Species living in harsh environments have low clade rank and are localized on former Laurasian continents: a case study of Willemia (Collembola)
Andreas Prinzing, Cyrille D’Haese, Sandrine Pavoine, Jean-François Ponge

To cite this version:

HAL Id: hal-00940083
https://hal.archives-ouvertes.fr/hal-00940083
Submitted on 31 Jan 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.
Species living in harsh environments have low clade rank and are localized on former Laurasian continents: a case study of Willemia (Collembola)

Andreas Prinzing,1,2* Cyrille A. D’Haese,3 Sandrine Pavoine,4,5 Jean-François Ponge6

1. Université de Rennes 1, CNRS UMR 6553 ECOBIO: Ecosystèmes, Biodiversité, Evolution; Campus de Beaulieu, 263 avenue du Général Leclerc, 35042 Rennes Cedex, France. e-mail: andreas.prinzing@univ-rennes1.fr
2. Alterra, Centre for Ecosystem Studies, WUR, PO Box 47, 6700AA Wageningen, The Netherlands.
3. Muséum National d'Histoire Naturelle, Département Systématique et Évolution, CNRS UMR 7205, CP 50, 45 rue Buffon, 75005 Paris, France. e-mail: dhaese@mnhn.fr
4. Muséum National d'Histoire Naturelle, Département Écologie et Gestion de la Biodiversité, CNRS-UPMC UMR 7204, 55-61 rue Buffon, 75005 Paris, France. e-mail: pavoine@mnhn.fr
5. Mathematical Ecology Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.
6. Muséum National d'Histoire Naturelle, Département Écologie et Gestion de la Biodiversité, CNRS UMR 7179, 4 avenue du Petit-Château, 91800 Brunoy, France. e-mail: ponge@mnhn.fr

Running title: Use of harsh environments across a phylogeny

Word count (Abstract - references, included): 8446 words
Printed page estimation: title/abstract etc.: 1 page, Introduction – Discussion: 5956 words ≈ 6 pages, 94 references ≈ 3 pages, tables and figures: 1.5 pages (table 1 could be shifted to Appendix)

Estimate of number of journal pages required by table and figures: 1.5 (Tab. 1 could also be moved to an Appendix if needed)

* Corresponding author
ABSTRACT

Aim Certain species have few living relatives, i.e., occupy low clade ranks. Hence, they possess high conservation value and scientific interest as unique representatives of ancient lineages. However, we do not know whether particular environments favour the maintenance of low clade ranks or whether the distribution of environments across the globe affects the global distribution of clade ranks and, hence evolutionary uniqueness. In this study, we tested whether and how harsh environments decrease the clade ranks of the species that inhabit them.

Location Global

Methods We described the phylogeny of the collembolan genus *Willemia* by a parsimonious method based on 52 morphological characters and estimated the species’ use of harsh environments (polar, high-mountain, desert, polluted, waterlogged, saline, and acidic) from 248 publications.

Results We found that the use of different types of harsh environments is maintained among close relatives and has similar phylogenetic signals (except for the use of salinity). The use of harsh environments might therefore affect the diversification of lineages. Correcting for the phylogenetic non-independence of species, we found that species using harsh environments have comparatively low clade ranks. We also found that species using harsh environments occur almost exclusively on former Laurasian continents and that as a statistical consequence, Laurasian species tend to have lower clade ranks.

Main Conclusions We suggest that harsh environments maintain low-clade-rank species by decreasing, simultaneously or successively, extinction and speciation, which may eventually explain the major variation in clade rank across the globe.
Keywords abiotic stress; diversification; habitat; niche conservatism; phylogenetic reconstruction; phylogenetic generalised least squares; phylogenetic principal components; phylogenetic permutation; root-skewness test; tropical
INTRODUCTION

Extant species can occupy very different clade ranks within a phylogenetic topology. Certain species have very few living relatives and thus occupy a low clade rank, whereas others have much higher clade ranks. Species of low clade rank are the sole extant representatives of their lineages and hence have a high evolutionary uniqueness: the loss of a low-clade-rank species implies the loss of more evolutionary history than the loss of a high-clade-rank species (Purvis et al., 2000; Prinzing et al., 2004; Colles et al., 2009). For this reason, it is important to understand whether low-clade-rank species are maintained to a greater extent in certain environments or regions than in others.

It has been suggested that species of low clade rank persist and accumulate in regions with low extinction rates (Willis, 1922; Condamine et al., 2012), notably due to relatively stable climates, especially in the tropics (Wiens & Donoghue, 2004; Jablonski et al., 2006; Hawkins et al., 2007; Donoghue, 2008; Buckley et al., 2010). Furthermore, low clade rank has been suggested to reflect a relatively low speciation rate, as has been proposed for the regions outside the tropics (Willis, 1922; Jablonski, 1993, 1999; Chown & Gaston, 2000; Jablonski et al., 2006). Overall, species of low clade rank can be expected to most likely accumulate in regions in which a low speciation (or immigration) rate very roughly outweighs a low rate of extinction (or emigration). There would be no net effect on species richness under such circumstances.

The distribution of clade ranks across regions has received considerable attention, whereas the distribution of clade ranks across types of environments (broadly, “habitats”) has received much less attention. Bartish et al. (2010) have recently shown that within a region, particular harsh environments might be colonised by species of particularly low clade rank: across 40 different habitats in the Netherlands, those with extremely high soil moisture and
extremely low soil pH were characterised by low mean clade ranks of their angiosperm inhabitants. However, this analysis was restricted to a single, small region and might not reflect (or influence) globally coherent patterns. Here, we define abiotically harsh environments as environments that tend to impose a major direct physiological stress on most species of a given lineage — i.e., a constraint on growth and reproduction (sensu Grime, 1977) — with the obvious exceptions of highly tolerant species and sub-lineages. We can speculate that harsh environments might indeed reduce extinction rates, as patches of harsh environments remain present at least locally at any given time in any region and maintain gene flow (Behrensmeyer et al. 1992), and species in harsh environments might rarely be driven to extinction by biotic interactions (Grime 1977, Callaway et al. 2002). In addition, harsh environments might possibly reduce speciation rates by reducing two of the major triggers of speciation: (i) recombination events, which become rarer with environmental harshness due to increased generation times (Grime 1977), and (ii) evolutionary arms races between prey and their natural enemies, which become less diverse with harshness due to a reduction in the number of trophic levels (Vermeij 1987). We therefore hypothesise that clade ranks are lower in harsh environments (whereas species richness is not affected).

Whatever the relationship between environmental harshness and clade rank, the existence of such a relationship requires that there is no trade-off between the capacity to use different types of harsh habitats; otherwise, the average harshness of the habitat used by any species would be intermediate (see Hoffmann & Parsons, 1997 for possible mechanisms). The existence of a relationship between harshness and clade rank would also require that the use of harsh and mesic environments is phylogenetically conserved so that related species tend to have similar affinities to harsh and mesic environments (see Prinzing et al., 2001, phylogenetic signal sensu Losos, 2008). Without such conservatism, past patterns of speciation and extinction in ancestral environments would not be transmitted to distributions.
of clade ranks in present-day environments (Condamine et al., 2012). However, certain authors suggest that the capacity to use harsh environments can evolve and disappear rapidly (see Hoffmann & Parsons, 1997 and Hoffmann & Willi, 2008 for mechanisms), particularly if expansions and constrictions of such harsh environments may trigger the acquisition or loss of the capacity to use these environments (Gaston, 1998) and if the use of harsh environments imposes a cost (but see Gaston, 2003). We hence hypothesise that the use of harsh habitats is not constrained by a trade-off between different types of harshness and is phylogenetically conserved.

Should particular environments maintain low-clade-rank species and particular regions maintain these particular environments and their inhabitants, we would expect to observe that these regions also maintain many low-clade-rank species (see also Donoghue, 2008). We can speculate that selection might have favoured the use of harsh habitats far from the tropics (Jablonski, 2008), especially in the Northern Hemisphere, where a steep latitudinal gradient of decreasing biodiversity can be observed today (Chown et al., 2004). Given that the clade rank of a species reflects the outcome of millions of years of evolutionary history, the past differentiation of regions, notably the distinction between Laurasia and Gondwana, might be more important than their present-day configuration. One might speculate that past environments were harsher, on average, in Laurasia than in Gondwana (Vršanský, 2005; Crisp et al., 2010), reflecting, among other differences, the larger surface area of the landmasses in northern compared to southern temperate regions. Larger landmass would produce more climatic and thereby edaphic extremes in Laurasia and its descendent land masses than in Gondwana, except for Antarctica (Chown et al., 2004). Additionally, even restricted periods of harsh conditions might leave a strong signal in a continental fauna if the vast majority of the continent became harsh and later recolonisation was slow. This regional origin and maintenance of species might still be reflected by a larger number of low-clade-rank species
in landmasses stemming from Laurasia (centres of origin, Myers & Giller, 1988). Moreover, independent of the above speculations, given what we know of the present worldwide distribution of soils (FAO-UNESCO, 2007), climates (World Climate Map, 2012) and human activities, harsh environments (e.g., soil acidity, drought, frost, waterlogging, heavy metal contamination), with the exception of seashore salinity, appear to be more widely distributed in ex-Laurasian than in ex-Gondwanan continents. Whatever the precise causes, we can hypothesise that non-tropical regions and those of Laurasian origin harbour more low-clade-rank species, largely as a consequence of the increased numbers of species using harsh habitats.

The predictions derived from these hypotheses can be tested across extant species belonging to monophyletic lineages that are ancient (having survived several ecological crises and dating back to the Laurasia/Gondwana epochs), highly diversified and distributed worldwide. Collembola (springtails) are one such old, diversified lineage dating back to the Early Devonian (Hirst & Maulik, 1926; Greenslade & Whalley, 1986; Grimaldi, 2010), and most actual forms, at the family or even genus level, are known from the Cretaceous (Christiansen & Pyke, 2002a, b). Among the cladistic studies conducted on Collembola, the genus *Willemia* deserves special attention given its wide distribution from both an ecological and a biogeographical point of view. The genus is monophyletic and comprises a large number of species for which phylogenetic trees can be reconstructed unambiguously (D’Haese & Weiner, 1998; D’Haese, 1998, 2000 for subtrees of the genus). In addition, certain species of *Willemia* are known for their preferential use of arid (Thibaud & Massoud, 1988), polar (Sørensen et al., 2006), mountainous (Loranger et al., 2001), acidic (Ponge, 1993), saline (D’Haese, 2000), or polluted (Filser & Hölscher, 1997) environments. The genus is also representative of many others in the absence of a time-calibrated phylogeny due to the scarcity of fossil records. This lack of information renders approaches based on branch
length inapplicable but does not affect approaches based on clade ranks. Obviously, Willemia is only one out of an almost infinite number of genera. However, studying one genus may help to develop a methodological approach for teasing apart the associations between the use of harsh environments and patterns of diversification within a phylogenetic context. This approach may then be applicable to other genera and larger taxonomic units.

