



HAL
open science

Widespread infection, diversification and old host associations of *Nosema* Microsporidia in European freshwater gammarids (Amphipoda)

Karolina Bacela-Spychalska, Remi Wattier, Maria Teixeira, Richard Cordaux, Adrien Quiles, Michal Grabowski, Piotr Wroblewski, Mykola Ovcharenko, Daniel Grabner, Dieter Weber, et al.

► To cite this version:

Karolina Bacela-Spychalska, Remi Wattier, Maria Teixeira, Richard Cordaux, Adrien Quiles, et al.. Widespread infection, diversification and old host associations of *Nosema* Microsporidia in European freshwater gammarids (Amphipoda). 2023. hal-04178588v1

HAL Id: hal-04178588

<https://hal.science/hal-04178588v1>

Preprint submitted on 8 Aug 2023 (v1), last revised 4 Sep 2023 (v3)

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.

1 Widespread infection, diversification and old host associations of *Nosema*
2 Microsporidia in European freshwater gammarids (Amphipoda)

3
4 Karolina Bacela-Spychalska^{1,*¶}, Remi Wattier^{2¶}, Maria Teixeira², Richard Cordaux^{3,¶a}, Adrien
5 Quiles^{1,2}, Michal Grabowski¹, Piotr Wroblewski⁴, Mykola Ovcharenko^{4,5}, Daniel Grabner⁶, Dieter
6 Weber^{7,8}, Alexander M Weigand⁸, and Thierry Rigaud²

7
8 ¹ Department of Invertebrate Zoology and Hydrobiology, Faculty of Biology and Environmental
9 Protection, University of Lodz, Lodz, Poland

10 ² Laboratoire Biogéosciences, UMR CNRS 6282, Université de Bourgogne, Dijon, France

11 ³ Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, UMR
12 CNRS 7267, Université de Poitiers, Poitiers Cedex 9, France

13 ⁴ Department of Ecology and Evolution of Parasitism, Witold Stefanski Institute of Parasitology,
14 Polish Academy of Science, Warsaw, Poland

15 ⁵ Institute of Biology and Earth Sciences, Pomeranian University in Slupsk, Slupsk, Poland

16 ⁶ Aquatic Ecology and Centre for Water and Environmental Research, University of Duisburg-
17 Essen, Essen, Germany

18 ⁷ Senckenberg Deutsches Entomologisches Institut Müncheberg, Germany

19 ⁸ Musée National d'Histoire Naturelle Luxembourg, Luxembourg, Luxembourg

20 ^{¶a}Current Address: CNRS, IRD, UMR Évolution Génomes Comportement Écologie, Université
21 Paris-Saclay, Gif-sur-Yvette, France

22
23 [¶]These authors contributed equally to this study

24 * Corresponding author: karolina.bacela@biol.uni.lodz.pl

25
26
27 **Short title:** Diversity of microsporidian *Nosema* spp. in freshwater gammarids (Amphipoda)

28 **Abstract**

29 The microsporidian genus *Nosema* is primarily known to infect insects of economic importance
30 stimulating high research interest, while other hosts remain understudied. *Nosema granulosis* is
31 one of the formally described *Nosema* species infecting amphipod crustaceans, being known to
32 infect only two host species.

33 Our first aim was to characterize *Nosema* spp. infections in different amphipod species from
34 various European localities using the small subunit ribosomal DNA (SSU) marker. Second, we
35 aimed to assess the phylogenetic diversity, host specificity and to explore the evolutionary history
36 that may explain the diversity of gammarid-infecting *Nosema* lineages by performing a
37 phylogenetic reconstruction based on RNA polymerase II subunit B1 (RPB1) gene sequences. For
38 the host species *Gammarus balcanicus*, we also analyzed sex ratios of progenies to test for sex
39 ratio distortion in relation with *Nosema* infection.

40 We identified *Nosema* spp. in 316 individuals from nine amphipod species being widespread in
41 Europe. The RPB1-based phylogenetic reconstruction using newly reported sequences and
42 available data from other invertebrates identified 39 haplogroups being associated with
43 amphipods. These haplogroups clustered into five clades (A-E) that did not form a single
44 amphipod-infecting monophyletic group. Closely related sister clades C and D correspond to
45 *Nosema granulosis*. Clades A, B and E might represent unknown *Nosema* species infecting
46 amphipods. Host specificity seemed to be variable with some clades being restricted to single
47 hosts, and some that could be found in several host species. We show that *Nosema* parasite richness
48 in gammarid hosts is much higher than expected, illustrating the advantage of the use of RPB1
49 marker over SSU. Finally, we found no hint of sex ratio distortion in *Nosema* clade A infecting *G.*
50 *balcanicus*. This study shows that *Nosema* spp. are abundant, widespread and diverse in European
51 gammarids. Thus, *Nosema* is as diverse in aquatic as in terrestrial hosts.

52

53 **Author Summary**

54 The intracellular microsporidian parasites of the genus *Nosema* are important pathogens as they
55 infect insects of economic importance such as pollinators or domestic silk moths. However, they
56 have been overlooked in other groups of invertebrates such as crustaceans, where only few *Nosema*
57 spp. were detected, including *Nosema granulosis* in two gammarid amphipod hosts. Here, using
58 molecular tools, we detected infections by *Nosema* spp. in nine freshwater amphipod species from
59 various European localities. These infections are common, widespread, and the phylogenetic
60 analysis revealed that some *Nosema* lineages are specific to certain host species, while the others
61 are not restricted to single hosts. We also confirms that there are several *Nosema* lineages infecting
62 gammarid crustaceans, not only *N. granulosis*, that can be new candidate species of *Nosema*. We
63 conclude that *Nosema* is as diverse in aquatic as in terrestrial hosts.

64 **Keywords:** Host-parasite relationships, co-diversification, phylogeny, RPB1, SSU *rDNA*,
65 *Gammarus*

66 Introduction

67 Since the mid-19th century, Microsporidia, a large group of obligate intracellular eukaryotic
68 microparasites related to fungi, has aroused interest in both basic and applied studies [1–3]. In this
69 context, the microsporidian genus *Nosema* (Nosematida) holds in a seminal place. The first
70 formally described microsporidium was *Nosema bombycis* [4]. Since then, 16 *Nosema* spp. or
71 closely-related species have been described infecting insects [5]. All are pathogenic to some
72 extent, with some species having a negative economic impact and some have been used for insect
73 pest control [6,7]. There are strains of *Nosema* that are horizontally transmitted (HT) [8], while
74 others are known to also use vertical transmission (VT) in their life cycle [9].

75 Few *Nosema* spp. are known to infect freshwater decapod or amphipod crustaceans based on
76 morphological and molecular data [10–18]. Many amphipod species are known to serve as hosts
77 for many other microsporidian parasites, such as *Dictyocoela* spp. and *Cucumispora* spp.
78 [summarised in 18], and one *Nosema* species, *N. granulosis*, which has been reported from only
79 two amphipod species. This parasite, unlike *Nosema* of insects, has not been recognized as a
80 pathogen so far, while in other crustaceans the impact on host fitness is unknown. *Nosema*
81 *granulosis* was first described in the amphipod *Gammarus duebeni* [10]. Infected *G. duebeni*
82 females transmit *N. granulosis* directly to their offspring through infected eggs [19–21]. Vertical
83 transmission occurs only via female hosts because eggs contain enough cytoplasm in which *N.*
84 *granulosis* can live and be transmitted [22]. Under the selective pressure of this transmission
85 asymmetry, traits increasing the proportion of female hosts is selectively advantageous for the
86 parasites. Therefore, the parasite's ability for turning genetic males into functional females
87 (feminization) was selected [10,23–26]. As a consequence of this feminization, female-biased sex
88 ratios can be found in populations of *G. duebeni* and *Gammarus roeselii* [20,21,23,27] and only a
89 few males were found to be infected by *N. granulosis* in *G. roeselii* (presumably resulting from
90 failed feminisation) [23]. These vertically-transmitted infections cause little pathogenicity to their

91 hosts [10,23,28,29], even having beneficial effects in *G. roeselii* [23]. Because of female excess
92 and positive effect on fitness-related traits, host populations infected with the feminizing *N.*
93 *granulosis* strains are predicted to have higher growth rate and may help the invasive host *G.*
94 *roeselii* in colonizing new territories [30]. Supporting this hypothesis, Quiles *et al.* [31] found that
95 the feminizing *N. granulosis* strain of *G. roeselii* is associated with the only host mitochondrial
96 genotype that invaded Western Europe after the last glaciation [32]. However, as feminizing
97 *Dictyocoela* microsporidians vary in their feminizing ability [21], it is likewise possible that the
98 feminization capacity of *N. granulosis* infecting *G. duebeni* and *G. roeselii* varies between parasite
99 strains and/or host species. The recent detection of *Nosema* spp. in populations of different
100 amphipod species with highly variable prevalence fuels this hypothesis [33,34], but more data are
101 necessary to understand *N. granulosis* diversity in amphipods. Studying microsporidian infections
102 in amphipods is facilitated by the fact that some hosts are subjected to intensive phylogenetic and
103 biogeographical research. For example, a study on New Zealand amphipods suggested that
104 amphipod-*Dictyocoela* associations could have preceded the split of Pangea some 180 million
105 years ago (MYA) and that these parasites co-differentiated with their hosts [35]. Similarly, the
106 resolution of the evolutionary histories of morphospecies *Gammarus balcanicus* and *G. roeselii*
107 [32,36,-38] allowed the study of co-differentiation between these hosts and some of their
108 microsporidian parasites, especially those belonging to the genus *Dictyocoela* [31,39,40].
109 However, data on *Nosema* infections remain too scarce in European gammarids to allow a general
110 analysis of their diversification. *Gammarus balcanicus* is a potential host of particular interest,
111 having begun its diversification in the early Miocene (c. 17 MYA) in the central Balkans, and for
112 which cryptic lineages are locally endemic due to paleo-habitat fragmentation and a complex
113 phylogeographic history [36,38]. Furthermore, the recent biogeographical study of the pan-
114 European morphospecies *Gammarus pulex* [41-43] and *Gammarus fossarum* [44] show high levels
115 of cryptic lineages diversity, with diversification starting some >10 MYA. The crown age of the
116 genus *Gammarus* itself has been estimated at 65 MYA [45]. All this offers the opportunity to

117 extend the analysis of *Nosema* diversity to a wider set of European gammarid species and to
118 decipher the age of infection in this group of hosts.

119 Besides the aforementioned advantages, the study of the phylogenetic diversity of microsporidian
120 parasites faces some problems with the resolution of available genetic markers. The small subunit
121 ribosomal DNA marker (SSU) is routinely used to PCR-screen hosts for microsporidian infections.
122 SSU sequencing data proved to be very useful for phylogenetic reconstructions at higher
123 taxonomic levels and for distinguishing major microsporidian clades (e.g. genera) [46], and
124 sometimes also for resolving phylogenetic relationships at lower taxonomic levels [40,47].
125 However, the use of SSU showed some unresolved phylogenetic relationships [e.g. *Dictyocoela*
126 *duebenum* / *muelleri* [47], and generated ambiguous results for the *Nosema* genus. On the one
127 hand, it enabled formal redefinition of the two genera *Nosema* and *Vairimorpha* within the family
128 Nosematidae [5]. On the other hand, phylogeography of *N. granulosis* in *G. roeselii* based on SSU
129 showed a low level of genetic variation despite a pan-European sampling [31]. Therefore, a more
130 variable marker is needed to unravel the phylogeography of *N. granulosis* and the evolution of
131 *Nosema*-amphipod associations. The genomes of the few economically relevant insect-infecting
132 *Nosema* spp. have been sequenced [48–50], as well as, recently, the genome of *N. granulosis* [51].
133 Based on these available genome data, the large subunit of the RNA polymerase II (RPB1) gene
134 has been shown to be a suitable marker to unravel microsporidian phylogenetic relationships in
135 cases where SSU was of limited value [5,52–54]. RPB1 is used for all mRNA synthesis in
136 eukaryotes [55], displaying a high level of synonymous variation. Its higher level of genetic
137 variation compared to SSU was already emphasized in a previous study on Microsporidia [52].

138 The aim of the present study was to deepen our understanding of the evolutionary history of
139 *Nosema* lineages infecting amphipod crustaceans, focusing mainly on *Gammarus* species, by using
140 both SSU and RPB1 markers. Specifically, we first aimed at elucidating the extent of molecular
141 variation associated with *Nosema* infections in amphipods (i.e., could all lineages identified as
142 *Nosema* based on SSU be considered as *Nosema granulosis sensu stricto*, or are they divergent

143 enough to be considered as different species?). Second, we wanted to infer host specificity and
144 explore evolutionary history scenarios that may explain the diversity of gammarid-infecting
145 *Nosema*. Finally, as new divergent lineages of *Nosema* were detected over the entire distribution
146 area of a new host, *Gammarus balcanicus*, the possible existence of sex ratio distortion associated
147 with vertical transmission was tested, using the crude proxy of an infection bias in females (as
148 already used in [20]). Such a test allowed us to explore whether sex ratio distortion is a trait that
149 appeared more than once in the evolution of amphipod *Nosema*.

150

151 **Materials and methods**

152 **Overview of dataset composition for *Nosema* infecting amphipods**

153 Our dataset combined information on 316 *Nosema*-infected amphipod individuals gathered from
154 three sources. The first source was data from the literature for which both SSU and RPB1
155 sequences were available. The second source was data for which SSU information is available
156 from the literature or GenBank and for which RPB1 information was tentatively gained from the
157 original DNA samples as part of this study. The third source was *de novo* sequencing data of
158 parasite SSU and RPB1 from individuals detected as infected by *Nosema* by SSU PCR screenings
159 (Table S1 and S2).

160 The first source is very limited, comprising of only two pairs of sequences, one associated with
161 infection of *G. duebeni* [56] and one associated with *G. roeselii* [51]. The second source included
162 samples of *G. fossarum*, *Niphargus schellenbergi* from Luxembourg and *Niphargellus arndti* from
163 Poland [34,57], *G. pulex* from Poland (Wroblewski unpublished), *G. pulex* and *G. fossarum* from
164 Germany [58], *G. duebeni* from Ireland [59] and *G. roeselii* from France [23]. A large number of
165 amphipod individuals from different countries were available for three taxa: *Dikerogammarus*
166 *villosus* [60], *G. roeselii* [31] and *G. balcanicus* [61] from 34, 94 and 88 sites, located in 8, 19 and
167 13 countries, with 1436, 1904 and 2225 host individuals tested, respectively. The third source was

168 associated with three host taxa we sampled during three local-scale surveys: *G. duebeni* (sites KER
169 and ROS in Brittany, France), *G. balcanicus* (site SK in Slovakia), and *G. pulex* (5 sites in France,
170 one in Germany and one in Poland). In addition, host individuals from two large-scale surveys we
171 conducted on *G. fossarum* and *G. balcanicus* were included (S1 Table). In *G. balcanicus* (unlike
172 other species), the sex of individuals was noted during dissection (S1 Table). This allowed us to
173 test if females are more often infected than males, which is an indication of sex-biased infection
174 by the parasites and suggestive of VT mode and sex ratio distortion, as previously shown in *G.*
175 *duebeni* and *G. roeselii* [20,23].

176

177 ***Nosema* infection status based on SSU ribosomal DNA sequencing**

178 All host individuals without previous information on microsporidian infection were first screened
179 to assess possible infection with *N. granulosis* based on the SSU marker. SSU was used for both
180 molecular screening for microsporidian infection based on specific PCR primers for Microsporidia
181 and assignment to *Nosema* based on BlastN searches [62] of sequenced PCR products against the
182 sequences available in GenBank. All molecular lab work conditions were as described in [31].

183

184 **Amplification and sequencing of *Nosema* RPB1**

185 All *Nosema*-infected individuals (based on SSU) were selected for RPB1 amplification (276
186 individuals). As all published primers tested [52,53,63,64] failed to produce reliable PCR products
187 of our samples, we designed new primers. Five RPB1 sequences were manually aligned, including
188 three *N. granulosis* sequences (associated with infection of *G. duebeni*: DQ996233 and EF119339
189 [52], and *G. roeselii*: SBJO01000442 [51]), and two sequences from *Nosema* infecting insects: *N.*
190 *bombycis* (DQ996231 [52]) and *Nosema empoascae* (DQ996232 [52]). Variable regions of interest
191 for phylogenetic reconstruction and conserved regions suitable for primer design were identified

192 from the alignment. Two sets of degenerated primers were designed using Geneious 10.2.2 [65],
193 targeting two non-overlapping fragments, named F2 and F4, starting at positions 386 and 1161 on
194 the RPB1 gene of the genome SBJO01000442, respectively (S1 Fig). The F2 primers were 5'-
195 GKT GTG GRA ATA AAC AGC-3' (forward, F2f) and 5'-TCT ACT CTC TTM CCC ATA AG-
196 3' (reverse, F2r), generating a 520 bp-long amplicon. The F4 primers were 5'-GAA AGA CAC
197 ATG CAG RAT G-3' (forward, F4f) and 5'-TTC CWG ACA TGA TYT CTC C-3' (reverse, F4r),
198 generating a 640 bp-long amplicon.

199 PCRs were performed in a volume of 30 μ l, containing 2.5 mM MgCl₂, 0.5 units of 5 PRIME
200 HotMaster Taq DNA polymerase (Qiagen, Hilden, Germany), 0.2 μ M dNTPs (MP Biomedicals
201 Europe, Illkirch, France), 0.2 μ M of forward and reverse primers each (Eurofins Genomics,
202 Ebersberg, Germany) and 2 ng DNA template. Amplification conditions were as follows: an initial
203 denaturing phase at 94°C for 2 min, 35 cycles at 94°C for 20s, annealing temperature was 50°C
204 for each fragment for 20 s and extension was at 65°C for 30 s, with a final extension step at 65°C
205 for 5 min. PCR products were purified and sequenced directly with the BigDye technology
206 (Genewiz, Leipzig, Germany) using the forward PCR primers. The chromatograms showed no
207 ambiguity in sequences and all the detected double peaks were of high quality. Nevertheless, seven
208 PCR products were bi-directionally sequenced to test for uncertain nucleotide position associated
209 with double peaks in chromatograms, and 17 products (from populations AL56, BF05, GR11,
210 H05, NIT, PA and PL8 (S1 Table), were sequenced twice with the forward primer to confirm
211 double peaks.

212 Using Geneious 10.2.2 [65], raw sequences were checked for microsporidian RPB1 specificity via
213 BlastN search [62] against sequences available in GenBank. Each sequence was manually edited,
214 and double peaks were called following the International Union of Pure and Applied Chemistry
215 (IUPAC) degenerate nucleotide code. The RPB1 gene is known to be a nuclear single-copy gene
216 in *Vairimorpha necatrix*, a close relative of *Nosema* [63], thus excluding paralogy as an
217 explanation for the observed double peaks. Instead, they could result from double infections with

218 *Nosema* individuals harbouring different haplogroups or from single infections by heterozygous
219 individuals. The F2 and F4 sequences were trimmed to the final size ranges of 282 to 455 bp and
220 280 to 562 bp, respectively, based on the quality level of the sequences. The RPB1 sequences were
221 translated into amino acids to confirm the absence of stop codons (S1A Fig). Sequences were
222 aligned after concatenation using MAFFT7.388 software [66,67] with the E-IIONS-I algorithm
223 using the legacy gap penalty option, incorporated in Geneious 10.2.2 [65]. To eliminate introns,
224 translated amino acid versions were also used as a backbone based on the reference the RPB1
225 sequence AF060234 from *Vairimorpha necatrix*.

226

227 **Haplogroup definition and phylogeny reconstruction for microsporidians**

228 As SSU and RPB1 sequence length was heterogeneous among individuals, definition of
229 haplotypes *sensu stricto* was not possible. Following the approach described in [31], sequences of
230 each marker were assigned to haplogroups as follows: two sequences were clustered in one
231 haplogroup based on 100% pairwise nucleotide identity in their shared part, each haplogroup being
232 defined by diagnostic SNPs [31]. Only few sequences could not be assigned to haplogroups, due
233 to reduced sequence length and lack of diagnostic features (S2 Table). Some RPB1 haplogroups
234 showed three short deletions in the F4 fragment (S1B Fig). Finally, the F2 or F4 fragments were
235 missing for some individuals, because amplification failed, and were coded as missing data (“N”).
236 The longest sequences of each haplogroup, were used for Bayesian phylogeny reconstruction (S2
237 Table). Haplogroup alignments used for building trees are provided in the Supplementary Data S1.
238 Compared to SSU, nucleotide variation for the RPB1 marker was high, although mostly
239 synonymous. Consequently, all codon positions were used for phylogenetic analyses.

240 For phylogenetic analyses, sequences of *Nosema* spp. infecting amphipods were complemented
241 by a set of 11 species of *Nosema/Vairimorpha* spp. infecting insects, according to the taxonomic
242 revision of [5], for which both SSU and RPB1 markers are available. We further included the two

243 following *Nosema* species (both infecting crayfish) in our analysis: *Nosema cheracis*, infecting the
244 Australian crayfish species *Cherax destructor* introduced in Europe and *Nosema austropotamobii*,
245 infecting *Austropotamobius pallipes*, a western European freshwater crayfish. *Ordospora*
246 *colligata* served as the outgroup [68]. All details are given in S2 Table.

