensitive to high metal concentrations, but at the same time they are able to develop metal tolerance and survive even at highly contaminated sites (Spurgeon and Hopkin 1999a; Spurgeon et al. 2006; Fisker et al. 2011). Earthworms constitute an important source of

food for animals of higher trophic levels, e.g., predatory insects, centipedes, lizards, mammals, and birds. The other two species selected for the study represent invertebrates that feed, among others, on earthworms and show much higher mobility than the earthworms. The centipedes and rove beetles are both predators with similar diet composition but differ substantially in dispersion rates. They both lay eggs into soil and have chitin-less epicuticle larvae. Therefore, we expected these two species to have similar routes of exposure to metals in soil.

Study sites

To test the research hypotheses we sampled the selected invertebrate species along a metal pollution gradient in the vicinity of the lead and zinc smelter 'Bolesław', close to Olkusz town in Southern Poland. The sampling area is rich in metal ores and has a long history of mining and metal smelting. The mining and smelting has started probably in the Middle Ages, but till the 20th century it was rather small scale activity. However, in 1969 a big smelter "Bolesław" was constructed. In 1970s, the emission of dust from the smelter was over 1000 t per year and caused contamination of soil over an extensive area of at least 500 km^2. The contamination persists till today and high concentrations of metals, mainly zinc, lead and cadmium, are found in the topsoil. The sites selected for the study are characterized by similar habitat type – mixed pine forest on acidic soils with a well-developed mor organic layer (Table 1). Besides the sites in the Olkusz area (OL1 – OL7), animals were collected also at two unpolluted sites: Trzemeśnia (TR) in Southern Poland and Kozienicka Forest (PK; after Polish Puszcza Kozienicka) in Central Poland (Fig. 5).

Table 1. Characteristics of the soil at the study sites. The distance from the smelter, soil pH, organic matter content (OM%), and metal concentrations [mg kg^{-1} dwt]: total concentrations (normal font) and 0.01 M $CaCl_2$-extractable concentrations (italics) are shown; mean ± SD (n = 3).

Site	Distance[km]	pH_{CaCl2}	OM [%]	Cd [mg kg^{-1}]	Pb [mg kg^{-1}]	Zn [mg kg^{-1}]
OL1	3.3	5.06 ± 0.06	45.1 ± 1.3	63.2 ± 3.0	3 041 ± 158	7 991 ± 536
				0.892 ± 0.018	*0.553 ± 0.014*	*54.4 ± 1.1*
OL2	2.5	4.12 ± 0.03	53.5 ± 0.4	49.1 ± 1.1	2 060 ± 37	3 960 ± 54
				3.69 ± 0.05	*1.98 ± 0.05*	*211 ± 2*
OL4	5.3	3.46 ± 0.02	54.2 ± 2.0	14.8 ± 0.2	847 ± 38	966 ± 22
				1.98 ± 0.02	*1.88 ± 0.02*	*84.3 ± 1.5*
OL5	7.7	4.29 ± 0.01	36.3 ± 0.7	12.1 ± 0.7	708 ± 12	756 ± 11
				0.688 ± 0.012	*0.526 ± 0.015*	*30.8 ± 0.1*
OL7	~32	4.25 ± 0.01	6.68 ± 0.08	1.36 ± 0.10	60.9 ± 3.5	88.6 ± 13.6
				0.393 ± 0.006	*0.231 ± 0.007*	*16.8 ± 0.7*
PK	~180	2.82 ± 0.01	35.4 ± 1.3	0.677 ± 0.102	56.5 ± 3.9	36.5 ± 3.8
				0.198 ± 0.069	*0.376 ± 0.032*	*6.98 ± 1.39*
TR	~65	5.33 ± 0.04	13.0 ± 0.1	1.77 ± 0.295	65.4 ± 1.10	170 ± 17
				0.073 ± 0.008	*0.005 ± 0.003*	*1.0 ± 0.3*

