REGULAR PAPER

Gene polymorphisms for elucidating the genetic structure of the heavy-metal hyperaccumulating trait in *Thlaspi caerulescens* and their cross-genera amplification in Brassicaceae

N. Basic · G. Besnard

Abstract Genetic polymorphism was investigated in *Thlaspi caerulescens* J. & C. Presl at 15 gene regions, of which seven have been identified to putatively play a role in heavy-metal tolerance or hyperaccumulation. Single nucleotide and length polymorphisms were assessed at four cleaved amplified polymorphic sequences (CAPS) and 11 simple sequence repeat (microsatellite) loci, respectively. The utility of these loci for genetic studies in *T. caerulescens* was measured among seven natural populations (135 individuals). Fourteen loci rendered polymorphism, and the number of alleles per locus varied from 2 to 5 and 1 to 27 for CAPS and microsatellites, respectively. Up to 12 alleles per locus were detected in a population. The global observed heterozygosity per population varied between 0.01 and 0.31. Additionally, cross-species/genera amplification of loci was investigated on eight other Brassicaceae (five individuals per population). Overall, 70% of the cross-species/genera amplifications were successful, and among them, more than 40% provided intraspecific polymorphisms within a single population. This indicates that such markers may, as well, allow comparative population genetic or mapping studies between and within several Brassicaceae, particularly for genes involved in traits such as heavy-metal tolerance and/or hyperaccumulation.

N. Basic · G. Besnard (✉)
Department of Ecology and Evolution, Biophore,
University of Lausanne, 1015 Lausanne, Switzerland
e-mail: gbesnard@unil.ch

Introduction

The development of studies on phytoremediation greatly increased the interest for heavy-metal-tolerant and hyperaccumulating plants because of their potential use in the cleaning of soils contaminated with toxic metals (Pollard et al. 2002). Several plant models have been described, particularly in Brassicaceae (Peer et al. 2003). *Thlaspi caerulescens* J. & C. Presl was recently considered as the most promising model species due to its physiological, morphological and genetic characteristics and its genetic proximity to *Arabidopsis thaliana* (L.) Heynh., a reference species in plant genetics (Assunçao et al. 2003a; Peer et al. 2003). The physiological and molecular bases of heavy-metal hyperaccumulation are not yet completely understood (Pollard et al. 2002; Bert et al. 2003; Lugon-Moulin et al. 2004), but several genes putatively involved in this character have been identified (e.g. Pence et al. 2000; Lombi et al. 2002; Bovet et al. 2003; Bernard et al. 2004; Roosens et al. 2005). For instance, genes or mRNAs encoding transporters for various metals such as Zn, Fe and putatively Cd have been isolated (e.g. ZNT and IRT gene families; Mäser et al. 2001; Lombi et al. 2002). Transcription of such genes can be highly influenced by heavy-metal presence in nutriments (Bovet et al. 2003; Bernard et al. 2004) and can vary considerably between metallicolous and nonmetallicolous ecotypes (Lombi et al. 2002).

In Europe, *T. caerulescens* is naturally distributed on soils with variable heavy-metal composition,

particularly for Zn and Cd. Tolerance and hyperaccu-mulation for these metals was found to be highly variable between populations (Roosens et al. 2003). Isozyme markers have been developed to analyse the systematic, the population genetic structure and the mating system of *T. caerulescens* (Koch et al. 1998; Dubois et al. 2003). These reports, as well as a phylogeographic reconstruction, may bring new in-sights into the understanding of the evolution of heavy-metal tolerance or hyperaccumulation in *T. caerulescens*. However, only few markers are cur-rently available. Thus, increasing the number of variable loci is needed for further genomic investi-gations. Development of such markers defined in genes may be also very useful because they can be genera-transferable and permit the comparison be-tween *T. caerulescens* and related species (e.g. Kuitti-nen et al. 2002). In addition, markers linked to coding regions, especially to those involved in heavy-metal tolerance and hyperaccumulation, should be of particular interest. Those genetic markers could be used to characterise natural populations and analyse how the genetic diversity structure is related to soil characteristics and plant hyperaccumulation efficien-cies. This, in turn, would allow the identification of genes under adaptive selection. Indeed, the higher genetic differentiation revealed at markers defined in genes potentially involved in heavy-metal tolerance and/or hyperaccumulation can be interpreted as sig-natures of natural selection (e.g. Beaumont and Nichols 1996). The population structure analysis combining such markers and markers unrelated to metal homeostasis could be used to detect which markers are under selective pressures related to the heavy-metal content of soil or the amount of metal accumulated in plant. Furthermore, polymorphisms in genes putatively involved in heavy-metal tolerance or hyperaccumulation may be investigated on prog-enies of controlled crosses. Detection of possible cosegregation of the hyperaccumulation trait with known quantitative trait loci (QTLs) would then be possible to test (Assunçao et al. 2003b, 2006; Frérot et al. 2003; Zha et al. 2004).

