Ecography

ECOG-02683
Eme, D., Zagmajster, M., Delić, T., Fišer, C., Flot, J.-F., Konecny-Dupré, L., Pálsson, S., Stoch, F., Zakšek, V., Douady, C. and Malard, F. 2017. Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. – Ecography doi: 10.1111/ecog.02683

Supplementary material

Appendix 1: COI data set.

COI data set for the 3028 specimens. MDSs: species hypotheses (SHs) delimited using morphology; TH: SHs delimited with the 16% COI divergence threshold; PTP (and their support): SHs delimited with the Poisson tree processes model; bPTP (and their support): SHs delimited with the Bayesian implementation of the Poisson tree processes model; Long.: longitude of the locality, in decimal degrees; Lat.: latitude of the locality, in decimal degrees; Dataset: "COI tree only" for sequences used only for species delimitation and "COI tree & Europe" for sequences used both for species delimitation and macroecological analysis. Range data: sequences used for the analysis of range size pattern.

The table is too long to be embedded as Appendix in the main document but it is available at the following address:

https://drive.google.com/open?id=0BxIuaQPA8MokbURaMzBjQ25mdmM

and this table will be available in Dryad if the paper is accepted.

Appendix 2: Supplementary tables and figures.

Table A1: Taxonomic comparison among the four sets of species hypotheses (SHs) generated using morphology (Morph.), a COI divergence threshold (TH), the Poisson tree processes model (PTP) and its Bayesian implementation (bPTP). Taxonomic inflation is defined as the number of SHs in set y divided by the number of SHs in set x. See text and Figure 2 for definitions of match, lump, split and reshuffling.

Dataset x		Dataset y		Taxonomic inflation	Shift from x to y				Shift from y to x			
Type	Number of SHs	Type	Number of SHs		Match	Lump	Split	Reshuffling	Match	Lump	Split	Reshuffling
Morph.	263	TH	519	1.97	134	1	348	36	134	69	21	39
Morph.	263	PTP	646	2.46	131	1	492	22	131	95	10	27
Morph.	263	bPTP	650	2.47	133	1	496	20	133	97	9	24
TH	519	PTP	646	1.24	428	1	214	3	428	83	3	5
TH	519	bPTP	650	1.25	425	2	221	2	425	86	4	4
PTP	646	bPTP	650	1.01	597	13	40	0	597	18	31	0

Table A2. Independent and shared contributions of historical climate variability (H), productive energy (E) and spatial heterogeneity (S) to variation in species richness of obligate groundwater crustaceans in Europe. . Ex_Var: total proportion of variance explained by the models (%). In abbreviations, colons denote shared variance between mechanisms. Variance partitioning is shown for the four different sets of species hypotheses delimited using morphology (Morph.), a COI divergence threshold (TH), the Poisson tree processes model (PTP) and its Bayesian implementation (bPTP). GLM.nb: generalized linear model with negative binomial error; OLS/GLS: ordinary/generalized least squares models.

Dataset	Model type	Ex_Var	Pure components			Shared components			
			H	E	S	H :E	E:S	H:S	H :S:E
Morph.	GLM.nb	28.54	2.02	13.45	4.51	5.44	7.04	-1.53	-2.39
	OLS	16.61	0.35	8.34	4.06	2.53	3.82	-0.99	-1.51
	GLS	18.11	0.53	8.41	4.66	3.11	3.67	-0.73	-1.55
TH	GLM.nb	27.54	0.66	11.56	5.47	4.15	8.83	-0.49	-2.65
	OLS	16.46	-0.36	8.51	4.02	1.75	4.10	-0.15	-1.40
	GLS	18.33	0.21	8.85	4.69	1.78	4.11	0.01	-1.31
PTP	GLM.nb	26.58	0.57	10.53	5.70	3.96	8.96	-0.44	-2.68
	OLS	16.44	-0.72	8.34	4.53	1.29	4.27	0.15	-1.41
	GLS	18.39	-0.10	8.91	5.29	1.13	4.02	0.25	-1.11
bPTP	GLM.nb	26.80	0.58	10.72	5.54	3.99	9.04	-0.47	-2.60
	OLS	16.32	-0.67	8.34	4.21	1.31	4.31	0.20	-1.37
	GLS	18.31	0.00	9.06	4.94	0.97	3.97	0.31	-0.94

