Implications of a phylogeographic approach for the selection of Ceutorhynchus assimilis as a potential biological control agent for Lepidium draba

Vincent Lesieur, Jean-François Martin, Hinz Hariet, Boris Fumanal, Sobhian Rouhollah, Marie-Claude Bon

Implications of a phylogeographic approach for the selection of *Ceutorhynchus assimilis* as a potential biological control agent for *Lepidium draba*

Authors:

Vincent Lesieur[1, 2]*, Jean-François Martin[2], Hariet L. Hinz[3], Boris Fumanal[4], Rouhollah Sobhian[1] and Marie-Claude Bon[1]

Affiliations:

[1] USDA ARS, European Biological Control Laboratory, 810 avenue du Campus Agropolis, 34980 Montferrier sur Lez, France

[2] Montpellier-SupAgro, UMR CBGP, 755 avenue du Campus Agropolis, 34980 Montferrier sur Lez, France

[3] CABI Switzerland, Rue des Grillons 1 CH-2800 Delémont, Switzerland

[4] Université Clermont Auvergne, INRA, PIAF, 63000 Clermont-Ferrand, France

*Corresponding author. Tel.: +33 499623041; fax: +33 499623049

E-mail address: vincent.lesieur@supagro.fr

Abstract

The root-galling weevil, *Ceutorhynchus assimilis* Paykull (Coleoptera, Curculionidae), has been selected as one of the candidates to control *Lepidium draba* L. (Brassicaceae), a highly invasive weed in western North America. Originally been described as oligophagous, also attacking crop species, previous studies indicated that a specific host race to *L. draba* exists in *C. assimilis*. We therefore explored the evolutionary history and phylogeographic structure of this weevil by sampling individuals from 67 populations, spanning 12 countries and five different host plant species including *L. draba*. To explore the genetic diversity of *C. assimilis*, we analysed the *COI* gene sequences of 458 individuals. Analysis revealed three distinct evolutionary lineages, one of which, the so-called *Lepidium* host race, was only found on *L. draba* and appears to occur in a restricted geographic area,

ranging between northern Spain and northern Italy. These results allow us to targeting sites for collection of the *Lepidium* host race and justify the prioritization of *C. assimilis Lepidium* host race as a candidate for the *L. draba* biological control.

Keywords

Invasive weed; Heart-podded hoary cress; Root-galling weevil; Host race; Genetic diversity; *COI* gene

1. Introduction

Heart-podded hoary cress, *Lepidium draba* L. ssp. *draba* [= *Cardaria draba* (L.) Desv.] (Brassicacae), is a perennial rhizomatous weed, introduced to the US as a contaminant of seed shipments from Eurasia in the late 19th century (Gaskin et al., 2005; Hinz et al., 2012). The species has since spread throughout North America and is especially problematic in the West, where it is a declared noxious weed in 15 US states, three Canadian provinces, and also in Mexico (Rice, 2014; USDA-NRCS, 2016). *Lepidium draba* competes with native plants, decreasing local biodiversity (Francis and Warwick, 2008; Puliafico et al., 2011). Besides its ecological impact, it causes economic losses by invading several crops (McInnis et al., 2003; Mulligan and Findlay, 1974) where it also poses a dual problem: as well as being a weed, it provides an alternative food source for major crop pests, such as the cabbage seedpod weevil, *Ceutorhynchus obstrictus* (Marsham) (Dosdall and Moisey, 2004; Fox and Dosdall, 2003). Furthermore, the weed, which is well adapted to sub-irrigated pastures and rangelands, is toxic to cattle due to alkaloid compounds, thus becoming a serious problem for livestock and forage production (Francis and Warwick, 2008; McInnis et al., 2003).

The invasive success of *L. draba* is related to its vegetative reproduction by an extensive rhizome system as well as its high seed production (Francis and Warwick, 2008; McInnis et al., 2003; Mulligan and Findlay, 1974). The use of herbicides is costly, not always feasible and mechanical control is not effective due to the extensive underground rhizome system of the weed (Miller et al., 1994). These limitations have motivated the exploration of alternative methods, and a classical biological control project was started in 2001 (Anonymous, 2004).

Among the known natural enemies of *L. draba* in its native range in Eurasia, *Ceutorhynchus assimilis* Paykull, 1792 [syn. *Ceutorhynchus pleurostigma* (Marsham 1802)] (Coleoptera, Curculionidae) (Colonnelli, 1993) has been prioritized as a of candidates for biological control (Fumanal et al., 2004a; 2004b; Hinz et al., 2013; Hinz et al., 2016; Virag et al., 2016). Adults lay most of their eggs in young and soft roots of the plant, causing the development of galls in which the three larval instars of the weevil develop before larvae leave the galls and pupate in the soil (Hoffman, 1954). Across its Eurasian distribution, the weevil has been assumed to display a broad host range and is listed as a pest of more than 13 plant species including crops (Dennis, 1987; Hoffman, 1954; Jourdheuil, 1963). However, Fumanal et al. (2004b) proposed that *C. assimilis* is probably a complex of morphocryptic entities that differ in their host-plant spectrum within the family Brassicaceae and one of these is a specialist on *L. draba* with a distribution in the warmer climatic regions of southern Western Europe (Fumanal et al., 2004a). Whilst the host-specificity and life-cycle of this specialist entity have been studied (Fumanal et al., 2004b; Hinz et al., 2016; Virag et al., 2016), population genetic structure and the evolutionary relationships between it and the others remain poorly depicted. Investigating the population genetic structure of potential candidate agents at the onset of biological control programs would streamline host-specificity testing of differentiated genetic groups of the agent and hence can help improve the safety and success of biological control of weeds (De Biase et al., 2016; Gaskin et al., 2011; Rauth et al., 2011; Rector et al., 2010).

