The complete mitochondrial genome of Gammarus roeselii (Crustacea, Amphipoda): insights into mitogenome plasticity and evolution

Alexandre Cormier, Rémi Wattier, Maria Teixeira, Thierry Rigaud, Richard Cordaux

The complete mitochondrial genome of *Gammarus roeselii* (Crustacea, Amphipoda): insights into mitogenome plasticity and evolution

Alexandre Cormier[1], Rémi Wattier[2], Maria Teixeira[2], Thierry Rigaud[2] and Richard Cordaux[1]

[1] Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Université de Poitiers, UMR CNRS 7267, Bât. B8, 5 rue Albert Turpin, TSA 51106, 86073 Poitiers Cedex 9, France

[2] Laboratoire Biogéosciences, Université Bourgogne Franche-Comté, UMR CNRS 6282, 6 boulevard Gabriel, 21000 Dijon, France

Corresponding author: richard.cordaux@univ-poitiers.fr

Abstract

This study presents the complete mitochondrial (mt) genome sequence, annotation and analysis of the amphipod *Gammarus roeselii*, a crustacean species widespread in European continental freshwaters. The circular mt genome has a total length of 16,073 bp and possesses the 37 canonical mt genes of bilaterians. Particularly noticeable is an unusual case of duplication of the full control region (CR). This duplication was confirmed experimentally in *G. roeselii* individuals caught in the wild, as it was found in all tested individuals from two distinct populations. Furthermore, comparing multiple mt haplotypes from closely related individuals from the western-most part of the distribution range of *G. roeselii* allowed us to identify single nucleotide polymorphisms and indels that may constitute valuable markers for phylogeographic analyses of *G. roeselii*. Finally, we performed a phylogenetic analysis that helped understanding the evolutionary dynamics of the CR.

Keywords

Control Region, Duplication, phylogeny, polymorphism

Introduction

The gammarid *Gammarus roeselii* (Amphipoda) is a freshwater peracarid crustacean (Väinölä et al., 2008). As the related morphospecies *Gammarus pulex* and *Gammarus fossarum*, *G. roeselii* is widespread in European continental freshwaters (Karaman & Pinkster, 1977), playing a key functional role in the ecosystem, through leaf litter recycling (Piscart et al., 2007). In addition, *G. roeselii* is often used in multi-topic research at the crossroads of ecophysiology, ecotoxicology, biological invasions, behavioral ecology and evolutionary parasitology (Haine et al., 2005; Moret et al., 2007; Tain et al., 2007; Sornom et al., 2010; Gerhardt et al., 2011; Böttger et al., 2012; Gismondi et al., 2012b, 2012c, 2012a, 2013; Gergs et al., 2013; Andreï et al., 2016).

Although *G. roeselii* is often considered as a native species all over Europe, Jażdżewski and Roux (1988) proposed that the Balkans should be considered as the native area of this species, as expansion outside this area is only recent, being associated with post-glacial expansion and/or even more recent human-mediated invasion. Recent invasion is likely as *G. roeselii* geographic distribution has been recorded to increase over the last 50 years (Grabowski, 2007; Paganelli et al., 2015). Consistently, based on mitochondrial (mt) *cytochrome c oxidase subunit I* (COI) DNA barcodes, it has been recently established that the morphospecies *G. roeselii* in the Balkans is characterized by extensive cryptic diversity, with 13 molecular operational taxonomic units (MOTUs) which diversification occurred from Miocene to Pleistocene epochs (Grabowski et al., 2017). By contrast, *G. roeselii* outside the Balkans is associated with a single COI-MOTU also present in the Northern Balkans (M. Grabowski and R. Wattier, unpublished results, see also Moret et al. 2007). However, deeper

exploration of *G. roeselii* expansion dynamics would benefit from the availability of additional mitochondrial markers.

Omics resources in amphipods are still relatively scarce or, at least, have been building up only recently (Kao et al., 2016), especially for the genus *Gammarus* (Krebes & Bastrop, 2012; Gismondi & Thomé, 2016; Trapp et al., 2016; Macher et al., 2017a). Amphipod nuclear genomes tend to be quite large, which constitutes a limiting factor for assembly processes (Kao et al., 2016). By contrast, the small and circular mt genome is more easily accessible. As a consequence, more than 20 amphipod mt genomes have been sequenced, although many are incomplete (Bauzà-Ribot et al., 2009, 2012; Ito et al., 2010; Ki et al., 2010; Kilpert & Podsiadlowski, 2010; Krebes & Bastrop, 2012; Shin et al., 2012; Pons et al., 2014; Romanova et al., 2014, 2015, 2016; Stokkan et al., 2015; Aunins et al., 2016; Lan et al., 2016; Macher et al., 2017a).

The mt genome of amphipods is a canonical bilaterian mitogenome, defined as a single circular molecule ranging in size from 14 to 18 kb, with a mean size around 16 kb (Boore, 1999). The gene content generally is highly conserved, with 37 genes encoding for 13 protein-coding genes, two rRNA genes, 22 tRNA genes and one control region (CR) containing regulatory elements for transcription and replication (Wolstenholme, 1992). In amphipods, mitogenomic features such as alterations in gene order, strand bias reversion, presence of additional tRNA genes and CR duplication have been reported and could be used to explore amphipod evolutionary history at various taxonomic levels (Bauzà-Ribot et al., 2009; Ito et al., 2010; Ki et al., 2010; Kilpert & Podsiadlowski, 2010; Krebes & Bastrop, 2012; Pons et al., 2014; Romanova et al., 2014; Aunins et al., 2016).

In this study, we present the complete mt genome sequence, annotation and analysis of the amphipod *G. roeselii*, obtained from individuals collected in France, i.e. in the west-most area of the species geographic range. A mitogenome of a *G. roeselii* individual collected in

Germany has recently been reported (Macher et al., 2017b) but it appears to be a partial sequence (see Results section below). Moreover, this mitogenome was produced as part of a methodological study and no annotation or analysis has been performed (Macher et al., 2017b). Our aims were: (i) to gain insights into mt genome evolutionary dynamics in *G. roeselii* with a focus on CR organization, and (ii) to perform a phylogenetic reconstruction testing for the relevance of mt coding genes to investigate relationships in the genus *Gammarus* and the family Gammaridae, also helping understanding the evolutionary dynamics of the CR.

Materials and Methods

Genome sequencing, assembly and annotation

The mt genome of *G. roeselii* was identified from data generated as part of an ongoing sequencing project in our laboratories. Briefly, total genomic DNA was extracted from two individuals (Ou3 and Ou53) sampled in Trouhans (Burgundy, France) in the Ouche river. For each individual, a paired-end library with ~500 bp inserts was prepared and sequenced on Illumina MiSeq (2x300 bp) and HiSeq2500 instruments (2x125 bp). Data quality was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and raw reads were trimmed and filtered using Trimmomatic v0.33 (Bolger et al., 2014). *De novo* assembly of the pooled reads was performed using SOAPdenovo2 with a k-mer size of 67 (Luo et al., 2012). Identification of mt sequences was performed by searching for sequence similarity between scaffolds and the available mt genome of *Gammarus duebeni* (NC_017760) (Krebes & Bastrop, 2012) using Blast (Altschul et al., 1990). This analysis recovered scaffolds which were subsequently used as seeds for independent assembly of Ou3 and Ou53 mt genomes.

Ou3 and Ou53 mt genomes were assembled using NOVOPlasty v2.6.3, with a k-mer size of 55 and previously identified scaffolds (Dierckxsens et al., 2017). Structural annotation was performed using the MITOS2 WebServer (Bernt et al., 2013). Reads were mapped against their respective assemblies using Bowtie2 v2.2.9 (Langmead & Salzberg, 2012) to identify and correct potential errors and to estimate genome sequencing depth. The sequencing depth for Ou3 and Ou53 was ~11,000x and ~18,000x, respectively. Reverse mapping was also used to identify structural variation between the two mt genomes. Because the complete mt genomes from both individuals were highly similar (see Results section

below), a single sequence was submitted to public databases. Accordingly, the complete mt genome sequence of Ou3 is available at DDBJ/ENA/GenBank under accession number MG779536.

To compare CR structure among Gammaridae, mt genomes listed in **Table 1** were manually reannotated using the same criteria (i.e. CR starting with a poly-T stretch and ending with the start/end of the next annotated gene). Repeat proportion in CRs was calculated using RepeatMasker 4.0.0 (developed by A.F.A. Smit, R. Hubley, and P. Green; see http://www.repeatmasker.org/) in slow search mode.