To evaluate the relationship between the use of harsh environments and clade rank, we tested whether (i) the use of different types of harsh environment is positively rather than negatively correlated (i.e., species tend to be able to tolerate either a broad range of harsh environments or none) and is phylogenetically conserved in the sense of being more similar among phylogenetically closely related species than among more distantly related species; and (ii) species using harsh habitats consistently occupy low clade ranks rather than being randomly scattered across the phylogeny, and these harsh environments tend to be the ancestral environments of such low-clade-rank species, which are as numerous as species absent from such harsh environments. To evaluate the relationship between geographic region, use of harsh environments and clade rank, we tested whether species outside the tropics occupy lower clade ranks than species within the tropics, due to a tendency of non-tropical species to use harsher habitats. We also tested whether species on former Laurasian land masses occupy lower clade ranks than species on former Gondwanan land masses, due to a tendency to use harsher habitats. In all analyses we accounted for the statistical non-independence of species. We also conducted character mapping to reconstruct ancestral stages and explore whether the environments and regions used by species are indeed ancestral to the respective (sub)lineage and hence might have influenced the clade rank of the respective species in that (sub)lineage.
MATERIALS AND METHODS

The genus Willemia and the reconstruction of its phylogeny

Within the Class Collembola, the genus Willemia belongs to the Hypogastruridae family. It differs from other hypogastrurid genera by the total lack of pigment or furcula and the small size of the slender body, which never exceeds 1 mm in length (Thibaud, 2004). According to their life form, all Willemia species belong to the euedaphobiont sub-category Bc3b (small size, slender body, no furcula) of Rusek (2007). The study addresses 42 of the 43 species currently known worldwide in this genus (list in Appendix S1b). The absent species was only described in 2011 by D’Haese & Thibaud, so its environmental or geographic distribution is still very far from being sufficiently documented. The genus is distributed worldwide, with 15 species recorded only in the tropics, 25 species outside the tropics and only 2 species present both in the tropics and elsewhere (details about the biogeographic distribution of species in Appendix S2e). A total of 13 species were recorded from continents and islands of Gondwanan origin vs. 20 of Laurasian origin and 9 of uncertain origin (Appendix S2e). As for most Collembola, dispersal modes are still unknown, although sea currents have been suspected to favour long-distance transport (Thibaud, 2007). Willemia species live in the soil (from litter to mineral soil, whether acid or alkaline), in psammic environments (beaches, sand dunes, deserts) and in caves, but not all of them are found in harsh environments (Table 1). Overall, the great variation in the biogeographic and ecological distributions of species, together with a sufficient but still-manageable number of species, makes this genus a good model for testing hypotheses about relationships between biogeography, ecology and the evolution of extant species.

The reconstruction of the phylogeny of the genus Willemia is explained in Appendix S1a-d. This reconstruction confirmed the monophyly of the genus already established by
D’Haese (2000). We were constrained to use morphological characters, as explained and justified in Appendix S1a. Obviously, speciation may not always leave morphological traces, and such “cryptic” speciation is overlooked if morphological characters alone are considered. This outcome is especially probable in lineages with morphological characters that are few in number or unstable in terms of shape and/or position (among Collembola, e.g., genera *Folsomia* and *Parisotoma*). *Willemia*, however, has numerous characteristics (e.g., hairs, sensilla, vesicles) of stable shape and position. Due to this feature of the genus, speciation is unlikely to be cryptic in *Willemia*. Cryptic speciation may be more frequent at the population level, but such ephemeral population phenomena were not of interest in our study. We also note that a dated phylogeny is not feasible for *Willemia* given the lack of dating points caused by the absence of fossils for this genus.

All analyses were run on each of the 6 most parcimonious phylogenetic trees as well as on a strict-consensus of 6 phylogenetic trees (detailed in Appendix S1: Phylogenetic reconstruction). The results from analyses run on the strict consensus tree are given in the Results section, those from the 6 individual trees in the corresponding appendices (detailed in Appendix S3: Reconstruction of ancestral states).

**Use of harsh environments and the biogeographic distribution of *Willemia* species**

The use of harsh environments (as defined in the Introduction) was indicated by the occurrence of *Willemia* species in environments known for at least one factor that is thought to be a major constraint for most soil-dwelling organisms (see Hopkin, 1997 for springtails): i.e., xeric, hydric, arctic, alpine, acidic, metallic or saline soils. See Appendix S2a for detailed definitions of these factors and literature research methods and Appendix S2b for references. A ‘harshness breadth’ index was estimated for every species based on the number of harsh
environments in which the species was recorded, scaled from 0 (no harsh environments recorded for the species) to 7 (all harsh environments recorded).

Biogeographic distributions were categorised as (i) tropical (between the tropic latitudes, Inter-Tropical Convergence Zone, ITCZ) or non-tropical (north or south of the tropic latitudes) and as (ii) Gondwanan or Laurasian following the maps by Christiansen & Bellinger (1995), as detailed in Appendix S2e. Appendix S2e also outlines the relationships between tropical/non-tropical and Laurasia/Gondwana classifications and between regions and harshness.

Statistical Analyses

The correlation among uses of different types of harsh environments across lineages was analysed by a phylogenetic Principal components analysis (pPCA), a multivariate method recently devised by Jombart et al. (2010b) by extending a methodology developed in spatial ecology and spatial genetics to the analysis of phylogenetic structures in biological features of taxa.

Phylogenetic conservatism is the tendency of closely related species to share similar values for a given trait (typically more similar than distantly related species, Wiens et al., 2010). We predicted phylogenetic conservation of the use of harsh environments, i.e., that related species tend to have similar harshness breadth index values. Here, harshness breadth varied from 0 to 7 harsh environments as defined above. Phylogenetic conservatism for harshness breadth was tested with the Pavoine et al. (2010) approach. Briefly, the total trait diversity of the lineage was decomposed across the nodes of a phylogenetic tree by attributing to each node a value measuring the differences among lineages descending from that node weighted by the proportion of species descending from it. Permutation tests (999 replicates)
allowed the attribution of a probability to the observed value. Our principal test was the root-

skewness test, which verifies whether phylogenetic distances between species correspond to
distances in a given trait. Supplementary tests in Pavoine et al. (2010) explore whether trait
variation is concentrated on a single node or on a few nodes, but that work cannot be
presented in detail here (usually, the former were not significant, whereas the latter were).
Phylogenetic distances were measured as nodal distances, i.e., the number of nodes on the
shortest path that connects two species on the phylogenetic tree. For measuring distances
between species in terms of habitat harshness (a multi-choice variable), we used the simple
matching coefficient (Sokal & Michener, 1958): (a+b)/(a+b+c+d), where ‘a’ is the number of
harshness types in common, ‘b’ is the number of harshness types with which neither of the
species are associated, and ‘c’ and ‘d’ are the numbers of harshness types unique to each of
the two species.

Losos (2008) has suggested that the term ‘phylogenetic conservatism’ should only be
used if trait change along the phylogeny is slower than Brownian motion and that the term
‘phylogenetic signal’ should be used for the wider phenomenon of correlations between trait
distance and phylogenetic distance. However, properly identifying the rate of trait change
would require a dated phylogeny, which is not feasible for Collembola. Moreover, whereas
change in the use of harsh environments may be considered ‘random’ and not ‘conservative’
from the point of view of phylogenetic reconstruction, it is still much more ‘conservative’
than the changes in the harshness of the environment surrounding the animals. For simplicity,
we therefore retain the term ‘conservatism’ and use it in the sense of a pattern, not of a
process (see Wiens et al., 2010 for further discussion on phylogenetic conservatism).

We then tested the link between harshness breadth as a response variable and (i)
phylogenetic clade rank, (ii) tropical/non-tropical classification and (iii) Gondwana/Laurasia
classification as explanatory variables. Any test of hypotheses on the environmental and
geographic correlates of clade ranks of species is vulnerable to the phylogenetic non-
independence of species. A classical approach to the problem of phylogenetic non-
independence is the use of sister-clade comparisons (Felsenstein, 1985), but this approach is
not applicable in our case because sister clades have the same clade ranks by definition. An
alternative and more general approach is to include co-matrices of phylogenetic proximity in a
phylogenetic generalised least square fit (PGLS) as described in Martins & Hansen (1997),
using the method of Grafen (1989) to estimate branch lengths. To increase the symmetry of
the harshness breadth distribution, this variable was square-root transformed.

We tested the relationships between the phylogenetic clade ranks of species as the
dependent variable and either the species’ tropical/non-tropical classification or their
Gondwana/Laurasia classification as the independent variable. We conducted alternative
analyses, either incorporating the uses of harsh habitats as covariables or not incorporating
these uses. As the dependent variable was directly inferred from the species’ phylogenetic
position and was, thus, a characteristic of the phylogeny, we incorporated phylogenetic non-
independence in the explanatory variables as follows. We computed the F-statistic associated
with each variable using an ordinary linear model. We then estimated a theoretical
distribution of this statistic via the method of phylogenetic permutation described by Lapointe
& Garland (2001, see also Harmon & Glor, 2010): closely related species are more likely to
exchange values than more distantly related species. The values of the explanatory variables
were permuted (999 times) by this approach, keeping the phylogenetic clade rank unchanged.
The P-value was then calculated as the proportion of theoretical values of the F-statistic that
were higher than or equal to the observed value. We note that species that occurred in both the
Laurasia and Gondwana biogeographic domains (6 species) or were on continents or islands
of uncertain origin (3 species) and species that were recorded both within and outside the
tropics (2 species) were discarded from geographic analyses.
All these methods were implemented in R (R Development Core Team, 2010), which was used for all statistical calculations. The packages used were ade4 (Dray & Dufour, 2007), adephylo (Jombart et al., 2010a), ape (Paradis et al., 2004), and nlme (Pinheiro et al., 2010). The R codes are available upon request.

For character mapping, we used Mesquite (Maddison & Maddison, 2011); specifically, we used parsimony and stochastic character mapping. As both approaches produced the same results, we only present the parsimony mapping, which suffers less from the absence of biogeographic classifications for individual species.
RESULTS

The characters and phylogeny of Willemia are provided in Appendix S1b and Appendix S1d, respectively. Note that there is no phylogenetic structure in the study intensity of Willemia species (i.e., the numbers of references), and hence, the analyses presented below are not biased by study intensity (Appendix S2c). Although branch lengths are not available, as explained in the Introduction, there is a major variation in the clade rank of species, ranging from 2 to 13, reflecting substantial variation in the number of extant close relatives across species.

Uses of most types of harsh habitats are positively correlated

The phylogenetic principal components analysis (pPCA) indicated strong positive correlations between all types of harsh habitats with the exception of salinity. The use of all other types of harsh habitats showed clear positive scores on the same principal component axis (Fig. 1). A visual examination of Table 1 (where species are arranged according to their order in the phylogenetic tree) confirms this result: with the exception of salinity, the use of all types of harsh habitats appears to be phylogenetically positively correlated. Given that the use of saline habitats corresponded to the absence of all of the other six types of harsh habitats, we excluded salinity from further analyses to better capture the overall syndrome of harsh habitat use (see also Appendix S2d).

Phylogenetic conservatism in the use of harsh habitats

The types of harshness under which species were found exhibited clear phylogenetic conservatism (root-skewness test P=0.002). Additionally, the number of types of harsh
habitats used by a species (harshness breadth) was phylogenetically conserved (root-skewness test; P=0.042). If salinity was excluded from the array of harsh habitats (given that it is not positively related to other harshness factors, see above), a stronger conservatism of harshness breadth was found (root-skewness test P=0.013). The presence/absence of Willemia species in harsh habitats (rather than the number of harsh habitats used) exhibited an even clearer phylogenetic conservatism if salinity was excluded from the array of harsh habitats (root-skewness test P=0.001). This phylogenetic structure was confirmed using parsimony character state reconstruction (Fig. 2 a). These results are based on the consensus tree but are robust to phylogenetic uncertainty and were found in all 6 of the equally parsimonious trees (Appendix S3).

Species using harsh habitats have lower clade ranks but are as numerous as species using mesic habitats

We found that the clade rank of a species decreased with the species’ harshness breadth (PGLS; P=0.002). If salinity was excluded, it appeared that the distance to the root of the phylogenetic tree was quite similar for all species using harsh habitats independently of the number of harsh habitat types they were using, with all harshness-tolerant species occupying a basal position (Fig. 3). The ancestral nature of life in harsh habitats was also visually assessed through parsimony character state reconstruction (Fig. 1 a), mapping the use of harsh habitats (salinity excluded) and other habitats on the cladogram. Most of the 22 harshness-tolerant species were located close to the root, i.e., they corresponded to a “basal” pool of species including all species from W. multilobata to W. denisi in Fig. 1 a. Only three species in this group did not confirm that harshness tolerance was an ancestral character: W. bellingeri, W. psammophila and W. unispina. Species belonging to the most advanced group, in contrast,
were not found in harsh habitats (except for saline habitats) with the exception of W. nepalensis. All these results are based on the consensus tree but are robust to phylogenetic uncertainty and were found in all 6 of the equally parsimonious trees (Appendix S3).

