Host-parasite dynamics in a natural system: revealing the evolutionary change in parasite populations infecting *Daphnia*

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Enrique González Tortuero

May 2016

Diese Dissertation wurde angefertigt
unter der Leitung von Prof. Dr. Justyna Wolinska
am Department Biologie II
an der Ludwig-Maximilians-Universität München

Erstgutachter: Prof. Dr. Justyna Wolinska

Zweitgutachter: Prof. Dr. Wolfgang Stephan

Tag der Abgabe: 12.05.2016

Tag der mündlichen Prüfung: 21.07.2016

To my grandfather,
Clemente Tortuero Bayo (1934 – 2013),
in memoriam

"*Ad inquisitionem tantorum aetas una non sufficit, ut tota caelo uacet* (...) *Veniet tempus quo posteri nostri tam aperta nos nescisse mirentur*."

"A single lifetime is not sufficient for the investigation of so great of things, even though the whole [lifetime] be at leisure for the heavens (...) A time will come when [in which] our descendants could be amazed that we did not know things so evident [to them]."

(Seneca, Nat. Quaest., 7, XXV)

Acknowledgements

I would like to thank my supervisor, Prof. Dr. Justyna Wolinska, for provided help to perform this PhD thesis possible. Her support and persistence over these years is deeply meritorious. She always provided her opinion to guide me in the main topic of this thesis, i.e., *Daphnia*–microparasites interactions (even in situations where we were in disagreement). Together with Dr. Sabine Gießler, Prof. Dr. Adam Petrusek and Inbar Maayan, they read many versions of my manuscripts and helped me with the English spelling. Additionally, I like to thank their group (Dr. Ramsy Agha, Dr. Elisabeth Funke, Federico Castro-Monzón, Johanna Griebel, Manja Saebelfeld and other people who were temporarily in the group) for discussing, sharing the scientific stories and for provided help (in general).

I thank all the people I met during my time in the Ludwig-Maximilians-University Munich, especially the Master Students from the Evolution, Ecology and Systematics Program. Additionally, I like to thank also the people from the Berlin Centre for Genomics in Biodiversity (BeGenDiv) and from the Leibniz-Institute of Freshwater Ecology and Inland Fisheries, both in Berlin. To all of them, thanks for the patience and for all the things I learnt (not only in the scientific aspect but also in the personal one).

Special thanks are deserved to the Dr. Michael Monaghan's lab team (Leibniz-Institute of Freshwater Ecology and Inland Fisheries), Dr. Alexandro Rodriguez-Rojas and Dr. Olga Makarova (Freie Universität Berlin) for the moral support, the social life and their interesting biological discussions. Although these discussions are not related with the main content of this PhD thesis, this collaboration allowed for several ideas that will improve my scientific career (or I hope so).

Finally, I thank my family and my good friends who supported me in the distance, which I miss them (a bier could go for you, guys!). I also thank Prof. Dr. Jose L. Bella from Autonomous University of Madrid and Dr. Francisco J. López de Saro (former PI in the Astrobiology Centre) for encouraging me to apply to this PhD position.

Declaration of contributions as a co-author

In this dissertation, I present the work of my doctoral research from January 2013 to April 2016. I alone conducted all of the bioinformatic and statistical analyses, as well as I also led the paper writing for all papers/manuscripts (Chapters 2, 3 and 4). Dr. Sabine Gießler provided feedback and revised some of the chapters, Prof. Dr. Adam Petrusek provided feedback and revised all the chapters, whereas Prof. Dr. Justyna Wolinska helped to improve and revised all the chapters. The work has resulted in three publications (Chapters 2, 3 and 4). The doctoral thesis consists of three manuscripts:

- González-Tortuero E., Rusek J., Petrusek A., Gießler S., Lyras D., Grath S., Castro-Monzón F., Wolinska J. (2015) The Quantification of Representative Sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*. *Molecular Ecology Resources*, 15: 1385 – 1395.
- González-Tortuero E., Rusek J., Turko P., Petrusek A., Maayan I., Pialek L., Tellenbach C., Gießler S., Spaak P., Wolinska J. (2016) *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. *Zoology*. In press.
- González-Tortuero E., Rusek J., Maayan I., Petrusek A., Pialek L., Laurent S., Wolinska J. (2016) Genetic diversity of two *Daphnia*-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region. *Parasites & Vectors*, 9: 293.

Table of contents

Abstract

The Red Queen Hypothesis postulates that reciprocal selection arising from host-parasite interactions should accelerate evolutionary rates through the need for continual adaptation and counter-adaptation. A process driving such rapid reciprocal adaptation is referred to as negative frequency-dependent selection, in which the most common genotypes decrease over time because they have a higher probability of becoming infected by coevolving parasites. This proposed mechanism of host-parasite coevolution was commonly tested in laboratory experiments under controlled conditions. Regarding field investigations of natural populations, temporal changes in relative frequencies of genotypes were mostly tested for host only, because tracking parasite dynamics over time remained difficult. As parasite population dynamics are highly sensitive to environmental changes, studies under natural conditions are essential to understand host-parasite coevolution.

The commonly explored model system to address coevolutionary questions are the water fleas of the genus *Daphnia* and their microparasites. In this PhD thesis, I analysed the population structure of two major microparasites of *Daphnia*: *Caullerya mesnili* (Chapters 2 and 3) and microsporidia (Chapter 4). First, in Chapter 2, I developed a new bioinformatic pipeline to analyse molecular data generated by next-generation-sequencing (NGS) platforms. *C. mesnili* populations from different water reservoirs in the Czech Republic were sequenced at the first internal transcribed spacer (ITS1) of the ribosomal gene cluster, analysed with this new pipeline and compared with published results from the same populations but using cloning and Sanger sequencing method. I detected that relative frequencies of *C. mesnili* ITS1 sequence types were similar when compared to other sequencing methods, thereby validating the bioinformatic pipeline, and showing the suitability of 454 platform to perform population biology analyses. After this validation, in Chapter 3, I analysed the population dynamics and host-genotype specificity of *C. mesnili*, in long-term samples collected from a single lake, and based on the sequence variations in the ITS1 region. I found that the most abundant *C. mesnili* ITS1 sequence type decreased, while rare sequences increased over the course of the study (4 years). The observed pattern is consistent with the negative frequency-dependent selection. However, only a weak signal of host-genotype specificity between *C. mesnili* and *Daphnia* genotypes was detected, which supports the lack of host-genotype specificity in this system. Finally, in Chapter 4, I described the patterns of geographical population structure, intraspecific genetic variation, and recombination of two *Daphnia*-infecting microsporidia: *Berwaldia schaefernai* and the unknown microsporidium MIC1. These patterns were used to predict the existence of secondary hosts in the life cycle of these microsporidia. I observed little variation among *B.*

schaefernai parasite strains infecting different host populations; in contrast, there was significant genetic variation among populations of MIC1. Additionally, ITS genetic diversity was lower in *B. schaefernai* than in MIC1. These findings suggest that the presumed secondary host for *B. schaefernai* is expected to be mobile, while in MIC1 the secondary host (if exists) does not appear to facilitate dispersal to the same degree. Finally, recombination analyses indicated cryptic sex in *B. schaefernai* and pure asexuality in MIC1. All these findings enable a more comprehensive understanding of the biology of *Daphnia*-infecting microparasites and the genetic basis of *Daphnia*-microparasites coevolution in natural populations.

Zusammenfassung

Die “Red Queen”-Hypothese besagt, dass die wechselseitige Selektion in Wirt-Parasit-Interaktionen und die hierdurch bedingte Notwendigkeit kontinuierlicher Anpassung zu einer Beschleunigung der Evolution solcher Systeme führen sollte. Ein Prozess, der eine solch rapide wechselseitige Anpassung antreibt, wird als negativ frequenzabhängige Selektion bezeichnet. In diesem Fall nimmt die Frequenz der häufigsten Genotypen mit der Zeit ab, da sie mit höherer Wahrscheinlichkeit von koevolvierenden Parasiten infiziert werden. Dieser angenommene Mechanismus für Wirt-Parasit-Koevolution wurde häufig in Laborexperimenten unter kontrollierten Bedingungen untersucht. Im Hinblick auf Freilandversuche mit natürlichen Populationen wurden zeitliche Veränderungen der relativen Genotypfrequenzen überwiegend für Wirtsorganismen untersucht, da die Untersuchung der Parasitendynamik im Zeitverlauf kompliziert blieb. Da die Dynamik von Parasitenpopulationen hochempfindlich auf Umweltveränderungen reagiert, sind Studien unter natürlichen Bedingungen essentiell für das Verständnis von Wirt-Parasit-Koevolution.

Das für Fragen der Koevolution am häufigsten untersuchte Modellsystem sind die Wasserflöhe der Gattung *Daphnia* und ihre Mikroparasiten. In dieser Doktorarbeit untersuchte ich die Populationsstruktur zweier wichtiger Mikroparasiten von *Daphnia*: *Caullerya mesnili* (Kapitel 2 und 3) und Microsporidia (Kapitel 4). Zuerst entwickelte ich eine neue bioinformatische Pipeline für die Analyse von durch Next Generation Sequencing (NGS)-Plattformen generierte molekulare Daten (Kapitel 2). Der erste interne transkribierte Spacer (ITS1) des ribosomalen Genclusters von *C. mesnili*–Populationen aus verschiedenen Wasserreservoiren in der Tschechischen Republik wurde sequenziert, mit der neuen Pipeline analysiert und mit auf Klonierung und Sanger-Sequenzierung beruhenden publizierten Ergebnissen derselben Populationen verglichen. Die relativen Frequenzen der *C. mesnili* ITS1-Sequenztypen waren mit den Ergebnissen der anderen Sequenziermethoden vergleichbar, wodurch die bioinformatische Pipeline sowie die Eignung der 454-Plattform für die Durchführung populationsbiologischer Analysen bestätigt wurden. Kapitel 3 beschreibt die Analyse der Populationsdynamiken und der Wirtsgenotyp-Spezifität von *C. mesnili* mithilfe von Langzeitproben eines einzelnen Sees auf der Basis von Sequenzvariationen in der ITS1-Region. Ich konnte zeigen, dass der häufigste *C. mesnili* ITS1-Sequenztyp im Verlauf der Studie (vier Jahre) abnahm, während seltene Sequenzen zunahmen. Das beobachtete Muster stimmt mit negativer frequenzabhängiger Selektion überein. Allerdings wurde nur ein schwaches Signal für Wirtsgenotyp-Spezifität zwischen *C. mesnili*- und *Daphnia*-Genotypen nachgewiesen, was das Fehlen von Wirtsgenotyp-Spezifität in diesem System unterstützt. Kapitel 4 beschreibt die geografische Populationsstruktur,

intraspezifische genetische Variation und Rekombination zweier *Daphnia*-infizierender Mikrosporidien, *Berwaldia schaefernai* und der unbekannten Mikrosporidium-Art MIC1. Diese Muster wurden verwendet, um die Existenz sekundärer Wirte im Lebenszyklus dieser Mikrosporidien vorherzusagen. Ich beobachtete geringe Variation zwischen *B. schaefernai*-Parasitenstämmen, die verschiedene Wirtspopulationen infizieren. Im Gegensatz dazu zeigte sich signifikante genetische Variation zwischen MIC1-Populationen. Zusätzlich war die genetische Diversität von ITS in *B. schaefernai* geringer als in MIC1. Diese Ergebnisse legen nahe, dass der angenommene sekundäre Wirt für *B. schaefernai* mobil ist, während der sekundäre Wirt von MIC1 (falls vorhanden) anscheinend die Verbreitung nicht im selben Maß ermöglicht. Abschließend deuten Rekombinationsanalysen auf verborgene geschlechtliche Fortpflanzung in *B. schaefernai* und reine Asexualität in MIC1 hin. Alle diese Ergebnisse verhelfen zu einem umfassenderen Verständnis der Biologie von *Daphnia*-infizierenden Mikroparasiten und der genetischen Basis von *Daphnia*-Mikroparasiten-Koevolution in natürlichen Populationen.

Chapter 1 – Introduction

1.1. Introductory notions to the Red Queen Hypothesis

One of the most important questions in Evolutionary Biology is why sexual reproduction exists even when it usually implies a 50% fitness disadvantage compared to asexual reproduction (Dawkins, 1989; Maynard Smith, 1978), assuming that there is no sex ratio distortion (Hurst and Pomiankowski, 1991; Werren, 1987). One of the widely accepted explanations for this is the Red Queen Hypothesis (RQH). This hypothesis stemmed from the observation that most population show constant extinction rates over time, which led to a proposal that extinction is more related to biotic factors, rather than abiotic ones (Van Valen, 1973). However, RQH is now more associated with debates related with the parasite-driven evolution of sex (reviewed in Brockhurst et al., 2014) and host-parasite coevolution (reviewed in Lively, 2010). According to the RQH, reciprocal selection arising from host-parasite interactions should accelerate evolutionary rates through the need for continual adaptation and counter-adaptation. Such rapid reciprocal adaptation can be driven by directional selection or negative frequency-dependent selection, being called “arms race dynamics” (ARD) and “fluctuating selection dynamics” (FSD) respectively (Brockhurst and Koskella, 2013; Gaba and Ebert, 2009; Gandon et al., 2008). In ARD, recurrent selective sweeps of novel host resistance and parasite infectivity alleles occur through time, leading to increases in the parasite's host range and host resistance traits (Gandon et al., 2008). Alternatively, in FSD, common genotypes of a given host have a higher probability of becoming infected by coevolving parasites than do rare genotypes. In such case, rare uninfected host genotypes display a fitness advantage and eventually take over the previously common host genotypes. At this point, parasite genotypes that are able to infect those surviving host genotypes will be favoured over those that are still adapted to previously common host (Fig. 1.1) (Clarke, 1976; Hamilton, 1980; Jaenike, 1978). Since sex is efficient at generating genetic diversity, this parasite-driven negative frequency-dependent selection could explain the long-term maintenance of sex in host and parasite populations (Bell, 1982; reviewed in Lively, 2010). However, a prerequisite for the FSD is the existence of genetic specificity in host-parasite interactions, where the outcome of infection depends on the genotypic identity of both host and parasite (Dybdahl et al., 2014).

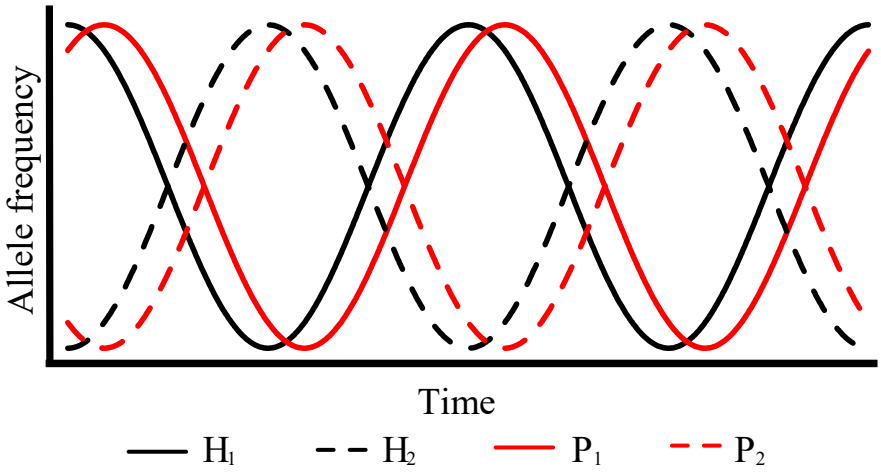

Figure 1.1. Allele frequency dynamics during host-parasite coevolution according to the FSD. Black lines indicate host genotypes and red lines indicate parasite genotypes. In this model, it is assumed that every host genotype has a specific parasite genotype. This assumption is represented by the type of line (continuous or discontinuous).

1.2. Fluctuating selection dynamics in natural populations

The strength and the response to parasite-mediated selection in nature vary across space and time. Such variation can be caused by selection mosaics across the landscape (Forde et al., 2004) or can result from temporal factors such as the seasonality of epidemics (Altizer et al., 2006). However, the majority of the studies about host-parasite dynamics to date were performed in the laboratory under controlled conditions (e.g. Decaestecker et al., 2007; Duffy and Sivars-Becker, 2007; Koskella and Lively, 2009; Schulte et al., 2010). Although such experiments are necessary to reduce environmental noise which can otherwise conceal important factors and processes, they tend to oversimplify natural conditions. Since parasite population dynamics are highly sensitive to environmental changes (reviewed in Wolinska and King, 2009), studies under natural conditions are essential to understand host-parasite coevolution.

1.2.1. Spatial scale of parasite populations

Most of the studies about parasites populations focus on the spatial genetic structure, which can affect the outcome of host-parasite interactions such as the parasite specialisation (Tripet et al., 2002) or the evolution of virulence (Thrall and Burdon, 1997). These two questions are relevant in epi-

demiological studies for disease control and prediction, as shown in clinical reports (e.g. Arnott et al., 2012; Brownstein et al., 2003; Raso et al., 2006). Specifically, estimates of parasite gene flow are fundamental to understand coevolutionary processes. Theoretical models show that the relative rate of gene flow could change parasite adaptation to local hosts and, thus, influence host-parasite dynamics (Frank, 1991). In fact, high parasite migration rates relative to those of their hosts are associated with parasite local adaptation by infectivity but not by virulence (Greischar and Koskella, 2007). For example, the introduction of novel alleles via gene flow among parasite populations may counteract host evolutionary responses (Slatkin, 1987). While in some cases limited parasite dispersal resulted in isolation-by-distance scenarios (e.g. Koop et al., 2014; Tanabe et al., 2013) in other cases parasite dispersal homogenised local patterns of population subdivision (e.g. Ocaña-Mayorga et al., 2010; Reuter et al., 2008).

1.2.2. Temporal changes in host populations

Field studies have demonstrated that host frequencies are tracked by parasites in a number of organismal systems, like in plants – fungi (e.g. Burdon and Thompson, 1995; Siemens and Roy, 2005), *Potamopyrgus antipodarum* (Gastropoda) – *Microphallus* sp. (Trematoda) (e.g. Jokela et al., 2009; King et al., 2009), *Cristatella mucedo* (Bryozoa) – *Tetracapsula bryozoides* (Myxozoa) (Vernon et al., 1996), and *Daphnia* spp. (Cladocera) – microparasites (e.g. Decaestecker et al., 2007; Little and Ebert, 1999; Wolinska and Spaak, 2009) systems. Below is a summary of the results in the mentioned systems.

Regarding the plant – fungi interactions, *Linum* spp. – *Melapsora lini* is one of the most well studied systems in phytopathology. In fact, this system was used to demonstrate the gene-for-gene relationship, in which the host resistance and parasite ability to cause disease in controlled by pairs of matching genes (Flor, 1971). Using this system, spatial differences between populations of *L. marginale* were found and such populations were described based on the proportion of resistant phenotypes. *M. lini* showed a similar spatial pattern based on the occurrence and frequency of its four most common races (Jarosz and Burdon, 1991). In a field study performed between 1981 and 1991 in Kiandra, New South Wales (Australia), the most common host genotypes decreased over time (Burdon and Thompson, 1995). These changes in frequencies of the most common host resistance genotypes were explained by linkage between resistance genes and the presence of *M. lini* races which are virulent for specific host phenotypes (Burdon et al., 1999). In other plant – fungi system (*Boechera holboelli – Puccinia monoica*) similar results were obtained. *B. holboelli* is a bi-

ennial or short-lived perennial apomictic (asexual by seed) plant distributed over Greenland and eastern Canada. Recently, its use for evolutionary and ecological genomics was proposed because it is phylogenetically related with *Arabidopsis thaliana* (Rushworth et al., 2011). This plant is attacked by a parasitic rust fungus, *P. monoica*, which inhibits flowering and transforms host morphology to facilitate its own sexual reproduction via pseudoflowers. In a ten-year field study, parasitism was lower in population with higher host genetic diversity as well as the evidence of local host adaptation in the fungus was detected. Although all these results supports the FSD, when herbivory by weevils was also incorporated, the parasitism decreased as host clone frequency and fitness increased, indicating that FSD is affected also by other interactions such as predation (Siemens and Roy, 2005).

P. antipodarum, the New Zealand mudsnail, is an ovoviviparous and parthenogenetic snail which is an invasive species. Native populations in New Zealand consist of diploid (sexual) and triploid parthenogenetically cloned females and sexually functional males (Dybdahl and Lively, 1995). These native populations of *P. antipodarum* are infected by *Microphallus* sp., a highly prevalent trematode which is able to completely sterilise infected snails (Jokela and Lively, 1995). In a field study performed between 1994 and 2005, the most common clones of snails were almost completely replaced by initially rare clones in the shallow and mid-water habitats of Lake Alexandrina (New Zealand), while sexual snails persisted in both habitats. The replacement of the most common clones by the rare clones was due to the susceptibility of the common clones to being infected by sympatric parasites (Jokela et al., 2009).

C. mucedo is a colonial, facultatively sexual freshwater bryozoan that has the unusual strategy of dispersing via asexually generated propagules. *C. mucedo* has a high genetic variability within populations, even in asexual populations (Freeland et al., 2000), and independently of the geographic distances among sites (Hatton-Ellis et al., 1998). In these populations, there is a highly prevalent parasitic myxozoa, *T. bryozoides*, that infects the body cavity and generates swelling, malformations and colony degeneration (Okamura, 1996). In a field study performed in the Bear Park Lake near Reading in Berkshire (United Kingdom) between 1992 and 1994, the two most common host clones varied in abundance and the significantly more common clone in the first year decreased its abundance in the third year. However, there was no evidence of high infection rate of the most common clone by the parasites which could be due to the existence of time-lags between parasite adaptations and the parasite-mediated decline of hosts (Vernon et al., 1996).

Finally, although the *Daphnia* spp. – microparasites system will be introduced in detail in

the section 1.4. of this thesis, here I focus on the evidences suggesting that *Daphnia* frequencies are tracked by parasites. In a field work performed during the summer of 1996 and spring of 1997 in six ponds between Northwest Switzerland and Northern Germany, temporal changes were observed in the proportion of parasitised *Daphnia* based on the host clonal frequencies and the population genetic characteristics. However, only three of these six ponds showed clonal dynamics consistent with FSD (Little and Ebert, 1999). In another field study, performed in spring and autumn of 2003 and 2004, the most common clone in *Daphnia* populations across several North Italian and Swiss lakes was often observed to decrease in frequency over time, supporting the FSD. In uninfected populations, however, the decrease was not observed (Wolinska and Spaak, 2009). Finally, a time-shift experiment using *Daphnia magna* and *Pasteuria ramosa*, a Gram-positive obligate intracellular bacteria showed that the parasite quickly adapts to its host. Also the observed temporal variation in parasite infectivity and virulence supported the FSD (Decaestecker et al., 2007).

1.2.3. Temporal changes in parasite populations

The majority of field studies investigated temporal changes in relative frequencies of host genotypes only, because tracking parasite dynamics over time often remained difficult in natural conditions (reviewed in Penczykowski et al., 2015). Surprisingly, except for the *Linum marginale-Melampsora lini* system (Thrall and Burdon, 2003), the Red Grouse-*Trichonstrongylus tenuis* system (Hudson and Dobson, 1997) and the St. Kilda Soay Sheep Project (Gulland and Fox, 1992; Gulland, 1992; Wilson et al., 2004), the majority of the studies on temporal changes in parasite populations and host-parasite specificity to date have been performed in the laboratory under controlled environmental conditions (reviewed in Penczykowski et al., 2015; Sadd and Schmid-Hempel, 2009). Studies under natural conditions are essential to understand host-parasite coevolution and to control diseases (Altizer et al., 2006).

1.2.4. Molecular identification of parasites

Traditionally, parasites were identified according to mainly morphological traits, although the information from epidemiology, host distribution and parasite physiology were also used to classify parasite strains (reviewed in McManus and Bowles, 1996). However, the existence of cryptic species of parasites questions the use of these traditional identification methods (e.g. Hanelt et al., 2015; Perkins, 2000; Vilas et al., 2005). Additionally, there are life stages such as eggs and larvae that are morphologically indistinguishable between phylogenetically related parasite species (re-

viewed in Criscione et al., 2005). For this reason, there is a trend to use molecular markers to identify parasites, especially their presence or absence (e.g. Djimde et al., 2001; Fischer et al., 2002; Knight et al., 1999). However, these molecular markers are often not able to differentiate among strains of the same parasite species, which would otherwise be necessary to study the parasite population dynamics at population level.

1.3. Internal transcribed spacer and its use as a molecular marker

The internal transcribed spacer (ITS) is the spacer DNA situated between the small and the large subunit ribosomal RNA (rRNA) genes. While in prokaryotes and in several Microsporidia (*Nosema* being an exception; see below) there is only one ITS region, in Eukaryotes there are two ITS regions; the ITS1 is located between the small subunit and 5.8S rRNA genes, while ITS2 is between 5.8S and the large subunit rRNA genes. ITS1 seems to be related to the prokaryotic ITS, while ITS2 is considered to be originated as an insertion that interrupted the ancestral large subunit rRNA gene (Lafontaine and Tollervey, 2001).

The rRNA gene structure is well preserved in all known phylogenetic groups, except in the *Nosema*/*Vairimorpha* group of Microsporidia (see below). For this reason, the ITS region is often used in molecular ecology and molecular systematics. In the former case, ITS region is mostly employed in Fungi, where it has been recommended as the universal fungal barcode sequence (Schoch et al., 2012). However, the combination of ITS sequences with other molecular markers is highly recommendable in some clinical molds (Balajee et al., 2009). Alternatively, as published ITS primers tend to amplify specific taxa to the detriment of other, the design of new primers for ITS sequences was suggested (Anderson et al., 2003; Bellemain et al., 2010). In the case of molecular systematics, ITS was employed at specific levels in bacteria (Boyer et al., 2001; Roth et al., 1998), and mainly in plants (e.g. Baldwin, 1992; Downie et al., 1998; Suh et al., 1993) and fungi (Iwen et al., 2002; Lloyd-MacGilp et al., 1996; Scorzetti et al., 2002). However, ITS substitution rate is highly variable between lineages in plants, and for this reason it is recommended to calibrate the molecular clock when ITS is used to study the evolution of a specific taxonomical group (Kay et al., 2006).

Although the rRNA gene structure is usually well preserved, such structure is highly variable in the *Nosema*/*Vairimorpha* group inside the Microsporidia. In fact, the rearrangement between small and large subunits and the presence of a 5S subunit at the end of the ribosomal RNA leads to the existence of an ITS2 between the small and the 5S subunits, as described in *N. bombycis*, *N. an-*

theraceae, *N. plutellae* and *N. spodopterae* (Huang et al., 2004; Tsai et al., 2005; Wang et al., 2006). Moreover, as the ribosomal RNA repeat unit is present in multiple copies throughout the genome, each copy has the potential for mutation, resulting in intragenomic variation. In fact, the presence of transcriptionally active fragmented copies of rRNA genes that coexist with the intact rRNA copies within the same genome was described in several isolates of *N. bombycis* (Iiyama et al., 2004). All this structural variation of the rRNA genes is a potential source of confusion in rRNA phylogenies (Ironside, 2007) and leads to a high variability in both ITS1 and ITS2 sequences in *Nosema/Vairimorpha* (Ironside, 2013). Such high variability could be due to recombination, which may be considered as evidence of "cryptic sex" in Microsporidia (Ironside, 2013; Krebes et al., 2014). Moreover, transposition events in ribosomal markers (including the ITS) are another source of high genetic variability (Iiyama et al., 2004; O'Mahony et al., 2007; Tsai et al., 2005).

1.4. *Daphnia*-microparasites as a host-parasite system

One of commonly explored model systems to address coevolutionary questions are the water fleas of the genus *Daphnia* (Crustacea: Cladocera) and their microparasites (Ebert, 2008; Gaba and Ebert, 2009; Stollewerk, 2010). *Daphnia* are small, largely transparent crustaceans that are found in most still freshwater bodies around the world. Such genus is probably the best-studied subjects in ecology, especially species *D. magna* and *D. pulex* (Ebert, 2005; Stollewerk, 2010). They can be kept as laboratory cultures on a diet of unicellular green algae. In fact, most members of this genus are able to reproduce both by parthenogenesis and sexually, switching a mode of reproduction depending on external factors (Ebert, 2008, 2005). The parthenogenetic reproduction allows obtaining clonal individuals that could be maintained in laboratory for years, with minimal genetic changes (Ebert, 2008).

As *Daphnia* spp. can be infected by several microparasites, such as fungi (including microsporidia), protozoa, oomycetes and bacteria (Ebert, 2008, 2005; Green, 1974), there is a need to decide on few microparasites as a model parasites. The best choice would be considering the most common microparasites infecting *Daphnia*. One of those is *Caullerya mesnili* (Ichthyosporea: Ichthyophonida; Bittner et al., 2002; Lohr et al., 2010a; Wolinska et al., 2007), which causes regular epidemics in large permanent lakes in Central Europe, reaching prevalences up to 40% (Wolinska et al., 2011a, 2007). *C. mesnili* is a highly virulent parasite that reduces the survival (from 36 days in healthy host to 21 days in infected host) and reproduction rate (from 23 to 2 offspring per healthy

and infected host, respectively) of its host (Lohr et al., 2010b). In *Daphnia* populations where *C. mesnili* is present, the existence of an acquired resistance to this parasite was described (Schoebel et al., 2010). *C. mesnili* was shown to be involved in driving frequencies of host species' abundance and genotypes (Wolinska et al., 2006) and its growth is mainly influenced by temperature (Schoebel et al., 2011) and the cyanobacteria abundance (Tellenbach et al., submitted). In a recent study, *C. mesnili* populations from seven water reservoirs in Czech Republic were shown to be significantly structured across space, and the frequency of *C. mesnili* genotypes varied significantly over time. Both observations suggest a limited dispersal of *C. mesnili* and a rapid evolutionary turnover (Wolinska et al., 2014).