The aims of the current study were to further elucidate the genetic diversity of *C. assimilis* across its native range and between different host plants, to better characterize the distribution of the different lineages or host races and to reconstruct its evolutionary history. To do so, extensive sampling was carried out across the weevil's native range, with samples collected from different known host-plants of *C. assimilis*. Preliminary sampling (the one used in Fumanal et al., 2004a) was enriched with new populations collected in eastern and continental Europe, but also with more individuals analyzed per population, especially those associated with *L. draba*.

The samples were used to analyze the genetic structure of *C. assimilis* by examining the mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene, a marker that has proven its usefulness and is widely used in biological control programs (Cheyppe-Buchmann et al., 2011; De Biase et al., 2016;

Nadel et al., 2012; Rauth et al., 2011; Tavares et al., 2015). The results were expected to provide essential information which would help to make decision regarding the suitability of the weevil as a biocontrol agent of *L. draba*, and in particular for selecting areas where the most-appropriate *L. draba* host race should be collected for further host specificity studies.

2. Materials and Methods

2.1. Insect sampling

Galls containing mature larvae of *C. assimilis* were collected in the field between 2001 and 2014 from the five taxa within the Brassicaceae known to be the most commonly attacked by the weevil: *L. draba* (n = 366), *Brassica oleracea* L. (n = 52), *Brassica napus* L. (n = 7), *Diplotaxis erucoides* Dc. (n = 35) and *Sinapis arvensis* L. (n = 18). Plants were sampled at least 1 m apart and only one larva per gall and per plant was collected. Some of the larvae were reared to adult for morphological confirmation by Enzo Colonnelli (University of Rome, Italy). The others were stored in absolute ethanol at -20°C until DNA extraction. Sampling covered a large part of the Eurasian distribution of *C. assimilis* and collection sites were distributed across 12 countries (Table 1). By combining host-plants and geographic collection sites, the study included no less than 67 populations. Between one and ten individuals from each population were analyzed. However, for some populations of interest, up to 28 individuals were analyzed.

2.2. *Molecular characterization protocol*

Genomic DNA was extracted from 458 individuals using a modified cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987) and more recently using Qiagen DNeasy Blood & Tissue kit following manufacturer's protocol.

To explore intraspecific diversity, the *COI* mitochondrial gene was partially amplified (831 bp) with the primers C1-J-2183 (5'-CAACATTTATTTTGATTTTTTGG-3') and TL2-N-3014

(5'-TCCAATGCACTAATCTGCCATATTA-3') (Simon et al., 1994). The PCR reaction mix contained, 1X Qiagen buffer, 0.2 µM of each dNTP, 0.3 µM of each primer and 2 U of *Taq* polymerase and 2 µL of genomic DNA diluted at 1/5 in a final volume of 25 µL. PCR were performed using the following program: 94°C for 3 min, followed by 35 cycles of 92°C for 30 s, 52°C for 30 s and 62°C for 1 min and finished by a final elongation at 62°C for 7 min. The amplicons were purified and sequenced in both directions by Genoscreen (Lille, France) using an ABI PRISM 3730XL DNA sequencer. Sequences were assembled into consensus contigs with CodonCode Aligner (www.codoncode.com) and then aligned using CLUSTAL W. In addition, sequences of closely related weevil species (*i.e*. other *Ceutorhynchus* species) were identified by a NCBI BLAST search in GenBank with one of the newly derived COI sequences of *C. assimilis*. Three *COI* sequences of two species (*C. obstrictus*, GenBank accession number: DQ058695 and JN163956; *C gallorhenanus*, GenBank accession number: DQ058700) were included as outgroups in the phylogenetic analysis.

2.3. Phylogenetic analysis and divergence time estimation

Relationships among *C. assimilis* from different host plants and geographical regions were reconstructed using both maximum likelihood and Bayesian inferences. First, jModelTest2 (Darriba et al., 2012) was used to test for the best-fit model of sequence evolution. The best-fit model was selected using the corrected Akaike Information Criterion (AICc) (Darriba et al., 2012). Maximum likelihood analyses were conducted with PhyML v3.0 (Guindon et al., 2010). Nodal support was assessed using non-parametric bootstrapping (1,000 replicates). Bayesian phylogeny was reconstructed using the software BEAST v1.8.0 (Drummond et al., 2012). We also used this software to estimate divergence time of lineages characterized within *C. assimilis*. In the absence of geological and/or fossil calibration points to estimate divergence times, we dated the clock-like *COI* phylogeny by applying a global mutation rate of 3.54% Myr^{-1}, a substitution rate calibrated for tenebrionid beetles (Papadopoulou et al., 2010). We used an uncorrelated lognormal (UCLN) relaxed clock under exponential growth population models. The analyses were performed using two independent runs of 20 million generations and trees were sampled every 1,000 iterations. Input files were generated with

BEAUTi (Drummond et al., 2012) v1.8.0. After checking for the effective sampling sizes (ESS) of parameters (a threshold of 200 was used) with Tracer v1.6 (available from http://tree.bio.ed.ac.uk/software/tracer), the first 25% of each run was discarded as burn-in phase for the estimation of the consensus topology and the computation of the posterior probability (PP) for each node. The intraspecific relationships might be better visualized with networks, allowing a more detailed display of population information than on evolutionary trees (Posada and Crandall, 2001) although both may be informative. A haplotype network was constructed in PopART (Leigh and Bryant, 2015) using TCS network. To visualize the spatial distribution of the different entities (*e.g.* *Lepidium* host race *versus* the other lineages), the genetic diversity of the populations was plotted as pie charts on a map.

