



# Continental-scale patterns of hyper-cryptic diversity within the freshwater model taxon *Gammarus fossarum* (Crustacea, Amphipoda).

Rémi Wattier, Tomasz Mamos, Denis Copilaș-Ciocianu, Mišel Jelić, Anthony Ollivier, Arnaud Chaumot, Michael Danger, Vincent Felten, Christophe Piscart, Krešimir Žganec, et al.

## ► To cite this version:

Rémi Wattier, Tomasz Mamos, Denis Copilaș-Ciocianu, Mišel Jelić, Anthony Ollivier, et al.. Continental-scale patterns of hyper-cryptic diversity within the freshwater model taxon *Gammarus fossarum* (Crustacea, Amphipoda).. Scientific Reports, 2020, 10 (1), pp.16536. 10.1038/s41598-020-73739-0 . hal-02970981

**HAL Id: hal-02970981**

**<https://hal.science/hal-02970981>**

Submitted on 10 Nov 2020

**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.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

# Continental-scale patterns of hyper-cryptic diversity within the freshwater model taxon *Gammarus fossarum* (Crustacea, Amphipoda)

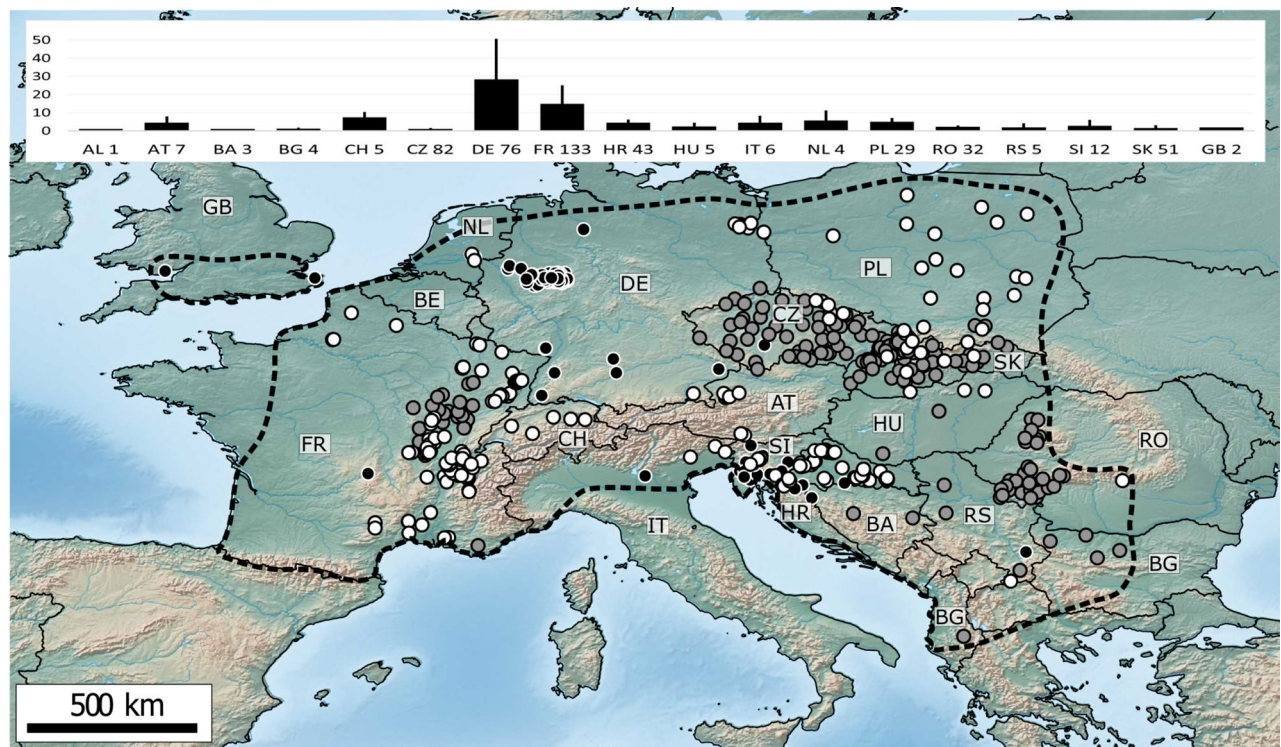
Remi Wattier<sup>1✉</sup>, Tomasz Mamos<sup>2,3</sup>, Denis Copilaş-Ciocianu<sup>4</sup>, Mišel Jelić<sup>5</sup>, Anthony Ollivier<sup>1</sup>, Arnaud Chaumot<sup>6</sup>, Michael Danger<sup>7</sup>, Vincent Felten<sup>7</sup>, Christophe Piscart<sup>8</sup>, Krešimir Žganec<sup>9</sup>, Tomasz Rewicz<sup>2,10</sup>, Anna Wysocka<sup>11</sup>, Thierry Rigaud<sup>1</sup> & Michał Grabowski<sup>2✉</sup>

Traditional morphological diagnoses of taxonomic status remain widely used while an increasing number of studies show that one morphospecies might hide cryptic diversity, i.e. lineages with unexpectedly high molecular divergence. This hidden diversity can reach even tens of lineages, i.e. hyper cryptic diversity. Even well-studied model-organisms may exhibit overlooked cryptic diversity. Such is the case of the freshwater crustacean amphipod model taxon *Gammarus fossarum*. It is extensively used in both applied and basic types of research, including biodiversity assessments, ecotoxicology and evolutionary ecology. Based on COI barcodes of 4926 individuals from 498 sampling sites in 19 European countries, the present paper shows (1) hyper cryptic diversity, ranging from 84 to 152 Molecular Operational Taxonomic Units, (2) ancient diversification starting already 26 Mya in the Oligocene, and (3) high level of lineage syntopy. Even if hyper cryptic diversity was already documented in *G. fossarum*, the present study increases its extent fourfold, providing a first continental-scale insight into its geographical distribution and establishes several diversification hotspots, notably south-eastern and central Europe. The challenges of recording hyper cryptic diversity in the future are also discussed.

In many areas of biology, including biodiversity assessments, eco-toxicology, environmental monitoring, and behavioural ecology, the species status of the studied organisms relies only on a traditional morphological definition<sup>1</sup>. However, a continually increasing number of studies of plants and animals show that one morphospecies might hide lineages with molecular divergences far exceeding what is expected at the intra-specific level in a given taxon<sup>2,3</sup>. This hidden diversity, nicknamed cryptic diversity (CD), can reach even tens of lineages for one morphospecies, i.e. showing hyper cryptic diversity (HCD)<sup>1</sup>. Even well-studied model-organisms may exhibit an overlooked CD<sup>4</sup>. This hidden diversity occurs at various geographic scales, and it has often been revealed to have pre-Pleistocene origins<sup>5</sup>.

Cryptic diversity has only recently been taken into account in both basic and applied ecology, as well as in evolutionary biology<sup>6–8</sup>. Even basic evolutionary studies were shown to be impacted by CD if two or more cryptic

<sup>1</sup>UMR CNRS 6282 Biogéosciences, Université Bourgogne Franche Comté, Dijon, France. <sup>2</sup>Department of Invertebrate Zoology and Hydrobiology, University of Lodz, Lodz, Poland. <sup>3</sup>Zoological Institute, University of Basel, Basel, Switzerland. <sup>4</sup>Institute of Ecology, Nature Research Centre, Vilnius Nature Research Centre, Institute of Ecology, Vilnius, Lithuania. <sup>5</sup>Department of Natural Sciences, Varaždin City Museum, Varaždin, Croatia. <sup>6</sup>Laboratoire d'écotoxicologie, INRAE, UR RiverLy, Villeurbanne, France. <sup>7</sup>UMR CNRS 73602 LIEC, Université de Lorraine, Metz, France. <sup>8</sup>Univ Rennes, CNRS, ECOBIO-UMR 6553, 35000 Rennes, France. <sup>9</sup>Department of Teacher Education Studies in Gospić, University of Zadar, Gospić, Croatia. <sup>10</sup>University of Guelph, Centre for Biodiversity Genomics, Guelph, ON, Canada. <sup>11</sup>Department of Genetics and Biosystematics, University of Gdansk, Gdansk, Poland. ✉email: remi.wattier@u-bourgogne.fr; michal.grabowski@biol.uni.lodz.pl



**Figure 1.** Map of *Gammarus fossarum* sampling sites. Dots indicate sampling sites (498); white with black contour represent new original data (206 sites), grey with black contour indicate sites associated with authors' previous publications (198 sites), black with white contour indicate sites for data derived from other published papers (94 sites). To prevent overlap of very closely located sites, some positions are slightly shifted. Thin blue-grey lines correspond to first-order river, and black lines are country borders. Dashed grey line represents the simplified distribution range limits for *G. fossarum*, adapted from Piscart and Bollache<sup>96</sup>. Countries (19) are indicated by their corresponding two-letter ISO codes: AL, Albania; AT, Austria; BA, Bosnia and Herzegovina; BE, Belgium; BG, Bulgaria; CH, Switzerland; CZ, Czechia; DE, Germany; FR, France; HR, Croatia; HU, Hungary; IT, Italy; NL, Netherlands; PL, Poland; RO, Romania; RS, Serbia; SI, Slovenia; SK, Slovakia; UK, United Kingdom. Top Inset represent sampling effort for each country: mean number of samples per site + S.E.. Number by the ISO code is the number of sites sampled. Map created with QGIS 3.4.5 (<https://www.qgis.org/fr/site/>).

species are merged into one analysis. Such is the case for sexual selection analyses<sup>9</sup>, studies upon host-parasite co-evolution<sup>6,10,11</sup>, functional ecology<sup>12,13</sup> or population genetics<sup>14</sup>, and is likely to impact genome assembly, even in the case of model organisms. Similarly, results of biodiversity assessments may be severely underestimated if ignoring the presence of CD<sup>15</sup>. At the macroecological scale, considering CD often generates a more nuanced view on determinants of biodiversity patterns<sup>8</sup>, and it should be taken into account while defining conservation units, especially in case of endemic cryptic lineages<sup>16,17</sup>. Simulations of distributional patterns under future climatic global change underestimate the real scale of biodiversity loss when CD is not taken into account<sup>18</sup>. Finally, ecotoxicological and biomonitoring studies are at risk of being flawed, as some authors reported that different cryptic species might differ in their tolerance to contaminants and ecosystem deterioration<sup>19–23</sup>.

While CD was detected across a wide range of taxa<sup>3</sup>, some may be more prone than others<sup>2</sup>. As a rule of thumb, morphospecies with broad geographic distributions, living in patchy habitats and having limited dispersal abilities tend to contain more divergent lineages<sup>2,24–26</sup>. For example, many freshwater macroinvertebrates<sup>27</sup> and especially amphipod crustaceans are known to be associated with a high level of CD<sup>5,19,28–39</sup>.

Here, we focus on the freshwater amphipod morphospecies *Gammarus fossarum* Koch, 1836, widespread in continental Europe and recently identified in the United Kingdom<sup>40</sup> (Fig. 1). The species is mostly epigeal, being present in a wide range of habitats, but sometimes is found even in shallow groundwaters. It is a keystone species in streams and rivers, where it occurs in high densities and plays a major role in the food web as the primary scavenger/shredder of organic matter as well as a food source for fish and macroinvertebrates (e.g.<sup>41,42</sup>). Thus, it is commonly used as a sentinel species in freshwater risk assessment of ecosystem quality, especially in the frame of the European Union's Water Framework Directive (EU WFD)<sup>43</sup>. It is also used in eco-physiology, e.g. in association with global warming studies<sup>44,45</sup>, as well as, extensively, in eco-toxicology<sup>46–49</sup>. This species has also been emerging as a model for molecular reproduction physiology<sup>50</sup>. Finally, *G. fossarum* is often used in evolutionary ecology studies such as behavioural ecology<sup>9</sup>, host-parasite relationships<sup>10,51</sup> or at the crossroads between parasitology and eco-toxicology<sup>45,52,53</sup>.