Another relevant group of *Daphnia* microparasites are microsporidia. These parasites belong mainly to the Clade I or "Aquasporidia" class, which infect freshwater animals (Vossbrinck and Debrunner-Vossbrinck, 2005). Microsporidian parasites of *Daphnia* have received considerable attention due to their complex life cycles (Refardt et al., 2008, 2002; Weigl et al., 2012). In the present thesis, the focus was put on two abundant microsporidia infecting *Daphnia* communities inhabiting large lakes and reservoirs in Central Europe: *Berwaldia schaefernai* and the microsporidium MIC1 (Wolinska et al., 2009). Both of these microsporidia species infect the body cavity of their host, where a massive amount of spores then proliferate (Vávra and Larsson, 1994). They are closely related to *Marssoniella elegans* (a parasite of the copepod *Cyclops vicinus*), *Senoma globulifera* (a parasite of the malaria mosquito *Anopheles messeae*), and other parasites of *Daphnia*, including *Larssonia obtusa*, *Gurleya vavrai* and *Binucleata daphniae* (Weigl et al., 2012). Unlike *S. globulifera* and *B. daphniae*, which are monoxenous parasites (Refardt et al., 2008; Simakova et al., 2005), *L. obtusa, G. vavrai* and *B. schaefernai* cannot be maintained in the laboratory, which could suggest the existence of an indirect life cycle with a secondary host (Refardt et al., 2002; Vávra and Larsson, 1994). Moreover, their relative species *M. elegans* is a dixenous parasite, having copepods as its main host and likely using mosquitoes or caddisflies as secondary hosts (Vávra et al., 2005; Vossbrinck et al., 2004). Although abiotic factors such as temperature are important to the aquatic microsporidian growth (Dunn et al., 2006; Grabner et al., 2014), it is assumed that the monoxenous life cycles of *S. globulifera* and *B. daphniae* are due to the loss of the ancestral character (i.e. the use of a secondary host) and it has been predicted that *B. schaefernai* and MIC1 may also have a secondary insect host, which would confer previously observed substantial potential for dispersal (Weigl et al., 2012; Wolinska et al., 2011b).

1.5. Outline of the thesis

The main goal of this PhD project was to analyse the population structure of two major endoparasites of *Daphnia*: *Caullerya mesnili* and microsporidia. To perform this kind of study, a new bioinformatic pipeline needed to be developed in order to be able to analyse molecular data generated by next-generation-sequencing (NGS) platform. NGS technologies provide a cheaper and faster alternative to sequencing DNA than traditional methods like cloning and Sanger sequencing (reviewed in Grada and Weinbrecht, 2013). However, the consideration of the number of taxon-specific reads as an indicator for the abundance of different taxa (semi-quantitative analysis) is still debated because NGS techniques produce several types of artefacts (e.g. Amend et al., 2010; Baldrian et al., 2013; Deagle et al., 2013), such as the presence of homopolymer and indel misinterpretations in 454 pyrosequencing platform (Margulies, 2005) and the increase of single-base errors and inverted repeats or GGC motifs in Illumina (Nakamura et al., 2011). In order to minimise these artefacts, the most abundant sequences per sequence cluster could be considered as the presumed ancestral allelic reference (Sommer et al., 2013). These sequences are called "representative sequences".

First, in Chapter 2, a new bioinformatic pipeline based on the concept of representative sequences was developed and applied to data of *C. mesnili* populations from different water reservoirs in the Czech Republic and compared with published results from the same populations but using Sanger sequencing method.

In Chapter 3, samples of infected *Daphnia* population from lake Greifensee (Switzerland) were collected during four consecutive years, and analysed with 454 pyrosequencing. Our new bioinformatic pipeline was then applied to analyse *C. mesnili* sequences and the population dynamics and host-genotype specificity of this parasite were assessed based on the sequence variation in the ITS1 region.

Finally, in Chapter 4, populations of *B. schaefernai* and the unknown microsporidium MIC1 sampled from seven water reservoirs in the Czech Republic, were sequenced on 454-platform and further analysed to describe their genetic diversity and the patterns of geographical population structure. All this information was useful to infer the biology of these parasite species. Also here, our new bioinformatic pipeline was employed.

The thesis concludes with Chapter 5, providing a general discussion and suggestions for future research.

Chapter 2 – The Quantification of Representative Sequences pipeline for amplicon sequencing: Case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*.

E. González-Tortuero[1,2,3,*], J. Rusek[3], A. Petrusek[4], S. Gießler[3], D. Lyras[3], S. Grath[3], F. Castro-Monzón[1], J. Wolinska[1].

Molecular Ecology Resources (2015) 15(6): 1385 – 1395

[1]Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin (Germany).

[2]Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin (Germany).

[3]Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, 82512 Planegg-Martinsried (Germany).

[4]Department of Ecology, Faculty of Sciences, Charles University in Prague, Viničná 7, 128 44 Prague (Czech Republic).

[*]**E-mail:** gonzalez@igb-berlin.de

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2015) **15**, 1385–1395 doi: 10.1111/1755-0998.12396

The Quantification of Representative Sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*

E. GONZÁLEZ-TORTUERO,*†‡ J. RUSEK,‡ A. PETRUSEK,§ S. GIEßLER,‡ D. LYRAS,‡ S. GRATH,‡ F. CASTRO-MONZÓN* and J. WOLINSKA*

**Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany, †Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin, Germany, ‡Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, 82512 Planegg-Martinsried, Germany, §Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague, Czech Republic*

Abstract

Next generation sequencing (NGS) platforms are replacing traditional molecular biology protocols like cloning and Sanger sequencing. However, accuracy of NGS platforms has rarely been measured when quantifying relative frequencies of genotypes or taxa within populations. Here we developed a new bioinformatic pipeline (QRS) that pools similar sequence variants and estimates their frequencies in NGS data sets from populations or communities. We tested whether the estimated frequency of representative sequences, generated by 454 amplicon sequencing, differs significantly from that obtained by Sanger sequencing of cloned PCR products. This was performed by analysing sequence variation of the highly variable first internal transcribed spacer (ITS1) of the ichthyosporean *Caullerya mesnili*, a microparasite of cladocerans of the genus *Daphnia*. This analysis also serves as a case example of the usage of this pipeline to study within-population variation. Additionally, a public Illumina data set was used to validate the pipeline on community-level data. Overall, there was a good correspondence in absolute frequencies of *C. mesnili* ITS1 sequences obtained from Sanger and 454 platforms. Furthermore, analyses of molecular variance (AMOVA) revealed that population structure of *C. mesnili* differs across lakes and years independently of the sequencing platform. Our results support not only the usefulness of amplicon sequencing data for studies of within-population structure but also the successful application of the QRS pipeline on Illumina-generated data. The QRS pipeline is freely available together with its documentation under GNU Public Licence version 3 at http://code.google.com/p/quantification-representative-sequences.

Keywords: amplicon sequencing, *Caullerya mesnili*, Ichthyosporea, ITS1 region, Next Generation Sequencing, pipeline

Received 16 June 2014; revision received 16 February 2015; accepted 19 February 2015

Introduction

Next generation sequencing (NGS) technologies generate substantially more data than Sanger sequencing, while being both cheaper and faster. Thus, NGS data sets can increase our understanding of biological phenomena as, for instance, linking observed microbial diversity with ecological functions (e.g. Huber *et al.* 2007; Edgcomb *et al.* 2011; Kautz *et al.* 2013). However, investigation of population or community changes over time and space requires not only identification of different genetic variants, but also reliable quantification of genotypes, operational taxonomic units (OTUs) or species. Until now, comparisons of results from cloning/Sanger sequencing and amplicon sequencing by NGS focused mostly on read length and sequencing accuracy (e.g. Huse *et al.* 2007; Harismendy *et al.* 2009; Liang *et al.* 2011), or on the correspondence at the qualitative level (Edgcomb *et al.* 2011; Sommer *et al.* 2013). The few analyses that consider the number of taxon-specific reads as an indicator for the abundance of the different taxa (i.e. semiquantitative analyses) are still debated (e.g. Amend *et al.* 2010; Baldrian *et al.* 2013; Deagle *et al.* 2013).

The correctness of quantitative assignments is still discussed because NGS methods produce several types of errors and artefacts. In the case of 454 pyrosequencing platform, homopolymer and indel misinterpretations (Margulies *et al.* 2005) are described as common errors. In contrast, reads from the Illumina platform are much

Correspondence: Enrique González-Tortuero, Fax: +49 30 83850784; E-mail: gonzalez@igb-berlin.de

more precise, but systematic base-calling biases are present (Erlich *et al.* 2008; Rougemont *et al.* 2008; Renaud *et al.* 2013). Specifically, there is a correlation between increased single-base errors and inverted repeats or GGC motifs (Nakamura *et al.* 2011). Moreover, error rates between reads vary between 0.3% and 4.6% (Dohm *et al.* 2008; Dolan & Denver 2008; Kozich *et al.* 2013) due to the different tiles of the sequencing plate. To correct for such errors, DNA from the same individuals can be amplified independently and information from different sequencing runs can be compared (Sommer *et al.* 2013). This, however, is impractical for large population data sets. In such cases, identification of true alleles can be optimized by grouping sequences containing minor errors using the most abundant sequence as the correct allelic reference. Such representative sequences could then be used in analyses of community or population structure.

Representative sequences from NGS data sets are usually inferred by distance-based methods and neighbour-joining clustering, as implemented in, for example, SESAME Barcode (Piry *et al.* 2012). These algorithms are fast but assume no reticulate relationships (Posada & Crandall 2001) and no recombination events (Posada & Crandall 2002). However, the task of identifying representative sequences requires discriminating between natural variants present in a population or organism and variants that result from methodological errors. Variation within an organism does not imply methodological errors as this variation can result from natural processes like incomplete concerted evolution of multicopy genes (e.g. Hugall *et al.* 1999; Koch *et al.* 2003; Lindner *et al.* 2013). To deal with this problem, network approaches such as statistical parsimony (Templeton *et al.* 1992) should be preferred. This algorithm involves grouping the sequences within networks in the most parsimonious way, while at the same time, it preserves the number of differences between sequences within groups below a specific threshold (Templeton *et al.* 1992; Posada & Crandall 2001). In fact, statistical parsimony has been proven to be more precise in identifying representative ITS1 sequences, that is the first internal transcribed spacer region of ribosomal DNA (rDNA), than distance-based methods and neighbour-joining clustering. Specifically, in a previous work (Giessler & Wolinska 2013), we compared the neighbour-joining method with statistical parsimony to address polymorphisms in the ichthyosporean *Caullerya mesnili*. Thus, half of the representative sequences, as assigned by statistical parsimony, were collapsed by the neighbour-joining method, indicating that statistical parsimony approach is more robust (Giessler & Wolinska 2013). Nevertheless, as far as large complex NGS data sets are concerned, no bioinformatic solution is readily available for the identification of representative sequences based on statistical parsimony.

In this study, we developed a new bioinformatic pipeline – quantification of representative sequences (QRS) – that infers the representative sequences from NGS data sets (based on neighbour joining or statistical parsimony) and calculates their frequencies. With this pipeline, we compared data from 454 amplicon sequencing ('454 data set') with data from Sanger sequencing of cloned ITS1 amplicons ('Sanger data set') of the ichthyosporean *Caullerya mesnili*, a common microparasite of the cladoceran *Daphnia* (Bittner *et al.* 2002; Lohr *et al.* 2010). The *C. mesnili–Daphnia* system has been used to study general evolutionary questions about host–parasite co-evolution (e.g. Wolinska *et al.* 2006; Wolinska & Spaak 2009; Schoebel *et al.* 2010). Using QRS, the frequencies of representative *C. mesnili* ITS1 sequences in the 454 data sets were scored and then compared against the outcome from Sanger-produced data sets using the same – but cloned – template DNA. Spatiotemporal variation of *C. mesnili* populations was then assessed, and the consistency of results from the two sequencing methods was compared to validate the usefulness of amplicon sequencing for semiquantitative analyses. Additionally, the robustness of results generated by the QRS pipeline was tested using a published Illumina amplicon sequence data set. The outcome of analysis was compared with that obtained from two well-established metabarcoding programs: mothur (Schloss *et al.* 2009) and UPARSE (Edgar 2013).

Materials and methods

Bioinformatic pipeline

General description. We developed an automatic and modular bioinformatic pipeline (implemented in Python 3.4) called QRS that determines genotypic structure within populations and communities from amplicon sequencing data. This pipeline can deal with any molecular marker but data sets with multiple markers (like multilocus sequence typing) should be demultiplexed (i.e. separated according to the combination of multiplexer identifiers, MIDs) before running QRS. Although our pipeline was developed specifically for analyses of 454 and Illumina data sets, it is applicable to other technologies, for example Ion Torrent (Rothberg *et al.* 2011) with minimal modifications. QRS processes all input files in four steps: (i) creating a reference data set, (ii) describing the NGS data set, (iii) processing the NGS data set to obtain an alignment and (iv) inferring representative sequences (Fig. 1). The complete manual of the program is provided in Data S1 (Supporting information).

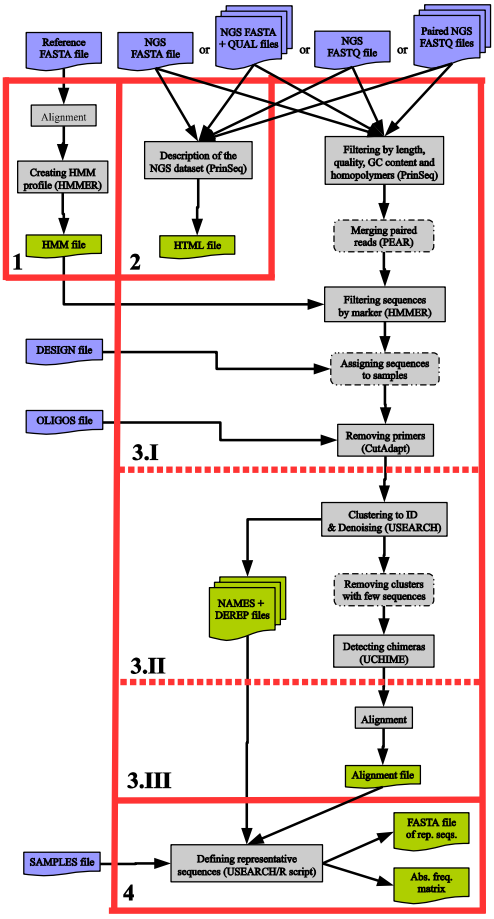

Fig. 1 Flowchart of the quantification of representative sequences (QRS) pipeline. Grey rectangles represent the different steps of the program (among those, broken-lined rectangles indicate optional steps; see main text and Data S1, Supporting information). Blue rectangles with a wavy base stand for input files and green rectangles indicate output files. Insets with red lines (numbered in 1–4) summarize steps of the QRS pipeline according to the main text and Data S1 (Supporting information). For the purpose of alignment, it is possible to use different tools (see main text).

Creating a reference data set. The program requires a reference FASTA file containing multiple sequences of the respective marker. Sequences within this reference file are aligned using one of several alignment programs (see below for details) according to the user criteria (Fig. 1: inset 1). Then, HMMER 3.1b (Finn *et al.* 2011) is called to infer a profile Hidden Markov Model (HMM; Churchill 1989) from the previously generated alignment. HMM is a probabilistic model that is used to search for a specific DNA sequence, and it has been proven that this method is as fast as BLAST but more accurate (Wheeler & Eddy 2013).

Describing the NGS data set. QRS uses a NGS FASTQ (single or paired FASTQ files) or a NGS FASTA file as input (it is also possible to add a NGS quality file when a FASTA file is used) (Fig. 1: inset 2). Sequences are processed in PrinSeq (Schmieder & Edwards 2011) to calculate basic statistics about length, GC content, base quality and sequence complexity distributions. All these measures are important for defining the best parameters in the subsequent analysis steps. The output of this step is an HTML file with all pertinent information regarding the NGS data set quality.

Processing the NGS data set to obtain an alignment. This step is divided in three processes: (i) filtering and assigning sequences to samples, (ii) denoising sequences and removing chimeras, and (iii) aligning all sequences (Fig. 1: inset 3).

Filtering and assigning sequences to samples (inset 3.I). Initially, the sequences from the NGS data sets are provided as input to PrinSeq to be filtered according to their quality, length, GC content and/or presence of homopolymers. If the NGS data sets are composed by paired FASTQ files, the deriving filtered sequences are merged in a single FASTA file using PEAR (Zhang *et al.* 2014). Then, all sequences are filtered according to the reference profile HMM, using nhmmer (Wheeler & Eddy 2013) from the HMMER 3.1b package. This software classifies each sequence to a specific marker according to the HMM profile based on the *e*-value. In our case, the threshold *e*-value was set to 10^{-10}, as in BLAST (Altschul *et al.* 1990), but this value can be adjusted interactively from the user interface. After that, the accepted sequences are assigned to the respective samples based on their MID sequences (if not already separated according to their MIDs). Primers are removed using CutAdapt (Martin 2011). To assign sequences to samples and to remove the primers, QRS uses the information from two plain text files: one provides a sample-specific combination of MID sequences (DESIGN file), and the other one contains information on the primers and MID sequences (OLIGOS file). All incomplete reads (not containing both forward and reverse MIDs) are discarded.

Denoising sequences and removing chimeras (inset 3.II). All accepted sequences from the 'filtering and assigning sequences to samples' process are joined at a given identity level, whereas all spurious nucleotides across all

sequences are removed, based on the CD-HIT-OTU approach (Li *et al.* 2012), using USEARCH. By default, the sequences are joined if they have only one mismatch or one or two gaps. The identity threshold can be modified by the user (see Data S1, Supporting information).

After this denoising step, the QRS pipeline can remove clusters that have fewer sequences than a specified threshold. By default, all clusters that have fewer than three sequences are removed. This threshold was determined by previous analyses (Data S2, Supporting information). Although clustering and subsequent removal of rare clusters result in less RAM consumption for the follow-up analyses, it is important to note that this step is optional and it can be disabled depending on users' preferences. Finally, to remove chimeric sequences, QRS calls the UCHIME algorithm (Edgar *et al.* 2011) implemented in USEARCH, using all sequences that passed through previous filters as a reference. This method is based on the assumption that abundant sequences in a data set have a lower probability of chimerism than rare ones (Quince *et al.* 2009). The use of all sequences that passed through previous filters as a reference is more efficient than using a reference alignment only (Schloss *et al.* 2011).

Aligning all sequences (inset 3.III). QRS can align all accepted sequences from the 'denoising and removal chimeras' process using one of several supported aligners: Clustal Omega (Sievers *et al.* 2011), FSA (Bradley *et al.* 2009), GramAlign (Russell *et al.* 2008), Kalign (Lassmann & Sonnhammer 2005), MAFFT (Katoh & Standley 2013), MUSCLE (Edgar 2004), PicXAA (Sahraeian & Yoon 2010) and PRANK (Löytynoja & Goldman 2005). QRS can then (optionally) execute ReformAlign (Lyras & Metzler 2014) to postprocess the alignment based on a profile-based meta-alignment approach.

Inferring representative sequences. QRS can assign sequences from the alignment of the previous step to representative sequences based on neighbour-joining clustering (Saitou & Nei 1987) or statistical parsimony (Templeton *et al.* 1992). When a neighbour-joining algorithm is needed, USEARCH is executed to cluster sequences. When a statistical parsimony algorithm is used to pool sequences into groups, QRS calls an embedded R script. Via this R 3.0 script, QRS can analyse a higher number of sequences compared to when the TCS software (Clement *et al.* 2000) is executed, because the latter requires more RAM when pooling huge data sets (results not shown). In our case, a 99.5% identity threshold (representing three connection steps) was used to assign sequences to representative sequences. We treated gaps as a fifth character (similar to Giessler & Wolinska 2013), but this can be changed by the user (see Fig. 1: inset 4, and Data S1, Supporting Information).

Independent of the used method (i.e. neighbour joining or statistical parsimony), QRS assigns the most likely representative sequence to each group and calculates the frequencies of each representative sequence in the different samples. To identify the most representative sequence in every group, QRS searches for the respective sequence header (according to the USEARCH output log file in case of neighbour joining or according to the R script analyses in case of statistical parsimony) in the original FASTA alignment. Then, all representative sequences are saved in a new FASTA file. Additionally, to quantify the frequencies of each representative sequence in the samples, a plain text file (SAMPLES file) is needed that contains all information about the samples (starting from the sample identifier which must be the same as in the DESIGN file). With this information, the USEARCH output log file (in case of neighbour joining) or the result from the R script analysis (in case of statistical parsimony) is processed to count how many representative sequences exist in each sample. This results in a matrix with the absolute frequencies (plain text file).

Case study: C. mesnili ITS1 sequence variation

Sampling design. For this study, we used the same parasite DNA samples as those previously cloned and Sanger sequenced in a methodological study by Giessler & Wolinska (2013), and in which the detailed sampling design is described. In short, *Daphnia* samples were collected from seven drinking water reservoirs in the Czech Republic: Brno, Římov, Stanovice, Trnávka, Vír, Vranov and Želivka (geographical locations and further characteristics of the reservoirs are provided in Seda *et al.* 2007). From all available samples (collected in autumn 2004, 2005, 2008 and 2009), those containing a substantial proportion of *C. mesnili*-infected *Daphnia* individuals were chosen (16 samples, one to four per lake; see Table S1, Supporting information and Giessler & Wolinska 2013). Twenty highly infected *Daphnia* individuals were selected and pooled per sample for DNA extraction.

Molecular analyses. DNA isolates from twenty pooled, infected *Daphnia* individuals (per sample) were processed as follows. First, PCR was run to amplify the ITS1 region in a 20 μl reaction volume, containing 1 μl of genomic DNA isolate, 0.25 mM of deoxynucleoside triphosphates, 1 μM of each primer, 1× buffer and 1.25 U of Dream Taq DNA polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Specific *C. mesnili* ITS1 primers (Forward: ACACCGCCCGTCACTACTAC and Reverse: TGG ATATACCACTCTCAAACAG), amplifying a fragment

approximately 425 bp long, were designed using Oligo Explorer 1.2 (Kuulasmaa 2010) based on the sequence alignment from Giessler & Wolinska (2013). The primers used in this previous study were 18S Cm1469For [AGCACAAGTCCTTAACTTGTGTT] and a 1:1 mix of 28S Cm1-1 Rev [CACTCGCCGTTACTGAGG GAATC] and 28S CM1-2 Rev [CATTCGCCATTACTA AGGGAATC] (Lohr *et al.* 2010). The newly designed *C. mesnili*-specific primer sequences were used as the 3′-portions of fusion primers for amplicon pyrosequencing, together with a 25-mer for binding to the DNA Capture Beads (Lib-A) at the 5′-end, and 10-base MIDs chosen from the 454 Standard MID Set (Roche/454 Life Sciences) in the middle of the fusion primer. Samples were tagged with a unique combination of MIDs at forward and reverse primers. The following PCR protocol was used as follows: an initial denaturation step of 3 min at 94 °C, followed by 40 cycles of 30 s at 94 °C, 30 s at 53 °C, 2 min at 72 °C and a final extension step of 7 min at 72 °C. All 16 samples were then pooled in an equimolar concentration of each product, with DNA concentrations determined by fluorescence measurements using PicoGreen (Invitrogen, Inc.). Emulsion-based clonal amplification of a DNA library was subsequently performed, following the GS Junior Titanium emulsion-based clonal amplification (emPCR) protocol for amplicons (Roche/454 Life Sciences) and using the recommended chemistry. Amplicon sequencing was carried out on the GS Junior System (Roche/454 Life Sciences), according to the manufacturer's instructions.

Comparison between 454 and Sanger data sets. To test whether 454 sequencing introduces quantitative bias into the assessment of population structure, we joined the 454 data set generated in this study ('filtered 454') with a data set previously obtained from the same DNA isolates by cloning and Sanger sequencing (Giessler & Wolinska 2013; Wolinska *et al.* 2014). The 'filtered 454' data set was produced by discarding all clusters that contained only one or two sequences after denoising and removing chimeras (see Data S2, Supporting information). Statistical parsimony was run jointly for all 16 samples (analysed both by 454 and by Sanger sequencing). It was possible to run this step with all 16 samples, because less than 2 MB of RAM was required to connect the reduced number of sequences after filtration. Specifically, 3245 sequences were present in the 'filtered 454' data set versus 16,100 in the 'raw 454' data set. After each sequence was assigned to a specific representative sequence, we compared the frequencies of representative *C. mesnili* sequences between the 'filtered 454' and Sanger data sets, separately for each of 16 analysed samples, by applying a Monte Carlo permutation test with 10,000 replicates as implemented in SPSS (version 20.0, IBM 2011). Sequential Bonferroni correction (Rice 1989) was used to correct *P*-values in these and all other statistical tests, if necessary. To assess the extent of spatial and temporal variation among *C. mesnili* population samples in both the 'filtered 454' and Sanger data sets, we applied analyses of molecular variance (AMOVA) using the distance method and pairwise differences, in ARLEQUIN 3.5 (Excoffier & Lischer 2010). AMOVA tests for spatial patterns were run separately for each year, from which more than two population samples were available (2004, 2005 and 2009; Table S1, Supporting information). Similarly, tests for temporal patterns were applied separately per lake, focusing on lakes that had been sampled more than twice (Římov, Vír and Vranov; Table S1, Supporting information). The significance of each AMOVA run was assessed relative to 1000 randomly permuted data sets.

Validation of the QRS pipeline on an Illumina data set

The QRS pipeline was also tested on an Illumina MiSeq data set using published data from 18S rDNA of intertidal meiofauna of Alabama (Brannock *et al.* 2014). In this validation test, we analysed only a subset of samples (i.e. 29 samples, which corresponds to ~ 8% of the entire data set), to restrict computational effort. This subset was downloaded from the SRA webpage (SRR codes: SRR1290551–SRR1290578, SRR1290580) and was processed in QRS, UPARSE (Edgar 2013) and mothur v. 1.34.4, inferring representative sequences at the 97% identity level, using neighbour-joining clustering. Later, all representative sequences were assigned to OTUs using SINA (Pruesse *et al.* 2012) and SILVA 119 SSU database as reference (Yilmaz *et al.* 2014). Finally, frequencies of all OTUs at phylum level were compared between the results from QRS, UPARSE and mothur after pooling all samples. This was performed by applying a Monte Carlo permutation test with 10 000 replicates using R 3.1.2 (R Core Team 2014). Sequential Bonferroni correction (Rice 1989) was used to correct *P*-values of post hoc comparisons. Detailed information on the validation of the QRS pipeline is presented in Data S3 (Supporting information).

Results

Comparison between 454 and Sanger data sets

There were no significant differences between the 'filtered 454' and 'raw 454' data sets (as assessed for 16 analysed cases) in the frequencies of representative sequences (Data S2, Supporting information). This allowed us to use the 'filtered 454' data set in all subsequent analyses of this study.

Specifically, the 3245 sequences recovered in the 'filtered 454' data set (see Data S2, Supporting information) were joined with 455 sequences obtained from cloning/Sanger sequencing runs (Giessler & Wolinska 2013). The length of this combined alignment (i.e. 'filtered 454 and Sanger') was 406 bp. Fifteen representative sequences (i.e. the most abundant sequence in each group; C2.1 to C2.15; Table S2, Supporting information) were identified in this combined data set. More than half of the sequences (53.9%) were assigned to a single representative sequence: C2.14. By contrast, ten representative sequences were present at proportions less than 0.5% (Table S2, Supporting information). In general, the same representative sequences were recovered by 454 and Sanger sequencing, at similar frequencies per sample (Fig. 2). There was only one instance where the relative frequencies of representative sequences differed significantly between the two sequencing approaches: Vranov 2009 (Fig. 2; Table S2, Supporting information). The consistency of results from the two sequencing methods was supported, even when the most abundant representative sequence was discarded in each pair of samples. In that case, the differences in representative sequence frequencies remained nonsignificant in all 16 pairwise comparisons (Table S4, Supporting information). Overall, the sequences present exclusively in one or the other data set were those of low frequencies. Specifically, seven rare representative sequences (i.e. at proportions less than 0.5%; Table S2, Supporting information) were present exclusively in the 454 data set and three in the Sanger data set only.

The largest amount of variation in both data sets, as revealed by AMOVA, was observed at the within-population level. However, both spatial and temporal components were still significant in the two data sets and explained comparable amounts of variation. Thus, 7–11% and 6–14% of the total variance were explained by the spatial, and 3–15% and 4–13% by the temporal component in the 454 and Sanger data sets, respectively (Table 1). Regarding the Sanger data set, the results

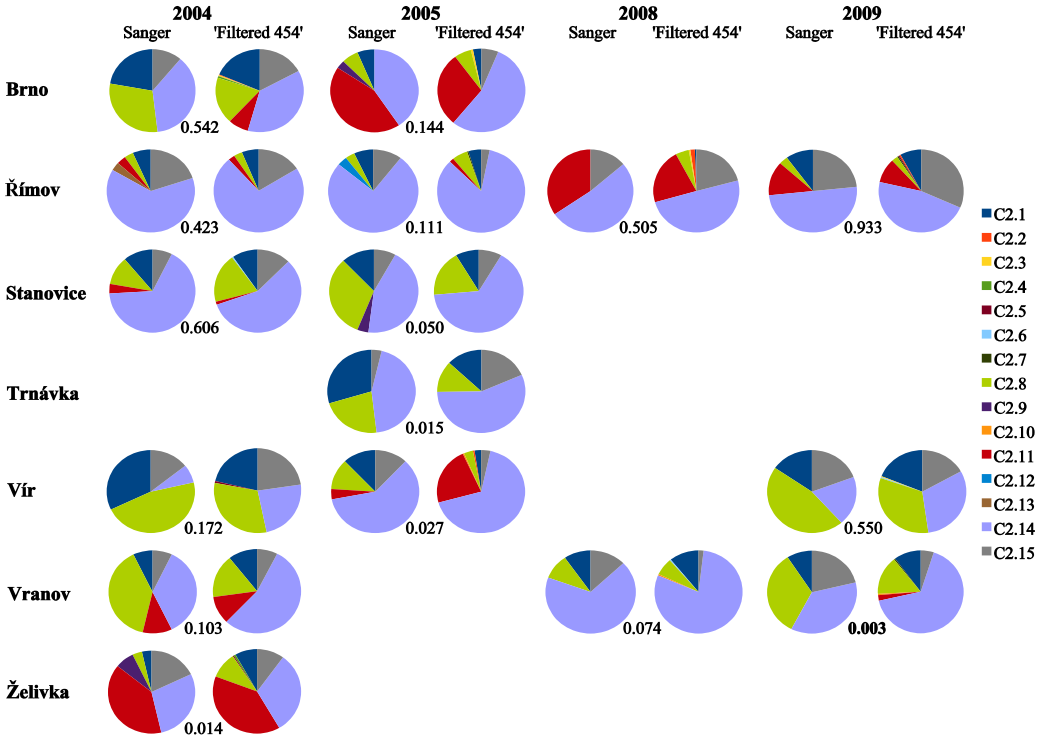

Fig. 2 Comparison of frequencies of *Caullerya mesnili* ITS1 representative sequences between 'Sanger' and 'filtered 454' data sets. Results of Monte Carlo permutation tests are shown between the respective pie charts. All calculations were performed using 10 000 replicates. The *P*-value that remained significant after sequential Bonferroni correction is shown in bold.

