



**HAL**  
open science

## Comparative genomics of strictly vertically transmitted, feminizing microsporidia endosymbionts of amphipod crustaceans.

Alexandre Cormier, Mohamed Amine Chebbi, Isabelle Giraud, Rémi Wattier, Maria Teixeira, Clement Gilbert, Thierry Rigaud, Richard Cordaux

### ► To cite this version:

Alexandre Cormier, Mohamed Amine Chebbi, Isabelle Giraud, Rémi Wattier, Maria Teixeira, et al.. Comparative genomics of strictly vertically transmitted, feminizing microsporidia endosymbionts of amphipod crustaceans.. *Genome Biology and Evolution*, 2021, 13 (1), pp.evaa245. 10.1093/gbe/evaa245 . hal-03142693

**HAL Id: hal-03142693**

**<https://hal.science/hal-03142693>**

Submitted on 27 May 2021

**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.

# Comparative Genomics of Strictly Vertically Transmitted, Feminizing Microsporidia Endosymbionts of Amphipod Crustaceans

Alexandre Cormier<sup>1</sup>, Mohamed Amine Chebbi<sup>1</sup>, Isabelle Giraud<sup>1</sup>, Rémi Wattier<sup>2</sup>, Maria Teixeira<sup>2</sup>, Clément Gilbert <sup>3</sup>, Thierry Rigaud<sup>2</sup>, and Richard Cordaux<sup>1,\*</sup>

<sup>1</sup>Laboratoire Ecologie et Biologie des Interactions, Equipe Ecologie Evolution Symbiose, Université de Poitiers, UMR CNRS 7267, France

<sup>2</sup>Laboratoire Biogéosciences, Université Bourgogne Franche-Comté, UMR CNRS 6282, Dijon, France

<sup>3</sup>Université Paris-Saclay, CNRS, IRD, UMR Évolution, Génomes, Comportement et Écologie, 91198 Gif-sur-Yvette, France

\*Corresponding author: E-mail: richard.cordaux@univ-poitiers.fr.

Accepted: 17 November 2020

## Abstract

Microsporidia are obligate intracellular eukaryotic parasites of vertebrates and invertebrates. Microsporidia are usually pathogenic and undergo horizontal transmission or a mix of horizontal and vertical transmission. However, cases of nonpathogenic microsporidia, strictly vertically transmitted from mother to offspring, have been reported in amphipod crustaceans. Some of them further evolved the ability to feminize their nontransmitting male hosts into transmitting females. However, our understanding of the evolution of feminization in microsporidia is hindered by a lack of genomic resources. We report the sequencing and analysis of three strictly vertically transmitted microsporidia species for which feminization induction has been demonstrated (*Nosema granulosis*) or is strongly suspected (*Dictyocoela muelleri* and *Dictyocoela roeselium*), along with a draft genome assembly of their host *Gammarus roeselii*. Contrary to horizontally transmitted microsporidia that form environmental spores that can be purified, feminizing microsporidia cannot be easily isolated from their host cells. Therefore, we cosequenced symbiont and host genomic DNA and devised a computational strategy to obtain genome assemblies for the different partners. Genomic comparison with feminizing *Wolbachia* bacterial endosymbionts of isopod crustaceans indicated independent evolution of feminization in microsporidia and *Wolbachia* at the molecular genetic level. Feminization thus represents a remarkable evolutionary convergence of eukaryotic and prokaryotic microorganisms. Furthermore, a comparative genomics analysis of microsporidia allowed us to identify several candidate genes for feminization, involving functions such as DNA binding and membrane fusion. The genomic resources we generated contribute to establish *Gammarus roeselii* and its microsporidia symbionts as a new model to study the evolution of symbiont-mediated feminization.

**Key words:** microsporidia, strict vertical transmission, feminization, endosymbiont, *Wolbachia*.

## Significance

Microsporidia are usually pathogenic and undergo horizontal transmission. However, some species evolved strict vertical transmission from mother to offspring and they may feminize male hosts into females. Our understanding of feminization in microsporidia is hindered by a lack of genomic resources. Here, we sequenced the genomes of three feminizing microsporidia, along with the genome of their host, the amphipod *Gammarus roeselii*. Comparison with feminizing *Wolbachia* bacterial endosymbionts of isopod crustaceans indicated that feminization evolved independently in eukaryotic and prokaryotic microorganisms. We also identified several candidate genes that may be involved in microsporidia-mediated feminization. The genomic resources we generated contribute to establish *G. roeselii* and its microsporidia symbionts as a new model to study the evolution of symbiont-mediated feminization.

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This article is published and distributed under the terms of the Oxford University Press, Standard Journals Publication Model ([https://academic.oup.com/journals/pages/open\\_access/funder\\_policies/chorus/standard\\_publication\\_model](https://academic.oup.com/journals/pages/open_access/funder_policies/chorus/standard_publication_model))

## Introduction

Obligate intracellular microorganisms exclusively replicate inside the cells of their host. Some of them are vertically transmitted from mother to offspring, infecting the cytoplasm of oocytes. From an endosymbiont perspective, males are dead ends and it is advantageous to be present in females. Consequently, several vertically transmitted microorganisms have evolved strategies to distort sex ratios of host progenies toward females, thereby increasing the frequency of the transmitting sex (Engelstädter and Hurst 2009; Cordaux et al. 2011). This phenomenon has been mostly studied through the lens of bacterial endosymbionts, the most widespread of which is the alphaproteobacterium *Wolbachia* (Werren et al. 2008). Being present in all major groups of arthropods, including at least 40% of insect species (Zug and Hammerstein 2012), *Wolbachia* is considered as the master manipulator of host reproduction. The evolutionary success of *Wolbachia* can be explained, at least partly, by its ability to distort the sex ratio of host progenies through various mechanisms, including male killing, parthenogenesis induction, and feminization of genetic males (Werren et al. 2008; Engelstädter and Hurst 2009; Cordaux et al. 2011).

Feminization converts genetic males into functional phenotypic females (Rigaud et al. 1997; Cordaux et al. 2011; Cordaux and Gilbert 2017). It is the most common phenotype induced by *Wolbachia* bacteria in terrestrial isopod crustaceans (Rigaud et al. 1997; Bouchon et al. 2008). To date, the molecular mechanisms underlying feminization have not been elucidated, although genomics approaches have yielded candidate genes (Pichon et al. 2012; Leclercq et al. 2016; Badawi et al. 2018; Chebbi et al. 2019). Symbiont-mediated feminization in crustaceans is not limited to terrestrial isopods, it has also been reported in amphipods. Strikingly, although feminization is induced by prokaryotes in isopods, the causative agents of feminization in amphipods are eukaryotes belonging to the phylum Microsporidia (Terry et al. 1998, 2004; Rodgers-Gray et al. 2004).

Microsporidia are unicellular eukaryotes related to fungi that are highly specialized obligate intracellular parasites. More than 1,300 species belonging to 187 genera have been reported that appear to infect a wide range of vertebrates and invertebrates and are particularly abundant in crustaceans and insects (Vávra and Lukeš 2013). Microsporidia are usually pathogenic but they vary greatly in how they exploit their hosts. Most studied microsporidia are horizontally transmitted and have a relatively uniform lifecycle in which the germinating spore, the only stage that can survive in the environment, is responsible for the dissemination of the parasite (Vávra and Lukeš 2013). Spores may also infect new cells of the same host, a phenomenon known as autoinfection (Weidner 1976, 1972; Frixione et al. 1992; Dunn et al. 2001). Autoinfectious spores may notably infect ovaries and parasites are subsequently vertically transmitted from mother

to offspring, through oocytes. A mix of horizontal and vertical transmission pathways has been recorded in some microsporidia (Zilio et al. 2018; Haag et al. 2020) or suggested for others (Rutrecht and Brown 2008; Traver and Fell 2012). Cases of nonpathogenic, strictly vertically transmitted microsporidia have also been reported (Terry et al. 1998, 2004; Dunn et al. 2001; Rodgers-Gray et al. 2004). They are transmitted transovarially from mother to offspring and they have lost the ability to undergo horizontal (infectious) transmission between hosts (Dunn et al. 2001). In the well-studied *Nosema granulosis*, spores developing in the host are not released in the environment (Terry et al. 1999). Instead, the spores located in the follicle cells surrounding oocytes inject sporoplasms into maturing oocytes, where meront division takes place. The division rate is moderate, as sporoplasm number is limited in freshly laid eggs (Terry et al. 1999; Haine et al. 2004). The microsporidian lifecycle then continues in developing host embryos, but symbionts target host germline during host development (Weedall et al. 2006).