In this context, we developed a set of primers amplifying microsatellite and cleaved amplified poly-morphic sequence (CAPS) loci located in several coding genes of *T. caerulescens*. We focused particu-larly on genes putatively involved in the heavy-metal responses. These markers were then used to characte-rise seven populations of *T. caerulescens*. Furthermore, the transferability of each gene marker was tested on eight other Brassicaceae species, of which several are related to heavy-metal-tolerant plants.

Materials and methods

Plant material and DNA extraction

Individuals of *T. caerulescens* ($2n = 2x = 14$) were prospected on five natural sites in Switzerland: three populations of Jura [coded J1 (West Jura), J8 and J12 (Central Jura)] and two in the Alps (coded A2 and A5). Soil and plant heavy-metal composition have been previously characterised at each site, and significant differences were observed in soil HNO_3-extractable and shoot Cd and Zn concentrations (Basic et al. 2006a, b). Between 18 and 21 individuals were sampled in each population. In addition, two metallicolous populations from Ganges (G; France) and Prayon (P; Belgium) differing in their capacity to accumulate Cd (Lombi et al. 2000; Roosens et al. 2003) were sampled with 20 and 17 individuals per population, respectively. For eight other Brassicaceae species, five individuals were sampled in natural fields in Switzerland: *Thlaspi arvense* L. ($2n = 2x = 14$), *Arabidopsis thaliana* (L.) Heynh. ($2n = 2x = 10$), *Brassica napus* L. ($2n = 4x = 20$), *Sinapis arvensis* L. ($2n = 2x = 18$), *Biscutella laevigata* L. ($2n = 2x = 18$), *Capsella bursa-pastoris* (L.) Medik. ($2n = 4x = 32$), *Arabis hirsuta* (L.) Scop. ($2n = 4x = 32$) and *Erophila praecox* (L.) Chev. ($2n = 4x = 36$). These species were chosen to be representative of a large taxonomical range in the Brassicaceae family (Yang et al. 1999; Koch et al. 2001). We extracted DNA of each individual from 100 mg of leaf using the FastDNA kit (Qbiogene, Inc., Carlsbad, CA, USA).

Genetic marker development

Investigated genes and primer design

We selected accessions of *Thlaspi* available in public DNA databanks (in June 2004). We looked for mi-crosatellite motifs using the findpatterns software (GCG package, http://www.accelrys.com/products/gcg_wisconsin_package/), and when several accessions were available for a gene, substitution variation in the coding sequence (cds) or noncoding sequence [i.e. untrancribed sequence of cDNA (UTR) and intron]. Based on these data, we looked for polymorphism in ten genes: ZNT1, ZNT2, ZNT5, IRT1, IRT2, E2F1, WRKY, AGAMOUS, CP and up1 (Table 1). Primers (Tables 2, 3) were defined for flanking regions of each region displaying a microsatellite motif (ZNT5, IRT1, WRKY, AGAMOUS, CP, up1) or a substitution leading to restriction-site polymorphism (ZNT1, ZNT2, IRT2, E2F1). When possible, primers were designed in conserved regions (particularly in exons)

Table 1 General features of the 15 genomic DNA regions (genes) characterised for sequence polymorphism in the present study

Gene code	Putative homologous gene in *Arabidopsis thaliana* and encoded protein	Genomic location in *Arabidopsis thaliana*
Tc-ZNT1	At1g10970 – Zn and Cd transporter	Chr 1 – U95973
Tc-ZNT2	At1g60960 – Putative Zn transporter	Chr 1 – AC018908
Tc-ZNT5	At1g05300 – Putative Zn transporter	Chr 1 – AC000098
Tc-E2F1	At5g22220 – E2F transcription factor-1	Chr 5 – AL589883
Tc-IRT1	At4g19690 – Putative Fe(II) transporter-1	Chr 4 – AL024486
Tc-IRT2	At4g19680 – Putative Fe(II) transporter-2	Chr 4 – AL024486
Tc-WRKY	At4g31550 – WRKY transcription factor	Chr 4 – AL080283
Tc-AGAMOUS	At4g18960 – Floral homeotic protein AGAMOUS	Chr 4 – AL021711
Tc-CP	At1g30630 – Putative coatomer protein	Chr 1 – AC007060
Tc-up1	At2g47440 – Unknown protein	Chr 2 – AC002535
Tc-up2 (Ap5)[a]	At3g01860 – Unknown protein	Chr 3 – AC010797
Tc-up3 (Ap6)	At1g16500 – Unknown protein	Chr 1 – AC006341
Tc-NOD (Ap7)	At4g30420 – Nodulin-like protein	Chr 4 – AF160182
Tc-up4 (Ap8)	At3g25410 – Unknown protein	Chr 3 – AB025639
Tc-bHLH (Na10-G10)	At5g04150 – bHLH transcription factor	Chr 5 – AL391716

Chr chromosome

[a]Marker code of the homologous locus in *Alliaria* (Durka et al. 2004) and *Brassica* (Lowe et al. 2002) is given in parentheses

Table 2 Characteristics of cleaved amplified polymorphic sequences (CAPS) loci: Genbank accession numbers, primer pairs used for polymerase chain reaction (PCR) amplification, specific PCR conditions [annealing temperature (T_a) and MgCl$_2$ concentration], location of the restriction site (RS) polymorphism, number (*N*) and size of alleles revealed