PCR validation and sequencing of duplicated control region

To experimentally test the presence of the CR duplication, PCR primers were designed as shown in **Figure 1** (Gr_mt2R: 5'-GCTTAAGCCGAGAATCATATATGTCA-3' and Gr_mt3F: 5'-TAGGCTAGCGCTGTTTCAGG-3'). This diagnostic PCR assay is expected to yield a single product of 371 bp if there are two copies of the CR, or no product if there is a single copy of the CR (**Figure 1**). DNA was extracted from 23 *G. roeselii* individuals sampled in the Ouche river at Trouhans and 15 individuals sampled in the Morthe river at Citey (Burgondy, France), using a EZ-10 96 Well Plate Genomic DNA Isolation kit (Biobasic Inc., Markham, Canada) according to manufacturer's instructions. PCR were performed in a total volume of 20 µL with 2 µL of genomic DNA, 200 nm of each primer, 200 µM of dNTPs, 1 U *Taq* DNA polymerase (HotMaster^{TM} *Taq* DNA polymerase, 5PRIME, Gaithersburg, USA) in 1× *Taq* buffer. Thermal cycling consisted of an initial denaturation at 95°C for 2 min, followed by 35 cycles (94°C for 20 s, 50°C for 20 s and 65°C for 90 s) and a final 5 min extension at 65°C.

PCR products were run on a 2 % agarose gel in 1X TAE buffer to be analyzed for size and subsequently sent for direct Sanger sequencing (Genewiz, UK), using forward and reverse primers.

Phylogenetic analysis

Phylogenetic reconstruction was performed using 20 species selected because of the availability of sequence information for the 13 mt protein-coding genes (**Table 1**). Specifically, 18 species represent the ingroup, including 7 species from the Gammaridae family and 11 species belonging to five families proposed by Hou and Sket (2016), based on partial COI sequences and 3 nuclear genes, to represent a broader definition of Gammaridae i.e. Gammarridae *sensu lato*. Two additional species, *Parhyale hawaiiensis* and *Pseudoniphargus sorbasiensis*, were used as outgroup to root the resulting tree (Romanova et al., 2016; Macher et al., 2017a) (**Table 1**). Data were obtained from either annotated mt genomes (NCBI GenBank), unannotated genome assemblies (NCBI WGS), unannotated transcriptome assemblies (NCBI TSA) and from raw transcriptome data assembled anew (NCBI SRA) as presented by Macher et al (2017a) (**Table 1**).

For each protein-coding gene, nucleotide sequences were aligned using MAFFT v7.299b in automatic mode (Katoh & Standley, 2013). Single-gene alignments were then concatenated into a single alignment using FASconCAT 1.0 (Kück & Longo, 2014), resulting in a total of 10,629 aligned nucleotides. The best model of nucleotide substitution was chosen with jModelTest2 (Darriba et al., 2012). A Bayesian phylogenetic tree was constructed using MrBayes v.3.2.6 (Ronquist & Huelsenbeck, 2003) with 0.5M iterations (Effective Sample Size > 200).

Results

General features of the *G. roeselii* mitochondrial genome

The complete mt genomes of Ou3 and Ou53 from *G. roeselii* are circular, with respective sizes of 16,073 and 16,072 bp. The two genomes only differ by a single nucleotide polymorphism (SNP) at position 1,736 (C/T) in the *cox2* gene and a 1-bp indel at position 14,844 (A/-) in a non-coding region. Because both mt genomes are nearly identical, all analyses and results presented hereafter in the manuscript are based on the Ou3 mt genome (unless specified otherwise). The *G. roeselii* mt genome is enriched in A and T nucleotides (66.8%). The mt gene set corresponds to the canonical bilaterian gene set: 13 protein-coding genes, 2 rRNA genes, and 22 tRNA genes (**Table 2**). Of these 37 genes, 23 are encoded on the forward strand and 14 on the reverse strand. Nineteen pairs of genes are overlapping by 1 to 60 bp: *trnL2-cox2*, *cox2-trnK*, *trnK-trnD*, *atp8-atp6*, *atp6-cox3*, *cox3-nad3*, *trnA-trnS1*, *trnN-trnE*, *trnE-trnR*, *trnR-trnF*, *trnF-nad5*, *trnH-nad4*, *trnT-trnP*, *nad6-cob*, *cob-trnS2*, *nad1-trnL1*, *trnV-rrnS*, *trnY-trnQ* and *nad2-trnW*. Protein-coding genes show the same order and transcriptional polarity compared to the reference pancrustacean ground pattern (Boore et al., 1998). However, the arrangement is modified for 6 tRNA genes: *trnS1*, *trnR*, *trnQ*, *trnQ*, *trnC* and *trnW* (**Figure 2**). The mt genome also contains 15 intergenic regions with length ranging from 1 to 1,029 bp.

A comparative analysis of gene order among *Gammarus* species mitogenomes indicated that protein-coding genes of all species present the same order and transcriptional polarity compared to the Pancrustacean ground pattern. The tRNA genes are more labile and the Pancrustacean ground pattern is not conserved within the *Gammarus* genus (**Figure 2**). The translocation of *trnG* forming the tRNA gene string *trnW-trnG*, the translocation of *trnS1*

and *trnR* forming the tRNA gene string *trnA-trnS1-trnN-trnE-trnR* and the translocations of *trnY, trnQ* and *trnC* forming the tRNA gene string *trnY-trnQ-trnC* next to the CR.

Unusual structure of the control region

The CR is the most important non-coding part of the mt genome and it is involved in replication and transcription (Saito et al., 2005). The CR usually consists of a single locus, which is typically associated with several specific features in invertebrates, including presence of a poly-T stretch, tandemly repeated sequences and AT-rich composition (Hassanin et al., 2005; Saito et al., 2005; Shao et al., 2005; Wei et al., 2010; Doublet et al., 2013). The putative CR of *G. roeselii* Ou3 and Ou53 is unusual in that it is duplicated (see below), spanning the two longest intergenic regions hereafter named CR1 and CR2 (**Figure 1**), located between *rrnS* and *trnaI* (**Figure 2**). CR1 and CR2 both start with a T-stretch of 13 bp and they are enriched in A and T nucleotides (78.4%). CR1 and CR2 are separated by a short region (169 bp) containing three tRNA genes (*trnY*, *trnQ* and *trnC*) (**Figure 1**, **Figure 2**). Interestingly, CR1 and CR2 are nearly identical, differing only by a SNP at position 13,312 (G/A) and a 2-bp indel (AA/--) at positions 13,621 (relative to CR1), strongly suggesting that the observed CR structure of *G. roeselii* is due to a duplication of the ancestral CR (**Figure 2**).

CR duplication was independently confirmed using both bioinformatic analyses and molecular biology experiments. Specifically, Inspection of the CR1/CR2 region of the mt genome showed that Illumina sequencing depth in this region is higher than in the rest of the mitogenome (**Figure 3**). For both Ou3 and Ou53 Illumina datasets, sequencing depth in the CR appeared to be nearly twice as high as mitogenome mean sequencing depth, assuming a single copy of the CR is present. Indeed, Ou3 mitogenome mean sequencing depth was

~11,000x while CR mean sequencing depth in CR was ~18,000x (**Figure 3A**). Similarly, Ou53 mitogenome mean sequencing depth was ~18,000x while CR mean sequencing depth was ~32,000x (**Figure 3B**). These convergent observations in two different individuals suggested that the CR duplication is unlikely to be an artifact of the *de novo* genome assembly process.

To independently test the authenticity of the CR duplication, we designed a diagnostic PCR assay yielding a single band (371 bp) if the duplication is present and no PCR product if the duplication is absent. Testing of 23 *G. roeselii* samples caught in the Ouche river resulted in all individuals amplifying a single band of expected 371 bp size (**Figure 4)**. Specific PCR amplification was confirmed by Sanger sequencing. This result independently validated the authenticity of the CR duplication and provided direct evidence that the CR is actually duplicated in all tested *G. roeselii* individuals. CR duplication was also observed in another population (Morthe river), in which all 15 individuals tested were found to amplify a single 371 bp band.

To further investigate the unusual *G. roeselii* CR structure, we compared the Ou3 mt genome to that produced by Macher et al. (2017b). The two genomes were highly similar at the nucleotide level, only differing by 6 SNPs and a 1-bp indel (**Table 3**). The only major difference between the two genomes was a 1,181 bp deletion in the Macher et al (2017b) genome relative to the Ou3 genome. This deletion encompassed tRNA genes *trnY*, *trnQ* and *trnC* and one full copy of the CR. We were unable to ascertain if the missing CR corresponds to the full CR1 or CR2, or parts of both CR1 and CR2, because of the high nucleotide similarity of CR1 and CR2. This configuration suggested that the Macher et al (2017b) assembly has been collapsed due to the CR1/CR2 duplication, leading to the concomitant and artefactual deletion of three tRNA genes. This explanation is supported by the fact that lacking three tRNA genes would be deleterious for the mt genome; the deletion thus most

likely cannot have a biological origin. Overall, these observations indicate that the unusual CR structure of the *G. roeselii* mt genome is authentic and that the mt sequence of *G. roeselii* generated by Macher et al. (2017b) is a partial assembly.