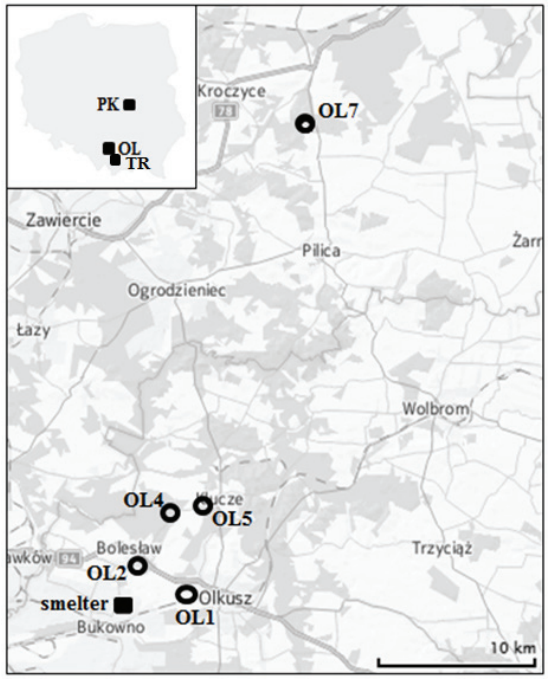


Fig. 5. Location of the study areas in Poland (insert): Kozienicka Forest (PK) in Central Poland, Olkusz area (OL) and Trzemeśnia (TR) in Southern Poland, and detailed location of the Olkusz sites (OL1-OL7) in the vicinity of the smelter 'Bolesław'.

Molecular methods

The selected invertebrates are not model species, with little genetic information available. In general, soil invertebrates naturally occurring in Poland do not include species with sequenced genomes. Thus, a choice of molecular methods must be adapted to limited genetic information that could be potentially used to develop molecular markers for population studies. A marker that is universal and can be applied in numerous species without costly developing of new primers, is the mtDNA, especially *cytochrome* c *oxidase* subunit 1 (COI), commonly used in barcoding (Hebert et al. 2003). Therefore, to estimate population genetic diversity we used the two mitochondrial genes COI and ATP6 (ATP synthase F0 subunit 6). However, as mtDNA is haploid and inherited only from mothers, and has a different evolution than nuclear DNA, it does not provide full description of populations and their evolutionary histories.

Recent developments of next-generation sequencing (NGS) technologies resulted in molecular approaches based on nuclear genome that are applicable also to non-model species (Baird et al. 2008; Peterson et al. 2012; Zieliński et al. 2014). These approaches include, i.a., Restriction site Associated DNA Sequencing (RADseq), already existing in several variations, e.g. ddRAD (Peterson et al. 2012), 2bRAD (Wang et al. 2012), ezRAD (Toonen et al. 2013), each having some advantages over the other. The general idea of RADseq is to sequence a reduced representation of a whole genome, obtained by digestion of genomic DNA with restriction enzymes. In double digest RADseq (ddRAD), which we have chosen as an approach to estimate genome-wide genetic diversity, genomic DNA is digested with two restriction enzymes. By the choice of specific enzymes the number of RAD tags in the final library can be controlled. Then, adapters with barcodes are ligated to obtained fragments of genomic DNA; unique barcodes are used to distinguish individuals. After ligation, size selection of fragments of specific length is performed to further control the number of RAD tags. This enables reduction of marker density and sequencing costs. The library is then amplified with PCR primers that can include indices. The in-line combination of indices and barcodes allows relatively cheap multiplexing of large number of individuals. At the end, such prepared RADseq library is sequenced on Illumina platform (Fig. 6). Depending on the number of RAD tags in the library and required coverage, a HiSeq or MiSeq platform can be chosen.