Some strictly vertically transmitted microsporidia have evolved the ability to feminize their hosts, as *Wolbachia* bacteria did. To the best of our knowledge, microsporidia-mediated feminization has been formally demonstrated in the two European amphipod species *Gammarus duebeni* and *Gammarus roeselii* (Terry et al. 1998; Haine et al. 2007) and it is suspected in several other gammarid species (Terry et al. 2004). In contrast with the *Wolbachia*/isopod system, several microsporidia species belonging to different phylogenetic clades have been found in excess in female hosts, whereas almost lacking in males (thus fitting the expected pattern of feminization): *N. granulosis*, *Pleistophora mulleri*, and *Dictyocoela duebenum* in the amphipod *Gammarus duebeni* (Terry et al. 2004; Jahnke et al. 2013; Bacela-Spychalska et al. 2018) and *N. granulosis*, *Dictyocoela muelleri*, and *Dictyocoela roeselium* in the amphipod *G. roeselii* (Haine et al. 2004; Bacela-Spychalska et al. 2018; Quiles et al. 2019). However, only *N. granulosis* and *Dictyocoela duebenum* have been firmly demonstrated to induce feminization (Haine et al. 2007; Jahnke et al. 2013). Being independent of the *Wolbachia*/isopod system, the microsporidia/amphipod system constitutes an excellent model to study the evolution of feminization. However, advances in our understanding of evolution of feminization in microsporidia is hindered by a lack of genomic resources.

In this study, we report the sequencing and analysis of three strictly vertically transmitted microsporidia species infecting the gammarid host *G. roeselii*, for which feminization induction has been demonstrated (*N. granulosis*) or is strongly suspected (*D. muelleri* and *D. roeselium*). We obtained genome assemblies for the three microsporidia, along with a draft genome assembly of their amphipod host *G. roeselii*, for which no genomic resource is currently available apart from a complete mitochondrial genome (Cormier et al. 2018). We performed a comparative genomics

analysis of microsporidia to identify candidate genes for feminization, taking advantage of the recently sequenced genome of the strictly vertically transmitted, albeit nonfeminizing, *Hamiltosporidium magnivora* symbiont of *Daphnia magna* (Haag et al. 2020). This allowed us to distinguish between feminization and strict vertical transmission for the candidate gene search. The genomic resources we generated contribute to establish *G. roeselii* and its microsporidia symbionts as a new model to study the evolution of symbiont-mediated feminization.

## Results

### De Novo Assembly of Feminizing Microsporidia Genomes

We sampled 165 *G. roeselii* females in the Ouche river at Trouhans (Côte d'Or, France) and screened them for microsporidia infection using a standard polymerase chain reaction (PCR) assay targeting the *Small Sub Unit (SSU) rRNA* gene, complemented by Sanger sequencing of PCR products (Quiles et al. 2019). Microsporidia prevalence was 86% (142/165). Specifically, 133 individuals were mono-infected with *N. granulosis* (20), *D. muelleri* (1), or *D. roeselum* (112). In addition, nine individuals were multi-infected (six with *N. granulosis* and an undetermined species, and three with the two *Dictyocoela* species). We selected four mono-infected individuals for Illumina sequencing, two with *N. granulosis* (Ou3 and Ou53), one with *D. muelleri* (Ou54), and one with *D. roeselum* (Ou19). Each individual was used to prepare a sequencing library.

Contrary to horizontally transmitted microsporidia that form environmental spores that can be purified, feminizing microsporidia can be difficult to isolate from their host cells, because spores and other microsporidian stages are found at low densities in amphipod ovaries. Thus, DNA extracts contained mixes of host and symbiont genomic DNA, making particularly challenging the correct identification of symbiont sequencing reads and genome assembly. To isolate symbiont reads from host reads, we took advantage of the fact that the three microsporidia are divergent from each other while being isolated from mono-infections of the same host species, thereby allowing us to subtract *in silico* host reads from each of the three mixed sequencing data sets (supplementary fig. S1, Supplementary Material online). It is noteworthy that although *G. roeselii* is a complex of highly divergent cryptic lineages across its European distribution range (Grabowski, Mamos, et al. 2017), populations from France such as the one sampled in this study belong to a single mitochondrial lineage characterized by low genetic diversity (Grabowski, Krzywozniak, et al. 2017). Consistently, the mitogenomes of the four sequenced individuals were nearly identical, with only three variable sites including a substitution in the *cox2* gene, an intergenic 1-bp indel and a structural variant in the control region (Cormier et al. 2018).

First, we *de novo* assembled each mix of host–symbiont reads independently, resulting in three mixed genome assemblies. Next, we mapped the host–*N. granulosis* sequencing reads against the mixed host–*D. roeselum* genome assembly and the host–*D. muelleri* and host–*D. roeselum* sequencing reads against the mixed host–*N. granulosis* genome assembly. We retained unmapped reads, assuming they were derived from the symbiont genome, specific host contaminations and/or low-quality reads. Unmapped reads accounted for 3.17%, 8.57%, and 6.68% of the initial raw reads from the host–*N. granulosis*, host–*D. muelleri*, and host–*D. roeselum* data sets, respectively. Unmapped reads of each of the three data sets were separately *de novo* assembled. We then used taxonomic assignment to identify microsporidia contigs in the three assemblies and to create reference sets enabling validation of a final step of filtering using metagenomics binning and oligonucleotide composition similarity (supplementary fig. S2, Supplementary Material online).

The final assembly for *N. granulosis* was composed of 1,755 contigs with  $N_{50}$  of 12,708 bp and a total assembly length of 8.8 Mb (table 1). Genome completeness was assessed using Benchmarking Universal Single Copy Orthologs (BUSCO) (Simão et al. 2015), revealing that 98.2% (510/518) conserved specific microsporidia genes were present in the *N. granulosis* assembly. This analysis supported that our assembly probably comprises the nearly complete *N. granulosis* genome. Comparison with other microsporidia showed that *N. granulosis* is among the most complete microsporidia genomes sequenced to date (fig. 1). We conclude that our sequencing and assembly strategy was highly efficient in recovering a high-quality symbiont genome assembly from mixed host–symbiont samples.

The final assembly for *D. muelleri* was of 41.9 Mb in length (table 1), with a very good level of completeness (BUSCO score: 87.2%). *Dictyocoela muelleri* is the second largest microsporidia genome sequenced to date, after *Edhazardia aedis* (51.3 Mb) (Desjardins et al. 2015). The final assembly for *D. roeselum* was much smaller (2.2 Mb), but this probably has no biological relevance because the assembly is far from complete (BUSCO score: 49.8%). Incompleteness of the *D. roeselum* assembly presumably results from insufficient sequencing depth ( $\sim 5\times$ ) due to low *D. roeselum* titer in its host, relative to the two other symbionts (Haie et al. 2004).

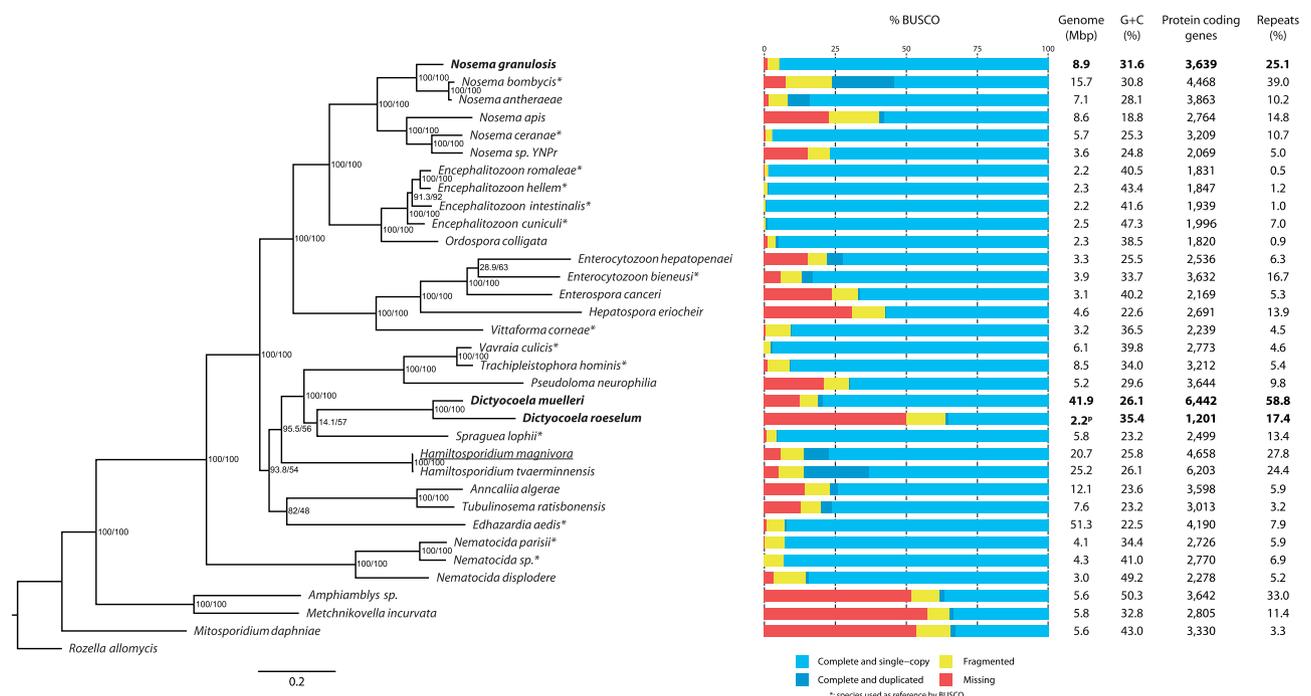
### Phylogenomic Analysis and Genome Comparisons