Locus name	Genbank accessions no.	Primer (5′ → 3′)	T_a (°C)	MgCl$_2$ (mM)	RS polymorphism location in the gene	N	Allele size (bp)
Tc-ZNT1	AF133267, AF275751, AJ313521, AJ746204, AM162547	*f.* TTCGTGCTCATGCAGCTCAC *r. Hex-*GACACAATCCCAAGCTCC	60	1.5	Intron – *Pag*I[a]	5	251 → 251 252 → 252 252 → 142 + 110 253 → 253 255 → 145 + 110
Tc-ZNT2	AF275752, AF292370, AJ538346	*f.* TTACCGGAGTTTCCTTGGAAG *r.* CAGAATGAGTAGTAGCTTCCC	57	3	CDS – *Hpa*II	2	141 → 141 141 → 105 + 36
Tc-E2F1	AJ746205, AJ746206	*f.* CCAGCCGCGGATCTGCCT**T**C[b] *r.* GGATGATACTGCCGCTTCGAAG	53	1.5	5′ UTR – *Taq*I	2	205 → 187 + 18 205 → 167 + 20 + 18
Tc-IRT2	AJ746209, AJ746210	*f.* CATGGTGTTGTGCTAGCAAC *r. Fam-*GAGATAGTCCAATGACCACAG	53	1.5	Intron – *Mse*I	3	231 → 94 + 5 + 19 + 113 231 → 99 + 19 + 113 231 → 118 + 113

[a]The restriction enzyme used to reveal polymorphism (nucleotide substitution) is indicated

[b]Bold nucleotide indicates a nucleotide change comparatively to the reference Accession (C → T) to create an absence/presence polymorphism of a *Taq*I restriction site

between *Thlaspi* and *Arabidopsis* to allow for a better cross-genera amplification of markers in Brassicaceae (Kuittinen et al. 2002). Seven of these genes (i.e. IRT1, IRT2, E2F1, WRKY, ZNT1, ZNT2 and ZNT5) were considered to be putatively implicated in the heavy-metal hyperaccumulation and tolerance responses (ZNT1, ZNT2 and ZNT5: Pence et al. 2000; IRT1 and IRT2: Lombi et al. 2002; E2F1: N.S. Pence and L.V. Kochian, unpublished results; WRKY: Susuki et al. 2001).

Table 3 Characteristics of microsatellite loci: GenBank accession numbers, primer pairs used for polymerase chain reaction (PCR) amplification, specific PCR conditions (annealing temperature (T_a) and $MgCl_2$ concentration), location of the microsatellite polymorphism, number (*N*) and size range of alleles revealed

Locus name	GenBank accessions no.	Primer (5′ → 3′)	T_a (°C)	$MgCl_2$ (mM)	Microsatellite location in the gene and repeated motif	*N*	Allele size (bp)
Tc-ZNT5	AF292029	f. AATCACACAAAACGTTAAGCTC r. Hex-AAGGTATGGCGGCGATCTTG	53	1.5	5′ UTR – $(CTT)_4$ $CAT(CT)_5$	1	157
Tc-IRT1	AJ746208	f. CTTGCGATATCGAGTCATTGC r. Fam-TCCAATGACCACAGAGTGAAC	53	1.5	Intron – $(AT)_{11}$	10	173–195
Tc-WRKY	AJ746211	f. TTCTCCGGAAAAGTCTCCGG r. Fam-CTCACGGTTCTCTTCATCCG	50	2.5	Intron – $(T)_8$ - $(A)_8$	5	438–441, 672
Tc-AGAMOUS	AY253254, AY253266	f. Hex-CCTCCATTGTTGTTAATGTCTG r. TACTCTCACTTACCATCACATG	53	1.5	Intron – $(TC)_5TT(TC)_4$	2	142–144
Tc-CP	AJ746244	f. TTTGGAGTTAGACACGGATCTG r. Hex- GTTGATCGCAGCTTGATAAGC	53	1.5	5′ UTR – $(GAA)_7$	5	151–164
Tc-up1	AJ746212	f. Fam-TGCTCTGTTTCTCTCCACATTC r. TTCCTTGCTTCTTCTCTTCCA	53	1.5	5′ UTR – $(CA)_5$ $(CT)_8CA(CT)_4$	8	132–170
Tc-up2	AJ746213	f. Hex-TGAGAAGAGGAGACACAGGAAC r. CACTTACCAAATCGAAAACTGCTCC	53	1.5	Intron – $(AG)_5$ - $(AG)_6$ - $(GA)_5$	3	234–244
Tc-up3	AJ746214	f. Hex-GAGGAGATCGCGAGTCATGAG r. CTGCCTAACGTACCGCATAACTG	53	5	5′ UTR – $(CT)_{10}$	22	172–288
Tc-NOD	AJ746215	f. AAGTACGTGTACGCCAACCG r. Fam-TGTACTCCTCTAACTTCCCC	53	5.5	5′ UTR – $(TC)_{11}$	27	216–312
Tc-up4	AJ746216	f. Fam-GTTTTGTCCGCTTTGCTTCC r. GCCATAGACTTTCTCATTGATTC	53	1.5	Intron – $(CT)_{13}$	8	255–266
Tc-bHLH	AJ746217	f. CTTGGAAACATTGGTGTTAAGG r. Fam-GATTCCATCTCAAATCCGGTC	53	1.5	Intron – $(TC)_6$	4	144–150