In total, there were 22 species using harsh habitats and 20 species using only mesic habitats (i.e., species that were never recorded in any of the harsh habitat types considered except salinity). This result is not significantly different from equality (df=1; Chi²=0.095; P=0.758).

**Geographical distribution: species using harsh habitats tend to be Laurasian, and as a statistical consequence, Laurasians tend to have low clade ranks**

We found that species using harsh and mesic habitats were distributed differently across the globe. Species using harsh habitats tended to occur in non-tropical regions, although the relationship was relatively weak (Fig. 4a; PGLS; df=30; t=2.29; P=0.029). At the same time, species using harsh habitats were strongly restricted to regions of Laurasian origin (Fig. 4b; PGLS; df=30; t=3.19; P=0.003). In fact, only two harshness-tolerant species were recorded outside former Laurasia (W. trilobata and W. namibae, Appendix S2e). These results are based on the consensus tree but are robust to phylogenetic uncertainty and were found in all 6 of the equally parsimonious trees (Appendix S3).

Low-clade-rank species were strongly restricted to present non-tropical regions (Fig. 5a; n=31 as in all tests of this paragraph, tests based on phylogenetic permutations, F=26.14; P=0.001, and see character state reconstruction, Fig. 1b). This relationship was maintained if harshness breadth was included as a co-variable: harshness breadth and presence in present-day non-tropical regions were both related to low clade rank (F=23.90; P=0.001 for harshness breadth, and F=9.76; P=0.007 for presence in present-day non-tropical regions; conclusions
were not impacted by the order in which the two explanatory variables were entered in the model; harshness first: F=29.17; P=0.001, harshness second: F=4.48; P=0.048). Therefore, the high harshness tolerance of non-tropical species did not explain their low clade rank. Species of low clade rank also tended to be restricted to former Laurasia regions (Fig. 5b; F=10.47; P=0.006 and see character state reconstruction, Fig. 1c). If occurrence in former Laurasia was included first in the model and harshness breadth second, both variables were significant (F=12.68; P=0.002 for occurrence in former Laurasia and F=7.33; P=0.014 for harshness breadth). However, this relationship disappeared if harshness breadth was added to the model first: in this case, occurrence in former Laurasia was no longer significant (F=1.31; P=0.262), whereas harshness breadth was still significantly related to low clade rank (F=18.69; P=0.001). Overall, this result indicates that species using harsh environments are particularly bound to former Laurasia continents (above paragraph) and that — as a statistical consequence — we find many species of low clade rank on former Laurasia continents. All these results are based on the consensus tree but are robust to phylogenetic uncertainty and were found in all 6 of the equally parsimonious trees (Appendix S3).

DISCUSSION

We found that uses of different types of harsh environments are maintained among close relatives and have similar phylogenetic signals with the exception of use of saline habitats. Use of saline habitats is thus not consistent with our hypotheses (see Introduction) and was excluded from further analyses. There are two opposing lines of explanation for this exception. First, one can argue that tolerance of saline conditions comes with a major cost in terms of tolerance of multiple other harshness factors, and thus, salinity tolerance will not increase the overall capacity to use harsh environments and will be lost rapidly through
evolution if no longer needed. Second, there are putative arguments suggesting that salinity is not a form of harshness from the point of view of Collembola (Appendix S2d). Pending more ecophysiological studies on the tolerance of springtails to sea water, it might be suggested that life in contact with dilute saline solutions (seashores but not dry saline environments) requires no special adaptation in basal Hexapoda, in contrast to fully terrestrial insects (higher insects).

Using a phylogenetic framework, we showed that among Willemia springtail species, the use of harsh habitats (tolerance of harshness) displayed a strong pattern of phylogenetic conservatism and no phylogenetic trade-off in the capacity to use different types of harsh habitat (except for salinity, see Appendix S2d). These results ultimately suggest that the use of harsh habitats can affect the mode of diversification of these lineages and, thereby, the distribution of clade ranks across habitats and eventually regions. In fact, we found that species using harsh habitats have significantly lower clade ranks.

Although the occurrence of archaic life forms has already been documented in specific exotic and harsh habitats such as volcanoes, hot springs, deserts and ice caps (Oliver et al., 2000; Nriagu, 2003; Labandeira, 2005; Oliver et al., 2005; Ehrhardt et al., 2007), a systematic link between species clade rank and habitat harshness has not, to our knowledge, been demonstrated previously. Lower clade ranks would be consistent with lower speciation rates in harsh habitats. However, we also found that species using harsh habitats are even slightly more numerous than species using mesic habitats. This combination of results is consistent with a scenario of decreased net extinction of species combined with decreased net speciation in harsh habitats. We stress, however, that we can only speculate on speciation and extinction averaged across time; we do not know whether this pattern reflects permanently low speciation rates combined with permanently low extinction rates or, for instance, initially high speciation and very low extinction rates combined with later very low speciation rates combined with high extinction rates. Both scenarios may produce the same low clade ranks of
extant species. Independently of the precise mechanisms, harsh habitats function today as
museums of low-clade-rank species, i.e., of species that are the last extant representatives of
ancient lineages.

Such a pattern of conservatism and of the preferential presence of low-clade-rank
species in harsh habitats may ultimately result in the global distribution of harsh-habitat
species driving the global distribution of low-clade-rank species, as predicted. In fact, we
found that species using harsh habitats exhibit distinct geographic distributions — they are
more numerous outside the tropics and, in particular, on former Laurasia — and that species
of low clade rank match these distributions. In addition, for Laurasia, the occurrence of low-
clade-rank species is, in fact, explained by the occurrence of species in harsh habitats. For this
reason, the geographically conserved use of harsh habitats decreases the phylogenetic clade
rank of species on the continents of former Laurasia.

Our results contribute to the debate on the evolutionary fate of ‘generalists’ versus
‘specialists’. Generalists are thought to be better able to overcome ecological crises (Raup,
1986; Zhou et al., 2005). More recently, however, Colles et al. (2009) showed that different
phylogenetic approaches may lead to diverging conclusions. Our present phylogenetic study
compares ‘harshness generalists’ to species putatively specialising in habitats lacking any of
the harshness factors. The results are consistent with a longer survival of ‘harshness
generalists’ and hence rather point towards generalists being, in fact, less susceptible to
environmental change than specialists. However, our study also indicates that reduced
extinction might possibly be coupled with reduced speciation, as this has been shown with
other traits, e.g. geographic range (Parker et al., 1999), leaving practically no net effect on the
numbers of species using harsh habitats.
Our results point to a common factor of tolerance of environmental harshness, which would be shared by a pool of low-clade-rank species, as uses of different harsh habitats were positively correlated with one another. This result is not surprising in view of the roles of soil acidity, altitude and latitude: the biologically active part of the soil (the topsoil) becomes more acidic as the decomposition rate of organic matter decreases, leaving the organic matter humified rather than mineralised (Ulrich, 1986; Ponge, 2003, 2012). However, it was more surprising to find relationships between these three factors and drought or metallic pollution. The only point in common among these five factors appears to be a decrease in biological activity caused by harsh environmental conditions. To explain these positive relationships among different types of harshness tolerance, we should first examine the cellular level. The integrity of cell membranes, combined with active cellular repair, is known to protect a wide array of organisms from death caused by environmental hazards (Nriagu, 2003; Oliver et al., 2005; Shi et al., 2008; Timmermans et al., 2009). This feature could be the most primitive mechanism of harshness tolerance, requiring no specialised adaptation to any particular harshness factor but rather the coordinated expression of genes functioning in common in a ‘general-purpose genotype’ (Weider, 1993; Niehrs & Pollet, 1999; Vohradsky et al., 2000; Van Doninck et al., 2002; Kaplan et al., 2004). For instance, it has been shown that drought acclimation produces cold tolerance in springtails (Holmstrup et al., 2002) and that long-chain fatty acids of phospholipidic membranes are involved in this phenomenon (Bayley et al., 2001). These findings are consistent with the evolutionary importance of membrane integrity (Jobson et al., 2010). Adaptation to harsh environments may occur rapidly if a phylogenetic lineage already possesses an efficient array of regulatory genes, known to be as evolutionarily important as changes in enzyme structures (Roelofs et al., 2006, 2010). However, the evolutionary acquisition of this gene array might be difficult and might only occur rarely. This
hypothesis is consistent with our observation of only a few cases of independent acquisitions of the capacity to use harsh habitats.

We showed that derived species of high clade rank were primarily tropical species (while not all tropical species are derived), an observation also made for marine bivalves by Harnik et al. (2010). This result is in contrast to the tropical conservatism hypothesis of Wiens & Donoghue (2004), which suggests that the tropics, representing supposedly more ancient environments, should harbour more ancient species characterised by low clade ranks (without excluding more recent species). Our results could be explained by a more ancient origin of environments observed today outside of the tropics, compared to today’s tropical environments (Smith et al., 2007). However, across the lifetime of the Collembola lineage, the tropical environments are usually considered to be at least as old as the non-tropical ones (e.g., Behrensmeyer et al., 1992). Alternatively, our results can be explained if the low clade rank of species results from a trait such as the capacity to use harsh habitats and if species with this trait tend to occur outside rather than inside the tropics. However, we found that the tropical/clade-rank relationship was independent of the use of harsh habitats. The observed high clade ranks of tropical species therefore remain unexplained. Potentially, traits other than the use of harsh habitats need to be considered, as well as an interplay between environmental filters (mainly operating out of the tropics) and evolutionary potential of lineages (mainly operating in the tropics), as postulated by Harnik et al. (2010). We note, however, that most tropical species are found on territories of Gondwanan origin and that the high clade rank of Gondwanan species, in turn, could be explained by their present-day use of mesic habitats. Our results may thus point to a link between niche conservatism, diversification, and biogeographic distribution leading to patterns that are inconsistent with those predicted by Wiens & Donoghue’s (2004) tropical conservatism hypothesis. It remains to be tested whether
the patterns we observed are a rare exception for a small group of species or can be found more frequently, e.g., in other groups of invertebrates.

By creating harsh environments (acid rain, polluted water, soil and atmosphere, vegetation-less areas) throughout the world and by accelerating species dispersal (alien introductions), humans select for particular major lineages to the detriment of others (Purvis, 2008; Ozinga et al., 2009; May, 2010). However, our results show that by rendering global environments increasingly harsh, humans also select for particular evolutionary scenarios to the detriment of others: increasingly abundant harsh environments appear to represent museum scenarios, i.e., environments that accumulate low-clade-rank species. It is possible that this outcome is due to low extinction and speciation rates. In contrast, environments of low harshness become rare and therefore potentially represent a cradle scenario, i.e., environments that accumulate high clade-rank species, due perhaps to high rates of speciation.

Another result is that the use of harsh habitats is apparently a character that, in most cases, does not originate de novo when needed but is carried by particular species for very long periods of time. A high priority should thus be given to the conservation of this pool of low-clade-rank species. However, other studies are necessary to ascertain the generality of the patterns we discerned in one ancient animal group, Collembola. These studies should focus on other ancient but highly diversified groups, such as mites and mosses, which are suspected to be relatively tolerant of harsh environments (Prinzing et al., 2000; Oliver et al., 2005).