Almost two decades ago, three cryptic lineages (A, B and C) were identified in German populations of *G. fossarum*<sup>54</sup>, and two of them were shown to have different habitat requirements<sup>13</sup> and, likely, sensitivity to

pollutants<sup>20,21</sup>. In recent years, the level of CD was further explored in several parts of Europe and has dramatically increased, establishing *G. fossarum* as a complex of highly divergent lineages that may represent cryptic species<sup>55–59</sup>. These studies revealed a substantial amount of CD occurring even at small geographical scales. Intriguingly, the number of lineages seems to be the highest in south-eastern Europe, but the highest phylogenetic diversity recorded so far is in central Europe<sup>59</sup>. However, the extent of diversity in *G. fossarum* is currently inadequately known and most likely underestimated at the continental level, because of numerous and wide sampling gaps. Indeed, most of the previous studies were restricted geographically compared to the known distribution of *G. fossarum* in Europe.

Thus, the present study is based on a large number of individuals (almost 5000) and a geographically extensive sampling scheme (ca. 500 sites), including both original and literature data, covering the morphospecies' geographic range in Europe. The mitochondrial DNA cytochrome C oxidase subunit I (COI) gene barcode region sequence<sup>60</sup> has been chosen to assess the extent of CD. The reasons for that were: (1) sequences of this marker were widely available for *G. fossarum* in the literature and (2) this marker was successfully used at identifying CD both in *G. fossarum*<sup>55–59</sup> and in other gammarids (e.g.<sup>28</sup>). Following Kekkonen and Hebert<sup>61</sup>, we used COI as a quantifier of diversity and an "efficient start for taxonomic workflow", only targeting here the delineation of Molecular Operational Taxonomic Units (MOTUs) as a way to propose testable "species hypotheses" (e.g.<sup>17</sup>).

Specifically, we aimed to:

- (1) estimate the level of hidden diversity within *G. fossarum* at the continental scale using MOTUs delimited with distance and tree-based methods,
- (2) reveal and interpret the critical temporal and geographic patterns of the observed diversification allowed by the pan-European sampling,
- (3) discuss HCD in relation to evolutionary, ecotoxicology, biodiversity surveys and biomonitoring as well as to the future of HCD assessment.

## Results

**A global overview of MOTU diversity.** COI sequences for the DNA-barcode region were either newly generated or derived from the literature for 4926 *Gammarus fossarum* individuals, from 498 sampling sites distributed throughout 19 European countries (Fig. 1), totalling 691 haplotypes (Table S1, Fig S1). An extraordinarily high level of cryptic diversity was evidenced by a total of 146, 152 and 84 MOTUs obtained with BIN, bPTP and ABGD methods, respectively (Fig. 2). The K2P distance between haplotypes ranged from almost zero up to 0.35 (Fig. 3a). Although the distribution of genetic distance values was generally bimodal, only a small proportion of values centred around 0.03, while most concentrated around 0.22, illustrating that most haplotypes are highly divergent (Fig. 3a). Only 35 of the 146 BINs have already been known from literature or derived from the sequences mined from GenBank by BOLD, resulting in 111 new BINs in the present study, a fourfold increase. In addition, eight short sequences (397–468 bp) derived from the literature, divergent enough from the existing BINs not to be ascribed to any of them, but too short to be attributed their own BINs (following BOLD standard), should be considered as eight extra BIN-level MOTUs, extending the total to 154 BIN-equivalent MOTUs. Both BIN- and ABGD-MOTUs, shared a bell-shaped distribution of K2P distances centred on 0.22, illustrating that most MOTUs are highly divergent (Fig. 3b,c). However, while the lowest K2P value among ABGD-MOTUs was 0.05, this value for BIN-MOTUs was occasionally down to <0.01.

In 62 instances, MOTUs based on BIN, bPTP and ABGD methods were concordant. Single ABGD-MOTUs comprised from two to up to fifteen BINs, clearly showing that BIN was the least conservative among the applied MOTU delimitation methods (Fig. 2).

We observed 32 lineages being both sister clades in the phylogeny (Fig. 2) and diverging by >0.2 K2P distance. These 32 Lagrue-MOTUs are a conservative estimate of the number of lineages for which we could expect prezygotic reproductive isolation as suggested by Lagrue et al.<sup>57</sup>.

**Spatial and temporal pattern of diversification.** The Bayesian maximum clade credibility chronogram (Fig. 2) based on BINs, each represented by one randomly chosen haplotype, revealed the phylogenetic relationships and divergence times between MOTUs of the *G. fossarum* morphospecies (Fig. 2). Three significant points are emphasized. First, the estimated age of the entire lineage complex is ca. 26 Mya, i.e. middle-Oligocene. Second, all the major-clades appeared already in the early-middle Miocene, ca. 21–13 Mya. These two features illustrate the ancient age of the *G. fossarum* complex and its early diversification. Third, the Lineage Through Time analysis (LTT, inset Fig. 2) indicated a rather continuous evolutionary diversification, although at a somewhat slower pace from ca. 12.5 Mya.

Six highly divergent major clades (Fig. 2) SEE, FRA-CE A, EE Q, CWE, CE B and CEE (see below for explanation of acronyms) were recovered according to the rationale presented in Material and Methods, matching the patterns observed in previous multilocus studies. All major clades diverged anciently, around 20 Mya (Figs. 2, 4b). In the following interpretation, a clade was considered as (1) "narrowly distributed" when it was present at only one sampling site or several geographically close sites (<100 km apart), (2) "broadly distributed" when sites harbouring this clade were >100 km and <1000 km apart, and (3) "widespread" when locations were spanning >1000 km. The same rule was applied to BIN- and ABGD-MOTUs.

Among these six clades, three clades, i.e. CWE (Central-Western Europe), CEE (Central-Eastern Europe) and SEE (South-Eastern Europe), are widespread and rich in BIN- and ABGD-MOTUs. However, the MOTUs harboured by these clades can drastically differ in geographic range size, ranging from narrowly endemic to widespread (Fig. 4a). The CWE clade included 44 BIN- or BIN-equivalent MOTUs and eight ABGD-MOTUs. The clade itself is widespread, ranging from the UK and France in Western Europe, to south-western Poland

**Figure 2.** Bayesian maximum clade credibility chronogram with MOTU designation for *G. fossarum*. The outgroup is removed for clarity. Each BIN is represented by one randomly chosen haplotype. Grey bars at key nodes represent 95%HPD (Highest Posterior Density) intervals of clade age. Circles at nodes represent Posterior Probabilities (PP) for Bayesian Inference (BI) and as bootstrap values (BS) for Maximum Likelihood (ML, see Fig. S1), according to the following colour code: black: PP  $\geq$  95%/BS  $\geq$  55%, grey: PP = 95–80%/BS = 54–30% and white: PP < 79%/BS < 30%, unlabeled = bipartition not recovered in both analyses. Precise values are given for key nodes. Bars annotated on the right represent results of the MOTU delimitation methods i.e. Barcode Index Number (BIN), the Bayesian implementation of the Poisson Tree Processes (bPTP), the Automatic Barcode Gap Discovery (ABGD) and Lagrue's reproductive isolation methods, respectively. Three ABGD-MOTUs, 1, 2 and 6, corresponding to types A, B and C as defined by Müller et al. (2000), respectively are specified. Six major clades were distinguished and highlighted by the following colours: SEE (South-Eastern Europe, purple), FRA-CE A (France and Central-European A, yellow), EE Q (Eastern-Europe Q, dark blue), CWE (Central-Western Europe), CE A (Central-European A, yellow) and CEE (Central-Eastern Europe). First inset: Image of a male *G. fossarum* sensu stricto, i.e. Type A which occurs at the type locality (Photograph: Denis Copilaş-Ciocianu). Second inset: Lineages through time (LTT) plot. Analysis was performed on the dataset reduced to BINs. Grey lines represent 95%HPD. Tree was generated using BEAST 2.4.8 (<https://www.beast2.org/>).

and western Slovakia in central Europe (Fig. 4). Four ABGD-MOTUs within this clade are broadly distributed. Among them, ABGD-MOTU 3 is restricted to France but split between north-western Alps and south-eastern Massif Central, two areas > 500 km away from each other, each harbouring a specific BIN (Table S1). The ABGD-MOTU 9 is present in Slovakia, Hungary and Poland. The three other ABGD-MOTUs 1, 2 and 6 are more widespread. They would (see “Discussion”) correspond to the *G. fossarum* types A, B and C, respectively, described by Müller (2000)<sup>54</sup> (Fig. 2) and is presented in detail below. Only three BINs in this CWE clade were presumably narrowly distributed (Fig. 4a,b, Table S1).

The CEE clade encompassed the highest number of MOTUs: 72 BIN-, or BIN-equivalent MOTUs and 48 ABGD-MOTUs (Fig. 2). Like CWE, the CEE clade is widespread although with a peculiar, discontinuous distribution. While it is present almost exclusively in Central and Eastern Europe, two BINs are nevertheless found in the southern Massif Central in France. The CEE clade includes seven ABGD-MOTUs that are broadly distributed to a various extent (Fig. 4a,b). The ABGD-MOTUs 48 and 61 were present in a few sampling sites ca 250 km away from each other. The ABGD-MOTUs 36, 40, 44 and 45 were present on each side of the Pannonian Basin, (in the south-eastern Alps and the Southern Carpathians), at least ca 500 km apart. The last ABGD-MOTU, 77 is the most widely distributed, being present in Central Europe, namely in Hungary, Slovakia, Poland and eastern Germany. The remaining 48 BINs are presumably endemic to eastern and south-eastern Europe (Fig. 4, Table S1).

The SEE clade harbours 35 BIN- or BIN-equivalent MOTUs and 28 ABGD-MOTUs (Fig. 2). This clade is present in seven countries, from northern Italy in the west, to Albania and Bulgaria in the southeast of Europe. Contrary to CWE and CEE clades, none of these MOTUs was either widespread or even broadly distributed to any extent (Fig. 4a,b).

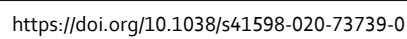
The major clade FRA-CE A (France-Central-Europe-A) was represented by two ABGD-MOTUs, each being also a single BIN-MOTU (Fig. 2). These two MOTUs are located in two distinct geographical regions, distant by > 1000 km: one in the northern French Alps, the other in western Slovakia and southern Poland (Western Carpathians) (Fig. 4a,b).

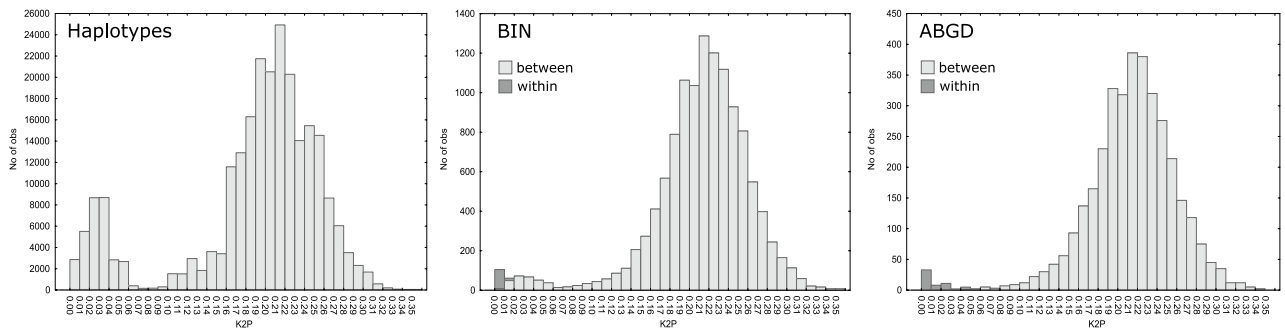
The two remaining major clades EE Q (Eastern-Europe-Q) and CE B (Central-Europe-B) were represented only by a single ABGD-MOTU, 78 and 80, respectively, each being also a single BIN-MOTU (Fig. 2). The CE B clade is present only in the Western Carpathians, while the EE Q clade is slightly more broadly distributed in eastern Czech-Republic and westernmost Slovakia (Fig. 4a,b).