Additionally, microsatellite loci isolated from Brassicaceae genera related to *Thlaspi* [*Brassica* spp. (Lowe et al. 2002) and *Alliaria petiolata* (Durka et al. 2004)] were recently available in public databanks. Gene sequences homologous to 14 (74%) of these loci were identified in *A. thaliana* using the BlastN software (http://www. ncbi.nlm.nih.gov): Ni2B01, Ni2-C12, Ni2-E04, Na10-B10, Na10-F06, Na10-G10, O109-A03 and O109-A06 from *Brassica* (Lowe et al. 2002) and Ap1, Ap3, Ap5, Ap6, Ap7 and Ap8 from *Alliaria* (Durka et al. 2004). Degenerated microsatellite motifs, containing at least five successive repeated dinucleotide motifs, were found again in *Arabidopsis* for loci Ap1, Ap5, Ap6, Ap7 and Ap8 (data not shown) whereas none was found at the nine other loci. Dinucleotide repeats were always located in untranscribed gene regions (intron or UTR). Conserved sequences between *Alliaria* or *Brassica* and *Arabidopsis* for flanking regions (particularly in exons) of untranscribed gene regions were detected for loci Ap5, Ap6, Ap7, Ap8, Ni2-E04 and Na10-G10. Specific primers were designed in these regions (Tables 1, 3). They were then used in polymerase chain reaction (PCR) amplification (see below for the PCR protocol) and direct sequencing of homologous loci in one individual of *T. caerulescens* (sequence accessions AJ746213 to AJ746217 and

AJ746248). Dinucleotide motifs were detected in five of the six investigated loci in *T. caerulescens*: Ap5, Ap6, Ap7 Ap8, and Na10-G10. These five loci were renamed Tc-up2, Tc-up3, Tc-NOD, Tc-up4 and Tc-bHLH, respectively (Table 1).

To summarise, 14 of the 15 investigated loci are located in untranscribed spacers (e.g. intron or UTR) of genes. Only the restriction site polymorphism at the Tc-ZNT2 locus is directly located in the coding sequence but is due to a synonymous substitution that consequently may not be directly under selection. Additionally, each of the 15 studied genes displays a high sequence identity with one expressed single-copy gene in the *Arabidopsis* genome (Table 1).

PCR protocol and polymorphism detection

For the PCR amplification of each locus, reaction mixtures contained 50 ng of DNA template, 1× reaction buffer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 0.2 µM of each oligonucleotide primer (one 5′ labelled with a fluorochrome; Applied Biosystems; Tables 2, 3) and 0.75 U of DNA polymerase (Qbiogene) in a total volume of 25 µl. Reaction mixtures were incubated in a thermocycler (T1, Biometra) firstly for 4 min at 94°C

and then for 36 cycles consisting of 45 s at 94°C, 45 s at the defined annealing temperature (Tables 2, 3) and 1 min at 72°C. The last cycle was followed by a 10 min extension at 72°C. For microsatellite loci, electrophoresis of PCR products was directly carried out on a denaturing 5% polyacrylamide gel using an automated sequencer (ABI 377; Applied Biosystems). For loci Tc-E2F1, Tc-ZNT1, Tc-ZNT2 and Tc-IRT2, PCR products were digested with a specific restriction enzyme (Table 2) to reveal CAPS. Restriction fragments were electrophoresed either on a 2.8% agarose gel and visualised by ethidium bromide staining (loci Tc-E2F1 and Tc-ZNT2) or on a denaturing 5% polyacrylamide gel, as previously described (loci Tc-IRT2 and Tc-ZNT1). The size of DNA fragments was estimated against the 100-bp ladder (Gibco-BRL) or the 500 ROX size standard (Applied Biosystems) on agarose and acrylamide gels, respectively.

Data analysis

For *T. caerulescens* populations, alleles were scored for each individual. Expected (H_e) and observed (H_o) heterozygosities were calculated and compared at each locus, and the fixation index within each population (F_{is}) across all loci was obtained with the GENETIX software (version 4.03; Belkhir et al. 2004). A test of the significance of association between genotypes at pairs of loci across all populations was performed with the FSTAT software (version 2.9.3, Goudet 2001). The excess of homozygotes at each locus and for each population was analysed with the MICRO-CHECKER software (Van Oosterhout et al. 2004). When an excess of homozygotes was observed, a null allele estimator for nonequilibrium populations was used to estimate null allele frequencies (Van Oosterhout et al. 2006). As the algorithm requires preliminary and independent estimate of fixation index (F_{is}), two F_{is} were used, 0.36 and 0.64, corresponding respectively to fixation indices found on allozymes in metalliferous (from France and Belgium) and nonmetalliferous (from France and Luxembourg) *T. caerulescens* populations (Dubois et al. 2003).