Table 1 Results of the analysis of molecular variance (AMOVA) to explore spatial and temporal population structure in *Caullerya mesnili* parasites. All calculations were based on the frequency of representative ITS1 sequences in parasite DNA

		'Filtered 454' data set			Sanger data set		
Level of variation	Source of variation	d.f.	% of explained variation	*P*-value	d.f.	% of explained variation	*P*-value
Spatial variation							
All 2004	Among lakes	5	8.70	<0.001	5	19.14	<0.001
	Within lakes	1208	91.30		162	80.86	
All 2005	Among lakes	4	7.27	<0.001	4	9.91	<0.001
	Within lakes	944	92.73		133	90.09	
All 2009	Among lakes	2	10.75	<0.001	2	6.70	0.005
	Within lakes	670	89.25		86	93.30	
Temporal variation							
All Římov	Among years	3	9.99	<0.001	3	4.11	0.027
	Within years	694	90.01		114	95.89	
All Vír	Among years	2	14.88	<0.001	2	12.84	0.001
	Within years	601	85.12		77	87.16	
All Vranov	Among years	2	3.62	<0.001	2	6.47	0.001
	Within years	576	96.38		89	93.53	

generated by the QRS pipeline are comparable with the outcome of AMOVA analyses published in Wolinska *et al.* (2014). Small differences in the amount of explained variation result from cutting sequence ends, necessary to standardize the alignment length for the comparison of Sanger and 'filtered 454' data set, whereas the slightly different number of analysed sequences is due to the removal of sequences that had ambiguous nucleotides.

Validation of the QRS pipeline on an Illumina data set

The QRS pipeline applied on an already published Illumina data set provided similar results regarding the frequencies of identified OTUs, as derived from two well-established metabarcoding programs: UPARSE and mothur. In fact, five phyla were only recovered by QRS, while there were no OTUs uniquely appearing in the results of mothur or UPARSE. Additionally, the percentage of 'unclassified' or prokaryotic OTUs was less in QRS than in UPARSE or in mothur. Detailed results on the validation of the QRS pipeline are presented in Data S3, Supporting information.

Discussion

Quantification of representative sequences, our newly developed bioinformatic pipeline, was used to estimate the frequency of representative variants of the ITS1 region from the *Daphnia* parasite, *Caullerya mesnili*. Data from cloning/Sanger sequencing and 454 amplicon sequencing from the same DNA isolates were compared to test the accuracy of the pipeline when using error prone NGS data. Results from both sequencing platforms were in good agreement with respect to the QRS as well as the outcome of population analyses focusing on spatial and temporal variation.

Development of the QRS pipeline

We developed a new pipeline to assign and quantify representative sequences from 454 data sets to analyse variation within populations using network-based approaches. Network algorithms are frequently used to analyse Sanger data sets at community or intraspecific levels (e.g. Knittweis *et al.* 2009; Chen *et al.* 2010; Wolinska *et al.* 2011). Missing, however, are toolkits which would allow the application of the same algorithm to large sequencing data sets, such as those obtained by NGS technologies. The majority of existing programs that use amplicon sequence data sets as input, like mothur (Schloss *et al.* 2009), QIIME (Caporaso *et al.* 2010) and UPARSE (Edgar 2013), were developed to describe, compare and analyse microbial communities in metagenomic studies, typically focusing on biodiversity at levels higher than the genus (e.g. Edgcomb *et al.* 2011; Kautz *et al.* 2013). SESAME BARCODE software (Piry *et al.* 2012) and a recently developed metabarcoding pipeline for fungal ITS1 sequences (Bálint *et al.* 2014) infer representative sequences using neighbour-joining clustering to classify amplicon sequences into operational taxonomic units for DNA barcoding purposes. However, representative sequences might be better inferred by network approaches to study within-population variability using multicopy genes or gene families. In fact, as shown in

our previous work (Giessler & Wolinska 2013), neighbour-joining clustering was not precise enough to identify representative ITS1 sequences in *C. mesnili* population samples because reticulate relationships and recombination events are not considered by this method (Posada & Crandall 2001, 2002). Therefore, we implemented in the QRS pipeline both neighbour joining and statistical parsimony (Templeton *et al.* 1992) as a clustering algorithm and a network approach, respectively, making it useful to identify and quantify representative sequences for a variety of purposes. However, the QRS pipeline is also open to the implementation of other network methods, such as reduced median networks (Bandelt *et al.* 1995) and median-joining networks (Bandelt *et al.* 1999).

Another advantage of the QRS pipeline is its flexibility to use a variety of alignment algorithms. After identifying the best-fitting multiple alignment program for the sequence data set, automatic tools can be applied to post-process the alignment. For the purpose of our evaluation, we used PRANK and ReformAlign because the combination of both methods generated more accurate alignments in terms of Modeler's and Cline shift scores than other aligners (results not shown). The Modeler's score computes the ratio of correctly aligned residue pairs with the length of the resulting alignment (Sauder *et al.* 2000), whereas the Cline shift score penalizes under- and over-alignment and considers regions in the generated alignment that may be shifted by a few positions according to the reference alignment (Cline *et al.* 2002).

Finally, the validation analysis suggested that the proposed pipeline was able to retrieve the same phyla and at similar frequencies as other state-of-the-art methods. These results demonstrated that the pipeline is compatible with Illumina-generated data sets (as we discuss in Data S3, Supporting information) and can be further adjusted to analyse Ion Torrent amplicon sequence data sets (e.g. Jünemann *et al.* 2012; Deagle *et al.* 2013). In the latter case, it would be necessary to implement tools that correct for the high number of erroneous sequences with high GC content (Quail *et al.* 2012; Deagle *et al.* 2013).

Comparison between 454 and Sanger data sets

The frequencies of representative ITS1 sequences derived from 454 and Sanger data sets were similar, with only one significant difference across the 16 studied cases. As expected, 454 pyrosequencing detected some additional rare (i.e. < 0.5%) representative sequences, because of the higher sampling power of the method (Huse *et al.* 2007; Kröber *et al.* 2009; Liang *et al.* 2011). Alternatively, the cloning bias, that is some sequences being less likely to be cloned than others due to the cell culture strain used (Forns *et al.* 1997; Sorek *et al.* 2007) and/or ligation conditions (Palatinszky *et al.* 2011; Zhuang *et al.* 2012), could also explain missing representative sequences in the Sanger results.

Surprisingly, some rare representative ITS1 sequences identified in the Sanger data set were not detected in the 454 data set. This discrepancy is consistent with reports from previous studies on fungal diversity which suggested poor sensitivity of the 454 method for certain, even frequent, taxa (Tedersoo *et al.* 2010; Kauserud *et al.* 2012). Moreover, differences between primers used in Sanger and in 454 could have produced further bias in the detection of rare representative ITS1 sequences. These additional sequences, however, might also result from artefacts. In both Sanger and 454 data sets, PCR errors like the inverse relationship between nucleotide quality and sequence length (Schröder *et al.* 2010) as well as chimeric sequence formation (Wang & Wang 1996; Ashelford *et al.* 2005; Haas *et al.* 2011) are well known. In fact, while up to 10% of sequences are chimeras in Sanger data sets, a frequency of up to 20% has been described for pyrosequencing (Haas *et al.* 2011). Thus, although the used here UCHIME algorithm to detect chimeras is highly sensitive and specific (Schloss *et al.* 2011), it is still possible that some chimeras could be overlooked and included in the subsequent analysis, leading to overestimation of sequence diversity (Reeder & Knight 2009). Finally, the nonproofreading DNA polymerases generate relatively high error rates, due to the lack of 3′–5′ exonuclease activity, resulting in an overestimation of sequence diversity (Malet *et al.* 2003). Because Dream Taq is a mixture of proofreading and nonproofreading polymerases (Thermo Fisher Scientific, Waltham, MA, USA), we cannot exclude the generation of similar artefacts although these artefacts should have minimal influence when the sequences are clustered at a specific threshold (like in the denoising step of QRS). To summarize, artefacts of both sequencing methods might lead to the detection of some rare artificial sequences. It is thus necessary to improve chimera detection programs. Also, proofreading polymerases should be used to avoid erroneous estimation of sequence frequencies.

Conclusion

Our results further support the idea that 454 generated semiquantitative results are reliable at intraspecific level. Here, based on the ITS1 marker, we were able to identify and characterize variation among closely related populations from the same species, obtaining patterns comparable with our previous Sanger sequencing-based study (Wolinska *et al.* 2014). Applying the QRS pipeline on cloning/Sanger sequencing and 454 pyrosequencing generated data sets, the same representative sequences were recovered at similar frequencies, despite the

differences between the two sequencing platforms and associated potential methodological errors. Consequently, our results suggest that amplicon sequencing is useful for studies of genetic population structure. Moreover, in response to the rapid development of more advanced NGS technologies, our pipeline was adjusted to analyse Illumina data sets. Its modular nature and flexible interfaces to bioinformatic tools at most steps allow for fast adjustment to data sets generated by future platforms.

Acknowledgements

We thank Sabine Radetzki who helped with molecular work and Christian Wurzbacher for discussions and suggestions for improvements of the QRS pipeline needed to adjust for the analysis of Illumina data sets. We also thank the editor (Travis Glenn) and four anonymous reviewers for their valuable comments which helped us to improve the manuscript. Finally, we thank the participants of the Priority Programme DFG-SPP 1399 meeting on Host–Parasite Coevolution in Blossin (Germany) for intensive discussion. This work was supported by DFG-SPP 1399 (WO 1587/2-2) and DFG-SNF (WO 1587/3-1) grants to J.W., and by the STRESSFLEA project of the European Science Foundation EUROCORES Programme EuroEEFG (funded through the Czech Science Foundation project EEF/10/E022 to A.P.). D.L. was supported by the German Science Foundation (DFG) through grant number ME 3134/4-1.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Amend AS, Seifert KA, Bruns TD (2010) Quantifying microbial communities with 454 pyrosequencing: does read abundance count? *Molecular Ecology*, **19**, 5555–5565.

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, **71**, 7724–7736.

Baldrian P, Větrovský T, Cajthaml T *et al.* (2013) Estimation of fungal biomass in forest litter and soil. *Fungal Ecology*, **6**, 1–11.

Bálint M, Schmidt PA, Sharma R, Thines M, Schmitt I (2014) An Illumina metabarcoding pipeline for fungi. *Ecology and Evolution*, **4**, 2642–2653.

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics*, **141**, 743–753.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bittner K, Rothhaupt KO, Ebert D (2002) Ecological interactions of the microparasite *Caullerya mesnili* and its host *Daphnia galeata*. *Limnology and Oceanography*, **47**, 300–305.

Bradley RK, Roberts A, Smoot M *et al.* (2009) Fast statistical alignment. *PLoS Computational Biology*, **5**, e1000392.

Brannock PM, Waits DS, Sharma J, Halanych KM (2014) High-throughput sequencing characterizes intertidal meiofaunal in northern Gulf of Mexico (Dauphin Island and Mobile Bay, Alabama). *Biological Bulletin*, **227**, 161–174.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chen H, Strand M, Norenburg JL *et al.* (2010) Statistical parsimony networks and species assemblages in cephalotrichid nemerteans (Nemertea). *PLoS ONE*, **5**, e12885.

Churchill G (1989) Stochastic models for heterogeneous DNA sequences. *Bulletin of Mathematical Biology*, **51**, 79–94.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cline M, Hughey R, Karplus K (2002) Predicting reliable regions in protein sequence alignments. *Bioinformatics*, **18**, 306–314.

Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN (2013) Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count? *Molecular Ecology Resources*, **13**, 620–633.

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, **36**, e105.

Dolan P, Denver D (2008) TileQC: A system for tile-based quality control of Solexa data. *BMC Bioinformatics*, **9**, 250.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, **10**, 996–998.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Edgcomb V, Orsi W, Bunge J *et al.* (2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME Journal*, **5**, 1344–1356.

Erlich Y, Mitra PP, delaBastide M, McCombie WR, Hannon GJ (2008) Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nature Methods*, **5**, 679–682.

Excoffier L, Lischer HEL (2010) ARLEQUIN suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, **39**, W29–W37.

Forns X, Bukh J, Purcell RH, Emerson SU (1997) How *Escherichia coli* can bias the results of molecular cloning: preferential selection of defective genomes of hepatitis C virus during the cloning procedure. *Proceedings of the National Academy of Sciences*, **94**, 13909–13914.

Giessler S, Wolinska J (2013) Capturing the population structure of microparasites: using ITS-sequence data and a pooled DNA approach. *Molecular Ecology Resources*, **13**, 918–928.

Haas BJ, Gevers D, Earl AM *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, **21**, 494–504.

Harismendy O, Ng P, Strausberg R *et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*, **10**, R32.

Huber JA, Mark Welch DB, Morrison HG *et al.* (2007) Microbial population structures in the deep marine biosphere. *Science*, **318**, 97–100.

Hugall A, Stanton J, Moritz C (1999) Reticulate evolution and the origins of ribosomal internal transcribed spacer diversity in apomictic *Meloidogyne*. *Molecular Biology and Evolution*, **16**, 157–164.

Huse S, Huber J, Morrison H, Sogin M, Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, **8**, R143.

IBM (2011) IBM SPSS Statistics for Windows Version 20.0. IBM Corp., Armonk, NY, USA.

Jünemann S, Prior K, Szczepanowski R *et al.* (2012) Bacterial community shift in treated periodontitis patients revealed by Ion Torrent 16S rRNA gene amplicon sequencing. *PLoS ONE*, **7**, e41606.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.

Kauserud H, Kumar S, Brysting A, Nordén J, Carlsen T (2012) High consistency between replicate 454 pyrosequencing analyses of ectomycorrhizal plant root samples. *Mycorrhiza*, **22**, 309–315.

Kautz S, Rubin BER, Russell JA, Moreau CS (2013) Surveying the microbiome of ants: comparing 454 pyrosequencing with traditional methods to uncover bacterial diversity. *Applied and Environmental Microbiology*, **79**, 525–534.

Knittweis L, Kraemer W, Timm J, Kochzius M (2009) Genetic structure of *Heliofungia actiniformis* (Scleractinia: Fungiidae) populations in the Indo-Malay Archipelago: implications for live coral trade management efforts. *Conservation Genetics*, **10**, 241–249.

Koch MA, Dobeš C, Mitchell-Olds T (2003) Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). *Molecular Biology and Evolution*, **20**, 338–350.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, **79**, 5112–5120.

Kröber M, Bekel T, Diaz NN *et al.* (2009) Phylogenetic characterization of a biogas plant microbial community integrating clone library 16S-rDNA sequences and metagenome sequence data obtained by 454-pyrosequencing. *Journal of Biotechnology*, **142**, 38–49.

Kuulasmaa T (2010) *Oligo Explorer 1.5*. University of Eastern Finland, Kuopio, Finland.

Lassmann T, Sonnhammer E (2005) Kalign – an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics*, **6**, 1–9.

Li W, Fu L, Niu B, Wu S, Wooley J (2012) Ultrafast clustering algorithms for metagenomic sequence analysis. *Briefings in Bioinformatics*, **13**, 656–668.

Liang B, Luo M, Scott-Herridge J *et al.* (2011) A comparison of parallel pyrosequencing and Sanger clone-based sequencing and its impact on the characterization of the genetic diversity of HIV-1. *PLoS ONE*, **6**, e26745.

Lindner DL, Carlsen T, Henrik Nilsson R *et al.* (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. *Ecology and Evolution*, **3**, 1751–1764.

Lohr JN, Laforsch C, Koerner H, Wolinska J (2010) A Daphnia parasite (*Caullerya mesnili*) constitutes a new member of the Ichthyosporea, a group of protists near the Animal-Fungi divergence. *Journal of Eukaryotic Microbiology*, **57**, 328–336.

Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 10557–10562.

Lyras D, Metzler D (2014) ReformAlign: improved multiple sequence alignments using a profile-based meta-alignment approach. *BMC Bioinformatics*, **15**, 265.

Malet I, Belnard M, Agut H, Cahour A (2003) From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *Journal of Virological Methods*, **109**, 161–170.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

Nakamura K, Oshima T, Morimoto T *et al.* (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Research*, **39**, e90.

Palatinszky M, Nikolausz M, Sváb D, Márialigeti K (2011) Preferential ligation during TA-cloning of multitemplate PCR products — A factor causing bias in microbial community structure analysis. *Journal of Microbiological Methods*, **85**, 131–136.

Piry S, Guivier E, Realini A, Martin J-F (2012) ISEISAMIE BARCODE: NGS-oriented software for amplicon characterization – application to species and environmental barcoding. *Molecular Ecology Resources*, **12**, 1151–1157.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution*, **16**, 37–45.

Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. *Journal of Molecular Evolution*, **54**, 396–402.

Pruesse E, Peplies J, Glöckner FO (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics*, **28**, 1823–1829.

Quail M, Smith M, Coupland P *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

Quince C, Lanzen A, Curtis TP *et al.* (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods*, **6**, 639–641.

R Core Team (2014) *R: A Language and Environmental for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reeder J, Knight R (2009) The "rare biosphere": a reality check. *Nature Methods*, **6**, 636–637.

Renaud G, Kircher M, Stenzel U, Kelso J (2013) freeIbis: an efficient basecaller with calibrated quality scores for Illumina sequencers. *Bioinformatics*, **29**, 1208–1209.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Rothberg JM, Hinz W, Rearick TM *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, **475**, 348–352.

Rougemont J, Amzallag A, Iseli C *et al.* (2008) Probabilistic base calling of Solexa sequencing data. *BMC Bioinformatics*, **9**, 431.

Russell D, Otu H, Sayood K (2008) Grammar-based distance in progressive multiple sequence alignment. *BMC Bioinformatics*, **9**, 306.

Sahraeian SME, Yoon B-J (2010) PicXAA: greedy probabilistic construction of maximum expected accuracy alignment of multiple sequences. *Nucleic Acids Research*, **38**, 4917–4928.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sauder JM, Arthur JW, Dunbrack RL Jr (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. *Proteins*, **40**, 6–22.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing MOTHUR: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE*, **6**, e27310.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**, 863–864.

Schoebel CN, Wolinska J, Spaak P (2010) Higher parasite resistance in *Daphnia* populations with recent epidemics. *Journal of Evolutionary Biology*, **23**, 2370–2376.

Schröder J, Bailey J, Conway T, Zobel J (2010) Reference-free validation of short read data. *PLoS ONE*, **5**, e12681.

Seda J, Petrusek A, Machacek J, Smilauer P (2007) Spatial distribution of the *Daphnia longispina* species complex and other planktonic crustaceans in the heterogeneous environment of canyon-shaped reservoirs. *Journal of Plankton Research*, **29**, 619–628.

Sievers F, Wilm A, Dineen D *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, **7**, 539.

Sommer S, Courtiol A, Mazzoni C (2013) MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics*, **14**, 542.

Sorek R, Zhu Y, Creevey CJ *et al.* (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, **318**, 1449–1452.

Tedersoo L, Nilsson RH, Abarenkov K *et al.* (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytologist*, **188**, 291–301.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.

Wang GCY, Wang Y (1996) The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. *Microbiology*, **142**, 1107–1114.

Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, **29**, 2487–2489.

Wolinska J, Spaak P (2009) The cost of being common: evidence from natural *Daphnia* populations. *Evolution*, **63**, 1893–1901.

Wolinska J, Bittner K, Ebert D, Spaak P (2006) The coexistence of hybrid and parental *Daphnia*: the role of parasites. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1977–1983.

Wolinska J, Spaak P, Koerner H *et al.* (2011) Transmission mode affects the population genetic structure of *Daphnia* parasites. *Journal of Evolutionary Biology*, **24**, 265–273.

Wolinska J, Petrusek A, Yin M, Koerner H, Seda J, Giessler S (2014) Population structure of a microparasite infecting *Daphnia*: spatio-temporal dynamics. *BMC Evolutionary Biology*, **14**, 207.

Yilmaz P, Parfrey LW, Yarza P *et al.* (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, **42**, D643–D648.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.

Zhuang F, Fuchs RT, Sun Z, Zheng Y, Robb GB (2012) Structural bias in T4 RNA ligase-mediated 3′-adapter ligation. *Nucleic Acids Research*, **40**, e54.

J.W. applied for funding and coordinated this project. J.W. and A.P. designed the study. J.R. designed sequencing primers and established the PCR protocol, while A.P. prepared and performed the pyrosequencing reactions. E.G.T. developed the QRS pipeline. D.L., S.Gr. and F.C.M. contributed to the development of QRS pipeline in earlier stages. E.G.T., with the help of S.G., performed statistical analyses and, together with J.W., S.G. and A.P., wrote the manuscript. D.L. and S.Gr. helped with writing the Supporting Information. All authors approved the final version of the manuscript and Supporting Information.

Data accessibility

Raw 454 data set from *C. mesnili* populations is available in Sequence Reads Archive (SRA) database under the Accession no. SRP052909. The representative ITS1 sequences are available in GenBank sequence database under the following Accession nos: KJ473407-KJ473421. DNA alignment of 3700 ITS1 sequences in FASTA format with a supporting information file is available in DRYAD (doi:10.5061/dryad.v7f7r). Finally, the QRS source code and updated manual are available at http://code.google.com/p/quantification-representative-sequences.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1 Quantification of Representative Sequences (QRS). Manual.

Data S2 Assessment of effects of the de-noising step in the QRS pipeline.

Data S3 Validation of the QRS pipeline on an Illumina dataset.

Table S1 Number of *Caullerya mesnili* ITS1 sequences generated in the 454 run as well as number of retrieved sequences in the 'filtered 454' and 'raw 454' datasets (provided separately per each of 16 analysed population samples).

Table S2 Type and number of representative ITS1 sequence variants as assigned by statistical parsimony analysis.

Table S3 Brief summary of all parameters of the QRS pipeline when it is executed in batch mode.

Table S4 Comparison of the frequencies of representative ITS1 variants between the 'filtered 454' and 'Sanger' datasets after excluding the most abundant representative ITS1 sequence type (C2.14) across all samples (provided separately per each of 16 analysed population samples).

Fig. S1 Comparison of the results generated by three pipelines (QRS, UPARSE and mothur) that were used for validation of the Illumina dataset.

Chapter 3 – *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes.

E. González-Tortuero[1,2,3,*], J. Rusek[3], P. Turko[4,5], A. Petrusek[6], I. Maayan[3], L. Piálek[6,7], C. Tellenbach[4], S. Gießler[3], P. Spaak[4,5], J. Wolinska[1,8].

Zoology (2016) In press

[1]Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin (Germany).

[2]Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin (Germany).

[3]Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, 82512 Planegg-Martinsried (Germany).

[4]Department of Aquatic Ecology, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland.

[5]Institute of Integrative Biology, ETH Zurich, 8092 Zurich, Switzerland.

[6]Department of Ecology, Faculty of Sciences, Charles University in Prague, Viničná 7, 128 44 Prague, Czech Republic.

[7]Department of Zoology, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic.

[8]Department of Biology, Chemistry and Pharmacy, Institute of Biology, Freie Universitat Berlin, Konigin-Luise-Straße 1-3, 14195 Berlin, Germany.

[*]**E-mail:** gonzalez@igb-berlin.de or enriquegleztortuero@gmail.com

G Model
ZOOL-25500; No. of Pages 8
ARTICLE IN PRESS

Zoology xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Zoology

journal homepage: www.elsevier.com/locate/zool

ZOOLOGY

Daphnia parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes

Enrique González-Tortuero[a,b,c,*], Jakub Rusek[c], Patrick Turko[d,e], Adam Petrusek[f], Inbar Maayan[c], Lubomír Piálek[f,g], Christoph Tellenbach[d], Sabine Gießler[c], Piet Spaak[d,e], Justyna Wolinska[a,h]

[a] Department of Ecosystem Research, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, D-12587 Berlin, Germany
[b] Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, D-14195 Berlin, Germany
[c] Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, D-82512 Planegg-Martinsried, Germany
[d] Department of Aquatic Ecology, Swiss Federal Institute of Aquatic Science and Technology (Eawag), Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
[e] Institute of Integrative Biology, ETH Zurich, Universitätstrasse 16, CH-8092 Zurich, Switzerland
[f] Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, CZ-12844 Prague, Czech Republic
[g] Department of Zoology, Faculty of Science, University of South Bohemia, Branišovská 31, CZ-37005 České Budějovice, Czech Republic
[h] Department of Biology, Chemistry and Pharmacy, Institute of Biology, Freie Universität Berlin, Königin-Luise-Straße 1-3, D-14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 20 November 2015
Received in revised form 2 March 2016
Accepted 20 April 2016
Available online xxx

Keywords:
Caullerya mesnili
Host-genotype specificity
Negative frequency-dependent selection
Parasite population dynamics

ABSTRACT

Studies of parasite population dynamics in natural systems are crucial for our understanding of host–parasite coevolutionary processes. Some field studies have reported that host genotype frequencies in natural populations change over time according to parasite-driven negative frequency-dependent selection. However, the temporal patterns of parasite genotypes have rarely been investigated. Moreover, parasite-driven negative frequency-dependent selection is contingent on the existence of genetic specificity between hosts and parasites. In the present study, the population dynamics and host-genotype specificity of the ichthyosporean *Caullerya mesnili*, a common endoparasite of *Daphnia* water fleas, were analysed based on the observed sequence variation in the first internal transcribed spacer (ITS1) of the ribosomal DNA. The *Daphnia* population of lake Greifensee (Switzerland) was sampled and subjected to parasite screening and host genotyping during *C. mesnili* epidemics of four consecutive years. The ITS1 of wild-caught *C. mesnili*-infected *Daphnia* was sequenced using the 454 pyrosequencing platform. The relative frequencies of *C. mesnili* ITS1 sequences differed significantly among years: the most abundant *C. mesnili* ITS1 sequence decreased and rare sequences increased over the course of the study, a pattern consistent with negative frequency-dependent selection. However, only a weak signal of host-genotype specificity between *C. mesnili* and *Daphnia* genotypes was detected. Use of cutting edge genomic techniques will allow further investigation of the underlying micro-evolutionary relationships within the *Daphnia–C. mesnili* system.

1. Introduction

One of the most important questions in evolutionary biology is why sexual reproduction persists when it usually has a 50% fitness disadvantage compared to asexual reproduction (Maynard Smith, 1978). One widely accepted explanation is that reciprocal selection arising from host–parasite interactions should accelerate evolutionary rates through the need for continual adaptation and counter-adaptation. Such rapid reciprocal adaptation can be driven by either directional selection or negative frequency-dependent selection (NFDS) (Gaba and Ebert, 2009; Brockhurst and Koskella, 2013; Papkou et al., 2016, current issue). In the former case, recurrent selective sweeps of novel host resistance and parasite infectivity alleles occur through time, leading to increases in the parasite's host range and in host resistance traits. In the latter case, common genotypes of a given host have a higher probability of becoming infected by coevolving parasites than do rare genotypes. Rare uninfected host genotypes thus tend to have a fitness advantage and eventually replace the

* Corresponding author at: Department of Ecosystem Research, Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, D-12587 Berlin, Germany.
E-mail addresses: gonzalez@igb-berlin.de, enriquegleztortuero@gmail.com (E. González-Tortuero).

http://dx.doi.org/10.1016/j.zool.2016.04.003
0944-2006/

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

previously common host genotypes. At this point, parasite genotypes that are able to infect surviving host genotypes will be favoured over those that are still adapted to previously common hosts. Parasite-driven NFDS could explain the long-term maintenance of sex in host and parasite populations (reviewed in Lively, 2010), since sex is efficient at promoting genetic diversity.

A prediction of NFDS is that common hosts (and thus, common parasites which are adapted to those common hosts) are at a disadvantage. Field studies of various systems have demonstrated that host frequencies are tracked by parasites, including the systems of *Daphnia* water fleas and microparasites (e.g., Decaestecker et al., 2007; Wolinska and Spaak, 2009), bryozoans and myxozoans (Vernon et al., 1996), plants and fungi (e.g., Burdon and Thompson, 1995; Siemens and Roy, 2005) and freshwater snails and trematodes (e.g., Jokela et al., 2009; King et al., 2009). However, the majority of field surveys investigated temporal changes in relative genotype frequencies only for hosts. This exclusion of parasites is striking because NFDS predicts changes to both host and parasite frequencies.

A prerequisite for NFDS is the existence of genetic specificity in host–parasite interactions, where the outcome of infection depends on the genotypic identity of both host and parasite (Dybdahl et al., 2014). Genotype-by-genotype interactions have been demonstrated in several host–parasite systems under experimental conditions (reviewed in Sadd and Schmid-Hempel, 2009). Thus, the pattern that might be expected under natural conditions is that parasite populations within hosts differ in their genetic composition depending on the host genotypes they infect (e.g., Lythgoe, 2002; Schmid-Hempel and Funk, 2004).