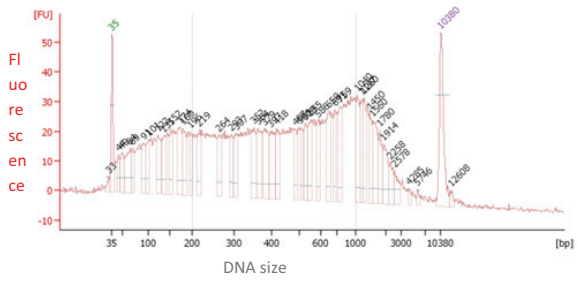
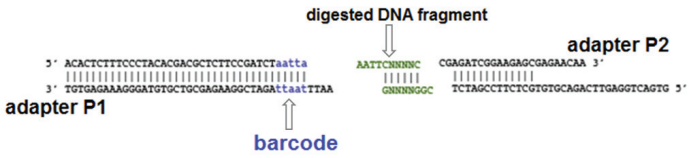
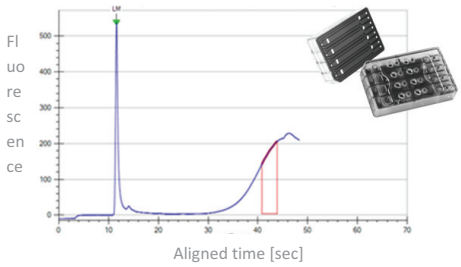
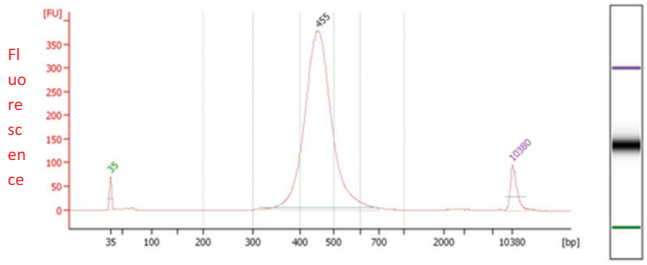
	digestion of genomic DNA with two restriction enzymes (picture shows size distribution of DNA fragments after digestion, analyzed with Bioanalyzer)
	ligation of adapters with barcodes
	size selection (e.g. with LabChip XT)
	PCR amplification (picture shows the final amplified and cleaned RADseq library analyzed with Bioanalyzer)
	Illumina sequencing (HiSeq 2000)

Fig. 6. RADseq workflow; from genomic DNA digestion to Illumina sequencing.

Not long ago, when deciding about sequencing strategy, one needed to keep in mind that RADseq libraries were low diversity libraries. This was a challenge for Illumina platforms because the diversity of the first 11 cycles was crucial for a proper work of Illumina algorithms. This diversity restriction applied to the RADseq library because there are barcodes (usually low number in a pool of samples) and restriction enzyme overhangs (identical in all fragments) within the first 11 nucleotides of the RAD tag fragment. Such composition of the RAD tag beginning made the RADseq library problematic for Illumina sequencing. Fortunately, after improvement of the Illumina algorithms that identify clusters and estimate the color matrix and phasing, low diversity libraries can now successfully be sequenced on Illumina platforms without PhiX spiking or control lanes.

Thesis outline

The issues stressed in this introduction are addressed in the four chapters of the thesis.

In **Chapter 1** the bioavailability of four metals (Cd, Pb, Cu, Zn) to the earthworm *L. rubellus* exposed to soils originating from the Olkusz gradient of metal pollution was estimated. The uptake and elimination kinetics of the metals were determined and the estimated values were related to such soil properties as organic matter content, pH and cation exchange capacity. After the demonstration of the availability of metals to soil organisms, in the following chapters effects of metal contamination on population genetic diversity and genetic structure were assessed. In **Chapter 2** the impact of soil pollution on genetic structure and genetic diversity of the earthworm *L. rubellus* was analyzed. It was also determined whether the divergent mitochondrial lineages of *L. rubellus* that occurred in sympatry at the studied sites in Southern Poland were reproductively isolated and represented cryptic species. In **Chapter 3** the effect of metal pollution on genetic diversity, population structure and cryptic speciation of the centipede *L. forficatus* was evaluated, while in **Chapter 4** populations of the rove beetle *S. erythropterus* in terms of their genetic diversity were analyzed. Then, the main results of the four chapters are summed up and their relation to dispersal abilities of the selected invertebrate species is discussed.