247 Phylogenetic reconstructions were performed separately for SSU and RPB1 using Bayesian
248 inference implemented in MrBayes [69] integrated in Geneious 10.2.2. The best-fitting model of
249 nucleotide substitution was determined with JModelTest-2.1.10. [70]. For both SSU and RPB1 we
250 used the General Time Reversible (GTR) model with gamma-distributed rate heterogeneity (+G)
251 and a proportion of invariable sites (+I). For each marker, four heated chains, each 1,100,000
252 iterations long, sampled every 100 iterations, were run. Runs reached satisfactory effective
253 sampling sizes (ESS > 200). Fifty percent majority-rule consensus trees were constructed after
254 removal of 10% ‘burn-in’ trees. For the RPB1 phylogeny reconstruction, a maximum likelihood-
255 tree was also constructed in MEGA version 11 [71] using the GTR+G+I model and node support
256 was assessed with 1,000 bootstrap replicates.

257

258 **Molecular species delimitation**

259 To explore the number of molecular operational taxonomic units (MOTUs) that may represent
260 potential species within the *Nosema* clade, we applied two methods: Assemble Species by
261 Automatic Partitioning (ASAP) [72] and the Poisson tree processes (PTP) model [73]. ASAP was
262 conducted using the ASAP webserver (<https://bioinfo.mnhn.fr/abi/public/asap/> accessed on June
263 2, 2023) based on the distance matrix generated through IQ-TREE analysis [72]. ASAP divides
264 species partitions based on pairwise genetic distances, and that was why the dataset for this
265 analysis included only the sequences covering both fragments of RPB1 (F2 and F4). ASAP also
266 computes a probability of panmixia (p-val), a relative gap width metric (W), and ranks results by
267 the ASAP score: the lower the score, the better the partitioning [72].

268 PTP modelling was performed using the PTP web server (<https://species.h-its.org/>; accessed on
269 June 1, 2023) with the Bayesian implementation (bPTP), which adds Bayesian support (pp) values
270 for putative species to branches in the input tree. The PTP method infers speciation events based
271 on a shift in the number of substitutions between internal nodes [73]. Analysis was run with
272 500 000 iterations, the run was examined and showed convergence. The sites reported on Figs 1
273 and S2 were plotted on a map from Natural Earth resources in QGIS 3.32.0-Lima [74].

274

275 **Testing sex-biased infection pattern**

276 In *G. balcanicus*, 334 males and 310 females from 16 populations were tested for estimating the
277 prevalence of *Nosema* (after removing the two populations in which no females were tested, S1
278 Table). A General Linearized Model (binomial distribution using a Logit transition function) was
279 executed, analyzing the effect of sex and population as factors, and their interactions, on the
280 infection rate by *Nosema*. A bias of infection towards females should be associated with sex ratio
281 distortion [20,23].

282

283 **Results**

284 Based on SSU sequences, 316 individuals from nine amphipod species were found to be infected
285 by *Nosema* spp. (S1 and S2 Table). Our new census yielded 181 new *Nosema* sequences: 89 from
286 *G. balcanicus*, 39 from *G. duebeni*, 47 from *G. fossarum*, five from *G. pulex* (S1 Table). They
287 extend a set of 136 sequences previously determined as *Nosema* (99 from *G. roeselii*, 18 from *G.*
288 *pulex*, 7 from *G. duebeni*, 6 from *G. fossarum*, two from *D. villosus*, two from *Eulimnogammarus*
289 *verrucosus*, one from *N. arndti* and one from *N. schellenbergi*) (S1 and S2 Tables).

290 The phylogenetic reconstruction inferred from the SSU marker was to a large extent in line with
291 previously published phylogenetic reconstructions (e.g. [31]), except that the clade containing the

292 *Nosema* spp. found in amphipods also included a crayfish-infecting species (Fig 2). Most
293 haplogroups (18/24) were found specifically infecting a single host species, while six were found
294 in two to four host species.

295 Amplification success of the F2 and F4 fragments of RPB1 gene was variable, including no success
296 (113 ind., 41%), single fragment amplified (57 ind., 21%) or both fragments amplified (105 ind.,
297 38%). Therefore, we newly obtained RPB1 sequence information for 163 individuals (depending
298 on DNA availability and sequencing success), including 60 from *G. roeselii*, 45 from *G.*
299 *balcanicus*, 38 from *G. fossarum*, 15 from *G. duebeni* and 4 from *G. pulex*. To this set, two RPB1
300 sequences of VT, feminizing, *N. granulosis* infecting *G. duebeni* from Scotland, Wales and
301 Sweden [48,53] were added which matched the F2 and F4 regions of RPB1 (GenBank accession
302 numbers, DQ996233 and JX213747). The geographic distribution of the parasites for which the
303 RPB1 sequences were obtained is shown in Fig 1.

304 Alignment of RPB1 sequences from *Nosema* infecting amphipods revealed 39 haplogroups (Fig
305 3) showing a higher variability compared to the phylogenetic relationships among haplogroups
306 based on SSU sequences (Fig 3). The amphipod-infecting *Nosema* spp. were all within a well-
307 defined and supported clade, which was the sister to a clade of insect-infecting *Nosema* spp. (*sensu*
308 [5]). Based on the available sequences, the amphipod-*Nosema* clade also contained three
309 sequences from non-amphipod hosts (Fig 3). All but one sequence of amphipod-infecting *Nosema*
310 spp. could be assigned to five well-supported clades (named A to E, Fig 3). The single remaining
311 sequence is associated with a branch (termed F) of uncertain phylogenetic position (highlighted in
312 grey in Fig 3).

313 Parasites from clade A were found only in *G. balcanicus* (Fig 3), despite extensive sampling. These
314 *Nosema* strains, particularly F2A2-3F4A2-3 and F2A2F4A2-3, were found in numerous
315 populations, exhibiting high prevalence in some of the populations (see S1 Table), and their
316 geographic distribution covered the entire geographic range of their host (Fig 1). However, they

317 were not found in other *Gammarus* species from the close area, and were particularly absent in *G.*
318 *roeselii*, the other abundant host species overlapping *G. balcanicus* distribution (Fig 1). We found
319 no clear pattern for co-diversification between *G. balcanicus* and these *Nosema* parasites. We were
320 not able to run a formal co-diversification analysis, because the asymmetry of amplification of F2
321 and F4 fragments among some hosts (S2 Fig) prevented the construction of a complete matrix of
322 genetic distance among parasite haplogroups of this relatively small clade. We nevertheless
323 observed only little congruence of phylogenies between hosts and parasites (S2 Fig). We tested
324 whether, with our dataset, we could provide clues suggesting the presence of VT/sex ratio
325 distortion in *Nosema* clade A infecting *G. balcanicus*. A General Linearized Model analysis
326 showed that the interaction sex*population was non-significant and the corrected Akaike
327 Information Criterion (AICc) of the whole model was quite high (LR $\chi^2 = 18.03$, 15 df., $P = 0.26$,
328 AICc = 553.19), so the interaction was removed from the model. The non-significance of this
329 interaction means that globally there was no difference in the infection rate between sexes among
330 the different populations. The model including sex and population was globally significant (LR χ^2
331 = 124.59, 16 df., $P < 0.0001$, AICc = 538.60). The infection rate significantly differed among
332 populations, ranging from 4.2% (in populations AL57 and AL69) to 64.6% (in population UA03)
333 (S1 Table). In total, males were more infected than females (26.9% vs. 14.8%, respectively, LR χ
334 $^2 = 17.67$, 1 df., $P < 0.0001$). In most populations, there was no significant difference between
335 sexes (12 populations), but in four populations, males were significantly more infected than
336 females (populations AL56, H07, H09, H11).

337 Parasites from clade B were found in *G. pulex* and *G. roeselii* hosts. *Gammarus roeselii* parasites
338 were restricted to south-eastern Europe (Greece and Albania), while the single haplogroup from
339 *G. pulex* was recorded in France (Fig 1). Clades A and B were more distantly related to clades C,
340 D and E.

341 Clade C included the parasites infecting *G. duebeni*, i.e. the host in which *Nosema granulosis* was
342 first described as a vertically-transmitted, feminizing sex ratio distorter [10]. Interestingly, clade
343 C also included *Nosema* infecting *G. roeselii*, in which vertical transmission and feminization has
344 also been demonstrated [23,24]. Overall, clade C haplogroups were widespread all over Europe,
345 and it must be noticed that a single haplogroup (F2C3F4C3) was frequent (52 occurrence) and
346 detected in various populations of *G. roeselii* from six countries and in one French population of
347 *G. fossarum* (Fig 2,3, S2 Table). Clade D, which appears to be closely related to clade C, consisted
348 of highly similar haplogroups of *Nosema* that were solely found in *G. fossarum* (Fig 3, S3 Table).
349 Clade D was restricted to France and Austria, and these parasites infected nine individuals from
350 four populations. All haplogroups from clades C and D shared two short deletions in the F4
351 fragment (deletions of 27 and 15 bp) (S1 Fig) and they were not distinguishable using the SSU
352 marker (Fig 2). Haplogroups of clade C included a supplementary 15 bp-long deletion.

353 Finally, clade E consisted of parasites infecting three host species (*G. fossarum*, *G. pulex* and *G.*
354 *duebeni*) sampled at French and Swiss sites (Fig 1).

355 In summary, our sequences grouped within clades A, B, C, D, E, with C and D being sister groups,
356 and a single sequence identified as lineage F. The ASAP molecular delimitation analysis revealed
357 four putative species within *Nosema* infecting amphipod crustaceans (excluding lineage F), mostly
358 overlapping with the aforementioned clades, but C and D grouped as one species, here interpreted
359 as *Nosema granulosis sensu stricto* (Fig 3). On the other hand, bPTP overall suggested from 19 to
360 29 putative species, with partitioning to 25 species with the highest support (lowest posterior
361 probability 0.50) (S2 Appendix), and 13 species within *Nosema* found in amphipods (Fig 3).

362

363 Discussion

364 Our results indicate that various microsporidian lineages of the genus *Nosema*, not only *N.*
365 *granulosis*, are abundant, widespread and diverse in European gammarid amphipods. Thus, we

366 deepen our knowledge of host range for this parasitic genus beyond the already well-known insects
367 and terrestrial hosts (e.g. [75,76,77]). Indeed, our phylogenetic reconstructions unambiguously
368 grouped freshwater amphipod parasites as a sister group to Lepidoptera and Hymenoptera-
369 infecting *Nosema* (*sensu* [5]). However, we cannot formally conclude that amphipod *Nosema* are
370 monophyletic. The phylogenetic tree reconstruction based on the RPB1 marker suggests that
371 *Nosema* spp. found in freshwater crustaceans are not limited to amphipod hosts and include one
372 terrestrial species, *Nosema empoascae*. Furthermore, *Nosema* isolates infecting crayfish are nested
373 within those infecting amphipods. This could suggest either that there has been an ancient *Nosema*
374 infection in Malacostraca (the crustacean group to which amphipods and decapods belong), or a
375 freshwater origin of this *Nosema* clade, with horizontal transfers between host species. The
376 divergence among crustacean *Nosema* strains reported here could appear low compared to host
377 divergence (Malacostraca diversified around 300 MYA [78]). Crustaceans colonized freshwater
378 at diverse times independently [79]. If so, the most likely explanation would be horizontal transfer
379 of parasites. However, the molecular phylogenetic reconstructions of the hosts and the parasites
380 could not be compared directly since different genetic markers were used. For example, the New-
381 Zealand amphipod hosts and their microsporidian parasites (*Dictyocoela*) seemed to have initiated
382 their diversification simultaneously several MYA, but their genetic divergence appears to have
383 reached different levels [35]. The presence of the insect-infecting *N. empoascae* within the *Nosema*
384 lineage infecting crustaceans is more puzzling and may reflect horizontal transfer between
385 crustaceans and insects. Consistently, horizontal transfer of *Wolbachia* intracellular
386 microorganisms between insects and crustaceans has previously been suggested [80]. However,
387 owing to the strict terrestrial life cycle of the insect host of *N. empoascae* (*Empoasca fabae*, a
388 small leafhopper of the Cicadellidae family [81]) a recent host shift is ecologically difficult to
389 conceive in the present state of knowledge. In addition, this parasite seems to be vertically
390 transmitted in its host [82], which may further reduce the likelihood of horizontal transfer between
391 hosts. However, the existence of additional vector/intermediate hosts such as water insects or

392 terrestrial insects drowning in water cannot be excluded, as suggested for *Nosema cheracis* [11],
393 even if no data is available up to now. Besides, microsporidian samples presently available for
394 insect hosts are far from representative of the global diversity of *Nosema* infecting these hosts.
395 Thus, future studies are necessary to address the problem of the potential transfer of *Nosema*
396 between aquatic and terrestrial ecosystems.

397 The improved phylogenetic resolution of the RPB1 marker compared to SSU allowed us to
398 distinguish five *Nosema* clades infecting amphipods [31,51]. Clade A was found only associated
399 with *G. balcanicus*, whose biogeographical history is well documented for both host and other
400 microsporidian parasites [36,39,40]. In this case, host specificity may be proposed. Indeed, clade
401 A parasites infected individuals in several locations throughout the entire host range. In addition,
402 they colonized diverse host cryptic lineages (S2 Fig.) but they were not found in other *Gammarus*
403 spp., conversely to most other *Nosema* parasites revealed in this study. An infection by *Nosema*
404 clade A therefore possibly occurred anciently in *G. balcanicus* and subsequently spread further
405 during host radiation. However, unlike other microsporidian parasites [40], no clear co-
406 diversification pattern was found between *Nosema* clade A and *G. balcanicus* (S2 Fig). Therefore,
407 ancient infection followed by several horizontal transfers may explain the present distribution of
408 *Nosema* in *G. balcanicus*. Parasites belonging to clade D were restricted to a few lineages of the
409 *G. fossarum* species complex [44]. However, the range of the host *G. fossarum* is pan-European,
410 and since we did not have access to the whole geographic range for our survey, we cannot firmly
411 conclude that the parasites of clade D are specific to this host. Because of the low diversity within
412 this clade, it is likely that these infections were recently acquired by the host species.

413 Prior to this study, microsporidians from gammarids showing high score values by BlastNN with
414 *Nosema granulosis* isolated from *G. duebeni* based on SSU sequences were generally assigned
415 either to *N. granulosis* [31,34], to *Nosema* sp. or to *Vairimorpha* sp. [83]. Here we found that the
416 vertically-transmitted (VT) and feminizing *N. granulosis* described from *G. duebeni* [10,19–21]
417 correspond to clade C, also includes the closely related *Nosema* from *G. roeselii*, which likely is a

418 VT sex ratio distorting parasite [23,24]. These infections are characterized by a high number of
419 infected individuals in each host species, numerous infected populations, and few haplogroups in
420 each species (e.g., in *G. roeselii*, 77 individuals from 12 populations share the same RPB1 variant).
421 As noted in [31], all these traits are compatible with the dynamics of VT-feminizing
422 microsporidians. Their spread in host populations is enhanced by parasite-induced female-biased
423 sex ratios of host progenies, which provides a selective advantage for infected vs. uninfected
424 females [30]. Theoretical models suggest that female-biased sex ratios result in populations with
425 high numbers of individuals and high demographic dynamics that may outcompete non-infected
426 populations [30]. Thus, the presence of this VT *N. granulosis* infection might have facilitated the
427 recent spread of the host species, *G. roeselii*, in western Europe [32,37], by directly enhancing
428 host invasion success through increased rates of population growth, as suggested for another
429 amphipod *Crangonyx pseudogracilis* and its VT-parasite, *Fibrillanosema crangonycis* [84,85].
430 Because all *Nosema* belonging to clade C are genetically very similar to the originally described
431 *N. granulosis* (0.02 ± 0.003 average genetic divergence over all RPB1 sequence pairs), we propose
432 that all parasites clustering into clade C should be referred to as *N. granulosis*. Furthermore,
433 because i) clades C and D are in a close sister group relationship, ii) clade C contains sequences
434 linked to type material used for *N. granulosis* description [10], iii) the monophyly of clades C + D
435 is statistically supported, and iv) clade C + D is also supported by the ASAP analysis, we propose
436 that clades C and D should be seen as *Nosema granulosis sensu stricto*. By contrast, we suggest
437 that parasites grouped into clades A, B and E and lineage F may be candidates to the status of new
438 *Nosema* species, based on their genetic divergence supported by the species delimitation analysis.
439 Remarkably, this divergence is larger than the range of divergence observed for *Nosema* taxa from
440 insects. It may be explained by the differences in mutation rate and evolutionary dynamics
441 between *Nosema* spp. infecting gammarids and insects. In our interpretation of the number of
442 putative *Nosema* species infecting gammarids, we followed a more conservative approach: we
443 based our outcomes on the ASAP results assuming fewer putative species than the bPTP analysis.

444 There are many studies demonstrating that ASAP is often in congruence with
445 morphological/biological species delimitation while bPTP method tends to overestimate the
446 species number [e.g. 47,86,87]. However, more studies including ultrastructural analysis are
447 needed to formally describe those clades as putative species. We also cannot exclude that our new
448 lineages can be *Nosema* spp. already described from different amphipods, including *G. pulex*,
449 based solely on morphology (light microscopy for older observations) [13–17,88]. However,
450 morphological descriptions of *Nosema* can be misleading [5], and using an integrative approach is
451 advisable and ultimately could lead to taxonomic revisions [89].

452 Interestingly, most haplogroups involving *Nosema* from clades B, D and E are present in a single
453 or few individuals within the same population. Because of the sharp contrast of this infection
454 pattern compared to clade C, it does not seem reasonable to us to propose that parasites of clades
455 B, D, and E are feminizing and VT microsporidia. Contrastingly, infections in *G. balcanicus* hosts
456 belonging to clade A show an infection pattern similar to those of clade C. For example, sequences
457 in populations from Slovakia, Hungary and Romania show closely-related or similar RPB1
458 haplogroups infecting numerous individuals (> 70) in a dozen of populations all over the host
459 geographic range. This pattern therefore resembles the VT-feminizing *N. granulosis* infection
460 patterns in *G. roeselii* (Fig 2, 3). However, the analysis of infection patterns relative to sexes
461 showed that there were no significant differences in prevalence between sexes in most populations.
462 Yet, parasites that have been shown to be sex ratio distorters - such as *N. granulosis* in *G. duebeni*
463 and *G. roeselii* or *Wolbachia pipientis* in the isopod *Armadillidium vulgare* - are always more
464 frequent in females than in males (which makes sense given that the parasites convert infected
465 males into females) [20,23,90]. The high prevalence and above all the homogeneity of *Nosema*
466 haplogroups across the geographic range of *G. balcanicus* is therefore puzzling. Based on this
467 survey and previous evidence, VT and feminization in *Nosema* infecting amphipods therefore
468 currently appear to be limited to *N. granulosis* belonging to clade C. It is noteworthy that clade C

469 also contains parasites infecting *G. fossarum* individuals, which opens the possibility that VT-
470 feminization might be found in a larger spectrum of amphipod species than previously thought.

471 In conclusion, our study illustrates the interest of using the RPB1 marker for the study of *Nosema*
472 parasites, allowing elucidating the parasitic richness in their gammarid hosts. We found that
473 *Nosema* diversity is much higher in amphipods than previously thought, and that the association
474 appears to be ancient in this host group. Furthermore, the new host species found to be infected by
475 *N. granulosis sensu stricto* opens the possibility that VT-feminization may be present in a larger
476 spectrum of amphipod species. Finally, given that gammarids are keystone species in the
477 functioning of freshwater environments, with their functions being affected by microsporidian
478 infections [91-94–], it is of crucial importance to correctly identify host-parasite relationships by
479 means of reliable markers.

480

481 **Acknowledgments**

482 We acknowledge the French National Research Agency (ANR) (Grant # ANR-15-CE32-0006) for
483 paying AQ salary, for contribution in designing the study, molecular analyses, statistical analyses,
484 writing the manuscript; as well as the ‘Polonium’ exchange program (PPN/BIL/2018/1/00138),
485 for funding travel between France and Poland in terms of analysing the results and writing the
486 manuscript by the authors. The sampling was performed during a series of scientific expeditions
487 in the years 2004–2021 supported by internal funds of the University of Lodz and partially by the
488 Polish Ministry of Science and Higher Education, grant number N N303 579439, as well as the
489 French National Research Agency (ANR) "Recolnat ANR-11-INBS-0004", Bourgogne-Franche-
490 Comté Region Recherche d'Excellence BFC "PIA-Recolnat", CNRS-INSU OSU-THETA
491 Franche-Comté Bourgogne (SRO 2022). We would like to thank Tomasz Mamos for sharing DNA
492 extracts from *G. balcanicus* used in Mamos et al. (2016). We also thank Tomasz Rewicz and Alicja
493 Konopacka for assistance with sampling.

494 We thank Dr. Joseph Ironside (Aberystwyth University) and two anonymous reviewers for their
495 valuable comments on the previous version of the manuscript, that significantly improved the
496 paper.