Extensive variation in CR length has previously been reported in Gammaridae *sensu lato* (**Figure 5**). Indeed, while average CR size is 1,234 bp, CR varies in size among species from 181 bp in *Eulimnogammarus cyaneus* to 2709 bp in *Brachyuropus grewingkii*. This effectively means that there is a 15-fold variation of CR size among species. We found that repeat content varies between CR (from 0 to 15%) but no correlation was observed between CR size and repeat percentage ($r = 0.35$, $p = 0.314$) (**Figure 5**). The AT nucleotide content also varies substantially between species, from 67% in *B. grewingkii* to 81% in *G. duebeni* and *Pallaseopsis kessleri*. No correlation between CR size and AT nucleotide content was observed ($r = -0.55$, $p = 0.078$).

Phylogenetic relationships of gammarids

The Bayesian phylogenetic reconstruction (**Figure 5**) showed that while *G. roeselii*, *G. fossarum* and *G. pulex* form a well-supported clade, the genus *Gammarus* is not monophyletic, as the three other *Gammarus* species (i.e. *G. lacustris*, *G. cheuvreuxi* and *G. duebeni*) are more closely related to species from families Eulimogammaridae, Acanthogammaridae, Micruropodidae, Pallaseidae and Crypturopodidae than to other *Gammarus* species (but see discussion). Second, the family Gammaridae is not monophyletic. Support for this conclusion comes from the aforementioned results on *Gammarus* species and the fact that *Echinogammarus veneris* is more related to *Pandorites podoceroides* (family Pontogammaridae) than to other Gammaridae species.

Discussion

The analysis of the mitogenome of *G. roeselii* reveals that the order and transcriptional polarity is conserved compared to the Pancrustacean ground pattern for protein-coding genes and modified for six tRNA genes. Despite a common structure, the mt genomes of *G. roeselii*, *G. duebeni* and *G. fossarum* show important size variation. This is mainly due to variation in the size of intergenic regions, and more specifically in the CR. The unusually large size of the *G. roeselii* mt genome is the result of an event of duplication of the CR. This is the first verified report of a duplicated CR in the superfamily Gammaridea and, overall, one of a relatively few examples described in animal mitogenomes.

Previously, one potential case of CR duplication was reported in the acanthogammarid *Garjajewia cabanisii* (Romanova et al., 2016). However, the CR was only partially sequenced in that study, which prevented any firm conclusion to be drawn (Romanova et al., 2016). In amphipods, CR duplication has been reported in two related species: *Caprella mutica* (Kilpert & Podsiadlowski, 2010) and *Caprella scaura* (Ito et al., 2010). As in *G. roeselii*, both *Caprella* species possess highly similar duplicated CR copies. Beyond amphipods, CR duplications have been observed in several animal species, including ticks (Black & Roehrdanz, 1998; Campbell & Barker, 1999), ostracods (Ogoh & Ohmiya, 2004), sea cucumbers (Arndt & Smith, 1998), katydids (Yang et al., 2016), birds (ABBOTT et al., 2005; Morris-Pocock et al., 2010; Schirtzinger et al., 2012), turtles (Parham et al., 2006a, 2006b; Peng et al., 2006), snakes (Kumazawa et al., 1996, 1998; Dubey et al., 2012) and fishes (Lee et al., 2001; Tatarenkov & Avise, 2007; Shi et al., 2014). Sometimes, CR duplications extend beyond the sole CR, as exemplified by terrestrial isopods in which the entire mt genome is duplicated (Marcadé et al., 2007; Chandler et al., 2015; Peccoud et al., 2017). Thus, although

reported in a diverse array of species, CR duplication appears to be a relatively rarely documented cause of mt structural variation, especially when considering the high number of animal mitogenomes that have been sequenced to date.

The mechanisms associated with the occurrence of CR duplications are not fully elucidated but several hypotheses have been proposed. Potential mechanisms that could duplicate CR include illegitimate recombination (Lunt & Hyman, 1997), dimeric mitogenome (Boore, 2000) and tandem duplication (Boore, 2000). Illegitimate recombination is characterized by recombination between non-homologous or short homologous sequences, but the exact associated biological mechanisms remain poorly understood (Rijk & Bloemendal, 2003). In the case of dimeric mitogenomes, dimeric molecules are formed by two monomers linked head-to-tail, which produces a fully duplicated mt genome. Another mechanism is tandem duplication, in which case an identical segment of a sequence is inserted next to the original sequence. Interestingly, the duplicated CR of *G. roeselii* show very high sequence similarity, a pattern that has also been noted in several other species with duplicated CR. It has been proposed that this may be the result of concerted evolution, which leads to sequence homogenization of duplicated CR (Liao, 2007; Li et al., 2015).

The potential advantages of the presence and conservation of duplicated CR remains largely unknown and unexplored. Based on studies in mammals and fruit flies, it has been proposed that mt genomes with duplicated CR may have a selective advantage over mt genomes with a single CR, by allowing a higher replication rate (Shao et al., 2005). Moreover, in some avian families, it was demonstrated that duplication of the CR forms a pseudo-CR which is associated with extended lifespan (Skujina et al., 2016). It is unclear whether any of these potential advantages may apply to *G. roeselii* and this issue represents an interesting avenue for future research in *G. roeselii* mitochondrial biology and, more

generally, to further investigate the evolutionary forces at work and implications of CR duplications in animal mt genomes.

Until recently (but see Ramonova et al 2016, Macher et al 2017a), the only mt markers used in Gammarid phylogeny and phylogeography were the COI barcode (~625 bp) and a small region *rnnS* (~300 bp) (e.g. Grabowsky et al 2017, Moret et al 2007). In *G. roeselii*, two complete mt genomes (Ou3 and Ou53 from France, this study) and one partial genome (from Germany, Macher et al 2017b) are now available. Despite the three genomes all sharing the same COI and *rrnS* haplotype, multiple SNPs were identified in four protein-coding genes and *rrnL* (**Table 3**). This is promising with respect to the relevance of next-generation sequencing in assessing polymorphism at very low taxonomic level and disentangling the evolutionary history of *G. roeselii*. The value of such a strategy for mitogenome sequencing as an approach to study intraspecific expansion/invasion is increasingly recognized (Rius et al., 2015) and it has been successfully applied to many organisms (e.g. Wu et al., 2017). The optimal strategy for *G. roeselii* remains to be found, being a compromise between sample size and genome coverage. The mitogenome enrichment strategy recently proposed by Macher et al. (2017b) may be highly relevant in this context. Alternatively, a more affordable strategy may consist in extending sampling in a pilot assay to identify mt genome regions with high information content, which may then be targeted by PCR using primers designed for indexed multiplex amplicon sequencing (Maddock et al., 2016; Meimberg et al., 2016).

In our phylogenetic analysis, *G. roeselii*, *G. fossarum* and *G. pulex* form a well-supported clade, a feature also observed in other studies based on either a limited set of partial mt DNA and nuclear DNA sequences (e.g. Hou et al 2011) or a set of 13 mt protein-coding genes (Macher et al., 2017a). However, the genus *Gammarus* does not appear to be monophyletic in the present study as the three other *Gammarus* species *G. lacustris*, *G. cheuvreuxi* and *G. duebeni* are more closely related to Baikalian families. However, the

phylogenetic status of *Gammarus* fluctuated over the recent years, sometimes being considered monophyletic (Hou et al., 2011), monophyletic but with low support (Macher et al., 2017a) or paraphyletic (Hou & Sket, 2016). These discrepancies might be caused by incomplete taxon sampling and/or limited informative content of used markers. Testing these hypotheses is beyond the scope of the present study but it is noteworthy that both the present study and Macher et al 2017a, were unable to provide a fully resolved tree for this taxonomic issue, despite use of 13 protein-coding genes and ~11,000 nucleotide-long datasets.

At a higher taxonomical level, the families included in our study are targets of molecular phylogenetic reassessments (Hou & Sket, 2016; Romanova et al., 2016). Our results, along with Macher et al (2017a), clearly support the conclusion that Gammaridae is not a monophyletic family. Based on partial mt and nuclear sequences, Hou and Sket (2016) proposed that the monophyly of Gammaridae could be assured but to avoid the paraphyly of the family, they proposed the omission of Pontogammaridae, Typhlogammaridae and all Baikalian families. Our results suggest that Gammaridae is paraphyletic with Pontogammaridae and Baikalian families (Eulimogammaridae, Acanthogammaridae, Microropodidae, Pallaseidae, Crypturopodidae). Based on 13 mt protein-coding genes, Romanova et al (2016) proposed that Baikalian families should be considered as monophyletic, which is consistent with our results. However, as the sister family Gammaridae is paraphyletic, further investigations are warranted to confirm this conclusion.

In conclusion, our study presents the first case of verified duplication of the CR in the *Gammarus* genus. Comparisons with other *G. roeselii* MOTUs and gammarid species will provide information on whether this type of structural variation is unique to MOTU of *G. roeselii* present outside the Balkans or a more general feature of gammarids. The discovery of this CR duplication also offers the opportunity to explore the molecular dynamics of mt genomes, for example with respect to regulation of mt genome replication or the origin of mt

gene translocations. Finally, the characterization of multiple mt haplotypes from closely related *G. roeselii* individuals constitutes a promising avenue to constitute a valuable set of markers for phylogeographic analyses of *G. roeselii*.