**ABGD-MOTUs latitudinal and altitudinal distribution.** The diversity of ABGD-MOTUs also varied according to altitude (Fig. 4c). Although numerous MOTUs inhabit the zone between 100 and 500 m asl, some ABGD-MOTUs were found below 100 m and many above 600 m, even up to 1500 m. Among the 15 ABGD-MOTUs found to be broadly distributed (Fig. 4d), most had a rather wide altitudinal span. For example, ABGD-MOTU 2 occurs from the sea level to as high as 1500 m asl, while MOTU 78 was found in sub-mountain areas and some ABGD-MOTUs (e.g. 36, 40, 44 and 77) are mainly restricted to lowlands (Fig. 4d).

The diversity of ABGD-MOTUs also varied according to latitude (Fig. 4e), with two peaks – the most prominent from 45.5°N to 46.5°N and another, much lower, from 48.5°N to 50°N. The first one is associated predominantly with the diversity hotspots in the southern part of *G. fossarum* distribution range, namely the western and south-eastern outskirts of the Alps, the south-western part of the Pannonian Basin and the southern part of the Western Carpathians. The other peak illustrates the lineage diversity present in the Western Carpathians predominantly. The lowest lineage diversity is observed between 50.5°N and 54.0°N, corresponding mainly to the western and central parts of the Great European Plain.

**Historical types A, B and C.** As shown on Figs. 2 and 4, the ABGD-MOTUs 1, 2 and 6 would correspond, respectively, to the *G. fossarum* types A, B and C initially defined by Müller<sup>54</sup> (see “Discussion” for details). Since these MOTUs were the most commonly used for various ecological and ecotoxicological studies (e.g.<sup>20,21</sup>), it is important to highlight their biogeographical history. They include large numbers of BINs, 12, 15 and 10, respectively (Figs. 2, 4, 5b–d for details) and their geographic ranges overlap either fully (B and C) or only partially (Fig. 5a). Numerous cases of sympatry between these types were also observed in France, especially in case of ABGD-MOTU 2 (type B) and 6 (type C), and ABGD-MOTU 2 (type B) and 6 (type C) (Fig. 5b–d). Altitudinal ranges for each BIN-MOTUs present in at least three sampled sites are shown in box-plot insets in Fig. 5b–d. The





**Figure 3.** Kimura 2 parameter (K2P) distances (a) between haplotypes, (b) within and between BIN-MOTUs and (c) within and between ABGD-MOTUs.

ranges observed at the ABGD-MOTU levels are overall as wide as the BIN-MOTU level for ABGD-MOTUs with high sampling effort. It shows that the ranges of ABGD-MOTUs were not artefacts resulting from combining BIN-MOTU ranges which would have non-overlapping distributions.

**Geographic overlap and co-occurrence of BIN-, ABGD-MOTUs and major clades.** We found a regional co-occurrence of BIN-MOTUs, ABGD-MOTUs and major clades (Figs. 4, 5). To further quantify within-site co-occurrence, sites with two or more sampled individuals (321 out of 498 sites) were used. Within-site co-occurrence of at least two BIN-MOTUs, ABGD-MOTUs and highly divergent clades was observed in 32.4%, 17.8% and 7.2% of the sites, respectively. Also, at least three BIN-MOTUs co-occurred in 19 locations and at least three ABGD-MOTUs in four sites (Table S1). Sympatry between these different genotypes is therefore quite frequent. Sampling size is likely to impact the ability to detect co-occurrence, particularly if MOTUs are unevenly distributed, which seems a regular feature for *G. fossarum* (e.g.<sup>55</sup>). For example, in our dataset, rare BIN-MOTU had a frequency of no more than 0.2 in 73.6% of the sites with more than ten individuals sampled, and no more than 0.1 in 49.1% of such sites.

**Geographic hotspots of divergence.** The landscape genetic map (Fig. 6a) of molecular divergence showed that the Carpathian, Dinaric and Pannonian regions exhibit the highest levels of lineage divergence. This pattern holds regardless if one or more of the six major clades are present in a given region.

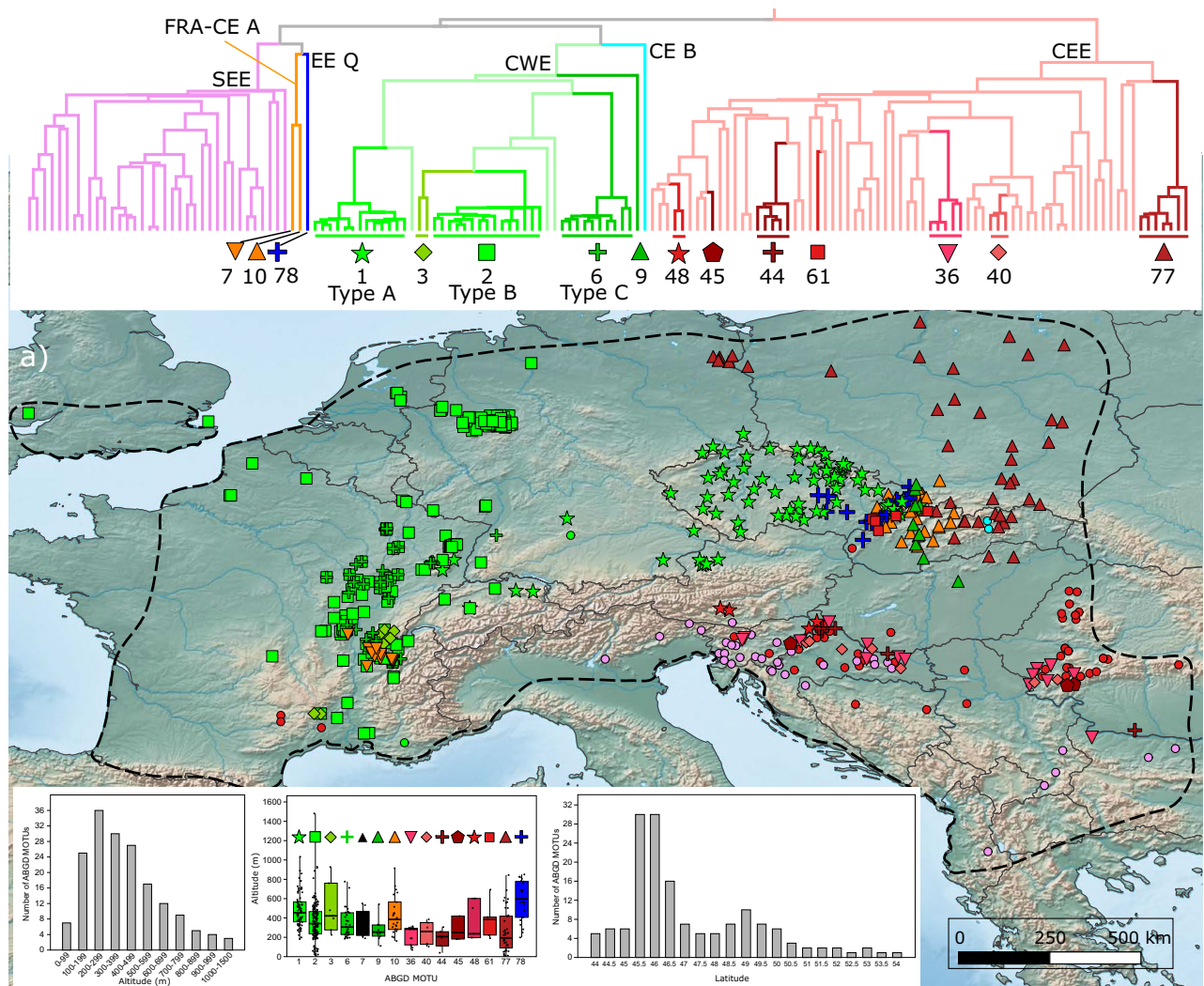
If highly divergent endemic Carpathian, Dinaric and Pannonian lineages are excluded (i.e. explaining most of the previously described hotspots), the Western Carpathians still appear as a hotspot of high divergence with five major clades being sympatric in this area: CEE (ABGD-MOTU 77), CWE, EE Q, FRA-CE A and CE B (Fig. 6b). Many regions in France also appeared as local hotspots either, associated with the sympatric occurrence of two major clades, e.g. north-western Alps (CWE and FRA-CE A) and southern Massif Central (CWE and CEE) or divergent ABGD-MOTUs from the same major clade (e.g. 1, 2 and 6, i.e. type A, B and C in central-eastern France) (Fig. 6b). On the contrary, the Great European Plain (northern France, Germany, Poland) of the current distribution of *G. fossarum* appeared as an area of low divergence (Fig. 6a,b).

## Discussion

Our results, based on extensive continental scale sampling of *G. fossarum*, revealed a remarkable level of hyper cryptic diversity (HCD) with 146, 152 and 84 MOTUs obtained with BIN, bPTP and ABGD methods, respectively. The age of the species complex was estimated at ca. 26 Mya and the divergence was predominantly high, most lineages exhibiting a K2P distance of 0.22. Finally, a relatively high level of sympatry was observed even at the site level. Although CD was already documented in *G. fossarum*<sup>54–59</sup>, including HCD<sup>58</sup>, the present study considerably increases its extent, offering for the first time the opportunity to discuss its origin on a continental scale.

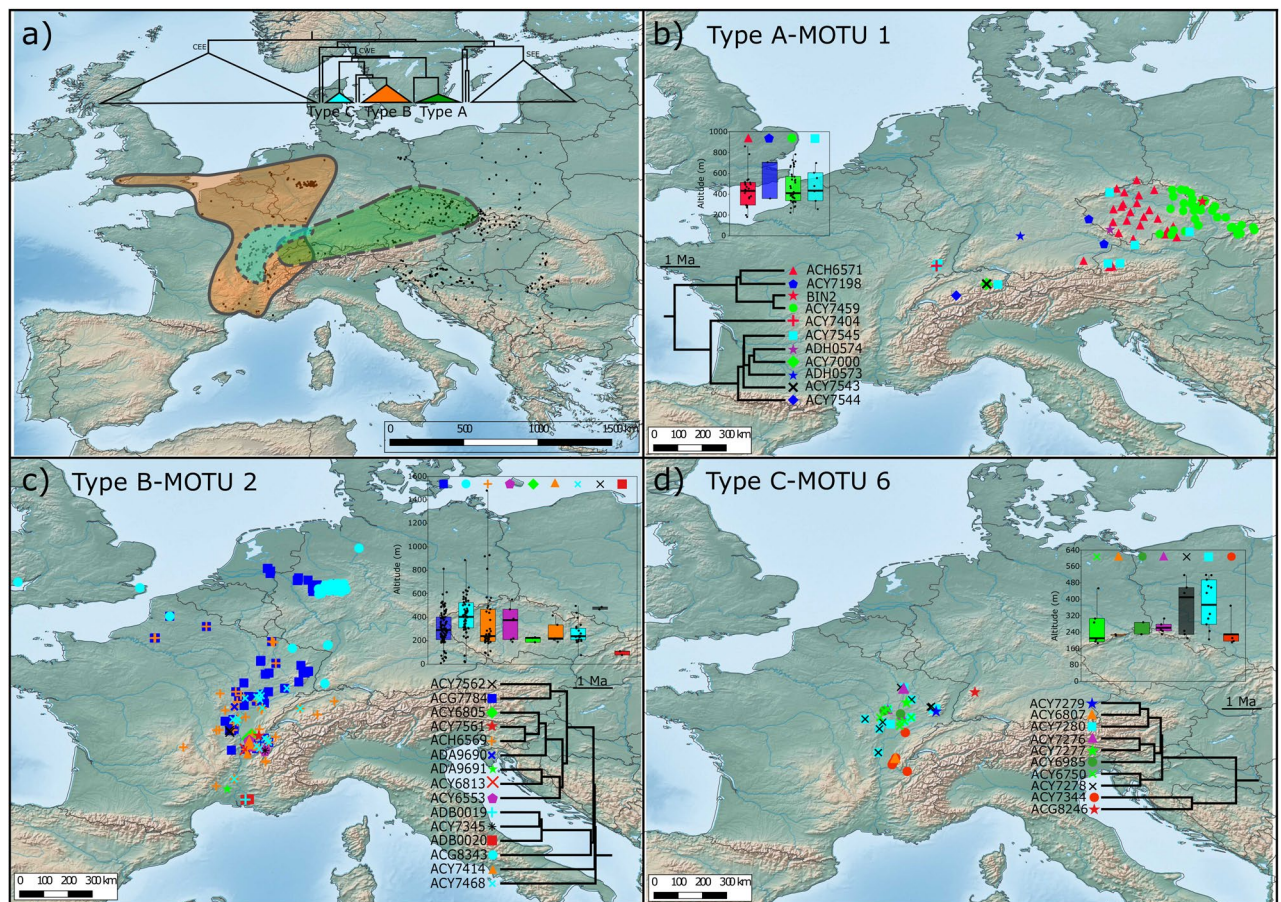
CD is a common phenomenon in freshwater morphospecies of the genus *Gammarus*, being reported in all Eurasian wide-ranged taxa studied thus far: e.g. *G. balcanicus*<sup>5,33,62</sup>, *G. komareki*<sup>32</sup>, *G. ochridensis*<sup>29</sup>, *G. pulex*<sup>57,63,64</sup>, *G. roeselii*<sup>28</sup>, *G. lacustris*<sup>32</sup>, *G. kischineffensis*<sup>64</sup>, *G. leopoliensis*<sup>65</sup>, but also in North American (e.g. *G. pecos*<sup>37</sup>) or East Asian (e.g. *G. nekkensis*<sup>66</sup>) taxa. HCD was detected in three studies<sup>5,28,32</sup>. Two studies identified 15 and 35 ABGD MOTUs in *G. roeselii*<sup>26</sup>, and in *G. komareki*<sup>30</sup>, respectively. A survey covering the entire European range of *G. balcanicus*<sup>5</sup> presented 49 highly divergent lineages (likely to represent ABGD-MOTUs given their degree of divergence). In comparison, a previous study<sup>62</sup> on the same taxon reported up to 22 lineages in the Carpathian Mountains alone. These studies shared a high sampling effort, both in the number of sampling sites and individuals, pointing out the need for appropriate sampling strategies to detect HCD. The age of these species complexes was usually shown to be old, ranging from 4–6 Mya in *G. kischineffensis*<sup>64</sup> or *G. leopoliensis*<sup>65</sup> to 15–17 Mya for *G. balcanicus*<sup>5</sup> and *G. roeselii*<sup>26</sup>. Intriguingly, the high frequency of sympatry observed between *G. fossarum* MOTUs was not detected in the other species complexes<sup>5,32</sup>.