To investigate the evolutionary relationships of feminizing microsporidia, we performed a phylogenomic analysis of microsporidia based on 33 single-copy orthologous genes (fig. 1). All six *Nosema* species for which a genome assembly is available (including *N. granulosis*) formed a well-supported monophyletic group (all bootstrap values at 100%). *Nosema granulosis* was closely related to *N. antheraeae* and

**Table 1**

Genome Assembly and Annotation Statistics for the Three Feminizing Microsporidia *Nosema granulosis*, *Dictyocoela muelleri*, and *Dictyocoela roeselum*

Genome Assembly Features	<i>Nosema granulosis</i>	<i>Dictyocoela muelleri</i>	<i>Dictyocoela roeselum</i>
Assembly size (Mb)	8.8	41.9	2.2
Contig number	1,754	11,512	1,033
$N_{50}$ (bp)	12,700	6,565	2,755
Sequencing depth (median)	44.64	118.68	4.65
G + C content (%)	31.58	26.14	35.37
Proportion of repeats (%)	25.14	58.80	17.41
Number of genes	3,639	6,442	1,201
Gene density (genes/kb)	0.41	0.15	0.55
Mean CDS length (bp)	926	796	852
G + C content in CDS (%)	34.77	27.99	40.53
% of genome covered by CDS	38.0	12.2	46.4
Median intergenic space distance (bp)	356	1,037	196
Genes with BlastP hits	1,812	2,444	918
Genes with Pfam domains	2,093	3,514	979
Genes with GO terms	1,478	1,986	726
Annotated genes	2,247	3,775	1,025
BUSCO ( $n = 518$ )			
Complete genes	490 (94.6%)	420 (81.0%)	188 (36.3%)
Complete and single-copy genes	490 (94.6%)	411 (79.3%)	183 (35.3%)
Complete and duplicated genes	0 (0%)	9 (1.7%)	5 (1.0%)
Fragmented genes	20 (3.9%)	32 (6.2%)	70 (13.5%)
Missing genes	8 (1.8%)	66 (12.8%)	260 (50.2%)



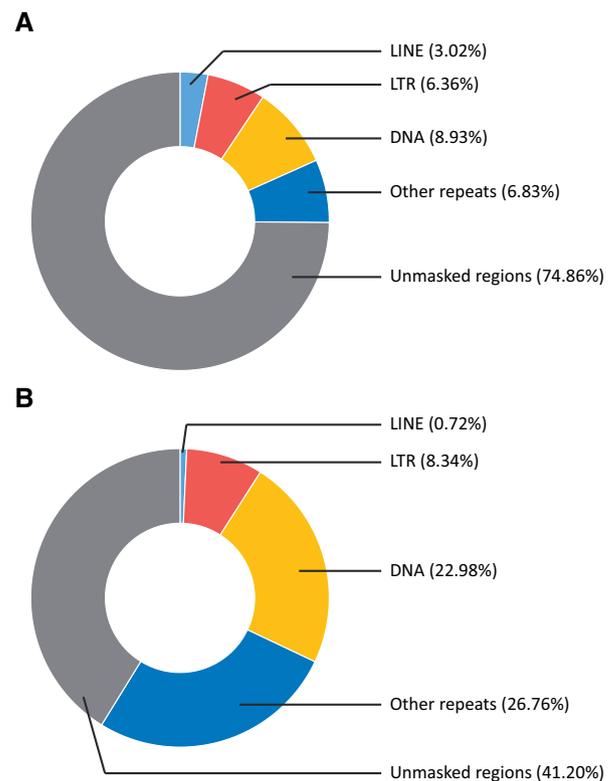
**FIG. 1.**—Phylogenetic tree of microsporidia (left) along with genome assembly completeness (BUSCO, middle) and features (right). The three strictly vertically transmitted and feminizing species sequenced in this study (*Nosema granulosis*, *Dictyocoela muelleri*, and *Dictyocoela roeselum*) are bolded. The strictly vertically transmitted albeit nonfeminizing species (*Hamiltosporidium magnivora*) is underlined. All other species undergo horizontal transmission to varying degrees. Bootstrap and SH-aLRT values (%) are indicated at each node of the tree. Scale bar indicates changes per site. \*Species used to create the microsporidia BUSCO reference set. <sup>P</sup>Partially assembled genome.

*N. bombycis* (Pan et al. 2013) and, to a lesser extent, to a clade comprising *N. ceranae* (Cornman et al. 2009), *N. apis* (Chen et al. 2013), and *Nosema* sp. YNPr (Xu et al. 2016). In contrast to the strictly vertically transmitted *N. granulosis*, all other *Nosema* species included in our analysis are insect pathogens known to mainly disperse by horizontal transmission (although vertical transmission may occasionally occur; Rutrecht and Brown 2008; Traver and Fell 2012). Thus, the branching of *N. granulosis* relative to other *Nosema* species is consistent with the view that the use of strict vertical transmission evolved secondarily from a horizontally transmitted ancestor, or from an ancestor using a mixed strategy combining both horizontal and vertical transmission.

Comparison among *Nosema* genomes indicated that most genomic features of *N. granulosis* are consistent with other *Nosema* species, such as genome size, repeat proportion, gene number and density, gene length, and median intergenic space distance (supplementary table S1, Supplementary Material online). The most distinctive feature was a higher GC content in *N. granulosis* (31.6%) relative to other *Nosema* genomes (18.8–30.8%), especially in coding sequences (CDS, 34.8% vs. 24.9–30.8%). Thus, there was no obvious genomic signature distinguishing strict vertical transmission, absence of virulence and feminization, based on these features.

Consistent with earlier evidence solely based on *SSU rRNA* gene sequencing (Bacela-Spychalska et al. 2018; Quiles et al. 2019), the phylogenetic position of *D. muelleri* and *D. roeselum* supported the monophyly of the *Dictyocoela* genus (fig. 1). However, the branching of *Dictyocoela* species relative to the most closely related microsporidia remains unclear due to several poorly supported nodes. This may be related to short internal branches contrasting with long external branches in this part of the tree. In any event, our phylogenomic analysis confirmed that *Dictyocoela* and *Nosema* genera were distantly related, indicating that feminization evolved at least twice during microsporidia evolution. Furthermore, strict vertical transmission apparently evolved at least three times independently in microsporidia, based on the phylogenetic positions of *N. granulosis*, the two *Dictyocoela* species and *H. magnivora* (fig. 1).

