Current use of and future needs for soil invertebrate functional traits in community ecology

To cite this version:

HAL Id: hal-01005001
https://hal.archives-ouvertes.fr/hal-01005001
Submitted on 11 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Current use of and future needs for soil invertebrate functional traits in community ecology

Benjamin PEY\textsuperscript{a,b}, Johanne NAHMANI\textsuperscript{c}, Apolline AUCLERC\textsuperscript{d}, Yvan CAPOWIEZ\textsuperscript{e}, Daniel CLUZEAU\textsuperscript{f}, Jérôme CORTET\textsuperscript{g}, Thibaud DECAËNS\textsuperscript{h}, Louis DEHARVENG\textsuperscript{i}, Florence DUBS\textsuperscript{j}, Sophie JOIMEL\textsuperscript{k}, Charlène BRIARD\textsuperscript{l}, Fabien GRUMIAUX\textsuperscript{1}, Marie-Angélique LAPORTE\textsuperscript{m}, Alain PASQUET\textsuperscript{n}, Céline PELOSI\textsuperscript{a}, Céline PERNIN\textsuperscript{i}, Jean-François PONGE\textsuperscript{o}, Sandrine SALMON\textsuperscript{o}, Lucia SANTORUFO\textsuperscript{k,p}, Mickaël HEDDE\textsuperscript{*,a}

\textsuperscript{a}INRA, UR251 PESSAC, RD 10, 78026 Versailles Cedex, France
\textsuperscript{b}CESAB/FRB, Domaine du Petit Arbois, Avenue Louis Philibert, 13545 Aix-en-Provence, France
\textsuperscript{c}Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), CNRS, Université de Montpellier II, Montpellier, France
\textsuperscript{d}University of Michigan, Department of Ecology and Evolutionary Biology, Ann Arbor, Michigan, USA
\textsuperscript{e}INRA, UR1115 « Plantes et Systèmes Horticoles », Domaine Saint-Paul, 84914 Avignon Cedex 09, France
\textsuperscript{f}Université de Rennes 1, UMR CNRS 6553 « EcoBio », Station Biologique, 35380 Paimpont, France
\textsuperscript{g}Université Paul Valéry Montpellier III, Centre d'Ecologie Fonctionnelle et évolutive, Laboratoire de Zoogéographie, UMR 5175 CEFE, route de Mende, 34199 Montpellier cedex 5, France
\textsuperscript{h}UFR Sciences et Techniques, EA 1293 « ECODIV », Université de Rouen, 76821 Mont Saint Aignan Cedex, France
\textsuperscript{i}CNRS, UMR 7205, Muséum National d’Histoire Naturelle, CP50, 45 rue Buffon, 75005 Paris, France
IRD, UMR BIOEMCO, Centre France Nord, 93143 Bondy Cedex, France

INRA/INPL, UMR 1120 « Laboratoire Sols et Environnement », Nancy-Université, 2 avenue de la Forêt de Haye, BP 172, 54505 Vandœuvre-lès-Nancy Cedex, France

Université de Lille 1, EA 4515 « Laboratoire Génie Civil & géo Environnement », Lille Nord de France, Ecologie Numérique et Ecotoxicologie - Bat SN3, 59655 Villeneuve d'Ascq Cedex, France

IRD, UMR 228 ESPACE-DEV, 500 rue Jean-François Breton, 34093 Montpellier Cedex, France

UR AFPA, Faculté des Sciences et Technologies, Université de Lorraine, Boulevard des Aiguillettes, BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France

CNRS, UMR 7179, Muséum National d'Histoire Naturelle, 4 Avenue du Petit-Château, 91800 Brunoy, France

Department of Structural and Functional Biology, University of Naples Federico II, Complesso Universitario di Monte Sant'Angelo, Via Cinthia, 80126 Naples, Italy

* Corresponding author. Tel.: +33 (0)6 22 13 54 78; fax: +33 (0)3 83 59 57 91.

E-mail address: mickael.hedde@versailles.inra.fr
Abstract

Soil invertebrates are assumed to play a major role in ecosystem dynamics, since they are involved in soil functioning. Functional traits represent one of the main opportunities to bring new insights into the understanding of soil invertebrate responses to environmental changes. They are properties of individuals which govern their responses to their environment. As no clear conceptual overview of soil invertebrate trait definitions is available, we first stress that previously-described concepts of trait are applicable to soil invertebrate ecology after minor modification, as for instance the inclusion of behavioural traits. A decade of literature on the use of traits for assessing the effects of the environment on soil invertebrates is then reviewed. Trait-based approaches may improve the understanding of soil invertebrate responses to environmental changes as they help to establish relationships between environmental changes and soil invertebrates. Very many of the articles are dedicated to the effect of one kind of stress at limited spatial scales. Underlying mechanisms of assembly rules were sometimes assessed. The patterns described seemed to be similar to those described for other research fields (e.g. plants). The literature suggests that trait-based approaches have not been reliable over eco-regions. Nevertheless, current work gives some insights into which traits might be more useful than others to respond to a particular kind of environmental change. This review also highlights methodological advantages and drawbacks. First, trait-based approaches provide complementary information to taxonomic ones. However the literature does not allow us to differentiate between trait-based approaches and the use of a priori functional groups. It also reveals methodological shortcomings. For instance, the ambiguity of the trait names can impede data gathering, or the use of traits at a species level, which can hinder scientific interpretation as intra-specific variability is not taken into account and may lead to some biases. To overcome these shortcomings, the last part aims at proposing some solutions and
prospects. It concerns notably the development of a trait database and a thesaurus to improve
data management.

**Keywords:** behaviour, community ecology, constraint, database management system,
disturbance, ecological preference, life-history trait, soil fauna, thesaurus
Zusammenfassung

von Merkmalen auf der Artebene, der die wissenschaftliche Interpretation erschweren kann, da die intraspezifische Variabilität nicht berücksichtigt wird und zu gewissen Verzerrungen führen kann. Um diese Unzulänglichkeiten zu überwinden, hat der letzte Teil zum Ziel, einige Lösungen und Ausblicke vorzuschlagen. Dies betrifft namentlich die Entwicklung einer Merkmalsdatenbank und eines Thesaurus' um die Datenverwaltung zu verbessern.
Introduction

The current biodiversity estimation of soil fauna assumes that soil is the third biotic frontier after tropical forest canopies and ocean abysses (Swift, Heal & Anderson 1979; André, Noti & Lebrun 1994; Giller 1996; Wolters 2001). The soil fauna encompasses both the obligate and facultative inhabitants of soil and soil annexes (Wolters 2001). Soil annexes are simple structures which diversify the soil surface (e.g. tree stumps)(Gobat, Aragno & Matthey 1998). The soil includes a variety of animals from almost all major taxa that compose the terrestrial animal communities and may represent as one quarter of all currently described biodiversity (Decaëns, Jimenez, Gioia, Measey & Lavelle 2006). Soil invertebrates are assumed to play a major role in ecosystem dynamics, since they are involved in soil functioning (e.g. carbon transformation and sequestration, regulation of microbial activity or community structure, nutrient turnover, aggregation). Consequently, soil invertebrates contribute to the provision of many ecosystem services such as nutrient cycling or soil structure maintenance (Lavelle, Decaëns, Aubert, Barot, Blouin et al. 2006; Barrios 2007; Kibblewhite, Ritz & Swift 2008).