497

498 **References**

- 499 1. Franzen C. Microsporidia: A review of 150 years of research. *Open Parasitol J.* 2008;2: 1–
500 34. doi:10.2174/1874421400802010001
- 501 2. Vávra J, Lukeš J. Microsporidia and ‘The art of living together.’ *Adv Parasitol.* 2013. pp.
502 253–319. doi:10.1016/B978-0-12-407706-5.00004-6
- 503 3. Mathis A. Microsporidia: emerging advances in understanding the basic biology of these
504 unique organisms. *Int J Parasitol.* 2000;30: 795–804. doi:10.1016/S0020-7519(00)00064-3
- 505 4. Naegeli K. Über die neue Krankheit der Seidenraupe und verwandte Organismen. *Botanische*
506 *Zeitung.* 1857;15: 760–761.
- 507 5. Tokarev YS, Huang W-F, Solter LF, Malysh JM, Becnel JJ, Vossbrinck CR. A formal
508 redefinition of the genera *Nosema* and *Vairimorpha* (Microsporidia: Nosematidae) and
509 reassignment of species based on molecular phylogenetics. *J Invertebr Pathol.* 2020;169:
510 107279. doi:10.1016/j.jip.2019.107279
- 511 6. Lockwood JA, Bomar CR, Ewen AB. The history of biological control with *Nosema*
512 *locustae*: Lessons for locust management. *Int J Trop Insect Sci.* 1999;19: 333–350.
513 doi:10.1017/S1742758400018968
- 514 7. Lewis LC, Bruck DJ, Prasifka JR, Raun ES. *Nosema pyrausta*: Its biology, history, and
515 potential role in a landscape of transgenic insecticidal crops. *Biol Control.* 2009;48: 223–
516 231. doi:10.1016/j.biocontrol.2008.10.009
- 517 8. Higes M, Martín-Hernández R, García-Palencia P, Marín P, Meana A. Horizontal
518 transmission of *Nosema ceranae* (Microsporidia) from worker honeybees to queens (*Apis*
519 *mellifera*). *Environ Microbiol Rep.* 2009;1: 495–498. doi:10.1111/j.1758-
520 2229.2009.00052.x
- 521 9. Han M-S, Watanabe H. Transovarial transmission of two microsporidia in the silkworm,
522 *Bombyx mori*, and disease occurrence in the progeny population. *J Invertebr Pathol.* 1988;51:
523 41–45. doi:10.1016/0022-2011(88)90086-9
- 524 10. Terry RS, Smith JE, Bouchon D, Rigaud T, Duncanson P, Sharpe RG, et al. Ultrastructural
525 characterisation and molecular taxonomic identification of *Nosema granulosis* n. sp., a
526 transovarially transmitted feminising (TTF) microsporidium. *J Eukaryot Microbiol.* 1999;46:
527 492–499. doi:10.1111/j.1550-7408.1999.tb06066.x
- 528 11. Moodie EG, Le Jambre LF, Katz ME. Ultrastructural characteristics and small subunit
529 ribosomal DNA sequence of *Vairimorpha cheracis* sp. nov., (Microspora: Burenellidae), a

- 530 parasite of the Australian yabby, *Cherax destructor* (Decapoda: Parastacidae). J Invertebr
531 Pathol. 2003;84: 198–213. doi:10.1016/j.jip.2003.11.004
- 532 12. Pretto T, Montesi F, Ghia D, Berton V, Abbadi M, Gastaldelli M, et al. Ultrastructural and
533 molecular characterization of *Vairimorpha austropotamobii* sp. nov. (Microsporidia:
534 Burenellidae) and *Thelohania contejeani* (Microsporidia: Thelohaniidae), two parasites of
535 the white-clawed crayfish, *Austropotamobius pallipes* complex (Decapoda: Astacidae). J
536 Invertebr Pathol. 2018;151: 59–75. doi:10.1016/j.jip.2017.11.002
- 537 13. Van Ryckeghem J. Les Cnidosporidies et autres parasites du *Gammarus pulex*. Cellule.
538 1930;39: 400–416.
- 539 14. Larsson R. On two microsporidia of the amphipod *Rivulogammarus pulex* light
540 microscopical and ultrastructural observations on *Thelohania muelleri* (Pfeiffer, 1895) and
541 *Nosema rivulogammari* n. sp. (Microspora, Thelohaniidae and Nosematidae). Zoologischer
542 Anzeiger. 1983;211: 299–323.
- 543 15. Ovcharenko M, Codreanu-Balcescu D, Grabowski M, Konopacka A, Wita I, Czaplinska U.
544 Gregarines (Apicomplexa) and microsporidians (Microsporidia) of native and invasive
545 gammarids (Amphipoda, Gammaroidea), occurring in Poland. Wiadomości
546 Parazytologiczne. 2009;55: 237–247.
- 547 16. Ovcharenko M, Wroblewski PP. An intracellular prokaryotic microorganism associated with
548 microsporidiosis of invasive amphipods *Pontogammarus robustoides*. Baltic Coastal Zone
549 Journal of Ecology and Protection of the Coastline. 2016; 2016–2020.
- 550 17. Lipa JJ. *Nosema kozhovi* sp. n., a new microsporidian parasite of *Brandtia lata lata*
551 (Crustacea, Gammaridae) of Baikal Lake. Acta Protozool. 1967;5: 93–96.
- 552 18. Bojko J, Ovcharenko M. Pathogens and other symbionts of the Amphipoda: taxonomic
553 diversity and pathological significance. Dis Aquat Organ. 2019;136: 3–36.
554 doi:10.3354/dao03321
- 555 19. Dunn AM, Smith JE. Microsporidian life cycles and diversity: The relationship between
556 virulence and transmission. Microbes Infect. 2001;3: 381–388. doi:10.1016/S1286-
557 4579(01)01394-6
- 558 20. Terry RS, Smith JE, Sharpe RG, Rigaud T, Littlewood DTJ, Ironside JE, et al. Widespread
559 vertical transmission and associated host sex-ratio distortion within the eukaryotic phylum
560 Microspora. Proc R Soc Lond B Biol Sci. 2004;271: 1783–1789. doi:10.1098/rspb.2004.2793
- 561 21. Ironside JE, Alexander J. Microsporidian parasites feminise hosts without paramyxean co-
562 infection: Support for convergent evolution of parasitic feminisation. Int J Parasitol. 2015;45:
563 427–433. doi:10.1016/j.ijpara.2015.02.001
- 564 22. Dunn A, Terry R, Smith J. Transovarial transmission in the microsporidia. Adv Parasitol.
565 Elsevier; 2001. pp. 57–100. doi:10.1016/S0065-308X(01)48005-5
- 566 23. Haine ER, Brondani E, Hume KD, Perrot-Minnot MJ, Gaillard M, Rigaud T. Coexistence of
567 three microsporidia parasites in populations of the freshwater amphipod *Gammarus roeseli*:
568 Evidence for vertical transmission and positive effect on reproduction. Int J Parasitol.
569 2004;34: 1137–1146. doi:10.1016/j.ijpara.2004.06.006

- 570 24. Haine ER, Motreuil S, Rigaud T. Infection by a vertically-transmitted microsporidian parasite
571 is associated with a female-biased sex ratio and survival advantage in the amphipod
572 *Gammarus roeseli*. *Parasitology*. 2007;134: 1363–1367. doi:10.1017/S0031182007002715
- 573 25. Jahnke M, Smith JE, Dubuffet A, Dunn AM. Effects of feminizing microsporidia on the
574 masculinizing function of the androgenic gland in *Gammarus duebeni*. *J Invertebr Pathol*.
575 2013;112: 146–151. doi:10.1016/j.jip.2012.11.008
- 576 26. Rodgers-Gray TP, Smith JE, Ashcroft AE, Isaac RE, Dunn AM. Mechanisms of parasite-
577 induced sex reversal in *Gammarus duebeni*. *Int J Parasitol*. 2004;34: 747–753.
578 doi:10.1016/j.ijpara.2004.01.005
- 579 27. Ironside JE, Dunn AM, Rollinson D, Smith JE. Association with host mitochondrial
580 haplotypes suggests that feminizing microsporidia lack horizontal transmission. *J Evol Biol*.
581 2003;16: 1077–1083. doi:10.1046/j.1420-9101.2003.00625.x
- 582 28. Kelly A, Dunn AM, Hatcher MJ. Incomplete feminisation by the microsporidian sex ratio
583 distorter, *Nosema granulosis*, and reduced transmission and feminisation efficiency at low
584 temperatures. *Int J Parasitol*. 2002;32: 825–831. doi:10.1016/S0020-7519(02)00019-X
- 585 29. Terry RS, Dunn AM, Smith JE. Cellular distribution of a feminizing microsporidian parasite:
586 a strategy for transovarial transmission. *Parasitology*. 1997;115: 157–163.
587 doi:10.1017/S0031182097001236
- 588 30. Rood ES, Freedberg S. Intragenomic conflict produces sex ratio dynamics that favor maternal
589 sex ratio distorters. *Ecol Evol*. 2016;6: 8085–8093. doi:10.1002/ece3.2498
- 590 31. Quiles A, Bacela-Spychalska K, Teixeira M, Lambin N, Grabowski M, Rigaud T, et al.
591 Microsporidian infections in the species complex *Gammarus roeselii* (Amphipoda) over its
592 geographical range: evidence for both host–parasite co-diversification and recent host shifts.
593 *Parasit Vectors*. 2019;12: 327. doi:10.1186/s13071-019-3571-z
- 594 32. Csapó H, Krzywoźniak P, Grabowski M, Wattier R, Bacela-Spychalska K, Mamos T, et al.
595 Successful post-glacial colonization of Europe by single lineage of freshwater amphipod
596 from its Pannonian Plio-Pleistocene diversification hotspot. *Sci Rep*. 2020;10: 18695.
597 doi:10.1038/s41598-020-75568-7
- 598 33. Grabner DS, Schertzinger G, Sures B. Effect of multiple microsporidian infections and
599 temperature stress on the heat shock protein 70 (hsp70) response of the amphipod *Gammarus*
600 *pulex*. *Parasit Vectors*. 2014;7: 170. doi:10.1186/1756-3305-7-170
- 601 34. Weigand AM, Kremers J, Grabner DS. Shared microsporidian profiles between an obligate
602 (*Niphargus*) and facultative subterranean amphipod population (*Gammarus*) at sympatry
603 provide indications for underground transmission pathways. *Limnologica*. 2016;58: 7–10.
604 doi:10.1016/j.limno.2016.01.005
- 605 35. Park E, Jorge F, Poulin R. Shared geographic histories and dispersal contribute to congruent
606 phylogenies between amphipods and their microsporidian parasites at regional and global
607 scales. *Mol Ecol*. 2020;29: 3330–3345. doi:10.1111/mec.15562
- 608 36. Mamos T, Wattier R, Burzyński A, Grabowski M. The legacy of a vanished sea: a high level
609 of diversification within a European freshwater amphipod species complex driven by 15 My
610 of Paratethys regression. *Mol Ecol*. 2016;25: 795–810. doi:10.1111/mec.13499

- 611 37. Grabowski M, Mamos T, Baćela-Spychalska K, Rewicz T, Wattier RA. Neogene
612 paleogeography provides context for understanding the origin and spatial distribution of
613 cryptic diversity in a widespread Balkan freshwater amphipod. PeerJ. 2017;5: e3016.
614 doi:10.7717/peerj.3016
- 615 38. Copilaş- Ciocianu D, Petrussek A. Phylogeography of a freshwater crustacean species
616 complex reflects a long-gone archipelago. J Biogeogr. 2017;44: 421–432.
617 doi:10.1111/jbi.12853
- 618 39. Quiles A, Rigaud T, Wattier RA, Grabowski M, Bacela-Spychalska K. Wide geographic
619 distribution of overlooked parasites: Rare microsporidia in *Gammarus balcanicus*, a species
620 complex with a high rate of endemism. Int J Parasitol: Parasites Wildl. 2021;14: 121–129.
621 doi:10.1016/j.ijppaw.2021.01.004
- 622 40. Quiles A, Wattier RA, Bacela-Spychalska K, Grabowski M, Rigaud T. *Dictyocoela*
623 microsporidia diversity and co-diversification with their host, a gammarid species complex
624 (Crustacea, Amphipoda) with an old history of divergence and high endemic diversity. BMC
625 Evol Biol. 2020;20: 149. doi:10.1186/s12862-020-01719-z
- 626 41. Hou Z, Sket B, Shuqiang L. Eocene habitat shift from saline to freshwater promoted Tethyan
627 amphipod diversification. PNAS. 2011; 108(135): 14533-14538.
628 Doi.org/10.1073/pnas.1104636108
- 629 42. Hupalo K, Mamos T, Wrzesinska W, Grabowski M. First endemic freshwater *Gammarus* from
630 Crete and its evolutionary history—an integrative taxonomy approach. PeerJ:6: e4457.
631 doi.org/10.7717/peerj.4457
- 632 43. Hupalo K, Karaouzaz I, Mamos T, Grabowski M. Molecular data suggest multiple origins and
633 diversification times of freshwater gammarids on the Aegean archipelago. Sci Rep. 2020: 10:
634 19813. doi.org/10.1038/s41598-020-75802-2
- 635 44. Wattier R, Mamos T, Copilaş-Ciocianu D, Jelić M, Ollivier A, Chaumot A, et al. Continental-
636 scale patterns of hyper-cryptic diversity within the freshwater model taxon *Gammarus*
637 *fossarum* (Crustacea, Amphipoda). Sci Rep. 2020;10: 16536. doi:10.1038/s41598-020-
638 73739-0
- 639 45. Hou Z, Sket B, Fiser C, Li S. Eocene habitat shift from saline to freshwater promoted Tethyan
640 amphipod diversification. PNAS. 2011;108: 14533–14538. doi:10.1073/pnas.1104636108
- 641 46. Vossbrinck CR, Debrunner-Vossbrinck BA, Weiss LM. Phylogeny of the Microsporidia.
642 Microsporidia: Pathogens of Opportunity: First Edition. Chichester, UK: John Wiley & Sons,
643 Inc.; 2014. pp. 203–220. doi:10.1002/9781118395264.ch6
- 644 47. Bacela-Spychalska K, Wróblewski P, Mamos T, Grabowski M, Rigaud T, Wattier R, et al.
645 Europe-wide reassessment of *Dictyocoela* (Microsporidia) infecting native and invasive
646 amphipods (Crustacea): Molecular versus ultrastructural traits. Sci Rep. 2018;8: 1–16.
647 doi:10.1038/s41598-018-26879-3
- 648 48. Xu J, Pan G, Fang L, Li J, Tian X, Li T, et al. The varying microsporidian genome: Existence
649 of long-terminal repeat retrotransposon in domesticated silkworm parasite *Nosema bombycis*.
650 Int J Parasitol. 2006;36: 1049–1056. doi:10.1016/j.ijpara.2006.04.010

- 651 49. Cornman RS, Chen YP, Schatz MC, Street C, Zhao Y, Desany B, et al. Genomic analyses of
652 the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. PLoS Pathog.
653 2009;5. doi:10.1371/journal.ppat.1000466
- 654 50. Chen Y ping, Pettis JS, Zhao Y, Liu X, Tallon LJ, Sadzewicz LD, et al. Genome sequencing
655 and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights
656 into host-parasite interactions. BMC Genomics. 2013;14: 451. doi:10.1186/1471-2164-14-
657 451
- 658 51. Cormier A, Chebbi MA, Giraud I, Wattier R, Teixeira M, Gilbert C, et al. Comparative
659 genomics of strictly vertically transmitted, feminizing Microsporidia endosymbionts of
660 amphipod crustaceans. Genome Biol Evol. 2021;13: evaa245. doi:10.1093/gbe/evaa245
- 661 52. Ironside JE. Multiple losses of sex within a single genus of Microsporidia. BMC Evol Biol.
662 2007;7: 48. doi:10.1186/1471-2148-7-48
- 663 53. Cheney SA, Lafranchi-Tristem NJ, Bourges D, Canning EU. Relationships of microsporidian
664 genera, with emphasis on the polysporous genera, revealed by sequences of the largest
665 subunit of RNA polymerase II (RPB1). J Eukaryot Microbiol. 2001;48: 111–7.
666 doi:10.1111/j.1550-7408.2001.tb00422.x
- 667 54. Tokarev YS, Timofeev SA, Malysh JM, Tsarev AA, Ignatieva AN, Tomilova OG, et al.
668 Hexokinase as a versatile molecular genetic marker for Microsporidia. Parasitology. 2019.
669 doi:10.1017/S0031182018001737
- 670 55. Cramer P, Bushnell DA, Kornberg RD. Structural basis of transcription: RNA polymerase II
671 at 2.8 ångstrom resolution. Science. 2001;292: 1863–1876. doi:10.1126/science.1059493
- 672 56. Ironside JE. Diversity and recombination of dispersed ribosomal DNA and protein coding
673 genes in Microsporidia. PLoS ONE. 2013;8: e55878. doi:10.1371/journal.pone.0055878
- 674 57. Grabner D, Weber D, Weigand AM. Updates to the sporadic knowledge on microsporidian
675 infections in groundwater amphipods (Crustacea, Amphipoda, Niphargidae). Subterr Biol.
676 2020;33: 71–85. doi:10.3897/subtbiol.33.48633
- 677 58. Prati S, Grabner DS, Pfeifer SM, Lorenz AW, Sures B. Generalist parasites persist in
678 degraded environments: a lesson learned from microsporidian diversity in amphipods.
679 Parasitology. 2022;149: 973–982. doi:10.1017/S0031182022000452
- 680 59. Krebs L, Blank M, Frankowski J, Bastrop R. Molecular characterisation of the
681 Microsporidia of the amphipod *Gammarus duebeni* across its natural range revealed hidden
682 diversity, wide-ranging prevalence and potential for co-evolution. Infect Genet Evol.
683 2010;10: 1027–1038. doi:10.1016/j.meegid.2010.06.011
- 684 60. Wattier RA, Haine ER, Beguet J, Martin G, Bollache L, Muskó IB, et al. No genetic
685 bottleneck or associated microparasite loss in invasive populations of a freshwater amphipod.
686 Oikos. 2007;116: 1941–1953. doi:10.1111/j.0030-1299.2007.15921.x
- 687 61. Quiles A. Evolutionary histories of symbioses between microsporidia and their amphipod
688 hosts: contribution of studying two hosts over their geographic ranges. PhD Thesis,
689 Bourgogne Franche-Comté. 2019.