Water fleas of the genus *Daphnia* (Crustacea: Cladocera) and their microparasites were recently proposed as a model system to address coevolutionary questions (Ebert, 2008; Gaba and Ebert, 2009). One of the most common microparasites infecting *Daphnia* of large European lakes is the ichthyosporean *Caullerya mesnili* (Wolinska et al., 2007; Lohr et al., 2010). *C. mesnili* is a highly virulent parasite that reduces the survival and reproduction rate of its hosts (Wolinska et al., 2006). It has also been shown to be involved in driving frequencies of host species' abundance and genotypes (Wolinska et al., 2006). Moreover, *Daphnia* genotype dynamics have been associated with NFDS; the most common host genotypes decreased in frequency in infected but not in uninfected populations (Wolinska and Spaak, 2009). However, *C. mesnili* population dynamics over longer time scales have not been explored in detail. In the present study, we analyse the population dynamics of *C. mesnili* and its host-genotype specificity in a natural *Daphnia* population during parasite epidemics of four consecutive years, based on the observed variation in the first internal transcribed spacer ITS1 marker.

2. Materials and methods

2.1. Study site

Greifensee (N 47°20′41″, E 8°40′21″) is a eutrophic peri-alpine lake in Switzerland. The cladoceran community of this lake is dominated by hybridising members of the *Daphnia longispina* complex (Brede et al., 2009). Epidemics of *C. mesnili* in Greifensee have been documented in previous studies (Wolinska et al., 2004, 2006). Ongoing work since 2002 demonstrates that parasite prevalence in this lake follows an epidemic pattern, with infection of 20–30% of the *Daphnia* population during autumn, and parasite presence dropping to undetectable levels during the rest of the year (Fig. 1) (Wolinska et al., 2006).

2.2. Daphnia *sampling*

Daphnia specimens from Greifensee have been regularly sampled for *C. mesnili* screening (since 2002) and for microsatellite host genotyping (since 2007), biweekly or monthly (in winter). At the deepest point of the lake, zooplankton samples were taken using a 250 μm net. In the laboratory, ~80 adult *Daphnia* individuals were randomly chosen and visually screened for *C. mesnili* infection, using a stereomicroscope (Lohr et al., 2010). Then, during each *C. mesnili* epidemic, ~80 additional infected *Daphnia* were collected (i.e. "infected sample"). These infected *Daphnia* were then genotyped using microsatellites (see Section 2.3). Here, we focus our analysis on four *C. mesnili* epidemic peaks, occurring from 2010 to 2013 (Fig. 1). *C. mesnili* DNA was obtained from infected *Daphnia* from those years; one infected sample per epidemic year was analysed (2010-08-10, 2011-09-13, 2012-09-06 and 2013-08-29).

2.3. *Selection of* Daphnia *hosts for characterisation of* C. mesnili *genetic structure*

In order to assign each infected *Daphnia* individual to a particular multilocus genotype (MLG) (Yin et al., 2010), all *Daphnia* specimens from each "infected sample" were characterised by ten polymorphic microsatellite markers (Brede et al., 2006; for details see Data S1 in the supplementary online Appendix). The MLGs were used to assess host genotype abundance distributions in the "infected sample" of *Daphnia* collected during four consecutive *C. mesnili* epidemics (Fig. S1 in the supplementary online Appendix). "Common" *Daphnia* genotypes were defined as those making up more than 5% of the sample, and "rare" genotypes as those which were only detected once in a sample. For each sampling period, 15 randomly selected rare genotypes, as well as all of the common genotypes, were taken for *C. mesnili* genetic analyses (genotypes neither "common" nor "rare" were excluded). If a common genotype was represented by more than ten individuals, ten specimens were randomly selected for subsequent analysis; otherwise all individual *Daphnia* belonging to a given common genotype were analysed.

2.4. *Molecular analyses of* C. mesnili *ITS1 region*

Primers amplifying the ITS1 region of *C. mesnili* were constructed by fusing a specific core primer sequence (forward: ACACCGCCCGTCACTACTAC and reverse: TGGATATACCACTCTCAAACAG) with a basal 25-mer for binding to the DNA capture beads (Lib-A) and a 10-base multiplex identifier (MID) chosen from the 454 Standard MID Set (Roche, Basel, Switzerland) targeting the approximately 425 bp long ITS1 region (González-Tortuero et al., 2015). DNA samples previously used for *Daphnia* microsatellite genotyping were purified by re-precipitation in 70% EtOH and resuspended in TE buffer. *C. mesnili* from infected *Daphnia* DNA samples was amplified using the following protocol: KAPA2 G Robust Ready Mix (Kappa Biosystems, Wilmington, MA, USA), 0.3 μM each of forward and reverse primer, 0.8 mM BSA, and 2 μl genomic DNA, for a total reaction volume of 14 μl. The PCR cycling parameters included initial denaturation at 94 °C for 3 min, followed by 40 cycles of 94 °C for 30 s, 53 °C for 30 s, and 72 °C for 30 s, followed by a final elongation at 72 °C for 7 min. The success of amplification was verified by agarose gel electrophoresis.

To construct a 454 library, PCR products were purified independently for each MID-labelled sample (i.e., isolated from each infected *Daphnia*) with a Gel/PCR DNA Fragments Extraction Kit (Geneaid Biotech, Taipei, Taiwan), their DNA concentration was measured on a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and then they were pooled in equimolar concentrations. The final poolplex was further purified by separation

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

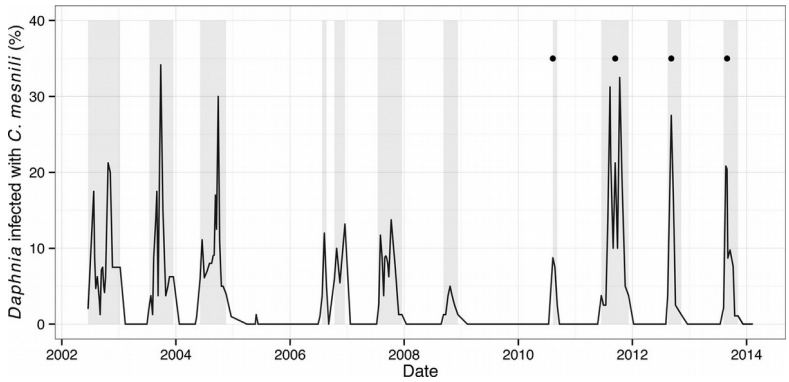

Fig. 1. *C. mesnili* prevalence in the *Daphnia* population of Greifensee, 2002 to 2014. Epidemic periods in which the prevalence of infection exceeded 2% are indicated in dark grey. Samples collected for genetic analyses are shown with a dot. X-axis labels indicate beginning of year.

on Pippin Prep (Pippin Prep Kit CSD2010; size-selection range 400–580 bp; Sage Science, Beverly, MA, USA) and purified with SPRI AMPure XP paramagnetic beads (Beckman Coulter Genomics, Danvers, MA, USA). Emulsion PCR and pyrosequencing were performed with the amplicon (Lib-A) kit, using GS Junior reagents and the manufacturer's protocols (454 Life Sciences, Branford, CT, USA), with the input ratio of DNA molecule-to-bead decreased to 0.4. DNA bead enrichment level was within the expected range (6%). The raw 454 dataset is available in the Sequence Read Archive (SRA) under accession number SRP065924.

2.5. *Identifying representative ITS1 sequences of* C. mesnili

The raw *C. mesnili* dataset was analysed using the Quantification of Representative Sequences (QRS) pipeline (González-Tortuero et al., 2015) to determine the number and frequency of representative sequence variants. A representative sequence is defined as the most abundant sequence per sequence cluster, and could be considered the presumed ancestral allelic reference (González-Tortuero et al., 2015). This reduction of variation is useful for population structure analysis when a multicopy marker (like the ITS1) is considered, as it helps mitigate potential overestimation of polymorphism (Giessler and Wolinska, 2013) as well as minor sequencing errors. These representative sequences were subsequently used instead of the raw data. The pipeline was run with default parameters (unless indicated otherwise; see Data S2 in the supplementary online Appendix). The most abundant ITS1 representative sequences are available in the GenBank sequence database under the accession numbers KU094678–KU094681.

2.6. *Haplotype network of* C. mesnili

A haplotype network was constructed for the abundant ITS1 representative sequences ("abundant" refers to representative sequences that reached overall frequencies higher than 0.5%). Connection distances between haplotypes were calculated using TCS (Clement et al., 2000) according to the statistical parsimony algorithm. The output was processed using a force-directed algorithm, implemented in Cytoscape 3.2.1 (Shannon et al., 2003). Additionally, to test if the ITS1 representative sequences detected here were present in previous studies, the abundant *C. mesnili* representative sequences obtained here were compared with those from a previous study (González-Tortuero et al., 2015). In that study, *C. mesnili*-infected *Daphnia* were sampled across seven reservoirs in the Czech Republic (González-Tortuero et al., 2015). Here, all abundant representative sequences were re-aligned using the MUSCLE algorithm (Edgar, 2004) and manually corrected. Subsequently, a haplotype network was created as described above.

2.7. *Temporal variation of* C. mesnili

To investigate temporal variation in *C. mesnili*, two types of analyses were performed. First, the frequencies of *C. mesnili* representative sequences were compared among years, using a Chi-squared test. Second, to describe the temporal trend, a Mann–Kendall trend test (Mann, 1945; Kendall, 1948) was performed on the residuals from the locally weighted scatterplot smoothing (LOWESS) (Cleveland, 1979), with the frequency of *C. mesnili* representative sequences as the dependent variable and time (year) as the independent variable. These statistical tests were performed in R (R Core Team, 2015). The *Kendall* (Hipel and McLeod, 2005) package was used for the Mann–Kendall trend test. For these analyses, all *C. mesnili* sequences were pooled per year, regardless of their host genotype.

2.8. *Host-genotype specificity of* C. mesnili

To assess differences in *C. mesnili* population structure between *Daphnia* genotypes, two types of statistical tests were performed. The tests were run by year (2010, 2011, 2012 and 2013) and only *C. mesnili* sequences detected on common *Daphnia* genotypes were considered. First, an analysis of molecular variance (AMOVA) was performed at three hierarchical levels: within a *Daphnia* individual, within a *Daphnia* genotype (i.e. among individuals that represented the same genotype) and among *Daphnia* genotypes. Second, to visualise the potential differences in the distribution of *C. mesnili* representative sequences sampled from different *Daphnia* genotypes, a non-metric multidimensional scaling (nMDS) plot was constructed, using the Bray–Curtis metric, and a permutational MANOVA was conducted (PERMANOVA) (Anderson, 2001). These statistical tests were performed in R using the *ade4* (Dray and Dufour, 2007) package for the AMOVA test and the *vegan* (Oksanen et al., 2015) and *MASS* (Venables and Ripley, 2003) packages for the nMDS and PERMANOVA.

Please cite this article in press as: González-Tortuero, E., et al., Daphnia parasite dynamics across multiple Caullerya epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

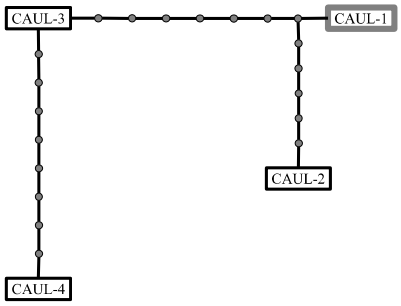

Fig. 2. Haplotype network of the abundant ITS1 representative sequences of *C. mesnili* in Greifensee. Each grey dot indicates a single connection step (i.e. a single mutation) between the ITS1 representative sequences. The grey-lined box indicates the most abundant ITS1 representative sequence.

3. Results

3.1. *Selection of* Daphnia *hosts for characterisation of* C. mesnili *genetic structure*

The genotype frequency distributions of infected *Daphnia* specimens fitted the expected pattern of many rare and few common genotypes (Fig. S1 in the supplementary online Appendix). There were 10 *Daphnia* genotypes that met the "common genotype" threshold: three in 2010 and 2013, and two in 2011 and 2012 (Table S1 in the supplementary online Appendix). An additional 15 *Daphnia* individuals per sampling year were randomly selected from the rare genotype pool. In total, *C. mesnili* ITS1 was separately sequenced from 145 infected *Daphnia* individuals (Table S1).

3.2. *Description of the* C. mesnili *ITS1 dataset*

Processing with the QRS pipeline yielded 20,645*C. mesnili* ITS1 sequences (out of 76,719 available sequences). The majority of the sequences removed were singletons. Moreover, if less than 10 sequences remained per *Daphnia* host, these *Daphnia* hosts were excluded from further analyses, resulting in 138 analysed *Daphnia* (out of the 145 that were originally sequenced; Table S1). The average number of *C. mesnili* ITS1 sequences retained per *Daphnia* was 149.5, with a standard deviation of 82.9. The length of the alignment was 460 bp (see the Fasta file in the supplementary online Appendix).

3.3. *Identifying representative ITS1 sequences of* C. mesnili

1,062 unique representative sequences were detected in the *C. mesnili* dataset. Across the entire dataset (i.e. all *Daphnia* genotypes and years pooled), the most abundant *C. mesnili* ITS1 representative sequences reached a frequency of 77.45% (CAUL-1), 7.89% (CAUL-2), 1.83% (CAUL-3) and 0.86% (CAUL-4). The remaining 1,058 representative sequences were present at proportions lower than 0.5% and were classified as rare.

3.4. *Haplotype network*

The haplotype network had an almost linear structure, with the most abundant *C. mesnili* ITS1 representative sequence (CAUL-1) placed between the other abundant representative sequences (Fig. 2). The second most abundant *C. mesnili* representative sequence (CAUL-2) matched the C2.14-type from a previous study (González-Tortuero et al., 2015), which was the most abundant type in that study. In a joint haplotype network of representative sequences (i.e. those from the present study as well as from the previous study by González-Tortuero et al., 2015), the two most abundant representative sequences (CAUL-1 and C2.14-type) were present at the centre of the network (Fig. S2 in the supplementary online Appendix).

Table 1
Results of AMOVA tests of the distribution of *C. mesnili* ITS1 representative sequences in Greifensee, per year. Only *C. mesnili* sequences originating from "common" *Daphnia* genotypes were included in these analyses.

Year	Source of variation	df	Percent variation	*P*
2010	Among *Daphnia* genotypes	2	−0.05	0.003
	Within a *Daphnia* genotype	23	1.65	<0.001
	Within a *Daphnia* individual	5235	98.40	<0.001
2011	Among *Daphnia* genotypes	1	0.21	0.007
	Within a *Daphnia* genotype	13	0.69	<0.001
	Within a *Daphnia* individual	2137	99.10	<0.001
2012	Among *Daphnia* genotypes	1	−0.06	0.067
	Within a *Daphnia* genotype	14	1.70	<0.001
	Within a *Daphnia* individual	2550	98.36	<0.001
2013	Among *Daphnia* genotypes	2	0.22	0.050
	Within a *Daphnia* genotype	20	2.94	<0.001
	Within a *Daphnia* individual	1900	96.83	<0.001

3.5. *Temporal variation of* C. mesnili

The relative frequencies of the *C. mesnili* ITS1 representative sequences differed among the years analysed (Chi-squared test: $\chi^2 = 1138.4$, $P < 2.2 \times 10^{-16}$; Fig. 3). A decrease in the most abundant *C. mesnili* ITS1 representative sequence (CAUL-1) and an increase in the rare representative sequences were the most pronounced trends. The significant changes in the relative abundance of representative frequencies over time were confirmed by a Mann–Kendall test (Fig. 4).

3.6. *Host-genotype specificity of* C. mesnili

Differences in the distribution of *C. mesnili* ITS1 representative sequences among different *Daphnia* genotypes were tested with an AMOVA (separately per epidemic year). The largest amount of variation was observed at the "within a *Daphnia* individual" level (up to 99.1% in 2011). Only a very small proportion of the variation was explained by differences in the frequencies of *C. mesnili* ITS1 representative sequences among *Daphnia* genotypes (this level of variation was significant in 2010 and 2011; Table 1). Differences in ITS1-based *C. mesnili* genetic structure between the *Daphnia* genotypes were visualised in nMDS plots, where the position of individual points is based on the frequencies of specific ITS1 representative sequences per *Daphnia* host (Fig. 5). The genotype centroids overlapped in 2010 (for two of the three analysed genotypes) and in 2011 (for both analysed genotypes). In 2012 and 2013, the group centroids were distinct. The results of the nMDS plots were supported by PERMANOVA tests, which indicated significant differences in the presence and abundance of representative *C. mesnili* ITS1 sequences among *Daphnia* genotypes in 2013 only (2010: $F = 0.782$, $P = 0.470$; 2011: $F = 0.590$, $P = 0.631$; 2012: $F = 2.220$, $P = 0.104$; 2013: $F = 2.342$, $P = 0.038$).

4. Discussion

In natural populations, tracking parasite dynamics over time remains difficult (reviewed in Penczykowski et al., 2015). Except for the *Trichostrongylus tenuis*–red grouse system (Hudson and Dobson, 1997) and the St. Kilda Soay Sheep Project (e.g., Wilson et al., 2004), the majority of studies about temporal changes in parasite populations to date have been performed in the laboratory under controlled conditions (e.g., Koskella and Lively, 2009; Schulte

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

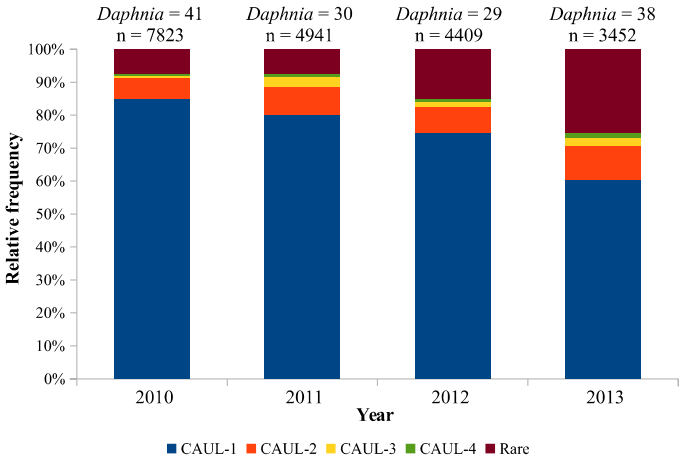

Fig. 3. Comparison of relative frequencies of the ITS1 representative sequences of *C. mesnili* among epidemic years in Greifensee, 2010 to 2013. The number of infected *Daphnia* individuals that were used for *C. mesnili* ITS1 sequencing and the number of analysed ITS1 sequences are shown above each stacked bar graph.

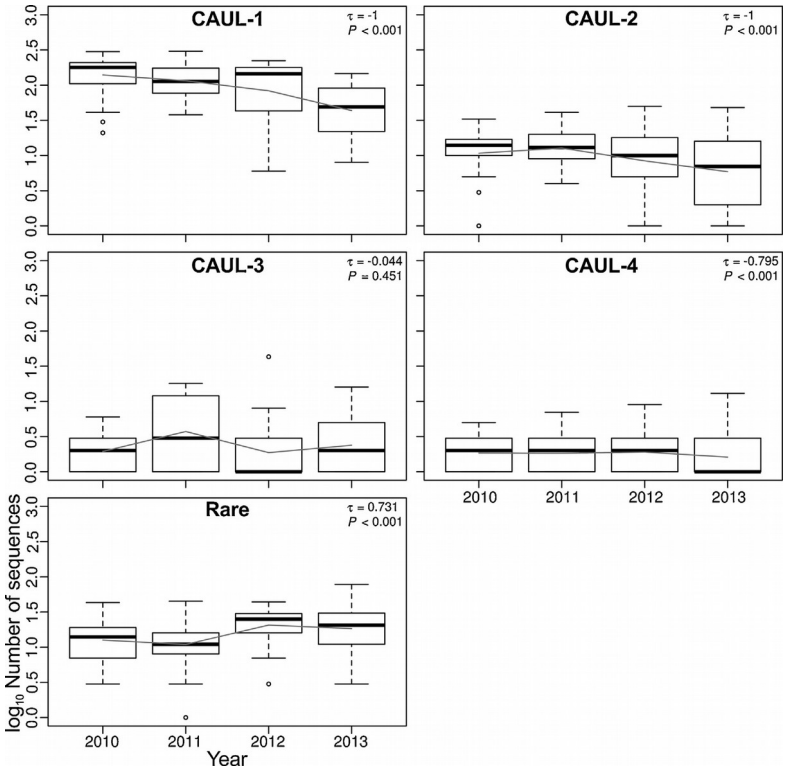

Fig. 4. Change over time in the number of *C. mesnili* representative sequences (i.e. CAUL-1, CAUL-2, CAUL-3, CAUL-4, and rare group) in Greifensee, 2010 to 2013. The grey line connecting the boxes describes the temporal trend of the representative sequences' abundance based on the LOWESS smoothing graph. Results from a Mann–Kendall trend test (i.e. Kendall's τ and P value) are shown in the top right corner.

et al., 2010). Although such experiments are necessary to reduce environmental noise which can otherwise conceal important factors and processes, they tend to oversimplify natural conditions. Since parasite population dynamics are highly sensitive to environmental changes (reviewed in Wolinska and King, 2009), studies

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

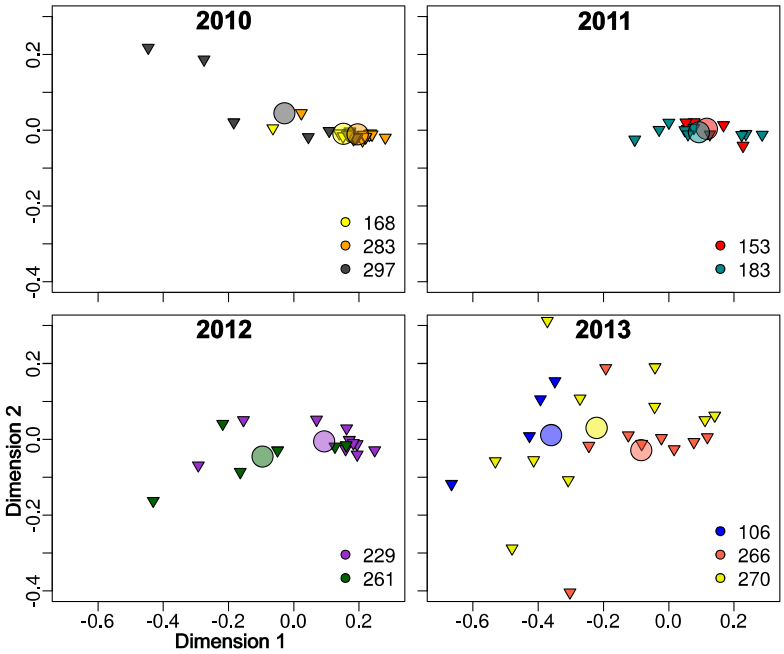

Fig. 5. Non-metric MDS plots of the structure of the *C. mesnili* population infecting *Daphnia* genotypes in Greifensee, 2010 to 2013. The analysis was based on the frequencies of ITS1 representative sequences of *C. mesnili* calculated per individual *Daphnia* host. Triangles represent individual *Daphnia* hosts while circles indicate the centroids per *Daphnia* genotype (different *Daphnia* genotypes are marked by different colours). "Stress" (the rank dissimilarities between the distance matrix and the plotted distances) was 0.079 for all plots.

under natural conditions are essential to understand host–parasite coevolution.

In the present study, we detected a decrease of the most abundant *C. mesnili* ITS1 representative sequence (CAUL-1) and an increase of the rare representative sequences in a natural *Daphnia* population over the four study years. This observation is in line with NFDS, as common parasites which presumably are adapted to the most common host genotypes are at a disadvantage and must then decrease in abundance. In a similar study examining seven drinking water reservoirs in the Czech Republic, the frequencies of *C. mesnili* ITS1 representative sequences changed across time as well (Wolinska et al., 2014). However, here we were able to analyse a much larger sequence dataset; the number of *C. mesnili* ITS1 representative sequences analysed per infected *Daphnia* was 25–33 in the study by Wolinska et al. (2014) in contrast to 3,452–7,823 sequences in the present study (due to the use of Sanger sequencing and 454 pyrosequencing platforms, respectively).

One alternative explanation that could have led to the trend of parasite genetic change over time observed here could be that different parasite genotypes are being favoured over time due to changes in the external environment (reviewed in Wolinska and King, 2009). In fact, interaction between parasites and temperature was significant in the *C. mesnili–Daphnia* system when assessed experimentally (Schoebel et al., 2011). However, there is no indication that conditions in Greifensee have changed in any particular direction within the four years examined. Another alternative explanation is that a yearly expansion following a bottleneck might explain the trend seen in the data. It is still unknown how *C. mesnili* survives between the epidemics; it may persist either at very low and undetectable densities in the *Daphnia* host, in as yet unknown alternative hosts (like fish) or in the sediment as spores. To be able to discard this potential explanation, population genetic tests are needed.

NFDS is associated with the parasite-driven evolution of sex (reviewed in Brockhurst et al., 2014). In fact, the existence of sexual cycles could potentially explain the origin of the rare *C. mesnili* ITS1 representative sequences. However, cryptic sexuality has only been described in the ichthyosporean shellfish symbiont *Sphaeroforma tapetis* so far (Marshall and Berbee, 2010). Alternative mechanisms might also account for the origin of rare variants. For instance, ribosomal DNA is present in multiple copies throughout the genome and each copy is a potential target for mutations leading to intragenomic variation. It has long been thought that such regions (including the ITS1) evolve under concerted evolution, i.e., become homogenised and evolve as a unit (reviewed in Liao, 1999). Although ITS1 polymorphism is not well studied in the class Ichthyosporea, high intragenomic ITS1 variability has been described in other protozoa such as the human intestinal parasite *Dientamoeba fragilis* (Bart et al., 2008) and the foraminiferan *Elphidium macellum* (Pillet et al., 2012). This high ITS1 variability argues against the concerted evolution of the ribosomal genes (at least in those particular taxa) and suggests a birth-and-death process, where new genes originate from gene duplication and some are maintained in the genomes while others are eliminated or become non-functional (Nei et al., 1997). To elucidate ribosomal gene evolution according to the above hypotheses (i.e. concerted evolution or birth-and-death processes), cytogenetic and genomic techniques should be implemented.

The weak signals of host-genotype specificity we found between *C. mesnili* and *Daphnia* genotypes support the lack of host-genotype specificity in this system, as indicated in a previous study in which we compared the distribution of *C. mesnili* ITS1 sequences between

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

Daphnia hybrids and their parental species (Wolinska et al., 2014). Two main arguments are commonly used to explain host specificity and its related factors: the specialisation of parasites and the relationship between host specificity and parasite transmission mode. Parasites tend to infect specific hosts that are phylogenetically and/or eco(physio)logically related (Adamson and Caira, 1994; Pedersen et al., 2005). In invertebrates, parasite specificity is associated with differential expression of the innate immune system, the genetic diversity of receptors or effectors, dosage effects and the host's microbiota (Schulenburg et al., 2007; Riddell et al., 2009; Koch and Schmid-Hempel, 2012). Host ecology is an important factor that affects specificity when the parasites interact minimally with the host's physiology and/or evade the host immune system (Adamson and Caira, 1994; Schmid-Hempel, 2009). Ecological parameters like composition of the community (e.g., Poulin, 1997; Marcogliese, 2002), host foraging strategies (Salathé and Schmid-Hempel, 2011) and the presence of vectors (reviewed in Hoberg and Brooks, 2008) have been associated with a lack of parasite specificity. In contrast, factors including similarity between habitats (Šimková et al., 2006) and host geographical distribution (reviewed in Poulin et al., 2011) could reinforce host-genotype specificity. Additionally, non-structured parasite populations could be the result of weak genetic interactions or of frequent genetic exchange (Schmid-Hempel and Funk, 2004; Bruyndonckx et al., 2009). Host specificity might also relate to parasite transmission mode; specific forms of transmission might facilitate encounters with multiple types of host (Pedersen et al., 2005). For instance, *Daphnia* become infected when they ingest *C. mesnili* spores during grazing (Lohr et al., 2010). This method of transmission, common in gut parasites, is expected to be a non-specific mode of parasite transmission (Marcogliese, 2002).

Finally, despite its many benefits, ITS1 may not be the optimal marker for host-genotype specificity tests. However, ITS1 sequences are the only known polymorphic marker in *C. mesnili* which could allow for the identification of different parasite strains or variants (Giessler and Wolinska, 2013). In a similar way, the use of ITS1 to discriminate strains or variants was successful in amoebozoa (Köhsler et al., 2006) and trichomonads (Ibáñez-Escribano et al., 2014). Nevertheless, the resolution of this marker is not fine enough to discriminate between strains or variants in oomycetes (Robideau et al., 2011), dinoflagellates (Stern et al., 2012) or in other protozoa (Homan et al., 1997; Lollis et al., 2011). For this reason, strains or variants identified with the ITS1 marker should be confirmed with other neutral markers which identify strains at a higher resolution.

In conclusion, we detected a decrease in the most abundant *C. mesnili* ITS1 representative sequence and an increase in the rare representative sequences over four consecutive epidemics. These findings are consistent with the assumptions of NFDS. However, only weak host-genotype specificity between *C. mesnili* and *Daphnia* was detected in our survey. In future studies, NFDS should be further confirmed by performing simultaneous genetic screening of host and parasite populations.

Acknowledgements

We thank Sabine Radetzki and Esther Keller for their help with molecular work. Esther Keller organised the collection and processing of *Daphnia* samples. We also thank the Editor (Joachim Kurtz) and two anonymous reviewers for their valuable comments which helped us to improve the manuscript. This work was supported by DFG-SPP 1399 (WO 1587/2-2) and a DFG-SNF lead agency grant to Justyna Wolinska (WO 1587/3-1) and Piet Spaak (310030L_135750).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.zool.2016.04.003.