To estimate the number of molecular operational taxonomic units (MOTUs) and to explore molecular species delimitation, two different methods were used. First, we used Automatic Barcode Gap Discovery (ABGD) developed by Puillandre et al. (2012). This tool uses pairwise distance to detect a barcode gap, this barcode gap is used as a threshold to separate the sequences into putative species (*i.e.* genetically homogeneous groups). The procedure is recursively applied to the previously obtained groups until no more partitioning can be done. The analysis was conducted through uploading a sequence alignment on the ABGD web server at http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html. We computed a matrix of pairwise distances using the K2P model (Kimura, 1980) and ABGD was run with the default settings (Pmin = 0.001, Pmax = 0.1, Steps = 10, X (relative gap width) = 1.5, Nb bins = 20). The Poisson tree processes (PTP) model for species delimitation (Zhang et al., 2013) was also used. This model uses coalescence theory and estimates the speciation rate directly from the number of substitutions. It assumes that each substitution has a small probability of generating a speciation event. Consequently, the number of substitutions between species is expected to be significantly higher than within species. The ML phylogeny obtained with PhyML was used as input. The analysis was run on a web server for PTP (available at http://species.h-its.org/ptp/) with 200,000 MCMC generations, a thinning value of 100 and a burn-in of 25%. As recommended by the developers, the convergence of the MCMC chain was confirmed visually.

To estimate divergence between the main genetic entities defined by the resulting phylogenetic tree, we used Mega version 7.0 (Kumar et al., 2016). We calculated the pairwise genetic *p*-distances among the phylogenetic entities. To get a rough overview of the historical demography for each lineage, the frequency-based indicators of a population expansion (or selection in non-neutral markers) Fu's *Fs* (Fu, 1997) were calculated with Arlequin 3.5 (Excoffier and Lischer, 2010). Fu's statistic is considered highly sensitive to detect population expansion (Fu, 1997; Ramos-Onsins and Rozas, 2002). The values of Fu's *Fs* are expected to have significantly negative values for population expansion.

The estimation of gene diversity (H_d) and nucleotide diversity (π) for each of the sampling locality was conducted with Arlequin 3.5 (Excoffier and Lischer, 2010) and allelic richness (*r*) was computed using the rarefaction method proposed by Petit et al. (1998) using Contrib (http://www.pierroton.inra.fr/genetics/labo/Software/Contrib/).

Finally, to analyze the genetic structure among populations and to identify potentially distinct units within the restricted area where the weevil *Lepidium* host race was found (see Results), we used a spatial analysis of molecular variance (SAMOVA) with SAMOVA 1.0 (Dupanloup et al., 2002). This approach was thus only used for individuals collected on *L. draba*. It defines groups of populations that are geographically homogeneous and maximally differentiated. The program was run for two to 10 differentiated groups ($K = 2$ to $K = 10$) using 10,000 permutations from 100 random initial conditions.

3. Results

The 458 sequenced individuals from 67 populations yielded a final alignment of 774 bp long uncovering a total of 66 haplotypes. All haplotype sequences were deposited in GenBank (accession numbers: KY486382-KY486447). From the final alignment, a total of 59 (7.6%) polymorphic nucleotide sites, of which 46 were parsimony-informative. Total gene diversity was 0.929 ± 0.006 and nucleotide diversity was 0.013 ± 0.003.

196 The most appropriate model of *COI* sequence evolution was TN93+I+G, and this model was
197 employed for further analyses. Phylogenetic reconstructions using both ML and BI revealed three
198 divergent genetic entities within *C. assimilis* (Fig. 1). The first of them, referred to hereafter as the
199 '*Lepidium* host race', was defined with strong support (BS = 94 and PP = 1.00). This evolutionary
200 lineage, composed of 12 haplotypes, was represented by weevils sampled only on *L. draba* ssp. *draba*
201 within a restricted area comprising northern Spain, southern France and the northwestern part of Italy
202 (Figs. 2 and 3). The second evolutionary lineage, called Lineage 2 (support values: BS = 83 and
203 PP = 0.99), was composed of seven haplotypes (Fig. 1). These haplotypes were recovered in southern
204 France, northern Italy and Greece from individuals collected on three different host plants: *D.*
205 *erucoides*, *S. arvensis* and *B. oleracea* (Figs. 2 and 3). This lineage was never found on *L. draba* in our
206 study. The remaining haplotypes clustered in Lineage 1 (Fig. 1). This evolutionary lineage was widely
207 distributed across the entire sampled area and from the whole spectrum of host plants considered in
208 this study (Fig. 3). On internal nodes, phylogeny is not fully resolved with low support values or
209 discrepancy between ML and BI (Fig. 1).
210 As observed on the phylogenetic tree, the haplotype network also revealed a clear split
211 between the *Lepidium* host race and the two other lineages (Fig. 3). The haplotype network showed
212 that the *Lepidium* host race and Lineage 1 differed from each other by at least 14 mutation steps
213 (Fig. 2).
214 The genetic divergence between sequences obtained from the *Lepidium* host race and the two
215 other lineages were 2.41% and 2.21%, respectively for Lineage 1 and Lineage 2 (Table 2). This is
216 much lower than divergence values obtained with the *Ceutorhynchus* species used as outgroups (11.38
217 – 12.95%). From the ABGD analysis, the partitions 1-3 (P = 0.001000-0.002783) matched the
218 phylogenetic tree and partitioned the dataset into three molecular operational taxonomic units
219 (MOTUs) corresponding to Lineage 1, Lineage 2 and *Lepidium* host race. However, ABGD results did
220 not fully support a marked barcoding gap (Supplementary material Fig. 1). The partition 4
221 (P = 0.004642) only resulted in two MOTUs while the partition 5 (P = 0.007743) suggested that *C.*
222 *assimilis* was a single species. The PTP analysis resulted in partitioning *C. assimilis* into two MOTUs,
223 the *Lepidium* host race on the one hand and the combination of the Lineages 1 and 2 on the other hand.

This analysis suggested that these MOTUs may represent potential cryptic species or subspecies within *C. assimilis*.

Estimation of time divergence was made for the two major splits within the *C. assimilis* tree (Fig. 1). Divergence between the *Lepidium* host race and the other lineages of *C. assimilis* (labeled with Roman numeral I on Fig. 1) was estimated to have occurred 1.17 Mya (95% HPD: 0.70–1.76), this estimation dates the split between the lower and the mid-Pleistocene. The Lineage 1 and Lineage 2 (referred on Fig. 1 as the Roman numeral II), diverged from each other around 0.87 Mya (95% HPD: 0.50–1.26). Demographic analyses suggest a population expansion in the three lineages (Table 2), although for the Lineage 2, negative value of Fu's *Fs* was not significant.