Acknowledgments

We thank Isabelle Giraud for preparation of DNA samples used for sequencing, Amine Chebbi and Bouziane Moumen for assistance with bioinformatic analyses, and Clément Gilbert for discussions. We also thank the Roscoff Bioinformatics platform ABiMS (http://abims.sb-roscoff.fr) for providing additional computational resources. We also thank the editors and reviewers for their comments on an earlier version of the manuscipt. This work was funded by Agence Nationale de la Recherche Grant ANR-15-CE32-0006 (CytoSexDet), Centre National de la Recherche Scientifique (CNRS) PEPS ExoMod Grant (MicroFem) to R.C. and T.R., the 2015–2020 State-Region Planning Contract and European Regional Development Fund, and intramural funds from the CNRS and the University of Poitiers.

References

ABBOTT, DOUBLE, T. JWH, ROBINSON, & COCKBU, 2005. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in Thalassarche albatrosses. Molecular Ecology 14: 3605–3613.

Altschul, Gish, Miller, Myers, & Lipman, 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403–410.

Andreï, Pain-Devin, Felten, Devin, Giambérini, Mehennaoui, Cambier, Gutleb, & Guérold, 2016. Silver nanoparticles impact the functional role of Gammarus roeselii (Crustacea Amphipoda). Environmental Pollution 208: 608–618.

Arndt, & Smith, 1998. Mitochondrial gene rearrangement in the sea cucumber genus Cucumaria. Molecular biology and evolution 15: 1009–1016.

Aunins, Nelms, Hobson, & King, 2016. Comparative mitogenomic analyses of three North American stygobiont amphipods of the genus Stygobromus (Crustacea: Amphipoda). Mitochondrial DNA Part B 1: 560–563.

Bauzà-Ribot, Jaume, Juan, & Pons, 2009. The complete mitochondrial genome of the subterranean crustacean Metacrangonyx longipes (Amphipoda): A unique gene order and extremely short control region. Mitochondrial DNA 20: 88–99.

Bauzà-Ribot, Juan, Nardi, Oromí, Pons, & Jaume, 2012. Mitogenomic Phylogenetic Analysis Supports Continental-Scale Vicariance in Subterranean Thalassoid Crustaceans. Current Biology 22: 2069–2074.

Bernt, Donath, Jühling, Externbrink, Florentz, Fritzsch, Pütz, Middendorf, & Stadler, 2013. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution 69: 313–319.

Black, & Roehrdanz, 1998. Mitochondrial gene order is not conserved in arthropods: prostriate and metastriate tick mitochondrial genomes. Molecular biology and evolution 15: 1772–1785.

Bolger, Lohse, & Usadel, 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

Boore, 1999. Animal mitochondrial genomes. Nucleic Acids Research 27: 1767–1780.

Boore, 2000. Computational Biology. Comparative Genomics 1: 133–147.

Boore, Lavrov, & Brown, 1998. Gene translocation links insects and crustaceans. Nature 392: 667–668.

Böttger, Schaller, & Mohr, 2012. Closer to reality — the influence of toxicity test modifications on the sensitivity of Gammarus roeselii to the insecticide imidacloprid. Ecotoxicology and Environmental Safety 81: 49–54.

Campbell, & Barker, 1999. The novel mitochondrial gene arrangement of the cattle tick, Boophilus microplus: fivefold tandem repetition of a coding region. Molecular Biology and Evolution 16: 732–740.

Chandler, Badawi, Moumen, Grève, & Cordaux, 2015. Multiple Conserved Heteroplasmic Sites in tRNA Genes in the Mitochondrial Genomes of Terrestrial Isopods (Oniscidea). G3: Genes|Genomes|Genetics 5: 1317–1322.

Darriba, Taboada, Doallo, & Posada, 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772–772.

Dierckxsens, Mardulyn, & Smits, 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18–e18.

Doublet, Helleu, Raimond, Souty-Grosset, & Marcadé, 2013. Inverted Repeats and Genome Architecture Conversions of Terrestrial Isopods Mitochondrial DNA. Journal of Molecular Evolution 77: 107–118.

Dubey, Meganathan, & Haque, 2012. Complete mitochondrial genome sequence from an endangered Indian snake, Python molurus molurus (Serpentes, Pythonidae). Molecular Biology Reports 39: 7403–7412.

Gergs, Schlag, & Rothhaupt, 2013. Different ammonia tolerances may facilitate spatial coexistence of Gammarus roeseliii and the strong invader Dikerogammarus villosus. Biological Invasions 15: 1783–1793.

Gerhardt, Bloor, & Mills, 2011. Gammarus: Important Taxon in Freshwater and Marine Changing Environments. International Journal of Zoology 2011: 1–2.

Gismondi, Beisel, & Cossu-Leguille, 2012a. Polymorphus Minutus Affects Antitoxic Responses of Gammarus Roeselii Exposed to Cadmium. PLoS ONE 7: e41475.

Gismondi, Cossu-Leguille, & Beisel, 2012b. Acanthocephalan parasites: help or burden in gammarid amphipods exposed to cadmium?. Ecotoxicology 21: 1188–1193.

Gismondi, Cossu-Leguille, & Beisel, 2013. Do male and female gammarids defend themselves differently during chemical stress?. Aquatic Toxicology 140: 432–438.

Gismondi, Rigaud, Beisel, & Cossu-Leguille, 2012c. Microsporidia parasites disrupt the responses to cadmium exposure in a gammarid. Environmental Pollution 160: 17–23.

Gismondi, & Thomé, 2016. Transcriptome of the freshwater amphipod Gammarus pulex hepatopancreas. Genomics Data 8: 91–92.

Grabowski, 2007. Alien Crustacea in Polish waters – Amphipoda. Aquatic Invasions 2: 25–38.
Grabowski, Mamos, Bącela-Spychalska, Rewicz, & Wattier, 2017. Neogene paleogeography provides context for understanding the origin and spatial distribution of cryptic diversity in a widespread Balkan freshwater amphipod. PeerJ 5: e3016.
Haine, Boucansaud, & Rigaud, 2005. Conflict between parasites with different transmission strategies infecting an amphipod host. Proceedings of the Royal Society of London B: Biological Sciences 272: 2505–2510.
Hassanin, Léger, Deutsch, & Collins, 2005. Evidence for Multiple Reversals of Asymmetric Mutational Constraints during the Evolution of the Mitochondrial Genome of Metazoa, and Consequences for Phylogenetic Inferences. Systematic Biology 54: 277–298.
Hou, & Sket, 2016. A review of Gammaridae (Crustacea: Amphipoda): the family extent, its evolutionary history, and taxonomic redefinition of genera. Zoological Journal of the Linnean Society 176: 323–348.
Hou, Sket, Fišer, & Li, 2011. Eocene habitat shift from saline to freshwater promoted Tethyan amphipod diversification. Proceedings of the National Academy of Sciences 108: 14533–14538.
Ito, Aoki, Yokobori, & Wada, 2010. The complete mitochondrial genome of Caprella scaura (Crustacea, Amphipoda, Caprellidea), with emphasis on the unique gene order pattern and duplicated control region. Mitochondrial DNA 21: 183–190.
Jażdżewski, K., & A.-L. Roux, 1988. Biogéographie de Gammarus roeselii Gervais en Europe, en particulier répartition en France et en Pologne. Crustaceana. Supplement 272–277.
Kao, Lai, Stamataki, Rosic, Konstantinides, Jarvis, Donfrancesco, Pouchkina-Stancheva, Semon, Grillo, Bruce, Kumar, Siwanowicz, Le, Lemire, Eisen, Extavour, Browne, Wolff, Averof, Patel, Sarkies, Pavlopoulos, & Aboobaker, 2016. The genome of the crustacean Parhyale hawaiensis, a model for animal development, regeneration, immunity and lignocellulose digestion. eLife 5: e20062.
Karaman, G., & S. Pinkster, 1977. Freshwater gammarus species from Europe, North Africa and adjacent regions of Asia (Crustacea-Amphipoda). Part II. Gammarus roeselii-group and related species. Bijdragen tot de Dierkunde 47: 165–196.
Katoh, & Standley, 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30: 772–780.
Ki, Hop, Kim, Kim, Park, & Lee, 2010. Complete mitochondrial genome sequence of the Arctic gammarid, Onisimus nanseni (Crustacea; Amphipoda): Novel gene structures and