References

Adamson, M.L., Caira, J.N., 1994. Evolutionary factors influencing the nature of parasite specificity. Parasitology 109, S85–S95.
Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 26, 32–46.
Bart, A., van der Heijden, H.M., Greve, S., Speijer, D., Landman, W.J., van Gool, T., 2008. Intragenomic variation in the internal transcribed spacer 1 region of *Dientamoeba fragilis* as a molecular epidemiological marker. J. Clin. Microbiol. 46, 3270–3275.
Brede, N., Thielsch, A., Sandrock, C., Spaak, P., Keller, B., Streit, B., Schwenk, K., 2006. Microsatellite markers for European *Daphnia*. Mol. Ecol. Notes 6, 536–539.
Brede, N., Sandrock, C., Straile, D., Spaak, P., Jankowski, T., Streit, B., Schwenk, K., 2009. The impact of human-made ecological changes on the genetic architecture of *Daphnia* species. Proc. Natl. Acad. Sci. U. S. A. 106, 4758–4763.
Brockhurst, M.A., Koskella, B., 2013. Experimental coevolution of species interactions. Trends Ecol. Evol. 28, 367–375.
Brockhurst, M.A., Chapman, T., King, K.C., Mank, J.E., Paterson, S., Hurst, G.D.D., 2014. Running with the Red Queen: the role of biotic conflicts in evolution. Proc. R. Soc B 281, 20141382.
Bruyndonckx, N., Henry, I., Christe, P., Kerth, G., 2009. Spatio-temporal population genetic structure of the parasitic mite *Spinturnix bechsteini* is shaped by its own demography and the social system of its bat host. Mol. Ecol. 18, 3581–3592.
Burdon, J., Thompson, J., 1995. Changed patterns of resistance in a population of *Linum marginale* attacked by the rust pathogen *Melampsora lini*. J. Ecol. 83, 199–206.
Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1659.
Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829–836.
Decaestecker, E., Gaba, S., Raeymaekers, J.A.M., Stoks, R., Van Kerckhoven, L., Ebert, D., De Meester, L., 2007. Host–parasite Red Queen dynamics archived in pond sediment. Nature 450, 870–873.
Dray, S., Dufour, A.B., 2007. The *ade4* package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Dybdahl, M.F., Jenkins, C.E., Nuismer, S.L., 2014. Identifying the molecular basis of host–parasite coevolution: merging models and mechanisms. Am. Nat. 184, 1–13.
Ebert, D., 2008. Host–parasite coevolution: insights from the *Daphnia*–parasite model system. Curr. Opin. Microbiol. 11, 290–301.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Gaba, S., Ebert, D., 2009. Time-shift experiments as a tool to study antagonistic coevolution. Trends Ecol. Evol. 24, 226–232.
Giessler, S., Wolinska, J., 2013. Capturing the population structure of microparasites: using ITS-sequence data and a pooled DNA approach. Mol. Ecol. Resour. 13, 918–928.
González-Tortuero, E., Rusek, J., Petrusek, A., Gießler, S., Lyras, D., Grath, S., Castro-Monzón, F., Wolinska, J., 2015. The Quantification of Representative Sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*. Mol. Ecol. Resour. 15, 1385–1395.
Hipel, K.W., McLeod, A.I., 2005. Time Series Modelling of Water Resources and Environmental Systems. Elsevier, Amsterdam, Electronic reprint of the book orginally published in 1994 http://www.stats.uwo.ca/faculty/aim/1994Book/.
Hoberg, E.P., Brooks, D.R., 2008. A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. J. Biogeogr. 35, 1533–1550.
Homan, W.L., Limper, L., Verlaan, M., Borst, A., Vercammen, M., van Knapen, F., 1997. Comparison of the internal transcribed spacer, ITS 1, from *Toxoplasma gondii* isolates and *Neospora caninum*. Parasitol. Res. 83, 285–289.
Hudson, P.J., Dobson, A.P., 1997. Transmission dynamics and host–parasite interactions of *Trichostrongylus tenuis* in red grouse (*Lagopus lagopus scoticus*). J. Parasitol. 83, 194–202.
Ibáñez-Escribano, A., Nogal-Ruiz, J.J., Arán, V.J., Escario, J.A., Gómez-Barrio, A., Alderete, J.F., 2014. Determination of internal transcribed spacer regions (ITS) in *Trichomonas vaginalis* isolates and differentiation among *Trichomonas* species. Parasitol. Int. 63, 427–431.
Jokela, J., Dybdahl, M.F., Lively, C.M., 2009. The maintenance of sex, clonal dynamics, and host–parasite coevolution in a mixed population of sexual and asexual snails. Am. Nat. 174, S43–S53.
Kendall, M.G., 1948. Rank Correlation Methods. Charles Griffin, Oxford.
King, K.C., Delph, L.F., Jokela, J., Lively, C.M., 2009. The geographic mosaic of sex and the Red Queen. Curr. Biol. 19, 1438–1441.
Koch, H., Schmid-Hempel, P., 2012. Gut microbiota instead of host genotype drive the specificity in the interaction of a natural host–parasite system. Ecol. Lett. 15, 1095–1103.

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

Köhsler, M., Leitner, B., Blaschitz, M., Michel, R., Aspöck, H., Walochnik, J., 2006. ITS1 sequence variabilities correlate with 18S rDNA sequence types in the genus *Acanthamoeba* (Protozoa: Amoebozoa). Parasitol. Res. 98, 86–93.
Koskella, B., Lively, C.M., 2009. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. Evolution 63, 2213–2221.
Liao, D., 1999. Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. 64, 24–30.
Lively, C.M., 2010. A review of Red Queen models for the persistence of obligate sexual reproduction. J. Hered. 101, 13–20.
Lohr, J.N., Laforsch, C., Koerner, H., Wolinska, J., 2010. A *Daphnia* parasite (*Caullerya mesnili*) constitutes a new member of the Ichthyosporea, a group of protists near the animal–fungi divergence. J. Eukaryot. Microbiol. 57, 328–336.
Lollis, L., Gerhold, R., McDougald, L., Beckstead, R., 2011. Molecular characterization of *Histomonas meleagridis* and other parabasalids in the United States using the 5.8S ITS-1, and ITS-2 rRNA regions. J. Parasitol. 97, 610–615.
Lythgoe, K.A., 2002. Effects of acquired immunity and mating strategy on the genetic structure of parasite populations. Am. Nat. 159, 519–529.
Mann, H., 1945. Nonparametric tests against trend. Econometrica 12, 245–249.
Marcogliese, D.J., 2002. Food webs and the transmission of parasites to marine fish. Parasitology 124, 83–99.
Marshall, W.L., Berbee, M.L., 2010. Population-level analyses indirectly reveal cryptic sex and life history traits of *Pseudoperkinsus tapetis* (Ichthyosporea, Opisthokonta): a unicellular relative of the animals. Mol. Biol. Evol. 27, 2014–2026.
Maynard Smith, J., 1978. The Evolution of Sex. Cambridge University Press, Cambridge.
Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc. Natl. Acad. Sci. U. S. A. 94, 7799–7806.
Oksanen, J., Guillaume Blanchet, F., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2015. *vegan*: Community Ecology R Package v 2.3 https://cran.r-project.org/web/packages/vegan/index.html.
Papkou, A., Gokhale, C.S., Traulsen, A., Schulenburg, H., 2016. Host–parasite coevolution: why changing population size matters. Zoology 119, ##–##.
Pedersen, A.B., Altizer, S., Poss, M., Cunningham, A.A., Nunn, C.L., 2005. Patterns of host specificity and transmission among parasites of wild primates. Int. J. Parasitol. 35, 647–657.
Penczykowski, R.M., Laine, A.-L., Koskella, B., 2015. Understanding the ecology and evolution of host–parasite interactions across scales. Evol. Appl. 9, 37–52.
Pillet, L., Fontaine, D., Pawlowski, J., 2012. Intra-genomic ribosomal RNA polymorphism and morphological variation in *Elphidium macellum* suggests inter-specific hybridization in Foraminifera. PLoS One 7, e32373.
Poulin, R., 1997. Parasite faunas of freshwater fish: the relationship between richness and the specificity of parasites. Int. J. Parasitol. 27, 1091–1098.
Poulin, R., Krasnov, B.R., Mouillot, D., 2011. Host specificity in phylogenetic and geographic space. Trends Parasitol. 27, 355–361.
R Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria http://www.R-project.org.
Riddell, C., Adams, S., Schmid-Hempel, P., Mallon, E.B., 2009. Differential expression of immune defences is associated with specific host–parasite interactions in insects. PLoS One 4, e7621.
Robideau, G.P., De Cock, A.W.A.M., Coffey, M.D., Voglmayr, H., Brouwer, H., Bala, K., Chitty, D.W., Désaulniers, N., Eggertson, Q.A., Gachon, C.M.M., Hu, C.-H., Küpper, F.C., Rintoul, T.L., Sarhan, E., Verstappen, E.C.P., Zhang, Y., Bonants, P.J.M., Ristaino, J.B., Lévesque, C.A., 2011. DNA barcoding of oomycetes with cytochrome *c* oxidase subunit I and internal transcribed spacer. Mol. Ecol. Resour. 11, 1002–1011.
Sadd, B.M., Schmid-Hempel, P., 2009. Principles of ecological immunology. Evol. Appl. 2, 113–121.
Salathé, R.M., Schmid-Hempel, P., 2011. The genotypic structure of a multi-host bumblebee parasite suggests a role for ecological niche overlap. PLoS One 6, e22054.
Schmid-Hempel, P., 2009. Immune defence, parasite evasion strategies and their relevance for macroscopic phenomena such as virulence. Phil. Trans. R. Soc. B 364, 85–98.
Schmid-Hempel, P., Funk, C.R., 2004. The distribution of genotypes of the trypanosome parasite, *Crithidia bombi*, in populations of its host, *Bombus terrestris*. Parasitology 129, 147–158.
Schoebel, C.N., Tellenbach, C., Spaak, P., Wolinska, J., 2011. Temperature effects on parasite prevalence in a natural hybrid complex. Biol. Lett. 7, 108–111.
Schulenburg, H., Boehnisch, C., Michiels, N.K., 2007. How do invertebrates generate a highly specific innate immune response? Mol. Immunol. 44, 3338–3344.
Schulte, R.D., Makus, C., Hasert, B., Michiels, N.K., Schulenburg, H., 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc. Natl. Acad. Sci. U. S. A. 107, 7359–7364.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.
Siemens, D.H., Roy, B.A., 2005. Tests for parasite-mediated frequency-dependent selection in natural populations of an asexual plant species. Evol. Ecol. 19, 321–338.
Šimková, A., Verneau, O., Gelnar, M., Morand, S., 2006. Specificity and specialization of congeneric monogeneans parasitizing cyprinid fish. Evolution. 60, 1023–1037.
Stern, R.F., Andersen, R.A., Jameson, I., Küpper, F.C., Coffroth, M.-A., Vaulot, D., Le Gall, F., Véron, B., Brand, J.J., Skelton, H., Kasai, F., Lilly, E.L., Keeling, P.J., 2012. Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. PLoS One 7, e42780.
Venables, W.N., Ripley, B.D., 2003. Modern Applied Statistics with S, 4th ed. Springer, New York.
Vernon, J.G., Okamura, B., Jones, C.S., Noble, L.R., 1996. Temporal patterns of clonality and parasitism in a population of freshwater bryozoans. Proc. R. Soc. Lond. B 263, 1313–1318.
Wilson, K., Grenfell, B.T., Pilkington, J.G., Boyd, H.E.G., Gulland, F.M.D., 2004. Parasites and their impact. In: Clutton-Brock, T.H., Pemberton, J.M. (Eds.), Soay Sheep: Dynamics and Selection in an Island Population. Cambridge University Press, Cambridge, UK, pp. 113–165.
Wolinska, J., King, K.C., 2009. Environment can alter selection in host–parasite interactions. Trends Parasitol. 25, 236–244.
Wolinska, J., Spaak, P., 2009. The cost of being common: evidence from natural *Daphnia* populations. Evolution 63, 1893–1901.
Wolinska, J., Keller, B., Bittner, K., Lass, S., Spaak, P., 2004. Do parasites lower *Daphnia* hybrid fitness? Limnol. Oceanogr. 49, 1401–1407.
Wolinska, J., Bittner, K., Ebert, D., Spaak, P., 2006. The coexistence of hybrid and parental *Daphnia*: the role of parasites. Proc. R. Soc. B 273, 1977–1983.
Wolinska, J., Keller, B., Manca, M., Spaak, P., 2007. Parasite survey of a *Daphnia* hybrid complex: host-specificity and environment determine infection. J. Anim. Ecol. 76, 191–200.
Wolinska, J., Petrusek, A., Yin, M., Koerner, H., Seda, J., Giessler, S., 2014. Population structure of a microparasite infecting *Daphnia*: spatio-temporal dynamics. BMC Evol. Biol. 14, 247.
Yin, M., Wolinska, J., Gießler, S., 2010. Clonal diversity, clonal persistence and rapid taxon replacement in natural populations of species and hybrids of the *Daphnia longispina* complex. Mol. Ecol. 19, 4168–4178.

Please cite this article in press as: González-Tortuero, E., et al., *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. Zoology (2016), http://dx.doi.org/10.1016/j.zool.2016.04.003

Chapter 4 – Genetic diversity of two *Daphnia*-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region.

E. González-Tortuero[1,2,3,*], J. Rusek[3], I. Maayan[3], A. Petrusek[4], L. Piálek[4,5], S. Laurent[6], J. Wolinska[1,7].

Parasites & Vectors (2016) 9: 293

[1]Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin (Germany).

[2]Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin (Germany).

[3]Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, 82512 Planegg-Martinsried (Germany).

[4]Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague (Czech Republic).

[5]Department of Zoology, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice (Czech Republic).

[6]School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne (Switzerland).

[7]Department of Biology, Chemistry and Pharmacy, Institute of Biology, Free University of Berlin, Königin-Luise-Straße 1-3, 14195 Berlin (Germany).

[*]**E-mail:** gonzalez@igb-berlin.de

González-Tortuero *et al. Parasites & Vectors* (2016) 9:293
DOI 10.1186/s13071-016-1584-4

Parasites & Vectors

RESEARCH Open Access

Genetic diversity of two *Daphnia*-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region

Enrique González-Tortuero[1,2,3*], Jakub Rusek[3], Inbar Maayan[3], Adam Petrusek[4], Lubomír Piálek[4,5], Stefan Laurent[6,8] and Justyna Wolinska[1,7]

Abstract

Background: Microsporidia are spore-forming obligate intracellular parasites that include both emerging pathogens and economically important disease agents. However, little is known about the genetic diversity of microsporidia. Here, we investigated patterns of geographic population structure, intraspecific genetic variation, and recombination in two microsporidian taxa that commonly infect cladocerans of the *Daphnia longispina* complex in central Europe. Taken together, this information helps elucidate the reproductive mode and life-cycles of these parasite species.

Methods: Microsporidia-infected *Daphnia* were sampled from seven drinking water reservoirs in the Czech Republic. Two microsporidia species (*Berwaldia schaefernai* and microsporidium lineage MIC1) were sequenced at the internal transcribed spacer (ITS) region, using the 454 pyrosequencing platform. Geographical structure analyses were performed applying Fisher's exact tests, analyses of molecular variance, and permutational MANOVA. To evaluate the genetic diversity of the ITS region, the number of polymorphic sites and Tajima's and Watterson's estimators of theta were calculated. Tajima's *D* was also used to determine if the ITS in these taxa evolved neutrally. Finally, neighbour similarity score and pairwise homology index tests were performed to detect recombination events.

Results: While there was little variation among *Berwaldia* parasite strains infecting different host populations, the among-population genetic variation of MIC1 was significant. Likewise, ITS genetic diversity was lower in *Berwaldia* than in MIC1. Recombination signals were detected only in *Berwaldia*.

Conclusion: Genetic tests showed that parasite populations could have expanded recently after a bottleneck or that the ITS could be under negative selection in both microsporidia species. Recombination analyses might indicate cryptic sex in *Berwaldia* and pure asexuality in MIC1. The differences observed between the two microsporidian species present an exciting opportunity to study the genetic basis of microsporidia-*Daphnia* coevolution in natural populations, and to better understand reproduction in these parasites.

Keywords: Cryptic sex, Genetic diversity, Internal transcribed spacer, Microsporidia, Recombination

* Correspondence: gonzalez@igb-berlin.de; enriquegleztortuero@gmail.com
[1]Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany
[2]Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin, Germany
Full list of author information is available at the end of the article

Background

Microsporidia are a phylum of spore-forming obligate intracellular parasites, and constitute one of the most phylogenetically divergent basal fungal clades [1]. This taxon comprises over 1500 known species distributed across more than 187 genera [2], although this is likely an underestimate of their true diversity [3]. Microsporidia are able to infect a wide range of eukaryotes, from protists of the Stramenopiles-Alveolata-Rhizaria (SAR) supergroup to the majority of animal lineages (including humans) [2]. Currently, Microsporidia are considered to be emerging pathogens [4] and a relevant threat to human health, as they are commonly found in immunocompromised patients [5]. There are also economically important pathogens among Microsporidia such as *Nosema bombycis*, which parasitises the silkworm [6, 7], or *N. ceranae*, one of the causes of the honeybee population decline [4, 8]. *Nosema* species have also been considered for use in pest biocontrol in place of parasitoids [9].

Despite the importance of microsporidians, little is known about patterns of genetic diversity in these parasites. Recent approaches to understanding these patterns have included intraspecific genome analysis as a way to investigate between-host genetic variation and the evolutionary history of parasite populations. Comparative genomics has also been used to predict mode of reproduction based on genes associated with meiosis [10] and/or recombination [11–13]. Reproduction appears to vary across microsporidians; while the existence of sexual reproduction was suggested in *Nematocida* spp. based on genomic evidence [12], *Nosema ceranae* populations across the world seem to be clonal [11]. Moreover, genomic approaches present an opportunity to search for new markers when substantial genetic variability between strains is discovered (e.g. in *Encephalitozoon cuniculi* [13]). However, genomic data are still lacking for most microsporidian taxa. Thus, another common approach to the analysis of intraspecific genetic diversity in Microsporidia is sequencing of a target genomic region; markers used for this purpose included single-copy loci such as those encoding the 70 kDa heat-shock protein [14, 15], the large subunit of the RNA polymerase II [15] and the polar tube proteins [16, 17], or a multi-copy marker such as the internal transcribed spacer, ITS [18–20].

The internal transcribed spacer (ITS) is the non-coding stretch of DNA situated between the small (16S) and the large (18S) subunit ribosomal RNA genes in the majority of microsporidian species. Intraspecific ITS variability differs markedly between microsporidian species: while low ITS variability has been described for *Enc. cuniculi* [21], *Enc. hellem* [19] and *Enc. intestinalis* [22], high variability was detected in *Enterocytozoon bieneusi* [23]. Although the ITS region is not related to the infection mechanism, it has been used to determine parasite genetic variability because the ITS variation is assumed to be neutral. Variability in the parasite is important because it increases the probability that the parasite will be able to successfully evade its host's immune response (reviewed in [24]), leading to survival and potential transmission (reviewed in [25]). Understanding the nature of genetic variation of a parasite is thus crucial to the understanding of host-parasite interactions.

The planktonic cladocerans of the genus *Daphnia* and their microparasites were recently proposed as a suitable host-parasite model system to study coevolutionary questions (e.g. [26, 27]). In addition, microsporidian parasites of *Daphnia* have received considerable attention due to their complex life-cycles [28–30]. In our study, we focused on two abundant microsporidians infecting *Daphnia* communities inhabiting large lakes and reservoirs in central Europe, classified as *Berwaldia schaefernai* and as the microsporidium MIC1 [31]. Both of these species infect the body cavity of their host, where a massive amount of spores then proliferate [32]. They are closely related to *Marssoniella elegans* (a parasite of the copepod *Cyclops vicinus*), *Senoma globulifera* (a parasite of the malaria-hosting mosquito *Anopheles messeae*), and other parasites of *Daphnia*, including *Larssonia obtusa*, *Gurleya vavrai* and *Binucleata daphniae* [30]; these relatives span a range of transmission and reproduction modes. *Marssoniella elegans* is a dixenous parasite which likely uses mosquitoes or caddisflies as secondary hosts [33, 34], while *S. globulifera* and *B. daphniae* are monoxenous parasites [29, 35]. It is thought that *L. obtusa*, *G. vavrai* and *B. schaefernai* may have an indirect life-cycle which involves a secondary host (similar to *M. elegans*), given that attempts to maintain them in the laboratory have proven consistently unsuccessful [28, 32]; however, failure to replicate relevant environmental conditions cannot be excluded [36, 37]. Alhough relatively limited, the available data indicate low genetic variability among populations of *Berwaldia*, pointing to a highly mobile secondary host or vector which is able to effectively homogenise the parasite population [38]. For MIC1, no previous data are available. The goal of the present work is to compare the patterns of geographic population structure, intraspecific genetic variation and recombination events of the ITS sequence of *Berwaldia* and MIC1, in order to infer the dispersal mechanism of these parasites. Recombination analyses were also used to investigate the potential presence of sexual reproduction in the life-cycle of the studied taxa. Taken together, this information will be used to better characterise the life-cycles and dispersal patterns of these parasite species.

Methods

Sampling design

Zooplankton samples, including the *Daphnia longispina* species complex, were collected from seven reservoirs in the Czech Republic (Brno, Římov, Seč, Stanovice, Trnávka, Vír, and Žlutice) in the summer and autumn of 2004 and 2005 by hauling a plankton net (mesh size 170 μm) through the water column. Geographical locations and further characteristics of the reservoirs are provided in [39]. In the present study, we focused on analysing parasite DNA from infected *Daphnia* host individuals which had been previously assessed for microsporidia infection [30]. Eighty-seven *Daphnia* infected with *Berwaldia* (sampled across six reservoirs: Římov, Seč, Stanovice, Trnávka, Vír, and Žlutice), and 28 *Daphnia* infected with the microsporidium MIC1 (sampled from two reservoirs: Brno and Stanovice) were included in this study. The number of infected *Daphnia* sampled per reservoir varied from 9 to 25 (see Additional file 1: Table S1). In a previous parasite survey of *Daphnia* populations from the aforementioned reservoirs, *Berwaldia* and MIC1 were the most abundant of the eight microsporidian taxa detected [30].

Molecular analyses

Primers amplifying the ITS regions to be used for 454 amplicon pyrosequencing were constructed by fusing a specific core primer sequence (*Berwaldia* forward: 5'-TGA TGR CGA TGC TCG ATG AGA G-3'; MIC1 forward: 5'- TTT GAC TCA ACG CGG GAM AAC TT-3'; reverse used for both species: 5'- CAA YTT CRC TCG CCG CTA CTA-3' [31]) with a basal 25-mer for binding to the DNA Capture Beads (Lib-A) and a 10-base multiplex identifier (MID) chosen from the 454 Standard MID Set (Roche, Basel, CH). After DNA isolation [30], the ITS region of 115 microsporidia-infected *Daphnia* was PCR-amplified using the following protocol: 1X Phusion HF Buffer, 0.5 U Phusion HF DNA polymerase (Thermo Fisher Scientific, Waltham, MA), 0.5 μM each of Forward and Reverse primer, 0.25 mM deoxynucleoside triphosphates, and 2 μl genomic DNA, for a total reaction volume of 25 μl. The success of amplification was evaluated by agarose gel electrophoresis. In cases when the initial PCR failed, DNA concentration in the reaction was varied (1 and 3 μl). To evaluate differences among (presumably) identical samples subjected to the same 454 sequencing run, three technical replicates were sequenced for each of three individual *Daphnia* (two individuals infected with *Berwaldia* and one individual infected with MIC1). These replicates were created by carrying out independent PCR reactions of the same DNA template using primers labelled with different multiplex identifiers.

To verify primer specificity, one *Berwaldia* and two MIC1 PCR products were randomly selected and cloned using a StrataClone PCR Cloning Kit (Agilent Technologies, La Jolla, CA), according to the manufacturer's protocol. Between two and four positive bacterial colonies were then selected for sequencing on an ABI 3730 DNA Analyzer using the BigDye 1.1 Terminator Sequencing Kit (both Applied Biosystems, Foster City, CA). The resulting electropherograms were visually inspected and manually corrected in MEGA6 [40]. NCBI's Nucleotide BLAST [41] was then used to verify PCR product sequence identity.

To create a 454 library, all PCR products were purified independently for each sample with a Gel/PCR DNA Fragments Extraction Kit (Geneaid Biotech Ltd., Taipei, TW), their concentrations were measured on a Qubit 2.0 Fluorometer, and then the samples were pooled in equimolar concentrations (*Berwaldia* and MIC1 separately). Fragments of the appropriate size (*Berwaldia*, 650 bp; MIC1, 800 bp) were subsequently separated from the pooled samples using the E-Gel platform with SizeSelect 2 % agarose gel kit, and the selected fragments were then purified with SPRI AMPure XP paramagnetic beads (Beckman Coulter Genomics, Danvers, MA) using a slightly modified standard protocol (isopropyl alcohol added to the sample in a 1:3 ratio in the binding step with beads). The pooled *Berwaldia* and MIC1 amplicon solutions were mixed at a 4-fold excess of the longer MIC1 amplicons, to compensate for the higher affinity of the shorter *Berwaldia* fragments to the sequencing beads during emulsion PCR. Emulsion PCR and pyrosequencing were performed with the amplicon (Lib-A) kit, using GS Junior Plus reagents and the manufacturer's protocols (454 Life Sciences, Branford, CT), with the ratio of DNA molecule-to-bead decreased to 0.35. DNA bead enrichment was slightly above the expected level (25 %). The raw 454 dataset is available in the Sequence Reads Archive (SRA) under accession number (GenBank: SRP056909).

Preparation of the dataset

The bioinformatic analyses (and subsequent statistical tests) were run separately for each of the two parasite species. The raw datasets were analysed using the Quantification of Representative Sequences (QRS) pipeline [42]. The pipeline was run with default parameters, unless indicated otherwise. The following sequences were discarded: those that contained more than two uncalled bases, those with GC content outside of the 43–49 % range, or those with length outside of the 570–660 bp (for *Berwaldia*) or 700–800 bp (for MIC1) range. Homopolymer error correction was performed using HECTOR [43]. After de-noising to minimise the presence of sequencing errors, only sequences present in at least three copies (or four copies, in case of genetic diversity and recombination analyses, see below) were retained.

Alignments were carried out using the MUSCLE algorithm [44] and manually corrected.

Defining representative sequences

In the analyses of geographical structure, phylogeny, and haplotype networks, "representative sequences" were used instead of raw data. A representative sequence is the most abundant sequence per group which is considered the correct or ancestral allelic reference [42]. Representative sequences are useful for the analysis of sequence variation when a multicopy marker like the ITS is used [45]. To obtain these representative sequences, the raw sequences were clustered using Statistical Parsimony [46] at 99.5 % of divergence (i.e. three connection steps) with gaps designated as a fifth state. These settings were consistent with the approach used to analyse ITS variation in *Berwaldia* in a previous study [38]. Additionally, statistical parsimony was used in these analyses because it is more robust than distance methods; the number of singletons is reduced using statistical parsimony compared to neighbour-joining because distance methods assume that there are no reticulate relationships between sequences and no recombination events [45]. The representative ITS sequences are available in the GenBank sequence database under accession numbers KR816811–KR816826.

Abundant *Berwaldia* representative sequences ("abundant" refers to representative sequences with overall frequencies higher than 0.5 %) were compared with representative sequences from [38] to assess whether any matched. In that previous study, *Berwaldia*-infected *Daphnia* were sampled from three Czech reservoirs (including two reservoirs studied here: Římov and Vír) and ITS sequence variation was assessed using Sanger sequencing of cloned PCR products [38]. All representative sequences (i.e. from the previous and present study) were re-aligned using the MUSCLE algorithm [44] and manually corrected. Subsequently, a haplotype network was created as described below. In the case of MIC1, such a comparison was not possible as neither intra- nor inter-population variation was evaluated in previous studies involving this parasite [30, 31].

Geographical structure

To investigate the geographical structure of genetic variation in *Berwaldia* and MIC1, four types of analyses were performed, all based on the ITS representative sequences identified in each species' dataset. First, the frequencies of abundant representative sequences were compared among populations using Fisher's exact test (representative sequences that did not reach a threshold abundance of 0.5 % in any population were pooled into the "rare" category). Secondly, an analysis of molecular variance (AMOVA) was run at different hierarchical levels: within individuals (i.e. within a *Daphnia* host), within populations (among *Daphnia* hosts within each reservoir), and among populations. To test whether the use of representative sequences produced results consistent with the original data, an AMOVA was also performed using the raw dataset in addition to the AMOVA with the representative sequences. Thirdly, a non-metric multidimensional scaling (nMDS) plot using Bray-Curtis dissimilarity was constructed in order to visualise genetic variation among microsporidian populations. The maximum number of iterations was set to 1000. The "stress" value, which represents the rank dissimilarities between the distance matrix and the plotted distances, was used to evaluate the reliability of the nMDS plots. Fourthly, a permutational multivariate analysis of variance (PERMANOVA) [47] was then performed to assess differences among microsporidian populations inhabiting the various reservoirs. All statistical tests were carried out at an alpha level of 0.05 and were performed in R 3.2.2 [48] using the *ade4* [49], *vegan* [50] and *MASS* [51] packages.

Phylogenetic analyses

To check if the phylogenetic position of *Berwaldia* and MIC1 obtained with ITS is congruent with conclusions derived from the SSU marker [30], neighbour-joining and maximum likelihood trees for the abundant ITS representative sequences of *Berwaldia* and MIC1 were constructed. The selection of other microsporidian taxa was based on similarity searches using BLAST [41]. A sequence of the basidiomycete *Agaricus bisporus* was used as an outgroup. All sequences were aligned using Opal v. 2.1.3 [52]. Poorly aligned, non-conserved, and highly-divergent regions were discarded using Gblocks 0.91b [53] set to less stringent settings, resulting in a 677 bp long alignment. Sequence similarity among taxa was inferred by the neighbour-joining method in rapidNJ [54] under the Kimura-2-Parameter model; branch support in the resulting tree was estimated using bootstrapping with 1000 pseudoreplicates. A maximum likelihood tree was built using RaxML v. 8.2.3 [55] under the GTR + G model selected by the corrected Akaike Information Criterion in jModeltest v. 2.1.7 [56]. Branch support in the resulting tree was estimated by the rapid bootstrapping algorithm [57] using 600 pseudoreplicates according to the *a posteriori* bootstrapping convergence test [58] based on the extended majority rule consensus tree.

Haplotype network

To study the relationships between the abundant ITS representative sequences, haplotype networks were created for *Berwaldia* and for MIC1. Connection distances between haplotypes were calculated using TCS [59] according to the statistical parsimony algorithm. Both

outputs were processed using a force-directed algorithm, implemented in Cytoscape 3.2.1 [60].