Most populations within sampling localities had one to three haplotypes. Haplotype H 21, belonging to Lineage 1, was the most common and the most widespread haplotype. It was found in 83 individuals from 17 sampling sites, distributed from France to Armenia. In the restricted area where the *Lepidium* host race was found, the most geographically widespread haplotype corresponded to H 1, and was found in 10 sampling sites (n = 35 individuals), and only on *L. draba*. Both genetic diversity and nucleotide diversity varied considerably from one sampling site to another in the restricted area of the *Lepidium* host race as well as in the other parts of Europe (Table 1).

Within the restricted area where the *Lepidium* host race lineage was found, and taking into account only the weevils collected on *L. draba*, the SAMOVA analysis suggested the presence of five clusters in the dataset. This configuration was retained because the largest increase in F_{CT} values was observed when the geographic sampling area was partitioned into five clusters (Supplementary material Fig. 2), although the F_{CT} values still slowly increased after K = 5, probably a response to the continuous decrease in F_{SC} until all sampling sites are separated (Dupanloup et al., 2002). Moreover, for K > 5, the newly formed groups only contained single populations, which indicates that the group structure disappeared (Magri et al., 2006). The five groups were geographically consistent, except group 4 that showed a disjunctive distribution, occurring mainly in France and Northern Spain but also in Northern Italy (Fig. 4). The population A84, located in Central Spain was mostly composed of individuals harboring the haplotype H55 that is only found in this sampling site (n = 26). This population was segregated from the other Spanish populations and considered as a distinct group

(group 5). Group 1, located in the northern continental range, clustered the sites A83 and A69. Those two sites are predominantly composed of individuals from the generalist Lineage 1 (Fig. 3), 14 individuals among 20 for site A83 and 12 individuals among 14 for site A69, respectively.

4. Discussion

4.1. Evolutionary history of C. assimilis

The results presented here, confirm the proposal (Fumanal et al., 2005; Fumanal et al., 2004b) that the weevil populations developing on *L. draba* in a restricted area of the western Mediterranean region should be considered as a *Lepidium* host race. Present results met most of the criteria proposed by Drès and Mallet (2002) in their definition of host race including genetically differentiated sympatric populations living on different hosts. Indeed, the lineage referred to here as the *Lepidium* host race represents a divergent evolutionary lineage strictly associated with *L. draba*. The genetic divergence between the host race and the other lineages of *C. assimilis* (2.21% with Lineage 1 and 2.41% with Lineage 2) corresponds to values lower than those usually observed between *Ceutorhynchus* species (Laffin et al., 2005a) as confirmed by the divergence observed with the congeneric species used as outgroups. However, the level of divergence detected here is in the same order of magnitude as the one observed between two sister species, *C. erysimi* and *C. contractus* (Stepanović et al., 2015). Likewise, although the ABGD procedure tends to consider the three different lineages as a single species, the PTP analysis suggests that the *Lepidium* host race may potentially represent cryptic species or subspecies. The potential for hybridization between the different genetic entities has never been investigated in the field, and only partly in lab conditions (Fumanal et al. 2004a).

Preliminary crossing experiments between individuals of *C. assimilis* from the *Lepidium* host race and individuals reared from *L. draba* but belonging to the generalist Lineage 1 showed an absence of prezygotic barriers but suggested partial postzygotic barriers. Indeed, fertilization seemed possible and egg laying from females was observed in some cases but it was not possible to test for

fertility of the resulting outcrossed progenies as they died before they reproduced. Crossing of individuals belonging to the *Lepidium* host race with individuals developing on other host plants (Lineage 2) resulted only in abnormal small galls and abortion of neonate larvae (Fumanal et al., 2004a). Taken as a whole, the available data (*i.e*. ecological, biological and genetic data) point to a limited gene flow between the *Lepidium* host race and the two other lineages. However, the extent of hybridization and gene flow between the different entities remain to be found under natural conditions. This could be tested in the Rhone Valley, where the *Lepidium* host race and more generalist entities co-occur on *L. draba*. The recent development of microsatellite markers for *C. assimilis* (Lesieur et al., 2016) will enable investigation of the gene flow between the different entities identified here.

The split between the *Lepidium* host race and the two other lineages was estimated to occur during the Pleistocene and, diversification within each lineage also started during this period. The Pleistocene and its climatic cycles, with the succession of glacial and interglacial periods, have had a great impact on species evolution and have shaped the current distribution of numerous organisms (Hewitt, 2000; Médail and Diadema, 2009; Schmitt, 2007). The western Mediterranean distribution of the *Lepidium* host race suggests one (or more) refugia in Spain or the south of France, a region identified as glacial refugium for several other species (Médail and Diadema, 2009).

Over its wide distribution, no clear geographic structure within Lineage 1 was found in the analysis. However, several glacial refugia for *C. assimilis* and further expansion of both lineages are suggested. In particular, neutrality tests support the hypothesis of a recent population expansion. The current distributions of Lineage 1 and the *Lepidium* host race reveal a secondary contact of those lineages in the Rhone Valley. Except for the sampling site A83 located in the Massif Central, where the *Lepidium* host race is not the main genetic lineage, the distribution of the *Lepidium* host race in France matches the Mediterranean biogeographical area (Condé et al., 2002). This would suggest that past but also present climatic conditions may shape the distribution of this lineage. The geographic distribution of the Lineage 2 (*i.e.* Greece, Southern France and also three individuals in Italy) is also consistent with southern refugia (Médail and Diadema, 2009). This lineage, only found on *D. erucoides*, *S. arvensis* and *B. oleracea* was never associated with *L. draba*. It may be more widespread than illustrated by our study, but this lack of representation may be linked with the lower

sampling effort on these plants at the global scale. Moreover, the moderate support values for Lineages 1 and 2 need further considerations for better resolving the evolutionary history of these two lineages.