Cryptic diversity is common in crustaceans, even overrepresented when compared to other metazoans<sup>2</sup>. The CD in amphipods is far from being restricted to the genus *Gammarus*. It has been reported for many other epigeic freshwater amphipod families, e.g. North American Hyaletellidae<sup>30,38,39,67</sup>, New Zealand Paracalliopiidae<sup>31</sup> or Australian Chiltonidae<sup>34</sup>. It is also extensively known from hypogean freshwater taxa,



**Figure 4.** Geographical and altitudinal distribution of major clades and AGBD-MOTUs in reference to their phylogenetic position. **(a)** Thin blue lines on map present first-order river, bold grey lines are country borders. Dashed black line represents a simplified distribution range limits for *G. fossarum*. **(b)** Simplified version of the Chronogram from Fig. 2. The six major clades distinguished in Fig. 2 are highlighted by the same colour code, both on the map and on top inset tree: SEE (South-Eastern Europe, purple), FRA CE A (France Central-Europe A, yellow), EE Q (Eastern-Europe Q, dark blue), CWE (Central-Western Europe, green), CE B (Central-Europe-B, turquoise blue) and CEE (Central-Eastern Europe, red). The 84 AGBD-MOTUs were split into two categories: MOTUs with narrow distribution based on our sampling (< 100 km between the most distant sites) and more broadly distributed MOTUs. The colour code for major clades was followed. Presence of the narrowly distributed MOTUs is indicated with a circle on the map, while the more broadly distributed MOTUs were attributed a specific symbol (see the top inset tree for details). To prevent overlap of symbols in sites with co-occurrence, some positions are slightly shifted. **(c)** The histogram presents the absolute frequency of AGBD-MOTUs according to altitude. **(d)** The altitudinal distribution of broadly distributed AGBD-MOTUs based on AGBD method is presented as box plot. The colour code follows major clade designation of the top inset tree (see Fig. 2 for details). The dots represent sampling sites, thick black bar represent the mean, grey box represents first and last quartile, thin vertical line represents min–max values. **(e)** Histogram representing the number of lineages per latitudinal band. Maps created with QGIS 3.4.5 (<https://www.qgis.org/fr/site/>) and trees with BEAST 2.4.8 (<https://www.beast2.org/>).

e.g. Niphargidae<sup>68,69</sup>, Paramelitidae<sup>70</sup>, or Crangonyctidae<sup>71</sup>, as well as in marine amphipods, including abyssal habitats, e.g. Epimeriidae<sup>72</sup>, Eusiridae<sup>73</sup>, Eurythenidae<sup>74,75</sup>. Among all these families, examples of both high/old divergence<sup>69,70,75</sup> and HCD<sup>34,38,70</sup> are not rare. Therefore amphipods, whatever the habitat, are highly prone to CD. The present, often decades old, morphological definitions of amphipod species seem insufficiently detailed in many cases. Either morphological stasis, convergent or parallel evolution could be reasons explaining this pattern, but discussing evolutionary processes in detail is beyond the scope of the present paper. It might also be

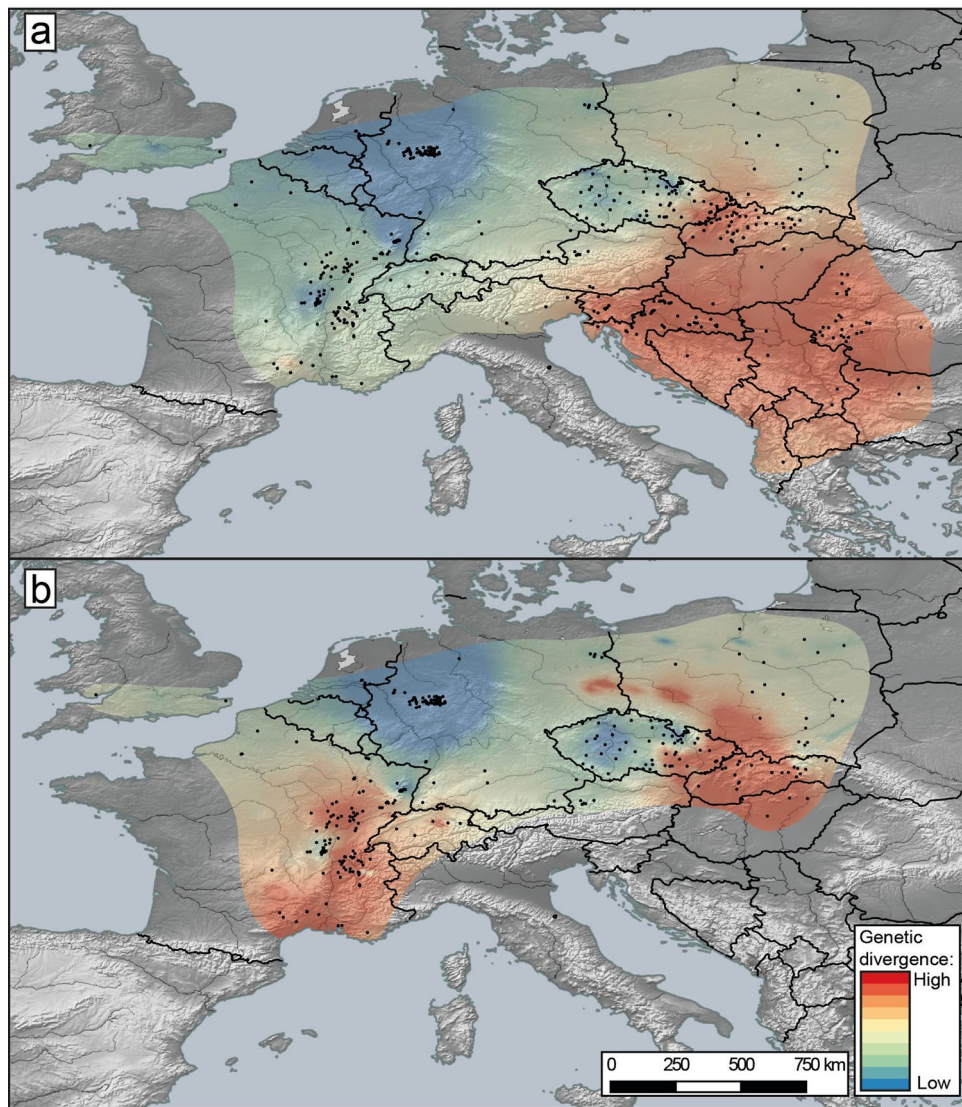


**Figure 5.** Geographical and altitudinal distribution of the three AGBD-MOTUs 1, 2 and 6, corresponding to types A, B and C as defined by Müller et al. (2000), and the contained BIN-MOTUs. **(a)** General distributional range of the three AGBD-MOTUs, 1, 2 and 6, corresponding to type A, B and C as defined by Müller et al.<sup>54</sup> (2000), respectively. Thin dashed line indicates the distribution of each AGBD-MOTU. The tree presented is a collapsed version of the chronogram highlighting the phylogenetic position of key MOTUs. **(b–d)** Distributional maps of BIN-MOTUs within each the three AGBD-MOTUs 1, 2 and 6, respectively. To prevent overlap of symbols in sites with co-occurrence, some positions are slightly shifted. Each presents both, the phylogenetic relationship between BIN-MOTUs and the symbols used in the map and in the box plot. Box plots show the altitudinal distribution of each BIN-MOTU. Dots represent sampling sites, thick black bar represent the mean, grey box represents first and last quartile, thin vertical line represents min–max values. Maps created with QGIS 3.4.5 (<https://www.qgis.org/fr/site/>) and Trees with BEAST 2.4.8 (<https://www.beast2.org/>).

a matter of taxonomic impediment, associated with an increasing shortage of taxonomists<sup>76</sup>, amphipods ranking amongst the least taxonomically studied taxa<sup>77</sup>.

Our data reveal an old and highly complex pattern of molecular diversity within a morphospecies at the local and the continental scale. For example, the ancestors of three clades likely had a pan-European distribution. First, while the CEE clade geographic distribution is centred in central and eastern Europe, three AGBD-MOTUs 68, 71 and 72 were located in southern France, suggesting that the ancestor of this clade had a pan-European distribution. Additional supporting evidence comes from AGBD-MOTU 71, which is in a well-supported sister relationship with the AGBD-MOTUs 65–67, all the latter currently restricted to Romania. The estimated age of this CEE group of lineages is ca 11 Mya, i.e. middle Miocene. Second and similarly, the two MOTUs of the clade FRA-CE A, one restricted to France, and the other to Czechia and Slovakia, are well-supported sister MOTUs, which diverged in Mid-Miocene, ca 12.5 Mya. Third, MOTUs of the CWE clade were present in Western Europe (ABGD-MOTUs 2, 3, 6) and Central Europe (ABGD-MOTUs 1, 9), and also diverged during the Miocene.