Genetic diversity

To evaluate the genetic diversity of the ITS region, raw sequences were used (in contrast to all aforementioned analyses, which were run using representative sequences). Only sequences present in at least four copies were retained. Calculations were carried out for pooled datasets (i.e. pooled across all six or two populations, for *Berwaldia* and MIC1, respectively). Three parameters, the number of polymorphic sites, Tajima's estimator of theta (π; [61]) and Watterson's estimator of theta (θ_w; [62]), were obtained using the package *PopGenome* [63]. While π is defined as the average number of nucleotide differences between two sequences [61], θ_w quantifies the level of variability as the total number of polymorphic sites [62]. Both estimators were divided by the alignment length to obtain the relative values per nucleotide. Changes in nucleotide diversity based on these summary statistics were also calculated within a sliding window of 50 bp with an increment of 25 bp. To determine if *Berwaldia* and MIC1 ITS sequences evolved neutrally (i.e. in mutation-drift equilibrium), Tajima's *D* [64] test was performed.

In order to evaluate whether the differences in sample size (i.e. six populations of *Berwaldia* but only two of MIC1) affected the results, *Berwaldia* sequences were re-sampled using the *sub.sample* function in mothur v. 1.36.1 [65] to obtain a dataset representing two populations and containing the same number of sequences as in the MIC1 dataset (one population with 2290 sequences and another one with 340 sequences). Then, genetic diversity was estimated across the two pooled, re-sampled populations. Re-sampling was repeated ten times. Tajima's *D* values obtained from the re-sampled sets were compared with the value obtained across all six *Berwaldia* populations using a two-sample Kolmogorov-Smirnov test.

Another potential source of error in population genetic tests is related to the fact that several structured subpopulations could produce more negative values of Tajima's *D* than the whole dataset, due to the "pooling effect" [66]. To rule out this possibility, a new dataset was created by randomly choosing a single microsporidian ITS sequence per *Daphnia* host and then calculating Tajima's *D*, as described above. This analysis was repeated ten times per microsporidian taxon, and then compared with the Tajima's *D* value obtained for the entire dataset (i.e. including multiple sequences per *Daphnia* host) using a two-sample Kolmogorov-Smirnov test.

Recombination

Raw sequences were used for the recombination analysis. However, as dereplication is a prerequisite for recombination tests, only one copy of each sequence was retained in the dataset (i.e. 606 out of 18,871 sequences were retained for *Berwaldia* and 138 out of 2630 sequences for MIC1). To detect recombination events, neighbour similarity score (NSS; [67]) and the pairwise homology index (PHI; [68]) with 1000 permutations were calculated in PhiPack [68]. Both NSS and PHI are based on compatibility of parsimoniously informative sites, i.e. sites that contain at least two types of nucleotides that occur twice [69]. While the PHI is defined as the minimum number of convergent or recurrent mutations (homoplasies) necessarily present on any tree describing the history of two sites [68], NSS is calculated as the fraction of adjacent parsimonious informative sites (independently of their compatibility) in an alignment [67]. A sliding window of 50 bp was used in these tests. When the results were significant, DnaSP 5.10.1 [70] was subsequently used to identify the minimum number of recombination events (Rm) according to the four-gamete test [71]. The RDP, GeneConv, Chimaera, MaxChi, BootScan and 3Seq algorithms were used to identify parental and recombinant sequences using the RDP4 Beta 4.46 interface [72]. A sliding window of 50 bp with an increment of 25 bp was used in the latter five tests. Gaps were not excluded in the recombination tests. To trace the origin of parental and recombinant sequences, raw sequences were classified into representative sequences according to Statistical Parsimony (see "Defining representative sequences") using TCS [59]. Then, the parental and recombinant sequences were tracked into the different representative sequences according to the log file.

Results

Description of the final dataset

After processing with the QRS pipeline, 19,681 ITS sequences were retrieved for *Berwaldia* (18,871 excluding tripletons) and 2906 (2630 excluding tripletons) for MIC1 (out of 49,947 and 32,369 sequences available for the respective taxa). These sequences originated from 80 *Berwaldia*- and 23 MIC1-infected *Daphnia* individuals; the remaining seven and five host individuals were discarded as they contributed less than ten parasite sequences each. The majority of the originally-generated sequences were discarded due to anomalous length, a product of the presence of two different forward primers at once (especially in MIC1). The length of the aligned sequences after the removal of primers was 546 bp for *Berwaldia* and 706 bp for MIC1 (Additional file 2: File S1 and Additional file 3: File S2, respectively). In the case of the *Berwaldia* sequences, the first 192 bp belonged to the 16S region of the rRNA and the final 194 bp belonged to the 18S region of the rRNA, which indicates that the ITS region is approximately 160 bp long (based on [28, 31]).

However, such an exact prediction is impossible to make for MIC1 as no published information about its rRNA gene structure is currently available.

Defining representative sequences

Twenty-six representative sequences were detected in the *Berwaldia* dataset and 32 in the MIC1 dataset. The most abundant representative sequence reached a frequency of 97.02 % in *Berwaldia* and 62.66 % in MIC1. The majority of representative sequences were classified as rare (i.e. were present at proportions lower than 0.5 %): 23 out of a total of 26 in *Berwaldia* and 19 out of 32 representative sequences in MIC1. The most abundant representative sequence in *Berwaldia*, BERW-1, matched exactly with the B3-type that was the most abundant in a previous study [38], confirming the dominance of the same ITS representative sequence in a larger set of lakes. The BERW-2 and BERW-3 sequences found here did not match any previously identified representative sequences. In a joint haplotype network of representative sequences from this and from a previous study [38], the BERW-1 (or B3-type) representative sequence was located in the centre of the network (Additional file 4: Figure S1), suggesting that it could be the ancestral type.

The 454 sequencing reaction repeatability assay produced mixed results. Neither of the two *Berwaldia*-infected individuals tested showed significant differences in the frequencies of representative sequences among three sequenced replicates (as assessed by Fisher's exact test; Additional file 5: Figure S2). On the other hand, one of the replicates of MIC1-infected *Daphnia* differed significantly in the proportion of representative ITS sequences from the other two (Additional file 5: Figure S2). This outlying replicate showed intermediate amplification success. In further analyses, only those samples with the greatest number of sequences (out of the three replicates) were considered.

Geographical structure

For both parasite species, the distribution of representative sequences differed significantly among populations, as assessed by Fisher's exact test (Fig. 1). In *Berwaldia*, the three most abundant representative sequences were present in all six surveyed populations. The relative abundance of the variant BERW-2 differed among populations; this variant had the highest frequency in Žlutice (3.57 %) and lowest in Seč (0.08 %) (Fig. 1). However, AMOVA tests did not detect any significant among-population genetic variation in *Berwaldia* (Table 1), indicating overall low effect of such variation in proportion of genetic variants across all screened populations. In MIC1, for which only two populations were tested, the differences in the distribution of representative sequences were much more pronounced. Although the frequency of the most abundant representative sequence (MIC1-1) was approximately the same in both populations, only five other representative sequences - out of twelve - were detected in both reservoirs (Fig. 1). These strong differences in the distribution of MIC1 representative sequences were confirmed by AMOVA. Specifically, genetic variation between the two MIC1 populations was significant and explained 3.02 % of total variance (Table 1). For both parasites, the largest amount of variation was observed at the within-individual level (i.e. within a *Daphnia* host): 93.7 % for *Berwaldia* and 93.6 % for MIC1. The results of AMOVA tests were similar between the raw and representative sequence datasets, indicating that the use of representative sequences did not bias the overall pattern (Additional file 1: Table S2).

The population structure of both microsporidia species was visualised using nMDS plots, where the position of individual points was based on the frequencies of specific ITS representative sequences per *Daphnia* host (Fig. 2). In *Berwaldia*, centroids of Žlutice and Římov populations overlapped and clustered away from the centroids of the four other populations. In MIC1, the centroids of the Brno and Stanovice populations were distinct, and there was no overlap among individual *Daphnia* hosts sampled from these two populations. The nMDS stress values in *Berwaldia* and in MIC1 were 0.022 and 0.055 respectively, indicating that both plots constituted a good representation of the original patterns of variation. The results of the nMDS plots were supported by PERMANOVA tests, which revealed significant differences in the presence and abundance of representative ITS sequences among populations for both microsporidian species (PERMANOVA (*Berwaldia*): $F = 0.166$, $df = 5$, $P = 0.006$; PERMANOVA (MIC1): $F = 0.378$, $df = 1$, $P = 0.001$).

Phylogenetic analyses

The ITS-based trees obtained for *Berwaldia*, MIC1, and reference parasite species using both methods produced identical topologies; thus, only the maximum likelihood tree is presented (Fig. 3). Microsporidian ITS sequences grouped in a single clade. *Berwaldia* ITS representative sequences clustered with several microsporidia known to infect *Daphnia* (including *L. obtusa* and *B. daphniae*), as well as the mosquito parasite *S. globulifera*. Similarly, MIC1 ITS representative sequences were grouped with *Gurleya daphniae* and *G. vavrai*.

Haplotype network

In the haplotype networks constructed for *Berwaldia* and for MIC1, the most abundant ITS representative sequences (BERW-1 or MIC1-1) were in the central position (Fig. 4). The different representative sequences did not cluster by populations of origin.

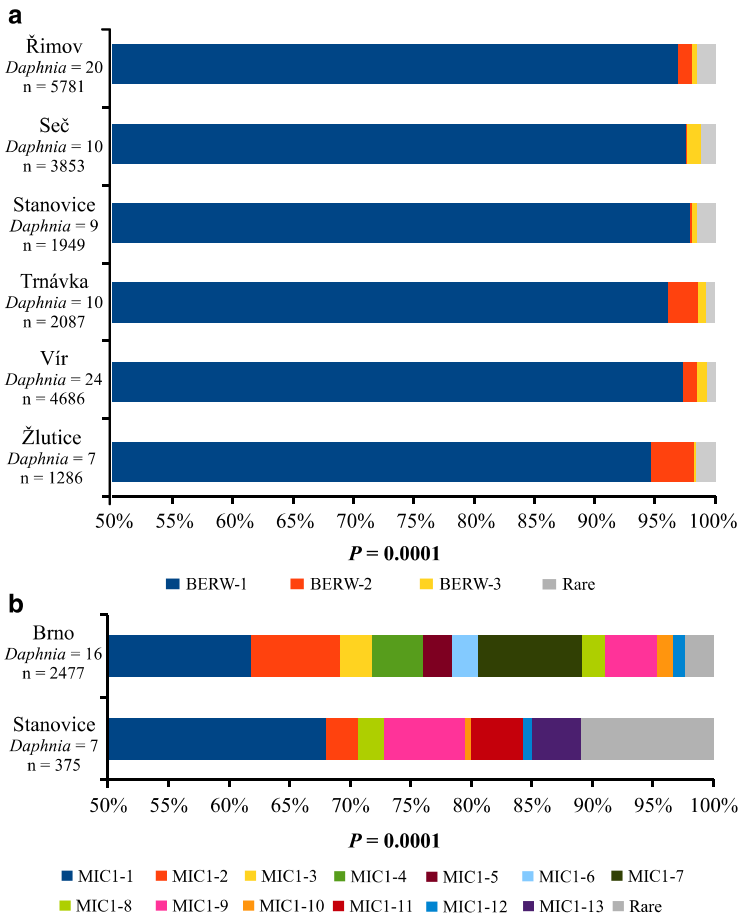

Fig. 1 Comparison of frequencies of the ITS representative sequences of (**a**) *Berwaldia* and (**b**) MIC1, among the different reservoirs. Number of screened *Daphnia* individuals and number of analysed ITS sequences are shown below each reservoir label. Results of Fisher's exact test are shown below each of the stacked bar charts. The "rare" category includes all ITS representative sequences present at a frequency lower than 0.5 %

Table 1 Results of the hierarchical analysis of molecular variance (AMOVA) of spatial population structure in *Berwaldia* and MIC1. Calculations were based on the frequency of representative ITS sequence variants, as detected within individual *Daphnia* host

Microsporidia	Source of variation	*df*	Variation explained (%)	*P*
Berwaldia	Among host populations	5	-0.11	0.145
	Within host population	74	6.42	< 0.001
	Within host individual	19,562	93.69	< 0.001
MIC1	Among host populations	1	3.02	< 0.001
	Within host population	24	3.37	< 0.001
	Within host individual	2826	93.61	< 0.001

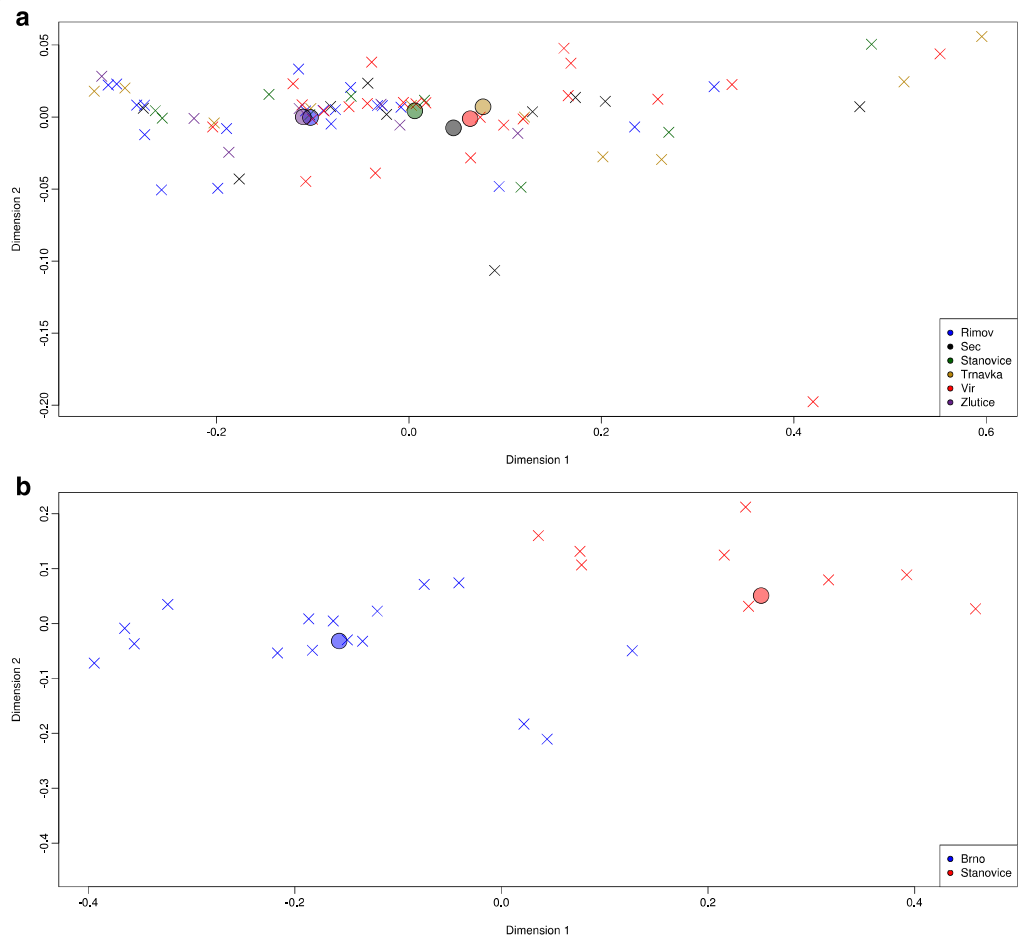

Fig. 2 Non-metric MDS plot of the population structure of (**a**) *Berwaldia* and (**b**) MIC1. The analysis is based on the frequency of ITS representative sequences of the parasites, calculated per individual *Daphnia* host. Crosses represent individual *Daphnia* hosts and circles indicate the centroids of each population

Genetic diversity

Summary statistics from Tajima's *D* neutrality test are reported in Table 2. *Berwaldia*'s ITS π value was lower than that of MIC1, while the reverse was true for θ_w, indicating lower genetic diversity in *Berwaldia* and a larger proportion of rare alleles in MIC1. In addition, sequence diversity per site showed different patterns for *Berwaldia* and MIC1 according to the sliding window analyses. In *Berwaldia*, π was always lower than θ_w, remaining near zero across all sliding windows (Fig. 5a). By contrast, for MIC1, π and θ_w followed similar patterns of variation across all sliding windows (Fig. 5b). Taken together, these statistics indicate that the ITS is more variable in MIC1 than in *Berwaldia*.

Differences in sample size between the two taxa did not have a substantial effect on the results of analyses, as demonstrated by comparison of the entire dataset with the smaller, re-sampled datasets. Tajima's *D* values from the ten re-sampled *Berwaldia* datasets (-2.444 ± 0.030) did not differ significantly from the value calculated using the entire dataset, -2.426 (Kolmogorov-Smirnov test: $D = 0.5$, $P = 0.977$; Additional file 1: Table S3). Likewise, Tajima's *D* values from the ten re-sampled datasets (per microsporidium taxon) which contained a single random sequence per *Daphnia* individual (*Berwaldia*: -2.397 ± 0.088; MIC1: -1.633 ± 0.252) did

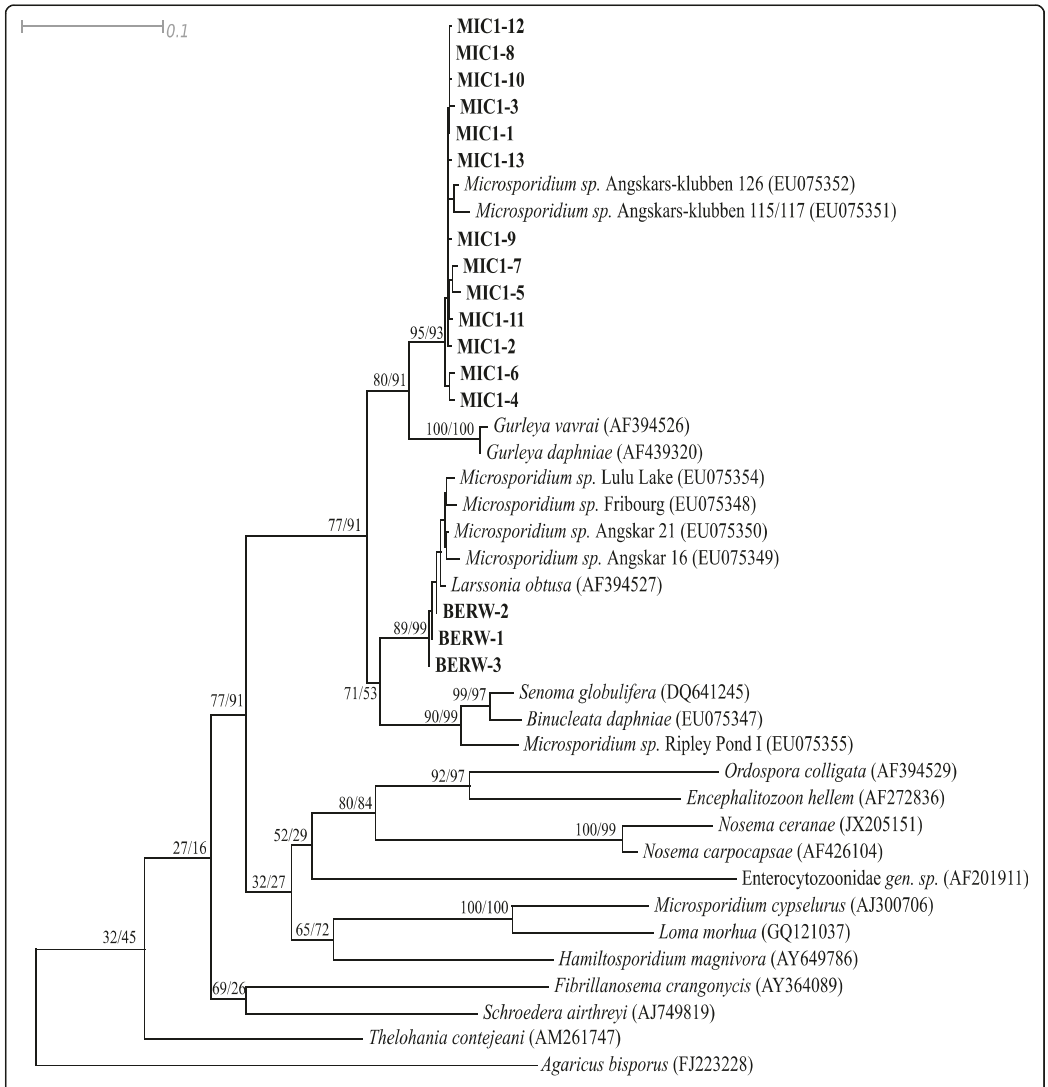

Fig. 3 Maximum likelihood tree of selected microsporidian taxa, including all abundant representative types of *Berwaldia* and MIC1 as detected in our study, based on ITS rDNA gene sequences. Maximum likelihood and neighbour-joining trees produced identical tree topologies. Support for each internal node is given as rapid bootstrap values and neighbour-joining bootstrap values, respectively. Branch lengths are based on the expected number of nucleotide substitutions per site. Bold labels indicate the most abundant ITS representative sequences of *Berwaldia* and MIC1 detected in the present study

not differ significantly from those obtained for the entire dataset: *Berwaldia*: -2.426; MIC1: -1.811 (Kolmogorov-Smirnov test (*Berwaldia*): $D = 0.6$, $P = 0.909$; Kolmogorov-Smirnov test (MIC1): $D = 0.8$, $P = 0.546$; Additional file 1: Table S4).

Recombination

The two recombination tests yielded contrasting results. A PHI test detected recombination signals in *Berwaldia* but not in MIC1 (PHI test (*Berwaldia*): $P = 0.002$; PHI test (MIC1): $P = 0.654$) whereas NSS tests found

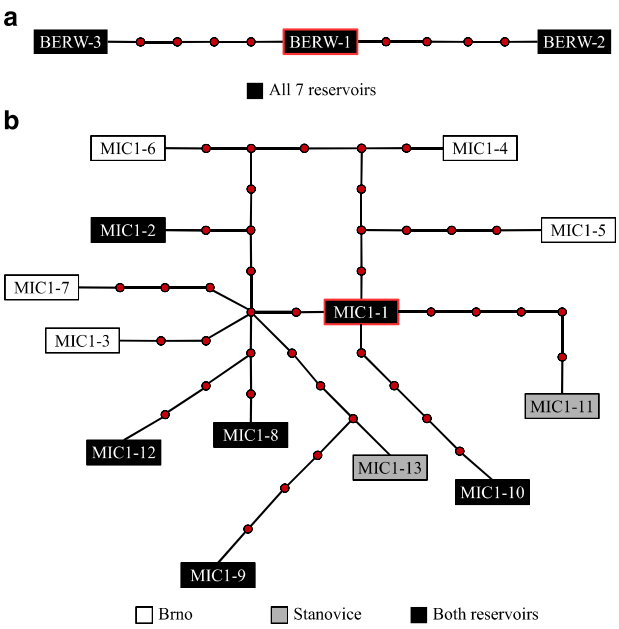

Fig. 4 Haplotype networks of the abundant ITS representative sequences of (**a**) *Berwaldia* and (**b**) MIC1. Each red circle indicates a single connection step (i.e. a single mutation) between the ITS representative sequences. Red-lined boxes indicate the most abundant ITS representative sequences for each parasite. Box colour indicates the reservoir(s) in which a given ITS representative sequences was detected

recombination in neither *Berwaldia* nor in MIC1 (NSS test (*Berwaldia*): $P = 0.133$; NSS test (MIC1): $P = 0.114$). Thus, further recombination tests were only carried out for *Berwaldia*. The minimum number of recombination events (Rm) was estimated to be ten in this species. Recombination breakpoints were detected in the middle of the alignment (Additional file 1: Table S5). Recombinant and parental sequences were detected mostly in *Berwaldia* sampled from different lakes (Additional file 1: Table S5). When the raw ITS sequences were classified into ITS representative sequences, parental and recombinant sequences belonged to BERW-1, BERW-3, and several rare ITS representative sequences (Additional file 1: Table S5).

Discussion

The ITS phylogenetic trees constructed here are consistent with those retrieved in previous studies that used the small subunit ribosomal DNA [30, 31, 33], even though the microsporidian ITS can sometimes generate incongruous phylogenies, as seen in the evolutionary history of the clade *Vairimorpha*/*Nosema* [73]. The positions of *Berwaldia* and MIC1 on the phylogenetic tree reinforce the prediction that both microsporidian species could be dixenous parasites, as previously suggested [30, 31]. Moreover, the negligible geographical variation observed in *Berwaldia* and the fact that recombinant and parental sequences originated from different lakes further support the hypothesis that this species can disperse with relative ease, and likely uses a mobile secondary host during its

Table 2 Summary statistics of the *Berwaldia* and MIC1 ITS marker

Microsporidia	Number of sequences	Number of segregating sites	π	θ_w	D
Berwaldia	18,871	171	1.470×10^{-3}	28.784×10^{-3}	**-2.426**
MIC1	2630	75	3.764×10^{-3}	12.291×10^{-3}	-1.811

Samples from different populations were pooled. Genetic diversity is calculated as the average heterozygosity per site (Tajima's estimator, π) and the average number of nucleotide differences per site (Watterson's estimator, θ_w). Bold number indicates a significant value in Tajima's D neutrality test

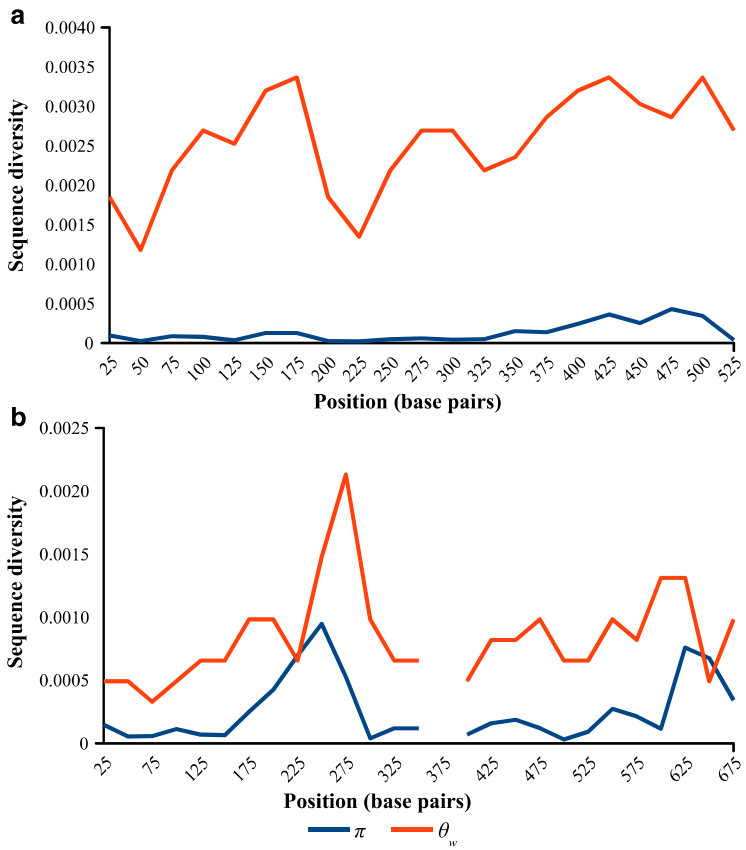

Fig. 5 Sliding window analysis of nucleotide diversity in the ITS marker of (**a**) *Berwaldia* and (**b**) MIC1. A sliding window of 50 base pairs was used, with an increment of 25 base pairs. Observed nucleotide diversity was calculated as the average heterozygosity per site (Tajima's estimator, π), while the expected nucleotide diversity was computed as the average number of nucleotide differences per site (Watterson's estimator, θ_w). The gap in positions 350–400 in the MIC1 graph is the result of complete sequence identity, which renders the estimator inapplicable

life-cycle [38]. Parasitic species with life-cycles where long-distance dispersal is needed generally show less population structure and diversity than those with a little ability to spread [74, 75]. In this way, the presence of a mobile secondary host in *Berwaldia* life-cycle could explain the obtained results in this study. Additionally, various measures of population genetic diversity strengthen the original observation (from a smaller set of localities) that the *Berwaldia* ITS marker has relatively low diversity [38]. These results are very similar to those obtained for *N. ceranae* across globally sampled populations. In that case, ITS microsporidian sequences from different host species (*Apis mellifera* and *A. cerana*) were identical, which indicates the ability of *N. ceranae* to infect both hosts [11].

ITS sequences are used as a universal fungal barcode due to the ease of PCR amplification and a higher probability of successful identification compared with protein-coding genes [76]. Consequently, variability of the ITS has been most frequently assessed to obtain information on genetic variation within microsporidian species [18, 19] as well as within one isolate [21]. Indeed, the ITS is the only known polymorphic marker in several microsporidians such as *Ent. bieneusi* [23] and *Nosema*/*Vairimorpha* spp. [77], although a low ITS variability was described in other microsporidians, e.g. *Enc. hellem*, *Enc. cuniculi* and *Enc. intestinalis* [19, 21, 22]. Nevertheless, low ITS variability in *Enc. cuniculi* [21] is not congruent with high intraspecific variability revealed by whole genome analysis [13]. This discrepancy is likely due to the short length of this species' ITS (28–45 bp, [21]), while in

other microsporidians the ITS sequence is considerably longer (e.g. *Ent. bieneusi*, 243 bp) [23]. In our study, we determined that *Berwaldia* ITS region was approximately 160 bp long (based on [28, 31]) while in MIC1 it was presumably longer, although Northern Blot analyses would be needed [78] for the experimental verification of the position of the ITS sequence in both microsporidia. Moreover, although the rRNA gene structure is usually well preserved, it is highly variable in the *Nosema/Vairimorpha* group within the Microsporidia. In fact, the rearrangement of the 18S and 16S subunits and the presence of a 5S subunit at the end of the ribosomal RNA results in an ITS1 located between the 18S and the 16S subunits and an ITS2 between 16S and the 5S subunits, as described in *N. bombycis*, *N. antheraceae*, *N. plutellae* and *N. spodopterae* [79–81]. In addition, because the ribosomal RNA repeat unit is present in multiple copies throughout the genome, each copy has the potential for mutation, resulting in further intragenomic variation. In fact, the existence of transcriptionally active but fragmented copies of rRNA genes that coexist with the intact rRNA copies within the same genome was described in several isolates of *N. bombycis* [82]. The structural variation of rRNA genes is a potential source of complication in rRNA phylogenies [83] and leads to a high variability in both ITS1 and ITS2 sequences in *Nosema/Vairiomorpha* [77]. The high variability of the ITS region could also be evidence of recombination (or so called "cryptic sex") in Microsporidia [77, 84]. Moreover, transposition events in ribosomal markers (including the ITS) are another source of high genetic variability in Microsporidia [80, 82, 85]. Finally, with multicopy gene markers one is not able to exclude the existence of co-infections, when DNA is isolated from an individual host and high variability in the multicopy gene region is observed, as was the case in our dataset. Thus, future analyses of population genetic diversity at the strain level will need to include additional nuclear genes [14–17], micro- or minisatellites [86, 87] or even whole genome analysis [12, 13] in order to make stronger predictions regarding the evolution of microsporidian populations and to evaluate their genetic diversity.