Results and discussion

Polymorphic loci in *T. caerulescens*

Polymorphism was found in different populations of *T. caerulescens* at 14 of the 15 loci (Tables 2, 3, 4). Locus Tc-ZNT5 was the only one to show no variation between populations (Table 3). The number of alleles

at CAPS and microsatellite loci varied from 2 to 5 and 1 to 27, respectively. Length-mutational events were revealed at ten microsatellite loci plus the Tc-ZNT1 locus. Polymorphism in genes putatively involved in the variable capacities of Cd hyperaccumulation was revealed, particularly at loci Tc-IRT1, Tc-IRT2, Tc-ZNT1 and Tc-ZNT2 (Lasat et al. 2000; Pence et al. 2000; Lombi et al. 2002; Bert et al. 2003). Furthermore, the locus pair Tc-E2F1 and Tc-AGAMOUS as well as Tc-IRT1 and Tc-IRT2 presented a significant genotypic linkage disequilibrium ($P < 0.05$). Linkage disequilibrium could be due to a physical linkage between loci. For example, the genes IRT1 and IRT2 are tightly linked in the *Arabidopsis* genome (chromosome 4; GenBank accession AL024486), and the genotypic linkage disequilibrium between these loci may indicate that this is also the case in *T. caerulescens*. But other factors could, as well, lead to the linkage disequilibrium observed, including inbreeding, population admixture or selection (Flint-Garcia et al. 2003; Gupta et al. 2005).

Intrapopulation gene polymorphism in *T. caerulescens*

The number of alleles per population (N), expected (H_e), observed (H_o) heterozygosities and fixation indices (F_{is}) within populations are shown in Table 4. We detected up to 12 alleles for two loci in a population (Tc-up3 and Tc-NOD in Ganges; Table 4). The genetic diversity of *T. caerulescens* populations varied between 0.31 (Ganges) and 0.17 (J8) whereas A2 was nearly fixed (0.01). The H_o was, in general, lower than H_e. Populations may depart from expected Hardy–Weinberg proportions as a result of local breeding structure or admixture. The F_{is} values for the Swiss populations was comprised between 0.33 (A5, J12) and 0.44 (J8; Table 4). Interestingly, the fixation index of Prayon obtained with the variable loci of the present study ($F_{is} = 0.42$) was close to the one reported with isozyme markers ($F_{is} = 0.51$; Dubois et al. 2003). In contrast, the lowest fixation index was found for Ganges ($F_{is} = 0.15$). The failure of amplification of particular alleles (i.e. null alleles) resulting from polymorphism on PCR priming site or an absence of large-allele amplification (large allele dropout) can also explain lower H_o than expected. Among the developed loci, Tc-ZNT1, Tc-ZNT2, Tc-IRT1, Tc-IRT2, Tc-up1, Tc-up3, Tc-NOD and Tc-up4 presented a significant excess in homozygotes in some populations (Table 4). However, when considering an inbreeding coefficient of 0.64, no loci showed a significant signal of null alleles (Table 5).

Table 4 Description of the genetic diversity in *Thlaspi caerulescens* populations: the number of alleles per population (N), observed (H_o) and expected (H_e) heterozygosities per population and fixation index (F_{is}) of populations are presented

Population	Tc-ZNT1	Tc-ZNT2	Tc-E2F1	Tc-IRT1	Tc-IRT2	Tc-WRKY	Tc-AGAMOUS	Tc-CP	Tc-up1	Tc-up2	Tc-up3	Tc-NOD	Tc-up4	Tc-bHLH	Total
J1															
N	3	2	1	4	2	2	1	1	2	2	2	5	3	1	
H_e	0.62	0.48	–	0.64	0.48	0.38	–	–	0.41	0.26	0.51	0.71	0.30	–	0.34
H_o	0.60	0.11[a]	–	0.40[a]	0.15[a]	0.20	–	–	0.25	0.20	0.50	0.25[a]	0.25	–	0.21
F_{is}															0.40***
J8															
N	3	2	2	3	2	1	2	1	4	2	3	8	2	2	
H_e	0.22	0.21	0.41	0.18	0.09	–	0.34	–	0.71	0.14	0.53	0.72	0.49	0.09	0.30
H_o	0.19	0.05[a]	0.35	0.10	0.00[a]	–	0.24	–	0.48[a]	0.05	0.24[a]	0.48[a]	0.10[a]	0.10	0.17
F_{is}															0.44***
J12															
N	3	2	1	3	1	1	2	1	3	3	3	2	3	1	
H_e	0.59	0.49	–	0.38	–	–	0.32	–	0.48	0.39	0.29	0.36	0.64	–	0.28
H_o	0.28[a]	0.41	–	0.33	–	–	0.28	–	0.39	0.17	0.22	0.33	0.24[a]	–	0.19
F_{is}															0.33***
A2															
N	1	1	1	1	1	1	1	1	2	2	2	2	1	1	
H_e	–	–	–	–	–	–	–	–	0.05	0.05	0.10	0.10	–	–	0.02
H_o	–	–	–	–	–	–	–	–	0.05	0.05	0.00	0.10	–	–	0.01
F_{is}															0.33*
A5															
N	1	2	1	3	2	1	2	1	3	2	2	4	3	1	
H_e	–	0.43	–	0.53	0.44	–	0.19	–	0.32	0.27	0.34	0.62	0.57	–	0.26
H_o	–	0.47	–	0.37	0.21	–	0.11	–	0.32	0.21	0.21	0.47	0.16[a]	–	0.18
F_{is}															0.33***
P															
N	2	1	1	3	1	2	1	2	2	2	6	7	4	3	
H_e	0.49	–	–	0.39	–	0.06	–	0.50	0.43	0.34	0.77	0.64	0.67	0.59	0.35
H_o	0.29	–	–	0.35	–	0.06	–	0.35	0.24	0.18	0.50	0.29[a]	0.08[a]	0.53	0.21
F_{is}															0.42***
G															
N	2	1	1	1	1	3	1	4	5	2	12	12	5	4	
H_e	0.10	–	–	–	–	0.34	–	0.46	0.67	0.38	0.93	0.90	0.72	0.60	0.36
H_o	0.10	–	–	–	–	0.20	–	0.30	0.55	0.40	0.91	0.82	0.57	0.50	0.31
F_{is}															0.15***