ACKNOWLEDGEMENTS

We acknowledge the work of countless collectors in the field and of the museum staff responsible for the collections, our discussions with Scot Wing and Conrad Labandeira and the technical support of Valerie Briand. AP profited from ATIP funding provided by the
 Comments by Michael Angilletta, Marl McPeek and two anonymous reviewers substantially improved the manuscript.
REFERENCES


SUPPORTING INFORMATION

Appendix S1 Phylogenetic reconstruction

Appendix S2 Literature research: definitions, references, exploration of biases

Appendix S3 Robustness of the analysis to variation in underlying phylogenetic trees

BIOSKETCH

Andreas Prinzing studies the interface of habitat environments, past evolutionary diversification of species pools, and how this diversification controls, and is maintained by, present ecological interactions.

Author contributions: JFP and AP conceived the ideas with the collaboration of all other authors, JFP collected the data, CDH constructed the phylogeny, SP analysed the data, and JFP and AP took the lead role in the writing.

Editor: Melodie McGeoch
Table 1. Use of harsh environments by *Willemia* species according to literature (references listed in Electronic Appendix S2). See text for more details on stress factors.

<table>
<thead>
<tr>
<th></th>
<th>Xeric</th>
<th>Hydric</th>
<th>Arctic</th>
<th>Alpine</th>
<th>Acid</th>
<th>Metallic</th>
<th>Saline</th>
<th>Tolerance breadth (salinity included)</th>
<th>Tolerance breadth (salinity excluded)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>W. multilobata</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><em>W. bellingeri</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. trilobata</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. namibiae</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. similis</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><em>W. christianseni</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>W. anophthalma</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td><em>W. dubia</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>W. scandinavica</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td><em>W. koreana</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. ichtacihuatensis</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>W. bedosae</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. unispina</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. psammophila</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. virae</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. tali</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. intermedia</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>W. shanghainensis</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. granulata</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><em>W. fjellbergi</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. arida</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><em>W. trisphaeræ</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><em>W. elisabethum</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. denisi</em></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><em>W. japonica</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. subbulbosa</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. persimilis</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. acantha</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. bulbosa</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. meybholaë</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. arenicola</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. nepalensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><em>W. neocaledonica</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. nadchatrami</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. brevispina</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. annapurna</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. buddenbrocki</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. wandae</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. nosyboraha</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><em>W. delamarei</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. setonychia</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>W. dehavergi</em></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
FIGURE CAPTIONS

Figure 1. Ordination of correlations between uses of different types of harsh habitats, analysed by pPCA (a multivariate analysis based on phylogenetic distances). The coordinates of harsh habitat types are given along the first principal component. Note that except for salinity, the uses of habitats of all types of harshness are positively correlated.

Figure 2. Parsimony reconstruction of the use of harsh habitats, i.e., any type of harsh habitat used except for salinity (A), tropical distribution (B), and distribution on former Gondwana continents (C). The corresponding character states are shown in black. The tree shapes are not exactly identical because biogeographic distributions could not be assigned to either of the categories for a few species.

Figure 3. Clade ranks of species, i.e., nodal distances to the root, using different numbers of harsh habitat types. Error bar=S.E. Note that species that do not use any harsh habitats are at a greater distance from the root (see text for analyses).

Figure 4. Harshness breadth of species from tropical vs. non-tropical regions (a) and of species of Gondwanan vs. Laurasian origin (b). The few species of intermediate localisation were not included. Error bar=S.E.

Figure 5. Clade ranks, i.e., distances to the root of the phylogenetic tree, in species from tropical vs. non-tropical regions (a) and from areas of Gondwanan vs. Laurasian origin (b). The few species with intermediate localisation were not included. Error bar=S.E. Note that the difference between Gondwanan vs. Laurasian origin disappears after including harshness breadth as a covariable (whereas harshness breadth is significant at p=0.011, see text).
Fig. 1
Fig. 2

A – use of harsh habitats
B – tropical distribution
C – former-Gondwana distribution
Fig. 3
Fig. 4

(a)

(b)

Laurasia

Gondwana

outside tropics

tropical

harshness breadth

(n = 23)  

(n = 20)  

(d.f. = 30)  

(t = 2.29)  

(P = 0.029)

(d.f. = 30)  

(t = 3.19)  

(P = 0.003)
Fig. 5
Appendix S1: Phylogenetic reconstruction

Appendix S1a: General approach

Within the Class Collembola the genus Willemia belongs to the Hypogastruridae family. It differs from other hypogastrurid genera by the total lack of pigment and furcula and the small size of the slender body, which never exceeds 1 mm in length (Thibaud, 2004). According to their life form all Willemia species belong to the euedaphobiont sub-category Bc3b (small size, slender body, no furcula) of Rusek (2007). The study deals with 42 of the 43 species worldwide known today in this genus (list below in Appendix S1b). The remaining species W. biseta Christiansen & Bellinger 1980 was excluded because it could not be studied in detail (see D’Haese 2000). Eight species have been added to the phylogenetic tree published by D’Haese (2000): W. bulbusa Bonet 1945, W. acantha Bonet 1945, W. unispina Fjellberg 2007, W. ichtacciuatlensis García Gómez & Cutz Pool 2008, W. tali Kaprus’ & Nevo 2003, W. psammophila Palacios-Vargas & Thibaud 2001, W. shanghaiensis Yue 1999, and W. nosybohara Thibaud 2008.

Fifty-two morphological characters, which have been detailed in D’Haese (2000), were used for the phylogenetic analysis. Question marks (Appendix S1 b) represent either inapplicable or unknown characters. Autapomorphies, i.e. derived traits that are unique to a given terminal group, were excluded, so that only potentially informative characters were kept. All character states were polarized through comparison with three species belonging to the Hypogastruridae as outgroup: Hypogastrura vernalis Carl 1901, Orogastrura dilatata Cassagnau 1954, and Xenyllogastrura octoculata Steiner 1955.

We used morphological characters for reconstructing the phylogeny of the genus Willemia, because most species are rare and known only as collection specimens kept in conditions improper for molecular studies, a situation that often impedes worldwide
invertebrate molecular phylogenies if we limit ourselves to molecular characters (Turner et al. 2010). Moreover, it has been shown that phylogenies based on molecular and morphological characters of extant species are in general largely congruent (Jablonski & Finarelli, 2009) despite obvious discrepancies in particular cases. Finally, none of the 52 characters which were used for the phylogeny of the genus *Willemia* (D’Haese, 2000) is related to use of harsh environments. These characters concern chaetotactic setup of body and legs, which is of high taxonomic value in most springtail families (including Hypogastruridae) but has little to do with ecological adaptation (Gisin, 1967; Nayrolles, 1998; Rusek, 2002). The only exception is pigmentation, absent from all *Willemia* species but present in the three outgroup species used *(H. vernalis, O. dilatata, X. octoculata)*. Thus there is no reason to believe that the use of harsh environments, or a set characters correlated with it, was included in the reconstruction of the phylogeny.

The data matrix (Appendix S1b) was analysed with POY version 4.1.2 (Varón et al., 2010) through standard parsimony (static homology) with non-additive character coding (unordered characters). The data matrix was submitted to a thousand replicates, ratchet perturbation i.e. selection of 15% of characters upweighted by a factor of 3 at each iteration (Nixon, 1999), tree fusing (Goloboff, 1999) and to a final branch swapping refinement keeping up to 20% longer trees. Consistency (Kluge & Farris, 1969) and Retention Indices (Farris, 1989) were calculated. Bremer (Bremer, 1988) and jackknife (Farris et al., 1996) branch support indices were computed.

For further calculations on phylogenetic relationships among *Willemia* species we used a strict majority consensus tree based on six fully-resolved trees. The distance of each species to the root was calculated by the number of nodes directly descending from the root to it.
All data matrices, character list and analysis scripts can be found below in Appendix S1b and S1c.

REFERENCES


Appendix S1b: Data matrix of 52 characters in 3 outgroups and 42 species of the genus Willemia.
Appendix S1c: Files used for the construction of the strict consensus phylogenetic tree.

Printout of files used by Cyrille d’Haese for the construction of the strict consensus phylogenetic tree (Fig. 1). The file will-bre1.trees needed to calculate bremer support is not provided because of its size (~58Mo). All these files are available upon request at dhaese@mnhn.fr.