Local presence of high CD has already been known for several, predominantly submontane regions of southern and central Europe<sup>58</sup>. We show here a general trend that confirms this pattern and provides further evidence for other local diversity hotspots in the south-western Pannonian Basin and the south-western outskirts of the Alps. However, we point out that the Balkan region is still poorly covered and likely harbours a significant amount of diversity. The probable middle Oligocene (ca. 26 Mya) to late Pliocene (ca. 3 Mya) timeframe of diversification for the main clades and MOTUs corresponds well with the temporal scheme obtained for other widespread European freshwater amphipods, such as *G. balcanicus*<sup>5</sup> or *G. roeselii*<sup>28</sup>. Our results advocate Alpine orogeny, i.e. simultaneous uplift of the Alps, Carpathians, and Dinarides, as the main factor shaping the phylogeography of *G. fossarum*. This period was also associated with the presence of extensive lacustrine systems and marshlands



**Figure 6.** A genetic landscape map based on COI mtDNA haplotypes overlaid onto a relief map of Europe. Black dots indicate sampling sites. Warmer colour (red) reflects high molecular divergence between neighbouring localities, while colder colour (blue) corresponds to areas of lower molecular divergence. **(a)** Map including all the sampling sites (498) and all haplotypes (691). **(b)** Map excluding highly divergent lineages endemic to the Carpathian, Dinaric and Pannonian regions (hotspots of most ancient divergence), to better illustrate the level of genetic divergence in the remaining area (469 haplotypes and 336 sites). Genetic distances were generated with Alleles in Space (AIS; <https://www.marksgeneticsoftware.net>) and maps created with QGIS 3.4.5 (<https://www.qgis.org/fr/site/>).

covering vast areas of the continent, subsequently fragmented as a result of the regional uplifts<sup>78</sup>. This could enhance the spread of some lineages over vast areas, promoting the subsequent diversification process. That is the case of the sister ABGD-MOTUs 7 (FRA) and 10 (CE A), the former being endemic to the western outskirts of Alps and the latter to the Western Carpathians, while both are in a sister relationship to another highly divergent ABGD-MOTU from the Western Carpathians. We also reveal the presence of wide areas of reduced genetic diversity, predominantly on the Great European Plain, that was partly covered by glaciers during the Last Glacial Maximum (LGM) and partly by a permafrost tundra with few scattered patches of forest<sup>79</sup>. Thus they provided no favourable conditions for most of the *Gammarus* species, which are known to play a major role in decomposing broadleaved tree debris<sup>41,80</sup>. In consequence, the area was probably colonised only after the LGM, as is evidenced already for two largely sympatric BINs of the ABGD-MOTU 77 that spread over the central part of the Great European Plain from the glacial refugium/diversity hotspot in the Western Carpathians<sup>81</sup>. Similarly, the other widespread ABGD-MOTU 2 is represented by three BINs in the western part of the Great European Plain. Overall, our findings stress the fundamental role played by a combination of regional historical and environmental factors shaping the local genetic diversity of the *G. fossarum* complex in different parts of Europe and point out to the need for precise identification of the sampled populations with DNA barcodes.

The *G. fossarum* types A, B and C were defined in by Müller<sup>54</sup> in 2000 based on 16S rRNA sequences and allozyme data from samples collected mainly in Germany. More recent publications referring to these three types were based on COI, but the sampling remained restricted to or centred around Germany<sup>20,55,56,63</sup>. Our study shows that samples from these studies<sup>20,55,56,63</sup> are associated exclusively with one or two BINs per type (type A: ADH0573<sup>20</sup>; type B: ACG7784 and ACG8343<sup>20,52,53</sup>; type C: ACG8246<sup>53</sup>). We conclude that our ABGD-MOTUs 1, 2 and 6 correspond to types A, B and C, respectively, each type comprising many BIN level MOTUs (11, 15 and 10, respectively). Our study confirms that the geographic distribution of type A does not extend westward more than already reported in the literature<sup>54,58,82</sup>. However, our results substantially extend the distribution of type B (but see<sup>55</sup>) and even more notably type C. Type B was known to be widely present in Germany, Switzerland and north-eastern France<sup>54,57,63</sup> and was recently recorded in the UK<sup>40</sup>. Our study extends its distribution to north-western and southern France. Initial studies restricted the distribution of type C to the French-German border marked by the Rhine River<sup>20,54,63,82</sup>. The present study shows it is widely present also in the whole eastern part of France. Some studies pointed out that geography could alone be predictive of the distribution of ABGD level MOTUs, e.g. type A and B within Germany<sup>54</sup>. However, the co-occurrence of individuals associated with different types was also pointed out in the literature (e.g.<sup>13</sup>). Based on our substantial sampling effort and extended geographic cover, we conclude that co-occurrence of MOTUs (ABGDs, BINs) may be the rule for the distributional pattern of *G. fossarum*. Initially proposed to be of pre-Pleistocene origin<sup>54,82</sup>, the evolutionary splits between types A, B and C are in our study suggested at early Miocene, matching the early diversification of the CWE clade itself. Phylogeographic history of type A was already extensively studied<sup>54</sup>. Detailed analysis of other types is beyond the scope of the present paper. However, intricate diversification patterns including multiple refugia and more or less extensive expansion during each interglacial in the Pleistocene and part of the Pliocene are likely, as was already proposed for two stygobiotic amphipods (*Niphargus virei*<sup>83</sup>, *N. rhenorhodanensis*<sup>84</sup>) in the same area.

It is essential that HCD is taken into account in both basic and applied studies while ignoring it might have multiple potential impacts. Recently, accumulating studies upon *G. fossarum* show that CD may impact various types of biological processes, including (1) sexual selection with impact on the outcome of mate choice and male-male competition<sup>9</sup>, (2) intensity of infection by acanthocephalan parasites<sup>10,11</sup> and parasite-induced mortality<sup>10</sup>, (3) functional ecology associated with habitat partitioning<sup>13</sup> and (4) ecotoxicology, with possibly different response to a contaminant<sup>20,21</sup>. It might also be largely involved in the within-species biodiversity effects, increasingly acknowledged in the biodiversity-functioning relationships<sup>85</sup>. Generally, how many different Evolutionary Significant Units (ESU) should be considered within *G. fossarum* is a relevant question and a challenge as exemplified by other amphipods<sup>35,86</sup>. On the one hand, our results suggest that individuals of different types will barely interbreed as they belong to different Lagrue-MOTUs. Indeed, it was revealed that individuals from BINs diverging by ca 16% K2P distance (ACY7276, type C vs either ACY7784 or ACH6569, both type B) predominantly discriminate between each other<sup>9,57</sup>. On the other hand, the same authors also found out that individuals from the two studied BINs within type B (ACY7784 or ACH6569) diverging by ca. 4% K2P distance, mate randomly, exhibiting no apparent pre-zygotic reproductive barriers. Our study illustrates that many BINs within each type are sympatric, so the question of how frequent between-MOTU mating remains open but testable. Applying nuclear markers, such as microsatellites (which are available for *G. fossarum*), might help to answer this question in the future<sup>14</sup>.

Our study points out a tremendous haplotypic (691 haplotypes) and MOTU-wise diversity as well as to the co-occurrence of MOTUs at the same sampling site. Also, such syntopy often includes rare MOTUs. Both phenomena will impact the future exploration of CD in *G. fossarum*, either for basic (e.g. phylogeography) or applied (e.g. monitoring or ecotoxicology) topics. Undoubtedly, the sampling strategy should include enough individuals per site to detect rare MOTUs. The Sanger sequencing of single PCR amplicons<sup>87</sup>, targeting the full COI barcode (as in the present study), might convey highly informative content but, simultaneously, it is a costly low throughput method<sup>87</sup>. Pyrosequencing of mitochondrial 16S targeting diagnostic Single Nucleotide Polymorphisms (SNPs) has been applied to identify type A, B and C<sup>82</sup>. Being high-throughput and fully valid in the restricted geographic area where it was initially implemented (mainly Switzerland), the method would seem inappropriate if applied on a continental scale. More recently, e-DNA-metabarcoding was used to detect the presence of *G. fossarum* in UK rivers. Environmental DNA (eDNA) was amplified for the entire macroinvertebrate community<sup>40</sup>, and a 333 bp COI mini-barcode was sequenced using high throughput sequencing (HTS<sup>88</sup>). The method proved (1) to be sensitive at detecting *G. fossarum* even at very low abundance (2.6% of total biomass), and (2) discriminant, as the mini-barcode proved informative enough compared to the full barcode to allow unambiguous phylogenetic assignment at the ABGD-MOTUs level (i.e. type B). Would this mini barcode be sufficiently informative to discriminate all BIN-MOTUs and ABGD-MOTUs diversity at the European scale? Although being outside the scope of the present paper, the large amount of data associated with the current study would easily allow testing it in silico.

Given the HCD described for *G. fossarum* in the present paper, we recommend any research, either related to applied or basic topics, to obtain both COI barcodes and BINs for the individuals under study. Our research also points out the necessity of a comprehensive DNA barcode reference library<sup>89</sup> for *G. fossarum* at the European scale. Even if our research is based on an extensive sampling effort, such sampling is heterogeneous, and many areas of *G. fossarum* distribution are unsampled or undersampled. Even if the task will be tremendous, particularly given the shortage of morpho-taxonomists, defining and naming the MOTUs would be desirable<sup>90</sup> as recently occurred for ABGD-MOTU 77, which was ascribed a formal taxonomic name: *Gammarus jazdzewskii*<sup>81</sup>. In the same line, as *G. fossarum* offers the opportunity to relatively easily set up crossing experiments<sup>57</sup>, a feature rarely explored in amphipods in the context of CD, it should be further explored<sup>37</sup>.

## Methods

**Data overview.** Sequences for COI were derived from a total of 4926 individuals collected from 498 sites (Fig. 1, Table S1). The sampling included 19 countries, covering the whole range of *G. fossarum* in continental Europe as well as two sites from United Kingdom, an area considered as recently invaded<sup>40</sup>. The collected data included: (1) original data (206 sites, 2071 sequences), (2) data derived from the co-authors' previously published studies<sup>57–59,91</sup> (198 sites, 682 sequences, although 463 sequences from<sup>54</sup> were initially unreleased in GenBank) and (3) data derived from other published papers<sup>20,40,55,56,63,92,93</sup> (94 sites and 2173 sequences, although 2086 of these sequences<sup>56</sup> originated from a very restricted area in Germany). All sequences retrieved from GenBank were checked to be phylogenetically part the *Gammarus fossarum* species complex using both BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi><sup>94</sup>) and assessing monophyly relative to sister clades<sup>93</sup>. Sampling size per site was on average 10 individuals but was highly variable with a standard error of 14. A total of 180 sites contained only one individual, 95% originating from already published studies (Table S1). Sampling effort was variable (Fig. 1). Some countries e.g. Slovakia and The Czech Republic were evenly and highly sampled but with low number of samples per site while other countries e.g. Germany accounted for few sites but heavily sampled.

**Processing of newly gathered material (sampling, taxonomic identification, molecular methods).** The newly gathered material (Fig. 1) was collected using kick-sampling with a benthic hand-net and fixed in 96% ethanol on site. In the laboratory, the animals were identified to the morphospecies level using characters described in available keys<sup>95–99</sup>. Genomic DNA was extracted from pieces of muscle tissue using either standard phenol-chloroform<sup>100</sup> or Chelex<sup>101</sup> protocols. For all individuals the barcoding region of the mtDNA cytochrome C oxidase subunit I (COI) was amplified for 659–710 bp long fragment (including primers) using any of the following primer pairs: LCO1490 and HCO2198<sup>102</sup>, UCOIF and UCOIR<sup>103</sup>, COIGrF and COIGrR2<sup>28</sup>, LCO1490-JJ and HCO2198-JJ<sup>104</sup>. Details about primer sequence and location are given in Table S2. All details of the molecular procedures followed those described by Mamos et al.<sup>5</sup>. The amplicons were sequenced using the BigDye sequencing protocol (Applied Biosystems 3730xl) by Macrogen Inc., Korea. Sequences were edited and aligned with CLUSTALW 2.0 (<https://www.clustal.org/>)<sup>105</sup> using either MEGA 7 (<https://www.megasoftware.net/>)<sup>106</sup> or Geneious 6.0.5 (<https://www.geneious.com/>). Based on a 530 nucleotides long alignment, haplotypes were retrieved using DnaSP v5 (<https://www.ub.edu/dnasp/>)<sup>107</sup>. All newly produced sequences were checked to be phylogenetically part the *Gammarus fossarum* species complex using both BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi><sup>94</sup>) and assessing monophyly in relation to sister clades<sup>93</sup>.