476 unusual control region features. Comparative Biochemistry and Physiology Part D: Genomics
477 and Proteomics 5: 105–115.
478 Kilpert, & Podsiadlowski, 2010. The mitochondrial genome of the Japanese skeleton shrimp
479 Caprella mutica (Amphipoda: Caprellidea) reveals a unique gene order and shared
480 apomorphic translocations with Gammaridea. Mitochondrial DNA 21: 77–86.
481 Krebes, & Bastrop, 2012. The mitogenome of Gammarus duebeni (Crustacea Amphipoda): A
482 new gene order and non-neutral sequence evolution of tandem repeats in the control region.
483 Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 7: 201–211.
484 Kück, & Longo, 2014. FASconCAT-G: extensive functions for multiple sequence alignment
485 preparations concerning phylogenetic studies. Frontiers in Zoology 11: 1–8.
486 Kumazawa, Ota, Nishida, & Ozawa, 1996. Gene rearrangements in snake mitochondrial
487 genomes: highly concerted evolution of control-region-like sequences duplicated and inserted
488 into a tRNA gene cluster. Molecular Biology and Evolution 13: 1242–1254.
489 Kumazawa, Ota, Nishida, & Ozawa, 1998. The complete nucleotide sequence of a snake
490 (Dinodon semicarinatus) mitochondrial genome with two identical control regions. Genetics
491 150: 313–329.
492 Langmead, & Salzberg, 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:
493 357–359.
494 Lan, Sun, Bartlett, Rouse, Tabata, & Qian, 2016. The deepest mitochondrial genome
495 sequenced from Mariana Trench Hirondellea gigas (Amphipoda). Mitochondrial DNA Part B
496 1: 802–803.
497 Lee, Miya, Lee, Kim, Park, Aoki, & Nishida, 2001. The complete DNA sequence of the
498 mitochondrial genome of the self-fertilizing fish Rivulus marmoratus (Cyprinodontiformes,
499 Rivulidae) and the first description of duplication of a control region in fish. Gene 280: 1–7.
500 Liao, 2007. Concerted Evolution: Molecular Mechanism and Biological Implications. The
501 American Journal of Human Genetics 64: 24–30.
502 Li, Shi, Munroe, Gong, & Kong, 2015. Concerted Evolution of Duplicate Control Regions in
503 the Mitochondria of Species of the Flatfish Family Bothidae (Teleostei: Pleuronectiformes).
504 PLOS ONE 10: e0134580.
505 Lunt, & Hyman, 1997. Animal mitochondrial DMA recombination. Nature 387: 247–247.
506 Luo, Liu, Xie, Li, Huang, Yuan, He, Chen, Pan, Liu, Tang, Wu, Zhang, Shi, Liu, Yu, Wang,
507 Lu, Han, Cheung, Yiu, Peng, Xiaoqian, Liu, Liao, Li, Yang, Wang, Lam, & Wang, 2012.
508 SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.
509 GigaScience 1: 1–6.

510 Macher, Leese, Weigand, & Rozenberg, 2017a. The complete mitochondrial genome of a
511 cryptic amphipod species from the Gammarus fossarum complex. Mitochondrial DNA Part B
512 2: 17–18.
513 Macher, Zizka, Weigand, & Leese, 2017b. A simple centrifugation protocol increases
514 mitochondrial DNA yield 140-fold and facilitates mitogenomic studies. bioRxiv 106583.
515 Maddock, Briscoe, Wilkinson, Waeschenbach, Mauro, Day, L. DTJ, Foster, Nussbaum, &
516 Gower, 2016. Next-Generation Mitogenomics: A Comparison of Approaches Applied to
517 Caecilian Amphibian Phylogeny. PLOS ONE 11: e0156757.
518 Marcadé, Cordaux, Doublet, Debenest, Bouchon, & Raimond, 2007. Structure and Evolution
519 of the Atypical Mitochondrial Genome of Armadillidium vulgare (Isopoda, Crustacea).
520 Journal of Molecular Evolution 65: 651–659.
521 Meimberg, Schachtler, Curto, Husemann, & Habel, 2016. A new amplicon based approach of
522 whole mitogenome sequencing for phylogenetic and phylogeographic analysis: An example
523 of East African white-eyes (Aves, Zosteropidae). Molecular Phylogenetics and Evolution 102:
524 74–85.
525 Moret, Bollache, Wattier, & Rigaud, 2007. Is the host or the parasite the most locally adapted
526 in an amphipod–acanthocephalan relationship? A case study in a biological invasion context.
527 International Journal for Parasitology 37: 637–644.
528 Morris-Pocock, Taylor, Birt, & Friesen, 2010. Concerted evolution of duplicated
529 mitochondrial control regions in three related seabird species. BMC Evolutionary Biology 10:
530 1–10.
531 Ogoh, & Ohmiya, 2004. Complete mitochondrial DNA sequence of the sea-firefly, Vargula
532 hilgendorfii (Crustacea, Ostracoda) with duplicate control regions. Gene 327: 131–139.
533 Paganelli, D., A. Gazzola, A. Marchini, & R. Sconfietti, 2015. The increasing distribution of
534 Gammarus roeseliii Gervais, 1835: first record of the non-indigenous freshwater amphipod in
535 the sub-lacustrine Ticino River basin. BioInvasions Records 4: 37–41.
536 Parham, Feldman, & Boore, 2006a. The complete mitochondrial genome of the enigmatic
537 bigheaded turtle (Platysternon): description of unusual genomic features and the
538 reconciliation of phylogenetic hypotheses based on mitochondrial and nuclear DNA. BMC
539 Evolutionary Biology 6: 1–11.
540 Parham, Macey, Papenfuss, Feldman, Türkozan, Polymeni, & Boore, 2006b. The phylogeny
541 of Mediterranean tortoises and their close relatives based on complete mitochondrial genome
542 sequences from museum specimens. Molecular Phylogenetics and Evolution 38: 50–64.
543 Peccoud, Chebbi, Cormier, Moumen, Gilbert, Marcadé, Chandler, & Cordaux, 2017.

544 Untangling Heteroplasmy, Structure, and Evolution of an Atypical Mitochondrial Genome by
545 PacBio Sequencing. Genetics 207: 269–280.
546 Peng, Nie, & Pu, 2006. Complete mitochondrial genome of Chinese big-headed turtle,
547 Platysternon megacephalum, with a novel gene organization in vertebrate mtDNA. Gene 380:
548 14–20.
549 Piscart, Webb, & Beisel, 2007. An acanthocephalan parasite increases the salinity tolerance of
550 the freshwater amphipod Gammarus roeselii (Crustacea: Gammaridae). Naturwissenschaften
551 94: 741–747.
552 Pons, Bauzà-Ribot, Jaume, & Juan, 2014. Next-generation sequencing, phylogenetic signal
553 and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea).
554 BMC Genomics 15: 1–16.
555 Rijk, & Bloemendal, 2003. Molecular Mechanisms of Exon Shuffling: Illegitimate
556 Recombination. Genetica 118: 245–249.
557 Rius, Bourne, Hornsby, & Chapman, 2015. Applications of next-generation sequencing to the
558 study of biological invasions. Current Zoology 61: 488–504.
559 Romanova, Aleoshin, Kamaltynov, Mikhailov, Logacheva, Sirotinina, Gornov, Anikin, &
560 Sherbakov, 2016. Evolution of mitochondrial genomes in Baikalian amphipods. BMC
561 Genomics 17: 291–306.
562 Romanova, Mikhailov, Logacheva, Kamaltynov, Aleoshin, & Sherbakov, 2014. The complete
563 mitochondrial genome of Baikalian amphipoda Eulimnogammarus vittatus Dybowsky, 1874.
564 Mitochondrial DNA 27: 1–3.
565 Romanova, Mikhailov, Logacheva, Kamaltynov, Aleoshin, & Sherbakov, 2015. The complete
566 mitochondrial genome of a deep-water Baikalian amphipoda Brachyuropus grewingkii
567 (Dybowsky, 1874). Mitochondrial DNA 27: 1–2.
568 Ronquist, & Huelsenbeck, 2003. MrBayes 3: Bayesian phylogenetic inference under mixed
569 models. Bioinformatics 19: 1572–1574.
570 Saito, Tamura, & Aotsuka, 2005. Replication Origin of Mitochondrial DNA in Insects.
571 Genetics 171: 1695–1705.
572 Schirtzinger, Tavares, Gonzales, Eberhard, Miyaki, Sanchez, Hernandez, Müeller, Graves,
573 Fleischer, & Wright, 2012. Multiple independent origins of mitochondrial control region
574 duplications in the order Psittaciformes. Molecular Phylogenetics and Evolution 64: 342–356.
575 Shao, Barker, Mitani, Aoki, & Fukunaga, 2005. Evolution of Duplicate Control Regions in
576 the Mitochondrial Genomes of Metazoa: A Case Study with Australasian Ixodes Ticks.
577 Molecular Biology and Evolution 22: 620–629.