Table A3: Independent and shared contributions of historical climate variability (H), habitat area/heterogeneity (S) and climate seasonality (S) to variation in median range size of obligate groundwater crustaceans in Europe. Ex_Var: total proportion of variance explained by the models (%). In abbreviations, colons denote shared variance between mechanisms. Variance partitioning is shown for the four different sets of species hypotheses delimited using morphology (Morph.), a COI divergence threshold (TH), the Poisson tree processes model (PTP) and its Bayesian implementation (bPTP). OLS/GLS: ordinary/generalized least squares models.

Dataset	Model type	Ex_Var	Pure components			Shared components			
			H	S	C	H:S	S:C	H:C	H:S:C
Morph.	OLS	19.42	15.325	0.144	-0.506	0.314	0.114	3.409	0.622
	GLS	18.19	15.244	-0.870	-1.411	-0.407	-0.165	4.139	1.660
TH	OLS	33.37	16.167	1.340	1.870	1.083	-1.469	11.726	2.651
	GLS	33.95	24.733	1.880	2.850	-7.956	-1.995	6.328	8.110
PTP	OLS	40.14	20.781	1.884	1.207	0.582	-0.995	13.610	3.073
	GLS	41.70	31.314	2.733	2.077	-9.742	-1.246	8.272	8.295
bPTP	OLS	40.90	19.766	2.406	1.444	1.146	-1.085	13.862	3.354
	GLS	42.78	28.734	3.532	2.553	-7.566	-1.209	11.489	5.243

Figure A1: Map showing the location of sampling sites and cells used for analyzing the determinants of species richness and median range size of groundwater crustaceans in Europe. The cell area of 0.9°-latitude cells was kept constant (10,000 km^2) all over the grid by adjusting the longitudinal divisions between adjacent cells in each latitudinal band (see Sastre, P. et al. 2009. A geoplatform for the accessibility to environmental cartography. – J. Biogeogr. 36: 568).