**Cryptic diversity—MOTU delimitation.** To explore the number of MOTUs that could represent putative cryptic species within *G. fossarum*, we applied four different approaches. Two methods were purely genetic distance-based methods, i.e. the Automatic Barcode Gap Discovery (ABGD)<sup>108</sup> and the Barcode Index Number (BIN)<sup>109</sup>. A third one was a tree based phylogenetic approach, using the Bayesian implementation of the Poisson Tree Processes (bPTP)<sup>110</sup>. The fourth method takes advantage of the study of Lagrue et al.<sup>57</sup> which combined genetic distance and reproductive isolation experiments. Both BIN and ABGD share the principle of clustering sequences into MOTUs according to their molecular divergence using algorithms aiming to find discontinuities, i.e. a so-called barcode gap separating intra- and interspecific genetic distances. For the ABGD (<https://www.wab.i.snv.jussieu.fr/public/abgd/>) method, we used primary partitions as a basis for group definition. Primary partitions are typically stable over a wide range of prior values, minimise the number of false-positive (over-split species) and, in the absence of cryptic diversity, are usually found to reflect the number of taxa described by morpho-taxonomists<sup>111</sup>. The Kimura two-parameter (K2P) substitution model was applied<sup>60</sup>, and the default value of 0.001 was used as the minimum intraspecific distance. Given that neither morphology<sup>32,55,103</sup> nor reproductive isolation<sup>57</sup> provides any consensus about which maximum intraspecific distance allows for reliable species delimitation in amphipods, we explored a set of values of up to 0.1. The BIN algorithm is implemented as part of The Barcode of Life Data Systems (BOLD; <https://www.boldsystems.org/>)<sup>112</sup>. In this algorithm, the newly submitted COI sequences are aligned and compared pairwise and also to each sequence already deposited in BOLD. Then, the initial single-linkage clustering is performed on the aligned sequence data in such a way that each cluster is allowed for a maximum intra-cluster distance of 0.022, and its distance to any other cluster is more than twice the threshold (> 0.044). In the subsequent step, the clusters are refined in the sense that clusters whose members show high sequence variation, but lack discontinuity remain as a single MOTU, while those in which sequence variation shows clear internal partitions are assigned to two or more MOTUs, even if their separation is less than 0.022. Then, finally, each refined cluster is assigned a globally unique and specific identifier. For details on the BIN system see Ratnasingham and Hebert (2003)<sup>109</sup>. Like ABGD, in the absence of cryptic diversity, BINs are claimed to be congruent with the number of taxa described by morpho-taxonomists<sup>109</sup>. In addition, BINs may be registered and publicly available through BOLD, allowing tractability and have similar properties as classic taxonomy descriptors (e.g. can be split or synonymised). Only sequences over 500 bp can be included in the BIN clustering, while shorter sequences which are over 300 bp may be only ascribed to an existing BIN, but will not create a BIN or split an existing one. For example, our data set included eight short sequences derived from the literature, divergent enough from the existing BINs not to be ascribed to any of them but too short to be ascribed their own BIN (following the BOLD standard). In the result, we considered them as separate BIN-equivalent MOTUs.

The tree-based bPTP method uses non-ultrametric phylogenies. It incorporates the number of substitutions in the model of speciation and assumes that the probability that a substitution gives rise to a speciation event follows a Poisson distribution. The branch lengths of the input tree are assumed to be generated by two independent classes of Poisson events, one corresponding to speciation and the other to coalescence<sup>110</sup>. For the input tree, we used a Maximum likelihood (ML) tree obtained with RAxML HPC 8.2.9 (<https://cme.h-its.org/>)

[exelixis/web/software/raxml/](#))<sup>113</sup> from the haplotype data. The ML analysis was run under a thorough tree search with the GTR + G substitution model applied to each codon partition. The bPTP (<https://species.h-its.org/ptp/>) analysis was performed with 500,000 iterations of MCMC and 10% burn-in. Three runs were performed. The convergence of each run was verified through stationary pattern of the MCMC iterations trace plot. All runs provided congruent results.

The fourth method is newly introduced in the present paper and will be designated as the Lagrue-MOTUs. It takes advantage of the study of Lagrue et al.<sup>57</sup>, which combined COI K2P distance and reproductive isolation in *Gammarus fossarum*. Lagrue et al.<sup>57</sup> tested two distance classes ca. 4% (for which reproductive isolation was absent) and ca 16% (for which reproductive isolation was almost exclusively present). Conservatively, sister clades that diverged by > 20% K2P distance in the present study were considered as different Lagrue-MOTUs and a surrogate of the minimum number of biological species.

**Time-calibrated phylogenetic tree.** The time-calibrated phylogenetic tree was reconstructed using Bayesian inference (BI) in BEAST 2.4.8 (<https://www.beast2.org/>)<sup>114</sup>. A *Gammarus roeselii* sequence (KP789695) was used as an outgroup for the analysis. Two sets of data were used, one at the haplotype level the other at the BIN level. The outgroup was removed after the analysis. Priors for the substitution models were selected using bModelTest (<https://github.com/BEAST2-Dev/bModelTest/>)<sup>115</sup>. The best-fitting model of substitution in both cases was the transversion model (TVM) with gamma-distributed rate heterogeneity (G) and a given proportion of invariable sites (I). The log-normal relaxed clock with the Birth–Death speciation model was set as priors following the results of path sampling selection<sup>116</sup>. The recent cross-validated analyses of molecular clock, based on geological events for *G. fossarum*<sup>58,59</sup> and other gammarids (e. g.<sup>5,28,29</sup>), indicated that the widely used arthropod COI rate of 0.0115 substitutions per site per My<sup>117</sup> is useful in linking present-day patterns with historical processes. Thus, we have used it to calibrate the molecular clock in our phylogeny reconstruction. Four MCMC chains were run for 50 M iterations and sampled every 5000 iterations. The Effective Sampling Size (ESS) of each parameter was verified to be above 200 in Tracer 1.7 (<https://beast.community/tracer>)<sup>118</sup>. The runs were combined in LogCombiner 2.4.8, with 25% burn-in, and the maximum clade credibility chronogram was annotated using TreeAnnotator 2.4.8, both programs being part of BEAST 2.4.8 (<https://www.beast2.org/>)<sup>114</sup>.

In order to provide additional support for the BI topology, we have also reconstructed a phylogeny using the Maximum Likelihood approach (ML) in RAXML HPC 8.2.9 (<https://cme.h-its.org/exelixis/web/software/raxml/>)<sup>113</sup>. The best-scoring ML tree was searched under the GTR + G substitution model which was applied to each codon partition. Statistical support was estimated with the GTRCAT model of rate heterogeneity and 1,000 rapid bootstrap (RBS) replicates<sup>119</sup>.

The history of diversification was explored with the help of a Lineage Through Time (LTT) plot, generated with Tracer 1.7<sup>118</sup>, using 1,500 post burn-in trees that resulted from the BEAST analysis on the BIN dataset.

**Definition of major clades.** We defined the major clades following studies from Copilaş-Ciocianu and Petrusek<sup>58</sup> and Copilaş-Ciocianu et al.<sup>59</sup> which were based on mitochondrial (COI and 16S) and nuclear markers (18S, 28S and EF1α). These studies have robustly recovered these deep clades in multilocus concatenation and species tree analyses. We acknowledge that, in the present study, node supports are not high in all cases as only the highly variable COI marker was used, which has reduced ability to resolve deep nodes (see “Results”). In addition, the names and acronyms were kept in order to maintain continuity with clade nomenclature from these previous studies<sup>58,59</sup>. Such clades refer to geography (i.e. SEE: south-eastern Europe), illustrating rather in which part of Europe a given clade has a centre of its distribution than willing to match any predefined strict geographic units. In addition, it is to be noted that extending the present study analysis at a continental scale might also have been associated with extending the original distribution of some clades.

**Geographic patterns of genetic diversity.** The spatial diversity pattern of *G. fossarum* in Europe was illustrated with genetic landscape shapes generated with Alleles in Space (AIS; <https://www.marksgeneticsoftware.net>) software<sup>120</sup>. The genetic landscape visualizes the abrupt transitions between populations and groups of populations characterized by divergent haplotypes. The analysis followed the approach already applied to amphipods (e.g.<sup>3</sup>), i.e. converting the genetic distances between sites into a hypsometric map. First, using the AIS software, the genetic distances between sampling sites were calculated based on the COI sequences longer than 496 bp (in order to obtain a gap-free alignment) and connected into a network based on the Delaunay Triangulation. The genetic distance values were set in the midpoints of each connection in the network. The raw genetic distances acquired from the program were interpolated afterwards and the matrix of the ‘elevation’ values, with their respective latitude and longitude coordinates, was then imported into QGIS (<https://qgis.osgeo.org>) software to produce a genetic divergence surface image using the Triangulated Irregular Network (TIN) algorithm. The resulting image was plotted onto a relief map of Europe to create a final map in which the hypsometric tints reflect the genetic distance between population pairs. Two maps were constructed, one that included all the sampling sites (498) and all haplotypes (691) and another one which excluded the highly divergent lineages endemic to the Carpathian, Dinaric and Pannonian regions (the hotspots of the most ancient divergences, see “Results”), to better illustrate the level of genetic divergence in the remaining area. A total of 336 sites and 469 haplotypes were included in this latter analysis.

**Ethical approval.** All the applicable international, national, and/or institutional guidelines for the care and use of animals were followed and were in agreement with recommendations from the “Comité d’Ethique et de l’Expérimentation Animale (C2EADijon Grand Campus) from Dijon University, France. No experiments on alive animals were performed for this study.

## Data availability

All newly produced COI sequences are released in GenBank under accession numbers: MT978656–MT980726. The 463 sequences from Lagrue et al.<sup>57</sup> which were initially unreleased in Genbank are now available under accession numbers: MT411018–MT411480. Sequences and metadata are available in BOLD dataset DS-GFOSCDEU ([dx.doi.org/10.5883/DS-GFOSCDEU](https://dx.doi.org/10.5883/DS-GFOSCDEU)). All the new material used in this study has been stored in the permanent collection of either Biogeosciences Laboratory, University of Burgundy-Franche-Comté or the Department of Invertebrate Zoology and Hydrobiology, University of Lodz, Poland.