- 690 62. Madden T. The BLAST Sequence Analysis Tool. The NCBI Handbook. National Center for
691 Biotechnology Information (US); 2003. pp. 1–15. doi:10.1086/597785
- 692 63. Hirt RP, Logsdon JM, Healy B, Dorey MW, Doolittle WF, Embley TM. Microsporidia are
693 related to Fungi: Evidence from the largest subunit of RNA polymerase II and other proteins.
694 PNAS. 1999;96: 580–585. doi:10.1073/pnas.96.2.580
- 695 64. Stiller JW, Hall BD. The origin of red algae: Implications for plastid evolution. *Evolution*.
696 1997;94: 4520–4525. doi:10.1073/pnas.94.9.4520
- 697 65. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic:
698 An integrated and extendable desktop software platform for the organization and analysis of
699 sequence data. *Bioinformatics*. 2012;28: 1647–1649. doi:10.1093/bioinformatics/bts199
- 700 66. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast
701 Fourier transform. *Nucleic Acids Res*. 2002;30: 3059–3066. doi:10.1093/nar/gkf436
- 702 67. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
703 Improvements in performance and usability. *Mol Biol Evol*. 2013;30: 772–780.
704 doi:10.1093/molbev/mst010
- 705 68. Refardt D, Canning EU, Mathis A, Cheney SA, Lafranchi-Tristem NJ, Ebert D. Small subunit
706 ribosomal DNA phylogeny of microsporidia that infect *Daphnia* (Crustacea: Cladocera).
707 *Parasitology*. 2002;124: 381–389. doi:10.1017/S0031182001001305
- 708 69. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees.
709 *Bioinformatics*. 2001;17: 754–755. doi:10.1093/bioinformatics/17.8.754
- 710 70. Darriba D, Taboada GL, Doallo R, Posada D. JModelTest 2: More models, new heuristics
711 and parallel computing. *Nat Methods*. 2012;9: 772. doi:10.1038/nmeth.2109
- 712 71. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version
713 11. *Molecul Biol Evol*. 2021;38: 3022-3027. doi:10.1093/molbev/msab120
- 714 72. Puillandre N, Brouillet S, Achaz G. ASAP: assemble species by automatic partitioning. *Mol*
715 *Ecol Res*. 2021, 21: 609–620. doi: 10.1111/1755-0998.13281
- 716 73. Zhang J, Kapli P, Pavlidis P, Stamatakis A. A General Species Delimitation Method with
717 Applications to Phylogenetic Placements. *Bioinformatics*, 2013. 29(22): 2869-2876. doi:
718 10.1093/bioinformatics/btt499
- 719 74. QGIS Development Team, 2023. QGIS Geographic Information System. Open Source
720 Geospatial Foundation Project. <http://qgis.osgeo.org>
- 721 75. Park E, Poulin R. Revisiting the phylogeny of microsporidia. *Int J Parasitol*. 2021;51: 855–
722 864. doi:10.1016/j.ijpara.2021.02.005
- 723 76. Wadi L, Reinke AW. Evolution of microsporidia: An extremely successful group of eukaryotic
724 intracellular parasites. *PLoS Pathog*, 2020. 16(2): e1008276. doi-
725 org.inee.bib.cnrs.fr/10.1371/journal.ppat.1008276
- 726 77. Bojko J, Reinke AW, Stentiford GD, Williams B, Rogers MSJ, Bass D. Microsporidia: a new
727 taxonomic, evolutionary, and ecological synthesis. *Trends Parasitol*. 2022;38: 642–659.
728 doi:10.1016/j.pt.2022.05.007

- 729 78. Oakley TH, Wolfe JM, Lindgren AR, Zaharoff AK. Phylotranscriptomics to bring the
730 understudied into the fold: Monophyletic Ostracoda, fossil placement, and pancrustacean
731 phylogeny. *Mol Biol Evol.* 2013;30: 215–233. doi:10.1093/molbev/mss216
- 732 79. Ahyong ST, Huang C. Colonization, adaptation, radiation, and diversity in fresh water. Poore
733 G. C. B. & Thiel M. *The Natural History of the Crustacea: Evolution and Biogeography of*
734 *the Crustacea.* Poore G. C. B. & Thiel M. Oxford University Press; 2020. pp. 303–330.
- 735 80. Cordaux R, Michel-Salzat A, Bouchon D. *Wolbachia* infection in crustaceans: novel hosts
736 and potential routes for horizontal transmission. *J Evol Biol.* 2001;14: 237–243.
737 doi:10.1046/j.1420-9101.2001.00279.x
- 738 81. Ni X, Backus EA, Maddox JV. A New microsporidium, *Nosema empoascae* n. sp., from
739 *Empoasca fabae* (Harris) (Homoptera: Auchenorrhyncha: Cicadellidae). *J Invertebr Pathol.*
740 1995;66: 52–59. doi:10.1006/jipa.1995.1060
- 741 82. Ni X, Backus EA, Maddox JV. Transmission Mechanisms of *Nosema empoascae*
742 (Microspora: Nosematidae) in *Empoasca fabae* (Homoptera: Cicadellidae). *J Invertebr*
743 *Pathol.* 1997;69: 269–275. doi:10.1006/jipa.1996.4646
- 744 83. Madyarova EV, Adelshin RV, Dimova MD, Axenov-Gribanov DV, Lubyaga YA,
745 Timofeyev MA. Microsporidian parasites found in the hemolymph of four Baikalian endemic
746 amphipods. *PLoS ONE.* 2015;10: e0130311. doi:10.1371/journal.pone.0130311
- 747 84. Galbreath Slothouber JGM, Terry RS, Becnel JJ, Dunn AM. Invasion success of
748 *Fibrillanosema crangonycis*, n.sp., n.g.: a novel vertically transmitted microsporidian
749 parasite from the invasive amphipod host *Crangonyx pseudogracilis*. *Int J Parasitol.* 2004;34:
750 235–244. doi:10.1016/j.ijpara.2003.10.009
- 751 85. Slothouber Galbreath JGM, Smith JE, Becnel JJ, Butlin RK, Dunn AM. Reduction in post-
752 invasion genetic diversity in *Crangonyx pseudogracilis* (Amphipoda: Crustacea): a genetic
753 bottleneck or the work of hitchhiking vertically transmitted microparasites? *Biol Invasions.*
754 2010;12: 191–209. doi:10.1007/s10530-009-9442-3
- 755 86. Mamos T, Jazdzewski K, Ciamporova-Zatovicova Z, Ciampor F Jr, Grabowski M. Fuzzy
756 species borders of glacial survivalists in the Carpathian biodiversity hotspot revealed using a
757 multimarker approach. *Sci Rep* 2021;11: 21629. doi.org/10.1038/s41598-021-00320-8
- 758 87. Guo B, Kong L. Comparing the efficiency of single-locus species delimitation methods within
759 Trochoidea (Gastropoda: Vetigastropoda). *Genes.* 2022;13(12): 2273.
760 doi.org/10.3390/genes13122273
- 761 88. Desportes I, Ginsburgervogel T, Zerbib C. Demonstration of 2 Microsporidia in amphipod
762 crustacean, *Orchestia-Gammarellus* (pallas) - *Nosema* and *Thelohania*. *J Protozool.* 1976;23:
763 A8–A9.
- 764 89. Ovcharenko MO, Bacela K, Wilkinson T, Ironside JE, Rigaud T, Wattier RA. *Cucumispora*
765 *dikerogammari* n. gen. (Fungi: Microsporidia) infecting the invasive amphipod
766 *Dikerogammarus villosus* : a potential emerging disease in European rivers. *Parasitology.*
767 2010;137: 191–204. doi:10.1017/S0031182009991119

- 768 90. Durand S, Lheraud B, Giraud I, Bech N, Grandjean F, Rigaud T, et al. Heterogeneous
769 distribution of sex ratio distorters in natural populations of the isopod *Armadillidium vulgare*.
770 Biol Lett. 2023;19: 20220457. doi:10.1098/rsbl.2022.0457
- 771 91. Gerhardt A. GamTox: A low-cost multimetric ecotoxicity test with *Gammarus* spp. for *In*
772 and *Ex Situ* application. Int J Zool. 2011;2011: 1–7. doi:10.1155/2011/574536
- 773 92. Bunke M, Alexander ME, Dick JTA, Hatcher MJ, Paterson R, Dunn AM. Eaten alive:
774 cannibalism is enhanced by parasites. R Soc Open Sci. 2015;2: 140369.
775 doi:10.1098/rsos.140369
- 776 93. Gismondi E, Rigaud T, Beisel J-N, Cossu-Leguille C. Effect of multiple parasitic infections
777 on the tolerance to pollutant contamination. PLoS ONE. 2012;7: e41950.
778 doi:10.1371/journal.pone.0041950
- 779 94. Gismondi E, Rigaud T, Beisel J-N, Cossu-Leguille C. Microsporidia parasites disrupt the
780 responses to cadmium exposure in a gammarid. Environ Pollut. 2012;160: 17–23.
781 doi:10.1016/j.envpol.2011.09.021
- 782
- 783

784 **Figure captions:**

785 **Figure 1.** Geographic distribution of *Nosema* spp. infecting a range of amphipod hosts based on
786 data listed in S1 and S2 Table. Assignment to clades or lineages is based on RPB1 (see Fig 3). The
787 infection of *Nosema* sp. in *Eulimnogammarus verrucosus* from the Lake Baikal, Russia is not
788 shown. The sites were plotted on a map from Natural Earth resources in QGIS 3.32.0-Lima [74].

789
790 **Figure 2.** Bayesian phylogenetic reconstruction of *Nosema* spp. based on small ribosomal subunit
791 (SSU) rDNA, and their association to different host species. *Ordospora colligata* (AF394529) was
792 used as an outgroup (not shown). Labels include: Genbank accession number (for previously
793 published sequences), microsporidian species name as given in [5] and for the newly produced
794 sequences the haplogroup name is given as in S1 and S2 Table. An X denotes an association and
795 a figure denotes the number of individuals infected, each amphipod species being represented by
796 a given geometric symbol (see also Fig 1). Colour code refers to lineages identified by the RPB1
797 marker (see Fig 3). PP: Bayesian Posterior Probability.

798
799 **Figure 3.** Bayesian phylogenetic reconstruction of *Nosema* spp. based on fragments F2 and F4 of
800 the RPB1 gene. For parasite sequences found in amphipods labels include: the GenBank accession
801 number in case of published sequences, the host species name (abbreviated for gammarids, see
802 below), the abbreviated SSU haplogroup name (as in S2 Table and Fig2), the RPB1 haplogroup
803 name, population name and the number of host individuals infected. Other sequences from
804 Genbank are representatives of further *Nosema* and *Vairimorpha* species (see S2 Table). PP =
805 Bayesian Posterior Probability, nodes supported by the Maximum Likelihood tree by bootstrap
806 values higher than 0.7 are indicated by red dots. Bars annotated on the right represent results of
807 the molecular species delimitation ASAP and bPTP methods. *– ASAP analysis was performed
808 using only the sequences covering both F2 and F4 RPB1 fragments; ? – haplogroup excluded from
809 the ASAP analysis because of lack of F2 or F4 fragment. Abbreviations for gammarid species
810 names: Gbal – *Gammarus balcanicus*, Gdue – *Gammarus duebeni*, Gfos – *Gammarus fossarum*,
811 Gpul – *Gammarus pulex*, Groe – *Gammarus roeselii*.

812 **Supporting information captions:**

813 **S1 Figure.** A. Overview of the amino-acid alignment for the RNA polymerase II largest subunit
814 (RPB1) gene haplogroups identified in amphipods. Grey colour stands for absence of amino acids.
815 B. Overview of the nucleotide alignment for the RPB1 gene haplogroups, for two fragments (F2
816 and F4). Grey colour stands for either absence of PCR product or shorter sequences for a given
817 fragment.

818
819 **S2 Figure. A.** Details of Clade A of *Nosema* phylogenetic reconstruction. Names in black are
820 *Nosema* haplogroups, names in red or blue are names of host clades to which they are associated.
821 Legends are similar to Fig 3 in the paper. Host clade names are given from the *Gammarus*
822 *balcanicus* phylogenetic tree simplified from [36] presented in (B). N and S represent the two
823 major *G. balcanicus* groups that differentiated around 18 MYA. C. Map showing sites where host
824 and parasites are coming from, where the limit between N and S host groups were redrawn after
825 [1]. Lines between A and B trees indicates the position of host individuals infected by *Nosema*
826 haplogroups on host phylogenetic tree. The sites were plotted on a map from Natural Earth
827 resources in QGIS 3.32.0-Lima [2].

828
829 **S1 Table.** A list of 85 sites where *Nosema* spp. were recorded in amphipod crustaceans. The names
830 of the SSU haplogroups are as in Fig 2 and for RPB1 clades/lineages as in Fig 1 and 3.

831
832 **S2 Table.** A list of samples used in the study, with host, site and country name, as well as GenBank
833 Accession numbers given. Sequences in blue are samples of *Nosema* infections outside the
834 amphipod hosts, used for the phylogenetic reconstructions. **sequences too short to be deposited
835 in GenBank (see S2 Appendix). References: Refence list is provided in the spread-sheet
836 REFERENCES. DS : Direct submission to Genbank, no published article associated.

837
838 **S3 Table.** Estimates of Average Evolutionary Divergence over Sequence Pairs within Groups. The
839 number of base substitutions per site from averaging over all sequence pairs within each group are
840 shown. Analyses were conducted using the Kimura 2-parameter model [1]. This analysis involved
841 36 nucleotide sequences. Codon positions included were 1st+2nd+3rd+noncoding. All ambiguous
842 positions were removed for each sequence pair (pairwise deletion option). There were 1025
843 positions in the final dataset. Evolutionary analyses were conducted in MEGA11 [2] The presence

844 of n/c in the results denotes cases in which it was not possible to estimate evolutionary distances.

845 SE - standard error.

846

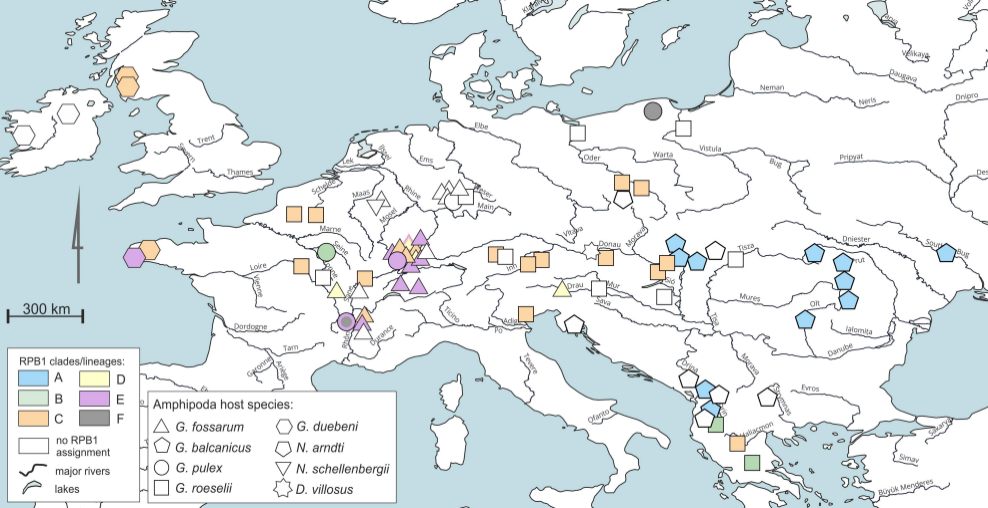
847 **S1 Appendix:** Alignment of F2 and F4 RPB1 fragments of *Nosema* spp., *Vairimorpha* spp. and
848 outgroup used for the Bayesian and Maximum Likelihood phylogeny reconstructions.

849

850 **S2 Appendix:** Small ribosomal subunit (SSU) rDNA sequences too short (<200bp) to be
851 deposited in GenBank.

852

853 **S3 Appendix:** Results of the ASAP and bPTP species delimitations.



● PP > 0.9

● PP > 0.7

I = Insects
 D = Decapoda
 A = Amphipoda
 Dv = *Dikerogammarus villosus*
 Ns = *Niphargus schellenbergi*
 Na = *Niphargellus arndti*
 Ev = *Eulimnogammarus verrucosus*
 Gb = *Gammarus balcanicus*
 Gd = *Gammarus duebeni*
 Gf = *Gammarus fossarum*
 Gp = *Gammarus pulex*
 Gr = *Gammarus roeselii*

0.07

Ngra p01
 KM657357 *N. granulosis* r03
 MK241534 *N. sp. L84*

Ngra H4
Ngra H1
Ngra H2
Ngra H3

Ngra p02
 MK241529 *N. sp. L105* b01
 MK241535 *N. sp. L146*

Ngra f04
 MK719277 *N. granulosis* r01

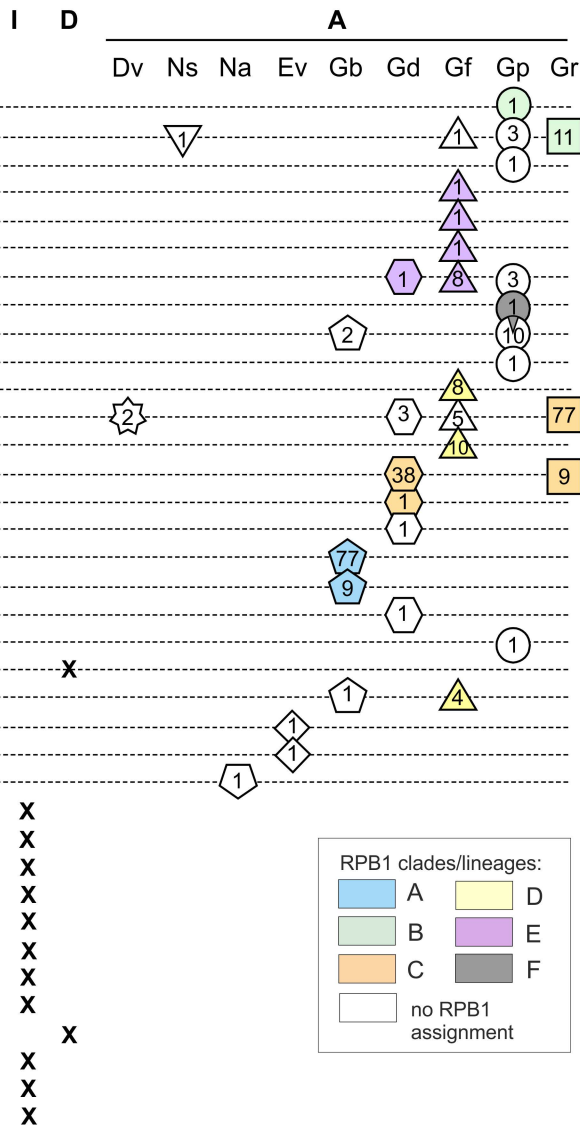
Ngra f03
 FN434087 *N. granulosis* r02
 AJ011833 *N. granulosis*

Ngra d01
Ngra b03
Ngra b04
 FN434088 *N. granulosis*
 KM657356 *N. granulosis* PT31
 MF348634 *N. austropotambonii*

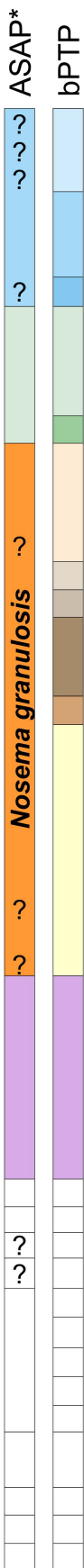
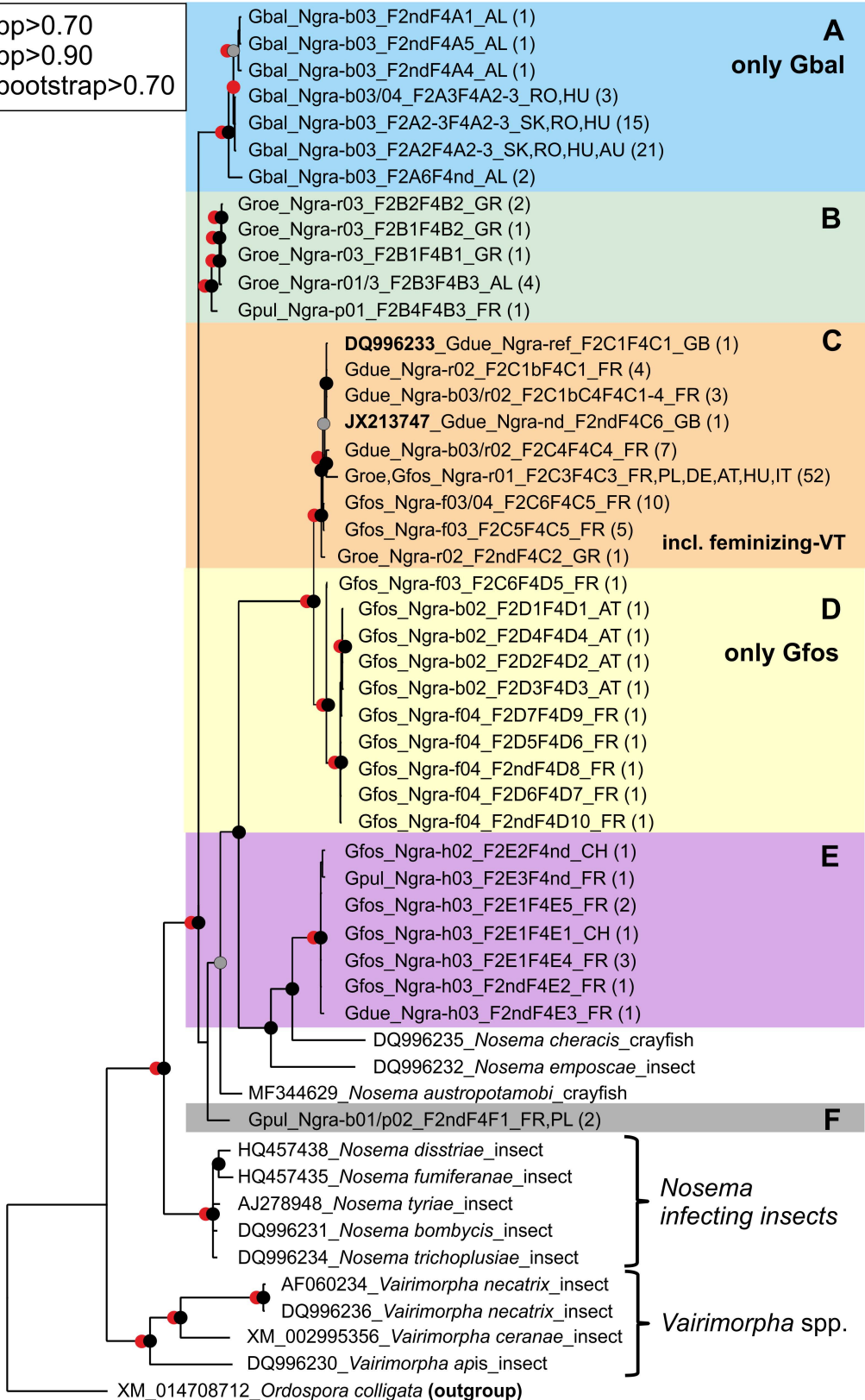
Ngra b02
 KM977840 *N. sp. Vr.28*
 KM977841 *N. sp. VR31*
 MN959960 *N. sp. B682*

EU864526 *N. antheraeae*
 EU219085 *N. distriae*
 KT020736 *N. fumiferanae*
 AY960987 *N. plutellae*
 D85503 *N. bombycis*
 NTU09282 *N. tricoplusiae*
 AJ012606 *N. tyriae*
 DQ996238 *N. empoascae*