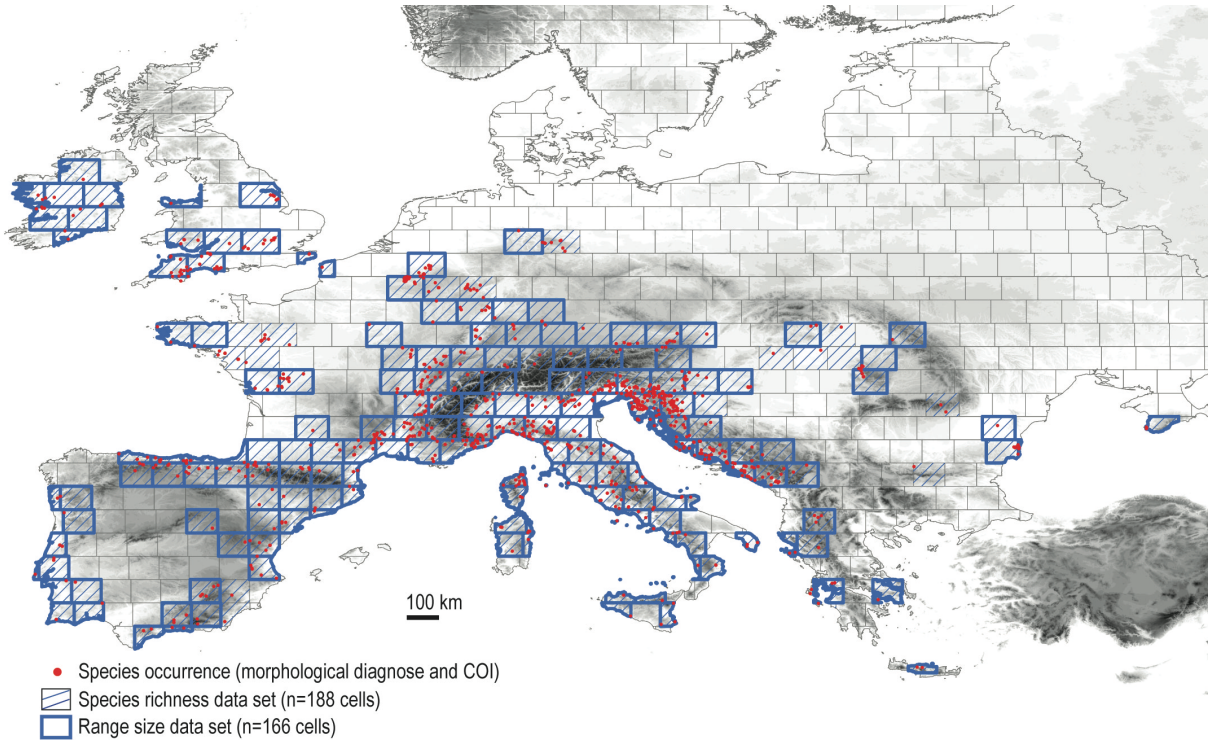


Figure A2: Plots of maximum linear extent (MLExt) (a) and latitudinal midpoints (b) of selected morphologically distinguishable species (MDSs, n=147) as measured from localities with COI sequences against the MLExt and latitudinal midpoints of MDSs as measured from occurrence data from the literature. The solid black line and broken red line in (a) represent the 1:1 equivalence line and the 0.75 interval line (i.e. the MLExt in y is 0.75 time the MLExt in x), respectively. Broken red lines in (b) represent the +/- 0.95 interval lines.

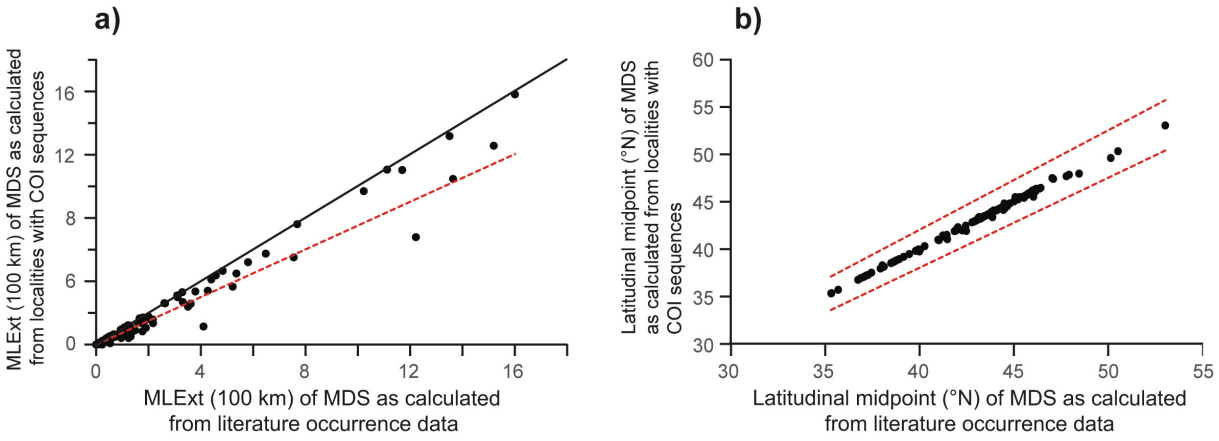


Figure A3: Pairwise relationships between species richness (log-transformed) and actual evapotranspiration (AET) and elevation range (n=188 cells). The red line represents the fit of an ordinary least squares model. Morphology: species hypotheses (SHs) delimited using morphology; TH: SHs delimited using a COI divergent threshold; PTP: SHs delimited using the Poisson tree processes; bPTP: SHs delimited using the Bayesian implementation of the Poisson tree processes.