The following file willemia2000.ss is the original data matrix from D’Haese (2000) in Hennig86 format:

```
xread
'Matrix of DHaese, C. Is psammophily an evolutionary dead end? a phylogenetic test in the genus Willemia (Collembola: Hypogastruridae) Cladistics, 2000, 16, 255-273'
52 37
Hypogastrura_vernalis
000000000?00000000000000000000000000000000
Orogastrura_dilatata
0000000000000000000000000000000001010001010000000
Xenyllogastrura_octoculata
00000001000?010000001011000000000001010010?100?0000
Willemia_annapurna
1003003111111013001310100011111101111111111101
Willemia_anophthalma
1000000110001010100011101000010001000100010001
Willemia_arenicola
02222221110211012101231000111111111101101010110
Willemia_arida
1332222110?100110101000?1010001100011011010110011
Willemia_bedosae
100000011000101010001110100011111001100111001
Willemia_belleri
100000011000101010000101110011000000000001010000000
Willemia_brevispina
13032031111111110101221010001101111111111111111101
Willemia_buddenbrocki
10030031111111210013310100011011111011111111111101
Willemia_christianseni
13000001100010101000111010001001000110011001
Willemia_deharvengi
20030031111131110131001231111110111111111110111111111
Willemia_delamarei
10030031111111111111111111111111110111111111111111111111
Willemia_denisi
111111110010111111011110111101111011110111101111111111111111
Willemia_dubia
1000000110001010100001110100011000000000011001000000
Willemia_elisabethum
1111211100101111010111101111111111101110111101100111
Willemia_fjellbergi
133222211001010100011101000111000110001001010110011
Willemia_granulata
10022011001010100100111101000110110110100100011
Willemia_intermedia
13000011001101010011000111110010111010010001
Willemia_japonica
1002221100101010011101100110011111101101111101111
Willemia_koreana
10022011001101010011000111110011011010100011
Willemia_meybholae
0222221111111110121110001111111111111111111111111
Willemia_multilobata
000201110010111010100011101000110000110110100100001
Willemia_nadchatrami
100300311111111101221000011011111111111111111111111
Willemia_namibiae
0000001100110110101000111010001101101101100111001
Willemia_neocaledonica
10030031111111111121001210100011111111111111111111111
Willemia_nepalensis
1002201100101101101100110011011001101101101101101
Willemia_scandinavica
10002201100010110101000111010001101111001101101000011
Willemia_setonychia
2003003111????113101110001101011010010100100100101
Willemia_similis
10000110011101100110011010001101100100001101101
Willemia_subbulbosa
1222221111112101101002310100011001011010101010101
Willemia_trilobata
0000001101011010111010100110010101111111111111111
Willemia_virae
1000000110110100101010011111100110110110100001101
Willemia_wandae
1000001100110110100101001111110111111111111111111111111

ccode - 0.51;
proc /;
;
6 Sensilla_i2_on_antennal_segment_IV subcylindrical_ (Fig._2A)
candle_flame-shaped_ (Fig._2B) globular_ (Fig._2C)
globular_in_cavity_ (Fig._2D);
(7)
Small_internal_sensilla_of_sensory_organ_of_antennal_segment_III_ (Figs.
_2A to _2D) free covered_in_part_by_a_tegumental_fold;
(8) Third_and_fourth_antennal_segments clearly_separated
ventrally_slightly_fused;
(9) Antennal_segment_II_with 12_setae 11_setae;
(10) Antennal_segment_I_with 7_setae 6_setae;
(11) Prelabral_chaetotaxy_with 4_setae 2_setae;
(12) Labral_chaetotaxy_with 5.5.4_setae 4.5.4_setae 2.5.4_setae
5.3.4_setae;
(13) Seta_a0_on_the_head present_ (Fig._1A) absent_ (Fig._1B);
(14) Number_of_setae_d_on_the_head: 5 + 5 4 + 4 (Figs._1A and _1B);
(15) Number_of_setae_v_on_the_head: 2 + 2 1 + 1 (Figs._1A and _1B);
(16) Setae_c1_on_the_head present absent;
(17) Corneules present absent;
(18) Postantennal_organ_with 4 to 9_vesicles_ (Fig._1A) 10_to_15_vesicles
more_than_15_vesicles_ (Fig._1B);
(19) Setae_a1_of_thoracic_tergum_III positioned on_m thoracic_tergum_III
between_thoracic_terga_II_and_III_ (Figs._1A and _1B);
(20) On_second_and_third_thoracic_terga,_setae_a1 present absent;
(21) On_thoracic_tergum_II_and_III,_setae_a2 present absent;
(22) On_thoracic_tergum_II_and_III,_setae_m3 present_ (Fig._1A)
absent_ (Fig._1B);
(23) Tibiotarsi_I_and_II_with 19_setae 17_setae 12_setae 11_setae;
(24) TibiotarsiIII_with 18_setae 16_setae 12_setae 11_or_10_setae;
(25) Tibiotarsi with_tenant_hair without_tenant_hair;
(26) On_thoracic_tergum_III,_setae_s (=m7) subcylindrical_and_acuminate
lanceolate;
(27) On_abdominal_terga_II_and_III,_setae_s
subcylindrical_and_acuminate_ (Fig._1A) lanceolate_ (Fig._1B);
(28) On_abdominal_tergum_IV,_setae_s subcylindrical_and_acuminate
lanceolate;
(29) On_abdominal_tergum_V,_setae_s subcylindrical_and_acuminate
lanceolate;
(30) On_abdominal_tergum_II_and_III,_setae_a2 present absent;
(31) On_abdominal_tergum_II_and_III,_setae_m2 present_ (Fig._1A)
absent_ (Fig._1B);
(32) On_abdominal_tergum_II_and_III,_setae_m3 present absent;
(33) On_abdominal_tergum_II_and_III, tegumentary_granulation normal
coarse_secondary_granulation_near_setae_s;
(34) On_abdominal_tergum_IV,_setae_m1 present_ (Fig._1A)
absent_ (Fig._1B);
(35) On_abdominal_tergum_IV,_setae_m2 present absent;
(36) On_abdominal_tergum_IV,_setae_m3 present_ (Fig._1A)
absent_ (Fig._1B);
(37) On_abdominal_tergum_IV,_setae_m3' present absent;
(38) On_abdominal_tergum_IV,_setae_p5 present_ (Fig._1A)
absent_ (Fig._1B);
(39) On_abdominal_tergum_V,_setae_a3 present_ (Fig._1A) absent_ (Fig._1B);
(40) On_abdominal_tergum_V,_setae_s
in_p3_position_i.e._p2 present_ (Fig._1A)
in_p2_position_i.e._p2 absent_ (Fig._1B);
(41) Ventral_tube_with_more_than_4 +_4_setae with_4_1_4_setae;
(42) On_abdominal_sternum_II,_setae_a3 present absent;
{43 On abdominal sternum IV, _setae_a1 present_ (Fig. 3A)
absent_ (Fig. 3B);

{44 On abdominal sternum IV, 4_rows_of_setae_ (Fig. 3A)
3_rows_of_setae_ (Fig. 3B);

{45 On abdominal sternum IV, _setae_m1 present absent;

{46 Furca present absent;

{47 On anal lobes, _setae_e present_ (Fig. 4A) absent_ (Fig. 4B);

{48 On anal lobes, _setae_z present_ (Fig. 4A) absent_ (Fig. 4B);

{49 On anal lobes, _distal_setae_of_hr_group present_ (Fig. 4A)
absent_ (Fig. 4B);

{50 Anal spines present absent;

{51 Pigmentation present absent;

;
The following file willemia2011_char.ss is the new matrix with 5 Willemia species added along with the character list:

xread
'Matrix of Ponge et al based on DHaese 2000 (Is psammophily an evolutionary dead end? Cladistics) with new species added'

Hypogastrura_vernalis

Orogastrura_dilatata

Xenyllogastrura_octoculata

Willemia_annapurna

Willemia_anophthalma

Willemia_arenicola

Willemia_nosyboraha

Willemia_arida

Willemia_bedosae

Willemia_bellingeri

Willemia_brevispina

Willemia_buddenbrocki

Willemia_christianseni

Willemia_deharvengi

Willemia_denesi

Willemia_dubia

Willemia_elisabethum

Willemia_fjellbergi

Willemia_unispina

Willemia_granulata

Willemia_intermedia

Willemia_iztacchiuatlensis

Willemia_japonica

Willemia_koreana

Willemia_meybholae

Willemia_multilobata

Willemia_nadchatrami

Willemia_namibiae
Willemia_neocaledonica
Willemia_neopapillata
Willemia_nepalensis
Willemia_persimilis
Willemia_psamaphila
Willemia_bulbosa
Willemia_acantha
Willemia_scandinavica
Willemia_setonychia
Willemia_shanghaiensis
Willemia_similis
Willemia_subbulbosa
Willemia_tali
Willemia_trilobata
Willemia_trisphaerae
Willemia_virae
Willemia_wandae

ccode - 0.51;
proc;

(0 On antennial_segment_IV, apical_vesicle large small absent;
1 Sensilla_e1_on_antennal_segment_IV subcylindrical_ (Figs. _2A_and_2D)
candle_flame-shaped_ (Fig. _2B) globular_ (Fig. _2C)
not differentiated_from_ordinary_setae;
2 Sensilla_e2_on_antennal_segment_IV subcylindrical_ (Figs. _2A_and_2D)
candle_flame-shaped_ (Fig. _2B) globular_ (Fig. _2C)
not differentiated_from_ordinary_setae;
3 Sensilla_e3_on_antennal_segment_IV subcylindrical_ (Fig. _2A)
candle_flame-shaped_ (Fig. _2B) globular_ (Fig. _2C)
globular_in_cavity_ (Fig. _2D);
4 Sensilla_d_on_antennal_segment_IV subcylindrical_ (Figs. _2A_and_2D)
candle_flame-shaped not differentiated_from_ordinary_setae_ (Fig. _2C);
5 Sensilla_i1_on_antennal_segment_IV subcylindrical_ (Figs. _2A_and_2D)
candle_flame-shaped_ (Fig. _2B)
not differentiated_from_ordinary_setae_ (Fig. _2C);
6 Sensilla_i2_on_antennal_segment_IV subcylindrical_ (Fig. _2A)
candle_flame-shaped_ (Fig. _2B) globular_ (Fig. _2C)
globular_in_cavity_ (Fig. _2D);
7 Small_internal_sensilla_of_sensory_organ_of_antennal_segment_III_ (Figs.
  _2A_to_2D) free covered_in_part_by_a_tegumental_fold;
8 Third_and_fourth_antennal_segments clearly separated
ventrally_slightly_fused;
9 Antennal_segment_II_with_12_setae 11_setae;
10 Antennal segment I with 7 setae 6 setae;
11 Prelabral chaetotaxy with 4 setae 2 setae;
12 Labral chaetotaxy with 5.5.4 setae 4.5.4 setae 2.5.4 setae
5.3.4 setae;
13 Seta a0 on the head present (Fig. 1A) absent (Fig. 1B);
14 Number of setae d on the head: 5 + 5 4 + 4 (Figs. 1A and 1B);
15 Number of setae v on the head: 2 + 2 1 + 1 (Figs. 1A and 1B);
16 Setae c1 on the head present absent;
17 Conules present absent;
18 Postantennal organ with 4 to 9 vesicles (Fig. 1A) 10 to 15 vesicles
more than 15 vesicles (Fig. 1B);
19 Setae al of thoracic tergum III positioned on thoracic tergum III
between thoracic terga II and III (Figs. 1A and 1B);
20 On second and third thoracic terga, setae al present absent;
21 On thoracic terga II and III, setae a2 present absent;
22 On thoracic terga II and III, setae m3 present (Fig. 1A)
absent (Fig. 1B);
23 Tibiotarsi I and II with 10 setae 12 setae 11 setae;
24 Tibiotarsus III with 18 setae 16 setae 12 setae 11 or 10 setae;
25 Tibiotarsi with tentent hair without tenent hair;
26 On thoracic tergum III, setae s (m7) subcylindrical and acuminate
lanceolate;
27 On abdominal tergum II and III, setae s
subcylindrical and acuminate (Fig. 1A) lanceolate (Fig. 1B);
28 On abdominal tergum IV, setae s subcylindrical and acuminate
lanceolate;
29 On abdominal tergum V, setae s subcylindrical and acuminate
lanceolate;
30 On abdominal tergum II and III, setae a2 present absent;
31 On abdominal tergum II and III, setae m2 present (Fig. 1A)
absent (Fig. 1B);
32 On abdominal tergum II and III, setae m3 present absent;
33 On abdominal tergum I II and III, tegumentary granulation normal
coarse secondary granulation near setae s;
34 On abdominal tergum IV, setae m1 present (Fig. 1A)
absent (Fig. 1B);
35 On abdominal tergum IV, setae m2 present absent;
36 On abdominal tergum IV, setae m3 present (Fig. 1A)
absent (Fig. 1B);
37 On abdominal tergum IV, setae m3' present absent;
38 On abdominal tergum IV, setae p5 present (Fig. 1A)
absent (Fig. 1B);
39 On abdominal tergum V, setae a3 present (Fig. 1A) absent (Fig. 1B);
40 On abdominal tergum V, setae s
in p3 position i.e. p2 present (Fig. 1A)
in p2 position i.e. p2 absent (Fig. 1B);
41 Ventral tube with more than 4 + 4 setae with 4.1 4 setae;
42 On abdominal sternum II, setae a3 present absent;
43 On abdominal sternum IV, setae a1 present (Fig. 3A)
absent (Fig. 3B);
44 On abdominal sternum IV, 4 rows of setae (Fig. 3A)
3 rows of setae (Fig. 3A);
45 On abdominal sternum IV, setae m1 present absent;
46 Furca present absent;
47 On anal lobes, setae e present (Fig. 4A) absent (Fig. 4B);
48 On anal lobes, setae z present (Fig. 4A) absent (Fig. 4B);
49 On anal lobes, distal setae of hr group present (Fig. 4A)
absent (Fig. 4B);
1266 (50 Anal_spines present absent;
1267 (51 Pigmentation present absent;
1268 ;
1269
1270
The following file willemia2011.ss is the same data matrix but without the character list to be directly readable by POY:

```plaintext
xread
'Matrix of Ponge et al based on DHaese 2000 (Is psammophily an evolutionary dead end? Cladistics) with new species added'

Hypogastrura_vernalis
```

<table>
<thead>
<tr>
<th>Character</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willemia_annapurna</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_anophthalma</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_arenicola</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_nosyboraha</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_bedosae</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_bellingeri</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_brevispina</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_buddenbrocki</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_christianseni</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_deharvengi</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_delamarei</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_denisi</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_dubia</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_elisabethum</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_fjellbergi</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_unispina</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_granulata</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_intermedia</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_iztacchuatlensis</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_japonica</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_koreana</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_meybholae</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_multilobata</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Willemia_nadchartrami</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Willemia_namibiae</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```
Willemia_neocaledonica
1003003111113111121001221010001101111111111111111111

Willemia_nepalensis
1003003111113111101210012211100011011111111111111111

Willemia_persimilis
0300220111?120110121010??101001110111110110100111001

Willemia_psammophila
0000000111010111000331111101000101?01100001?0001

Willemia_bulbosa 0322222111?1201101[0-
1]1010??10100111011110110100111001

Willemia_acantha 0300220111?120110121010??101001110111110110100111011

Willemia_scandinavica
10022011000101101010001110?000110110100010000100001

Willemia_setonychia
200300311?11101310????1111101111111111??11111

Willemia_shangaiensis
10022011????????101000??100000110000??011????1????1

Willemia_similis 10000001100110110100110001101000100101000100001

Willemia_subbulbosa
1222222111120110100002310001110101101010101101

Willemia_tali 1111?01100010110101000111110011110110000110001

Willemia_trilobata
00000001110011010210000000100110000?1001????101

Willemia_trisphaerae
131122110100111?101000111010100110000111011101100111

Willemia_virae 1000000110001101010100011101000110111100112000100001

Willemia_wandae 1003003111113111121001331100011011111111111111

ccode - 0.51;
proc/;
The following file will.sh is the main analysis script for POY:

```bash
wipe ()
read ("willemia2011.ss")
set (root: "Hypogastrura_ernalis")
build (1000)
select (unique)
perturb (iterations:2, ratchet:(0.15,3), swap(tbr, trees:2))
select (unique)
fuse ()
select (unique)
swap (threshold:20)
swap (all, visited:"will-brel.trees", timeout:3600)
select ()
report ("will.tre", trees:(total))
report ("will_results.ss", phastwinclad, trees:(hennig, total))
report ("will_cons.tre", consensus, "will-cons", graphconsensus)
report ("will-stats.txt", treestats, ci, ri)
report ("will-diag.txt", diagnosis)
extit ()
```
will-cons.pdf (consensus in pdf format), will-cons.tre (consensus in text format), will-stats (tree length ci and ri), and will.tre (the six equally parsimonious trees) are the results of the analysis.
Strict Majority Consensus Tree

Trees Found:

<table>
<thead>
<tr>
<th>Tree length</th>
<th>Number of hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>154</td>
<td>6</td>
</tr>
</tbody>
</table>

CI Statistics:

<table>
<thead>
<tr>
<th>Tree Cost CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>46.1038961039</td>
</tr>
</tbody>
</table>

RI Statistics:

<table>
<thead>
<tr>
<th>Tree Cost RI</th>
</tr>
</thead>
<tbody>
<tr>
<td>82.3404255319</td>
</tr>
</tbody>
</table>
[Hypogastrura vernalis, (Orogastrura dilatata, (Xenyllogastrura octoculata, (Willemia multilobata, Willemia bellingeri), (Willemia trilobata, Willemia namibiae), (Willemia similis, (Willemia christianseni, Willemia anopthalmal, (Willemia dubia, (Willemia scandinavica, Willemia koreana), (Willemia iztaccihuatensis, (Willemia bedosae, (Willemia unispina, Willemia psammophila), (Willemia virae, Willemia tali))))))]), (Willemia intermedia, (Willemia shanghaiensis, (Willemia granulata, (Willemia fjellbergi, Willemia arida), (Willemia trisphaerae, Willemia bulbosa, (Willemia persimilis, Willemia acantha))))), (Willemia nepalensis, (Willemia neocaledonica, (Willemia nadchatrami, Willemia brevispina))),(Willemia annapurna, (Willemia buddenbrocki, (Willemia wandae, (Willemia nosyboraha, (Willemia delamarei, (Willemia setonychia, Willemia deharvengi)))))))]

[154.]

(Hypogastrura vernalis, (Orogastrura dilatata, (Xenyllogastrura octoculata, (Willemia multilobata, Willemia bellingeri), (Willemia trilobata, Willemia namibiae), (Willemia similis, (Willemia christianseni, Willemia anopthalmal, (Willemia dubia, (Willemia scandinavica, Willemia koreana), (Willemia iztaccihuatensis, (Willemia bedosae, (Willemia unispina, Willemia psammophila), (Willemia virae, Willemia tali))))))]), (Willemia intermedia, (Willemia shanghaiensis, (Willemia granulata, (Willemia fjellbergi, Willemia arida), (Willemia trisphaerae, Willemia bulbosa, (Willemia persimilis, Willemia acantha))))), (Willemia nepalensis, (Willemia neocaledonica, (Willemia nadchatrami, Willemia brevispina))),(Willemia annapurna, (Willemia buddenbrocki, (Willemia wandae, (Willemia nosyboraha, (Willemia delamarei, (Willemia setonychia, Willemia deharvengi)))))))]

[154.]

(Hypogastrura vernalis, (Orogastrura dilatata, (Xenyllogastrura octoculata, (Willemia multilobata, Willemia bellingeri), (Willemia trilobata, Willemia namibiae), (Willemia similis, (Willemia christianseni, Willemia anopthalmal, (Willemia dubia, (Willemia scandinavica, Willemia koreana), (Willemia iztaccihuatensis, (Willemia bedosae, (Willemia unispina, Willemia psammophila), (Willemia virae, Willemia tali))))))]), (Willemia intermedia, (Willemia shanghaiensis, (Willemia granulata, (Willemia fjellbergi, Willemia arida), (Willemia trisphaerae, Willemia bulbosa, (Willemia persimilis, Willemia acantha))))), (Willemia nepalensis, (Willemia neocaledonica, (Willemia nadchatrami, Willemia brevispina))),(Willemia annapurna, (Willemia buddenbrocki, (Willemia wandae, (Willemia nosyboraha, (Willemia delamarei, (Willemia setonychia, Willemia deharvengi)))))))]

[154.]

(Hypogastrura vernalis, (Orogastrura dilatata, (Xenyllogastrura octoculata, (Willemia multilobata, Willemia bellingeri), (Willemia trilobata, Willemia namibiae), (Willemia similis, (Willemia christianseni, Willemia anopthalmal, (Willemia dubia, (Willemia scandinavica, Willemia koreana), (Willemia iztaccihuatensis, (Willemia bedosae, (Willemia unispina, Willemia psammophila), (Willemia virae, Willemia tali))))))]), (Willemia intermedia, (Willemia shanghaiensis, (Willemia granulata, (Willemia fjellbergi, Willemia arida), (Willemia trisphaerae, Willemia bulbosa, (Willemia persimilis, Willemia acantha))))), (Willemia nepalensis, (Willemia neocaledonica, (Willemia nadchatrami, Willemia brevispina))),(Willemia annapurna, (Willemia buddenbrocki, (Willemia wandae, (Willemia nosyboraha, (Willemia delamarei, (Willemia setonychia, Willemia deharvengi)))))))]

[154.]

(Hypogastrura vernalis, (Orogastrura dilatata, (Xenyllogastrura octoculata, (Willemia multilobata, Willemia bellingeri), (Willemia trilobata, Willemia namibiae), (Willemia similis, (Willemia christianseni, Willemia anopthalmal, (Willemia dubia, (Willemia scandinavica, Willemia koreana), (Willemia iztaccihuatensis, (Willemia bedosae, (Willemia unispina, Willemia psammophila), (Willemia virae, Willemia tali))))))]), (Willemia intermedia, (Willemia shanghaiensis, (Willemia granulata, (Willemia fjellbergi, Willemia arida), (Willemia trisphaerae, Willemia bulbosa, (Willemia persimilis, Willemia acantha))))), (Willemia nepalensis, (Willemia neocaledonica, (Willemia nadchatrami, Willemia brevispina))),(Willemia annapurna, (Willemia buddenbrocki, (Willemia wandae, (Willemia nosyboraha, (Willemia delamarei, (Willemia setonychia, Willemia deharvengi)))))))]

[154.]
will_results.ss is the data matrix with the resulting optimal trees readable by Winclada (obtained by the will_results.sh script run in POY):

```
xread 'Generated by POY 4.0' 52 45
Willemia_wandae
Willemia_virae
Willemia_trisphaerae
Willemia_trilobata
Willemia_tali
Willemia_subbulbosa
Willemia_similis
Willemia_shanghaiensis
Willemia_setonychia
Willemia_scandinavica
Willemia_acantha
Willemia_bulbosa
Willemia_psammophila
Willemia_persimilis
Willemia_neopseudosetosa
Willemia_nepalensis
Willemia_nadchatrami
Willemia_multilobata
Willemia_meybholae
Willemia_koreana
Willemia_japonica
Willemia_iztaccihuatlensis
Willemia_intermedia
Willemia_unispina
Willemia_fjellbergi
Willemia_deharvengi
Willemia_delamarei
Willemia_denisi
Willemia_christianseni
Willemia_buddenbrocki
Willemia_christianseni
```
Willemia_brevispina
Willemia_bellingeri
Willemia_bedosae
Willemia_arida
Willemia_nosyboraha
Willemia_arenicol
Willemia_anophthalma
Willemia_annapurna
Xenyllogastrura_octoculata
Orogastrura_dilatata
Hypogastrura_vernalis
cc - 0.51;
proc /;
#
$;

cc - 0.51;

tread (Hypogastrura_vernalis (Orogastrura_dilatata (Xenyllogastrura_octoculata (((Willemia_multilobata Willemia_bellingeri) ((Willemia_trilobata Willemia_namibiae) ((Willemia_similis Willemia_christianseni Willemia_anophthalma) ((Willemia_dubia (Willemia_scandinavica Willemia_koreana) ((Willemia_virae Willemia_unispina Willemia_psammophila) (Willemia_bedosae ((Willemia_unispina Willemia_psammophila) (Willemia_virae Willemia_unispina) (Willemia_intermedia (Willemia_granulata ((Willemia_fjellbergi Willemia_arida) (Willemia_trisphaerae (Willemia_elisabethum Willemia_denisi) (Willemia_japonica (Willemia_subbulbosa (((Willemia_persimilis Willemia_acantha) (Willemia_bulbosa (Willemia_meybholae Willemia_arenicola))) (Willemia_nepalensis (Willemia_neocaledonica (Willemia_nadchatrami Willemia_brevispina)) (Willemia_annapurna (Willemia_buddenbrocki (Willemia_wandae (Willemia_nosyboraha (Willemia_delamarei (Willemia_setonychia (Willemia_deharvengi))))))))))))))))))))))))))))))))) * (Hypogastrura_vernalis (Orogastrura_dilatata (Xenyllogastrura_octoculata (((Willemia_multilobata Willemia_bellingeri) ((Willemia_trilobata Willemia_namibiae) ((Willemia_similis Willemia_christianseni Willemia_anophthalma) ((Willemia_dubia (Willemia_scandinavica Willemia_koreana) ((Willemia_virae (Willemia_nosyboraha (Willemia_delamarei (Willemia_setonychia (Willemia_deharvengi))))))))))))))))))))))))))))))))))))))))))))))))))))))))) * (Hypogastrura_vernalis (Orogastrura_dilatata (Xenyllogastrura_octoculata (((Willemia_multilobata Willemia_bellingeri) ((Willemia_trilobata Willemia_namibiae) ((Willemia_similis Willemia_christianseni Willemia_anophthalma) ((Willemia_dubia (Willemia_scandinavica Willemia_koreana) ((Willemia_virae (Willemia_nosyboraha (Willemia_delamarei (Willemia_setonychia (Willemia_deharvengi)))))))))))))))))))))))))))))))))))))))))))))))))))))) * (Hypogastrura_vernalis (Orogastrura_dilatata (Xenyllogastrura_octoculata (((Willemia_multilobata Willemia_bellingeri) ((Willemia_trilobata Willemia_namibiae) ((Willemia_similis Willemia_christianseni Willemia_anophthalma) ((Willemia_dubia (Willemia_scandinavica Willemia_koreana) ((Willemia_virae (Willemia_nosyboraha (Willemia_delamarei (Willemia_setonychia (Willemia_deharvengi))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))}}}
Willemia_acantha) (Willemia_bulbosa (Willemia_meybholae
Willemia_arenicola)) ((Willemia_nepalensis (Willemia_nosyboraha
Willemia_delamarei)) (Willemia_setonychia Willemia_deharvengi)))
(Willemia_annapurna (Willemia_buddenbrocki (Willemia_wandae
Willemia_neocaledonica (Willemia_nadchatrami
Willemia_brevispina))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))))}
bremer_w.sh and jackknife_w.sh are the scripts for bremer and jackknife calculation respectively; willone.tre file contains one of the six optimal tree for the purpose of these calculations;

bremer_w.sh:

```bash
wipe ()
read ("willemia2011.ss")
set (root: "Hypogastrura_vernalis")
read ("willone.tre")
report ("bremerw.txt", supports:bremer:"will-brel.trees")
report ("bremerw", graphsupports:bremer:"will-brel.trees")
exit ()
```

jackknife_w.sh:

```bash
wipe ()
read ("willemia2011.ss")
set (root: "Hypogastrura_vernalis")
read ("willone.tre")
calculate_support(jackknife:(remove:0.50, resample:1000), build(),
swap(tbr, trees:5))
report ("jackknife_w", graphsupports:jackknife)
exit ()
```

will-one.tre:

```
(Hypogastrura_vernalis,(Orogastrura_dilatata,(Xenyllogastrura_octoculata
,(Willemia_multilobata,Willemia_bellingeri),((Willemia_trilobata
,Willemia_namibiae),((Willemia_similis,((Willemia_christianseni
,Willemia_anophthalma),((Willemia_dubia,(Willemia_scandinavica
,Willemia_koreana)),(Willemia_iztaccihuatlensis,(Willemia_bedosae
,(Willemia_unispina,Willemia_psammophila),((Willemia_virae,Willemia_tal
i(result)))
),(Willemia_intermedia,(Willemia_shanghaiensis,(Willemia_granulata
,((Willemia_fjellbergi,Willemia_arida),((Willemia_trisphaerace
,(Willemia_elisabethum,Willemia_denisi))),(Willemia_japonica
,((Willemia_subbulbosa,((Willemia_persimilis,Willemia_acantha
,(Willemia_bulbosa,(Willemia_meybholae,Willemia_arenicola))
,(Willemia_nepalensis,(Willemia_neocaledonica,((Willemia_nadchatrami
,Willemia_brevispina))),(Willemia_annapurna,(Willemia_buddenbrocki
,(Willemia_wandae,(Willemia_nosyboraha,(Willemia_delamarei
,(Willm_ea_setonychia,Willemia_deharvengi))))))))))))))))))))))))))){154.};
```
Figure Appendix S1 shows the strict consensus of 6 most parsimonious trees with a tree length of 154 steps, with jackknife and Bremer indices indicated at each node. Consistency Index is 46 and Retention Index is 82. All trees, jackknife, statistics and Bremer supports can be found in Electronic Appendix S1c. The addition of the 5 new *Willemia* species did not change the general topology found in D’Haese (2000). The monophyly of the genus *Willemia* was confirmed and strongly supported (jackknife resampling percentage of 100% and Bremer support value 3). The *anophthalma*, *denisi* s.str. and *buddenbrocki* groups were retrieved monophyletic again and the *denisi* group sensu Potapov is still paraphyletic (with the addition of *W. shanghaiensis* between *W. intermedia* and *W. granulata* in a paraphyletic assemblage). The new species (for the analysis) *W. tali*, *W. izardcuatlensis*, *W. psammophila* and *W. unispina* belong to the *anophthalma* group. *W. acantha* is sister group to *W. persimilis*, in the same group, *W. bulbosa* is sister group to *W. arenicola + W. meybhola*. Lastly, *W. nosyboraha* is included in the *buddenbrocki* group. Table 1 and Table in Appendix S2e show biogeographic and ecological attributes of species, respectively. In these tables species were arranged in the order given by the phylogeny.

REFERENCES

**Figure Appendix S1.** Strict majority consensus of six phylogenetic trees of the genus *Willemia*, with three outgroup species (see text for more details). Jacknife resampling percentages and Bremer support values are indicated above and below each node, respectively.
Appendix S2: Literature research: definitions, references, exploration of biases

Appendix S2a – Definitions of factors of harshness

- ‘xeric’ was defined as an environment displaying permanent or seasonal aridity: deserts (whether warm or cold), matorrals, sand dunes with poor vegetation cover, karst landscapes and dry Mediterranean areas.

- ‘hydric’ was defined as an environment displaying excess of water and associated lack of soil atmosphere: waterlogged soils, flooded areas.

- ‘arctic’ was defined as an environment displaying a long period of frost, generally more than six months, reflecting a latitudinal temperature gradient: polar, arctic and subarctic environments.

- ‘alpine’ was defined as an environment displaying a long period of frost, generally more than six months, reflecting an altitudinal temperature gradients: high mountains, including alpine and subalpine climates, above 1,000 m in nordic countries, above 2,000 m in temperate areas, above 3,000 or 4,000 m in tropical areas.

- ‘acid’ was defined as a soil at pH_{water} less than 5, where free aluminium and high phenolic concentrations are likely to occur: raw humus, podzols.

- ‘metallic’ was defined as a soil polluted with heavy metals: areas in the vicinity of smelter or mining activities.

- ‘saline’ was defined as a soil with a high osmotic pressure due to a high charge in salts (beaches and salt deserts).

Every species could be assigned to several harsh habitats according to (i) multiple records of the same species in a variety of habitats, or (ii) habitats belonging to several harshness.
categories. For instance, over the whole array of literature consulted, the same species can be recorded both in waterlogged and arid soils or the same habitat can be ‘arctic’ and ‘acid’, but only when this was indicated. In our census we did not try to discern whether some factors were dominant above others in a given record unit. Rather we considered them as independent causes of harshness, with which species must cope within a given habitat throughout their evolutionary history (Jablonski, 2008).

Caves were excluded given that (i) they constitute refuges where more sensitive species can live in a more buffered environment, whatever the nature of the surrounding environment, (ii) there was no evidence they were detrimental to growth and reproduction of most soil invertebrate species (Moseley, 2007), and (iii) some *Willemia* species were reported to occur in caves (some of them were even described for the first time in caves) but they were also found in other environments, suggesting the absence of specialisation for cave life in this genus (Christiansen, 1965). Psammophily sensu D’Haese (2000), i.e. life in sand, was partitioned in ‘saline’ and ‘xeric’ categories according to whether sea water or drought was the prevailing factor. For instance seashore dunes were considered ‘xeric’ but not ‘saline’ while the intertidal zone was considered ‘saline’ but not ‘xeric’. Experimental studies focusing on individual *Willemia* species as well as on entire soil communities were also taken into account. The ‘harshness breadth’ index was estimated for every species by the number of harsh environments in which the species was recorded, thus scaling from 0 (no harsh environment recorded for the species) to 7 (all harsh environments recorded).

Obviously, no place on earth will show all types of harshness. This is why we considered multiple types of harshness, and why we quantified the harshness breadth as the number of harshness types occupied by a given species. Moreover, we tested whether occupation of different types of harshness is positively or negatively correlated, i.e. whether species that can tolerate certain kinds of harsh environments tend to be able to tolerate a broad
range of harsh environments (except those affected by salinity). We found that occupations of
all but one type of harshness are positively correlated (Results). This means that a species that
is found in environments that are harsh in one type tend to be found under different types of
harsh environments, too. Inversely, other species occupy environments lacking any of the
types of harshness. This result justifies talking broadly about “species using harsh
environments” as we repeatedly do it.

The largest possible corpus of literature was examined for the assessment of use of harsh environments in *Willemia* species. As defined above (Introduction), harsh environments are those which limit the rate of growth and reproduction of most species, except a few ones tolerating it. This does not mean that harsh environments are poorer in Collembolan individuals, since better adapted species may benefit from the alleviation of competition caused by the resulting decrease in species richness and thereby may locally proliferate (Usher, 1985). A total of 248 references were used for the census (Appendix S2b). Languages available to us were English, French, German, Italian, Portuguese and Spanish, embracing most literature published in books and scientific journals. Other languages (Russian, Japanese, Hungarian) were cautiously discarded, in order to avoid misinterpretation, but they comprised less than 1% of published literature. When information about soil or climate was not given for a record, but the plant community was sufficiently described in the vicinity of the sample, then vegetation was used as a basis for estimating the harshness of the environment: in several instances dominant ericaceous or coniferous vegetation was considered as indicative of soil acidity given the recalcitrance of its litter and its known acidifying influence on the environment occupied by *Willemia* (Ponge, 2000), while grassland species were indicative of more fertile and less acid soils (Miles, 1985; Falkengren-Grerup, 1986). In two instances, the geographic locality was sufficiently precise and was used to derive environmental conditions
prevailing at the sampling site. In the absence of any clue on environmental conditions, records were disregarded.

The above classification might miss species that are able to use harsh environments but did not happen to encounter any. However, given that we accounted for a large range of harsh environments, species in all regions should have encountered at least some of them (and use of one kind of harsh environment turn out to be positively correlated to the use of all others in most cases, see Results). Not occurring in harsh environments is hence true evidence for not being able to use them.

REFERENCES


### 1999 Appendix S2b - References used for the assessment of use of harsh environments in *Willemia*

2000 species.

2001

<table>
<thead>
<tr>
<th>Species</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>acantha</td>
<td>1 Bonet (1945)</td>
</tr>
<tr>
<td>arida</td>
<td>2 Fjellberg (1991), Addison et al. (2006)</td>
</tr>
<tr>
<td>bellerina</td>
<td>1 Palacios-Vargas &amp; Vázquez (1988)</td>
</tr>
<tr>
<td>bulbosa</td>
<td>2 Bonet (1945), Cuzh-Pool et al. (2007)</td>
</tr>
<tr>
<td>christianseni</td>
<td>1 d'Haese (1998)</td>
</tr>
<tr>
<td>dehavengi</td>
<td>1 d'Haese &amp; Weiner (1998)</td>
</tr>
<tr>
<td>delamarei</td>
<td>1 Prabhuo (1971)</td>
</tr>
</tbody>
</table>

2002

REFERENCES

2003

2003


Arbea, J.I., Jordana, R., 1985b. Estudio ecológico de la colembofauna de los suelos del Macizo de Quinto Real (Pirineos Occidentales) y descripción de dos nuevas especies: Anurida flagellata n. sp. y. Onychiurus subedinensis n. sp. (Insecta, Collembola). Boletín de la Estación Central de Ecología 14, 57-80.


Hågvar, S., 1984. Ecological studies of microarthropods in forest soils, with emphasis on relations to soil acidity. Norwegian Forest Research Institute, Ås.


Kováč, L., Miklisová, D., 1995. Collembolan communities in winter wheat-clover cropping system on two different soil types. Polskie Pismo Entomologiczne 64, 365-381.


Palacios-Vargas, J.G., 1981. Collembola asociados a Tillandsia (Bromeliaceae) en el Derrame Lávico del Chichinautzin, Morelas, México. The Southwestern Entomologist 6, 87-98.


Appendix S2c – Exploration of bias due to differences between species in numbers of records in the literature.

Methods: Because species differed in the number of records from which harshness-breadth was estimated (see above Appendix S2b, column 2), we also verified whether the number of records was ‘phylogenetically conserved’. For this, we also used the root-skewness test with Euclidean distances among species. This test aims to highlight whether our results were biased by differences among species in the amount of knowledge on their environmental distribution. This approach is conservative: the intensive study of a species may increase the number of records in harsh environments (bias), but a species’ use of harsh environments may also increase the chance of being observed in many of the environments studied (non-bias).

Results: There was no phylogenetic structure in the number of literature records (root-skewness test P-value = 0.326) even after log-transformation (root-skewness test P-value = 0.291) indicating that the analyses done in the main text on the link between phylogeny and properties of species do not depend on the extent to which these species were studied in the literature.
We showed that in the genus *Willemia* tolerance to salinity (use of haline habitats) was negatively correlated to other types of harshness and that ‘haline’ species exhibited on average a higher clade rank than species tolerant of other harshness factors such as drought, frost, acidity, waterlogging (anaerobiosis) or heavy metals. In the absence of ecophysiological studies on the genus *Willemia*, which is probably explained by the small size of these animals and their concealed way of life (d’Haese, 2000), only putative arguments can be given for rejecting salinity as a harshness factor to which *Willemia* should adapt for life on the seashore. A strong attraction to sodium (compared to potassium) has been shown to occur in *Heteromurus nitidus*, another soil-dwelling springtail species (Salmon et al., 2002). Thibaud (2007), on the base of biogeographic arguments, postulated that transport by sea currents could be responsible for the occurrence of the same *Willemia* species on remote seashores, whether insular or continental. Witteveen *et al.* (1987) showed that the ionic concentration of the haemolymph of inland Collembola equilibrated rapidly with sea water when transferred to seashore conditions. It could be suggested that permanent contact with dilute saline solutions (seashores, but not dry saline environments) needs no special adaptation in basal Hexapoda, contrary to fully terrestrial insects (higher insects), for which potassium is an essential element which replaces sodium (Wigglesworth, 1965). The ancestry of life in sea water versus freshwater has been shown to be the rule in aquatic invertebrates (Lee & Bell, 1999). The separation we performed between tolerance to drought and to salinity, based on ecophysiological and biocenotic grounds (Hartmut *et al.*, 1995; Thibaud, 2007), and supported by studies on other animal groups (Gomez-Mestre & Tejedo, 2005), might explain why psammophily (life in sand), which confounds both adaptation to drought to salinity, exhibits a high degree of reversal along phylogenetic trees of the genus *Willemia* (d’Haese, 2000).