A distinguishing feature of the *D. muelleri* genome was a very large size (41.9 Mb) (fig. 1). Unfortunately, incompleteness of the *D. roeselum* genome precludes any conclusion on whether large genome size is a specific feature of *D. muelleri* or it may be a shared feature of *Dictyocoela* species. Interestingly, we annotated 58.8% of the *D. muelleri* genome as repeats (table 1), which is the most repeat-rich microsporidia genome sequenced to date (fig. 1). Most repeats corresponded to transposable elements, including DNA transposons (23.0%), long terminal repeat retrotransposons (8.3%) and unclassified elements (26.7%) (fig. 2). In addition, 40.2% (2,591/6,442) of *D. muelleri* annotated genes can be clustered in only 233 orthogroups. In sum, the unusually large



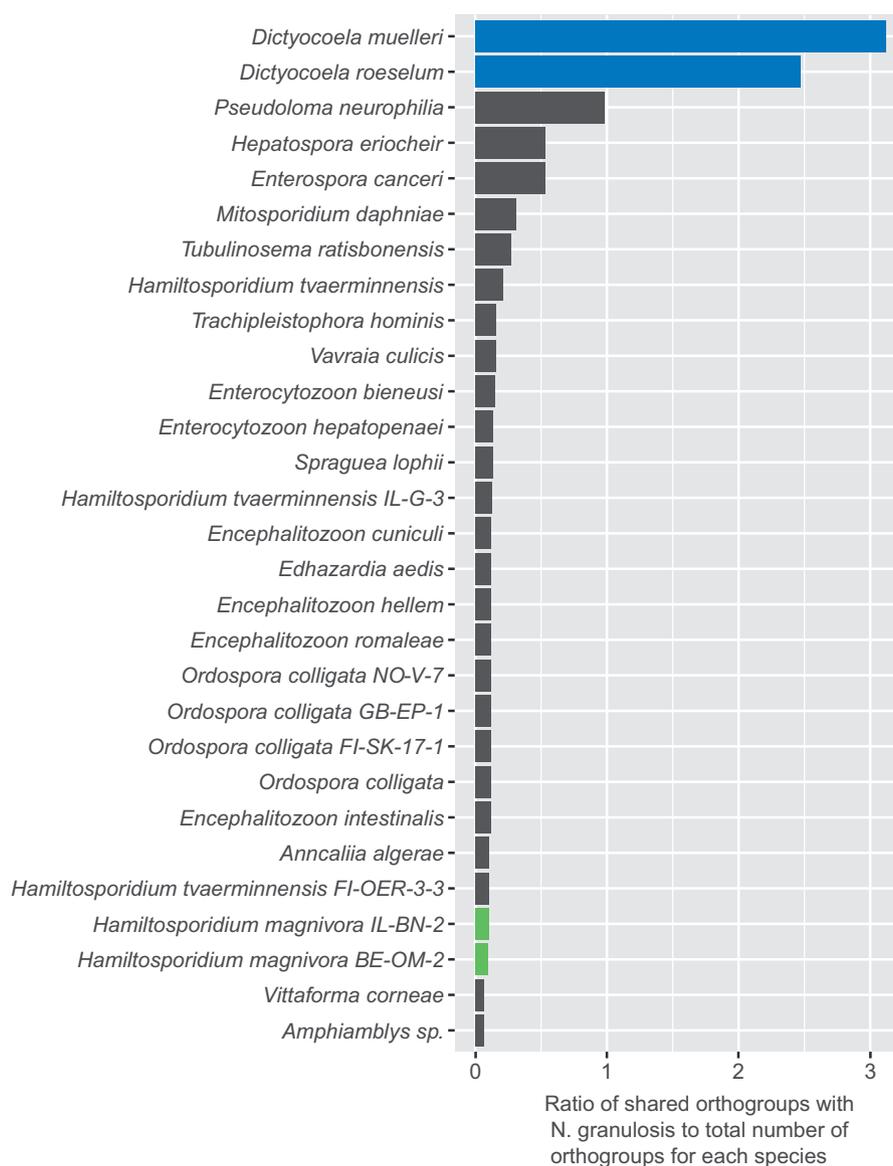
**Fig. 2.**—Repeat content of the *Nosema granulosis* (A) and *Dictyocoela muelleri* (B) genome assemblies. Long interspersed nuclear elements (LINE), long terminal repeat retrotransposons (LTR), DNA transposons (DNA), and other repeats are shown. Unmasked regions correspond to nonrepeated regions of the genome.

genome size of *D. muelleri* appears to result from transposable element expansions and massive gene duplications.

### Identification of Candidate Genes for Feminization

To identify candidate genes associated with feminization and/or strict vertical transmission, we performed an analysis of gene gain and loss during *Nosema* evolution. We searched for specific gene gains and losses in *N. granulosis*, the sole feminizing and strictly vertically transmitted species in the *Nosema* genus. We also took advantage of the *Dictyocoela* genomes to investigate genes specifically shared with *N. granulosis*, or uniquely lacking in *D. muelleri* and *N. granulosis*, as the two feminizing taxa are phylogenetically distantly related.

We performed the gene gain and loss analysis using 38 microsporidia genomes and 13 nonmicrosporidia eukaryote outgroups, based on Dollo parsimony (supplementary fig. S3, Supplementary Material online). As previously reported on smaller data sets, we observed a massive loss of genes at the early stages of microsporidia evolution presumably related to the transition to obligate intracellular lifestyle (Heinz et al. 2012; Nakjang et al. 2013). With respect to the 3,639 genes



**FIG. 3.**—Prevalence of orthogroups shared between *Nosema granulosis* and other microsporidia, but that are not present in any other *Nosema* genome. Strictly vertically transmitted and feminizing species are shown in blue. Strictly vertically transmitted albeit nonfeminizing genomes are shown in green. Other species are shown in gray. The x axis indicates the percentage of orthogroups shared with *N. granulosis* out of all orthogroups identified in each species.

of *N. granulosis*, we found that 315 single-copy genes and 398 duplicated genes (belonging to 107 orthogroups) were specific to this genome. However, only 133 of these genes were functionally annotated, mainly associated with transposition functions (supplementary table S2, Supplementary Material online). The remaining 2,926 *N. granulosis* genes clustered in 1,828 orthogroups, of which 1,738 were shared with at least one other *Nosema* species and 90 were only shared with non-*Nosema* microsporidia. Analysis of these 90 orthogroups revealed that none was shared by all three taxa of strictly vertically transmitted microsporidia (i.e., *N. granulosis*, *H. magnivora*, and at least one of the two *Dictyocoela* species), despite the inclusion

of two *H. magnivora* genomes in the analysis (Haag et al. 2020). In sharp contrast, the feminizing *D. muelleri* and *D. roeselum* were the two major non-*Nosema* microsporidia sharing orthogroups with *N. granulosis* (fig. 3). Specifically, 70 of the 90 orthogroups were shared with *D. muelleri* and/or *D. roeselum*, including 48 exclusively shared by *N. granulosis* and at least one of the two *Dictyocoela* species. Of the 48 orthogroups, 17 were functionally annotated, all of which were associated with transposition functions (supplementary table S3, Supplementary Material online). The 31 other orthogroups were without functional annotation and they represent candidate genes for feminization.

We also identified 201 orthogroups corresponding to gene losses in *N. granulosis* relative to other microsporidia. Of particular interest were seven orthogroups absent in *N. granulosis* but present in all other *Nosema* genomes (Supplementary table S4, Supplementary Material online). It turns out that four of these seven orthogroups were present in *D. muelleri* (*D. roeselium* was not included in this comparison due to assembly incompleteness), suggesting that their specific absence in *N. granulosis* may be unrelated to feminization or strict vertical transmission. By contrast, the three remaining orthogroups were also lacking in *D. muelleri*. One orthogroup (OG0000654) was involved in basic cellular functions, suggesting it can be ruled out as a candidate for feminization and/or strict vertical transmission. The two remaining orthogroups encoded proteins involved in membrane fusion (OG0000922) and DNA binding (OG0001239). Interestingly, OG0001239 is present in *H. magnivora*, suggesting that its loss in both *N. granulosis* and *D. muelleri* may be associated with feminization rather than strict vertical transmission.

As feminization is also mediated by *Wolbachia* bacterial endosymbionts, we tested the hypothesis that feminizing *Wolbachia* and microsporidia may share genomic regions acquired by horizontal gene transfer. No *Wolbachia*-like sequence was identified in *N. granulosis* and the two *Dictyocoela* genomes. In addition, no sequence was assigned to *Wolbachia* during the taxonomic assignment performed on the initial genome assemblies. Furthermore, no microsporidia-like sequence was identified in *Wolbachia* genomes. Similarly, no sequence of the bacterial genera *Cardinium*, *Rickettsia*, *Spiroplasma*, and *Arsenophonus* which include sex-ratio-distorting symbionts (Engelstädter and Hurst 2009; Cordaux et al. 2011) was detected in *N. granulosis* and the two *Dictyocoela* genomes or during the taxonomic assignment of the initial genome assemblies. These results indicated that no detectable horizontal gene transfer occurred between feminizing *Wolbachia* (or other sex-ratio-distorting symbionts) and microsporidia.