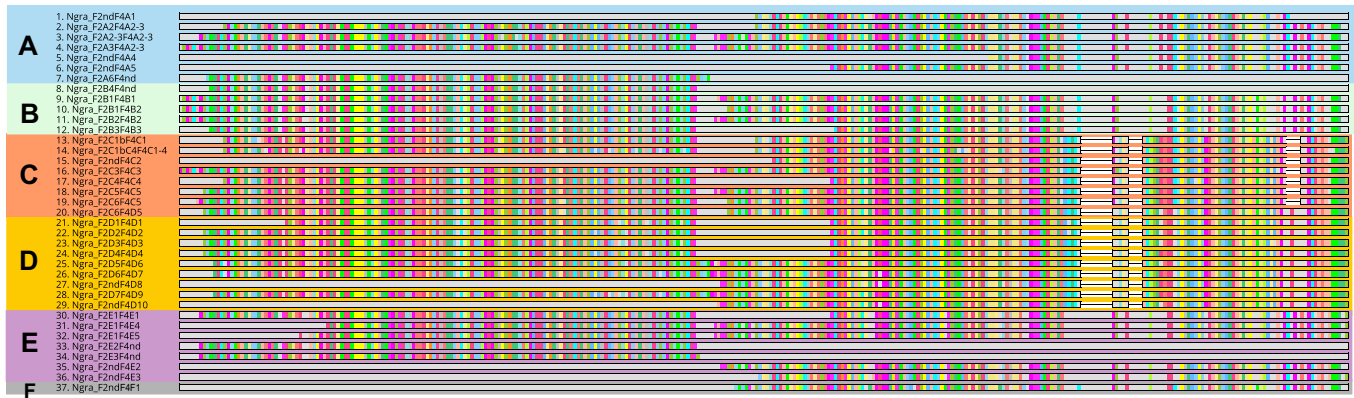
AF327408 *N. cheracis*
 AU26534 *V. apis*
 NCU26533 *V. ceranae*
 Y00266 *V. necatrix*



● pp>0.70
 ● pp>0.90
 ● bootstrap>0.70



A



B

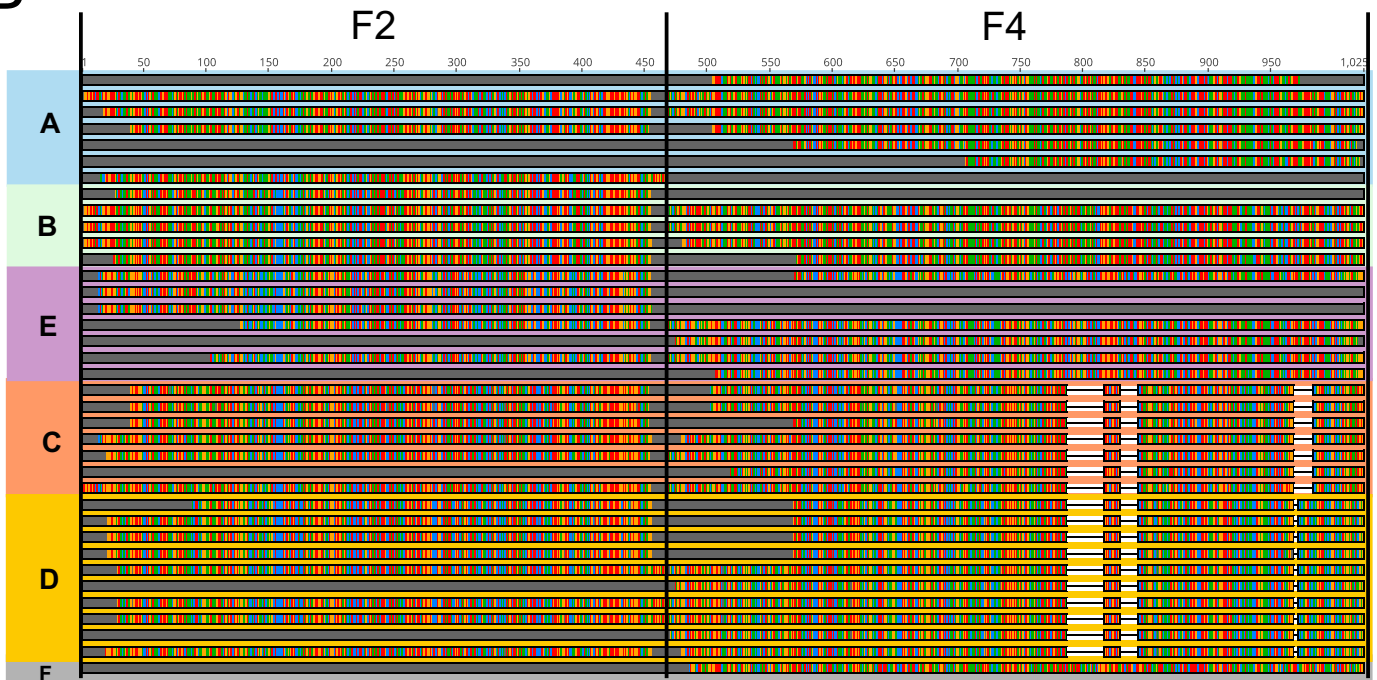
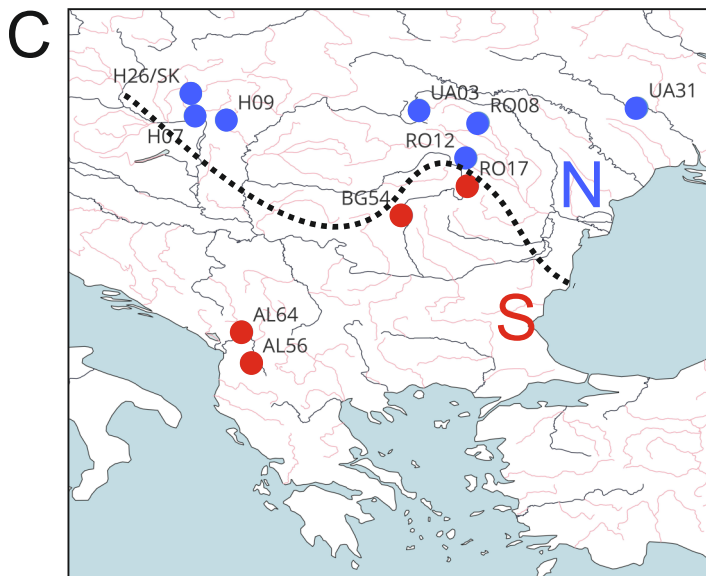
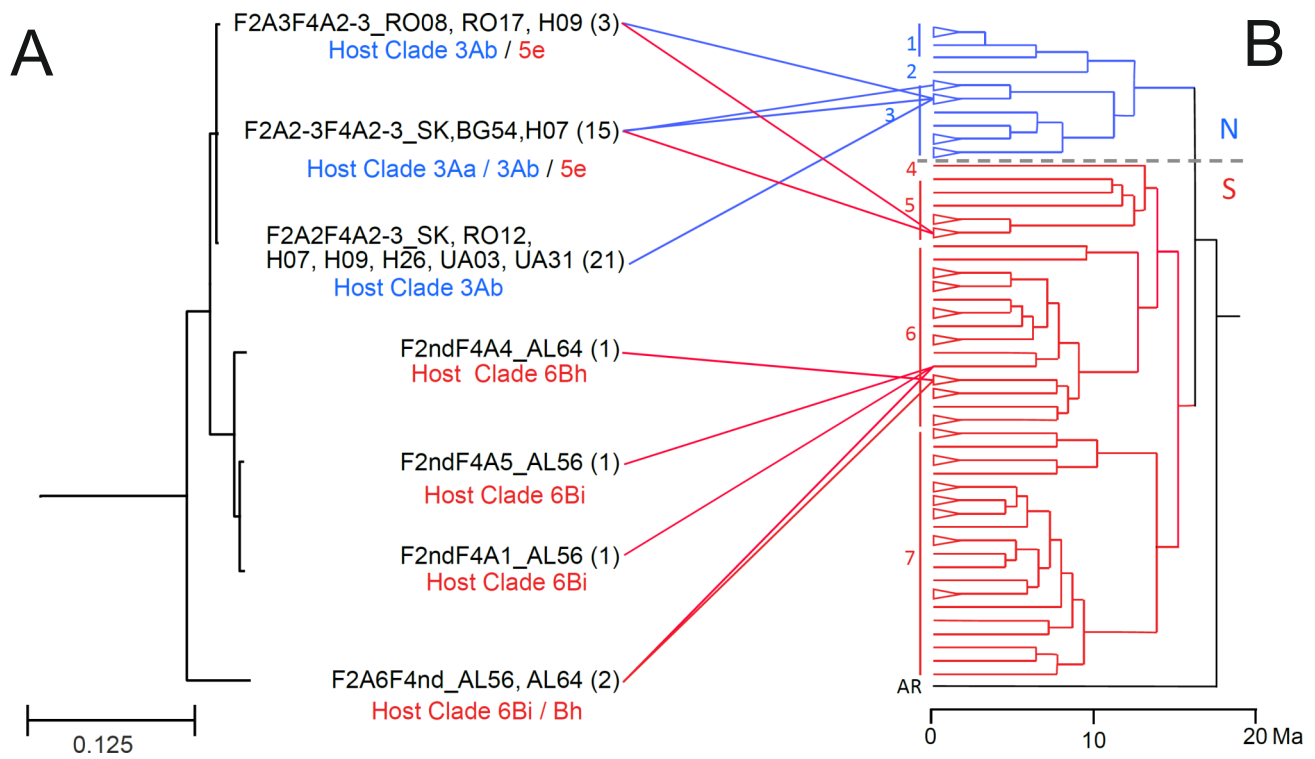


Figure S1.

A. Overview of the amino-acid alignment for the RNA polymerase II largest subunit (RPB1) gene haplogroups identified in amphipods. Grey color stands for absence of amino acids.

B. Overview of the nucleotide alignment for the RPB1 gene haplogroups, for two fragments (F2 and F4). Grey color stands for either absence of PCR product or shorter sequences for a given fragment.



S2 Figure. A. Details of Clade A of *Nosema* phylogenetic reconstruction. Names in black are *Nosema* haplogroups, names in red or blue are names of host clades to which they are associated. Legends are similar to Fig 3 in the paper. Host clade names are given from the *Gammarus balcanicus* phylogenetic tree simplified from [1] presented in (B). N and S represent the two major *G. balcanicus* groups that differentiated around 18 MYA. C. Map showing sites where host and parasites are coming from, where the limit between N and S host groups were redrawn after [1]. Lines between A and B trees indicates the position of host individuals infected by *Nosema* haplogroups on host phylogenetic tree. The sites were plotted on a map from Natural Earth resources in QGIS 3.32.0-Lima [2].

[1] 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.* 2016;25: 795–810. doi:10.1111/mec.13499

[2] QGIS Development Team, 2023. QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>

Table S1 : 85 sites were *Nosema* where recorded in amphipod crustaceans. The names of the SSI**References:** Reference list is provided in the spread-sheet REFERENCES. DS : Direct submission to Gen

Population Acronym	Country	ISO Country	River/Lake	Geographic Latitude	Geographic Longitude
SJL	France	FR	Saone	47.1000	5.2600
SPI	Austria	AT	Danube	48.3600	15.4200
LYS	Russia	RU	Baikal	51.8500	104.8600
AL56	Albania	AL	Drin	41.5100	20.3000
AL57	Albania	AL	Drin	41.7380	20.3020
AL60	Albania	AL	Drin	41.9670	20.3850
AL64	Albania	AL	Drin	42.3150	20.0500
AL69	Montenegro	ME	Danube	42.8360	19.5690
BG25	Bulgaria	BG	Struma	42.2290	22.8880
BG54	Romania	RO	Danube	45.3460	24.2050
H07	Hungary	HU	Danube	47.9360	18.8320
H09	Hungary	HU	Danube	47.8270	19.6450
H11	Hungary	HU	Danube	47.9830	20.4400
H26	Slovakia	SK	Danube	48.5210	18.7260
HR30	Croatia	HR	Vretenica	45.1490	14.5600
RO08	Romania	RO	Danube	47.7520	26.2010
RO12	Romania	RO	Danube	46.8380	25.8820
RO17	Romania	RO	Danube	46.1120	25.9200
SK	Slovakia	SK	Danube	48.5210	18.7260
UA03	Ukraine	UA	Prypiat	48.0840	24.6560
UA31	Ukraine	UA	Southern Buh	48.1410	30.3260
CUM	UK Scotland	UK	Small river - no name	55.7600	-4.9200
KER	France	FR	Small river no name	47.9800	-4.3800
KRE1	R. of Ireland	IE	Eme	54.1800	-7.4800
KRE2	R. of Ireland	IE	Lough Corrib	53.4500	-9.3200
ROS	France	FR	Tributary Elorn	48.3900	-3.9700
TER	UK Scotland	UK	White Bay	55.7900	-4.9100
38.1-GALA	France	FR	Galaveyson	45.2739	5.1186
39.1	France	FR	Guiers-Mort	45.3617	5.7544
40.3-COMB	France	FR	Agy	45.5300	5.3075
57.1-MOUG	France	FR	Mouge	46.4397	4.7564
58.1	France	FR	Liepvrette	48.2203	7.1441
58.2	France	FR	Tributary Liepvrette	48.2295	7.1699
58.4	France	FR	Tributary Liepvrette	48.2270	7.1667
58.5	France	FR	Tributary Liepvrette	48.2465	7.1741
59.4	France	FR	Tributary Liepvrette	48.2580	7.2207
60.2	France	FR	Rombach	48.2759	7.2780

60.3	France	FR	TributaryLiepvrette	48.2667	7.2917
60.4	France	FR	Tributary Liepvrette	48.2647	7.2734
61.2	France	FR	Rösselbach	48.2762	7.3080
61.4	France	FR	TributaryLiepvrette	48.2727	7.3394
ALPM04	Austria	AT	Small stream near Treffen	46.6997	13.8243
ALPM22	Switzerland	CH	Thunersee	46.7039	7.6576
CHGF	Switzerland	CH	Alpine Lake	nd	nd
GRU1	Germany	DE	Gründau	50.2599	9.1745
GRU2	Germany	DE	Gründau	50.2407	9.1441
KIN2	Germany	DE	Kinzig	50.3427	9.5525
SAL2	Germany	DE	Salz	50.3143	9.3669
STE1	Germany	DE	Steinebach	50.3323	9.4665
HUD	Luxembourg	LU	Fledermaustunnel	50.1500	6.0200
11.8	France	FR	Le vernidard	47.3895	4.0904
52P-ARDI	France	FR	Ardière	46.1867	4.5211
54P	France	FR	Cance	45.1736	4.5011
59.4	France	FR	Tributary Liepvrette	48.2580	7.2207
OMDE	Germany	DE	Tributary Rhine	51.3403	6.5149
STE1	Germany	DE	Steinebach	50.3323	9.4665
WRO	Poland	PL	Slupia river	54.4520	17.0330
AL35	Albania	AL	Tributary Shkumbin	41.0746	20.4879
ALP41	Germany	DE	Amper/Isar	48.0766	11.1339
ALP43	Germany	DE	Loisach/Isar	47.6567	11.3556
ALP49	Germany	DE	Chiemsee	47.8838	12.4176
ALP50	Germany	DE	GötzingerAchen/Salzach	47.9539	12.7487
BF04	France	FR	Somme	49.9238	2.2294
BF05	France	FR	Omignon	49.8918	3.1707
COU	France	FR	Yonne	47.5272	3.5408
GR11	Greece	GR	Aliakmonas	40.2889	21.4511
GR25	Greece	GR	Sofaditikos	39.1762	22.0460
GRU	Poland	PL	Osa	53.5370	18.7968
GRU2	Germany	DE	Gründau	50.2407	9.1441
GSSL	Slovenia	SI	Tributary Drava	46.5442	15.6144
H05	Hungary	HU	Tributary Danube	47.2947	18.0536
H16	Hungary	HU	Sajo / Tisza	47.9664	21.0499
HUS10	Hungary	HU	Tributary Danube	46.2877	18.2737
KIN1	Germany	DE	Kinzig	50.3427	9.5806
NIT	Italie	IT	Sile	45.6557	12.2195
OU	France	FR	Ouche	47.1483	5.2775
PA	Austria	AT	Pielach	48.0290	15.4331
PL1	Poland	PL	Tywa	53.2259	14.4973
PL8	Poland	PL	Oławain Siechnice	51.0421	17.1589
PRO	Poland	PL	Kaczawa	51.2790	16.3642
SLB	Slovakia	SK	Hron	47.9182	18.6459
STG	France	FR	Loire	47.6997	2.5422
STE1	Germany	DE	Steinebach	50.3323	9.4665
NifARN	Poland	PL	aff Pelczina	50.8170	16.3304
HUD	Luxembourg	LU	Fledermaustunnel	50.1500	6.0200

M - males

F - females

nd - no data

xxx - not counted

In Blue - sequences too short to be firmly assigned to a given haplogroup ('?' denotes possible haplotype)

logroups)

Reference	RPB1 CLADES/LINEAGE:	A	B	C	D	E	F
		Wattier et al 2007	nd				
Wattier et al 2007	nd						
Madyarova et al 2015	nd						
This Study	A	3					
This Study	nd						
This Study	nd						
This Study	A	2					
This Study	nd						
This Study	nd						
This Study	A	1					
This Study	A	6					
This Study	A	5					
This Study	nd						
This Study	A	7					
This Study	nd						
This Study	A	1					
This Study	A	1					
This Study	A	1					
This Study	A	15					
This Study	A	2					
This Study	A	1					
Ironside 2013	C			1			
This study	E					1	
Krebes et al 2010	nd						
Krebes et al 2010	nd						
This study	C			14			
Terry et al 1999	C			1			
This study	E					1	
This study	nd						
This study	C			1			
This study	nd						
This study	C,D			12	1		
This study	C			2			
This study	nd						
This study	E					1	
This study	E					2	
This study	C			1			

This study	E				1		
This study	C,D,E		2	1	2		
This study	D			3			
This study	E				1		
This study	D			4			
This study	E				1		
This study	E				1		
Prati et al 2022	nd						
Prati et al 2022	nd						
Prati et al 2022	nd						
Prati et al 2022	nd						
Prati et al 2022	nd						
Weigand et al 2016	nd						
This study	B	1					
This study	D			1			
This study	E,F				1	1	
This study	E				1		
Prati et al 2022	nd						
Prati et al 2022	nd						
Wroblewski, DS.	F					1	
Quiles et al 2019	B	4					
Quiles et al 2019	C		1				
Quiles et al 2019	nd						
Quiles et al 2019	C		3				
Quiles et al 2019	C		1				
Quiles et al 2019	C		2				
Quiles et al 2019	C		8				
Haine et al 2004	nd						
Quiles et al 2019	C		1				
Quiles et al 2019	B	5					
Quiles et al 2019	nd						
Prati et al 2022	nd						
Quiles et al 2019	nd						
Quiles et al 2019	C		2				
Quiles et al 2019	nd						
Quiles et al 2019	nd						
Prati et al 2022	nd						
Quiles et al 2019	C		1				
Quiles et al 2019	C		13				
Quiles et al 2019	C		5				
Quiles et al 2019	nd						
Quiles et al 2019	C		10				
Quiles et al 2019	C		1				
Quiles et al 2019	C		3				
Quiles et al 2019	C		1				
Prati et al 2022	nd						
Grabner et al 2020	nd						
Weigand et al 2016	nd						
	nd	45	10	86	10	13	2