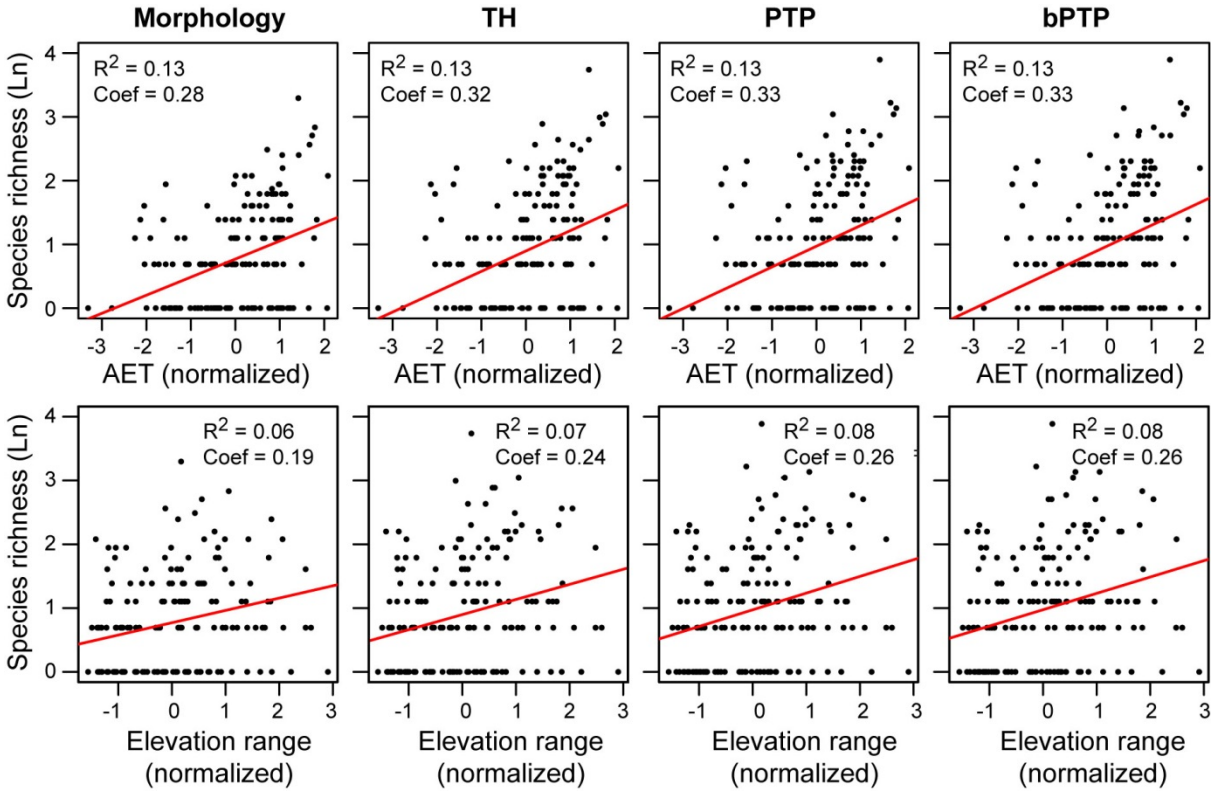
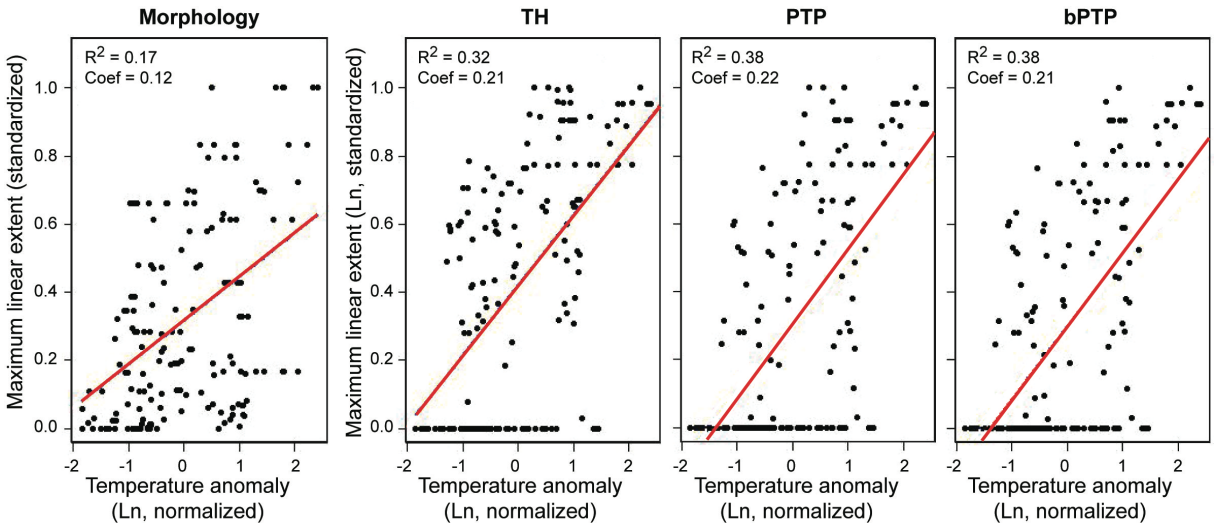