Studying soil invertebrate responses to environmental changes is of great interest. In various research fields (e.g. plant ecology), functional components of communities have revealed valuable insights into the understanding of organisms' responses to the environment (McGill, Enquist, Weiher & Westoby 2006; Garnier & Navas 2012). Originally, taxa were grouped into *a priori* functional groups based on certain “characteristics” which they shared. The classification into such functional groups is based on subjective expert judgment. For instance, several plant functional types existed, based on their life form or growth form (Lavorel, McIntyre, Landsberg & Forbes 1997). Conclusions were drawn from these *a priori* functional groups’ richness (Villéger, Mason & Mouillot 2008). However these approaches led to several limitations (Villéger et al. 2008) such as (i) a loss of information by imposing a discrete structure on functional differences between taxa, which are usually continuous (Gitay
& Noble 1997; Fonseca & Ganade 2001), (ii) a non-robust way of obtaining results depending on the choice of the functional group types in the analysis (Wright, Naeem, Hector, Lehman, Reich et al. 2006) and sometimes (iii) a failure to take account of abundance (Díaz & Cabido 2001). As an alternative to the taxonomic and a priori functional group approaches, trait-based approaches have been developed (Lavorel & Garnier 2002; McGill et al. 2006). Traits can be divided into response and effect traits. An effect trait is an individual property which affects an upper level of organization (e.g. ecosystem processes). Response traits, also called functional traits, are properties of individuals which govern their responses to their environment (Statzner, Hildrew & Resh 2001; Violle, Navas, Vile, Kazakou, Fortunel et al. 2007). In the following, traits will mean response traits. Unlike a priori functional groups, trait-based approaches are based on objective relations between individual properties (= traits) and the environment. In other research fields, notably for plants, trait-based approaches have brought several new insights to the understanding of organisms' responses to environmental changes, by improving predictability and reducing context dependence (Webb, Hoeting, Ames, Pyne & LeRoy Poff 2010; Garnier et al. 2012). Prediction involves that a relationship must be found between soil invertebrates and environmental changes through their traits. It has been demonstrated that community assembly mechanisms are governed by rules. The literature tends to support the existence of environmental filters which filter a sub-set of individuals of the regional pool to form local communities (Keddy 1992; McGill et al. 2006). Furthermore, environmental filters can be categorized according to the scale on which they work. From larger scales to smaller ones, filters are (i) dispersal filters which select individuals according to their dispersal capacity, (ii) abiotic filters which select individuals according to their capacity to live under certain abiotic conditions and (iii) biotic filters which represent the selection resulting from the interactions between individuals (Belyea & Lancaster 1999; Garnier et al. 2012). Reducing context dependency implies that trait-based
approaches have to be: (i) generic over eco-regions and (ii) reliable whatever kind of environmental change is considered. Enough trait-based approach studies have been made on plants to associate one or more traits with one or more environmental changes in any eco-region (Garnier et al. 2012). For instance, “leaf area” responds gradually to complex environmental change such as climate change over eco-regions (Thuiller, Lavorel, Midgley, Lavergne & Rebelo 2004; Moles, Warton, Warman, Swenson, Laffan et al. 2009).

To our knowledge, attempts to relate terrestrial invertebrate responses in terms of their “characteristics” to environmental stress began at the end of the nineteenth century (Statzner et al. 2001). In 1880, Semper (in Statzner et al. 2001) assessed the temperature-induced switch from parthenogenetic to sexual reproduction in aphids. During the following years, authors were convinced that environmental stress and “characteristics” of terrestrial insects were linked (Shelford 1913; Buxton 1923; Hesse 1924; Pearse 1926 - all in Statzner et al. 2001). For instance, Buxton (1923 - in Statzner et al. 2001) related “characteristics” of terrestrial insects such as the presence of wings or the tolerance of larvae to a lack of food and water to harsh environmental conditions of deserts (e.g. drought, torrential rain, whirlwinds).

Despite this early interest, no clear conceptual and methodological overview has been made for such “characteristics” of soil invertebrates, which are now called traits. Originally, as for plants, most previous studies assessed soil invertebrate responses to their environment using taxonomic structure and/or composition of communities. As soil invertebrate taxonomic diversity is huge, authors tried to simplify it by grouping together individuals by shared properties. The grouping also dealt with the lack of knowledge of taxonomy. For instance, eco-morphological groups, such as epigeic, anecic and endogeic groups of earthworms (Bouché 1972), epiedaphic, hemiedaphic and euedaphic groups of springtails (Gisin 1943) or terrestrial isopods (Schmallfuss 1984) and functional guilds such as the distinction between ecosystem engineers, litter transformers and micropredators (Lavelle & Spain 2001) were
used. For instance, eco-morphological groups bring together individuals based on subjective expert judgments of some of the ecological or biological “characteristics” they share. For instance, epigeic earthworms are pigmented and live near the soil surface, whereas endogeic earthworms are unpigmented and live deep in the soil. As for plants, all of these groupings have been used as *a priori* functional groups and should present the same disadvantages (see above). Experience in other research fields led us to think that using functional trait-based approaches for soil invertebrates represents one of the main opportunities to bring new insights into the understanding of soil invertebrate responses to the environment.

To our knowledge, no attempt has been made to clearly define functional trait concepts for soil invertebrates. The concept already existed but was used in other research fields. As a consequence, we first determine whether the actual definitions around the notion of traits are applicable to soil invertebrates. Second, to summarise the current advances in the understanding of soil invertebrate responses to the environment through their traits, a one-decade literature review was made. It also aimed to focus on current methodological advantages and drawbacks of soil invertebrate trait-based approaches. The last part envisages solutions and prospects for overcoming current conceptual and methodological drawbacks. It notably deals with the development of eco-informatics tools.