<i>Nosema granulosis</i>	Ngra	SK100	SK	SK	662	YES	Ngra b03	OR166005	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142010	OR166653	This Study	
<i>Nosema granulosis</i>	Ngra	SK15	SK	SK	690	YES	Ngra b03	OR165994	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142011	OR166654	This Study	
<i>Nosema granulosis</i>	Ngra	SK2	SK	SK	690	YES	Ngra b03	OR165991	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142011	OR166670	This Study	
<i>Nosema granulosis</i>	Ngra	SK20	SK	SK	662	YES	Ngra b03	OR165995	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142012	OR166671	This Study	
<i>Nosema granulosis</i>	Ngra	SK27	SK	SK	693	YES	Ngra b03	OR165996	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142013	OR166672	This Study	
<i>Nosema granulosis</i>	Ngra	SK41	SK	SK	720	YES	Ngra b03	OR165997	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141991	OR166674	This Study	
<i>Nosema granulosis</i>	Ngra	SK49	SK	SK	320	YES	Ngra b03	OR165998	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141992	OR166673	This Study	
<i>Nosema granulosis</i>	Ngra	SK5	SK	SK	690	YES	Ngra b03	OR165992	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142005	OR166656	This Study	
<i>Nosema granulosis</i>	Ngra	SK58	SK	SK	390	YES	Ngra b03	OR165999	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142009	OR166657	This Study	
<i>Nosema granulosis</i>	Ngra	SK6	SK	SK	690	YES	Ngra b03	OR165993	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR142006	OR166658	This Study	
<i>Nosema granulosis</i>	Ngra	SK69	SK	SK	697	YES	Ngra b03	OR166000	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	NO							
<i>Nosema granulosis</i>	Ngra	SK82	SK	SK	691	YES	Ngra b03	OR166001	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR142008	OR166659	This Study	
<i>Nosema granulosis</i>	Ngra	SK92	SK	SK	685	YES	Ngra b03	OR166002	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141993	OR166660	This Study	
<i>Nosema granulosis</i>	Ngra	SK94	SK	SK	690	YES	Ngra b03	OR166003	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141994	-	This Study	
<i>Nosema granulosis</i>	Ngra	SK98	SK	SK	693	YES	Ngra b03	OR166004	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141995	OR166661	This Study	
<i>Nosema granulosis</i>	Ngra	SLB-05	SLB	PL	687	YES	Ngra r01	MK719268	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	YES	F2C3F4nd	H05-02	C	OR142080	-	This Study	
<i>Nosema granulosis</i>	Ngra	SLB-08	SLB	PL	687	YES	Ngra r01	MK719269	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	YES	F2C3F4nd	H05-02	C	OR142081	-	This Study	
<i>Nosema granulosis</i>	Ngra	SLB-12	SLB	PL	317	YES	Ngra r01	MK719264	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	YES	F2C3F4nd	H05-02	C	OR142082	-	This Study	
<i>Nosema granulosis</i>	Ngra	SPI-40	SPI	AT	Shorter/365	YES	Ngra r01	EF091823	NO	Amphipoda	<i>Dikerogammarus villosus</i>	Dvil	Wattier et al 2007	NO							
<i>Nosema granulosis</i>	Ngra	STE1-A	STE1	DE	299	YES	Ngra r01	ON113518	NO	Amphipoda	<i>Gammarus fossarum</i>	Gfos	Prati et al 2022	NO							
<i>Nosema granulosis</i>	Ngra	STE1-B	STE1	DE	299	YES	Ngra r01	ON113518	NO	Amphipoda	<i>Gammarus pulex</i>	Gpul	Prati et al 2022	NO							
<i>Nosema granulosis</i>	Ngra	STE1-C	STE1	DE	299	YES	Ngra r01	ON113518	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Prati et al 2022	NO							
<i>Nosema granulosis</i>	Ngra	STG-03	STG	FR	480	YES	Ngra r01	MK719434	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	NO							
<i>Nosema granulosis</i>	Ngra	STG-09	STG	FR	480	YES	Ngra r01	MK719435	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	YES	F2C3F4nd	H05-02	C	OR142083	-	This Study	
<i>Nosema granulosis</i>	Ngra	STG-12	STG	FR	304	YES	Ngra r01	MK719432	NO	Amphipoda	<i>Gammarus roeselii</i>	Groe	Quiles et al., 2019	NO							
<i>Nosema granulosis</i>	Ngra	TER-1	TER	GB (SCO)	1143/1168	YES	AJ011833	YES	AJ011833	YES	Amphipoda	<i>Gammarus duebeni</i>	Gdue	Terry et al., 1999	YES	F2C1F4C1	DQ996233	C	DQ996233	-	Ironside 2007
<i>Nosema granulosis</i>	Ngra	UA03-07M	UA03	UA	726	YES	Ngra b03	OR166006	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2-3F4A2-3	H07-04	A	OR141999	-	This Study	
<i>Nosema granulosis</i>	Ngra	UA03-12M	UA03	UA	726	YES	Ngra b03	OR166007	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4A2-3	SK41	A	OR141996	OR166669	This Study	
<i>Nosema granulosis</i>	Ngra	UA31-22M	UA31	UA	717	YES	Ngra b03	OR166008	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	YES	F2A2F4nd	SK41	A	OR141997	-	This Study	
<i>Nosema granulosis</i>	Ngra	UA31-M58	UA31	UA	1011	YES	Ngra b03	OR166009	NO	Amphipoda	<i>Gammarus balcanicus</i>	Gbal	This Study	NO							
<i>Nosema granulosis</i>	Ngra	HUD	LU	LU	337	YES	Ngra r03	KT633991*	NO	Amphipoda	<i>Gammarus fossarum</i>	Gfos	Weigand et al., 2016	NO							
<i>Nosema granulosis</i>	Ngra	HUD	LU	LU	337	YES	Ngra r03	KT633991*	NO	Amphipoda	<i>Niphargus schellenbergi</i>	Nsch	Weigand et al., 2016	NO							
<i>Nosema empaoscae</i>	Nemp	-	-	-	1189	YES	DQ996238	YES	DQ996238	NO	Hemiptera	<i>Empoasca fabae</i>	Efab	Ironside 2007	YES	DQ996232	DQ996232				Ironside 2007
<i>Vairimorpha austrapotamobii</i>	Vaus	-	-	-	1139	YES	MF344634	YES	MF344634	NO	Decapoda	<i>Austropotamobius pallipes</i>	Apal	Pretto et al 2017	YES	MF344629	MF344629				Pretto et al 2017
<i>Vairimorpha cheracis</i>	Vche	-	-	-	1127	YES	AF327408	YES	AF327408	NO	Decapoda	<i>Cherax destructor</i>	Cdes	Moodie et al 2003	YES	DQ996235	DQ996235				Ironside 2007
<i>Vairimorpha necatrix</i>	Vnec	-	-	-	-	-	-	-	-	NO	Lepidoptera	<i>Pseudaletia unipuncta</i>	Puni	Ironside 2007	YES	DQ996236	DQ996236				Ironside 2007
<i>Nosema antheraeae</i>	Nant	-	-	-	1232	YES	EU864526	YES	EU864526	YES	Lepidoptera	<i>Antheraea pernyi</i>	Aper	Shen DS	NO						
<i>Nosema plutellae</i>	Nplu	-	-	-	1230	YES	AY960987	YES	AY960987	YES	Lepidoptera	<i>Plutella xylostea</i>	Pxyl	Ku et al 2007	NO						
<i>Nosema bombycis</i>	Nbom	-	-	-	1232	YES	D85503	YES	D85503	YES	Lepidoptera	<i>Bombyx mori</i>	Bmor	Hatakeyama DS	YES	DQ996231	DQ996231				Ironside 2007
<i>Nosema disstriae</i>	Ndis	-	-	-	1232	YES	EU219085	YES	EU219085	YES	Lepidoptera	<i>Malacosoma disstria</i>	Mdis	Kyei-Poku et al 2008	YES	HQ457438	HQ457438				Kyei-Poku unpub
<i>Nosema trichoplusiae</i>	Ntri	-	-	-	1233	YES	NTU09282	YES	U09282	YES	Lepidoptera	<i>Trichoplusia ni</i>	Tni	Fries et al 1996	YES	DQ996234	DQ996234				Ironside 2007
<i>Nosema tyriae</i>	Ntyr	-	-	-	1233	YES	AJ012606	YES	AJ012606	YES	Lepidoptera	<i>Tyria jacobaeae</i>	Tjac	Canning et al 1999	YES	AJ278948	AJ278948				Cheney et al unpub
<i>Ordaspora colligata</i>	Ocol	-	-	-	1973	YES	AF394529	YES	AF394529	YES	Diplostraca	<i>Daphnia magna</i>	Dmag	Refardt et al 2002	YES	XM_014708712	XM_014708712				Pombert et al 2015
<i>Vairimorpha opis</i>	Vapi	-	-	-	1242	YES	NAU26534	YES	NAU26534	YES	Hymenoptera	<i>Apis mellifera</i>	Amel	Fries et al 1996	YES	DQ996230	DQ996230				Ironside 2008
<i>Vairimorpha ceranae</i>	Vcer	-	-	-	1259	YES	NCU26533	YES	U26533	YES	Hymenoptera	<i>Apis cerana</i>	Acer	Fries et al 1996	YES	XM_002995356	XM_002995356				Cormann et al 2009
<i>Vairimorpha necatrix</i>	Vnec	-	-	-	1244	YES	Y00266	YES	Y00266	YES	Lepidoptera	<i>Lacanobia oleracea</i>	Lola	Vossbrinck et al 1987	YES	AF060234	AF060234				Hirt et al 1999

Table S3. Estimates of Average Evolutionary Divergence over Sequence Pairs within Groups

The number of base substitutions per site from averaging over all sequence pairs within each group are shown.

Analyses were conducted using the Kimura 2-parameter model [1]. This analysis involved 36 nucleotide sequences.

Codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

There were a total of 1025 positions in the final dataset.

Evolutionary analyses were conducted in MEGA11 [2] The presence of n/c in the results denotes cases in which it was not possible to estimate evolutionary distances.

WITHIN groups		SE
A	0.02299468	0.003771363
B	0.020679952	0.004310813
E	0.008799772	0.002409033
C	0.022969903	0.00358103
D	0.003240104	0.001249442
F	n/c	n/c

AMONG groups	SE						
	A	B	E	C	D	F	
A			0.0159	0.0255	0.0253	0.0320	0.0237
B		0.1920		0.0230	0.0208	0.0234	0.0209
E		0.3492	0.2993		0.0262	0.0260	0.0392
C		0.3247	0.2868	0.3304		0.0128	0.0291
D		0.3735	0.3284	0.3381	0.1308		0.0395
F		0.1981	0.1776	0.3793	0.3149	0.3728	

1. Kimura M. (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

2. Tamura K., Stecher G., and Kumar S. (2021). MEGA 11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msab120>.

>AF060234

AAAGGAGAAGAAGAAAGTGATGGTAAAGTAATTTTGAATGGAGAAAGAGTTCATAATATT
CTGAAGAAGATTGTAAATGAAGATGCTGTGTTTTTGGGTTTTGATCAGAAGTTTACTAAG
CCTGAGTGGCTTATTTTACTGTTCTTTTAGTTCTCTCTCTCTCTGTAAGGCCTAGTATT
GTCATGGAAGGAATGTTGAGAGCAGAAGATGATTTGACTCATAAATTGGCAGATATAGTA
AAAGCTAATACTTATCTTAAGAAATATGAATTAGAAGGGGCACCTGGTCATGTAGTAAGG
GATTATGAACAATTACTCCAATTTTCATATTGCCACAATGATTGACAATGACATAAGTGGC
CAACCACAGGCACTTCAGAAAAGTGGGCGCCCTTTAAAAAGCATTCTGCTCGACTGAAA
GGCAAAGAAGGGCGAGTCCGGGGAAATTTAATGGGAAAAAGGGTAGATATGAGCATGATG
GCTCATTTTGTCCGAGTCATGGAAGGCCAAAACCTTTAGACTAAATTTGTCTTGTGTCTCG
CCCTATAACGCGGATTTTACGGGGACGAAATGAATCTTCACATGCCGCAAAGTTACAAC
AGTAAGGCGGAATTAGAAGAATTATGTTTAGTAAGTAAACAAGTATTAAGTCCACAATCT
AATAAACCAGTAATGGGGATAGTACAAGACAGTTTAAACAGCTCTAAGACTTTTTACATTA
CGAGATTCATTTTTTGTATCGTAGAGAAACAATGCAATTATTATATTCTGTAAATATTAAT
AATTATGAGTTTACTGATTCGTCGAAATTAATTATGACTCATGATGACTCATTGGTAAT
AATTTACATACTGAAGAATCTTCAAATATTATGAAAATATTAATTTCCAGCTATTTCT
TATCCTAAGAAATTGTGGACTGGTAAACAATCTTAAGTTACATTTTACCAAATACAATT
TATAATGGAAAATCTAATGAGCACAACGAAGAAGACTTGGAAAATGTCTGAAGATTCTTAC
GTCATAATTAGAAATGG

>DQ996236

AAAGGAGAAGAAGAAAGTGATGGTAAAGTAATTTTGAATGGAGAAAGAGTTCATAATATT
TTAAGAAGATTGTAAATRAAGATGCTGTGTTTTTGGGTTTTGATCAGAAGTTTACTAAG
CCCGAGTGGCTTATTTTACTGTTCTTTTAGTTCTCTCTCTCTCTGTAAGGCCTAGTATT
GTCATGGAAGGAATGTTGAGAGCAGAAGATGATTTGACTCATAAATTGGCAGATATTGTG
AAAGCTAATACTTATCTTAAGAAATATGAATTAGAAGGGGCACCTGGTCATGTAGTAAGG
GATTATGAACAATTACTCCAATTTTCATATTGCCACAATGATTGACAATGACATAAGTGGC
CAACCACAGGCACTTCAGAAAAGTGGGCGCCCTTTAAAAAGCATTCTGCTCGACTGAAA
GGCAAAGAAGGGCGAGTCCGGGGAAATTTAATGGGAAAAAGGGTAGATATGAGCATGATG
GCTCATTTTGTCCGAGTCATGGAAGGCCAAAACCTTTAGACTAAATTTGTCTTGTGTCTCG
CCGTATAACGCGGATTTTACGGGGACGAAATGAATCTTCACATGCCGCAAAGTTACAAC
AGTAAGGCGGAATTAGAAGAATTATGTTTAGTAAGTAAACAAGTATTAAGTCCACAATCT
AATAAACCAGTAATGGGGATAGTACAAGACAGTTTAAACAGCTCTAAGACTTTTTACATTA
CGAGATTCATTTTTTGTATCGTAGAGAAACAATGCAATTATTATATTCTGTAAATATTAAT
AATTATGAGTTTACTGATTCGTCGAAATTAATTATGACTCATGATGACTCATTGRTAAT
AATTTACATACTGAAGAATCTTCAAATATTATGAAAATATTAATTTCCAGCTATTTCT
TATCCTAAGAAATTGTGGACTGGTAAACAATCTTAAGTTACATTTTACCAAATACAATT
TATAATGGAAAATCTAATGAGCACAACGAAGAAGACTTGGAAAATGTCTGAAGATTCTTAC
GTCATAATTAGAAATGG

>XM_002995356

AAAGGTGAAGAAGACAGTGATGGAAAAGTTATTTTAAATGGTGAAAGGGTGCATAATATT
TTAAGAAAATTTCTGATGAAGATTCAACATATTTAGGATTTGATACAAAATATAGTAAA
CCGGATTGGCTTGTATAACTGTACTTCTTGTCTCTCCACCATCAGTTAGACCATCTATT
GTTATGGAAGGGATGCTTAGAGCAGAAGATGATTTGACACATAAATTGGCTGATATAGTT
AAAGCTAACACGTATTTAAAAAATATGAATTAGAAGGAGCACCAGGTCATGTGATAAGA
GATTATGAACAGTTATTACAATTTTCATATTGCTACTATGATTGATAATGATTTAAGTGG
CAACCACAAGCTTTACAAAAAAGTGGAAAGACCAATAAAATCCATTAGTGCTAGATTA
GGTAAAGAAGGTCGTGTTAGAGGAAATTTAATGGGAAAAAGAGTAGATATGTCATGATG
GGACATTTTCGTTTCGAGTTATGGAAGGTAAAACCTTTAGATTAATTTAAGTTGTACGTCA
CCGTATAATGCGGATTTTCGATGGAGACGAAATGAATCTTCATATGCCACAGAGTTACAAT
ACTAAAGCAGAATTAGAAGAATTGTGTTTGGTTAGTAAACAAGTTCTAAGTCTCAATCA
AACAGACCTGTAATGGGTATTGTACAGGATTCTTTAACCAGTCTAAGGCTTTTTTACACTT
AGAGATACTTTTTTTGATAAAAGAGAAACAATGCAATTACTTTATTCCATAAACTTAAGC
AAATATGAAAATTTTAAT-----
AATGAAGATATGCATAAACT---TCTATTATGGATTTATTAATTTATCCAACAATTTCT
TATCCTAAGAAATTATGGACAGGAAAGCAAATCTTTAGTTATATTTTACCGAATATTGTA
TATTCAGCTGTGTCTAATGAACACGATGAAAATGATCTTGAAAATTTAAAAGATACAGTA
GTAATCATTTCGTAATGG

>DQ996230

AAAGGGGAAGAAAACCTCCGACGGTAAAGTTATTTTAAATGGAGAAAGGGTTTTGAGTATT
TTGAAAAAATTTCTGAACAAGATGCTAATTATATGGGATTTGATTTTGTAAATAGTAAA
CCCGAATGGTTAATTTTACTGTTTTGCTAGTTCCCTCCTCCTTCTGTAAGACCTTCTATT
GTTATGGAGGGAATGTTAAGAGCAGAAGATGATCTAACACACAAATTGGCTGATATTATA
AAAGCAAATACATATTTAAAAAATATGAGTTGGAAGGCGCTCCTGGTCATGTTATAAGG
GATTATGAACAATTACTACAATTTTCATATGGCTACAATGGTAGATAATGATTTAAGTGG
CAACCACAAGCTTTACAAAAAAGTGGAAAGGCCTATAAAGTCGATTAGTGCTAGATTA
GGAAAAGAAGGTCGAGTACGAGGAAATTTGATGGGCAAAAGGGTAGATATGTCTATGATG
GGTCATTACGTCAGAGTGATGGAAGGAAAAACRTTTTAGATTAATTTGAGTGTCACA
CCGTATAATGCAGATTTTACGGAGATGAAATGAATTTACACATGCCACAATCACATAAT
TCAAGAGCAGAGCTTGAAGAACTCTGTCTGGTATCTAAACAAGTATTGACACCACAAGCT
AATCGTCTATAATGGGTATTGTACAAGAYACACTAACTGGACTTCGTTTGTACTTTA
AGAGACTCGTTTTTTGATAAACGAGAAACGATGCTATTACTTTATTCAGTAGATCTTGAG
AAATATGAAAATTTA-----

-----AAAGAAGGT---GGAATTTGTGAATTTATGAATYGTCCAGCAATTTCA
TATCCAAAGMAATATTGGACAGGAAAACAAATTATCAGTTTTATTTACCTGATATAATT
TTAATGGAAATTCAAATGAATTTGATGARGAGGATTTAGAAAATGTTAAAGATTTAAA
ATTTATTTAAGAAACGG

>AJ278948

AAAGGAGAAGATAACTCAGAAGGGAAGGTCATTTTAAATGGAGAGAAAGTTCACAGTATT
TTTAAAAAGATTTCTGATGAAGATATTGAAGTTTTAGGTTTTGATCTTGCTTACAGTACA
CCCGAATGGCTTATTTTAACTGTCTTATTGGTACCTCCACCATCGGTAAGACCTTCGATT
GTTATGGAAGGGATGCTAAGAGCAGAAGATGATCTTACACACAAACTTTCTGATATTATC
AAATCCAACACTTACCTTAAGAAATACGAAATGGAAGGAGCACCAGGGCATATAATTCGA
GATTATGAACAACCTTTTACAATTTTCATGTGGCTACTTTAATAGACAATGACATTAGTGG
CAACCACAAGCCCTTCAGAAAAGTGGTAGACCTTTGAAATCTATTTCTGCCCGACTTAAA
GGAAAGGAAGGACGTGTTAGAGGAAATCTTATGGGTAAGAGAGTGGACATGTCTATGATG
GT---CACAGTTCGTGTGATG---AGGAAGACTTTCGTTT---AATTGAGTGTCACTTCT
CCNTATAATGCCGATTTTGTGATGGAGATGAAATGAATCTTCATATGCCACAAAATTACAAT
TCCATTGCAGAATTAGAAGAAATTTGTATGGTACCCAAACAAATTCTTGGACCTCAAAGT
AACAAACCTGTCATGGGAATCGTTCAAGATACTCTCACTGGTCTAAGATTTTTTACAGTT
CGAGATGCATTTTTTTGATAAAAATGAGATGATGCAAGTCTTTTATTCCATAGATTTTGAG
AAATATTATGATATAGGT---CTCGACAGTGTCATTA AAAAGGGTAAAAAATTGGA----
--CATGGGAGCTAAGGAGTAC---AGTTTAATGGGGTTACTCTCAAACCAGCTATTCA
AAACCAAACAGTTGTGGACAGGCAAACAAATTTAAGTTTTATTTTCCTAATGTCTAT
TATAAAAATTCGTCCAATGAAAGTCGGGAAAATGACTTAGAAAATGTCTCGGACACGTCT
GTAGTAATATGCGGAGG

>DQ996231

AAAGGAGAAGATAACTCAGAAGGGAAGGTTATTTTAAATGGAGAGAAAGTTCACAGTATT
TTTAAAAAGATTTCTGATGAAGATATTGAAGTTTTAGGTTTTGATCTTGCTTACAGTAGA
CCCGAATGGCTTATTTTAAACCGTCTTGTTGGTACCTCCACCATCAGTAAGACCTTCGATT
GTTATGGAAGGGATGTTAAGAGCAGAGGATGATCTTACACACAAACTTTCTGATATTATT
AAATCCAACACTTACCTTAAGAAATACGAAATGGAAGGAGCACCAGGGCATATAATTCGA
GATTATGAACAGCTTTTACAATTTTCATGTGGCTACTTTAATAGACAACGACATTAGTGG
CAGCCACAAGCACTTCAGAAAAGTGGTAGACCTTTGAAATCTATTTCTGCCCGACTTAAA
GGAAAGGAAGGGCGTGTTAGAGGAAATCTTATGGGTAAGAGAGTTGACATGTCTATGATG
GGCCACAAAGTTCGTGTGATGGAAGGGAAGACTTTCCGTTTAAATTTGAGTGTCACTTCT
CCTTATAATGCCGATTTTGTGATGGAGATGAAATGAATCTTCATATGCCACAAAATTACAAT
TCAATTGCTGAATTAGAAGAAATTTGTATGGTACCCAAACAAATTCTTGGACCTCAAAGT
AACAAACCTGTCATGGGGATCGTTCAAGATACTCTCACTGGTCTTAGATTTTTTACAGTT
CGAGATGCATTTTTTTGATAAAAATGAGATGATGCAAGTCTTTTATTCCATAGATTTTCGAT
AAATATTATGATATAGGT---CTTGACAGTGTCATTA AAAAGGGTAAAAAATTGGA----
--CATGGGTGCTAAGGAGTAC---AGTTTAATGGGGTTACTCTCAAACCAGCTATTCAA
AAACCAAACAGTTGTGGACAGGTAAGCAAATTTAAGTTTTATTTTCCTAATGTCTTT
TACAAAATTCGTCCAATGAAAGTCGGGAAAATGACTTAGAAAATGTCTCGGATACGTCT
GTAGTAATATGTGGAGG