Within the area where the *Lepidium* host race is present, Médail and Diadema (2009) defined several 'phylogeographical hotspots' favorable to the evolutionary divergence of different plant species. In line with this, the weevil's genetic divergence correlates with the differentiation of the European *L. draba* ssp *draba.* Fumanal et al. (2004a) found two different plant genetic lineages within the European *L. draba*, one restricted to Southern Europe and one widely distributed in Europe. As for *C. assimilis* in which the lineages were morphologically similar, a comparison of several plant traits (such as shoot density, individual shoot height) of European populations of *L. draba* from several regions did not reveal any differences (Hinz et al., 2012). However, the authors did not include in their study any population of the weed from the area where the *C. assimilis Lepidium* host race occurs. In our study, the divergence found between the *Lepidium* host race and the other genetic entities of *C. assimilis* can also be linked with biological features of weevil populations. Differences in phenology, sex ratio, cold hardiness and overwintering survival have been observed between the different entities (Fumanal et al., 2004b; Virag et al., 2016). For instance, Fumanal et al. (2004b) observed in the area where the *Lepidium* host race is found that sympatric weevil populations on *L. draba* developed faster and pupation started up to two months earlier than individuals on *S. arvensis*. Phenological differences are an important factor in sympatric speciation or, at least, host plant or habitat specialization (Drès and Mallet, 2002; Roy et al., 2016; Santos et al., 2007). These biological differences could partly explain the genetic divergence between sympatric populations living on different host- plants as observed in the two sites in southern France (*i.e.* Q-VV-VVS and A49-L-B (Table 1)) where the weevils were collected synchronously.

Despite numerous taxonomic revisions of the genus (Colonnelli, 1993; Colonnelli, 2004; Hoffman, 1954; Tempère and Péricart, 1989), taxonomical and nomenclatural issues represent the main constraints in studying this important weevil group, as highlighted by Stepanović et al. (2015). In this context, although it was not the main aim of this study, we confirmed the distinction between the cabbage seedpod weevil, *C. obstrictus* (Marsham) [= *C. assimilis* (Paykull)], a serious European pest

of canola or oil seed rape, which is invasive in North America (Dosdall and Mason, 2010; Laffin et al., 2005b), and *C. assimilis* [= *C. pleurostigma*]. Colonnelli (1993) determined that *C. obstrictus* is the correct name for the cabbage seedpod weevil, however confusion still exists in some publications where *C. obstrictus* is still cited as *C. assimilis*. Moreover, *C. assimilis* has been listed as a pest of several cultivated crops such as cabbage or rapeseed (Hoffman, 1954; Jourdheuil, 1963).

4.2. Implications for biological control

The present distinction between the specialist (*i.e.* the *Lepidium* host race) and the other lineages has serious implications from a biological control perspective. Indeed, the *Lepidium* host race should be considered as beneficial and a potential biological control agent for *L. draba* whereas the others, particularly those developing on economic crops, should still be regarded as pests. First, the results allow targeting precise sites for collection of the *Lepidium* host race. We showed that the *Lepidium* host race of *C. assimilis* is restricted to a relatively small geographic area. On the 122 individuals collected on *L. draba* within a triangular area including Valls in Cataluna (Spain), San Remo in Ligury (Italy) and Montélimar in the Rhone Valley (France), 116 individuals (ca. 95%) corresponded to the *Lepidium* host race.

Secondly, within the restricted distribution area where the *Lepidum* host race occurs in sympatry with the other lineages, Lineage 1 was observed on *L. draba* at a low frequency and Lineage 2 has never been found on *L. draba* but on other Brassicaceae. Moreover, when sampling was carried out synchronously on several host plant species, the *Lepidium* host race has never been found on plants other than *L. draba,* thus reducing the risk of collecting generalist individuals. Indeed, these other plant species do not constitute acceptable hosts for the *Lepidium* host race (Fumanal et al., 2004b). Thus, when restricting sampling to sites/areas where the *Lepidium* host race predominates and only collecting galls/adults from *L. draba,* one can be close to certain to only collect individuals of the specialist *Lepidium* host race. Nevertheless, as previously mentioned, the different entities are morphologically identical, discriminating the *Lepidium* host race from the other lineages remains an issue. However, Fumanal et al. (2005) developed a non-invasive technique based on direct-PCR of

fecal DNA combined with double strand conformation polymorphism (DSCP) typing. This method allows to quickly discriminating the *Lepidium* host race from the other entities without killing the insects and could be adapted for systematic screenings with current metabarcoding methods.

Thirdly, based on our results, the split between the *Lepidium* host race and the other more generalist lineages is relatively old (1.17 Mya, Pleistocene) suggesting that the *Lepidium* host race belongs to a long-standing *L. draba*-associated entity rather than a lineage prone to a recent host shift. As such it should remain specific to *L. draba*, which justifies its prioritization as a candidate for biological control of *L. draba*.

Even if the results presented here justify the prioritization of the *Lepidium* host race as a candidate for the *L. draba* biological control, studies on the biology and the specificity of the host race must be undertaken to evaluate its potential host spectrum in North America. Host specificity tests, including North American endemic plant species, are currently conducted by CABI at Délemont (Switzerland).

Genetic diversity within the *Lepidium* host race was highly variable from one sites to another, and in some sampling sites the genetic diversity was very low (*e.g.* sites A55; A69; A72 (Table 1)). At the sampling site scale, only one to three haplotypes were observed, even when the sampling size was large ($n > 25$). Collecting individuals from several nearby sites should ensure maximizing genetic diversity. Preserving this diversity in released populations is of importance as genetic variation in ecologically important traits is hypothesized to be important in successful establishment, allowing adaptation to the novel environment (Dlugosch et al., 2015; Szucs et al., 2014).