MIC1 is closely related to the genus *Gurleya* ([30]; this study), which is predicted to have a complex life-cycle [28]. However, MIC1 exhibits greater ITS genetic diversity than *Berwaldia*, and the differences between populations of the two microsporidians could indicate that MIC1 life-cycle does not facilitate dispersal among *Daphnia* host populations as efficiently as is seen in *Berwaldia*. However, even with dispersal, small effective parasite population size, highly aggregated distribution among hosts, high host specificity, and patchy spatial and temporal parasite niche distribution could potentially contribute to increased genetic variability of a parasite [74, 75].

According to the results of Tajima's *D* test, the ITS regions of both *Berwaldia* and MIC1 could have evolved following recent population expansions after a bottleneck, under negative selection and/or multiple mergers [88]. However, as the results from Tajima's *D* test could be affected by recombination signals [89], recombination tests were conducted. These tests could indicate the presence of cryptic sex in *Berwaldia* and pure asexuality in MIC1. Although gene recombination tests have been used to demonstrate cryptic sex in microsporidia in past studies [77, 90], such analyses may not be sufficient to unambiguously confirm the presence of sexual cycles in this group, because multiple and heterogeneous copies of rDNA could recombine non-homologously [91]. Thus, other evidence is needed, such as confirmation of the existence of polyploid stages. While no diplokaryotic cells have been observed in *Berwaldia*, the existence of bi- or tetranucleated cells has been described [32]. If there is any sexual cycle present, the different sets of chromosomes in polyploid species must sort during meiosis to produce balanced gametes. Recently, genes relevant to meiosis were detected in two microsporidian parasites of mosquitoes (*Edhazardia aedis* and *Vavraia culicis*; [10]). These genes were only expressed in *E. aedis*, confirming the existence of sexuality in this species [92] and providing an explanation for the lack of sexual cycles in *V. culicis* [93]. To investigate whether *Berwaldia* and MIC1 have sexual cycles, cytogenetic, flow cytometry, genomic and transcriptomic studies should be considered, even though such studies may be challenging with intracellular parasites.

Conclusions

The biology and life history of the presumed secondary hosts of *Berwaldia* and MIC1 likely differ. While the presumed secondary host for *Berwaldia* is expected to be mobile, in MIC1 the secondary host (if it exists) does not appear to facilitate dispersal to the same degree. Further, the recombination tests might suggest that there is cryptic sex in *Berwaldia* and pure asexuality in MIC1. These predictions should be confirmed in future experiments using modern laboratory techniques (cytogenetics, flow cytometry, genomics and/or transcriptomics). This would allow a more comprehensive understanding of the biology of *Daphnia*-infecting microsporidians and of the genetic basis of microsporidia-*Daphnia* coevolution in natural populations.

Additional files

Additional file 1: Table S1. Number of *Daphnia* host individuals sampled from each reservoir. **Table S2.** Comparison between the results of the hierarchical analysis of molecular variance (AMOVA) of spatial population structure for *Berwaldia* and MIC1, using the dataset before and after clustering with Statistical Parsimony (using QRS pipeline; [42]). **Table S3.** Comparison of the summary statistics of the ten re-sampled, smaller *Berwaldia* ITS marker datasets simulating only two populations

("R-1" to "R-10": 2630 sequences each) and summary statistics calculated for the entire *Berwaldia* ITS dataset, including all six populations ("Entire": 18,871 sequences). **Table S4.** Comparison between the summary statistics of the ten smaller *Berwaldia* and MIC1 ITS datasets ("B-1" to "B-10" and "M-1 to M-10; 80 and 26 sequences per dataset, respectively) and the summary statistics calculated for the entire *Berwaldia* and MIC1 ITS datasets ("B-Entire" and "M-Entire"; 18,871 and 2630 sequences, respectively). **Table S5.** Recombination events detected in the *Berwaldia* ITS alignment using RDP4. (XLS 31 kb)

Additional file 2: File S1. DNA alignment of the 26 *Berwaldia* ITS representative sequences in FASTA format. (FASTA 14 kb)

Additional file 3: File S2. DNA alignment of the 32 MIC1 ITS representative sequences in FASTA format. (FASTA 23 kb)

Additional file 4: Figure S1. Haplotype network of the most abundant ITS representative sequences of *Berwaldia*. Each red circle represents a single connection step (i.e. a single mutation) between the ITS representative sequences. Red-outlined boxes indicate the abundant ITS representative sequences. White boxes indicate *Berwaldia* ITS representative sequences from [30]. Grey boxes indicate *Berwaldia* ITS representative sequences from this study. The black box represents the *Berwaldia* ITS representative sequences that was present in both studies. (PDF 24 kb)

Additional file 5: Figure S2. Comparison of frequencies of (A) *Berwaldia* and (B) MIC1 ITS representative sequences in the replicated samples. "ID" refers to identity of the *Daphnia* individual that was processed. The total number of ITS representative sequences (per replicate) is shown in each "replicate" label (as "n"). Results of Fisher's exact test are shown below each stacked bar chart. *P*-values that remained significant after sequential Bonferroni correction are shown in bold. The "rare" category includes all ITS representative sequences that were present at a frequency lower than 0.5 % (calculated per parasite taxon). (PDF 61 kb)

Acknowledgements
We thank Sabine Radetzki for assistance with the molecular work. Samples could be obtained thanks to a project coordinated by Jaromír Seďa. EGT wants to thank Tamara Gómez-Moracho for discussions and suggestions for improvements of the manuscript. We also thank the Editor (Kevin Tyler) and two anonymous reviewers for their valuable comments which helped us to improve the manuscript. This work was supported by DFG-SPP 1399 (WO 1587/2-2) grant to JW.

Authors' contributions
JW applied for funding and coordinated this project. JW and JR designed the study. AP contributed to sampling. JR designed sequencing primers, and JR and IM established and carried out the PCR protocol. LP and AP designed, and LP performed the pyrosequencing reactions. EGT performed all bioinformatics and statistical analyses. SL contributed to the interpretation of the genetic tests' results. EGT and JW wrote the manuscript, with comments and editing by AP and IM. All authors approved the final version of the manuscript and Supporting Information.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Ecosystem Research, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Müggelseedamm 301, 12587 Berlin, Germany. [2]Berlin Centre for Genomics in Biodiversity Research (BeGenDiv), Königin-Luise-Straße 6-8, 14195 Berlin, Germany. [3]Department of Biology II, Ludwig Maximilians University, Großhaderner Straße 2, 82512 Planegg-Martinsried, Germany. [4]Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague, Czech Republic. [5]Department of Zoology, Faculty of Science, University of South Bohemia, Branišovská 31, 370 05 České Budějovice, Czech Republic. [6]School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [7]Department of Biology, Chemistry and Pharmacy, Institute of Biology, Free University of Berlin, Königin-Luise-Straße 1-3, 14195 Berlin, Germany. [8]Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland.

Received: 29 March 2016 Accepted: 10 May 2016
Published online: 20 May 2016

References

1. Capella-Gutiérrez S, Marcet-Houben M, Gabaldón T. Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biol. 2012;10:47.
2. Vávra J, Lukeš J. Microsporidia and "The Art of Living Together.". Adv Parasitol. 2013;82:253–320.
3. Stentiford GD, Feist SW, Stone DM, Bateman KS, Dunn AM. Microsporidia: diverse, dynamic, and emergent pathogens in aquatic systems. Trends Parasitol. 2013;29:567–78.
4. Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. Emerging fungal threats to animal, plant and ecosystem health. Nature. 2012;484:186–94.
5. Didier ES, Weiss LM. Microsporidiosis: current status. Curr Opin Infect Dis. 2006;19:485–92.
6. Ishihara R. The life cycle of *Nosema bombycis* as revealed in tissue culture cells of *Bombyx mori*. J Invertebr Pathol. 1969;14:316–20.
7. Ma Z, Li C, Pan G, Li Z, Han B, Xu J, Lan X, Chen J, Yang D, Chen Q, Sang Q, Ji X, Li T, Long M, Zhou Z. Genome-wide transcriptional response of silkworm (*Bombyx mori*) to infection by the microsporidian *Nosema bombycis*. PLoS One. 2013;8:e84137.
8. Higes M, Meana A, Bartolomé C, Botías C, Martín-Hernández R. *Nosema ceranae* (Microsporidia), a controversial 21st century honey bee pathogen. Environ Microbiol Rep. 2013;5:17–29.
9. Simões RA, Feliciano JR, Solter LF, Delalibera I. Impacts of *Nosema* sp. (Microsporidia: Nosematidae) on the sugarcane borer, *Diatraea saccharalis* (Lepidoptera: Crambidae). J Invertebr Pathol. 2015;129:7–12.
10. Desjardins CA, Sanscrainte ND, Goldberg JM, Heiman D, Young S, Zeng Q, Madhani HD, Becnel JJ, Cuomo CA. Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nat Commun. 2015;6:7121.
11. Pelin A, Selman M, Aris-Brosou S, Farinelli L, Corradi N. Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen *Nosema ceranae*. Environ Microbiol. 2015;17:4443–58.
12. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, Didier ES, Fan L, Heiman DI, Levin JZ, Young S, Zeng Q, Troemel ER. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. Genome Res. 2012;22:2478–88.
13. Pombert JF, Xu J, Smith DR, Heiman D, Young S, Cuomo CA, Weiss LM, Keeling PJ. Complete genome sequences from three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian *Encephalitozoon cuniculi*. Eukaryot Cell. 2013;12:503–11.
14. Haag KL, Traunecker E, Ebert D. Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. Mol Ecol. 2013;22:314–26.
15. Gómez-Moracho T, Maside X, Martín-Hernández R, Higes M, Bartolomé C. High levels of genetic diversity in *Nosema ceranae* within *Apis mellifera* colonies. Parasitology. 2014;141:475–81.
16. Gómez-Moracho T, Bartolomé C, Bello X, Martín-Hernández R, Higes M, Maside X. Recent worldwide expansion of *Nosema ceranae* (Microsporidia) in *Apis mellifera* populations inferred from multilocus patterns of genetic variation. Infect Genet Evol. 2015;31:87–94.
17. Maside X, Gómez-Moracho T, Jara L, Martín-Hernández R, De la Rúa P, Higes M, Bartolomé C. Population genetics of *Nosema apis* and *Nosema ceranae*: one host (*Apis mellifera*) and two different histories. PLoS One. 2015;10:e0145609.
18. Gresoviac SJ, Khattra JS, Nadler SA, Kent ML, Devlin RH, Vivares CP, Fuente E, Hedrick RP. Comparison of small subunit ribosomal RNA gene and internal transcribed spacer sequences among isolates of the intranuclear microsporidian *Nucleospora salmonis*. J Eukaryot Microbiol. 2000;47:379–87.
19. Haro M, Del Águila C, Fenoy S, Henriques-Gil N. Intraspecies genotype variability of the microsporidian parasite *Encephalitozoon hellem*. J Clin Microbiol. 2003;41:4166–71.
20. Santín M, Fayer R. *Enterocytozoon bieneusi* genotype nomenclature based on the internal transcribed spacer sequence: a consensus. J Eukaryot Microbiol. 2009;56:34–8.
21. Selman M, Sak B, Kváč M, Farinelli L, Weiss LM, Corradi N. Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. Eukaryot Cell. 2013;12:496–502.

22. Liguory O, Fournier S, Sarfati C, Derouin F, Molina J-MM. Genetic homology among thirteen *Encephalitozoon intestinalis* isolates obtained from human immunodeficiency virus-infected patients with intestinal microsporidiosis. J Clin Microbiol. 2000;38:2389–91.
23. Henriques-Gil N, Haro M, Izquierdo F, Fenoy S, del Águilat C. Phylogenetic approach to the variability of the microsporidian *Enterocytozoon bieneusi* and its implications for inter- and intrahost transmission. Appl Environ Microbiol. 2010;76:3333–42.
24. Schmid-Hempel P. Immune defence, parasite evasion strategies and their relevance for "macroscopic phenomena" such as virulence. Philos Trans R Soc B Biol Sci. 2009;364:85–98.
25. Frank SA. Models of parasite virulence. Q Rev Biol. 1996;71:37–78.
26. Ebert D. Host-parasite coevolution: Insights from the *Daphnia*-parasite model system. Curr Opin Microbiol. 2008;11:290–301.
27. Stollewerk A. The water flea *Daphnia* - a "new" model system for ecology and evolution? J Biol. 2010;9:21.
28. Refardt D, Canning EU, Mathis A, Cheney SA, Lafranchi-Tristem NJ, Ebert D. Small subunit ribosomal DNA phylogeny of microsporidia that infect *Daphnia* (Crustacea: Cladocera). Parasitology. 2002;124:381–9.
29. Refardt D, Decaestecker E, Johnson PTJ, Vávra J. Morphology, molecular phylogeny, and ecology of *Binucleata daphniae* n. g., n. sp. (Fungi: Microsporidia), a parasite of *Daphnia magna* Straus, 1820 (Crustacea: Branchiopoda). J Eukaryot Microbiol. 2008;55:393–408.
30. Weigl S, Körner H, Petrusek A, Seda J, Wolinska J. Natural distribution and co-infection patterns of microsporidia parasites in the *Daphnia longispina* complex. Parasitology. 2012;139:870–80.
31. Wolinska J, Giessler S, Koerner H. Molecular identification and hidden diversity of novel *Daphnia* parasites from European lakes. Appl Environ Microbiol. 2009;75:7051–9.
32. Vávra J, Larsson JIR. *Berwaldia schaefernai* (Jírovec, 1937) comb. n. (protozoa, microsporida), fine structure, life cycle, and relationship to *Berwaldia singularis* Larsson, 1981. Eur J Protistol. 1937;1994(30):45–54.
33. Vossbrinck CR, Andreadis TG, Vavra J, Becnel JJ. Molecular phylogeny and evolution of mosquito parasitic microsporidia (Microsporidia: Amblyosporidae). J Eukaryot Microbiol. 2004;51:88–95.
34. Vávra J, Hyliš M, Oborník M, Vossbrinck CR. Microsporidia in aquatic microcrustacea: The copepod microsporidium *Marssoniella elegans* Lemmermann, 1900 revisited. Folia Parasitol (Praha). 2005;52:163–72.
35. Simakova AV, Pankova TF, Tokarev YS, Issi IV. *Senoma* gen. n., a new genus of microsporidia, with the type species *Senoma globulifera* comb. n.(syn. *Issia globulifera* Issi et Pankova 1983) from the malaria mosquito *Anopheles messeae* Fall. Protistology. 2005;4:135–44.
36. Dunn AM, Hogg JC, Hatcher MJ. Transmission and burden and the impact of temperature on two species of vertically transmitted microsporidia. Int J Parasitol. 2006;36:409–14.
37. Martín-Hernández R, Meana A, García-Palencia P, Marín P, Botías C, Garrido-Bailón E, Barrios L, Higes M. Effect of temperature on the biotic potential of honeybee microsporidia. Appl Environ Microbiol. 2009;75:2554–7.
38. Wolinska J, Spaak P, Koerner H, Petrusek A, Seda J, Giessler S. Transmission mode affects the population genetic structure of *Daphnia* parasites. J Evol Biol. 2011;24:265–73.
39. Seda J, Petrusek A, Machacek J, Smilauer P. Spatial distribution of the *Daphnia longispina* species complex and other planktonic crustaceans in the heterogeneous environment of canyon-shaped reservoirs. J Plankton Res. 2007;29:619–28.
40. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.
41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
42. González-Tortuero E, Rusek J, Petrusek A, Gießler S, Lyras D, Grath S, Castro-Monzón F, Wolinska J. The quantification of representative sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*. Mol Ecol Resour. 2015;15:1385–95.
43. Wirawan A, Harris RS, Liu Y, Schmidt B, Schröder J. HECTOR: a parallel multistage homopolymer spectrum based error corrector for 454 sequencing data. BMC Bioinformatics. 2014;15:131.
44. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.
45. Giessler S, Wolinska J. Capturing the population structure of microparasites: Using ITS-sequence data and a pooled DNA approach. Mol Ecol Resour. 2013;13:918–28.
46. Templeton AR, Crandall KA, Sing CF. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III Cladogram estimation Genetics. 1992;132:619–33.
47. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26:32–46.
48. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna Austria 2015:{ISBN} 3–900051–07–0.
49. Dray S, Dufour AB. The *ade4* package: implementing the duality diagram for ecologists. J Stat Softw. 2007;22:1–20.
50. Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. vegan: community ecology package. 2015.
51. Venables WN, Ripley BD. Modern Applied Statistics With S. 4th ed. New York: Springer; 2003.
52. Wheeler TJ, Kececioglu JD. Multiple alignment by aligning alignments. In: Bioinformatics, vol. 23. 2007. p. i559–68.
53. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17:540–52.
54. Simonsen M, Mailund T, Pedersen CS. Rapid Neighbour-Joining. In: Crandall KA, Lagergren J, editors. Algorithms in Bioinformatics, vol. 5251. Berlin Heidelberg: Springer; 2008. p. 113–22. Lecture Notes in Computer Science.
55. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.
56. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 2012;9:772–2.
57. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 2008;57:758–71.
58. Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A. How many bootstrap replicates are necessary? J Comput Biol. 2010;17:337–54.
59. Clement M, Posada D, Crandall KA. TCS: a computer program to estimate gene genealogies. Mol Ecol. 2000;9:1657–9.
60. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007;2:2366–82.
61. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;105:437–60.
62. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975;7:256–76.
63. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. *PopGenome*: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31:1929–36.
64. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.
65. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing *mothur*: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75:7537–41.
66. Städler T, Haubold B, Merino C, Stephan W, Pfaffelhuber P. The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. Genetics. 2009;182:205–16.
67. Jakobsen IB, Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Comput Appl Biosci. 1996;12:291–5.
68. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006;172:2665–81.
69. Fitch WM: Toward finding the tree of maximum parsimony. Proc. 8th Int. Conf. Numer. Taxon. 1975, 189–230.
70. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.
71. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics. 1985;111:147–64.
72. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics. 2010;26:2462–3.
73. Choi Y, Lee Y, Cho KS, Lee S, Russell J, Choi J, Jeong G. Chimerical nature of the ribosomal RNA gene of a *Nosema* species. J Invertebr Pathol. 2011;107:86–9.
74. Nadler SA. Microevolution and the genetic structure of parasite populations. J Parasitol. 1995;81:395–403.
75. Huyse T, Poulin R, Théron A. Speciation in parasites: a population genetics approach. Trends Parasitol. 2005;21:469–75.

76. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci U S A. 2012;109:6241–6.
77. Ironside JE. Diversity and recombination of dispersed ribosomal DNA and protein coding genes in Microsporidia. PLoS One. 2013;8:e55878.
78. Peyretaillade E, Peyret P, Metenier G, Vivares CP, Prensier G: The identification of rRNA maturation sites in the microsporidian *Encephalitozoon cuniculi* argues against the full excision of presumed ITS1 sequence. J Eukaryot Microbiol 2001;48:60S–62S.
79. Huang W-F, Tsai S-J, Lo C-F, Soichi Y, Wang C-H. The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. Fungal Genet Biol. 2004;41:473–81.
80. Tsai S-J, Huang W-F, Wang C-H. Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene. J Eukaryot Microbiol. 2005;52:52–4.
81. Wang L-L, Chen K-P, Zhang Z, Yao Q, Gao G-T, Zhao Y. Phylogenetic analysis of *Nosema antheraeae* (Microsporidia) isolated from Chinese oak silkworm, *Antheraea pernyi*. J Eukaryot Microbiol. 2006;53:310–3.
82. Iiyama K, Chieda Y, Yasunaga-Aoki C, Hayasaka S, Shimizu S. Analyses of the ribosomal DNA region in *Nosema bombycis* NIS 001. J Eukaryot Microbiol. 2004;51:598–604.
83. Ironside JE. Multiple losses of sex within a single genus of Microsporidia. BMC Evol Biol. 2007;7:48.
84. Krebes L, Zeidler L, Frankowski J, Bastrop R. (Cryptic) sex in the microsporidian *Nosema granulosis* - Evidence from parasite rDNA and host mitochondrial DNA. Infect Genet Evol. 2014;21:259–68.
85. O'Mahony EM, Tay WT, Paxton RJ. Multiple rRNA variants in a single spore of the microsporidian *Nosema bombi*. J Eukaryot Microbiol. 2007;54:103–9.
86. Li W, Cama V, Feng Y, Gilman RH, Bern C, Zhang X, Xiao L. Population genetic analysis of *Enterocytozoon bieneusi* in humans. Int J Parasitol. 2012;42:287–93.
87. Li W, Cama V, Akinbo FO, Ganguly S, Kiulia NM, Zhang X, Xiao L. Multilocus sequence typing of *Enterocytozoon bieneusi*: Lack of geographic segregation and existence of genetically isolated sub-populations. Infect Genet Evol. 2013;14:111–9.
88. Tellier A, Lemaire C. Coalescence 2.0: a multiple branching of recent theoretical developments and their applications. Mol Ecol. 2014;23:2637–52.
89. Thornton K. Recombination and the properties of Tajima's *D* in the context of approximate-likelihood calculation. Genetics. 2005;171:2143–8.
90. Wilkinson TJ, Rock J, Whiteley NM, Ovcharenko MO, Ironside JE. Genetic diversity of the feminising microsporidian parasite *Dictyocoela*: New insights into host-specificity, sex and phylogeography. Int J Parasitol. 2011;41:959–66.
91. Lee SC, Heitman J, Ironside JE: Sex and the Microsporidia. In: Microsporidia: Pathogens of Opportunity. Edited by Weiss LM, Becnel JJ. Chichester, UK: John Wiley & Sons, Inc.; 2014:231–243.
92. Becnel JJ, Sprague V, Fukuda T, Hazard EI. Development of *Edhazardia aedis* (Kudo, 1930) n. g., n. comb. (Microsporida: Amblyosporidae) in the mosquito *Aedes aegypti* (L.) (Diptera: Culicidae). J Protozool. 1989;36:119–30.
93. Diarra K, Toguebaye BS. On the development cycle and ultrastructure of *Vavraia culicis* Weiser, 1947 (microsporida, pleistophoridae) with comments on the taxonomy of the genus *Vavraia* Weiser, 1977. Eur J Protistol. 1947;1991(27):134–40.

Chapter 5 – Discussion

5.1. General discussion

Using molecular techniques and bioinformatic analyses, this PhD project aimed to obtain a better understanding of population structure of *C. mesnili* and microsporidian parasites from natural populations of *Daphnia*. Such information would allow a comprehensive understanding of the biology of *Daphnia*-infecting microparasites as well as population dynamics of host-parasite under natural conditions.

5.1.1. The use of Next Generation Sequencing platforms in population biology

Before discussing the parasite population structure and its dynamics, accurate methods for identification of parasites' strains or variants are needed. As discussed in Chapter 2, there is a controversy about the use of Next Generation Sequencing (NGS) platforms in ecological studies, especially considering the number of taxon-specific reads as an indicator for the abundance of the different taxa (e.g. Amend et al., 2010; Baldrian et al., 2013; Deagle et al., 2013). Other pipelines such as SESAME Barcode (Piry et al., 2012) and the unnamed Bálint's automated metabarcoding pipeline for Fungi (Bálint et al., 2014) were also developed to identify and (semi-)quantify variation within populations using distance methods and neighbour-joining clustering. However, distance and neighbour-joining methods are less robust and precise than network methods to identify representative sequences (Giessler and Wolinska, 2013). For this reason, I developed the Quantification of Representative Sequences (QRS) pipeline as a modular pipeline to analyse variation within population using network-based approaches, which consider the existence of reticulations and polytomies (Posada and Crandall, 2001).

As provided in Chapter 2, QRS was validated by doing two different comparisons. First, the frequencies of representative ITS1 sequences derived from 454 and Sanger datasets from *C. mesnili* Czech Republic populations were compared with each other (see Chapter 2 for details). In this comparison, the frequencies were similar between the sequencing methods, indicating the suitability of 454 to perform population biology analyses. Second, a subset of a published dataset from 18S rDNA of intertidal meiofauna of Alabama (Brannock et al., 2014) was processed for metagenomics applications using three different metabarcoding programs: QRS, *mothur* (Schloss et al., 2009) and

UPARSE (Edgar, 2013). Obtained results were similar between the pipelines when all samples were pooled (Data S3 from González-Tortuero et al., 2015). However, if we consider each sample separately, the results differ between the employed pipelines (Fig. 5.1). This finding is consistent with recent publications about the lack of reproducibility in metabarcoding studies, which it is due to mainly the use of greedy clustering algorithms to obtain the different operational taxonomic units (OTUs; Chen et al., 2013; He et al., 2015; Koskinen et al., 2014). In fact, QRS was the only bioinformatic pipeline able to detect phyla that were not present in other datasets, which could be consi-

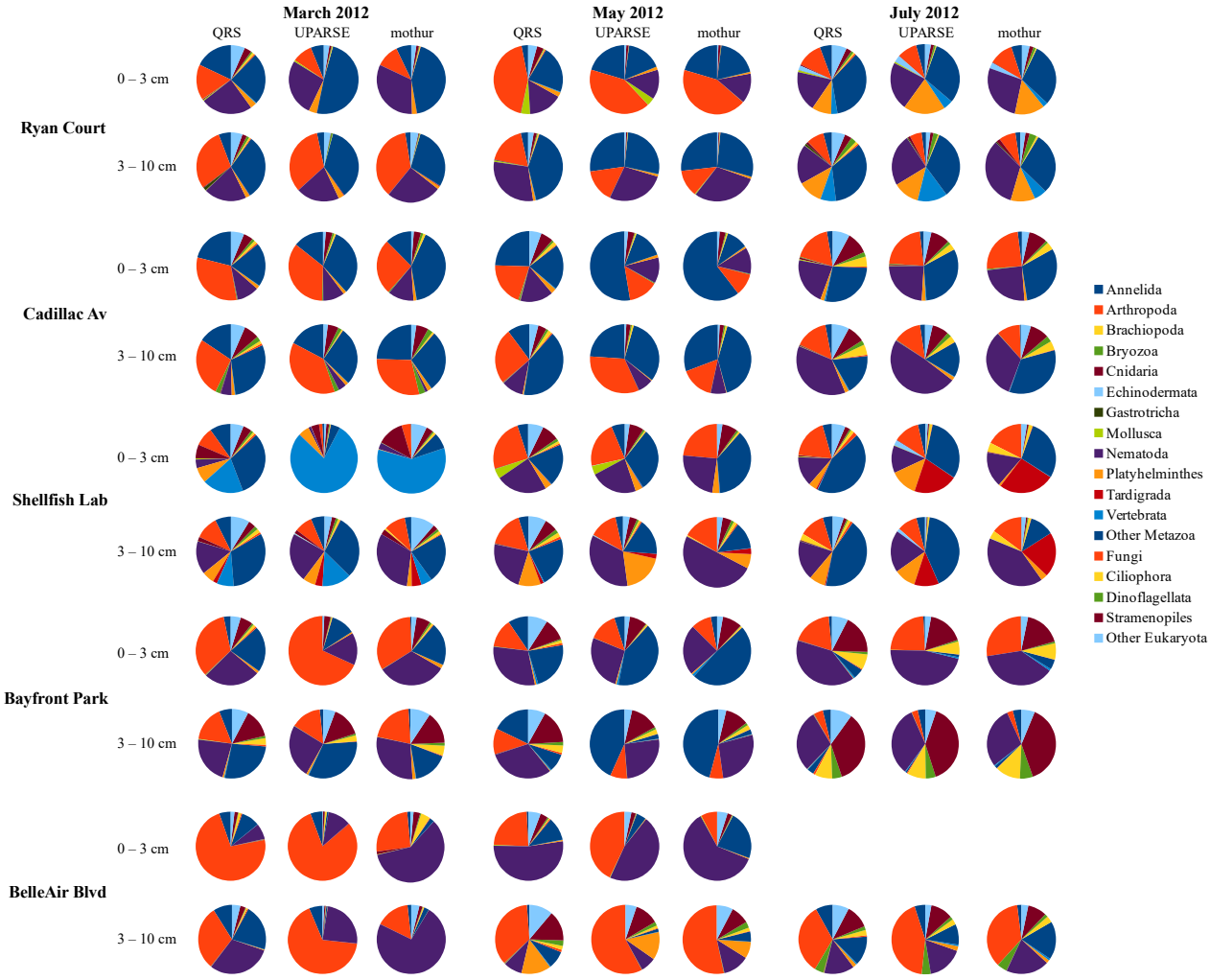

Figure 5.1. Comparison of the results generated by three pipelines (QRS, UPARSE and *mothur*) that were used for validation of the Illumina dataset. Shown are the frequencies of the OTUs as retrieved from the 18S dataset of intertidal meiofauna (Brannock et al., 2014) per sample and depth.

dered as an indicator that the QRS pipeline is more sensitive in clustering sequences to OTUs than other standard metabarcoding programs. These differences between results obtained by different pipelines could lead to biases in the interpretation of alpha and beta diversity due to the over- or underestimation of the number of real OTUs in the samples (Chen et al., 2013; He et al., 2015). Several solutions such as the implementation of the distribution of the sequences across samples in clustering algorithms (Preheim et al., 2013) and the use of graph theory-based algorithms (Wang et al., 2013) were proposed as potential alternatives to greedy clustering algorithms. Nevertheless, there are also other potential biases that could lead a misrepresentation of OTUs such as the misidentification of chimeric sequences (Edgar, 2013) and the different outputs from de-noising algorithms (Koskinen et al., 2014).