>DQ996234

AAAGGAGAAGATAACTCAGAAGGGAAGGTCATTCTAAATGGAGAGAAAGTCCACAGTATT
TTTAAAAAGATTTCTGATGAAGATATTGAAGTTTTAGGTTTTGATCTTGCTTACAGTAGA
CCCGAATGGCTTATTTTAACCGTCTTGTTGGTACCTCCACCATCAGTAAGACCTTCGATT
GTCATGGAAGGGATGCTAAGAGCAGAGGATGATCTTACACACAACTTTCTGATATTATT
AAATCTAACACTTACCTTAAGAAATACGAAATGGAAGGAGCACCGGGACATATAATTCGA
GATTATGAACAACCTTTTACAATTTTCATGTGGCTACTTTAATAGACAACGACATTAGTGGA
CAACCACAAGCCCTTCAGAAAAGTGGTAGACCTTTGAAATCTATTTCTGCCCGACTTAAA
GGAAAGGAAGGACGTGTTAGAGGAAATCTTATGGGTAAGAGAGTGGACATGTCTATGATG
GGTCACAAAGTTCGTGTGATGGAAGGGAAGACTCTCCGTTTAAATTTGAGTGTCACTTCT
CCTTATAATGCCGATTTTGTGATGGAGATGAAATGAATCTTCATATGCCACAAAATTACAAT
TCCATTGCAGAATTGGAAGAAATTTGTATGGTACCCAAACAAATTCTTGGACCTCAAAGT
AACAAACCTGTCATGGGGATCGTTCAAGATACTCTCACTGGTCTTAGATTTTTTTACAGTT
CGAGATGCATTTTTTCGATAAAAATGAGATGATGCAAGTCCTTTATTCCATAGATTTTTGAT
AAATATTATGATATAGGT---CTTGACAGTGCATTAATAAAGGGTAAAAAATTGGA----
--CATGGGAGCTAAAGAGTAC---AGTTTAATGGGGTACTCTCAAACCAGCCATTCAA
AAACCAAACAGTTGTGGACAGGTAACAAATTTTAAGTTTTATTTTTCCCTAATGTCTTT
TACAAAATTCGTCCAATGAAAGTCGGGAAAATGACTTAGAAAATGTCTCGGATACCTCT
GTAGTAATATGTGGAGG

>HQ457438

AAAGGGGAAGATAACTCGGAAGGGAAGGTTATTTTAAACGGAGAAAAAGTTCATAGTATT
TTTAAAAAGATATCTGACGAAGACATTGAAGTTTTAGGTTTTGATCTTGCTTACAGTAGA
CCTGAATGGCTTATTTTAACCTGTTTTATTGGTTCCTCCACCATCTGTAAGACCTTCGATT
GTTATGGAAGGGATGTTAAGAGCAGAAGATGATCTTACACACAACTTTCTGATATTATT
AAATCCAACACTTACCTTAAGAAATACGAAATGGAAGGAGCACCGGTCATATAATTCGA
GATTATGAGCAACTTTTACAATTTTCATGTAGCAACTTTAATTGACAACGACATTAGTGGC
CAACCACAGGCCCTTCAGAAAAGTGGGCGACCTTTGAAATCTATTTCTGCTCGACTTAAA
GGAAAGGAAGGACGTGTTAGAGGAAATCTTATGGGTAAGAGAGTGGATATGTCTATGATG
GGGCACAAAGTCCGTGTGATGGATGGAAAGACTTTCCGTCTAAATTTGAGTGTCACTTCT
CCTTATAATGCCGATTTTGTGATGGAGATGAAATGAATCTTCATATGCCACAAAATTACAAT
TCTATTTCTGAACTAGAAGAAATTTGTATGGTTCCTCAAACAAATTCTTGGACCTCAAAGT
AATAAACCTGTCATGGGAATTGTTCAAGATACTCTCACTGGTCTAAGATTTTTTTACAGTT
CGAGATGCATTTTTTGATAAAAATGAGATGATGCAAGTCCTTTATTCTATAGATTTTTGAG
AAATACTATGATATAGGT---CTTGATAGTGCATTAATAAAGGGTAAAAAGTTTGA----
--CATGGGGGCTAAAGAGTAC---AGTTTAATGGGTTACTCTCAAACCAGCCATTCAA
AAACCAAACAGTTGTGGACAGGCAAACAAATTTTAAGTTTTATTTTTCCCAATGTCTTT
TACAAAATTCGTCCAATGAAAGTCGGGAAAATGACCAAGAAAATGTTTCTGATACGTCT
GTGGTAATATGCGGAGG

>HQ457435

AAAGGAGAAGATAACACAGAAGGAAAAGTCATTTTAAACGGAGAGAAAGTTCATAGTATT
TTCAAAAAGATTTTAGATGAAGACATTGAAGTTTTAGGTTTTGATCTCGCCTACAGTAGA
CCAGAATGGCTTATTTTAACCGTTTTGTTGGTGCCTCCACCGTCAGTAAGACCTTCTATT
GTCATGGAAGGGATGTTAAGAGCAGAGGATGATCTTACTCACAACTTTCTGATATTATT
AAATCTAATACTTACCTTAAGAAATACGAAATGGAAGGAGCACCGGGCATATTATTCGA
GATTATGAACAACCTGTTACAATTTTCATGTGGCCACTTTAATAGACAACGACATTAGTGGA
CAACCACAGGCCCTTCAGAAAAGTGGTAGGCCTTTGAAATCTATTTTCAGCTCGACTTAAA
GGAAAGGAAGGACGTGTTAGAGGAAATCTTATGGGTAAGAGAGTGGACATGTCTATGATG
GGGCACAAAGTTCGTGTGATGGCAGGGAAGACTTTCCGTTTAAATTTAAGTGTACTTCT
CCTTACAACGCCGATTTTGTGATGGAGATGAAATGAACCTTCATATGCCACAAAATTACAAT
TCTATTTCCGAGTTAGAAGAAATTTGTATGGTACCCAAACAAATTCTCGGACCTCAAAGT
AACAAACCCGTCATGGGAATCGTTCAAGATACTCACCAGTCTAAGATTTTTTTACAGTT
CGAGATGCATTTTTTGATAAAAATGAGATGATGCAAGTACTTTATTCTATAGATTTTCGAG
AAATACTATGATATTGGT---CTTGATAGTGCATTAATAAAGGGTAAAAAGTTGGA----
--CATGGGTGCTAAAGAATAT---AGTTTAATGGCTTTACTCTCAAACCAGCCATTCAA
AAACCAAACAGTTGTGGACAGGTAAGCAAATTTTAAGTTTTATTTTTCCCTAATGTCTTT
TACAAAATTCGTCTAATGAAAGTCGGGAAAATGATTTAGAAAATGTTTCGGTTACATCT
GTCGTAATATGCGGAGG

>F2A3F4A2-3

NNGGGAGAAGACAATTCAGAAGGAAAAGTCATCCTTAATGGTGAAAGAGTTCACAGTATT
TCAAGAAAATTTCTGATGAAGATATTGTTATGTTAGGATTTGATTTAGAATTTAGCAGG
CCTGAATGGCTTATTCTTACTGTTCTTTTGGTACCACCACCAGCAGTTAGACCTTCTATT
GTTATGGAAGGAATGTTAAGAGCTGAAGATGATCTAACACATAAACTTGCTGATATTATT
AAATCAAATACATATTTGAAGAAGTATGAGCTTGAAGGTGCTCCAGGACATATTATTAGA
GATTATGAACAGCTTTTACAATCCACATTGCAACTCTCATTGATAATGATATTAGTGA
CAACCACAGGCTTTGCAAAAAGCGGTAGACCGTTGAAATCGATCTCAGCAAGACTCAAG
GGGAAAGAAGGAAGAGTCAGAGGAAATCTTATGGGNNNNNNNNNNNNNNNATGTCTATGATG
GGCCATAGAGTACGTGTAATGTCAGGAAAACTTTTAGATTGAATTTGAGTGTAACATCT
CCATATAATGCTGATTTTGTATGGTATGAAATGAATTTACACATGCCACAGAGTTATAAT
ACTATTTCTGAACTTGAAGAAAATTTGTATGGTTTCTAACAGGTTACTAGGACCACAAAGC
AACAAACCTGTAATGGGAATTGTTCAAGATACTCTAACGGGCTTGAGATTTTTTACAATT
AGAGATTCATTCTTTGATAAATGTGAGATGATGCAAGTTTATATTCTATTGATTTAGAT
AAGTACAATGATATTGGC---TTAGATAGTATTACAAGAGAAGGTAAGAAATTGGA----
--TATTAACCAAAGAATAT---AATCTGTTGAAGCTCCTTAATACTCCAGCAATTCAA
AAGCCAAAACAGCTTTGGACAGGTAAACAAATTCTAAGTTTATTTTTCCGAATATTTTC
TATCAAAATGTTTCTAATGAAAATTTTGA AAAAGATATTGAAA ACTTTTCGGATACTCAA
GTAGTAATCTGTGAGGG

>F2ndF4A4

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
AGAGATTCTTTCTTTGATAAATGTGAGATGATGCAGGTTTATATTCTATTGATTTAGAT
AAGTACAATGATATTGGC---TTAGATAGTATTACAAGAGAAGGTAAGAAATTRGA----
--TATTAACCAAAGAATAT---AATCTCTTGAAGCTCCTCAATACTCCAGCAATTCAA
AAGCCAAAACAGCTTTGGACAGGCAAACAAATTCTAAGTTTATTTTTCCGAATATTTTC
TATCAAAATGTTTCCAATGAAAATTTTGA AAAAGACCTTGAAA ACTTTTCGGATACTCAA
GTAGTCATCTGTGNNNN

>F2ndF4A5

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
ACTATTTCTGAACTAGAAGAAATTTGTATGGTTTCTAACAGGTTACTAGGACCGCAAAGC
AACAAACCTGTAATGGGAATTGTTCAAGATACTTTAACGGGCTTGCGATTTTTTACAATT
AGAGATTCTTTCTTTGATAAATGTGAGATGATGCAGGTTTATATTCTATTGATTTAGAT
AAGTACAATGATATTGGC---TTAGATAGTATTACAAGAGAAGGTAAGAAATTAGA----
--TATTAACCAAAGAATAT---AATCTGTTGAAGCTCCTCAATACTCCAGCAATTCAA
AAGCCAAAACAGCTTTGGACAGGYAAACAAATTCTAAGTTTATTTTTCCGAATATTTTC
TATCAAAATGTTTCTAATGAAAATTTTGA AAAAGACCTTGAGAACTTTTCGGATACTCAA
GTAGTCATCTGNNNNNN

>F2B2F4B2

NNGGGAGAAGATAACTCGGAAGGAAAAGTTATTCTTAATGGTGAGAGAGTCCACAGTATA
TTTAAGAAGATTTCCGACGAAGATATTGAAATGTTAGGCTTTGATTTGGGATATAGTAGA
CCTGAATGGCTTATTCTAACAGTCTTATTGGTACCGCCCCCTGCAGTTCGACCTTCTATC
GTCATGGAAGGAATGCTTAGAGCTGAAGATGATCTAACTCACAACTAGCCGATATTATA
AAATCGAATACGTATCTAAAGAAGTATGAGCTTGAAGGAGCACCAGGTCATATCATTAGG
GATTATGAACAGCTTCTTCAGTTTCATATAGCAACTCTTATTGACAATGACATCAGTGGA
CAACCACAGGCCTTGCAAAAAGAGCGGAAGACCGTTAAAATCAATCTCAGCCAGACTCAAG
GGGAAGGAAGGAAGAGTTAGAGGTAATCTTATGGGNNNNNNNNNNNNNNNNATGTCTATGATG
GGCACAGAGTAAAAGTAATGTCAGGTAAGACTTTTAGACTGAATTTGAGTGTAACATCC
CCATACAATGCGGACTTTGATGGTGATGAAATGAATTTACATATGCCACAGAGTTATAAT
ACTATATCAGAACTTGAAGAAATTTGTATGGTCTCTAAGCAAGTTTLAGGCCACAAAGC
AACAAACCAGTTATGGGAATTGTCCAGGATACTCTTACTGGTCTAAGATTTTTTACTATT
AGAGATTCATTCTTTGATAAGAATGAAATGATGCAAGTTTTGTATTCCATCGACCTAGAC
AAGTATAATGATATTAAT---CTTGATTGTATCACCAGAGAAGGTAAGAAAATAAA----
--TATAAAGCCAAAAGAATAC---AATCTTATCAAACCTTCTTAATACACCAGCAATTCAA
AAGCCAAAACAGTTATGGACAGGTAACAAATTTTAAGCTTTATTTTCCCGAATGTCTTT
TACAAAGCTTCATCCAATGAACACTACGAAAATGATACGGAAAACCTTCTCCGATACACAA
GTAGTCATTTGTGAAGG

>F2B1F4B2

NNGGGAGAAGATAACTCGGAAGGAAAAGTTATTCTTAATGGTGAGAGAGTCCACAGTATA
TTTAAGAAGATTTCCGACGAAGATATTGAAATGTTAGGCTTTGATTTGGAATATAGTAGA
CCTGAATGGCTTATTCTAACAGTCTTATTGGTACCGCCCCCTGCAGTTCGACCTTCTATC
GTCATGGAAGGAATGCTTAGAGCTGAAGATGATCTAACTCACAACTAGCCGATATTATA
AAATCGAATACGTATCTAAAGAAGTATGAGCTTGAAGGAGCACCAGGTCATATCATTAGG
GATTATGAACAGCTTCTTCAGTTTCATATAGCAACTCTTATTGACAATGACATCAGTGGA
CAACCACAGGCCTTGCAAAAAGAGCGGAAGACCGTTAAAATCAATCTCAGCCAGACTCAAG
GGGAAGGAAGGAAGAGTTAGAGGTAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNN
GGGCACAGAGTAAAAGTAATGTCAGGTAAGACTTTTAGACTGAATTTGAGTGTAACATCC
CCATACAATGCGGACTTTGATGGTGATGAAATGAATTTACATATGCCACAGAGTTATAAT
ACTATATCAGAACTTGAAGAAATTTGTATGGTCTCTAAGCAAGTTTLAGGCCACAAAGC
AACAAACCAGTTATGGGAATTGTCCAGGATACTCTTACTGGTCTAAGATTTTTTACTATT
AGAGATTCATTCTTTGATAAGAATGAAATGATGCAAGTTTTGTATTCCATCGACCTAGAC
AAGTATAATGATATTAAT---CTTGATTGTATCACCAGAGAAGGTAAGAAAATAAA----
--TATAAAGCCAAAAGAATAC---AATCTTATCAAACCTTCTTAATACACCAGCAATTCAA
AAGCCAAAACAGTTATGGACAGGTAACAAATTTTAAGCTTTATTTTCCCGAATGTCTTT
TACAAAGCTTCATCCAATGAACACTACGAAAATGATACGGAAAACCTTCTCCGATACACAA
GTAGTCATTTGTGAAGG

>F2B3F4B3

NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAGTTATTCTTAATGGTGAGAGAGTCCACAGTATA
TTTAAGAAGATTTCCGACGAAGATATTGAAATGTTAGGCTTTGATTTGGAATATAGTAGA
CCTGAATGGCTTATTCTAACAGTCTTATTGGTACCGCCCCCTGCAGTTCGACCTTCTATC
GTCATGGAAGGAATGCTTAGAGCTGAAGATGATCTAACTCACAACTAGCCGATATTATA
AAATCGAATACGTATCTTAAGAAGTATGAGCTTGAAGGAGCACCAGGTCATATCATTAGG
GATTATGAACAGCTTCTTCAGTTTCATATAGCAACTTATTGACAATGACATCAGTGGA
CAACCACAGGCCTTGCAAAAAGAGCGGAAGACCGTTAAAATCAATCTCAGCCAGACTCAAG
GGGAAGGAAGGAAGAGTTAGAGGTAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNN
NN
NN
NN
ACTATATCAGAACTTGAAGAAATTTGTATGGTCTCTAAGCAAGTTTLAGGCCACAAAGC
AACAAACCAGTCATGGGAATTGTTCAAGATACTCTTACTGGTCTAAGATTTTTTACTATT
AGAGATTCATTCTTTGATAAGAATGAAATGATGCAAGTTTTGTATTCCATCGACCTAGAC
AAGTATAATGATATTAAT---CTTGATTGTATCACCAGAGAAGGTAAGAAAATAAA----
--TATAAAGCCAAAGGAATAC---AATCTTATCAAACCTTCTTAATACACCAGCAATTCAA
AAGCCAAAACAGTTATGGACAGGTAACAAATTTTAAGCTTTATTTTCCAAATGTCTTT
TACAAAGCTTCATCCAATGAACACTACGAAAATGATACGGAAAACCTTCTCCGATACACAA
GTAGTCATTTGTGAAGG

>F2C1F4C1_DQ996233

AAAGGAGAAGACAACCTCAGAGGGGAAGGTGATTCTGAATGGAGAAAGGGTCCACAGTATT
TTCAAGAAGATATCTGATGAGGATATTCTTAGGCTTGGGTTTGATTTAGAATTTAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GAAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCTAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCACCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTGATAGACAATGATATTTCTGGA
CAACCCCAAGCTCTACAAAAGAGTGGGAGACCTCTGAAATCCATCTCAGCTAGATTGAAA
GGGAAAGAAGGGAGAGTAAGAGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNATGTCTATGATG
GGCCACAGAGTCAAAGTAATGWCTGGGAAGACTTCTAGATTAAATCTCAGTGTGACATCT
CCTTACAATGCAGATTTTCGATGGAGATGAAATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATCTCCGAACTAGAAGAAATTTGTATGGTTTCTAAGCAGGTCCCTAGGGCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGATACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
AAGCCCAAAGAGCTCTGGACAGGCAAGCAGATACTAAGCTTCATCTTCCCCAGTGTATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTGGTAATTATAGGAGG

>F2C1bF4C1

NNNGGAGAAAGGGTCCACAGTATT
TTCAAGAAGATYTTCTGATGAGGATATTCTTAGGCTTGGGTTTGATTTAGAATTTAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCTAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCMCCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTGATAGACAATGATATTTCTGGA
CAACCCCAAGCTCTACAAAAGAGTGGGAGACCTCTGAAATCCATCTCAGCTAGATTGAAA
GGGAAAGAAGGGAGAGTAAGAGGGAATCTTATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCTGGGAAGACTTCTAGATTAAATCTCAGTGTGACATCT
CCTTACAATGCAGATTTTCGATGGAGATGAAATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATCTCCGAACTAGAAGAAATTTGTATGGTTTCTAAGCAGGTCCCTAGGGCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGATACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
AAGCCCAAAGAGCTCTGGACAGGCAAGCAGATACTAAGCTTCATCTTCCCCAGTGTATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTGGTAATTATAGGAGG

>F2C1bC4F4C1-4

NNNGGAGAAAGGGTCCACAGTATT
TTCAAGAAGATTTCTGAYGAGGATATTCTTAGGCTTGGGTTTGATTTTGAATTTAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCYAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCMCCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTRATAGACAATGATATTTCTGGA
CAACCMCAAGCTCTACARAAGAGTGGGAGACCTCTGAAATCYATCTCAGCTAGATTGAAA
GGRAAAGAAGGGAGAGTAAGAGGRAATCTTATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNCTGGGAAGACTTCTAGATTAAATCTCAGTGTGACATCT
CCTTACAATGCAGATTTTCGATGGAGATGAAATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATCTCCGAACTAGAAGAAATYTGATGGTTTCTAAGCAGGTCCCTAGGRCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTSCAAGAYACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
AAGCCCAAAGAGCTCTGGACAGGCAAGCAGATACTAAGCTTCATCTTCCCCAGTGTATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTRGTAATTATAGGAGG

>F2C4F4C4

NNGGAGAAAGGGTCCACAGTATT
TTCAAGAAGATTTCTGACGAGGATATTCTTAGGCTTGGGTTTGATTTGGAATTTAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCCAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCACCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTAATAGACAATGATATTTCTGGA
CAACCACAAGCTCTACAGAAGAGTGGGAGACCTCTGAAATCTATCTCAGCTAGATTGAAA
GGAAAAGAAGGGAGAGTAAGAGGAAATCTTATGNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATCTCCGAAGTAGAAGAAATCTGTATGGTTTCTAAGCAGGTCCTAGGACCCCAGAGC
AATAAGCCAGTGATGGGGATAGTGCAAGACACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
ARGCCCAAAGAGCTCTGGACAGGCAAGCAGATACTAAGCTTCATCTTCCCAGTGATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTAGTAATTATAGGAGG

>F2C6F4C5

NN
TTCAAGAAGATTTCTGACGAGGACATTCTTAGGCTTGGGTTTGATTTGGAATTCAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCCAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCACCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTRATAGACAATGATATTTCTGGA
CAACCCCAAGCTCTACAAAAGAGTGGGAGACCTTTGAAATCCATCTCAGCCAGATTGAAA
GGTAAAGAAGGGAGAGTAAGAGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNNNG
GGCCACAGAGTCAAAGTAATGTCTGGGAAGACCTTTAGATTAAATCTCAGTGTGACATCT
CCTTACAATGCAGATTTTCGATGGAGACGAAATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATCTCCGAAGTAGAAGAAATTTGTATGGTTTCTAAGCAGGTCCTAGGGCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGATACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
AAGCCCAAAGAGCTCTGGACAGGCAAGCAGATCCTAAGCTTCATCTTCCCAGTGATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTAGTAATTATAGGAGG

>F2C5F4C5

NN
TTCAAGAAGATTTCTGACGAGGACATTCTTAGGCTTGGGTTTGATTTGGAATTCAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCCAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCACCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTAATAGACAATGATATTTCTGGA
CAACCCCAAGCTCTACAAAAGAGTGGGAGACCTTTGAAATCCATCTCAGCCAGATTGAAA
GGTAAAGAAGGGAGAGTAAGAGGGAATCTTATGGNNNNNNNNNNNNNNNNNNNNATGTCTATGATG
GGCCACAGAGTCAAAGTAATGTCTGGGAAGACCTTTAGATTAAATCTCAGTGTGACATCT
CCTTACAATGCAGATTTTCGATGGAGACGAAATGAATCTCCACATGCCCCAGAGCTACAAC
ACSATCTCCGAAGTAGAAGAAATTTGTATGGTTTCTAAGCAGGTCCTAGGGCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGATACTCTCACAGGGTTGAGATTCTTTACGATT
CGAGATGCTTTCTTTGATAAGAGAGAGATGATGCAAGTTCTGTACTCCATAGAGCAGAAT
ATAAATA-----GAAGACAGGACAGT-----
-----AGCAATTAT---GGGATTCTAGATCTTCTGAATAGACCAGCCATCCAG
AAGCCCAAAGAGCTCTGGACAGGCAAGCAGATCCTAAGCTTCATCTTCCCAGTGATTC
TACCAGACACAGTCTGAGGA-----TCCTGAGAATCTCTCGGACTCTGCA
GTAGTAATTATAGGAGG

>F2D1F4D1

NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCAAACCTCGGGTTTGGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTTACAGTTCTACTTGTGCCACCCCCAGCAGTCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACWTATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCACAAGCTCTTCAAGAAGAGTGGGAGACCCTGAAGTCGATCTCAGCMAGACTGAAA
GGCAAGGAAGGCAGAGTCMGAGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGATGAATCTCCACATGCCCCAGAGCTACAAY
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTCACTGGTCTGAGATTCTTTACTATC
CGAGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAW
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTAGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2D4F4D4

NNNNNNNNNNNNNNNNNNNNNNNNNNNGGGAAGGTGATTCTCAATGGAGAAAGAGTCCACAGTATT
TTCAAGAAGATTTTCGGATGAAGATATTCTCAAACCTCGGGTTTGGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTTACAGTTCTACTTGTGCCACCCCCAGCAGTCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACATATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCMCAAGCTCTTCAAGAAGAGTGGGAGACCCTGAAGTCGATCTCAGCMAGACTGAAA
GGCAAGGAAGGCAGAGTCMGAGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGATGAATCTCCACATGCCCCAGAGCTACAAY
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTCACTGGTCTGAGATTCTTTACTATC
CGRGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAW
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCCTAGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2D2F4D2

NNNNNNNNNNNNNNNNNNNNNNNNNNNGGGAAGGTGATTCTCAATGGAGAAAGAGTCCACAGTATT
TTCAAGAAGATTTTCGGATGAAGATATTCTCAAACCTCGGGTTTGGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTTACAGTTCTACTTGTGCCACCCCCAGCAGTCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACWTATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCMCAAGCTCTTCAAGAAGAGTGGGAGACCWCTGAAGTCGATCTCAGCAAGACTGAAA
GGCAAGGAAGGCAGAGTCAGAGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTCACTGGTCTGAGATTCTTTACTATC
CGRGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAW
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTAGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2D3F4D3

NNNNNNNNNNNNNNNNNNNNNNNNNNNGGGAAGGTGATTCTCAATGGAGAAAGAGTCCACAGTATT
TTCAAGAAGATTTCCGGATGAAGATATTCTCAAAGCTCGGGTTTGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTTACAGTTCTACTTGTGCCACCCCAAGAGTCCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACATATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCACAAGCTCTCAGAAGAGTGGGAGACCWCTGAAGTCGATCTCAGCAAGACTGAAA
GGCAAGGAAGGCAGAGTTCMAGGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN
NNNGATGAATCTCCACATGCCCAAGAGCTACAAY
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACTCTCACTGGTCTGAGATTCTTTACTATC
CGAGATGCCTTCTTTGATAAGAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTRGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTC
TACCAGACCAWTCTGAGRA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2D5F4D6

NN
TTCAAGAAGATTTCCGGATGAAGATATTCTCAAAGCTCGGGTTTGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTCACAGTTCTACTTGTGCCACCCCAAGAGTCCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACATATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCACAAGCTCTCAGAAGAGTGGGAGACCWCTGAAGTCGATCTCAGCCAGACTGAAA
GGAAAGGAAGGCAGAGTCCAGAGGGAATCTTATGGGTAAAGAGAGTAGAANNNGTCGATGATG
GGCCACAGAGTGAAAGTGAATGTCAGGGGAAAGACYTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCAAGAGCTACAAY
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTTACTGGTCTGAGATTCTTTACTATC
CGAGATGCCTTCTTYGATAAGAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTRGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTC
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2ndF4D8

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
GGCCACAGAGTGAAAGTGAATGTCAGGGGAAAGACYTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCAAGAGCTACAAY
ACCATTTCTGAATTRGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTTACTGGTCTGAGATTCTTTACTATC
CGRGATGCCTTCTTYGATAAGAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCTTGGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTC
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATTGGTGG

>F2D6F4D7

NNGATTCTCAATGGWGAAAGAGTCCACAGTATT
TTCAAGAAGATTTTCGGATGAAGATATTCTCAAACCTCGGGTTTGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTCACAGTTCTACTTGCGCCACCCCCAGCAGTCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACATATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCACAAGCTCTTCAGAAGAGTGGGAGACCTCTGAAGTCGATCTCAGCCAGACTGAAA
GGMAAGGAAGGCAGAGTCAGAGGGAATCTTATGGGTAAGAGAGTAGAAATGTCGATGATG
GGCCACAGAGTGAAAGTGTATGTCAGGGAAGACYTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCCAGAGCTACAAY
ACCATTTCTGAATTRGAAGAGATCTGCATGGTCTCCAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACWCTTACTGGTCTGAGATTCTTTACTATC
CGGGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCTTGGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2D7F4D9

NNGATTCTCAATGGTGAAAGAGTCCACAGTATT
TTCAAGAAGATTTTCGGATGAAGATATTCTCAAACCTCGGGTTTGATTTGGAATTCAGTAGA
CCAGAGTGGTTGATCCTCACAGTTCTACTTGCGCCACCCCCAGCAGTCAGACCATCCATA
GTCATGGAAGGGATGCTCAGGGCAGAGGATGATCTCACCCACAAACTGGCAGATATCATC
AAAGCCAATACATATCTCAAGAAGTACGAACTAGAAGGAGCCCCAGGCCATATAATCCGA
GACTATGAGCAATTACTCCAGTTCCACATAGCCACTCTCATAGACAATGATATCTCAGGA
CAACCMCAAGCTCTTCARAAGAGTGGGAGACWCTGAAGTCGATCTCAGCCAGACTGAAA
GGMAAGGAAGGCAGAGTCAGAGGGAATCTTATGGGTAAGAGAGTAGAAATGTCGATGATG
GGCCACAGAGTGAAAGTGTATGTCAGGGAAGACYTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACACTCACTGGTCTGAGATTCTTTACTATC
CGAGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCCTAGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2ndF4D10

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
GGCCACAGAGTGAAAGTGTATGTCAGGGAAGACTTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATTTCTGAATTRGAAGAGATCTGCATGGTCTCCAACAGGTCTTGGGTCCCCAGAGC
AATAAGCCAGTGATGGGGATAGTCCAAGACACACTTACTGGTCTGAGATTCTTTACTATC
CGRGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTRGATCTTCTTAGTAGACCTGCCATTCAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCCAGTGTTTTT
TACCAGACCCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>F2C6F4D5

NNNNNNNNNNNNNNNNNNNNNNAGGGGAAGGTGATTCTGAATGGAGAAAGGGTCCACAGTATT
TTCAAGAAGATTTCTGACGAGGACATTCTTAGGCTTGGGTTTGATTTGGAATTCAGTAGA
CCAGAATGGTTGATTCTTACAGTGCTACTGGTACCACCCCCAGCTGTGAGACCTTCTATA
GTAATGGAAGGGATGCTCAGGGCAGAAGATGATCTTACCCACAACTAGCAGATATTATC
AAAGCCAATACCTATCTCAAGAAGTACGAACTAGAAGGAGCACCAGGCCACATTATCAGA
GATTACGAACAATTACTCCAATTCCACATAGCTACTCTRATAGACAATGATATTTCTGGA
CAACCCCAAGCTCTACAAAAGAGTGGGAGACCTTTGAAATCCATCTCAGCCAGATTGAAA
GGTAAAGAAGGGAGAGTAAGAGGGGAATCTTATGGGNNNNNNNNNNNNNNNNNNNNNNNG
GGCCACAGAGTGAAAGTGTGTCAGGGAAGACTTTCAGACTGAATCTCAGTGTGACTTCT
CCGTACAATGCTGATTTTCGATGGGGATGAGATGAATCTCCACATGCCCCAGAGCTACAAC
ACCATTTCTGAATTAGAAGAGATCTGCATGGTCTCCAAACAGGTCTTGGGTCCCCAGAGC
ATAAGCCAGTGATGGGGATAGTCCAAGACACWCTTACTGGTCTGAGATTTCTTACTATC
CGRGATGCCTTCTTTGATAAGAGAGAGATGATGCAAGTCTTGTACTCAATAGACCAGAAT
ATCAATA-----GGAGACAGGACAGT-----
-----AGCAACTAT---GGGATCYTRGATCTTCTTAGTAGACCTGCCATTGAG
AAGCCCAAAGAGCTTTGGACTGGGAAACAGATTCTCAGCTTCATCTTCCCAGTGTTTTT
TACCAGACCAGTCTGAGGA---TCTCTTGGACTCTTCTGAGAATCTCTCGGACTCTTCA
GTGGTGATCATCGGTGG

>MF344629

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NNNNNNNNNNNNNNNNNNNNNNNGAGATATGCATGGTATCTAAGCAAGTACTGGGACCCCAAAGC
AATAAGCCTGTAATGGGAATTGTTCAAGATACTCTTACAGGCCTGAGATTCTTCACTATC
AGAGATTCCTTCTTTGATAAAAATGAGATGATGCAGGTTTTGTACTCCATCGACTTAGAC
AATAACAATGATATAAAT---CTCGACAGTATCACCAGAGAAGGTAAGAAAATAA----
--TATCAAGCCAAAGGAATAC---AGCCTTATCAAAGTCTCAATACACCAGCAATATCA
AAGCCTAAACAGTTATGGACTGGTAAACAAATCTTAAGCTTATTTTCCCAGATGTTTTT
TACAAAGCGTCATCTAATGAAAATTTGAGAATGACACAGAGAATTTTCTGATACTCAG
GTAGTCATCTGTGAGGG

>F2ndF4F1

NN
NN
NN
NN
NN
NN
NN
NN
NN
NN
NNNNNCAGGGTCAAAGTTATGGGAGGAAAACCTTTTAGATTGAATCTAAGTGTGACATCT
CCTTATAACGCGGATTTTCGATGGAGATGAAATGAATCTACACATGCCTCAGAGTTACAAT
TCTATTTCTGAACTTGAAGAGATCTGTATGGTATCTAAACAAATCCTTGGACCTCAAAGC
AATAAACAGTAATGGGAATTGTTCAAGATACTCTTACTGGTTTTAAGATTTTACTATT
AGAGATTCTTCTTTGATAAGAATGAAATGATGCAAGTTCTATATTCTATTGATTTAGAT
AATAACAATGATATAAAT---CTTACTGCATCACGAGGGAAAGGTAAGAAATTGAA----
--CATAAGGCCAAAGGAATAC---AATCTTGTAAAACCTTCTAATACTCCAGCAATCCAA
AAACCAAACAGCTTTGGACAGGCAAACAGATTCTAAGCTTATCTTCCGATGTTTTT
TACAAGGCTGCATCTAATGAAAACCTTGTAGAATGATACCGAGAATTTTCTGATACACAA
GTAGTCATCTGTGAGGG

>XM_014708712_Ordospora_colligata

NNNGGAGAGGAAAGCGGAGAGGGTAAGGTGATGCTGAATGGCGAGCGTGTATACAGCATA
TTCAAGAAAATAAGCGATGCAGACTGTGAACACATGGGATTTGACCTGAAGTACAGTCGA
CCGGAATGGATGATTCTTACTGTGCTTCTTGTTCCACCACGGTGAGGCCATCTATT
GTGATGGAAGGCTCGCTGCGTGGAGAGGATGACCTAACACACAAGCTTGCAGACATAATC
AAGTCTAATGGATACCTGAAGAAGTATGAACAAGAAGGAGCACCTGGGCATATATTGCGT
GATTACGAGCAGCTGCTGCAGTTTCATGTTGCCACGCTGATTGACAATGACATTGGAGGG
CAACCACAAGCGCTGCAAAAAGAGTGGAAGACCGTTGAAATCAATATCAGCGAGGCTGAAA
GGAAAAGAAGGGCGGATAAGAGGAAACCTGATGGGTAAAAGAGTGGACATGTCTATGATG
GCACACTATGTAAGGGTTATGGATAACAAGACGTTTCAGGCTGAACTTGAGTGTTACATCT
CCATATAATGCAGACTTTGATGGAGACGAGATGAATTTGCACATGCCACAGAGTTATACG
TCGAAGTCTGAGCTTGAAGAATTGGTTCTTGTGCCGAACCAGATTATATCGCCGCAGTCT
AACAAACCTGTGATGGGGATAGTCCAAGACACACTTACAGGGCTCAGACTCCTTACGCTT
AGAAACACATTTCTAGACGAACGAGAGACAATGTCGCTACTTTATGCGGTGAATTC----
-----TCAGTTTGATGATATGCCAGTCGGGGAGAATGCAAGAGCAGAG
TTCAAGAGGACTAAGGAGTAT---GACATCATGAAGGTTCTAGCAAGACCAAGTATAGCA
AAACCAAAGAAGCTGTGGACAGGAAAGCAGGTGCTGAGTTTTATTCTACCATGCCTGAAT
TATTCTGGATTCTCATCTGAGCATAATGATGAAGATAAAGAAAATGCAAATGACACGGTT
GTGCGAATTCAGGATGG

>57.1-MOUG-8-Gfos_SSU

CCGGAGAAGGAGCCTGAGAGATTGCTACTAAGTCTAAGGATTGCAGCAGGGGCGAAACTTGACCTAT
GATATGATATTGAGGCAGTTATGAGTAGTATTTTTAATTATTGTAGTATTGTAAGTATATACTACAAGA
TAAATCGGAGGGCAAATCGAGTGCC

>57.1-MOUG-10-Gfos_SSU

ACAAGACTATGACGGATAACGGTATTACTTTGTAATATTCCGGAGAAGGAGCCTGAGAGATTGCTACT
AAGTCTAAGGATTGCAGCAGGGGCGAAACTTGACCTATGATATGATATTGAGGCAGTTATGAGTAGTA
TTTTTAATTATTGTAGTATTGTAAGTATATACTACAAGATAAATCGGAGGGCAAATCGAGTGCC

>ROS76_Gdue_SSU

GTTAGTTGTTAAGGTAATGGCTTAACAAGACTATGACGGATAACGGTATTACTTTGTAATATTCCGGAG
AAGGAGCCTGAGAGATTGCTACTAAGTCTAAGGATTGCAGCAGGGGCGAAACTTGACCTATGATATGA
TATTGAGGCAGTTATGAGTAGTATTTTT

Results of ASAP analysis:

Nb of subsets	asap-score	P-val (rank)	W (rank)	Threshold dist.
11	2.50	6.54e-04 (1)	2.58e-03 (4)	0.112635
11	4.00	1.41e-03 (2)	2.46e-03 (6)	0.186576
* 17	4.00	4.71e-01 (7)	3.91e-03 (1)	0.030950
* 15	5.00	7.93e-03 (3)	2.10e-03 (7)	0.040951
15	6.50	2.50e-01 (5)	1.72e-03 (8)	0.053581
11	8.00	8.88e-01 (11)	2.50e-03 (5)	0.223591
* 16	8.50	9.40e-01 (14)	2.76e-03 (3)	0.035194
* 20	10.00	3.97e-01 (6)	9.24e-04 (14)	0.007322
13	10.50	9.00e-01 (12)	1.68e-03 (9)	0.064135
* 19	11.00	8.30e-01 (9)	9.83e-04 (13)	0.009095

Partition 1

Asap-Score: 2.500000

Proba: 1.399791e-03

Nb subsets with recursion: 11 (without recursion: 10)

W (rank) 2.87e-03 (4)

Threshold dist. 0.112099

Subset[1] n: 2 ;id: AF060234 DQ996236

Subset[2] n: 1 ;id: XM_002995356

Subset[3] n: 1 ;id: DQ996230

Subset[4] n: 5 ;id: AJ278948 DQ996231 DQ996234 HQ457438 HQ457435

Subset[5] n: 3 ;id: F2A3F4A2-3 F2A2-A3F4A2 F2A2F4A2

Subset[6] n: 4 ;id: F2B1F4B1 F2B1F4B2 F2B2F4B2 F2B3F4B3

Subset[7] n: 1 ;id: DQ996235

Subset[8] n: 3 ;id: F2E1F4E1 F2E1F4E4 F2E1F4E5

Subset[9] n: 1 ;id: DQ996232

Subset[10] n: 15 ;id: F2C1F4C1_DQ996233 F2C1bF4C1 F2C1bC4F4C1-4 F2C4F4C4 F2C6F4C5 F2C5F4C5

F2C3F4C3 F2D1F4D1 F2D4F4D4 F2D2F4D2 F2D3F4D3 F2D5F4D6 F2D6F4D7 F2D7F4D9 F2C6F4D5

Subset[11] n: 1 ;id: XM_014708712_Ordospora_colligata

Results of bPTP analysis

Acceptance rate: 0.09894400000000000004

Merge: 249995

Split: 250005

Estimated number of species is between 19 and 29

Mean: 23.85

Most supported partition found by simple heuristic search

Species 1 (support = 1.000)

XM_014708712_Ordospora_colligata

Species 2 (support = 1.000)

DQ996230

Species 3 (support = 1.000)

DQ996232

Species 4 (support = 1.000)

MF344629

Species 5 (support = 1.000)
F2ndF4F1

Species 6 (support = 0.994)
AF060234,DQ996236

Species 7 (support = 1.000)
XM_002995356

Species 8 (support = 1.000)
DQ996235

Species 9 (support = 0.954)
F2E1F4E1,F2E2F4nd,F2E3F4nd,F2E1F4E4,F2ndF4E2,F2E1F4E5,F2ndF4E3

Species 10 (support = 1.000)
HQ457438

Species 11 (support = 1.000)
HQ457435

Species 12 (support = 1.000)
DQ996234

Species 13 (support = 1.000)
DQ996231

Species 14 (support = 1.000)
AJ278948

Species 15 (support = 1.000)
F2A6F4nd

Species 16 (support = 0.994)
F2B4F4B3

Species 17 (support = 0.968)
F2B1F4B1,F2B2F4B2,F2B1F4B2,F2B3F4B3

Species 18 (support = 0.909)
F2D1F4D1,F2D4F4D4,F2D2F4D2,F2D3F4D3,F2D7F4D9,F2D5F4D6,F2ndF4D8,F2D6F4D7,F2ndF4D10

Species 19 (support = 0.995)
F2C6F4D5

Species 20 (support = 0.658)
F2C6F4C5,F2C5F4C5,F2ndF4C2

Species 21 (support = 0.608)
F2C1F4C1_DQ996233,F2C1bF4C1,F2C1bC4F4C1-4,F2ndF4C6_JX213747

Species 22 (support = 0.649)
F2C4F4C4

Species 23 (support = 0.649)
F2C3F4C3

Species 24 (support = 0.500)

F2ndF4A1,F2ndF4A5,F2ndF4A4

Species 25 (support = 0.514)

F2A3F4A2-3,F2A2-A3F4A2,F2A2F4A2