Lepidium draba is invasive in a large part of North America, including areas with a wide spectrum of climatic conditions (Rice, 2014; USDA-NRCS, 2016). A climate match model suggests that the *Lepidium* host race may establish and persist in some areas where the weed is problematic, while establishment in more continental climates could be possible, but more difficult (Virag et al., 2016). However, relatively rapid adaptation after release may occur as observed for the flea beetle *Longitarsus jacobaeae* (Waterhouse) (Szucs et al., 2012). Preserving the genetic variation of the *Lepidium* host race in released populations may increase their adaptive potential in the new

environment (Phillips et al., 2008), and therefore also potential adaptation to more continental climate, such as eastern Colorado, Wyoming and central Montana.

The *L. draba* populations that invaded North America displayed genetic differentiation and originated from multiple Eurasian origins including the southern European region where the weevil host race occurs (Gaskin et al., 2005). Likewise, within its restricted distribution area, the genetic diversity of the *Lepidium* host race appears as different units as highlighted by the SAMOVA. Different populations of potential biological control agents may have different abilities to develop and damage different introduced genotypes of the target weed, as it was observed for the eriophyid mite, *Floracarus perrepae* Knihinicki and Boczek, to control the fern, *Lygodium microphyllum* (Cav.) R. Br. in Florida (Goolsby et al., 2006). Because of the common pattern of differentiation between the *Lepidium* host race of *C. assimilis* and *L. draba*, future investigations should evaluate the impact of the different units of the *Lepidium* host race on different populations of *L. draba*, including North American populations.

Acknowledgements

We thank the Director of EBCL, Lincoln Smith, for reviewing the manuscript and providing useful comments. We thank Camille Rouvière for her help in extracting DNA and sequencing. We also thank Fatiha Guermache and Mélanie Jeanneau (USDA-ARS-EBCL) for technical assistance and field collections. We are also indebted to many collaborators who provided samples. Enzo Colonnelli (University of Roma, Italy) is gratefully acknowledged for identification of the weevils. We would also like to thank Dr. G. Kergoat (INRA, UMR CBGP) for helpful discussions on data analyses and interpretation. Mention of any trade names or commercial products in this publication is done solely for the purpose of providing specific information, and does not imply recommendation or endorsement by the USDA; the USDA is an equal opportunity provider and employer. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Anonymous, 2004. Whitetop / Hoary Cress. In: Coombs, E.M., Clark, J.K., Piper, G.L., Cofrancesco Jr., A.F. (Eds.), Biological Control of Invasive Plants in the United States. Oregon State University Press. Oregon State University Press, pp. 467.

Cheyppe-Buchmann, S., Bon, M.C., Warot, S., Jones, W., Malausa, T., Fauvergue, X., Ris, N., 2011. Molecular characterization of *Psyttalia lounsburyi*, a candidate biocontrol agent of the olive fruit fly, and its *Wolbachia* symbionts as a pre-requisite for future intraspecific hybridization. Biocontrol 56**,** 713-724.

Colonnelli, E., 1993. The Ceutorhynchinae types of I.C. Fabricius and G. von Paykull (Coleoptera: Curculionidae). Koleopterologische Rundschau 63**,** 299-310.

Colonnelli, E., 2004. Catalogue of Ceutorhynchinae of the world, with a key to genera. (Insecta: Coleoptera: Curculionidae).

Condé, S., Richard, D., Liamine, N., Leclère, A.-S., Framstad, E., Erikstad, K.E., Yoccoz, N., 2002. The Mediterranean biogeographical region. European Environment Agency, Sweden.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature methods 9**,** 772-772.

De Biase, A., Colonnelli, E., Belvedere, S., La Marca, A., Cristofaro, M., Smith, L., 2016. Genetic and morphological studies of *Trichosirocalus* species introduced to North America, Australia and New Zealand for the biological control of thistles. Bulletin of entomological research 106**,** 99-113.

Dennis, S., 1987. Agricultural insect pest of temperate regions and their control. Cambridge University Press, London, England.

Dlugosch, K.M., Anderson, S.R., Braasch, J., Cang, F.A., Gillette, H.D., 2015. The devil is in the details: genetic variation in introduced populations and its contributions to invasion. Molecular Ecology 24**,** 2095-2111.

Dosdall, L.M., Mason, P.G., 2010. Key pests and parasitoids of oilseed rape or canola in North America and the importance of parasitoids in integrated management. In: Williams, I.H., (Ed.), Biocontrol-based integrated management of oilseed rape pests. Springer, Dordrecht, The Netherlands, pp. 167-213.

Dosdall, L.M., Moisey, D.W., 2004. Developmental biology of the cabbage seedpod weevil, *Ceutorhynchus obstrictus* (Coleoptera: Curculionidae), in spring canola, Brassica napus, in western Canada. Annals of the Entomological Society of America 97**,** 458-465.

Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19**,** 11-15.

Drès, M., Mallet, J., 2002. Host races in plant-feeding insects and their importance in sympatric speciation. Philosophical Transactions of the Royal Society B-Biological Sciences 357**,** 471-492.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution 29**,** 1969-1973.

Dupanloup, I., Schneider, S., Excoffier, L., 2002. A simulated annealing approach to define the genetic structure of populations. Molecular Ecology 11**,** 2571-2581.

Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular ecology resources 10**,** 564-567.

Fox, A.S., Dosdall, L.M., 2003. Reproductive biology of *Ceutorhynchus obstrictus* (Coleoptera: Curculionidae) on wild and cultivated Brassicaceae in southern Alberta. Journal of Entomological Science 38**,** 533-544.

Francis, A., Warwick, S., 2008. The biology of Canadian weeds. 3. *Lepidium draba* L., *L. chalepense* L., *L. appelianum* Al-Shehbaz (updated). Canadian Journal of Plant Science 88**,** 379-401.

Fu, Y.-X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147**,** 915-925.

Fumanal, B., Martin, J.F., Sobhian, R., Gaskin, J., Bon, M., 2004a. Evolutionary biology as a tool towards a more customized biological control strategy of weeds: *Lepidium draba* as a case study. AFPP, XII Colloque international sur la Biologie des mauvaises herbes, Dijon, France, pp. 421-426.