#### Draft Genome of the Amphipod Crustacean *G. roeselii*

As part of our sequencing strategy, we assembled the genome of the amphipod crustacean *G. roeselii* from each mix of the three host–microsporidia sequencing data sets we generated. We deposited in GenBank (under accession number SDVV00000000) the assembly derived from the Ou19 sample because it showed the highest contiguity and completeness of all three host assemblies (Supplementary table S5, Supplementary Material online). The final draft assembly of *G. roeselii* is composed of 1.1 million contigs and scaffolds with  $N_{50}$  of 4.8 kb and contains only 0.02% of undetermined nucleotides. The total length of the assembly was 3.2 Gb, which is close to the estimated genome size from the k-mer frequency distribution of paired-end reads (3.4 Gb). Genome completeness assessment using BUSCO revealed that 855 of

1,066 (80.2%) conserved specific Arthropoda genes are present in the final assembly of *G. roeselii* (Supplementary table S5, Supplementary Material online). Contiguity and completeness of the draft genome of *G. roeselii* appear to be fairly good, considering that it was assembled exclusively with Illumina short reads and that it is a very large and highly repeated genome (51% based on sequencing reads and 45% directly from the genome).

#### Discussion

To investigate the evolution of feminization in microsporidia, we sequenced the genomes of three species, for which feminization induction has been demonstrated (*N. granulosis*) or is strongly suspected (*D. muelleri* and *D. roeselium*). Although the *D. roeselium* assembly is partial, our assemblies of *N. granulosis* and *D. muelleri* show very good levels of completeness. With a BUSCO score of 98.2%, the *N. granulosis* assembly is even among the most complete microsporidia genomes sequenced to date. Accessing genomic information for *N. granulosis*, *D. muelleri*, and *D. roeselium* was particularly challenging because they do not produce environmental spores. Consequently, they cannot be easily isolated from host cells, unlike horizontally transmitted microsporidia whose lifecycle involves a free-living stage outside of host cells enabling the extraction of genomic DNA free of host contamination. Thus, we cosequenced the genomic DNA of the microsporidia and their host, the amphipod *G. roeselii*. As applying a standard assembling pipeline to these sequencing data sets would have yielded erroneous, chimeric assemblies, we opted for a strategy based on in silico subtraction of host reads and de novo assembly of unmapped reads, followed by taxonomic assignment and filtering using metagenomics binning and oligonucleotide composition similarity. The resulting genome assemblies we have obtained for *N. granulosis* and *D. muelleri* provide proof of principle that this strategy can be very efficient to recover the genomes of microbial endosymbionts from mixed DNA samples. Nevertheless, we acknowledge that some microsporidian contigs may have been missed by this approach. However, the use of less stringent criteria to increase the likelihood of recovering additional microsporidian contigs would have inevitably led to an increased inclusion of false positive contigs (i.e., nonmicrosporidian contigs) in the genome assemblies. As we aimed at being conservative, we opted to minimize false positive contigs. The ongoing development of new computational tools (e.g., using assembly graphs, Mallawaarachchi et al. 2020) and molecular biology protocols (e.g., single-cell sequencing, Ahrendt et al. 2018) promises further improvements toward genome assembly completeness.

Another benefit of our sequencing and assembling strategy is that it gave us the opportunity to concomitantly generate a draft assembly of the host genome. The *G. roeselii* genome appears to be very large (~3.2 Gb), which is probably

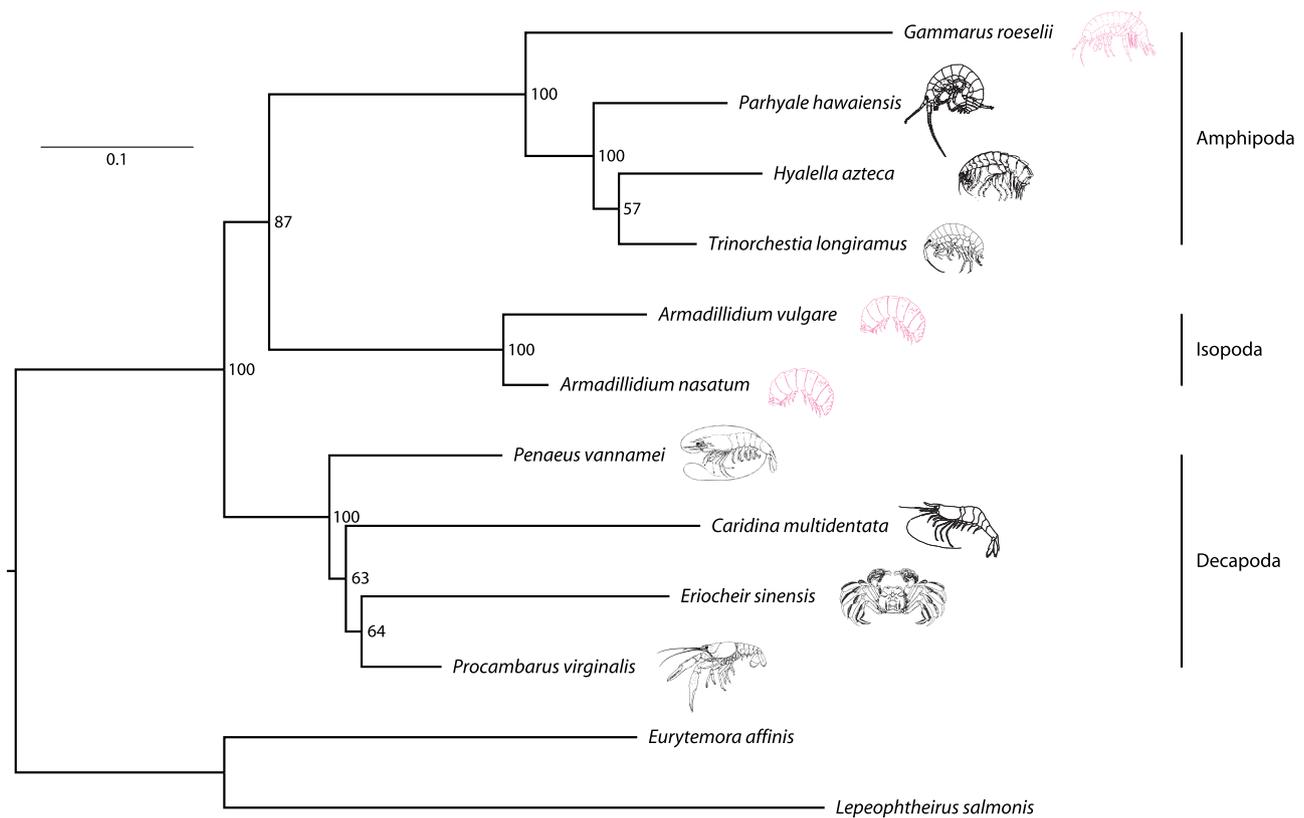
related to its very high repeat content (~50%). Yet, we obtained a quite contiguous ( $N_{50}$  of ~5 kb) and complete (BUSCO score of 80%) assembly, despite the sole use of Illumina reads. Together with *Parhyale hawaiiensis* (Kao et al. 2016), *Hyalella azteca* (Poynton et al. 2018), and *Trinorchestia longiramus* (GenBank accession number VCRD00000000.1), the *G. roeselii* genome assembly serves as a resource for studying the biology and evolution of amphipods in particular, and crustaceans in general.

The sequencing of the *N. granulosis*, *D. muelleri*, and *D. roeselium* genomes allowed us to start investigating the genetic basis and evolution of feminization in microsporidia. As symbiont-mediated feminization is also known to be caused by *Wolbachia* bacteria in terrestrial isopod crustaceans (Rigaud et al. 1997; Cordaux et al. 2011; Cordaux and Gilbert 2017), we tested the hypothesis that feminization evolved only once at the molecular genetic level and was then horizontally transferred to explain the evolution of feminization in both microsporidia and *Wolbachia*. If so, one would expect feminizing microsporidia and *Wolbachia* to exhibit genomic regions with high levels of similarity. As we did not find any such genomic region shared between feminizing microsporidia and *Wolbachia*, we conclude that there is no evidence that horizontal transfer of feminization (or vertical transmission) genes occurred between the two taxa. This is consistent with the fact that feminizing microsporidia and *Wolbachia* are not known to coinfect the same crustacean host species (experimental transfections also failed, Dunn and Rigaud 1998), although *Wolbachia* infection has been reported in some amphipod species (Cordaux et al. 2001, 2012). Thus, feminization and vertical transmission most likely evolved independently in microsporidia and *Wolbachia* at the molecular genetic level, thereby representing a remarkable evolutionary convergence of prokaryotic (*Wolbachia*) and eukaryotic (microsporidia) microorganisms.