578 Shi, Miao, & Kong, 2014. A novel model of double replications and random loss accounts for
579 rearrangements in the Mitogenome of Samariscus latus (Teleostei: Pleuronectiformes). BMC
580 Genomics 15: 1–9.
581 Shin, Cho, Lee, Ahn, Lee, & Park, 2012. Complete mitochondrial genome of the Antarctic
582 amphipod Gondogeneia antarctica (Crustacea, amphipod). Mitochondrial DNA 23: 25–27.
583 Skujina, McMahon, L. VPE, Gkoutos, & Hegarty, 2016. Duplication of the mitochondrial
584 control region is associated with increased longevity in birds. Aging (Albany NY) 8: 1781–
585 1788.
586 Sornom, Felten, Médoc, da, Rousselle, & Beisel, 2010. Effect of gender on physiological and
587 behavioural responses of Gammarus roeselii (Crustacea Amphipoda) to salinity and
588 temperature. Environmental Pollution 158: 1288–1295.
589 Stokkan, Jurado-Rivera, Juan, Jaume, & Pons, 2015. Mitochondrial genome rearrangements
590 at low taxonomic levels: three distinct mitogenome gene orders in the genus Pseudoniphargus
591 (Crustacea: Amphipoda). Mitochondrial DNA Part A 27: 3579–3589.
592 Tain, Perrot-Minnot, & Cézilly, 2007. Differential influence of Pomphorhynchus laevis
593 (Acanthocephala) on brain serotonergic activity in two congeneric host species. Biology
594 Letters 3: 69–72.
595 Tatarenkov, & Avise, 2007. Rapid concerted evolution in animal mitochondrial DNA.
596 Proceedings of the Royal Society of London B: Biological Sciences 274: 1795–1798.
597 Trapp, Armengaud, Gaillard, Pible, Chaumot, & Geffard, 2016. High-throughput proteome
598 dynamics for discovery of key proteins in sentinel species: Unsuspected vitellogenins
599 diversity in the crustacean Gammarus fossarum. Journal of Proteomics 146: 207–214.
600 Väinölä, W. JDS, Grabowski, Bradbury, Jazdzewski, & Sket, 2008. Global diversity of
601 amphipods (Amphipoda; Crustacea) in freshwater. Hydrobiologia 595: 241–255.
602 WATTIER, ENGEL, SAUMITOU-LAPRADE, & VALERO, 1998. Short allele dominance
603 as a source of heterozygote deficiency at microsatellite loci: experimental evidence at the
604 dinucleotide locus Gv1CT in Gracilaria gracilis (Rhodophyta). Molecular Ecology 7: 1569–
605 1573.
606 Wei, Shi, Chen, Sharkey, Achterberg, Ye, & He, 2010. New Views on Strand Asymmetry in
607 Insect Mitochondrial Genomes. PLoS ONE 5: e12708.
608 Wolstenholme, 1992. International Review of Cytology. 141: 173–216.
609 Wu, Kumagai, Cen, Chen, Wallis, Polek, Jiang, Zheng, Liang, & Deng, 2017. Analyses of
610 Mitogenome Sequences Revealed that Asian Citrus Psyllids (Diaphorina citri) from California
611 Were Related to Those from Florida. Scientific Reports 7: s41598-017-10713–3.

612 Yang, Ye, & Huang, 2016. Mitochondrial genomes of four katydids (Orthoptera:
613 Phaneropteridae): New gene rearrangements and their phylogenetic implications. Gene 575:
614 702–711.
615

616 **Table 1** Summary of amphipod mitogenomes used in this study. (p) indicates partial genome.

617

Species name	Family	Accession Number	Size (bp)
Acanthogammarus victorii	Acanthogammaridae	KX341962	17,424 (p)
Brachyuropus grewingkii	Acanthogammaridae	KP161875	17,118
Crypturopus tuberculatus	Cryturopodidae	KX341963	13,864 (p)
Echinogammarus veneris	Gammaridae	TSA GARO01000000	n/a
Eulimnogammarus cyaneus	Eulimnogammaridae	KX341964	14,370
Eulimnogammarus verrucosus	Eulimnogammaridae	KF690638	15,315
Eulimnogammarus vittatus	Eulimnogammaridae	KM287572	15,534
Gammarus chevreuxi	Gammaridae	TSA HADC01000000	n/a
Gammarus duebeni	Gammaridae	JN704067	15,651
Gammarus fossarum	Gammaridae	KY197961	15,989
Gammarus lacustris	Gammaridae	SRA SRR3467069	n/a
Gammarus pulex	Gammaridae	WGS FJVI01000000	n/a
Gammarus roeselii	Gammaridae	MG779536	16,073
Garjajewia cabanisii	Acanthogammaridae	KX341965	17,576 (p)
Gmelinoides fasciatus	Microropodidae	KX341966	18,114
Linevichella vortex	Microropodidae	KX341967	11,444 (p)
Pallaseopsis kesslerii	Pallaseidae	KX341968	15,759 (p)
Pandorites podoceroides	Pontogammaridae	SRA SRR3467097	n/a
Parhyale hawaiiensis	Hyalidae	AY639937	12,224 (p)
Pseudoniphargus sorbasiensis	Speudoiphargidae	LN871175	15,460 (p)

618

619

620 **Table 2** Annotation of the complete mitochondrial genome of *Gammarus roeselii*.

621

Gene	Position	Strand	Size (bp)	Start codon	Stop codon	Intergenic nucleotides
cox1	1-1,540	+	1,540	ATT	T	0
trnL2	1,544-1,603	+	60			3
cox2	1,601-2,285	+	685	ATA	T	-3
trnK	2,283-2,341	+	59			-3
trnD	2,341-2,405	+	65			-1
atp8	2,406-2,564	+	159	ATA	TAA	0
atp6	2,558-3,229	+	672	ATG	TAA	-7
cox3	3,229-4,014	+	786	ATG	TAG	-1
nad3	4,011-4,364	+	354	ATA	TAG	-4
trnA	4,372-4,432	+	61			7
trnS1	4,432-4,483	+	52			-1
trnN	4,486-4,548	+	63			2
trnE	4,546-4,608	+	63			-3
trnR	4,603-4,662	+	60			-6
trnF	4,661-4,720	-	60			-2
nad5	4,691-6,404	-	1,714	ATT	T	-30
trnH	6,423-6,482	-	60			18
nad4	6,466-7,695	-	1,230	ATG	TAA	-17
nad4l	7,788-8,078	-	291	ATG	TAG	92
trnT	8,082-8,141	+	60			3
trnP	8,141-8,201	-	61			-1
nad6	8,213-8,707	+	495	ATT	TAA	11
cob	8,707-9,843	+	1,137	ATG	TAG	-1
trnS2	9,842-9,896	+	55			-2
nad1	9,926-10,864	-	939	ATG	TAG	29
trnL1	10,859-10,920	-	62			-6
rrnL	10,921-11,897	-	977			0
trnV	11,901-11,952	-	52			12
rrnS	11,952-12,597	-	646			-1
CR1	12,594-13,622		1,029			-4
trnY	13,623-13,686	-	64			0
trnQ	13,683-13,739	-	57			-4
trnC	13,740-13,793	-	54			0
CR2	13,792-14,818		1,027			0

trnI	14,851-14,909	+	59			32
trnM	14,911-14,971	+	61			1
nad2	15,014-16,016	+	1,003	ATT	T	42
trnW	15,957-16,019	+	63			-60
trnG	16,020-16,072	+	53			0

622

623

624 **Table 3** Sequence variation between *G. roeselii* mitochondrial haplotypes. Ou3 is used as a

625 reference sequence.

626

Nucleotide coordinate	Type	Variation	Genomic locus	Haplotype with variation
1,736	SNP	C/T	*cox2*	Ou53
4,836	SNP	C/A	*nad5*	Macher et al 2017b
5,877	SNP	C/A	*nad5*	Macher et al 2017b
6,961	SNP	C/T	*nad4*	Macher et al 2017b
9,442	SNP	G/A (Gly/Ser)	*cob*	Macher et al 2017b
11,143	SNP	G/A	*rrnL*	Macher et al 2017b
11,207	SNP	C/T	*rrnL*	Macher et al 2017b
14,843	Indel	A/-	Intergenic region	Macher et al 2017b; Ou53

627

628

Figure legends

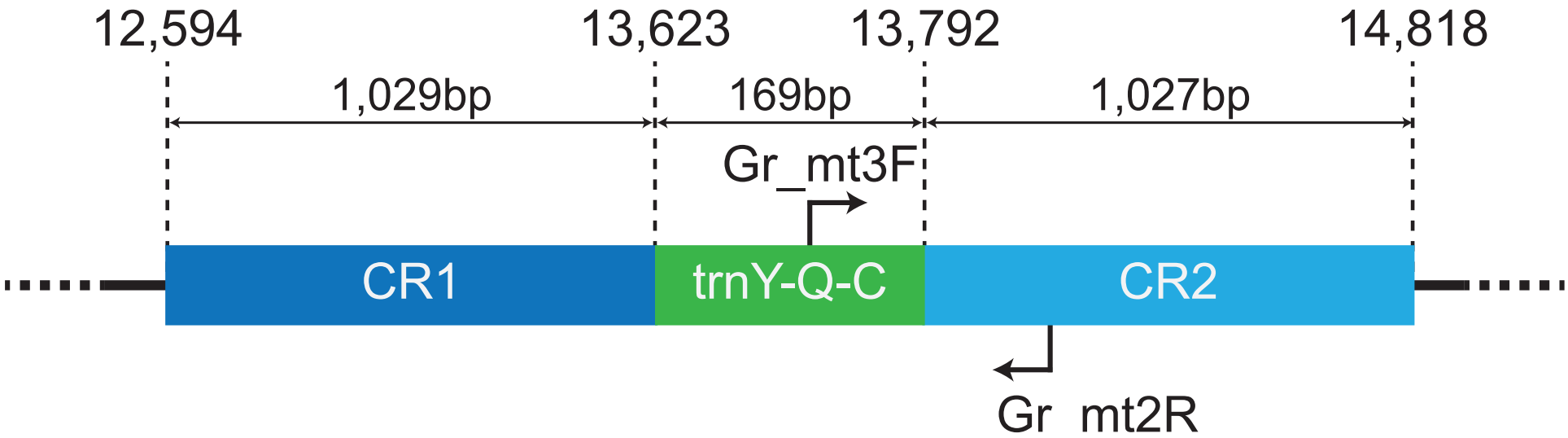
Figure 1 Control region (CR) duplication in the mitogenome of *Gammarus roeselii*. Coordinates are relative to the Ou3 mitogenome. The CR duplication consists of two CR copies (CR1 and CR2) separated by three tRNA genes (trnY-Q-C).Arrows show the targets of primers Gr_mt3F and Gr_mt2R used for PCR analyses (see main text).