Fumanal, B., Martin, J.F., Bon, M.C., 2005. High through-put characterization of insect morphocryptic entities by a non-invasive method using direct-PCR of fecal DNA. Journal of Biotechnology 119**,** 15-19.

Fumanal, B., Martin, J.F., Sobhian, R., Blanchet, A., Bon, M.C., 2004b. Host range of *Ceutorhynchus assimilis* (Coleoptera : Curculionidae), a candidate for biological control of *Lepidium draba* (Brassicaceae) in the USA. Biological Control 30**,** 598-607.

Gaskin, J.F., Bon, M.-C., Cock, M.J., Cristofaro, M., De Biase, A., De Clerck-Floate, R., Ellison, C.A., Hinz, H.L., Hufbauer, R.A., Julien, M.H., 2011. Applying molecular-based approaches to classical biological control of weeds. Biological Control 58**,** 1-21.

Gaskin, J.F., Zhang, D.Y., Bon, M.C., 2005. Invasion of *Lepidium draba* (Brassicaceae) in the western United States: distributions and origins of chloroplast DNA haplotypes. Molecular Ecology 14**,** 2331-2341.

Goolsby, J.A., De Barro, P.J., Makinson, J.R., Pemberton, R.W., Hartley, D.M., Frohlich, D.R., 2006. Matching the origin of an invasive weed for selection of a herbivore haplotype for a biological control programme. Molecular Ecology 15**,** 287-297.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59**,** 307-321.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405**,** 907-913.

Hinz, H., Blair, L., Closca, C., Diaconu, A., 2013. Biological control of whitetops, *Lepidium draba*, *L. chalepense* and *L. appelianum.* Annual Report 2012. CABI Switzerland.

Hinz, H.L., Cloşca, C., Castellan, I., Heijs, W., Morelon, S., Stahlke, A., 2016. Biological control of whitetops, *Lepidium draba*, *L. chalepense* and *L. appelianum*. Annual Report 2015. CABI Switzerland.

Hinz, H.L., Schwarzländer, M., McKenney, J.L., Cripps, M.G., Harmon, B., Price, W.J., 2012. Biogeographical comparison of the invasive *Lepidium draba* in its native, expanded and introduced ranges. Biological Invasions 14**,** 1999-2016.

Hoffman, A., 1954. Faune de France. Office Central de Faune publ., Paris, France.

Jourdheuil, P., 1963. *Ceutorhynchus pleurostigma* Marsham. In: Balachowsky, A.S., (Ed.), Entomologie appliquée à l'agriculture (Coléoptères). Masson et Cie.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of molecular evolution 16**,** 111-120.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33, 1870-1874.

Laffin, R., Dosdall, L., Sperling, F., 2005a. Population structure and phylogenetic relationships of *Ceutorhynchus neglectus* (Coleoptera: Curculionidae). The Canadian Entomologist 137**,** 672-684.

Laffin, R.D., Dosdall, L.M., Sperling, F.A.H., 2005b. Population structure of the cabbage seedpod weevil, *Ceutorhynchus obstrictus* (Marsham) (Coleoptera Curculionidae): Origins of North American introductions. Environmental Entomology 34**,** 504-510.

Leigh, J.W., Bryant, D., 2015. POPART: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6**,** 1110-1116.

Lesieur, V., Jeanneau, M., Martin, J., Bon, M., 2016. Development and characterization of 11 microsatellite markers in the root-gall-forming weevil, *Ceutorhynchus assimilis* (Coleoptera: Curculionidae). Applied Entomology and Zoology**,** 1-6.

Magri, D., Vendramin, G.G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., Latałowa, M., Litt, T., Paule, L., Roure, J.M., 2006. A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. New Phytologist 171**,** 199-221.

McInnis, M.L., Kiemnec, G.L., Larson, L.L., Carr, J., Sharratt, D., 2003. Heart-podded hoary cress. Rangelands Archives 25**,** 18-23.

Médail, F., Diadema, K., 2009. Glacial refugia influence plant diversity patterns in the Mediterranean Basin. Journal of Biogeography 36**,** 1333-1345.

Miller, R.F., Svejcar, T.J., Rose, J.A., McInnis, M.L., 1994. Plant development, water relations, and carbon allocation of heart-podded hoary cress. Agronomy Journal 86**,** 487-491.