**Are existing trait definitions applicable to soil invertebrates?**

From work on terrestrial plants (Lavorel, Díaz, Cornelissen, Garnier, Harrison et al. 2007) or aquatic invertebrates (Bonada, Prat, Resh & Statzner 2006), traits are being defined as properties of organisms measured at the individual level (Violle et al. 2007). Furthermore, a trait is qualified as “functional” when it influences the organism’s performance and consequently its fitness (Southwood 1977; Nylin & Gotthard 1998; Blanck, Tedesco & Lamouroux 2007; Violle et al. 2007; Webb et al. 2010). Some authors distinguish the
performance traits from morphological, phenological and physiological traits (“M-P-P” traits).

Performance traits describe growth, reproduction and survival, considered as being the three components of fitness (Arnold 1983; McGill et al. 2006; Violle et al. 2007). Three main performance traits are recognized in plant ecology: vegetative biomass, reproductive output and measured plant survival (Violle et al. 2007). Conversely, “M-P-P” traits are supposed to influence fitness indirectly by influencing performance traits. In addition, plant abiotic preferences are denominated “Ellenberg’s numbers” and reflect optima/ranges in environmental gradients (Ellenberg 1988). In aquatic invertebrate ecology, traits are usually split into biological and ecological traits (Dolédec, Statzner & Bournard 1999). Biological traits include M-P-P and life-history traits, while ecological traits reflect behaviour and ecological optima/ranges in environmental gradients.

Regarding soil fauna, many functional traits considered in the literature are related to morphology, physiology or phenology (Ribera, Doledec, Downie & Foster 2001; Barbaro & van Halder 2009; Vandewalle, de Bello, Berg, Bolger, Dolédec et al. 2010; Pérès, Vandenbergulcke, Guernion, Hedde, Beguiristain et al. 2011) matching the definition proposed by Violle et al. (2007). The literature used, for instance, carabid beetle eye diameter or wing form for morphology, carabid beetle breeding season for phenology (Ribera et al. 2001; Vandewalle et al. 2010) or springtail reproductive mode for physiology (Malmstrom 2012). However, behaviour, such as “hunting strategy” (Langlands, Brennan, Framenau & Main 2011), is a crucial component in animal fitness that was not taken into account in Violle’s definition as the definition was stated for plants. For animals other than soil invertebrates, behaviour was semantically included (i) in a “biological traits” group, (ii) in an “ecological traits” group or (iii) in a semantically dedicated “behavioural traits” group (Relya 2001; Bonada, Dolédec & Statzner 2007; Frimpong & Angermeier 2010). Behaviour can be defined as an organized and directed biological response to variations in the environment to suit the
individual’s requirements (adapted from (Barnard 2004))). The environment refers both to the biocenosis and the biotope. We propose to extend Violle et al.’s (2007) definition of a functional trait for soil invertebrates as follows: “any morphological, physiological, phenological or behavioural (MPPB) feature measurable at the individual level, from the cell to the whole-organism level, without reference to any other level of organization” (Table 1). Furthermore, as for plants, we can distinguish MPPB traits from performance traits. The performance traits for soil invertebrates could be for instance: biomass, offspring output and measured survival. Population parameters can be derived from the median, mean and/or breadth of distribution of a trait (aggregated values of a MPPB or a performance trait, Table 1).

In addition, some of the functional traits used in the literature refer to properties of the environment in which individuals of a given species live. For instance, authors used the term “soil moisture preferences” (Makkonen, Berg, van Hal, Callaghan, Press et al. 2011) to express the breadth of the occurrence distribution of individuals of a species along a soil moisture gradient. We propose to call “ecological preference” any value which results from the optimum and/or the breadth of distribution of a trait along an environmental gradient (Table 1).

Finally, authors called “life-history traits” (Stearns 1992) or “life-cycle traits” a wide range of data such as moisture preference (Bokhorst, Phoenix, Bjerke, Callaghan, Huyer-Brugman et al. 2012), adult daily activity (Barbaro et al. 2009) or body size estimated for a species (Malmstrom 2012). Life-history traits need to be renamed, depending on their nature. In our examples, moisture preference will be classified as an “ecological preference”, while adult daily activity and body size estimated for a species are “population parameters derived from a trait”.

**Trait-based approaches for soil invertebrate community ecology**
Methods for literature review

A literature review was made from the ISI Web of Knowledge research platform using the search terms “trait” and “soil” or “ground” with each vernacular or taxonomic name of four groups: earthworms, ground beetles, spiders and springtails. The taxonomic groups were chosen because they represent a wide range of biological strategies and were often used as bio-indicators. Papers were selected according to several criteria described below. The term “trait” must have directly concerned soil invertebrates. To keep the scope of our study as restricted as possible, we only selected studies dealing with the effects of environmental changes on soil invertebrates. We did not include approaches exclusively dealing with other ecological questions or dedicated to evolutionary questions (e.g. adaptation, speciation). However, we are aware that ecological and evolutionary questions can overlap, notably when considering links between phylogeny and trait conservation (Cavender-Bares, Kozak, Fine & Kembel 2009). Reviews (with no original data) and methodological papers were excluded. Searches were limited to papers published since 2000 as the use of the term “trait” in soil invertebrate ecological studies is quite recent. We may have failed to find some papers as the word “trait” was not used in some papers even though a trait-based approach was used. This highlights the fact that the trait concept suffers from semantic inconsistency for soil invertebrates as stated in the previous section. However, we chose to look for literature on measurable criteria (as mentioned above), especially by using the search word “trait”, rather than on studies based on expert knowledge, even though this meant excluding a considerable number of papers. For instance, some studies using a trait-based approach have not been collected, e.g. for carabid beetles (Vanbergen, Woodcock, Koivula, Niemelä, Kotze et al. 2010), springtails (Ponge, Dubs, Gillet, Sousa & Lavelle 2006), earthworms (Jimenez, Decaëns & Rossi 2012), spiders (Lambeets, Vandegehuchte, Maelfait & Bonte 2008; Le Viol, Julliard, Kerbiriou, de Redon, Carnino et al. 2008; Lambeets, Vandegehuchte, Maelfait &
Bonte 2009; Cristofoli, Mahy, Kekenbosch & Lambeets 2010) and for multiple groups (Bell, Mead, Skirvin, Sunderland, Fenlon et al. 2008; Jennings & Pocock 2009; Moretti & Legg 2009; De Lange, Lahr, Van der Pol & Faber 2010; Hedde, van Oort & Lamy 2012). However, we are confident in the representativeness of the literature review, which found 4, 17, 4 and 6 papers for earthworms, ground beetles, spiders and springtails respectively (Table 2).