[a] Significant excess of homozygotes ($P < 0.05$)

F_{is} values: * $P < 0.05$, *** $P < 0.001$

Table 5 Null allele frequencies estimated according to Van Oosterhout et al. (2006) for each locus presenting a significant excess of homozygotes per population. Two fixation indices previously calculated for metallicolous ($F_{is}^M = 0.36$) and nonmetallicolous ($F_{is}^{NM} = 0.64$) *Thlaspi caerulescens* populations were used for simulations (Dubois et al. 2003)

Population	Tc-ZNT1	Tc-ZNT2	Tc-IRT1	Tc-IRT2	Tc-up1	Tc-up3	Tc-NOD	Tc-up4
J1								
F_{is}^M	–	0.191	–	0.177	–	–	0.273	–
F_{is}^{NM}	–	–	–	–	–	–	–	–
J8								
F_{is}^M	–	0.096	–	0.035	–	0.146	–	0.197
F_{is}^{NM}	–	–	–	–	–	–	–	–
J12								
F_{is}^M	0.221	–	–	–	–	–	–	0.262
F_{is}^{NM}	–	–	–	–	–	–	–	–
A5								
F_{is}^M	–	–	–	–	–	–	–	0.186
F_{is}^{NM}	–	–	–	–	–	–	–	–
P								
F_{is}^M	–	–	–	–	–	–	0.251	0.324
F_{is}^{NM}	–	–	–	–	–	–	–	–

– means that the assessed null allele frequency is inferior to 1% for the considered F_{is} value

Table 6 Cross-species/genera amplification of loci in Brassicaceae: size range (in bp) of polymerase chain reaction (PCR) products

Locus	TA	AT	AH	BN	SA	EP	CBP	BL
Tc-ZNT1	249	370	281	241–272	247–262	257	418–430	–
Tc-ZNT2	139	139	–	m	–	m	139	–
Tc-ZNT5	157	160	154–157	154–157	157	98	154–157	157
Tc-CP	145	139–161	149	152	136–162	121	159–185	146–159
Tc-E2F1	–	–	–	–	–	–	–	–
Tc-IRT1	166	163	155	202	161–226	163–164	151–153	136
Tc-IRT2	238	210	–	234	212–232	218–234	231	–
Tc-WRKY	461	412	369	440	359	440	440	325
Tc-AGAMOUS	141	146	146–156	135	135–139	138	162–193	127–151
Tc-up1	111	–	126–127	–	–	107	–	–
Tc-up2	242	234–240	–	218–230	233	182	271–312	–
Tc-up3	195	184–195	–	–	176–182	–	–	–
Tc-NOD	209–213	209	–	–	–	–	–	–
Tc-up4	191	191–192	–	186–256	352–365	–	256–258	244–274
Tc-bHLH	149	133	135–143	134–238	110	143	143	95–105

Species code: TA *Thlaspi arvense*, AT *Arabidopsis thaliana*, AH *Arabis hirsuta*, BN *Brassica napus*, SA *Sinapis arvensis*, EP *Erophila praecox*, CBP *Capsella bursa-pastoris*, BL *Biscutella laevigata*

m multi-band pattern, – null or weak amplification

Utility of markers in *T. caerulescens*

We propose to use the present gene markers for specific applications in *T. caerulescens*. Compared with isozyme markers, which present a limited level of variation (maximum of four alleles; Dubois et al. 2003), use of the present markers will be more suitable to study *T. caerulescens* natural populations. However, several additional remarks have to be considered for their application in population genetics. First, some of the present markers may correspond to nonneutrally evolving loci, and this would violate population genetics assumptions. Indeed, for such loci, strong

physical linkage between the polymorphism investigated and mutations under selection may result in "hitchhiking" effects (Gupta et al. 2005). To identify loci under selection, a combination of different markers [amplified fragment length polymorphism (AFLPs), microsatellites and CAPS] has to be recommended, and simulations should be performed to model neutral loci and allow the identification of outlying loci (Beaumont and Nichols 1996). Based on this methodology, Swiss populations were recently characterised (N. Basic et al., in preparation). Second, some loci (particularly Tc-up4) could display null alleles in some populations (Tables 4, 5), and such markers have

to be used with caution. Some limitations may result, particularly for inbreeding or gene-flow measurements. Furthermore, our gene makers may be used in genetic mapping and quantitative genetic studies to investigate in particular if some of them (e.g. Tc-ZNT1, Tc-ZNT2, Tc-IRT1 or Tc-IRT2) colocate with QTLs for heavy-metal hyperaccumulation and could allow marker-assisted manipulation of a particular trait (e.g. Lakshmi et al. 2005; Assunçao et al. 2006).