529 Mulligan, G.A., Findlay, J.N., 1974. The biology of Canadian weeds. 3. *Cardaria draba*, *C.*
530 *chalepensis*, and *C. pubescens*. Canadian Journal of Plant Science 54**,** 149-160.
531 Nadel, R.L., Wingfield, M.J., Scholes, M.C., Lawson, S.A., Noack, A.E., Neser, S., Slippers, B., 2012.
532 Mitochondrial DNA diversity of *Cleruchoides noackae* (Hymenoptera: Mymaridae): a
533 potential biological control agent for *Thaumastocoris peregrinus* (Hemiptera:
534 Thaumastocoridae). Biocontrol 57**,** 397-404.
535 Papadopoulou, A., Anastasiou, I., Vogler, A.P., 2010. Revisiting the insect mitochondrial molecular
536 clock: the mid-Aegean trench calibration. Molecular Biology and Evolution 27**,** 1659-1672.
537 Petit, R.J., El Mousadik, A., Pons, O., 1998. Identifying populations for conservation on the basis of
538 genetic markers. Conservation biology 12**,** 844-855.
539 Phillips, C., Baird, D., Iline, I., McNeill, M., Proffitt, J., Goldson, S., Kean, J., 2008. East meets west:
540 adaptive evolution of an insect introduced for biological control. Journal of Applied Ecology
541 45**,** 948-956.
542 Posada, D., Crandall, K.A., 2001. Intraspecific gene genealogies: trees grafting into networks. Trends
543 in Ecology & Evolution 16**,** 37-45.
544 Puillandre, N., Lambert, A., Brouillet, S., Achaz, G., 2012. ABGD, Automatic Barcode Gap
545 Discovery for primary species delimitation. Molecular ecology 21**,** 1864-1877.
546 Puliafico, K.P., Schwarzländer, M., Price, W.J., Harmon, B.L., Hinz, H.L., 2011. Native and exotic
547 grass competition with invasive hoary cress (*Cardaria draba*). Invasive Plant Science and
548 Management 4**,** 38-49.
549 Ramos-Onsins, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against population
550 growth. Molecular biology and evolution 19**,** 2092-2100.
551 Rauth, S.J., Hinz, H.L., Gerber, E., Hufbauer, R.A., 2011. The benefits of pre-release population
552 genetics: a case study using *Ceutorhynchus scrobicollis*, a candidate agent of garlic mustard,
553 *Alliaria petiolata*. Biological Control 56**,** 67-75.
554 Rector, B.G., De Biase, A., Cristofaro, M., Primerano, S., Belvedere, S., Antonini, G., Sobhian, R.,
555 2010. DNA fingerprinting to improve data collection efficiency and yield in an open-field
556 host-specificity test of a weed biological control candidate. Invasive Plant Science and
557 Management 3**,** 429-439.
558 Rice, P., 2014. Invaders database system. http://invader.dbs.umt.edu/
559 Roy, L., Bon, M.C., Cesarini, C., Serin, J., Bonato, O., 2016. Pinpointing the level of isolation
560 between two cryptic species sharing the same microhabitat: a case study with a scarabaeid
561 species complex. Zoologica Scripta 45, 407-420.
562 Santos, H., Rousselet, J., Magnoux, E., Paiva, M.R., Branco, M., Kerdelhue, C., 2007. Genetic
563 isolation through time: allochronic differentiation of a phenologically atypical population of
564 the pine processionary moth. Proceedings of the Royal Society B-Biological Sciences 274**,**
565 935-941.
566 Schmitt, T., 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends.
567 Frontiers in zoology 4**,** 1.
568 Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and
569 phylogenetic utility of mitochondrial gene sequences and a compilation of conserved
570 polymerase chain reaction primers. Annals of the entomological Society of America 87**,** 651-
571 701.
572 Stepanović, S., Kosovac, A., Krstić, O., Jović, J., Toševski, I., 2015. Morphology versus DNA
573 barcoding: two sides of the same coin. A case study of *Ceutorhynchus erysimi* and *C.*
574 *contractus* identification. Insect science 23, 638-648.
575 Szucs, M., Melbourne, B.A., Tuff, T., Hufbauer, R.A., 2014. The roles of demography and genetics in
576 the early stages of colonization. Proceedings of the Royal Society B-Biological Sciences 281.
577 Szucs, M., Schaffner, U., Price, W.J., Schwarzlaender, M., 2012. Post-introduction evolution in the
578 biological control agent *Longitarsus jacobaeae* (Coleoptera: Chrysomelidae). Evolutionary
579 Applications 5**,** 858-868.
580 Tavares, C., Roques, A., Courtial, B., Branco, M., Jactel, H., Lopez-Vaamonde, C., 2015.
581 Phylogeography of the ladybird *Iberorhyzobius rondensis*, a potential biological control agent
582 of the invasive alien pine bast scale *Matsucoccus feytaudi*. Biocontrol 60**,** 59-69.

583 Tempère, G., Péricart, J., 1989. Faune de France, Coléoptères Curculionidae Quatrième partie.
584 Fédération Française des Sociétés de Sciences Naturelles, Paris, France.
585 USDA-NRCS, 2016. The plants database. United States Department of Agriculture, National
586 Resources Conservation Service.
587 Virag, A., Bon, M., Cloşca, C., Diaconu, A., Haye, T., Weiss, R., Müller-Schärer, H., Hinz, H., 2016.
588 Phenology and temperature-dependent development of *Ceutorhynchus assimilis*, a potential
589 biological control agent for *Lepidium draba*. Journal of Applied Entomology 141, 219-230.
590 Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A., 2013. A general species delimitation method with
591 applications to phylogenetic placements. Bioinformatics 29**,** 2869-2876.
592
593

Figures and Tables Caption

Fig 1. Bayesian phylogenetic tree from analysis of the *COI* sequences of *Ceutorhynchus assimilis*. Maximum likelihood bootstrap values (> 80%) are listed above the branches, followed by Bayesian posterior probabilities (> 0.75). Branch lengths represent expected substitutions per site. The scale bar indicates the expected number of substitutions per site. *Ceutorhynchus obstrictus* and *C. gallorhenanus* were used as outgroup. Roman numerals refer to estimated divergence times given in the Results section. Grey vertical bars represent the results of the species delimitation methods (*i.e*. ABGD (Puillandre et al., 2012) and PTP (Zhang et al., 2013) procedures).

Fig 2. *COI* mitochondrial haplotype network of *Ceutorhynchus assimilis*. Each circle corresponds to one haplotype; circle size gives the proportion of individuals belonging to the haplotype. The color inside each circle represents the host and indicates the proportion of individuals sampled in the different hosts. Small crossing lines indicate the number of mutations separating haplotypes. Black circles represent missing intermediate haplotypes.

Fig 3. Geographic distribution of the different lineages of *Ceutorhynchus assimilis*. Pie chart sizes are proportional to sample size.

Fig 4. Geographical distribution of the five SAMOVA derived groups identified within the 23 populations of the *Lepidium* host race of *Ceutorhynchus assimilis* in the western Mediterranean region. Color codes of the populations correspond to the five groups identified by SAMOVA.

Table 1. Sampling details and summary statistics of genetic diversity of the sampled *Ceutorhynchus assimilis* populations used in this study.

Table 2. Estimates of evolutionary divergence over the different genetic entities (*p*-distance, expressed as a %) and results of neutrality tests for the three lineages identified within *Ceutorhynchus assimilis*.