**Scientific advances and drawbacks**

All the literature showed, as for other research fields, that some environmental filters filter a sub-set of individuals from a regional pool to form local communities according to some of their traits. Most of the studies were dedicated to assess soil invertebrate response to some kind of stress (Table 2). For instance, Barbaro et al. (2009), Driscoll et al. (2005) and Ribera et al. (2001) assessed mechanisms of carabid beetle responses to habitat types according to their traits (e.g. body size, wing development, Table 2). Underlying mechanisms of assembly rules were sometimes assessed. For instance, Decaëns et al. (2008) demonstrated that some abiotic environmental filters led to a trait convergence for earthworms. Decaëns et al. (2011) revealed that the variability of morphological earthworm traits was lower in the regional species pool and higher in the local species pool compared to what would have been expected by chance. As very few examples were given, such patterns cannot be used as general patterns for soil invertebrate assembly rules. However, the patterns described seemed to be similar to those described in the introduction for other research fields. These results claimed that soil invertebrate trait-based approaches help to improve predictability of community assembly in relation to environmental changes as they materialise relationships between traits and environmental changes.

Almost all of the studies assessed the responses of soil invertebrates in relation to only one kind of environmental change. Some exceptions were found. For instance, Gobbi et al. (2010) aimed to assess both the abiotic effect of deglaciation and the biotic effect of plant
communities on carabid beetle communities. While individual studies usually dealt with a single change, environmental changes studied were diverse among studies. They included "natural" changes such as habitat type, fire, flooding or climatic events and also "anthropic" changes such as invasive tree species or human practices on cultivated fields or forests (Table 2). In addition, studies were geographically limited to the regional scale (sensu Belyea et al. 1999). Some exceptions occurred, e.g. Vandewalle et al. (2010) who sampled carabid beetles in several European countries. They assumed that the responses of functional diversity indices calculated from traits (e.g. Rao index of diversity, Botta-Dukat 2005) to habitat composition and landscape heterogeneity were consistent across geographical regions.

To conclude, we cannot be confident in trait genericity over eco-regions, as this was rarely studied (Vandewalle et al. 2010). Despite these shortcomings in reducing the context dependence, the literature currently gives us some insights as to which traits might be more useful than others to respond to a particular kind of environmental change. For instance, it has been shown that ground beetle wing development varies with habitat type in different contexts (Ribera et al. 2001; Driscoll & Weir 2005; Gobbi & Fontaneto 2008; Barbaro et al. 2009; Gobbi, Caccianiga, Cerabolini, Bernardi, Luzzaro et al. 2010; Vandewalle et al. 2010). To make the trait-based approaches reliable whatever the kind of environmental changes, we have to establish relationships between each kind of environmental change with one or several traits.

Methodological advantages and drawbacks

Complementarity with other approaches

From a methodological point of view, trait-based approaches bring new insights into the understanding of soil invertebrate responses to stress, compared to taxonomic approaches (Cole, McCracken, Dennis, Downie, Griffin et al. 2002; Gobbi et al. 2008; Langlands et al. 2011). First, inverse trends between results obtained by trait-based and taxonomic approaches
were reported. For example, Gerisch, Agostinelli, Henle & Dziock (2012) showed that the species diversity of ground beetle communities increased whereas functional diversity (functional evenness and divergence) decreased with increasing flooding disturbances. This combined approach led the authors to conclude that flooding disturbance increased the number of species but that species were functionally redundant. Otherwise, Gobbi et al. (2008) showed that ground beetle traits such as wing morphology, diet and body size responded to habitat diversity, while species richness and a taxonomic diversity index based on phylogeny did not. The authors therefore claimed that trait-based approaches should be favoured for assessing mechanisms of carabid beetle responses to habitat disturbance rather than taxonomic approaches. In other cases, trait-based approaches complemented the conclusions based on taxonomic approaches. For instance, in a study by Fournier, Samaritani, Shrestha, Mitchell & Le-Bayon (2012), community-weighted means of earthworm traits (e.g. body length and width, pH optimum and range) were more strongly correlated with environmental variables (e.g. total carbon, gravel sizes, type of cover, such as mosses, woody debris) than species composition and taxonomic diversity. However, no study aimed at comparing approaches based on a priori functional groups (e.g. eco-morphological groups) with trait-based approaches.

**Deficiencies in trait definitions, data treatment and gathering structure**

The literature review revealed semantic inconsistencies for trait names. For instance, the type of materials eaten by soil invertebrates (e.g. carnivorous) and the way they feed on them (e.g. as predators, i.e. by killing their preys). However, the literature revealed several categorical traits whose attributes could describe several of the above concepts simultaneously. For instance, “food of the adult” (Cole et al. 2002; Ribera et al. 2001) referred both to the type of food eaten (e.g. plant, springtails) but also to the way it was eaten (e.g. generalist predators) whereas “diet” (Barbaro et al. 2009) refers only to the first one. Such drawbacks occurred
They can hinder data gathering in so far as they can cast doubt on a trait’s scientific meaning.

At the moment, soil invertebrate trait-based approaches used traits at the species level. Such a process can lead to two main biases. A first bias occurs when the trend of the relationship between the mean trait of N species and an environmental gradient is in the opposite direction to the relationships between this environmental gradient and individual trait values. The second bias is that using traits at the species level hides individual heterogeneity.

Traits can be described in two formats, numerical data (e.g. eye diameter, (Ribera et al. 2001)) or by text (e.g. pigmentation, wing form, (Vandewalle et al. 2010)). Format heterogeneity and the missing data impeded the use of traits. It has been suggested that traits should be encoded into a limited number of subsets (Chevenet, Dolédec & Chessel 1994; Hedde et al. 2012). For all of these reasons, some authors discretized data into attributes, e.g. by fuzzy coding procedures (e.g. body size classes, (Jelaska, Jesovnik, Jelaska, Pirnat, Kucinic et al. 2010) or diet, (Pérès et al. 2011). When working on one or several taxonomic groups, it was crucial to be able to deal with different data formats. However when this was done, the way data were transformed by fuzzy coding was not clearly explained. This impedes the comparison between studies using a trait shared by one or several groups but not necessarily using the same coding procedure (e.g. different categories for the diet) (Barbaro et al. 2009; Gerisch 2011). It also limits the reuse of an encoded trait from the literature as readers do not know exactly how the trait was encoded.

Exploiting existing literature was preferred to time-consuming trait measurements on sampled specimens. Whatever the methodology, the review of literature underlined the lack of a data-compilation structure for soil invertebrate traits. Depending on the author, a trait could be described from different literature sources. Cole et al. (2002) and Karen, O'Halloran, Breen, Giller, Pithon et al. (2008) described body size trait values for *Nebria brevicollis* (Fabricius)
from two different literature sources. As a consequence, works do not benefit each other as no
data-compilation allows authors to have access on existing trait data.