Figure A4: Pairwise relationships between maximum linear extent and temperature anomaly (n=166 cells). The red line represents the fit of an ordinary least squares model. Morphology: species hypotheses (SHs) delimited using morphology; TH: SHs delimited using a COI divergent threshold; PTP: SHs delimited using the Poisson tree processes; bPTP: SHs delimited using the Bayesian implementation of the Poisson tree processes.



Appendix 3: Supplementary methods.

Predictors of species richness

Spatial heterogeneity was represented by elevation range and habitat diversity. Elevation range is a proxy for topographic heterogeneity (Leprieur et al. 2011) and was calculated as the highest difference in elevation between any two locations in a cell using elevation data from the pan-European river and catchment database (Vogt et al. 2007). Habitat diversity was estimated for each cell using the Shannon's diversity index defined as $H' = -\Sigma\, p_i \times \ln p_i$, where p_i represented the areal proportion of 12 groundwater habitat types. The areal proportions of each habitat were computed from the European vector map of groundwater habitats (Cornu et al. 2013). To represent climate/productivity, we used mean annual actual evapotranspiration (AET) as an index of water-energy dynamics that subsumes ambient energy (temperature) and water availability (precipitation) (O' Brien, 2006). Data on AET were obtained from the 30 arc-second resolution world map released by Trabucco and Zomer (2010). To test for the effect of history, we computed temperature and precipitation anomalies, defined as the differences in mean annual temperature and annual precipitation between the present and the Last Glacial Maximum (LGM) (Leprieur et al. 2011). Data for the LGM were extracted from two global circulation models, CCSM and MIROC2 (Hijmans et al. 2005), and values from both were averaged to account for variation between models.

Predictors of median range size

For range size, we used seven predictors to test for the role of three influential mechanisms – habitat area/heterogeneity, climate seasonality and historical climate variability. Habitat area/heterogeneity was represented by elevation range, AET, aquifer area, and climatic rarity. Aquifer area was calculated using the European vector map of groundwater habitats as the

total area of aquifer available within a 1000-km radius around a cell (Cornu et al. 2013). To compute climatic rarity, we extracted and averaged for each grid cell data on mean annual temperature and all eight bioclimatic precipitation layers of the WorldClim dataset, except precipitation seasonality (30 arcsec resolution; Hijmans et al. 2005). A normalized principal components analysis (PCA) was then performed on grid-cell averaged data, log-transformed when appropriate to approximate a normal distribution. The first two PCA axes accounted for 86.8% of the climatic variance. They were used to calculate the average Euclidian distance in climatic space between each cell and all neighboring cells within a radius of 500 km. High average distance values correspond to cells showing rare climates relative to their neighboring cells (Morueta-Holme et al. 2013). For climate seasonality, we computed the per-cell-average of precipitation seasonality from the WorldClim dataset (Hijmans et al. 2005) as a surrogate of present intra-annual environmental variability because there is very little thermal seasonality in groundwater.