Within microsporidia, phylogenetic analyses suggest that feminization may have evolved at least twice, as it occurs in the distantly related *Dictyocoela* and *Nosema* genera. This observation again raises the question whether feminization evolved only once at the molecular genetic level, followed by horizontal gene transfer. Interestingly, genes specifically shared by *N. granulosis* and non-*Nosema* microsporidia preferentially involve *D. muelleri* and *D. roeselium*, suggesting that horizontal gene transfer may have occurred between the two genera. If so, one would expect the genes specifically shared by *N. granulosis* and the *Dictyocoela* species to exhibit higher similarity than standard, vertically inherited genes. Consistently, we found that specifically shared genes were significantly more similar than the highly conserved BUSCO genes (Welch test,  $P=0.001$ ) (supplementary fig. S4 and table S6, Supplementary Material online). Furthermore, most genes (68%) are located on contigs that contain other microsporidia genes (supplementary table S7, Supplementary Material online) and none of the genes showed significant

nucleotide similarity to the *G. roeselii* genome assembly in BlastN searches, thus ruling out host contamination as an explanation for these observations. Overall, our results are consistent with the hypothesis of horizontal gene transfer, although it remains unclear when the transfers may have occurred and whether they involve multiple donors. Most genes specifically shared by *N. granulosis* and *Dictyocoela* with functional annotation comprised transposable elements. This is consistent with the fact that the genomes of *N. granulosis* and *D. muelleri* are particularly enriched in transposable elements, *D. muelleri* even being the most repeat-rich microsporidia genome sequenced to date. As *N. granulosis* and *D. muelleri* infect the same host species *G. roeselii*, occasional intracellular coinfection of individuals by both microsporidia species may offer opportunities for the symbionts to exchange genetic material, as proposed by the “intracellular arena” hypothesis (Bordenstein and Wernegreen 2004). Involvement of transposable elements in exchanges is perhaps not surprising, given the mobile nature of these genetic elements. Exchanges of mobile genetic elements between intracellular microorganisms have been previously reported in bacterial symbionts (Bordenstein and Wernegreen 2004; Cordaux et al. 2008) and our results suggest that they may also occur between intracellular eukaryotic symbionts. However, transposable elements may not be the sole type of genetic material exchanged between *N. granulosis* and *D. muelleri*, as we identified 31 orthogroups lacking any functional annotation. Interestingly, the occurrence of these orthogroups correlates with feminization, but not strict vertical transmission, as they are not present in *H. magnivora*. Therefore, they represent candidates for involvement in feminization, under the scenario of a single origin in microsporidia at the molecular genetic level.

Alternatively to evolution through gain-of-function, it is also conceivable that loss-of-function underlies the evolution of feminization in *N. granulosis* and *D. muelleri*. For example, the loss of a gene whose product is involved in molecular interactions with host cells may cause regulatory or physiological changes in the host. If such a change happened to alter the fate of host sexual differentiation and drive feminization, it could have been selected by vertically transmitted microsporidia. To investigate this possibility, we analyzed genes lacking in *N. granulosis* and *D. muelleri* while present in other microsporidia. We identified two orthogroups constituting good candidates, inferred to encode proteins involved in membrane fusion (OG0000922) and DNA binding (OG0001239). It is noteworthy that these two candidates encode relevant functions in the perspective of a potential involvement in molecular interactions with the host. Strikingly, candidate genes for feminization by *Wolbachia* bacteria in the isopod *Armadillidium vulgare* have also been inferred to involve functions such as DNA binding and membranes (Badawi et al. 2018). Therefore, the independent evolution of feminization in microsporidia and *Wolbachia* may involve the parallel recruitment of analogous functions.



**FIG. 4.**—Maximum-likelihood phylogenetic tree of malacostraca based on 12 single-copy nuclear genes extracted from nine sequenced genomes available in NCBI and *Gammarus roeselii* (this study). Pink, species reported to be infected by feminizing endosymbionts (microsporidia in *G. roeselii* and *Wolbachia* in the two isopods). Bootstrap values (%) are indicated at each node. Scale bar indicates changes per site.

In conclusion, the sequencing of the genomes of three feminizing microsporidia represents a milestone toward the establishment of the amphipod *G. roeselii* and its microsporidia symbionts as a new model to study the evolution of symbiont-mediated feminization, a remarkable evolutionary convergence of prokaryotic (*Wolbachia*) and eukaryotic (microsporidia) microorganisms (fig. 4).

## Materials and Methods

### Animal Sampling

We sampled 165 *G. roeselii* reproductive pairs in the Ouche river at Trouhans (Côte d'Or, France). *Gammarus roeselii*, like many gammarids, form precopula pairs several days before mating, during which the male holds on to the female to ensure paternity of the future brood (Pöckl 1993). Therefore, the 165 females were about to lay eggs when they were sampled. Pairs of individuals were isolated in individual containers and checked daily until eggs were laid in the female's brood pouch. Eggs were flushed out from the pouch with a gentle water flow and immediately transferred to an Eppendorf tube prior to freezing in liquid nitrogen. Eggs with more than eight cells were discarded. The mothers were then anesthetized and dissected. One ovary of each female was

transferred into pure ethanol and then used to screen for microsporidia infection, using a standard PCR assay targeting the microsporidian *SSU rRNA* gene, complemented by Sanger sequencing of PCR products, as described previously (Quiles et al. 2019). PCR screening and sequencing were performed using the microsporidia-specific primers V1f and 530r (Haine et al. 2004).

### Genome Sequencing, Assembly, and Annotation of Microsporidia

Genomic DNA was extracted from eggs using the Qiagen DNeasy Blood and Tissue kit, according to the protocol for animal tissues (3 h of incubation in proteinase K at 56 °C and 30 min of digestion with RNase at 37 °C). Eggs with a maximum of eight cells were used for genome sequencing instead of ovaries to maximize the ratio of symbiont to host DNA. Indeed, infected eggs contain on average  $8 \pm 1.5$  *D. roeselium* cells (maximum 115),  $173 \pm 12.3$  *D. muelleri* cells (maximum 600), and  $122 \pm 8.2$  *N. granulosis* cells (maximum 275) (Haine et al. 2004).

For each sample, a paired-end library with ~500-bp inserts was prepared using the NEB Next Ultra II FS DNA Library Prep Kit for Illumina and sequenced on Illumina HiSeq2500

(2× 125 bp) and MiSeq (2× 300 bp) instruments for *N. granulosis* (Ou3 and Ou53) and on an Illumina HiSeq3000 instrument (2× 150 bp) for *D. muelleri* (Ou54) and *D. roeselium* (Ou19). We obtained a total of 584,354,251 paired reads for *N. granulosis* (104,526,454 with MiSeq and 479,827,797 with HiSeq), 821,178,855 paired reads for *D. muelleri*, and 797,742,808 paired reads for *D. roeselium*. Following adapter removal by the sequencing company, data quality was assessed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>; last accessed December 3, 2020).

Assemblies of symbiont genomes were performed in three steps using a subtraction approach. First, an assembly of each mixed host–symbiont data set was obtained using MaSuRCA v3.2.7 (Zimin et al. 2013) with Illumina only assembly parameters and raw reads, resulting in three genome assemblies: *G. roeselii/N. granulosis*, *G. roeselii/D. muelleri*, and *G. roeselii/D. roeselium*. Next, reads of each data set were mapped to another mixed genome using Bowtie2 v2.2.9 (Langmead and Salzberg 2012). Unmapped reads were extracted using Samtools v1.5 (Li et al. 2009) and assembled using SPAdes v3.12.0 (Bankevich et al. 2012), with default parameters except -k 55,77,99,127. Finally, each symbiont assembly was filtered with a combination of taxonomic assignment using BlobTools v1.0 (Laetsch and Blaxter 2017), metagenomic binning using MaxBin v2.2.4 (Wu et al. 2014), and oligonucleotide composition similarity using PhylOligo v1.0 (Mallet et al. 2017). Completeness of the assemblies was evaluated based on expectations of gene content from near-universal single-copy orthologs using BUSCO v3.0 (Simão et al. 2015) with the microsporidia data set.