A general shortcoming which is not often considered in the current literature is the fact that
traits used in a study can be inter-correlated (“trait syndromes”) (Poff, Olden, Vieira, Finn,
Simmons et al. 2006). Inter-correlation can therefore cause that traits appear decoupled from
environmental changes (Statzner, Dolédec & Hugueny 2004; Poff et al. 2006). Generally, trait
selection for analyses was *a priori* justified on the basis of the biological function they are
supposed to be linked with. For instance, (Langlands et al. 2011) selected the body shape of
spiders, as spiders with flattened bodies are supposed to shelter better from fire. Apart from
this view, no analysis has been described to identify “trait syndromes” before performing
linking traits to environmental variables. Exception was made for certain studies (Gobbi et al.
2008).

**Future needs: eco-informatics at a crossroad**

The following prospects are not limited to the four taxa used in the literature search. They are
suitable for all the soil invertebrate taxa. Large amounts of data from multiple data sources
need to be characterized and integrated into a unified corpus in order to improve soil
invertebrate trait-based approaches. Current eco-informatics literature provides a basis for a
global scheme to structure ecological data (Madin, Bowers, Schildhauer, Krivov, Pennington
et al. 2007; Garnier et al. 2012). Between non-robust data storage by scientists (*e.g.*
spreadsheets, relational database systems) (Jones, Schildhauer, Reichman & Bowers 2006)
and their exploitation by software tools (*e.g.* “R Statistical Package”) (R Development Core
Team 2010), an intermediate level is needed. It requires linking data with metadata, which are
information used to document and interpret data (Jones et al. 2006). Such a level would
greatly enhance data management (storage, integrating, querying, and analysing) by
producing robust traceability. One way is to construct a database management system
(DBMS) for soil invertebrate traits which could associate metadata with data. First are
“scientific” metadata describing scientific data (*e.g.* information usually provided in the
Materials and methods section). Scientific metadata provide all the necessary information for
acquiring, interpreting and using scientific data. Second are “computer” metadata required for
computerisation (*e.g.* metadata required for the database structure, semantic metadata). They
principally allow acquisition and automated input, analysis and processing of scientific data
by the computer (Michener 1997; Michener 2006). Associating data to metadata in a DBMS
provides several advantages. Data longevity (data history) and quality (control of the nature of
data) are increased. Data could be easily reused and integrated. Finally data sharing is
facilitated (Jones et al. 2006; Michener 2006). DBMS *per se* possesses sorting, indexing and
querying functions which increase data interpretation and use (Porter 1998). A few databases
for soil invertebrates already exist: for instance, Edaphobase (Russell, Vorwald, Franzke,
Höfer, Horak et al. 2012), Coltrait (Salmon & Ponge 2012), the Dutch soil invertebrate trait
database (from M.P. Berg) (Makkonen et al. 2011), Macrofauna (Lapied, personal
communication), and Ant Profiler (Bertelsmeier, Luque, Confais & Courchamp 2012).
Nevertheless, they do not always contain trait data or are not always in a format which allows
collaborative data sharing. Even if they fulfil such criteria, they tend to be concerned with a
small part of the whole diversity of soil invertebrates (usually a single group is concerned).
Computer science solutions currently exist to gather data from different sources (Jones et al.
2006; Michener 2006), so previous soil invertebrate databases should not be seen as isolated
islands (Jones et al. 2006) but as complementary bricks which can be combined to create new
soil invertebrate trait databases. However, combining data from different formats, especially
from spreadsheets, is not easy (Jones et al. 2006).
Among the existing solutions, semantic data integration is a promising way which preserves
the scientific meaning of data. Semantic approaches deal with the differences in the terms
used (terminology) and the scientific concepts formulated by soil invertebrate experts over
time (Madin et al. 2007; Laporte, Mougenot & Garnier 2012). To achieve this, the soil
invertebrate scientific community is required to standardize meaningful and precise terms that
cover their domain of interest. Trait names are especially concerned, taking a central position
in trait-based approaches in the context of the responses of soil invertebrates to their
environment. A thesaurus of a particular domain reflects a community agreement on a set of
terms established in a given area and its organization through a well-designed structure.
Furthermore, a thesaurus is recognized as a knowledge organization system and bypasses
ambiguity issues in natural language, controlling and clarifying the access and exchange of
information and facilitating communication. The main concern focuses on access, sharing and
dissemination of information within the soil invertebrate scientific community. First, a soil
invertebrate trait thesaurus can serve as a stable reference resource, specifically when
published in RDF (Resource Description Framework) language (Manola & Miller 2004) and
available as linked data on the web. A second prospect is to include such a thesaurus in soil
invertebrate trait databases to facilitate data management. A third, more long-term prospect,
involves the use of the thesaurus as a prerequisite for the construction of a soil invertebrate
trait ontology. To conclude, it would be of major assistance for the soil invertebrate scientist
community to have access to knowledge-based models enabling the efficient answering of
questions, which, for example, may require the data aggregation of different traits from
several taxa.

Effort on data management using eco-informatics tools will fill some gaps revealed by the
literature review. First, it will strengthen current scientific advances. By increasing the
collection of trait data and associated environmental parameters, it will offer the possibility of
considering the actions of several environmental filters on different spatial and temporal
scales (see section “Scientific advances and drawbacks”). It will also aim to establish
consistent “population parameters derived from traits” and “ecological preferences” (Table 1) by increasing the number of literature sources informing trait values used to calculate them. All of this will contribute to a better general understanding of soil invertebrate responses to the environment from local to biogeographical scales, which was not always possible from independent single studies. The data gathering structure should also improve knowledge of soil invertebrate group interactions, since it will become possible to work on several groups and taxa with several comparable traits.

Second, it will help with some methodological shortcomings. It will improve the possibility of dealing with (i) inter-correlation of traits and (ii) bias when using traits on the species level (see section “Deficiencies in trait definitions, data treatment and gathering structure”). On the one hand (i), “trait syndromes” could be more easily revealed because the data gathering structure should provide a large body of available documented traits. We recommend testing for inter-correlation of traits before drawing conclusions (*e.g.* fuzzy correspondence analysis, “ade4” R package, (Chessel, Dufour & Thioulouse 2004)). One other solution which has not been tested for soil invertebrates since not enough trait data have yet been gathered, is the screening method (Bernhardt-Römermann, Römermann, Nuske, Parth, Klotz et al. 2008). This allows the best combination of traits to be found for an environmental change. On the other hand (ii), with the increasing number of trait values measured on individuals rather than compiled at species or higher taxonomic level, it will provide the opportunity to put much more intraspecific variability into the assessment of functional diversity. It is a way to overrule bias when using traits at a species level.

Although the data gathering structure will enable the collection of data documenting traits from all sources (*e.g.* articles, books) and from all formats, *i.e.* numerical data (*e.g.* body size distribution) and literal data (*e.g.* text descriptions of diets), it will not deal with the definition of similar fuzzy coding protocols (see section “Deficiencies in trait definitions, data treatment
and gathering structure”). For instance, we propose two main protocols: one for traits
described by numerical values and another for traits described by textual data (see Appendix
A).