Statistical procedure to take into account spatial autocorrelation in the residuals of species richness and median range size models

We performed generalized least square (GLS) models in order to take into account spatial autocorrelation in the residuals of the OLS models (Beale et al. 2010). Using the OLS model including all predictors, we selected the best spatial variogram structure of the residuals and tested for the presence of a nugget effect using restricted maximum-likelihood optimization based on the AICc criterion (Zuur et al. 2009). The GLS model was fitted using maximum-likelihood optimization with the best spatial residual structure. Then, variance partitioning and multimodel inference were performed as described in the section "Determinants of spatial variation in species richness and median range size" of the Materials and Methods. As a measure of explained variance, we used pseudo-R^2, defined as the squared Pearson correlation

51 coefficient between the fitted and observed values of species richness (Ballesteros-Meija et al.
52 2013).
53
54 **REFERENCES**

55 Ballesteros-Mejia, L. et al. 2013. Mapping the biodiversity of tropical insects: species
56 richness and inventory completeness of African sphingid moths. – Global Ecol. Biogeogr.
57 22: 586–595.
58 Beale, C. M. et al. 2010. Regression analysis of spatial data. – Ecol. Lett. 31: 246–264.
59 Cornu, J.-F. et al. 2013. The distribution of groundwater habitats in Europe. – Hydrogeol. J.
60 21: 949–960.
61 Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land
62 areas. – Int. J. Climatol. 25: 1965–1978.
63 O' Brien, E. M. 2006. Biological relativity to water-energy dynamics. – J. Biogeogr. 33:
64 1868–1888.
65 Morueta-Holme, N. et al. 2013. Habitat area and climate stability determine geographical
66 variation in plant species range sizes. – Ecol. Lett. 16: 1446–1454.
67 Trabucco, A. and Zomer, R. J. 2010. Global soil water balance geospatial database. CGIAR
68 Consortium for Spatial Information. Published online. Available from: URL
69 http://www.cgiar-csi.org (Last accessed 15 January 2016).
70 Vogt, J. et al. 2007. A pan-European river and catchment database. EUR 22920 EN. European
71 Commission – Joint Research Centre, Luxembourg. Available from: URL
72 http://ccm.jrc.ec.europa.eu/php/index.php?action=view&id=23 (Last accessed 15 January
73 2016).
74 Zuur, A. F. et al. 2009. Mixed effects models and extension in ecology with R. Springer, New
75 York.