Received: 15 May 2020; Accepted: 16 September 2020

Published online: 06 October 2020

## References

- Adams, M., Raadik, T. A., Burrige, C. P. & Georges, A. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room?. *Syst. Biol.* **63**, 518–533 (2014).
- Pérez-Ponce de León, G. & Poulin, R. Taxonomic distribution of cryptic diversity among metazoans: not so homogeneous after all. *Biol. Lett.* **12**, 20160371. <https://doi.org/10.1098/rsbl.2016.0371> (2016).
- Pfenninger, M. & Schwenk, K. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evol. Biol.* **7**, 121. <https://doi.org/10.1186/1471-2148-7-121> (2007).
- Keshavmurthy, S. et al. DNA barcoding reveals the coral “laboratory-rat”, *Stylophora pistillata* encompasses multiple identities. *Sci. Rep.* **3**, 1520. <https://doi.org/10.1038/srep01520> (2013).
- Mamos, T., Wattier, R., Burzyński, A. & Grabowski, M. The legacy of a vanished sea: a high level of diversification within a European freshwater amphipod species complex driven by 15 My of Paratethys regression. *Mol. Ecol.* **25**, 785–810 (2016).
- Bickford, D. et al. Cryptic species as a window on diversity and conservation. *Trends Ecol. Evol.* **22**, 148–155 (2007).
- Brodersen, J. & Seehausen, O. Why evolutionary biologists should get seriously involved in ecological monitoring and applied biodiversity assessment programs. *Evol. Appl.* **7**, 968–983 (2014).
- Eme, D. et al. Do cryptic species matter in macroecology? Sequencing European groundwater crustaceans yields smaller ranges but does not challenge biodiversity determinants. *Ecography* **41**, 424–436 (2018).
- Galipaud, M., Gauthey, Z., Turlin, J., Bollache, L. & Lagrue, C. Mate choice and male–male competition among morphologically cryptic but genetically divergent amphipod lineages. *Behav. Ecol. Sociobiol.* **69**, 1907–1916 (2015).
- Galipaud, M., Bollache, L. & Lagrue, C. Variations in infection levels and parasite-induced mortality among sympatric cryptic lineages of native amphipods and a congeneric invasive species: Are native hosts always losing?. *Int. J. Parasitol. Parasites Wildl.* **6**, 439–447 (2017).
- Westram, A. M., Baumgartner, C., Keller, I. & Jokela, J. Are cryptic host species also cryptic to parasites? Host specificity and geographical distribution of acanthocephalan parasites infecting freshwater Gammarus. *Infect. Genet. Evol.* **11**, 1083–1090 (2011).
- Fišer, Ž., Altermatt, F., Zakšek, V., Knapic, T. & Fišer, C. Morphologically cryptic amphipod species are “ecological clones” at regional but not at local scale: a case study of four *Niphargus* species. *PLoS ONE* **10**, e0134384. <https://doi.org/10.1371/journal.pone.0134384> (2015).
- Eisenring, M., Altermatt, F., Westram, A. M. & Jokela, J. Habitat requirements and ecological niche of two cryptic amphipod species at landscape and local scales. *Ecosphere* **7**, e01319. <https://doi.org/10.1002/ecs2.1319> (2016).
- Westram, A. M., Jokela, J. & Keller, I. Hidden biodiversity in an ecologically important freshwater amphipod: Differences in genetic structure between two cryptic species. *PLoS ONE* **8**, e69576. <https://doi.org/10.1371/journal.pone.0069576> (2013).
- Mutanen, M., Kaila, L. & Tabell, J. Wide-ranging barcoding aids discovery of one-third increase of species richness in presumably well-investigated moths. *Sci. Rep.* **3**, 2901. <https://doi.org/10.1038/srep02901> (2013).
- Cook, B. D., Page, T. J. & Hughes, J. M. Importance of cryptic species for identifying ‘representative’ units of biodiversity for freshwater conservation. *Biol. Conserv.* **141**, 2821–2831 (2008).
- Fišer, C., Robinson, C. T. & Malard, F. Cryptic species as a window into the paradigm shift of the species concept. *Mol. Ecol.* **27**, 613–635 (2018).
- Bálint, M., Barnard, P. C., Schmitt, T., Ujvárosi, L. & Popescu, O. Differentiation and speciation in mountain streams: a case study in the caddisfly *Rhyacophila aquitanica* (Trichoptera). *J. Zool. Syst. Evol. Res.* **46**, 340–345 (2008).
- Major, K., Soucek, D. J., Giordano, R., Wetzel, M. J. & Soto-Adames, F. The common ecotoxicology laboratory strain of *Hyaella azteca* is genetically distinct from most wild strains sampled in eastern North America: common lab strain of *H. azteca* is distinct from wild strains. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.2355> (2013).
- Feckler, A., Thielsch, A., Schwenk, K., Schulz, R. & Bundschuh, M. Differences in the sensitivity among cryptic lineages of the *Gammarus fossarum* complex. *Sci. Total Environ.* **439**, 158–164 (2012).
- Feckler, A. et al. Cryptic species diversity: an overlooked factor in environmental management?. *J. Appl. Ecol.* **51**, 958–967 (2014).
- Caputo, D. R., Robson, S. C., Werner, I. & Ford, A. T. Complete transcriptome assembly and annotation of a critically important amphipod species in freshwater ecotoxicological risk assessment: *Gammarus fossarum*. *Environ. Int.* **137**, 105319 (2020).
- Cogne, Y. et al. De novo transcriptomes of 14 gammarid individuals for proteogenomic analysis of seven taxonomic groups. *Sci. Data* **6**, 184. <https://doi.org/10.1038/s41597-019-0192-5> (2019).
- Costa-Silva, G. J., Rodriguez, M. S., Roxo, F. F., Foresti, F. & Oliveira, C. Using different methods to access the difficult task of delimiting species in a complex Neotropical hyperdiverse group. *PLoS ONE* **10**, e0135075 (2015).
- Dincă, V. et al. DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Sci. Rep.* **5**, 12395. <https://doi.org/10.1371/journal.pone.0135075> (2015).
- Trontelj, P. et al. A molecular test for cryptic diversity in ground water: how large are the ranges of macro-stygobionts?. *Freshw. Biol.* **54**, 727–744 (2009).
- Jackson, J. K. et al. Cryptic biodiversity in streams: a comparison of macroinvertebrate communities based on morphological and DNA barcode identifications. *Freshw. Sci.* **33**, 312–324 (2014).
- Grabowski, M., Mamos, T., Bacela-Spychalska, K., Rewicz, T. & Wattier, R. A. Neogene paleogeography provides context for understanding the origin and spatial distribution of cryptic diversity in a widespread balkan freshwater amphipod. *PeerJ* <https://doi.org/10.7717/peerj.3016> (2017).
- Grabowski, M., Wysocka, A. & Mamos, T. Molecular species delimitation methods provide new insight into taxonomy of the endemic gammarid species flock from the ancient Lake Ohrid. *Zool. J. Linn. Soc.* <https://doi.org/10.1093/zoolinnean/zlw025> (2017).
- Hogg, I. D., Larose, C., de Lafontaine, Y. & Doe, K. G. Genetic evidence for a *Hyaella* species complex within the Great Lakes—St. Lawrence River drainage basin: implications for ecotoxicology and conservation biology. *Can. J. Zool.* **76**, 1134–1140 (1998).
- Hogg, I. D., Stevens, M. I., Schnabel, K. E. & Ann Chapman, M. Deeply divergent lineages of the widespread New Zealand amphipod *Paracalliope fluviatilis* revealed using allozyme and mitochondrial DNA analyses. *Freshw. Biol.* **51**, 236–248 (2006).

32. Katouzian, A.-R. *et al.* Drastic underestimation of amphipod biodiversity in the endangered Irano-Anatolian and Caucasus biodiversity hotspots. *Sci. Rep.* **6**, 22507. <https://doi.org/10.1038/srep22507> (2016).
33. Mamos, T., Wattier, R., Majda, A., Sket, B. & Grabowski, M. Morphological vs. molecular delineation of taxa across montane regions in Europe: the case study of *Gammarus balcanicus* Schäferna, (Crustacea: Amphipoda). *J. Zool. Syst. Evol. Res.* **52**, 237–248 (2014).
34. Murphy, N. P., Adams, M., Guzik, M. T. & Austin, A. D. Extraordinary micro-endemism in Australian desert spring amphipods. *Mol. Phylogenet. Evol.* **66**, 645–653 (2013).
35. Murphy, N. P., King, R. A. & Delean, S. Species, ESUs or populations? Delimiting and describing morphologically cryptic diversity in Australian desert spring amphipods. *Invertebr. Syst.* **29**, 457 (2015).
36. Seidel, R. A., Lang, B. K. & Berg, D. J. Phylogeographic analysis reveals multiple cryptic species of amphipods (Crustacea: Amphipoda) in Chihuahuan Desert springs. *Biol. Conserv.* **142**, 2303–2313 (2009).
37. Sutherland, D. L., Hogg, I. D. & Waas, J. R. Phylogeography and species discrimination in the *Paracalliopefluviatilis* species complex (Crustacea: Amphipoda): can morphologically similar heterospecifics identify compatible mates?: Mate discrimination in *P. fluviatilis*. *Biol. J. Linn. Soc.* **99**, 196–205 (2009).
38. Witt, J. D. S., Threlloff, D. L. & Hebert, P. D. N. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Mol. Ecol.* **15**, 3073–3082 (2006).
39. Witt, J. D. S. & Hebert, P. D. N. Cryptic species diversity and evolution in the amphipod genus *Hyaella* within central glaciated North America: a molecular phylogenetic approach. *Can. J. Fish. Aquat. Sci.* **57**, 12 (2000).
40. Blackman, R. *et al.* Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples—first record of *Gammarus fossarum* in the UK. *Aquat. Invasions* **12**, 177–189 (2017).
41. Dangles, O., Gessner, M. O., Guerold, F. & Chauvet, E. Impacts of stream acidification on litter breakdown: implications for assessing ecosystem functioning. *J. Appl. Ecol.* **41**, 365–378 (2004).
42. Felten, V., Tixier, G., Guérol, F., De Crespín De Billy, V. & Dangles, O. Quantification of diet variability in a stream amphipod: implications for ecosystem functioning. *Fundam. Appl. Limnol. Arch. Für Hydrobiol.* **170**, 303–313 (2008).
43. Besse, J.-P., Geffard, O. & Coquery, M. Relevance and applicability of active biomonitoring in continental waters under the Water Framework Directive. *TrAC Trends Anal. Chem.* **36**, 113–127 (2012).
44. Schmidlin, L., von Fumetti, S. & Nagel, P. Temperature effects on the feeding and electron transport system (ETS) activity of *Gammarus fossarum*. *Aquat. Ecol.* **49**, 71–80 (2015).
45. Labaude, S., Rigaud, T. & Cézilly, F. Additive effects of temperature and infection with an acanthocephalan parasite on the shredding activity of *Gammarus fossarum* (Crustacea: Amphipoda): the importance of aggregative behavior. *Glob. Change Biol.* **23**, 1415–1424 (2017).
46. Kunz, P. Y., Kienle, C. & Gerhardt, A. *Gammarus* spp. in aquatic ecotoxicology and water quality assessment: Toward integrated multilevel tests. In *Reviews of Environmental Contamination and Toxicology* Vol. 205 (ed. Whitacre, D. M.) 1–76 (Springer, New York, 2010).
47. Mehennaoui, K. *et al.* *Gammarus fossarum* (Crustacea, Amphipoda) as a model organism to study the effects of silver nanoparticles. *Sci. Total Environ.* **566–567**, 1649–1659 (2016).
48. Besse, J.-P. *et al.* Caged *Gammarus fossarum* (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: towards the determination of threshold values. *Water Res.* **47**, 650–660 (2013).
49. Gouveia, D. *et al.* Ecotoxic-Proteomics for aquatic environmental monitoring: First in situ application of a new proteomics-based multibiomarker assay using caged amphipods. *Environ. Sci. Technol.* **51**, 13417–13426 (2017).
50. Trapp, J. *et al.* Proteogenomics of *Gammarus fossarum* to document the reproductive system of amphipods. *Mol. Cell. Proteomics* **13**, 3612–3625 (2014).
51. Sanchez-Thirion, K. *et al.* High food quality increases infection of *Gammarus pulex* (Crustacea: Amphipoda) by the acanthocephalan parasite *Pomphorhynchus laevis*. *Int. J. Parasitol.* **49**, 805–817 (2019).
52. Bigot-Clivot, A. *et al.* Bioaccumulation of toxoplasma and cryptosporidium by the freshwater crustacean *Gammarus fossarum*: involvement in biomonitoring surveys and trophic transfer. *Ecotoxicol. Environ. Saf.* **133**, 188–194 (2016).
53. Chen, H.-Y., Grabner, D. S., Nachev, M., Shih, H.-H. & Sures, B. Effects of the acanthocephalan *Polymorphus minutus* and the microsporidian *Dictyocoela duebenum* on energy reserves and stress response of cadmium exposed *Gammarus fossarum*. *PeerJ* **3**, e1353. <https://doi.org/10.7717/peerj.1353> (2015).
54. Müller, J. Mitochondrial DNA variation and the evolutionary history of cryptic *Gammarus fossarum* types. *Mol. Phylogenet. Evol.* **15**, 260–268 (2000).
55. Weiss, M., Macher, J. N., Seefeldt, M. A. & Leese, F. Molecular evidence for further overlooked species within the *Gammarus fossarum* complex (Crustacea: Amphipoda). *Hydrobiologia* **721**, 165–184 (2014).
56. Weiss, M. & Leese, F. Widely distributed and regionally isolated! Drivers of genetic structure in *Gammarus fossarum* in a human-impacted landscape. *BMC Evol. Biol.* **16**, 153 (2016).
57. Lagrue, C. *et al.* Confrontation of cryptic diversity and mate discrimination within *Gammarus pulex* and *Gammarus fossarum* species complexes. *Freshw. Biol.* **59**, 2555–2570 (2014).
58. Copilaş-Ciocianu, D. & Petrusek, A. The southwestern Carpathians as an ancient centre of diversity of freshwater gammarid amphipods: insights from the *Gammarus fossarum* species complex. *Mol. Ecol.* **24**, 3980–3992 (2015).
59. Copilaş-Ciocianu, D., Rutová, T., Pafil, P. & Petrusek, A. Epigean gammarids survived millions of years of severe climatic fluctuations in high latitude refugia throughout the Western Carpathians. *Mol. Phylogenet. Evol.* **112**, 218–229 (2017).
60. Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, S96–S99 (2003).
61. Kekkonen, M. & Hebert, P. D. N. DNA barcode-based delineation of putative species: efficient start for taxonomic workflows. *Mol. Ecol. Resour.* **14**, 706–715 (2014).
62. Copilaş-Ciocianu, D. & Petrusek, A. Phylogeography of a freshwater crustacean species complex reflects a long-gone archipelago. *J. Biogeogr.* **44**, 421–432 (2017).
63. Grabner, D. S. *et al.* Invaders, natives and their enemies: distribution patterns of amphipods and their microsporidian parasites in the Ruhr Metropolis, Germany. *Parasit. Vectors* **8**, 419 (2015).
64. Copilaş-Ciocianu, D., Zimţă, A. & Petrusek, A. Integrative taxonomy reveals a new *Gammarus* species (Crustacea, Amphipoda) surviving in a previously unknown southeast European glacial refugium. *J. Zool. Syst. Evol. Res.* **57**, 272–297 (2019).
65. Copilaş-Ciocianu, D., Zimţă, A.-A., Grabowski, M. & Petrusek, A. Survival in northern microrefugia in an endemic Carpathian gammarid (Crustacea: Amphipoda). *Zool. Scr.* **47**, 357–372 (2018).
66. Hou, Z., Li, J. & Li, S. Diversification of low dispersal crustaceans through mountain uplift: a case study of *Gammarus* (Amphipoda: Gammaridae) with descriptions of four novel species: Diversification of *Gammarus* species. *Zool. J. Linn. Soc.* **170**, 591–633 (2014).
67. Wellborn, G. A. & Cothran, R. D. Niche diversity in crustacean cryptic species: complementarity in spatial distribution and predation risk. *Oecologia* **154**, 175–183 (2007).
68. Delić, T., Svára, V., Coleman, C. O., Trontelj, P. & Fišer, C. The giant cryptic amphipod species of the subterranean genus *Niphargus* (Crustacea, Amphipoda). *Zool. Scr.* **46**, 740–752 (2017).