Acknowledgements
The authors wish to thank the CESAB (Centre de Synthèse et d’Analyses sur la Biodiversité)
and the FRB (Fondation pour la Recherche sur la Biodiversité) for their financial support. We
also thank Baptiste Laporte (FRB/CESAB) for his advice on computer science. Finally, we
kindly thank M. Berg and three anonymous reviewers for greatly improving the scientific
content of the manuscript through their comments.
## Tables

Table 1. Definitions of trait concepts for soil invertebrates.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPPB trait</strong></td>
<td>Any morphological, physiological, phenological or behavioural (MPPB) feature measurable at the individual level, from the cell to the whole-organism level, without reference to any other level of organization</td>
</tr>
<tr>
<td><strong>Performance trait</strong></td>
<td>Performance traits describe growth, reproduction and survival, considered as being the three components of fitness (Violle et al. 2007). For soil invertebrates there are for instance: biomass, offspring output and survival.</td>
</tr>
<tr>
<td><strong>Ecological preference</strong></td>
<td>The optimum and/or the breadth of distribution of a trait on an environmental gradient.</td>
</tr>
<tr>
<td><strong>Population parameters derived from traits</strong></td>
<td>The median, mean and/or breadth of distribution of a trait (aggregated values of a MPPB or a performance trait).</td>
</tr>
</tbody>
</table>
Table 2. Results of the literature review and some of the properties of the selected articles. LIT: trait data from the literature, OMS: original measurements of traits. Without any specific information, we assumed that trait data had been derived from the literature.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Soil invertebrate group</th>
<th>Environmental change</th>
<th>LIT or OMS</th>
<th>Traits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decaëns et al. (2011)</td>
<td>Earthworms</td>
<td>Habitat (different aged pastures)</td>
<td>LIT</td>
<td>(Ecological category), body length, diameter, weight, epithelium type, pigmentation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Size, weight, pigmentation, (ecological categories, ecological features)</td>
</tr>
<tr>
<td>Decaëns et al. (2008)</td>
<td>Earthworms</td>
<td>Habitat</td>
<td>LIT</td>
<td>Length, width, weight, segment number, pH optimum, pH range, prostomium type, (ecological type), C/N (soil) preference</td>
</tr>
<tr>
<td>Fournier et al. (2012)</td>
<td>Earthworms</td>
<td>Flooding</td>
<td>LIT + OMS</td>
<td>Body pigmentation, body wall thickness, food, reproductive strategy, rarity</td>
</tr>
<tr>
<td>Pérès et al. (2011)</td>
<td>Earthworms</td>
<td>Contamination and land use</td>
<td>LIT</td>
<td>Mean size, local activity-density, niche breath, ballooning, seasonal activity, generation time, diurnal activity</td>
</tr>
<tr>
<td>Bonte et al. (2006)</td>
<td>Spiders</td>
<td>Sand dynamics</td>
<td>LIT + OMS</td>
<td>Burrowing, body size (length), cephalothorax heavy sclerotisation, abdominal scutes, ballooning, time to maturity, phenology, hunting strategy, diet specialization (ants), flattened body</td>
</tr>
<tr>
<td>Buchholz (2010)</td>
<td>Spiders</td>
<td>Climate (drought)</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Langlands et al. (2011)</td>
<td>Spiders</td>
<td>Fire</td>
<td>LIT + OMS</td>
<td>(Life form), biomass, body length, moisture preference, vertical stratification</td>
</tr>
<tr>
<td>Tropek et al. (2008)</td>
<td>Spiders</td>
<td>Stone quarry</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Bokhorst et al. (2012)</td>
<td>Springtails</td>
<td>Climate (winter warming)</td>
<td>LIT + OMS</td>
<td></td>
</tr>
</tbody>
</table>
Huebner et al. (2012) Springtails Fire LIT
Dente shape, eye number, total body length, furcula, pigmentation, body scales, PAO, antennae length, antennal organ, sexual dimorphism

Lindberg & Bengtsson (2005) Springtails Climate (drought) LIT + OMS
Depth distribution, reproductive mode, habitat specialization, (ecological category)
Ocelli number, body size, body pigmentation level, body pigmentation pattern, modified hairs or scales, furca development, antenna/body, moisture preference, habitat width

Makkonen et al. (2011) Springtails Climate LIT
Ocelli number, body size, body pigmentation level, body pigmentation pattern, modified hairs or scales, furca development, antenna/body, moisture preference, habitat width

Malmstrom (2012) Springtails Fire LIT + OMS
Habitat (vertical stratification), body size, reproductive mode, dispersal traits

Vandewalle et al. (2010) Springtails Invasive tree species LIT
Ocelli, antenna length, furca, hairs/scales, pigmentation