Appendix 4: Supplementary acknowledgments

The following persons and institutions contributed to the collection of specimens used in the present study: Ait Boughrous A., Alamichel F., ALES caving club, Alhama de Granada municipality, Altermatt F., Alther R., Amont B., Angyal D., Argano R., Ariano D., Arndt E., Arnuš U., Association Libanaise d'Etudes Spéléologiques, Augros X., Bahrdt S., Baković P., Balázs G., Baratti M., Bedek J., Bedjanič M., Bertochio P., Bertrand J.-Y., Besson J.-P., Bianco D., Bilandžija H., Blant M., Bodon M., Bogic B., Bollache L., Borissov I., Borko Š., Borreguero M., Bottazzi E., Bou C., Bouillon M., Boutin C., Bračko G., Brancelj A., Breuss W., Bric B., Bur M., Bürger K., Buser S., Caillon M., Calcagno M., Camacho A., Canale E., Candresse T., Caoduro G., Capderrey C., Carreira I., Carter J., Centa M., Chardon P., Chatelier B., Chatelier J., Chatelier N., Chiesi M., Christian E., Churcheward B., Cianfanelli S., Cianferoni F., Coculo L., Coineau N., Colson-Proch C., Comotti G., Creuzé des Châtelliers M., Cuccui A.L., Cueva de Valporquero, Čuković T., Ćukušić A., Culver D.C., Cuves de Sassenage, Cvitanović H., Danielopol D., Dányi L., Datry T., Dattagupta S., Deiana M., Delangle M., Della Toffola R., Delmastro G.B., Dethier M., Dole-Olivier M.-J., Dorigo L., Drevenšek A., Đud L., Dumnicka E., Dupre E., Eisenring C., Ekologgruppen i Landskrona, Erker J., Esmaeili S., Espanol C., Faille A., Farine V., Ferrandini J., Ferrandini M., Fiedler S., Fišer Ž., Flury M., Flury M., Fong D., Frajman B., François C., Frau S., Freiburghaus M., Fuschs A., Galassi D., Gerecke R., Gérino M., Giachino P., Giardini M., Gigante M., Glasnovič P., Glavaš I., Gnezda P., Golob K., Gonser T., Gorički Š., Gottstein S., Gouffre de Padirac, Gramaglia J.- P., Grazioli F., Grenier J.-P., Griebler Ch., Grottes de la Balme, Gutjahr S., Haenselmann P., Hahn H.S., Heide S., Hidding B., Hmura D., Hobbs H., Hudoklin A., Hugon B., Illés A., Inguscio S., Innocenti G., Jaklič G., Jalžić B., Jann B., Jazbec K., Jean P., Jugovic J., Kač K., Kapla A., Kastelic L., Kaufmann B., Kavanaugh K., Kavšček M., Kirin A., Klenk P., Kljun F., Knab O., Knapič T., Knight L., Komerički A., Konrad P.K., Küry D., Lah L., Lamanna F., Lana E., Landertshammer J., Lasnik K., Laurent S., Le Pennec R., Lefébure T., Lefebvre F., Lepretre B., Lescher-Moutoué F., Lesénéchal T., Likozar L., Lipps J., Lukić M., Maazouzi C., Magniez G., Malard A., Mallet E., Marcia P., Marmonier P., Marrone F., Martin D., Matičič J., Matthijs S., Matthys S., Mazza G., Mede M., Meleg I.., Mermillod-Blondin F., Meyssonnier M., Michel G., Mihaljević Z., Mirzajani A., Mock A., Mora A., Mori E., Morvan C., Morvan D., Morvan J.-P., Mulargia M., Nassar E., Nassar-Simon N., Niederreiter R., Notenboom J., Nyiro A., Obu J., Ozimec R., Pagliano S., Palatov D., Pamio A., Parc National du Mercantour, Parco Naturale Alta Valle Pesio e Tanaro, Parque Natural Saja-Besaya, Pascutto T., Pastorelli A., Pavlek M., Perne M., Petrov B., Piscart C., Piva E., Plan L., Polak S., Presetnik P., Prevorčnik S., Prié V., Proudlove G., Puch C., Pysarczuk S., Quilichini Y., Rađa T., Rade P., Radej B., Radoš M., Reboleira A.S., Renault D., Reš R., Réserve Naturelle de Sixt Fer à Cheval, Saclier N., Sambugar B., Šarac B., Sauve municipality, Scherrer E., SCV caving club, Sella R., Seminara M., Simon L., Simonič M., Šinigoj D., Sivec N., Sket B., Škufca D., Sotgia M.G., Stéfanato J.-P., Stein H., Stucki P., Stutz L., Sulzbacher D., Taddei P., Talenti E., Tanatmiş M., Tegzes Z., Tester R., Thys E., Toffola R.D., Tomasin G., Toniello V., Tramte J., Trontelj P., Tuekam Kayo R., Turjak M., Turk T., Vanni S., Verovnik R., Vilfan N., Vittori M., von Fumetti S., Vujčić Karlo S., Weidner A., Wildberger A. and Wood P.J.