Cross-genera PCR amplification of loci

Overall, 70% of the cross-species/genera amplifications were successful (Table 6). Among them, more than 40% provided polymorphic markers within a population (Table 6). All the loci except Tc-E2F1 amplified in *T. arvense* Tc-E2F1 (for which the forward primer was developed in a nonconserved DNA region) was not amplified in any other tested Brassicaceae apart from *T. caerulescens*. Comparatively to other investigations (e.g. Plieske and Struss 2001), our relatively high cross-species/genera amplification of gene markers in Brassicaceae may be related to the design of most of primers in highly conserved sequences (particularly in exon sequences) between distantly related species (i.e. *Thlaspi/Arabidopsis*, *Brassica/Arabidopsis* or *Alliaria/Arabidopsis*). We also observed that microsatellite loci isolated from genomic libraries (i.e. *Brassica*, *Alliaria* and *Lepidium*; Lowe et al. 2002; Durka et al. 2004; Bon et al. 2005) frequently correspond to expressed regions in *A. thaliana* (about 70% for the three cited studies). Consequently, public microsatellite genomic libraries can be considered as a source of variable and transferable gene markers in Brassicaceae. Cross-genera sets of markers may allow the comparison of genetic polymorphism between several Brassicaceae at genes putatively involved in traits of interest, such as heavy-metal tolerance and/or hyperaccumulation. They will be also useful for comparative mapping studies or population genetic analyses among and within species in the Brassicaceae family (Clauss et al. 2002; Boivin et al. 2004; Lowe et al. 2004).

References

Assunçao AGL, Schat H, Aarts MGM (2003a) *Thlaspi caerulescens*, an attractive model species to study heavy metal hyperaccumulation in plants. New Phytol 159:351–360

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

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

Basic N, Keller C, Fontanillas P, Vittoz P, Besnard G, Galland N (2006a) Cadmium hyperaccumulation and reproductive traits in natural *Thlaspi caerulescens* populations. Plant Biol 8:64–72

Basic N, Salamin N, Keller C, Galland N, Besnard G (2006b) Genetic differentiation of *Thlaspi caerulescens* natural populations in relation to their cadmium hyperaccumulation capacity. Biochem Syst Ecol 34: DOI 10.1016/j.bse.2006.04.001

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc Roy Soc Lond Ser B 263:1619–1626

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.03, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions CNRS UMR 5000, Université de Montpellier II, Montpellier (France). http://www.univ-montp2.fr/~genetix/genetix/genetix.htm

Bernard C, Roosens N, Czernic P, Lebrun M, Verbruggen N (2004) A novel CPx-ATPase from the cadmium hyperaccumulator *Thlaspi caerulescens*. FEBS Lett 569:140–148

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 Soil 249:9–18

Boivin K, Acarkan A, Mbulu RS, Clarenz O, Schmidt R (2004) The *Arabidopsis* genome sequence as a tool for genome analysis in Brassicaceae: a comparison of the *Arabidopsis thaliana* and *Capsella rubella* genomes. Plant Physiol 135:735–744

Bon MC, Hurard C, Gaskin J, Risterucci AM (2005) Polymorphic microsatellite markers in polyploidy *Lepidium draba* L. ssp. *draba* (Brassicaceae) and cross-species amplification in closely related taxa. Mol Ecol Notes 5:68–70

Bovet L, Eggmann T, Meylan-Bettex M, Polier J, Kammer P, Marin E, Feller U, Martinoia E (2003) Transcript levels of *AtMRPs* after cadmium treatment: induction of *AtMRP3*. Plant Cell Environ 26:371–381

Clauss MJ, Cobban H, Mitchell-Olds T (2002) Cross-species microsatellite markers for elucidating population genetic structure in *Arabidopsis* and *Arabis* (Brassicaeae). Mol Ecol 11:591–601

Dubois S, Cheptou PO, Petit C, Meerts P, Poncelet M, Vekemans X, Lefèbvre C, Escarré J (2003) Genetic structure and mating systems of metallicolous and nonmetallicolous populations of *Thlaspi caerulescens*. New Phytol 157:633–641

Durka W, Bossdorf O, Gautschi B (2004) Isolation and characterization of microsatellite loci in the invasive *Alliaria petiolata* (Brassicaceae). Mol Ecol Notes 4:173–175