<table>
<thead>
<tr>
<th>Author(s) (Year)</th>
<th>Taxa</th>
<th>Habitat (Type)</th>
<th>Literature Type</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbaro et al. (2009)</td>
<td>Ground beetles</td>
<td>Habitat (fragmentation)</td>
<td>LIT</td>
<td>European trend, European rarity, regional rarity, biogeographic position, daily activity, diet, overwintering, body colour, breeding season, body size (mm), wing development, adult activity period</td>
</tr>
<tr>
<td>Cole et al. (2002)</td>
<td>Ground beetles</td>
<td>Habitat (agricultural management)</td>
<td>LIT</td>
<td>Size (length), overwintering, life cycle duration, adult food, daily activity, breeding season, emergence, main activity, wing morphology, locomotion</td>
</tr>
<tr>
<td>Driscoll et al. (2005)</td>
<td>Ground beetles</td>
<td>Habitat (fragmentation)</td>
<td>LIT</td>
<td>Flight, trophic group, adult primary position, size</td>
</tr>
<tr>
<td>Gerisch et al. (2012)</td>
<td>Ground beetles</td>
<td>Flooding</td>
<td>LIT</td>
<td>Wing morphology, overwintering strategy (reproduction season), body size</td>
</tr>
<tr>
<td>Gerisch (2011)</td>
<td>Ground beetles</td>
<td>Flooding</td>
<td>LIT</td>
<td>Body size, wing morphology, reproduction period, overwintering stage, daily activity, colour elytra,</td>
</tr>
<tr>
<td>Study</td>
<td>Life stage</td>
<td>Habitat Type</td>
<td>Methodology</td>
<td>Life stage Notes</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Gobbi et al. (2010)</td>
<td>Ground beetles</td>
<td>Deglaciated terrain and plants</td>
<td>/</td>
<td>Brachypterous, autumn-breeding, predators, average body length</td>
</tr>
<tr>
<td>Gobbi et al. (2008)</td>
<td>Ground beetles</td>
<td>Habitat</td>
<td>LIT</td>
<td>Wing morphology, body length, diet</td>
</tr>
<tr>
<td>Grimbacher &amp; Stork (2009)</td>
<td>Ground beetles</td>
<td>Climate (seasonality)</td>
<td>LIT + OMS</td>
<td>Feeding ecology, body size, diet, habitat strata, mean period of activity</td>
</tr>
<tr>
<td>Jelaska et al. (2010)</td>
<td>Ground beetles</td>
<td>Habitat (natural temperate forests)</td>
<td>LIT + OMS</td>
<td>Body size</td>
</tr>
<tr>
<td>Karen et al. (2008)</td>
<td>Ground beetles</td>
<td>Habitat (forest cycle plantation)</td>
<td>LIT</td>
<td>Broad habitat associations, body size, wing-type, microhabitat associations</td>
</tr>
<tr>
<td>Liu et al. (2012)</td>
<td>Ground beetles</td>
<td>Habitat (human practices on semi-natural habitats and cultivated fields)</td>
<td>LIT &amp; OMS</td>
<td>Trophic status, body size</td>
</tr>
<tr>
<td>Ribera et al. (2001)</td>
<td>Ground beetles</td>
<td>Habitat (land disturbance)</td>
<td>LIT + OMS</td>
<td>Eye diameter, antenna length, pronotum maximum width, pronotum maximum depth,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>elytra maximum width, metatarsus length, metatarsus length, eye diameter,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>pronotum shape, overwintering, adult food, daily activity, breeding season,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>main period of adult emergence, main period of adult activity</td>
</tr>
<tr>
<td>Silva et al. (2011)</td>
<td>Ground beetles</td>
<td>Habitats (orchard and riparian)</td>
<td>LIT</td>
<td>Moisture preferences</td>
</tr>
<tr>
<td>Tropek et al. (2008)</td>
<td>Ground beetles</td>
<td>Stone quarry</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>Vandewalle et al. (2010)</td>
<td>Ground beetles</td>
<td>Habitat (composition and selection)</td>
<td>LIT</td>
<td>Wing form, body pubescence, body length, elytra width, elytra</td>
</tr>
<tr>
<td>Reference</td>
<td>Study</td>
<td>Habitat (removal of topsoil on former agricultural fields)</td>
<td>Literature (LIT)</td>
<td>Habitat preference (characterization and amplitude), dispersal capacity (flying), occurrence, size</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------</td>
<td>------------------------------------------------------------</td>
<td>-------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Verhagen et al. (2008)</td>
<td>Ground beetles</td>
<td>landscape heterogeneity</td>
<td>LIT</td>
<td>Habitat preference (characterization and amplitude), dispersal capacity (flying), occurrence, size</td>
</tr>
<tr>
<td>Warnaffe &amp; Dufrene (2004)</td>
<td>Ground beetles</td>
<td>Habitat (forest management)</td>
<td>LIT</td>
<td>Habitat preference (characterization and amplitude), dispersal capacity (flying), occurrence, size</td>
</tr>
</tbody>
</table>
References


Appendix A. Supplementary data

The heterogeneity of data informing trait (numeric, textual), the missing data and finally the heterogeneity of individuals within and between taxa led authors to transform trait raw data to (i) allow their comparison and (ii) reduce continuous data into a limited number of subsets. The fuzzy coding is one of the techniques which allowed the synthesis of data informing trait from different formats by their discretization into trait sub-classes called “categories” (Chevenet, Dolédec & Chessel 1994).

In soil invertebrate ecology, the main aim of the fuzzy coding is to discretize data informing a trait for a taxon into categories. Category choice is \textit{a priori} done and must be though out from sound biological hypotheses and from the accuracy of data informing trait. For instance, the “integument sclerotization” trait could be divided into 2 categories: unsclerotized or sclerotized. The thesaurus described in our review can serve as a stable semantic reference which fixes both trait names but also their category number and names (also category name synonyms).

Once trait categories were defined, the information for a taxon obtained from a literature source must be coded for each trait category by an affinity score. For instance, affinity scores can range from 0 to 3, from no to very high affinity of the taxon to a trait category. Then, affinity scores of several sources of a trait were converted to a percentage for building a trait profile for the considered taxon.

The way an affinity score is attributed to the categories was generally not clearly explained in the literature review. That is a point we want to fix in this review by proposing detailed protocols depending on the data format informing the trait.


Trait informed by numerical data

Literature sources commonly proposed one or several values for a trait. We propose to only use the minimum and the maximum values. Then, affinity scores are attributed to each *a priori* category. For each category, it is done according to “the space of the interval between the maximum and the minimum category values” occupied by “the interval between the maximum and the minimum trait values”. For this, each category interval was divided into three sub-categories:

- \([a ; a + (b-a)*1/3 [\) which represents the first 33\% of the category interval
- \([a + (b-a)*1/3 ; a + (b-a)*2/3 [\) which represents the middle 33\% of the category interval
- \([a + (b-a)*2/3 ; b [\) which represents the last 33\% of the category interval

where \(a\) and \(b\) are the minimum and the maximum values of the category.

Then if the interval between the maximum and the minimum trait values is included into

- No sub-category, then the affinity will be 0.
- One sub-category, then the affinity score will be 1.
- Two sub-categories, then the affinity score will be 2.
- Three sub-categories, then the affinity score will be 3.

Then, affinities scores of several sources for a trait were converted to percentages building a trait profile for the considered taxon.

Example:
We took the maximum and minimum body length of *Lumbricus castaneus* from different literature sources (Table 1).

<table>
<thead>
<tr>
<th>Source</th>
<th>Minimum body length (mm)</th>
<th>Maximum body length (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bouché (1972)</td>
<td>25</td>
<td>60</td>
</tr>
<tr>
<td>Sims &amp; Gerard (1985)</td>
<td>30</td>
<td>45</td>
</tr>
</tbody>
</table>

We proposed to divide the body length trait into 6 categories: [20;35[, [35;50[, [50;100[, [100;150[, [200;400[. Then we attribute the affinity scores by category following the rules above described (Table 2). For instance, the Bouché’s interval (25-60) is not included into the first sub-category [20;25[ but is included into the two other sub-categories [25;30[ and [30;35[ of the first category [20;35[. As a consequence, affinity score for the first category for Bouché is 2. The second category [35;50[ is completely filled by the Bouché’s interval (25-60). That is why the affinity score is 3.