Gene structures were predicted using GeneMarkS 4.30 with intronless eukaryotic parameter (Besemer and Borodovsky 2005). The functional annotation for predicted proteins was performed using the UniProtKB/Swiss-Prot database (UniProt Consortium 2016) and Interproscan (Jones et al. 2014). Both annotations were merged using Blast2GO 3.0 (Götz et al. 2008). Repetitive elements were identified using RepeatModeler v1.0.8 and the resulting consensus sequences were mapped across the genome using RepeatMasker v1.323 (<http://www.repeatmasker.org>; last accessed December 3, 2020).

### Identification of Orthogroups and Phylogenetic Analysis

Orthologous gene families were identified using OrthoFinder v2.2.7 (Emms and Kelly 2015). The analysis was performed using 39 microsporidian species (supplementary table S8, Supplementary Material online). Phylogenetic reconstruction was performed with one representative genome per species using the 33 single-copy orthologs identified at least in 32 genomes. Protein sequences were aligned using MAFFT v7.299b in automatic mode (Kato and Standley 2013). Alignments were trimmed using Gblocks (Talavera and

Castresana 2007) (-t = p; -p = n; -b3 = 8; -b4 = 2; -b5 = h). Each alignment was then concatenated into a single alignment using FASconCAT 1.0 (Kück and Longo 2014) resulting in a total of 9,142 aligned amino acid sites. The best fitting substitution model selection and the phylogenetic reconstruction using maximum likelihood were performed with iqTree (Nguyen et al. 2015). Nodal support was assessed using bootstrap and SH-aLRT (both with 1,000 replicates). The endoparasitic fungus *Rozella allomyces* was used to root the resulting tree.

### Gene Gain and Loss Analysis and Comparison with *Wolbachia*

To construct protein families, we selected the previously used microsporidian genomes minus *D. roeselium* (due to genome assembly incompleteness) plus 13 representative fungal and other related genomes (supplementary table S9, Supplementary Material online) and performed protein orthology clustering using OrthoFinder, as previously described. Evolution of protein families was modeled using the OrthoFinder-generated orthology groups and the Dollo parsimony method of the Count software (Csuo 2010). Orthogroups absent in *N. granulosis* but present in all other *Nosema* species were further subjected to TBlastN searches against the *N. granulosis* genome assembly to avoid false positives. Potential horizontal gene transfer between *Wolbachia* bacteria and microsporidia was investigated using BlastP between the proteomes of all *Wolbachia* genomes available in NCBI as of April 2019 and the proteomes of all microsporidia listed in supplementary table S9, Supplementary Material online. Of note, the *Wolbachia* genome set included the feminizing strain wVulC of *Armadillidium vulgare* (Cordaux et al. 2004) (GenBank accession number ALWU000000000). Additional searches were also performed with the 207 genomes available in GenBank for the following bacterial genera: *Cardinium*, *Rickettsia*, *Spiroplasma*, and *Arsenophonus* (supplementary table S10, Supplementary Material online).

### Assembly and Analysis of the *G. roeselii* Genome

As previously explained, we obtained three assemblies of the genome of *G. roeselii*. We selected the assembly resulting from the *D. roeselium* data set because it showed the highest contiguity and completeness of all three host assemblies. To remove contaminants from the genome assembly, contigs were searched for similarities against the Uniref90 database (release October 2018) (Suzek et al. 2015) using diamond v0.9.12 (Buchfink et al. 2015). Sequencing coverage was estimated by mapping reads to the genome with Bowtie2. Contigs were then visualized with Blobtools v0.9.17 using Taxon-annotated-GC-Coverage (TAGC) plots (Laetsch and Blaxter 2017). We removed all contigs annotated as prokaryotic sequences or assigned to microsporidia. Unassigned

contigs were retained, as the absence of hits for these contigs could be explained by the lack of sequenced genomes from closely related species in public databases. Assembly completeness was evaluated based on expectations of gene content from near-universal single-copy orthologs using BUSCO v3.0 (Simão et al. 2015) with the Arthropoda data set. The proportion of repetitive elements in the host genome was evaluated using RepeatExplorer (Novák et al. 2013) directly from raw reads and from the filtered assembly using RepeatMasker v1.323.

We constructed a phylogeny of malacostraca by identifying orthologous gene families of ten species with available genomes using BUSCO v3.0 (Simão et al. 2015). Phylogenetic reconstruction was performed using the 12 single-copy orthologs identified in the ten genomes. Protein sequences were aligned using MAFFT v7.299b in automatic mode (Kato and Standley 2013) and trimmed using Gblocks (Talavera and Castresana 2007) (-t = p; -p = n; -b3 = 8; -b4 = 2; -b5 = h). Each alignment was then concatenated into a single alignment using FASconCAT 1.0 (Kück and Longo 2014), resulting in a total of 3,733 aligned amino acid sites. The best fitting substitution model was selected with ProtTest v3.4.1 (Darriba et al. 2011). The phylogenetic tree was reconstructed using maximum likelihood with RAxML v8.2.9 under the VT + GAMMA model (Stamatakis 2014). Nodal support was assessed using 1,000 bootstrap replicates. The related *Eurytemora affinis* and *Lepeophtheirus salmonis* crustacean species were used to root the resulting tree.

## Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

## Acknowledgments

We thank Bouziane Moumen for assistance with bioinformatics analyses and the Roscoff Bioinformatics platform ABIMS (<http://abims.sb-roscoff.fr>; last accessed December 3, 2020) for providing additional computational resources. This work was funded by Agence Nationale de la Recherche (Grant No. ANR-15-CE32-0006 [CytoSexDet] to R.C. and T.R.), 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.

## Data Availability

Genome projects are available for *Nosema granulosis*, *Dictyocoela muelleri*, *Dictyocoela roeselium*, and *Gammarus roeselii* at DDBJ, EMBL, and GenBank under accessions

SBJO00000000, SDVU00000000, SDVW00000000, and SDVV00000000, respectively. Raw sequence data sets are available in the SRA repository under accession numbers: SRR8481863–SRR8481878, SRR8494490–SRR8494493, and SRR8494486–SRR8494489.

## Literature Cited

- Ahrendt SR, et al. 2018. Leveraging single-cell genomics to expand the fungal tree of life. *Nat Microbiol.* 3(12):1417–1428.
- Bacela-Spychalska K, et al. 2018. Europe-wide reassessment of Dictyocoela (Microsporidia) infecting native and invasive amphipods (Crustacea): molecular versus ultrastructural traits. *Sci Rep.* 8(1):8945.
- Badawi M, Moumen B, Giraud I, Grève P, Cordaux R. 2018. Investigating the molecular genetic basis of cytoplasmic sex determination caused by *Wolbachia* endosymbionts in terrestrial isopods. *Genes* 9(6):290.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477.
- Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Res.* 33(Web Server):W451–W454.
- Bordenstein SR, Wernegreen JJ. 2004. Bacteriophage flux in endosymbionts (*Wolbachia*): infection frequency, lateral transfer, and recombination rates. *Mol Biol Evol.* 21(10):1981–1991.
- Bouchon D, Cordaux R, Grève P. 2008. Feminizing *Wolbachia* and the evolution of sex determination in isopods. Vol. 3. Boca Raton (FL): CRC Press/Taylor and Francis Group LLC.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12(1):59–60.
- Chebbi MA, et al. 2019. The Genome of *Armadillidium vulgare* (Crustacea, Isopoda) provides insights into sex chromosome evolution in the context of cytoplasmic sex determination. *Mol Biol Evol.* 36(4):727–741.
- Chen Y. P, et al. 2013. Genome sequencing and comparative genomics of honey bee microsporidia, *Nosema apis* reveal novel insights into host-parasite interactions. *BMC Genomics.* 14(1):451.
- Cordaux R, Bouchon D, Grève P. 2011. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genetics.* 27(8):332–341.
- Cordaux R, Gilbert C. 2017. Evolutionary significance of *Wolbachia*-to-animal horizontal gene transfer: female sex determination and the element in the isopod *Armadillidium vulgare*. *Genes* 8(7):186.
- Cordaux R, Michel-Salzat A, Bouchon D. 2001. *Wolbachia* infection in crustaceans: novel hosts and potential routes for horizontal transmission. *J Evol Biol.* 14(2):237–243.
- Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D. 2004. Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. *Heredity* 93(1):78–84.
- Cordaux R, et al. 2008. Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. *Mol Biol Evol.* 25(9):1889–1896.
- Cordaux R, et al. 2012. Widespread *Wolbachia* infection in terrestrial isopods and other crustaceans. *Zookeys* 176:123–131.
- Cormier A, Wattier R, Teixeira M, Rigaud T, Cordaux R. 2018. The complete mitochondrial genome of *Gammarus roeselii* (Crustacea, Amphipoda): insights into mitogenome plasticity and evolution. *Hydrobiologia* 825(1):197–210.
- Cornman RS, et al. 2009. Genomic analyses of the microsporidian *Nosema ceranae*, an emergent pathogen of honey bees. *PLoS Pathog.* 5(6):e1000466.
- Csuos M. 2010. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26(15):1910–1912.

- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27(8):1164–1165.
- Desjardins CA, et al. 2015. Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat Commun.* 6(1):7121.
- Dunn A, Terry R, Smith J. 2001. Advances in parasitology. *Adv Parasitol.* 48:57–100.
- Dunn AM, Rigaud T. 1998. Horizontal transfer of parasitic sex ratio distorters between crustacean hosts. *Parasitology* 117(1):15–19.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16(1):157.
- Engelstädter J, Hurst GDD. 2009. The ecology and evolution of microbes that manipulate host reproduction. *Annu Rev Ecol Evol Syst.* 40(1):127–149.
- Frixione E, et al. 1992. Dynamics of polar filament discharge and sporoplasm expulsion by microsporidian spores: dynamics of microsporidian spore discharge. *Cell Motil Cytoskeleton* 22(1):38–50.
- Götz S, et al. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36(10):3420–3435.
- Grabowski M, Krzywoznik P, Rewicz T, Bacela-Spychalska K, Wattier R. 2017. *Gammarus roeselii* Gervais, 1835 (Gammaridae) in Western and Central Europe: post-glacial colonisation or human mediated introduction? *Biodivers J.* 8:525–526.
- Grabowski M, Mamos T, Bacela-Spychalska K, Rewicz T, Wattier RA. 2017. Neogene paleogeography provides context for understanding the origin and spatial distribution of cryptic diversity in a widespread Balkan freshwater amphipod. *PeerJ.* 5:e3016.
- Haag KL, et al. 2020. Microsporidia with vertical transmission were likely shaped by nonadaptive processes. *Genome Biol Evol.* 12(1):3599–3614.
- Haine ER, Motreuil S, Rigaud T. 2007. Infection by a vertically-transmitted microsporidian parasite is associated with a female-biased sex ratio and survival advantage in the amphipod *Gammarus roeselii*. *Parasitology* 134(10):1363–1367.
- Haine ER, et al. 2004. Coexistence of three microsporidia parasites in populations of the freshwater amphipod *Gammarus roeselii*: evidence for vertical transmission and positive effect on reproduction. *Int J Parasitol.* 34(10):1137–1146.
- Heinz E, et al. 2012. The genome of the obligate intracellular parasite *Trachipleistophora hominis*: new insights into microsporidian genome dynamics and reductive evolution. *PLoS Pathog.* 8(10):e1002979.
- Jahnke M, Smith JE, Dubuffet A, Dunn AM. 2013. Effects of feminizing microsporidia on the masculinizing function of the androgenic gland in *Gammarus duebeni*. *J Invertebr Pathol.* 112(2):146–151.
- Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
- Kao D, et al. 2016. The genome of the crustacean *Parhyale hawaiiensis*, a model for animal development, regeneration, immunity and lignocellulose digestion. *Elife* 5:e20062.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30(4):772–780.
- Kück P, Longo GC. 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Front Zool.* 11(1):81.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Research* 6:1287.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9(4):357–359.
- Leclercq S, et al. 2016. Birth of a W sex chromosome by horizontal transfer of *Wolbachia* bacterial symbiont genome. *Proc Natl Acad Sci U S A.* 113(52):15036–15041.
- Li H, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Mallawaarachchi V, Wickramarachchi A, Lin Y. 2020. GraphBin: refined binning of metagenomic contigs using assembly graphs. *Bioinformatics* 36(11):3307–3313.
- Mallet L, Bitard-Feildel T, Cerutti F, Chiapello H. 2017. PhylOligo: a package to identify contaminant or untargeted organism sequences in genome assemblies. *Bioinformatics* 33(20):3283–3285.
- Nakjang S, et al. 2013. Reduction and expansion in microsporidian genome evolution: new insights from comparative genomics. *Genome Biol Evol.* 5(12):2285–2303.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29(6):792–793.
- Pan G, et al. 2013. Comparative genomics of parasitic silkworm microsporidia reveal an association between genome expansion and host adaptation. *BMC Genomics.* 14(1):186.
- Pichon S, et al. 2012. The expression of one ankyrin pk2 allele of the WO prophage is correlated with the *Wolbachia* feminizing effect in isopods. *BMC Microbiol.* 12(1):55.
- Pöckl M. 1993. Reproductive potential and lifetime potential fecundity of the freshwater amphipods *Gammarus fossarum* and *G. roeselii* in Austrian streams and rivers. *Freshwater Biol.* 30(1):73–91.
- Poynton HC, et al. 2018. The toxicogenome of *Hyalomma azteca*: a model for sediment ecotoxicology and evolutionary toxicology. *Environ Sci Technol.* 52(10):6009–6022.
- Quiles A, et al. 2019. Microsporidian infections in the species complex *Gammarus roeselii* (Amphipoda) over its geographical range: evidence for both host–parasite co-diversification and recent host shifts. *Parasit Vectors* 12(1):327.
- Rigaud T, Juchault P, Mocquard J-P. 1997. The evolution of sex determination in isopod crustaceans. *BioEssays* 19(5):409–416.
- Rodgers-Gray TP, Smith JE, Ashcroft AE, Isaac RE, Dunn AM. 2004. Mechanisms of parasite-induced sex reversal in *Gammarus duebeni*. *Int J Parasitol.* 34(6):747–753.
- Rutrecht ST, Brown MJF. 2008. Within colony dynamics of *Nosema bombi* infections: disease establishment, epidemiology and potential vertical transmission. *Apidologie* 39(5):504–514.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Suzek BE, et al. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31(6):926–932.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 56(4):564–577.
- Terry RS, Smith JE, Dunn AM. 1998. Impact of a novel, feminising microsporidium on its crustacean host. *J Eukaryot Microbiol.* 45(5):497–501.
- Terry RS, et al. 1999. Ultrastructural characterisation and molecular taxonomic identification of *Nosema granulosis* n. sp., a transovarially transmitted feminising (TTF) microsporidium. *J Eukaryot Microbiol.* 46(5):492–499.
- Terry RS, et al. 2004. Widespread vertical transmission and associated host sex-ratio distortion within the eukaryotic phylum Microspora. *Proc R Soc Lond Ser B Biol Sci.* 271(1550):1783–1789.
- Traver BE, Fell RD. 2012. Low natural levels of *Nosema ceranae* in *Apis mellifera* queens. *J Invertebr Pathol.* 110(3):408–410.

- UniProt Consortium. 2016. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 45:D158–D169.
- Vávra J, Lukeš J. 2013. Advances in parasitology. *Adv Parasitol.* 82:253–319.
- Weedall RT, Robinson M, Smith JE, Dunn AM. 2006. Targeting of host cell lineages by vertically transmitted, feminising microsporidia. *Int J Parasitol.* 36(7):749–756.
- Weidner E. 1972. Ultrastructural study of microsporidian invasion into cells. *Z Parasitenkd.* 40(3):227–242.
- Weidner E. 1976. The microsporidian spore invasion tube. The ultrastructure, isolation, and characterization of the protein comprising the tube. *J Cell Biol.* 71(1):23–34.
- Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nat Rev Microbiol.* 6(10):741–751.
- Wu Y-W, Tang Y-H, Tringé SG, Simmons BA, Singer SW. 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2(1):26.
- Xu J, et al. 2016. The genome of *Nosema* sp. isolate YNPr: a comparative analysis of genome evolution within the *Nosema/Vairimorpha* clade. *PLoS One* 11(9):e0162336.
- Zilio G, Thiévent K, Koella JC. 2018. Host genotype and environment affect the trade-off between horizontal and vertical transmission of the parasite *Edhazardia aedis*. *BMC Evol Biol.* 18(1):59.
- Zimin AV, et al. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29(21):2669–2677.
- Zug R, Hammerstein P. P. 2012. Still a host of hosts for *Wolbachia*: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS One* 7(6):e38544.

**Associate editor:** Nancy Moran