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Frerot H, Petit C, Lefebvre C, Gruber W, Collin C, Escarré J (2003) Zinc and cadmium accumulation in controlled crosses between metallicolous and nonmetallicolous populations of *Thlaspi caerulescens* (Brassicaceae). New Phytol 157:643–648

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485

Koch M, Mummenhoff K, Hurka H (1998) Systematics and evolutionary history of heavy metal tolerant *Thlaspi caerulescens* in Western Europe: evidence from genetic studies based on isozyme analysis. Biochem Syst Ecol 26:823–838

Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: evidence from coding plastidic *matK* and nuclear *Chs* sequences. Am J Bot 88:534–544

Kuittinen H, Aguadé M, Charlesworth D, Haan ADE, Lauga B, Mitchell-Olds T, Oikarinen S, Ramos-Onsins S, Stranger B, Van Tienderen P, Savolainen O (2002) Primers for 22 candidate genes for ecological adaptations in Brassicaceae. Mol Ecol Notes 2:258–262

Lakshmi PK, Arumugam N, Gupta V, Mukhopadhyay A, Sodhi YS, Pental D, Pradhan AK (2005) Mapping and tagging of seed coat colour and the identification of microsatellite markers for marker-assisted manipulation of the trait in *Brassica juncea*. Theor Appl Genet 111:8–14

Lasat MM, Pence NS, Garvin DF, Ebbs SD, Kochian LV (2000) Molecular physiology of zinc transport in the Zn hyperaccumulator *Thlaspi caerulescens*. J Exp Bot 51:71–79

Lombi E, Zhao FJ, Dunham SJ, McGrath SP (2000) Cadmium accumulation in populations of *Thlaspi caerulescens* and *Thlaspi goesingense*. New Phytol 145:11–20

Lombi E, Tearall KL, Howarth JR, Zhao FJ, Hawkesford MJ, McGrath SP (2002) Influence of iron status on cadmium and zinc uptake by different ecotypes of the hyperaccumulator *Thlaspi caerulescens*. Plant Physiol 128:1359–1367

Lowe AJ, Jones AE, Raybould AF, Trick M, Moule CL, Edwards KJ (2002) Transferability and genome specificity of a new set of microsatellite primers among *Brassica* species of the U triangle. Mol Ecol Notes 2:7–11

Lowe AJ, Moule C, Trick M, Edwards KJ (2004) Efficient large-scale development of microsatellites for marker and mapping applications in *Brassica* crop species. Theor Appl Genet 108:1103–1112

Lugon-Moulin N, Zhang M, Gadani F, Rossi L, Koller D, Krauss M, Wagner GJ (2004) Critical review of the science and options for reducing cadmium in tobacco (*Nicotiana tabacum* L.) and other plants. Adv Agron 83:111–180

Mäser P, Thomine S, Schroder JI, Waed JM, Hirschi K, Sze H, Talke IN, Amtmann A, Maathuis FJ, Sanders D, Harper JF, Tchieu J, Gribskov M, Persans MW, Salt DE, Kim SA, Guerinot ML (2001) Phylogenetic relationships within cation transporter families of *Arabidopsis*. Plant Physiol 126:1646–1667

Peer WA, Mamoudian M, Lahner B, Reeves RD, Murphy AS, Salt DE (2003) Identifying model metal hyperaccumulating plants: germplasm analysis of 20 Brassicaceae accessions from a wide geographical area. New Phytol 159:421–430

Pence NS, Larsen PB, Ebbs SD, Letham DL, Lasat MM, Garvin DF, Eide D, Kochian LV (2000) The molecular physiology of heavy metal transport in the Zn/Cd hyperaccumulator *Thlaspi caerulescens*. Proc Natl Acad Sci USA 98:9995–10000

Plieske J, Struss D (2001) Microsatellite markers for genome analysis in *Brassica*. I. development in *Brassica napus* and abundance in *Brassicaceae* species. Theor Appl Genet 102:689–694

Pollard AJ, Powell KD, Harper FA, Smith JAC (2002) The genetic basis of hyperaccumulation in plants. Crit Rev Plant Sci 21:539–566

Roosens N, Leplae R, Bernard C, Verbruggen N (2005) Variations in plant metallothioneins: the heavy metal hyperaccumulator *Thlaspi caerulescens* as a case study. Planta 222:716–729

Roosens N, Verbruggen N, Meerts P, Ximénez-Embún P, Smith JAC (2003) Natural variation in cadmium tolerance and its relationship to metal hyperaccumulation for seven populations of *Thlaspi caerulescens* from western Europe. Plant Cell Environ 26:1657–1672

Suzuki N, Koizumi N, Sano H (2001) Screening of cadmium-responsive genes in *Arabidopsis thaliana*. Plant Cell Environ 24:1177–1188

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538

Van Oosterhout C, Weetman D, Hutchinson WF (2006) Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. Mol Ecol Notes 6:255–256

Yang YW, Lai KN, Tai PY, Ma DP, Li WH (1999) Molecular phylogenetic studies of *Brassica, Rorippa, Arabidopsis* and allied genera based on the internal transcribed spacer region of 18S–25S rDNA. Mol Phylogenet Evol 13:455–462

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