<table>
<thead>
<tr>
<th>Source</th>
<th>[20;35[</th>
<th>[35;50[</th>
<th>[50;100[</th>
<th>[100;150[</th>
<th>[200;400[</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bouché (1972)</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sims et al. (1985)</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Finally, affinity scores of several sources were converted to percentages to create a trait profile (Table 3). For instance, for the first category, the percentage is calculated as being the ratio between the sum of affinity scores for this category (2+1) and the sum of all the affinity scores (10).

<table>
<thead>
<tr>
<th>Taxon</th>
<th>[20;35[</th>
<th>[35;50[</th>
<th>[50;100[</th>
<th>[100;150[</th>
<th>[200;400[</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Lumbricus castaneus</em></td>
<td>0.3</td>
<td>0.6</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Trait informed by textual data**

The text informing a trait generally contains two kinds of information: (i) some words which correspond to category names or their synonyms and (ii) adjectives which qualify categories.
For instance, the diet description for *Amara aenea* from (Brandmayr 1972) “Italy, in the field: mostly seeds” contains the word “seed” which corresponds to a category of the “diet” trait and “mostly”, an adverb. An affinity score must be attributed to each *a priori* category. We propose to do this according to (i) the number of categories embodied in a given literature source and (ii) the strength of adverbs characterizing each category. To do this, we first build a correspondence table (Table 4) between the frequently found adverbs in the literature and their strength.

<table>
<thead>
<tr>
<th>Strong</th>
<th>Medium</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolutely</td>
<td>almost</td>
<td>absent</td>
</tr>
<tr>
<td>almost entirely</td>
<td>common(ly)</td>
<td>accidentally</td>
</tr>
<tr>
<td>almost exclusively</td>
<td>frequently</td>
<td>episodically</td>
</tr>
<tr>
<td>always</td>
<td>generally</td>
<td>exceptionally</td>
</tr>
<tr>
<td>constantly</td>
<td>moderately</td>
<td>few</td>
</tr>
<tr>
<td>continually</td>
<td>moderately</td>
<td>infrequently</td>
</tr>
<tr>
<td>decidedly</td>
<td>more (in, during)</td>
<td>less frequently</td>
</tr>
<tr>
<td>definitely</td>
<td>mostly</td>
<td>maybe</td>
</tr>
<tr>
<td>entirely</td>
<td>often frequent</td>
<td>mentioned</td>
</tr>
<tr>
<td>especially</td>
<td>particularly numerous</td>
<td>missing</td>
</tr>
<tr>
<td>essentially</td>
<td>preferably</td>
<td>more or less</td>
</tr>
<tr>
<td>exceptionally</td>
<td>quite often</td>
<td>never</td>
</tr>
<tr>
<td>fully</td>
<td>regularly</td>
<td>occasionally</td>
</tr>
<tr>
<td>mainly</td>
<td>to like</td>
<td>optionally</td>
</tr>
<tr>
<td>majority</td>
<td>to prefer</td>
<td>partially</td>
</tr>
<tr>
<td>maximum</td>
<td>to require</td>
<td>possibly</td>
</tr>
<tr>
<td>numerically dominant in</td>
<td>usually</td>
<td>presumed to be</td>
</tr>
<tr>
<td>overwhelmingly</td>
<td>well known</td>
<td>probably</td>
</tr>
<tr>
<td>particularly</td>
<td>rarely</td>
<td></td>
</tr>
<tr>
<td>persistently</td>
<td>scarcely</td>
<td></td>
</tr>
<tr>
<td>predilection for</td>
<td>slightly</td>
<td></td>
</tr>
<tr>
<td>predominantly</td>
<td>sometimes</td>
<td></td>
</tr>
<tr>
<td>purely</td>
<td>sometimes</td>
<td></td>
</tr>
<tr>
<td>significantly</td>
<td>somewhat</td>
<td></td>
</tr>
<tr>
<td>specialist</td>
<td>sporadically</td>
<td></td>
</tr>
<tr>
<td>specially</td>
<td>to seem to</td>
<td></td>
</tr>
<tr>
<td>specifically</td>
<td>to tolerate</td>
<td></td>
</tr>
<tr>
<td>strictly</td>
<td>transiently</td>
<td></td>
</tr>
<tr>
<td>strong</td>
<td>only</td>
<td></td>
</tr>
<tr>
<td>strongly favoured</td>
<td></td>
<td></td>
</tr>
<tr>
<td>strongly influenced by</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Then, we develop rules to determine affinity scores:

- If there is no adverb:
  - and words of a given literature source correspond to more (>2) categories, then the category will receive an affinity score of 2
  - and words of a given literature source corresponds to 1 or 2 categories, then the category will receive an affinity score of 3

- If there are adverbs whose:
  - strength is “weak”, then the category will receive an affinity score of 1
  - strength is “medium”, then the category will receive an affinity score of 3
  - strength is “strong”
    - and words of a given literature source correspond to more (>2) categories, then the category will receive an affinity score of 3
    - and words of a given literature source correspond to 1 or 2 categories, then the category will receive an affinity score of 4

Example:
The diet of *Loricera pilicornis* is described in different literature sources (Table 5).

<table>
<thead>
<tr>
<th>Source</th>
<th>Diet description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies (1953)</td>
<td>England, in the field: food of adults: collembola, spiders and pollen</td>
</tr>
<tr>
<td>Ribera, Doledec, Downie &amp; Foster (2001)</td>
<td>Mostly collembola</td>
</tr>
</tbody>
</table>
We proposed to divide the diet trait into 5 categories: detritivore, microbivore-microphytophagous, phytophagous, geophagous, zoophagous. Then we attribute the affinity scores by category following the rules above described (Table 6). In the first source, the sentence contains the words corresponding to 2 categories: “collembolan” and “spiders” for the zoophagous category and “pollen” for the phytophagous category. As there are no adverb, affinity scores are 3 for both categories. In the second source, the sentence contains a word “collembola” corresponding to the zoophagous category. A “strong” adjective is present. The affinity score is therefore 4.

Table 6. Affinity scores for *Loricera pilicornis* from two literature sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Detritivore</th>
<th>Microbivore</th>
<th>Phytophagous</th>
<th>Geophagous</th>
<th>Zoophagous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Davies (1953)</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Ribera et al. (2001)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Finally, affinity scores of several sources were converted to percentages to create a trait profile (Table 7). For instance, for the zoophagous category, the percentage is calculated as the ratio between the sum of affinity scores for this category (3+4) and the sum of all the affinity scores (10).

Table 7. Trait profile for *Lumbricus castaneus*

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Detritivore</th>
<th>Microbivore</th>
<th>Phytophagous</th>
<th>Geophagous</th>
<th>Zoophagous</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Loricera pilicornis</em></td>
<td>0</td>
<td>0</td>
<td>0.3</td>
<td>0</td>
<td>0.7</td>
</tr>
</tbody>
</table>

References for the Appendix