The above biogeographic classifications correspond to very general trends in environmental harshness. For species that are particularly sensitive to moisture deficits, such as the majority of Collembola, regions receiving a high amount of rainfall will on average be less harsh. For species that are not able to increase their body temperature above that of the immediate environment, such as again Collembola, regions of high temperatures will on average be less harsh. In that sense, the Inter-Tropical Convergence Zone, with a reasonably warm (cloudy) and moist climate, will be favorable to most sensitive species. Besides coldness and aridity, acid and metal-contaminated soils were also considered as harsh.
habitats. Indeed, most tropical soils are acid, at least they exhibit pH values less than 5. However, if we except areas degraded by human activities in the tropics (pasture, mining,…), tropical rain forests are not characterized by nutrient shortage, due to (i) high organic inputs from the overstory, (ii) a high rate of mineral weathering in the soil. Tropical soils are fragile but not poor in nutrients, at least in the top few centimeters where most soil animals (*Willemia* included) are living (see Dos Santos Neves *et al*., 2010; Brookshire *et al*., 2012). Obviously, tropical areas also include harsher environments (which have been coded as such in our database), more especially in mountains above 3000 m (the Andes), white sands and seashores. As explained in the Introduction, during much of the geological past, environments might have been on average harsher on Laurasia rather than on Gondwana paleocontinents (Vršanský, 2005; Crisp *et al*., 2010), among others reflecting the larger surface of landmasses of northern than of southern temperate regions which may lead to more climatic and thereby edaphic extremes in Laurasia and descendant land masses (except for Antarctica) (Chown *et al*., 2004).

**REFERENCES**


---

**Appendix 3 – Table.** Biogeographic provenance of *Willemia* species. Biogeographic provinces according to Christiansen & Bellinger (1995). Gondwana-inherited provinces (Australasia) are in bold type, Eurasia-inherited provinces in italic type, non-defined (Macaronesia and Central America) in normal type. Species belonging to both old continents were classified as ‘inherited provinces in italic type, non-inherited provinces in roman bold type, Eurasia-inherited provinces in italic type, non-defined (Macaronesia and Central America) in normal type. Species belonging to both old continents were classified as ‘inherited provinces in italic type, non-inherited provinces in roman bold type.’

| Species | 1 | 2a | 2b | 3a | 3b | 4 | 5 | 6 | 7a | 7b | 8 | 10 | 14 | 15 | 17 | 18 | 19 | 20 | 21 | 24a | 24b | 25 | 27 | 28 | 29 | 36 | Gondwana | Laurasia | Non-defined |
|---------|---|----|----|----|----|---|---|---|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| multilobata | √ | √ | √ | √ | | | | | | | | | | | | | | | | | | | | | | | | | |
| bellingeri | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| tribulata | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| namibiæ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| similis | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| christianseni | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| anopthalma | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| dubia | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| scandinavica | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| koreana | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| iztachuatlensis | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| bedosae | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| unsipina | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| psammophila | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| vireae | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| tali | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| intermediæ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| shanghaïensis | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| granulata | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| fjellbergi | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| arida | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| trisphaeræ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| elisabethæ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| denisi | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| japonica | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| subbulbosa | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| parsimilis | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| acantha | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Columns corresponding to tropical provinces are shaded.
<table>
<thead>
<tr>
<th>Species</th>
<th>2720</th>
<th>2721</th>
</tr>
</thead>
<tbody>
<tr>
<td>bulbosa</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>mayolholae</td>
<td></td>
<td></td>
</tr>
<tr>
<td>arenicola</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nepalensis</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>neocaledonica</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nadchatrami</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>brevispina</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>annapurna</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>buddingbrodki</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>wandiae</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>nosyboraha</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>delamarei</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>setonychia</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>deharvengi</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Appendix S3. Robustness of the analysis to variation in underlying phylogenetic trees

We have performed again all analyses with the 6 most parsimonious trees. The 6 trees are given below.

According to all these analyses, the results obtained with the 6 most parsimonious trees are equivalent to those obtained with the resulting tree used in the main text.
Phylogenetic conservatism in the use of harsh habitats

With all phylogenetic trees, the types of harshness under which species were found exhibited clear phylogenetic conservatism (root-skewness test P values varied from 0.001 to 0.003 for a statistic varying from 0.616 to 0.647, 1 meaning that the whole trait variation is at the root node). Also the number of types of harsh habitats used by a species was phylogenetically conserved (root-skewness test P values varied from 0.013 to 0.043, for a statistic varying from 0.623 to 0.660). When salinity was excluded from the array of harsh habitats a stronger conservatism was registered (root-skewness test P values varied from 0.001 to 0.019, for a statistic varying from 0.659 to 0.699). The presence/absence of *Willemia* species in harsh habitats (instead of their number used) exhibited an even clearer phylogenetic conservatism.
when salinity was excluded from the array of harsh habitat (root-skewness test with all P
values equal to 0.001, for a statistic varying from 0.754 to 0.786).

**Uses of most types of harsh habitats are positively correlated**

Phylogenetic principal components analysis (pPCA) showed strong positive correlations
between all types of harsh habitats, with the exception of salinity. Use of all other types of
harsh habitats scored distinctly positively on the same Principal Component (Fig. 1).
Appendix S3 Table. Ordination of correlations between uses of different types of harsh habitats, analysed by pPCA analysis (multivariate analysis based on phylogenetic distances).

The table gives coordinates of harsh habitat types along the first principal component for the 6 dichotomous phylogenies provided above. Note that except for salinity, uses of habitats of all types of harshness are positively correlated.

<table>
<thead>
<tr>
<th></th>
<th>Phy1</th>
<th>Phy2</th>
<th>Phy3</th>
<th>Phy4</th>
<th>Phy5</th>
<th>Phy6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xeric</td>
<td>-.343</td>
<td>-.398</td>
<td>-.396</td>
<td>-.414</td>
<td>-.426</td>
<td>-.395</td>
</tr>
<tr>
<td>Hydric</td>
<td>-.183</td>
<td>-.177</td>
<td>-.190</td>
<td>-.176</td>
<td>-.181</td>
<td>-.188</td>
</tr>
<tr>
<td>Arctic</td>
<td>-.389</td>
<td>-.395</td>
<td>-.380</td>
<td>-.420</td>
<td>-.398</td>
<td>-.373</td>
</tr>
<tr>
<td>Alpine</td>
<td>-.482</td>
<td>-.453</td>
<td>-.461</td>
<td>-.432</td>
<td>-.450</td>
<td>-.470</td>
</tr>
<tr>
<td>Acid</td>
<td>-.581</td>
<td>-.559</td>
<td>-.566</td>
<td>-.548</td>
<td>-.552</td>
<td>-.567</td>
</tr>
<tr>
<td>Metallic</td>
<td>-.201</td>
<td>-.198</td>
<td>-.197</td>
<td>-.202</td>
<td>-.190</td>
<td>-.195</td>
</tr>
<tr>
<td>Saline</td>
<td>.296</td>
<td>.312</td>
<td>.303</td>
<td>.305</td>
<td>.291</td>
<td>.299</td>
</tr>
</tbody>
</table>

Species using harsh habitats have lower clade ranks but are as numerous as species using mesic habitats.

We found that the clade rank of species decreased with species’ harshness breadth (PGLS; df=40, P values varied from 0.002 to 0.003 depending on the dichotomous phylogeny used). When salinity was excluded it appeared that the distance to the root of the phylogenetic tree was quite similar for all species using harsh habitats, independent of the number of harsh...
habitat types they were using, all of the harshness-tolerant species being placed in a basal position (Appendix S4 Figure).

Appendix S3 Figure. Clade ranks, i.e. nodal distances to the root, of species using different numbers of harsh habitat types. Error bar = S.E. Note that species that do not use any harsh habitat are at a higher distance from the root (see text for analyses). This analysis has been done for each dichotomous phylogenetic tree given above. Species having a single harsh habitat type have been separated into those affected by salinity (red) and those affected by another harsh habitat type (black).
Geographical distributions: Species using harsh habitats tend to be Laurasian, and as a statistical consequence Laurasians tend to have low clade ranks.

We found that species using harsh and mesic habitats were distributed differently across the globe. Species using harsh habitats tended to occur in present non-tropical regions, albeit the relationship was relatively weak (PGLS; df = 30; P values varying from 0.0004 to 0.0553). At the same time, species using harsh habitats were strongly restricted to former Laurasia regions (PGLS; df = 30; P values varying from 0.0002 to 0.0103).

Low clade-rank species were strongly restricted to present non-tropical regions (n = 31 in all tests of this paragraph, tests based on phylogenetic permutations, all P values equal 0.001 whatever the dichotomous phylogenetic tree used). This relationship was maintained when including harshness breadth as a co-variable: harshness breadth and presence in present-day tropical regions were both related to low clade rank (P values were equal to 0.001 for harshness breadth, and they varied from 0.001 to 0.009 for presence in present-day tropical regions depending on the phylogenetic tree used; conclusions were not impacted by the order in which the two explanatory variables were entered in the model: changing the order gave P varying from 0.034 to 0.051 for harshness breadth, and equal 0.001 for presence in present-day tropical regions). Therefore the high harshness-tolerance of non-tropical species did not explain their low clade rank. Species of low clade rank also tended to be restricted to former Laurasia regions (P varied from 0.005 to 0.013 depending on the phylogenetic tree used). When including first occurrence in former Laurasia and second harshness breadth, both variables were significant (P varied from 0.002 to 0.004 for occurrence in former Laurasia, and from 0.007 to 0.017 for harshness breadth). But this relationship disappeared when including first harshness breadth in the model: occurrence in former Laurasia was not significant (P varying 0.239 from to 0.294) while harshness breadth was significantly related...
to low clade rank (P varying from 0.001 to 0.002). Overall, this indicated that species using harsh environments are particularly bound to former Laurasia continents (above paragraph) and that - as a statistical consequence - we find many species of low clade rank on former Laurasia continents.
Reconstructions of ancestral states

The below graphs show the reconstructions of the three traits analysed for each of the six trees. Visual inspection shows that the different trees lead to the same conclusions on which trait states are ancestral and when they have changed throughout evolution.
A – use of harsh habitats  
Tree 3

B – tropical distribution

C – former-Gondwana distribution

A – use of harsh habitats  
Tree 4

B – tropical distribution

C – former-Gondwana distribution