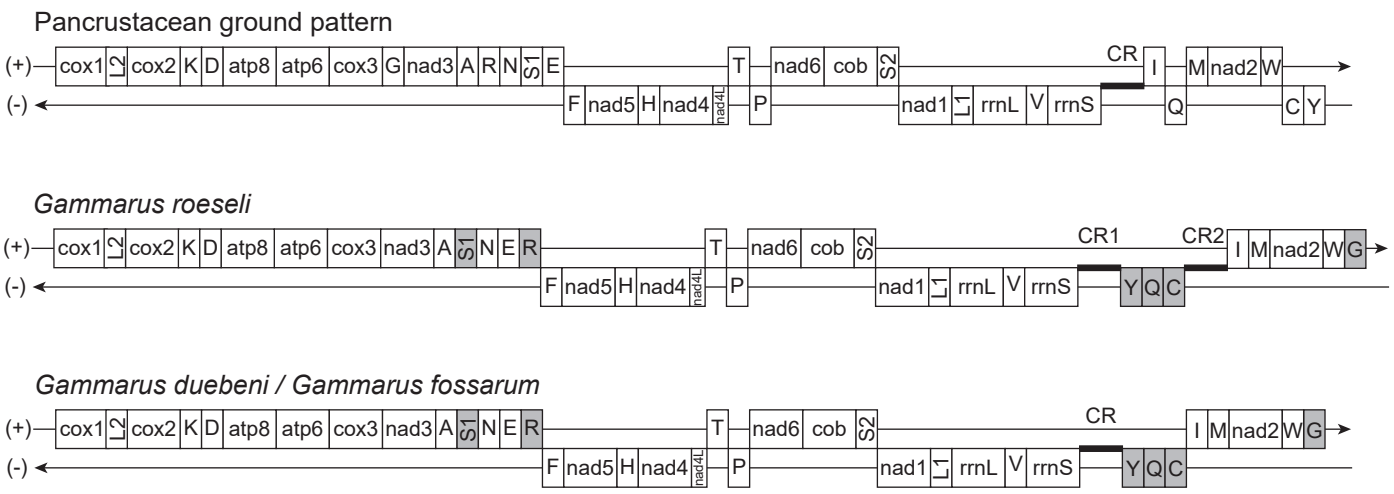
637 **Figure 2** Organization of mitochondrial genomes in three species of the genus *Gammarus*.
638 Gene features with altered location in comparison to the pancrustacean ground pattern are
639 shown in grey color. (+) indicates forward DNA strand and (-) indicates reverse DNA strand.
640 Transfer RNAs genes are labeled by their single-letter amino acid code.
641

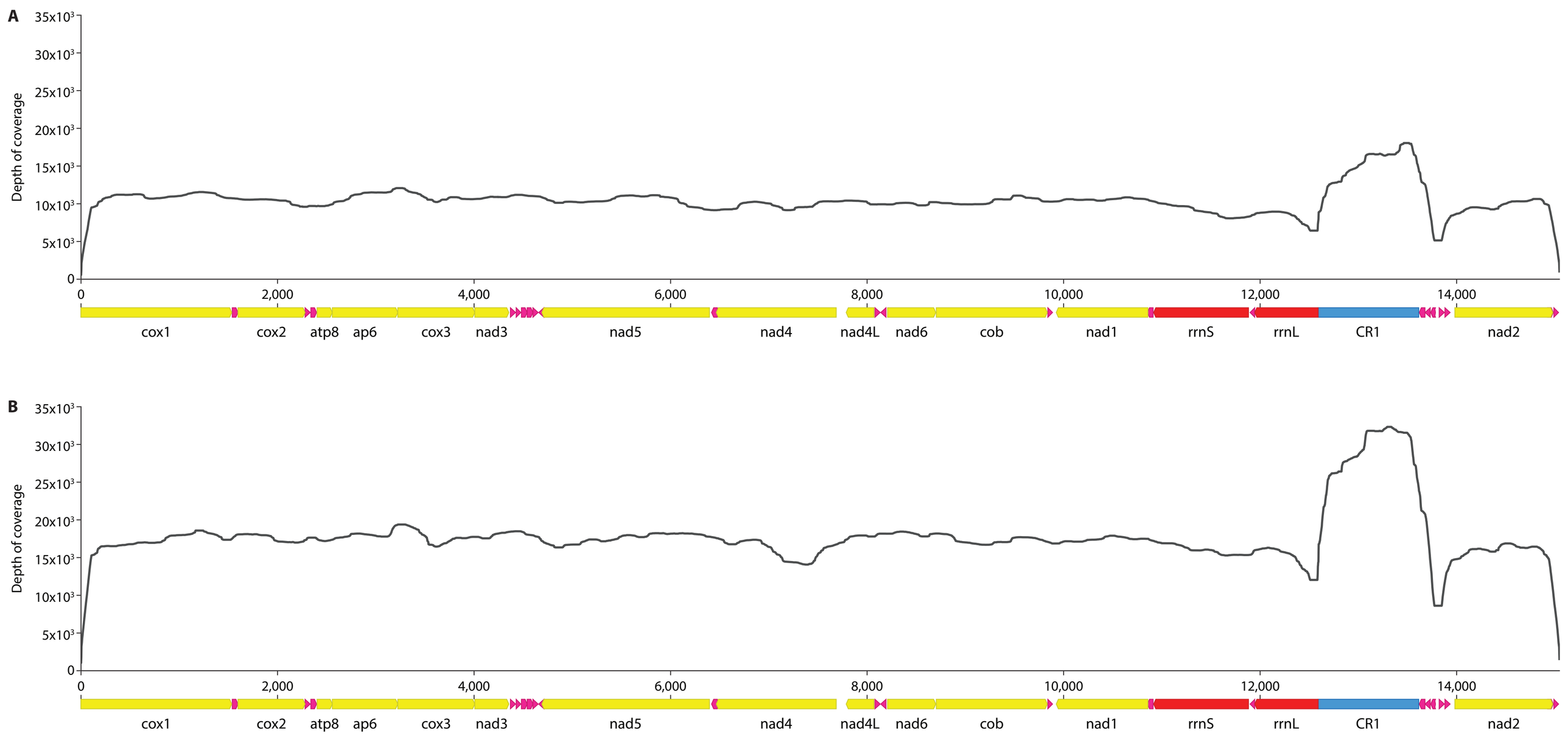
Figure 3 Per nucleotide sequencing depth of *G. roeselii* mitochondrial genome assuming a single copy of the control region. A) Sequencing depth for Ou3 individual B) Sequencing depth for Ou53 individual. Structural annotation is represented on the x axis by boxes colored according to features (protein-coding genes: yellow; tRNA: pink; rRNA: red; control region: blue). Grey curves represent the sequencing depth.