69. McNerney, C. E. *et al.* The ancient Britons: groundwater fauna survived extreme climate change over tens of millions of years across NW Europe. *Mol. Ecol.* **23**, 1153–1166 (2014).
70. Cooper, S. J. B. *et al.* Subterranean archipelago in the Australian arid zone: mitochondrial DNA phylogeography of amphipods from central Western Australia: Phylogeography of subterranean amphipods. *Mol. Ecol.* **16**, 1533–1544 (2007).
71. Ethridge, J. Z., Gibson, J. R. & Nice, C. C. Cryptic diversity within and amongst spring-associated *Stygobromus* amphipods (Amphipoda: Crangonyctidae): *Stygobromus* amphipod cryptic diversity. *Zool. J. Linn. Soc.* **167**, 227–242 (2013).
72. Beermann, J. *et al.* Cryptic species in a well-known habitat: applying taxonomics to the amphipod genus *Epimeria* (Crustacea, Peracarida). *Sci. Rep.* **8**, 6893. <https://doi.org/10.1038/s41598-018-25225-x> (2018).
73. Baird, H. P., Miller, K. J. & Stark, J. S. Evidence of hidden biodiversity, ongoing speciation and diverse patterns of genetic structure in giant Antarctic amphipods: Genetic diversity in antarctic amphipods. *Mol. Ecol.* **20**, 3439–3454 (2011).
74. Havermans, C. *et al.* Genetic and morphological divergences in the cosmopolitan deep-sea amphipod *Eurythenes gryllus* reveal a diverse abyss and a bipolar species. *PLoS ONE* **8**, e74218. <https://doi.org/10.1371/journal.pone.0074218> (2013).
75. Havermans, C. Have we so far only seen the tip of the iceberg? Exploring species diversity and distribution of the giant amphipod *Eurythenes*. *Biodiversity* **17**, 12–25 (2016).
76. Coleman, C. O. Taxonomy in times of the taxonomic impediment—examples from the community of experts on amphipod crustaceans. *J. Crustac. Biol.* **35**, 729–740 (2015).
77. Arfianti, T., Wilson, S. & Costello, M. J. Progress in the discovery of amphipod crustaceans. *PeerJ* **6**, e5187 (2018).
78. Popov, S. V. *et al.* Lithological-Paleogeographic maps of Paratethys—10 maps late Eocene to Pliocene. *Courier Forschung-Institut Senckberg* 1–46 (2004).
79. Frenzel, B., Pécsi, M. & Velichko, A. A. *Atlas of Paleoclimates and Paleoenvironments of the Northern Hemisphere: Late Pleistocene, Holocene* (Geographical Research Institute Hungarian Academy of Science; G. Fischer, Budapest, 1992).
80. Macneil, C., Dick, J. T. A. & Elwood, R. W. The trophic ecology of freshwater *Gammarus* spp (Crustacea; Amphipoda): Problems and perspective concerning the functional feeding group concept. *Biol. Rev.* **72**, 349–364 (2007).
81. Rudolph, K., Coleman, C. O., Mamos, T. & Grabowski, M. Description and post-glacial demography of *Gammarus jazdzewskii* sp. nov. (Crustacea: Amphipoda) from Central Europe. *Syst. Biodivers.* **16**, 587–603 (2018).
82. Westram, A. M., Jokela, J., Baumgartner, C. & Keller, I. Spatial distribution of cryptic species diversity in European freshwater amphipods (*Gammarus fossarum*) as revealed by pyrosequencing. *PLoS ONE* **6**, e23879. <https://doi.org/10.1371/journal.pone.0023879> (2011).
83. Lefebvre, T. *et al.* Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments: cryptic and dynamic evolution in subsurface. *Mol. Ecol.* **15**, 1797–1806 (2006).
84. Lefebvre, T., Douady, C. J., Malard, F. & Gibert, J. Testing dispersal and cryptic diversity in a widely distributed groundwater amphipod (*Niphargus rhenorhodanensis*). *Mol. Phylogenet. Evol.* **42**, 676–686 (2007).
85. Raffard, A., Santoul, F., Cucherousset, J. & Blanchet, S. The community and ecosystem consequences of intraspecific diversity: a meta-analysis: the ecological effects of intraspecific diversity. *Biol. Rev.* **94**, 648–661 (2019).
86. Jabłońska, A., Wrzesińska, W., Zawal, A., Pešić, V. & Grabowski, M. Long-term within-basin isolation patterns, different conservation units, and interspecific mitochondrial DNA introgression in an amphipod endemic to the ancient Lake Skadar system, Balkan Peninsula. *Freshw. Biol.* **65**, 209–225 (2020).
87. Stein, E. D., Martinez, M. C., Stiles, S., Miller, P. E. & Zakharov, E. V. Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States?. *PLoS ONE* **9**, e95525. <https://doi.org/10.1371/journal.pone.0095525> (2014).
88. Valentini, A., Pompanon, F. & Taberlet, P. DNA barcoding for ecologists. *Trends Ecol. Evol.* **24**, 110–117 (2009).
89. Weigand, H. *et al.* DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Sci. Total Environ.* **678**, 499–524 (2019).
90. Delić, T., Trontelj, P., Rendoš, M. & Fišer, C. The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Sci. Rep.* **7**, 3391. <https://doi.org/10.1038/s41598-017-02938-z> (2017).
91. Wysocka, A. *et al.* Origin of the Lake Ohrid gammarid species flock: ancient local phylogenetic lineage diversification. *J. Biogeogr.* **41**, 1758–1768 (2014).
92. Esmaeili-Rineh, S., Sari, A., Deliç, T., Moškrić, A. & Fišer, C. Molecular phylogeny of the subterranean genus *Niphargus* (Crustacea: Amphipoda) in the Middle East: a comparison with European Niphargids. *Zool. J. Linn. Soc.* **175**, 812–826 (2015).
93. Hou, Z., Sket, B., Fišer, C. & Li, S. Eocene habitat shift from saline to freshwater promoted Tethyan amphipod diversification. *Proc. Natl. Acad. Sci.* **108**, 14533–14538 (2011).
94. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. 8.
95. Karaman, G. & Pinkster, S. Freshwater *Gammarus* species from Europe, North Africa and adjacent regions of Asia (Crustacea-Amphipoda). Part I. *Gammarus pulex*-group and related species. *Bijdragen tot de Dierkunde* 1–97 (1977).
96. Piscart, C. & Bollache, L. *Crustacés amphipodes de surface : gammarus d'eau douce* (Association Française de Limnologie, Thonon-les-Bains, 2012).
97. Eggers, T. & Martens, A. A key of the freshwater Amphipods of Germany. *Lauterbornia* 1–68 (2001).
98. Jazdzewski, K. Morfologia, taksonomia i występowanie w Polsce kielży z rodzajów *Gammarus* Fabr. i *Chaetogammarus* Mart. (Crustacea, Amphipoda). *Acta Universitatis Lodziansis* 1–187 (1975).
99. Karaman, G. S. *Crustacea* (Amphipoda di acqua dolce. Calderini, Bologna, 1993).
100. Hillis, D. M., Moritz, C. & Mable, B. K. *Molecular Systematics* (Sinauer Associates Inc, Sunderland, 1996).
101. Casquet, J., Thebaud, C. & Gillespie, R. G. Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders: Quick and easy dsDNA extraction for barcoding. *Mol. Ecol. Resour.* **12**, 136–141 (2012).
102. Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol. Mar. Biol. Biotechnol.* **6**, 294–299 (1994).
103. Costa, F. O., Henzler, C. M., Lunt, D. H., Whiteley, N. M. & Rock, J. Probing marine *Gammarus* (Amphipoda) taxonomy with DNA barcodes. *Syst. Biodivers.* **7**, 365–379 (2009).
104. Astrin, J. J. & Stüben, P. E. Phylogeny in cryptic weevils: molecules, morphology and new genera of western Palearctic Cryptorhynchinae (Coleoptera:Curculionidae). *Invertebr. Syst.* **22**, 503–522 (2008).
105. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
106. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
107. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
108. Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Mol. Ecol.* **21**, 1864–1877 (2012).
109. Ratnasingham, S. & Hebert, P. D. N. A DNA-based registry for all animal species: the barcode index number (BIN) System. *PLoS ONE* **8**, e66213. <https://doi.org/10.1371/journal.pone.0066213> (2013).
110. Zhang, J., Kapli, P., Pavlidis, P. & Stamatakis, A. A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**, 2869–2876 (2013).

111. Puillandre, N. *et al.* Large-scale species delimitation method for hyperdiverse groups: large-scale species delimitation. *Mol. Ecol.* **21**, 2671–2691 (2012).
112. Ratnasingham, S. & Hebert, P. D. N. BOLD: The Barcode of life data system (<https://www.barcodinglife.org>). *Mol. Ecol. Notes* **7**, 355–364 (2007).
113. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
114. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537> (2014).
115. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42 (2017).
116. Baele, G., Lemey, P. & Vansteelandt, S. Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinform.* **14**, 85 (2013).
117. Brower, A. V. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci.* **91**, 6491–6495 (1994).
118. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
119. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
120. Miller, M. P. Alleles In Space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *J. Hered.* **96**, 722–724 (2005).

## Acknowledgements

The authors thank the French National Agency for Research for funding (ANR-13-BSV7-0004 ANR MULTISTRESS; ANR-14-CE21-0006 PROTEOGAM), the Adour-Garonne water agency (program Dylaq) and the French Office for Biodiversity (OFB) for funding. Study was financially supported by the Polish Ministry of Science and Education, grant number N N303 579439 and partially from the internal funds of the University of Lodz. Additional financing for MJ's stay in Poland was provided by the Short-Term Scientific Mission (STSM) of COST Action CA15219-Developing new genetic tools for bioassessment of aquatic ecosystems in Europe—DNAqua-Net (Grant no. 39779)".

## Author contributions

Designed the study: R.W. and M.G. Sampling: R.W., M.G., V.F., C.P., A.C., M.D., M.J., K.Z., T.M. and D.C.-C. Performed the laboratory work: R.W., A.O., M.J., D.C.-C., T.M., T.Re. and A.W. Contributed reagents: R.W., T.Ri and M.G. Data analysis: R.W., T.M., D.C.-C., M.G., and M.J. Discussion and interpretation of the results: R.W., T.M., M.G., D.C.-C., T. Ri and M.J. Wrote the manuscript: R.W. and M.G. All authors have read and approved the final version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-73739-0>.

**Correspondence** and requests for materials should be addressed to R.W. or M.G.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020