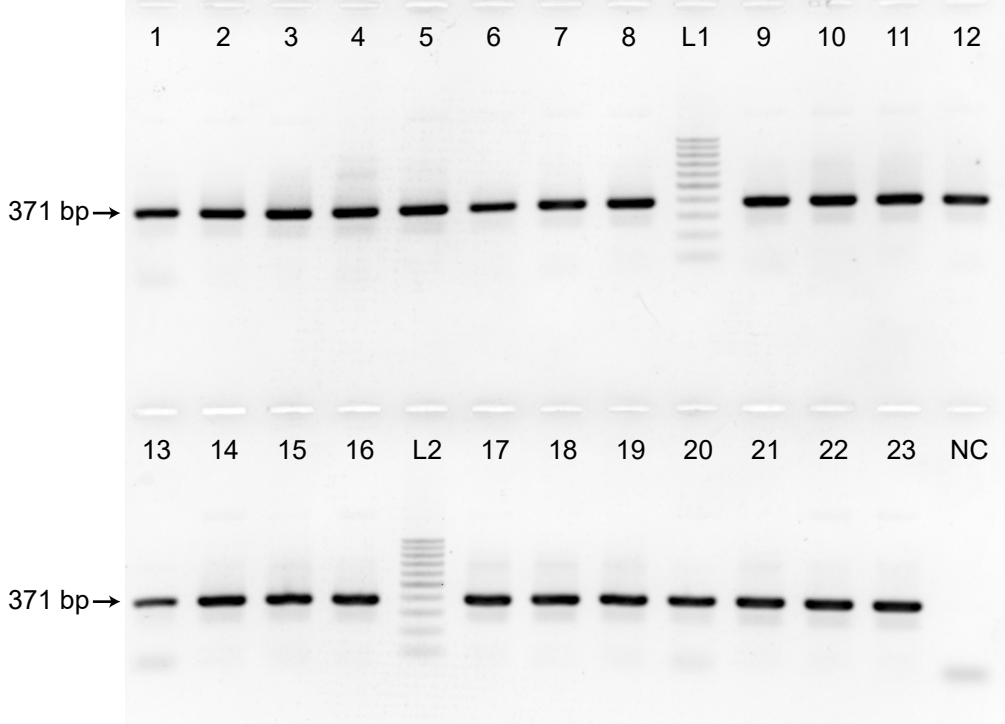
Figure 4 PCR test assessing the presence or absence of the duplicated control region in *G. roeselii* individuals. Electrophoresis gel of PCR products obtained using Gr_mt3F and Gr_mt2R primers, from 23 individuals (1 to 23) from the Ouche river (France). All individuals amplify a 371 bp-long product indicating the presence of two CR copies. L1: ladder 1 (100 bp Euromedex), L2: ladder 2 (lambda DNA EcoRI/HindIII), NC: negative control (PCR mix without DNA template). Lane 3 corresponds to the Ou3 individual from which the mt genome was obtained.

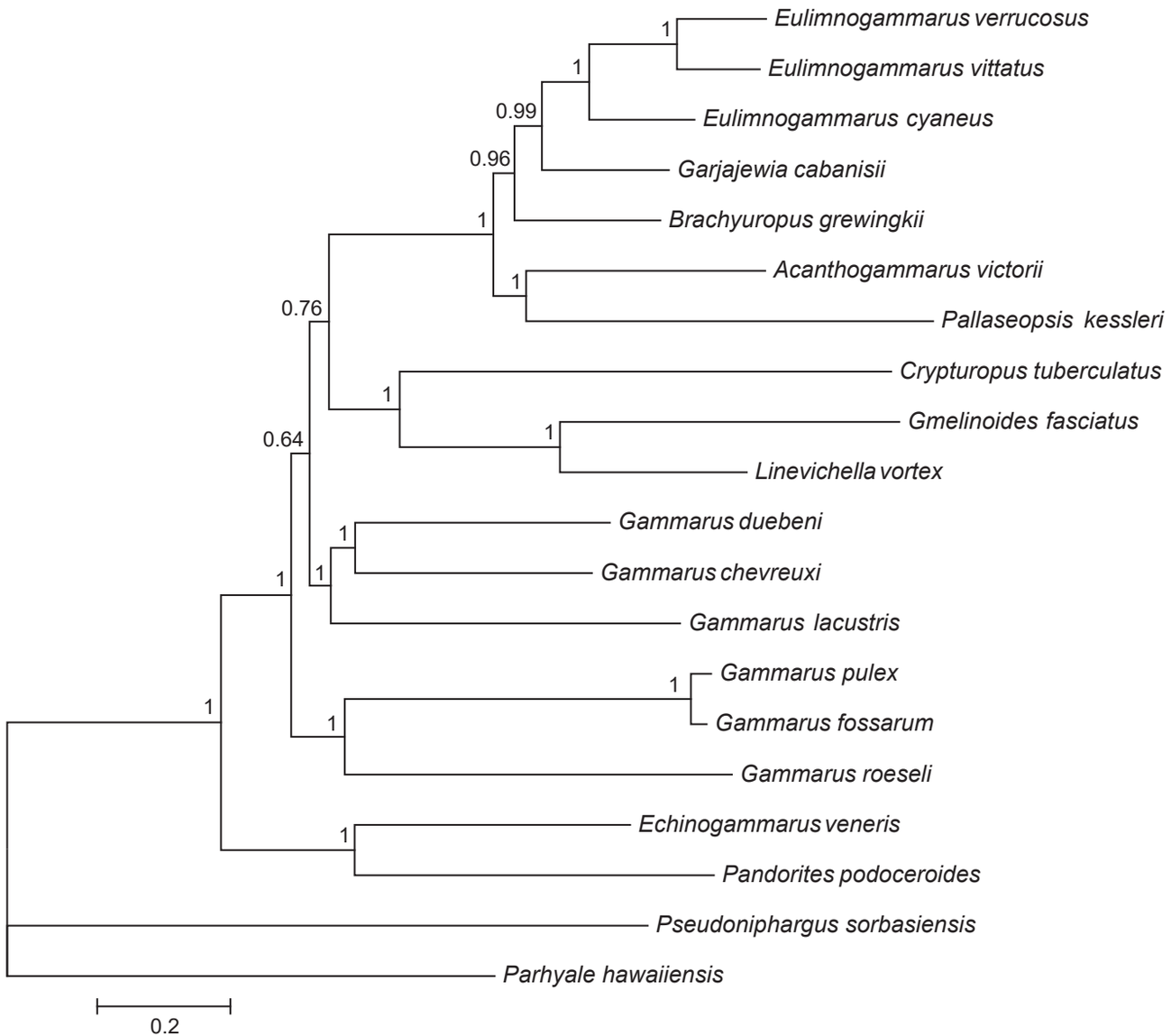
Figure 5 Bayesian phylogenetic tree of 20 gammarid species inferred from 13 mt protein-coding gene sequences. *Parhyale hawaiiensis* and *Pseudoniphargus sorbasiensis* were used as an outgroup to root the tree. Numbers above branches indicate Bayesian posterior probabilities. Scale bar below the tree indicates evolutionary distances. * Poly-T stretch refers to the size (bp) of the poly-T sequence at the start of the control region (CR).


12,594
13,623
13,792
14,818
1,029bp
169bp
1,027bp
Gr_mt3F
CR1
trnY-Q-C
CR2
Gr_mt2R


Pancrustacean ground pattern
(+)
(-)
cox1
L2
cox2
K
D
atp8
atp6
cox3
G
nad3
A
R
N
S1
E
T
nad6
cob
S2
CR
I
M
nad2
W
F
nad5
H
nad4
nad4L
P
nad1
L1
rrnL
V
rrnS
Q
C
Y
Gammarus roeseli
CR1
CR2
S1
N
E
R
Y
Q
C
Gammarus duebeni / Gammarus fossarum
G


A
B
Depth of coverage
35x10³
30x10³
25x10³
20x10³
15x10³
10x10³
5x10³
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
cox1
cox2
atp8
ap6
cox3
nad3
nad5
nad4
nad4L
nad6
cob
nad1
rrnS
rrnL
CR1
nad2


1
2
3
4
5
6
7
8
L1
9
10
11
12
371 bp
13
14
15
16
L2
17
18
19
20
21
22
23
NC
371 bp



	Family	CR size (bp)	%AT genome | CR	Poly-T stretch*	Repeats in CR (bp | %)
Eulimnogammarus verrucosus	Eulimogammaridae	1,100	69 | 80	13	168 bp | 15.27 %
Eulimnogammarus vittatus	Eulimogammaridae	1,053	67 | 79	13	46 bp | 4.37 %
Eulimnogammarus cyaneus	Eulimogammaridae	181	68 | 79	9	0 bp | 0.00 %
Garjajewia cabanisii	Acanthogammaridae	1,212	65 | 76	13	43 bp | 3.55 %
Brachyuropus grewingkii	Acanthogammaridae	2,709	62 | 67	8	0 bp | 0.00 %
Acanthogammarus victorii	Acanthogammaridae	1,390	59 | 71	13	0 bp | 0.00 %
Pallaseopsis kessleri	Pallaseidae	340	63 | 81	10	0 bp | 0.00 %
Crypturopus tuberculatus	Crypturopodidae	n/a	58 | n/a	n/a	n/a
Gmelinoides fasciatus	Micruropodidae	1,014	66 | 79	13	50 bp | 4.93 %
Linevichella vortex	Micruropodidae	n/a	65 | n/a	n/a	n/a
Gammarus duebeni	Gammaridae	1,626	64 | 81	19	19 bp | 1.17 %
Gammarus chevreuxi	Gammaridae	n/a	n/a	n/a	n/a
Gammarus lacustris	Gammaridae	n/a	n/a	n/a	n/a
Gammarus pulex	Gammaridae	n/a	n/a	n/a	n/a
Gammarus fossarum	Gammaridae	1,729	65 | 79	15	166 bp | 9.60 %
Gammarus roeseli	Gammaridae	2,225	68 | 78	13	166 bp | 8.07 %
Echinogammarus veneris	Gammaridae	n/a	n/a	n/a	n/a
Pandorites podoceroides	Pontogammaridae	n/a	n/a	n/a	n/a