5.1.2. *C. mesnili* population dynamics and host-genotype specificity

In Chapter 3, *C. mesnili* population dynamics were tracked in a natural *Daphnia* population, by exploring samples collected over four years, from a single study lake, Greifensee. I tested if the most common parasite genotype decreased over time; this being expected according to the fluctuating selection dynamics; see Chapter 1 for more information. Under the assumption of this theory, common parasites, which presumably are adapted to the most common host genotypes, are in disadvantage and must then decrease in frequency (Clarke, 1976; Hamilton, 1980; Jaenike, 1978). In the *C. mesnili* population of lake Greifensee, I detected a decrease of the most common ITS1 representative sequence and an increase of the rare representative sequences. Although this observation may suggest a fluctuating selection dynamics, there are other potential explanations that could explain the decrease of the most common sequence, such as changes in the external environment and a parasite yearly expansion after a bottleneck. Specifically, an alternative explanation for the parasite genetic change in frequency over time could be that different parasite genotypes are being favoured over time due to changes in the external environment (reviewed in Wolinska and King, 2009). Interaction between parasites and temperature was significant in the *C. mesnili-Daphnia* system when assessed experimentally (Schoebel et al., 2011). However, there is no indication that conditions in Greifensee have changed in any particular direction within the four years examined. Another alternative explanation is that a yearly expansion following a bottleneck might explain the parasite genetic change in frequency seen in the lake Greifensee. This potential explanation will be discussed in more detail in the Section 5.2.3.

In Chapter 3, only a weak signal in the host-parasite specificity between *C. mesnili* and

Daphnia genotypes was detected. This observation is supported by a previous study where the distribution of *C. mesnili* ITS1 sequences did not differ between *Daphnia* hybrids and their parental species (Wolinska et al., 2014). The lack of specificity between hosts and parasites could be attributed also to high levels of parasite gene flow (e.g. Olival et al., 2013) and the influence of symbionts in host-parasite interactions (Bordenstein and Theis, 2015; Kwiatkowski et al., 2012). In the latter case, the symbionts and hosts interact in a biological market, in which the symbionts are able to outcompete parasites if the host bring resources to the symbionts (Noë and Hammerstein, 1995; Sachs and Simms, 2006). Moreover, as discussed in Chapter 3, there are two other explanations to clarify host specificity and its related factors: the specialisation of parasites and the relationship between host specificity and parasite transmission mode.

Parasites tend to infect specific hosts that are phylogenetically (e.g. Bellec et al., 2014; Pedersen et al., 2005; Sasal et al., 1999) and/or eco(physio)logically related (Adamson and Caira, 1994). The phylogenetic relationship between hosts and parasites are used at a macroevolutionary scale to study the divergence of each species and to test if this divergence in host and parasites is mirrored, having a parallel speciation (Fahrenholz, 1912). However, this kind of studies cannot offer direct insight into the coevolutionary process or rate at which change is occurring within populations (de Vienne et al., 2013; Penczykowski et al., 2015). The physiological relationship between both players results from the interaction between parasites' virulence factors and host's resistance genes. Usually, genetic specificity is explained according to theoretical models such as the gene-for-gene (Flor, 1971; reviewed in Thompson and Burdon, 1992), the matching alleles (e.g. Grosberg and Hart, 2000), the inverse gene-for-gene (Fenton et al., 2009) and the hologenome (Bordenstein and Theis, 2015) hypotheses. In most host-parasite systems, virulence factors and resistance loci are unknown although their identification is necessary in order to reveal their interaction (reviewed in Penczykowski et al., 2015). In clinical studies, virulence factors are often used to identify parasite strains instead of neutral markers (e.g. Mooi et al., 1999; Ruiz et al., 2002; Xiang et al., 1995). Virulence factors are normally present in multiple copies and they could be recombined due to the presence of hypervariable regions (e.g. Baldo et al., 2005; Bhat et al., 1991; Zhang et al., 1997). The presence of different copies of virulence factors allows the parasite to infect more potential hosts, increasing the parasite transmission rate (Frank, 1996). All these genes could evolve under balanced selection to optimise parasite virulence because highly virulent parasites kill hosts too quickly to allow their transmission (Alizon et al., 2009). The same applies for host resistance genes, which are hypervariable in order to protect against new potential parasites (e.g. Barrangou et al., 2007; Par-

niske et al., 1997; Schatz, 2004). However, the expression of resistance genes could imply a reduction of the reproductive traits supporting a trade-off between the immunity and reproduction (French et al., 2007; Zuk and Stoehr, 2002). For this reason, the identification of virulence factors and resistance genes is a proper approach to explain the genetic specificity between hosts and parasites, but not to identify parasite strains or variants. The latter must be conducted with neutral markers.

5.1.3. Existence of sexuality in *Daphnia*-infecting microsporidia

In Chapter 4, the patterns of geographic population structure, intraspecific genetic variation, and recombination events of the ITS sequence were compared between two microsporidia taxa: *Berwaldia* and MIC1. This was done in order to better understand the biology of these parasite species. The lack of geographic variation in *Berwaldia* suggested the existence of a mobile secondary host, as well as the geographical differences between the two MIC1 populations could indicate that the secondary host (if exists) does not appear to facilitate dispersal to the same degree. Moreover, the recombination tests pointed out that there is cryptic sex in *Berwaldia* and pure asexuality in MIC1. Although gene recombination tests have been used to demonstrate cryptic sex in Microsporidia (Ironside, 2013; Wilkinson et al., 2011), such analyses may not be sufficient to unambiguously confirm the presence of sexual cycles in this group, because multiple and heterogeneous copies of rDNA could recombine nonhomologously (Lee et al., 2014). Thus, other evidence is needed, such as confirmation of the existence of polyploid stages or the presence of meiosis-related genes.

For a long time, microsporidia were presumed to be asexual organisms. However, if we consider that they could have a fungi-like ancestor, it is reasonable that such ancestor could have sexual reproduction (Dyer, 2008; Lee et al., 2008). Probably, sex has been eliminated in several microsporidian species as a consequence of the morphological miniaturisation and genomic compaction (Corradi and Slamovits, 2011; Haag et al., 2014), which are considered adaptations to the parasitic lifestyle (Poulin and Randhawa, 2015). Nevertheless, the existence of diplokaryotic cells (i.e. cells with binucleated nucleus; e.g. Freeman and Sommerville, 2009; Ironside, 2007; Vávra et al., 2005) or plurinucleated cells (e.g. Bylén and Larsson, 1994; Leiro et al., 1996; Lom et al., 1989; Vagelli et al., 2005; Vávra and Larsson, 1994) indicated the existence of diploid stages in the life cycles of several microsporidian species. Due to these observations, unusual meiotic processes in microsporidia were proposed (reviewed in Lee et al., 2014). In fact, if there is any sexual cycle present, the different sets of chromosomes in polyploid species must be sorted during meiosis to produce bal-

anced gametes.

Although microsporidian genomes are the most reduced ones across eukaryotes (Katinka et al., 2001; Slamovits et al., 2004), the presence of genes related with meiosis was described in some microsporidial species such as *Antonosema locustae*, *Encephalitozoon cuniculi*, *Enterocytozoon bieneusi* and *Nematocida parisii*. However, the majority of these genes are also related with DNA reparation, questioning their relevance in microsporidian cryptic sexuality (reviewed in Lee et al., 2014). Recently, the presence of a sex locus in Zygomycota was described in several microsporidia, including *Enc. cuniculi*, *Enc. intestinalis*, *Enc. hellem*, *Ent. bieneusi*, *A. locustae* and *Nosema ceranae* (Lee et al., 2010, 2008). The sex locus consists of a triose phosphate transporter, an unknown additional ORF, a high-mobility-group (HMG) transcription factor and an RNA helicase gene. The HMG genes are named *sexP/M*, and these genes are encoded in the positive and the negative strand of the DNA, respectively. Both genes play a key role in the mating system, although the RNA helicase gene is not necessarily linked to the sex-related locus (Lee et al., 2010). A similar sex locus was recently detected in two microsporidian parasites of mosquitoes (*Edhazardia aedis* and *Vavraia culicis*; Desjardins et al., 2015). This sex locus was only expressed in *E. aedis*, confirming the existence of sexuality in this species (Becnel et al., 1989) and providing an explanation for the lack of sexual cycles in *V. culicis* (Diarra and Toguebaye, 1991). The presence of this gene cluster in *Berwaldia* and MIC1 should be tested using genomic and transcriptomic analyses to corroborate the existence of cryptic sexuality in *Berwaldia* and pure asexuality in MIC1.

5.2. Future directions

5.2.1. Implementing new algorithms and techniques in metabarcoding studies

In the present PhD thesis, I developed the QRS pipeline (as described in the Chapter 2) to identify intraspecific variation in NGS datasets. Currently, this bioinformatic pipeline only uses Neighbour Joining (Saitou and Nei, 1987) as a clustering algorithm, and Statistical Parsimony (Templeton et al., 1992) as a network approach (as defined in Section 5.1.1). However, other network methods such as reduced median networks (Bandelt et al., 1995) and median-joining networks (Bandelt et al., 1999) are also being used to identify new variants or strains in population ecology and they should be implemented in future versions of the QRS pipeline.

Moreover, NGS platforms are evolving fast, and nowadays new molecular sequencing techniques such as Illumina (e.g. Brannock et al., 2014; Degnan and Ochman, 2012; Schmidt et al.,

2013), Ion Torrent (e.g. Deagle et al., 2013; Jünemann et al., 2012), PacBIO (Fichot and Norman, 2013; Mosher et al., 2014, 2013) and MinION (Greninger et al., 2015; Mikheyev and Tin, 2014) are used to perform metabarcoding or population biology studies. Therefore, although the QRS pipeline is able to deal with Illumina datasets (as indicated in Chapter 2), it would be necessary to update this bioinformatic software to deal with other emerging platforms. In the case of Ion Torrent, implementing tools that are able to correct for the high number of erroneous sequences with high GC content (Deagle et al., 2013; Quail et al., 2012) is needed. Another needed improvement for its application in Ion Torrent datasets is the reduction of homopolymer errors. Although the homopolymers correction in the QRS pipeline was originally based on low-complexity masking algorithms (Morgulis et al., 2006), a new tool for minimising the effect of homopolymers (HECTOR; Wirawan et al., 2014) was implemented in a recent version of the mentioned pipeline (and was used in Chapters 3 and 4 to correct such errors). Additionally, in future versions of the pipeline, a new algorithm based on the Poisson binomial distribution (Puente-Sánchez et al., 2015) will be implemented to avoid the use of the de-noising step, because de-noising algorithms negatively influences in taxonomic affiliation (Koskinen et al., 2014). In respect to the PacBIO sequencing platform, there is very little work done in its application to metagenomics and the established protocols are simple adaptations from 454 Pyrosequencing and Illumina (Fichot and Norman, 2013; Mosher et al., 2014). Recently, a new bioinformatic protocol was developed to deal with PacBIO, but only a small fraction of the sequences are really informative due to the high error rates (Schloss et al., 2016). This suggests that this sequencing platform should be improved in order to be applied in metagenomics. Finally, MinION was recently developed to shorten the time between specimen collection to first sequence data without relying on PCR. However, this platform is currently unsuitable for genotyping due to their high error rates and, as in the case of PacBIO, improving MinION platform is needed for its use in population biology (Mikheyev and Tin, 2014).

5.2.2. Testing the suitability of the internal transcribed spacers for parasites' strains identification

As indicated in the Chapter 1, there is a controversy about the suitability of the ITS1 sequences in molecular ecology. In the present PhD thesis, ITS1 sequences were used to discriminate representative sequences in *C. mesnili* and in two microsporidia taxa—all common parasites of *Daphnia*. ITS1 sequences are the only known polymorphic sequences in *C. mesnili*, which could allow for the iden-

tification of different microparasitic strains or variants (Giessler and Wolinska, 2013). In a similar way, the use of ITS1 to discriminate strains or variants was successful in amoebozoa (Köhsler et al., 2006) and trichomonads (Gaspar da Silva et al., 2007; Ibáñez-Escribano et al., 2014). However, such markers might not be suitable to identify parasite strains, as it happened in oomycetes (Robideau et al., 2011), dinoflagellates (Stern et al., 2012) or in other protozoa like apicomplexa (Homan et al., 1997) and parabasalids (Lollis et al., 2011). Moreover, as discussed in Chapter 4 and in Wolinska et al. (2011b), ITS variability may not be a good estimator of the intraspecific variability in Microsporidia. In fact, a low ITS variability was described in *Encephalitozoon cuniculi* (Selman et al., 2013) whereas these microsporidia have a high intraspecific variability according to the whole genome analysis (Pombert et al., 2013). These differences are probably due to the sequence length, as *E. cuniculi* presents a short ITS sequence (28-45 bp., Selman et al., 2013) while in other microsporidia like *Enterocytozoon bieneusi* ITS sequence are 243 bp long. The ITS is the only known polymorphic marker also in *E. bieneusi* (Henriques-Gil et al., 2010). Nevertheless, strains or variants identified with the ITS1 marker should be confirmed with other neutral markers (see below) which might enable the identification of strains with higher resolution.

Ribosomal DNA is present in multiple copies throughout the genome and each copy is a potential target for mutations leading to intragenomic variation (as explained in Chapter 1). For long time, it was considered that such regions (including the ITS1 sequences) evolved under concerted evolution. Under this hypothesis, ribosomal genes become homogenised by unequal crossing over and gene conversion (reviewed in Liao, 1999). As unequal crossing over is more common in homologous recombination than in non-homologous end joining (Goldman, 1996), concerted evolution would be stronger for tandem arrays of genes than for dispersed repeated DNA regions. Thus, differences in obtained sequences between strains could be due to the number of ITS1 copies found in tandem in the same chromosomes, rather than because of the simple ITS1 sequence variability (as detected in *Nosema apis*; Gatehouse and Malone, 1998). However, the high intragenomic ITS1 variability in several microorganisms such as the human intestinal protozoan parasite *Dientamoeba fragilis* (Bart et al., 2008) and the foraminiferan *Elphidium macellum* (Pillet et al., 2012) argues against the concerted evolution of ribosomal genes. In fact, the high intragenomic ITS1 variability could suggest a birth-and-death process, where new genes originate from gene duplication and some are maintained in the genomes while others are eliminated or become non-functional (Nei et al., 1997). Birth-and-death processes are expected to happen when the multi-copy DNA sequences are dispersed throughout the genome, as observed in the microsporidian *Encephalitozoon cuniculi*

(Peyretaillade et al., 1998), *Nosema bombi* (O'Mahony et al., 2007; Tay et al., 2005) and *N. bombycis* (Liu et al., 2008). To elucidate ribosomal gene evolution according to the above hypotheses (concerted evolution and birth-and-death process), the spatial distribution of the ITS1 in the chromosomes would have to be evaluated by chromosomes mapping (Biderre et al., 1997; Peyretaillade et al., 1998) and by *de novo* sequencing of the genomes of the microparasites (reviewed in Keeling et al., 2014).

The latter analysis brings also the opportunity to search for other neutral markers that could be used for population genetic tests to study the evolution of parasite populations and to evaluate their genetic diversity. As indicated above, ITS might evolve under birth-and-death processes, which could explain the high intragenomic variability of these sequences. To avoid this matter, it is convenient to use single copy genes which are neutral markers such as the 70kDa heat shock protein (hsp70; Haag et al., 2013a, 2013b). The use of these alternative neutral markers will result in obtaining better predictions of the genetic diversity and the evolution of parasite populations (Gómez-Moracho et al., 2015, 2014; Roudel et al., 2013). Other alternatives is the use of micro- or minisatellites (Li et al., 2013, 2012) or even whole genome analysis (Cuomo et al., 2012; Pelin et al., 2015; Pombert et al., 2013) to study the genetic diversity and the evolution of parasite populations.

5.2.3. Deciphering the life cycle of *C. mesnili* and *Daphnia*-infecting microsporidia

As indicated in the Chapter 1, the life cycles of the most common microparasites infecting *Daphnia* are not well known. In the present PhD thesis, I was focusing on the description of population dynamics and host-genotype specificity of *C. mesnili* (Chapter 3) and the genetic diversity and the patterns of geographic population structure in two microsporidian parasites: *Berwaldia* and MIC1 (Chapter 4). All this information could help to predict the biology of these parasites.

In chapter 3, an alternative explanation to the observed decrease of the most common genotype in *C. mesnili*, beside a postulated mechanism of negative frequency-dependant selection, is a yearly expansion following a bottleneck. It is still unknown how *C. mesnili* survives between the epidemics, as this microparasite repeatedly appears in later summer and disappears towards the end of autumn. *C. mesnili* may persist either under very low (and, thus, undetectable) densities in *Daphnia* host, in as of yet unknown alternative hosts (like fish) or in the sediment as spores. Population genetic tests such as Tajima's *D* (Tajima, 1989), Fu's *FS* (Fu, 1997) and/or Ramos-Onsins' and

Rozas' R_2 (Ramos-Onsins and Rozas, 2002) are necessary in order to test for the existence of an expansion after a bottleneck in the parasite populations.

Ichthyosporean parasites are known to infect the digestive organs of various marine vertebrates (mainly fishes) and invertebrates (reviewed in Suga and Ruiz-Trillo, 2015). To verify the existence of reservoir hosts in *C. mesnili* life cycle, it should be recommendable to screen freshwater animals and search for the presence of this microparasite via nested PCR of the SSU rRNA genes. These kind of approaches have been successfully applied for the detection of intracellular microorganisms such as *Wolbachia* (e.g. Arthofer et al., 2009a; Taylor and Hoerauf, 1999; Werren et al., 1995). However, nested PCR is prone to generate a high number of false positives and contamination (Arthofer et al., 2009a, 2009b). Due to this, other approaches such as PCR followed by consequent hybridization (Arthofer et al., 2009b; Schneider et al., 2014) and quantitative PCR (Gosiewski et al., 2014; Mee et al., 2015) were proposed to detect the presence of microorganisms even in low titre infections. However, *C. mesnili* is able to survive in laboratory cultures (Bittner et al., 2002; Lohr et al., 2010a), which could indicate the lack of necessity of a reservoir host in the *C. mesnili* life cycle.

Recently, analysis of environmental rRNA libraries suggested the existence of free living species or free living stages of ichthyosporea (del Campo and Ruiz-Trillo, 2013; Mendoza et al., 2002; Suga and Ruiz-Trillo, 2015; van Hannen et al., 1999). In order to test if *C. mesnili* is able to have any diapause stages in the sediments (as proposed also in Chapter 3), molecular screening of sediments should be used. In this kind of studies, sediment cores are processed to describe the microorganism diversity using universal primers, which amplifies different hypervariable regions of the small subunit of the ribosomal genes (e.g. Berney et al., 2004; Dawson and Pace, 2002; Lawley et al., 2004). Nevertheless, these hypervariable regions exhibit different degrees of sequence diversity, and no single hypervariable region is enough to differentiate between all living organisms, as seen in bacteria (Chakravorty et al., 2007). For this reason, to detect only the presence or absence of *C. mesnili* spores in sediments, it should be convenient to apply PCR with hybridization and quantitative PCR (as described before) using specific primers for *C. mesnili* small subunit of the ribosomal gene.

In Chapter 4, I predicted that *Berwaldia* may have a mobile secondary host whereas MIC1 might not have a secondary host (or if it exists it should not facilitate its dispersal) according to the geographic variation, genetic diversity and recombination tests. However, these results are just bioinformatic predictions and they need to be demonstrated in field and/or laboratory work. To

verify them, sampling benthic insects and searching for insects that are positive for *Berwaldia* and/or MIC1 (as described above for the *C. mesnili* problematics) should be performed. I hypothesise that mosquitoes and/or caddisflies are the secondary hosts in *Berwaldia*, as described for copepods-infecting microsporidia (Vávra et al., 2005; Vossbrinck et al., 2004). Although the transmission of *Amblyospora* spp. between copepods and mosquitoes was described in previous studies (Andreadis, 1994; Sweeney et al., 1990), the release of spores to the environment (during mosquito life or after mosquito death) allows that copepods would be randomly infected with these microsporidia during grazing.

However, I cannot discard that strains of *Berwaldia* or MIC1 are not maintained in the laboratory due to the influence of abiotic factors such as temperature (Dunn et al., 2006; Martín-Hernández et al., 2009). Although higher temperatures cause the increase of the number of spores in *Nosema apis* and *N. ceranae* (Martín-Hernández et al., 2009), in aquatic microsporidia it seems that reduced temperatures has a greater impact on the parasite burden in host cells (Dunn et al., 2006). Also the transmission of aquatic microsporidia is influenced by the temperature because high temperatures increases the vertical transmission and low temperatures increases the horizontal transmission of *N. granulosis* and *Dictyocoela duebenum* (Dunn et al., 2006). To test the influence of temperature on the life cycle of *Berwaldia* and MIC1, parasite transmission experiments using different temperature treatments should be performed.

References

Adamson, M.L., Caira, J.N., 1994. Evolutionary factors influencing the nature of parasite specificity. *Parasitology* **109**, S85–S95.

Alizon, S., Hurford, A., Mideo, N., Van Baalen, M., 2009. Virulence evolution and the trade-off hypothesis: history, current state of affairs and the future. *J. Evol. Biol.* **22**, 245–259.

Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., Rohani, P., 2006. Seasonality and the dynamics of infectious diseases. *Ecol. Lett.* **9**, 467–484.

Amend, A.S., Seifert, K.A., Bruns, T.D., 2010. Quantifying microbial communities with 454 pyrosequencing: Does read abundance count? *Mol. Ecol.* **19**, 5555–5565.

Anderson, I.C., Campbell, C.D., Prosser, J.I., 2003. Potential bias of fungal 18S rDNA and internal transcribed spacer polymerase chain reaction primers for estimating fungal biodiversity in soil. *Environ. Microbiol.* **5**, 36–47.

Andreadis, T.G., 1994. Ultrastructural characterization of meiospores of six new species of *Amblyospora* (Microsporida: Amblyosporidae) from Northern *Aedes* (Diptera: Culicidae) mosquitoes. *J. Eukaryot. Microbiol.* **41**, 147–154.

Arnott, A., Barry, A.E., Reeder, J.C., 2012. Understanding the population genetics of *Plasmodium vivax* is essential for malaria control and elimination. *Malar. J.* **11**, 14.

Arthofer, W., Riegler, M., Avtzis, D., Stauffer, C., 2009a. Evidence for low-titre infections in insect symbiosis: *Wolbachia* in the bark beetle *Pityogenes chalcographus* (Coleoptera, Scolytinae). *Environ. Microbiol.* **11**, 1923–1933.

Arthofer, W., Riegler, M., Schneider, D., Krammer, M., Miller, W.J., Stauffer, C., 2009b. Hidden *Wolbachia* diversity in field populations of the European cherry fruit fly, *Rhagoletis cerasi* (Diptera, Tephritidae). *Mol. Ecol.* **18**, 3816–3830.

Balajee, S.A., Borman, A.M., Brandt, M.E., Cano, J., Cuenca-Estrella, M., Dannaoui, E., Guarro, J., Haase, G., Kibbler, C.C., Meyer, W., O'Donnell, K., Petti, C.A., Rodriguez-Tudela, J.L., Sutton, D., Velegraki, A., Wickes, B.L., 2009. Sequence-based identification of *Aspergillus*, *Fusarium*, and mucorales species in the clinical mycology laboratory: where are we and where should we go from here? *J. Clin. Microbiol.* **47**, 877–884.

Baldo, L., Lo, N., Werren, J.H., 2005. Mosaic nature of the *Wolbachia* surface protein. *J. Bacteriol.* **187**, 5406–5418.

Baldrian, P., Větrovský, T., Cajthaml, T., DobiáŠová, P., Petránková, M., Šnajdr, J., Eichlerová, I., 2013. Estimation of fungal biomass in forest litter and soil. *Fungal Ecol.* **6**, 1–11.

Baldwin, B.G., 1992. Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: An example from the compositae. *Mol. Phylogenet. Evol.* **1**, 3–16.

Bálint, M., Schmidt, P.A., Sharma, R., Thines, M., Schmitt, I., 2014. An Illumina metabarcoding pipeline for fungi. *Ecol. Evol.* **4**, 2642–2653.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48.

Bandelt, H.-J., Forster, P., Sykes, B.C., Richards, M.B., 1995. Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712.

Bart, A., van der Heijden, H.M., Greve, S., Speijer, D., Landman, W.J., van Gool, T., 2008. Intragenomic variation in the internal transcribed spacer 1 region of *Dientamoeba fragilis* as a molecular epidemiological marker. *J. Clin. Microbiol.* **46**, 3270–3275.

Becnel, J.J., Sprague, V., Fukuda, T., Hazard, E.I., 1989. Development of *Edhazardia aedis* (Kudo, 1930) n. g., n. comb. (Microsporida: Amblyosporidae) in the mosquito *Aedes aegypti* (L.) (Diptera: Culicidae). *J. Protozool.* **36**, 119–130.

Bell, G., 1982. *The Masterpiece of Nature: The Evolution and Genetics of Sexuality.* Cambridge University Press. ISBN-13: 978-0520045835

Bellec, L., Clerissi, C., Edern, R., Foulon, E., Simon, N., Grimsley, N., Desdevises, Y., 2014. Cophylogenetic interactions between marine viruses and eukaryotic picophytoplankton. *BMC Evol. Biol.* **14**, 59.

Bellemain, E., Carlsen, T., Brochmann, C., Coissac, E., Taberlet, P., Kauserud, H., 2010. ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiol.* **10**, 189.

Berney, C., Fahrni, J., Pawlowski, J., 2004. How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol.* **2**, 13.

Bhat, K.S., Gibbs, C.P., Barrera, O., Morrison, S.G., Jähnig, F., Stern, A., Kupsch, E.-M., Meyer, T.F., Swanson, J., 1991. The opacity proteins of *Neisseria gonorrhoeae* strain MS11 are encoded by a family of 11 complete genes. *Mol. Microbiol.* **5**, 1889–1901.

Biderre, C., Duffieux, F., Peyretaillade, E., Glaser, P., Peyret, P., Danchin, A., Pagès, M., Méténier, G., Vivarès, C.P., 1997. Mapping of repetitive and non-repetitive DNA probes to chromosomes of the microsporidian *Encephalitozoon cuniculi*. *Gene* **191**, 39–45.

Bittner, K., Rothhaupt, K.-O., Ebert, D., 2002. Ecological interactions of the microparasite *Caullerya mesnili* and its host *Daphnia galeata*. *Limnol. Oceanogr.* **47**, 300–305.

Bordenstein, S.R., Theis, K.R., 2015. Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biol.* **13**, e1002226.

Boyer, S.L., Flechtner, V.R., Johansen, J.R., 2001. Is the 16S-23S rRNA Internal Transcribed Spacer region a good tool for use in molecular systematics and population genetics? A case study in Cyanobacteria. *Mol. Biol. Evol.* **18**, 1057–1069.

Brannock, P.M., Waits, D.S., Sharma, J., Halanych, K.M., 2014. High-throughput sequencing characterizes intertidal meiofaunal communities in northern gulf of Mexico (Dauphin Island and Mobile Bay, Alabama). *Biol. Bull.* **227**, 161–174.

Brockhurst, M.A., Chapman, T., King, K.C., Mank, J.E., Paterson, S., Hurst, G.D.D., 2014. Running with the Red Queen: the role of biotic conflicts in evolution. *Proc. R. Soc. B* **281**, 20141382.

Brockhurst, M.A., Koskella, B., 2013. Experimental coevolution of species interactions. *Trends Ecol. Evol.* **28**, 367–375.

Brownstein, J.S., Holford, T.R., Fish, D., 2003. A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environ. Health Perspect.* **111**, 1152–1157.

Burdon, J.J., Thrall, P.H., Brown, A.H.D., 1999. Resistance and virulence structure in two *Linum marginale-Melampsora lini* host-pathogen metapopulations with different mating systems. *Evolution* **53**, 704–716.

Burdon, J., Thompson, J., 1995. Changed patterns of resistance in a population of *Linum marginale* attacked by the rust pathogen *Melampsora lini*. *J. Ecol.* **83**, 199–206.

Bylén, E.K.C., Larsson, J.I.R., 1994. *Napamichum cellatum* n. sp. (Microspora, Thelohaniidae), a new parasite of midge larvae of the genus *Endochironomus* (Diptera, Chironomidae) in Sweden. *J. Eukaryot. Microbiol.* **41**, 450–457.

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J. Microbiol. Methods* **69**, 330–339.

Chen, W., Zhang, C.K., Cheng, Y., Zhang, S., Zhao, H., 2013. A comparison of methods for clustering 16S rRNA sequences into OTUs. *PLoS One* **8**, e70837.

Clarke, B.C., 1976. The ecological genetics of host-parasite relationships, in: Taylor, A.E.R., Muller, R. (Eds.), *Genetic aspects of host-parasite relationships.* Blackwell, Oxford, pp. 87–103.

Corradi, N., Slamovits, C.H., 2011. The intriguing nature of microsporidian genomes. *Brief. Funct. Genomics* **10**, 115–124.

Criscione, C.D., Poulin, R., Blouin, M.S., 2005. Molecular ecology of parasites: Elucidating ecological and microevolutionary processes. *Mol. Ecol.* **14**, 2247–2257.

Cuomo, C.A., Desjardins, C.A., Bakowski, M.A., Goldberg, J., Ma, A.T., Becnel, J.J., Didier, E.S., Fan, L., Heiman, D.I., Levin, J.Z., Young, S., Zeng, Q., Troemel, E.R., 2012. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res.* **22**, 2478–2488.

Dawkins, R., 1989. *The Selfish Gene*. Oxford University Press. ISBN-13: 978-0192860927

Dawson, S.C., Pace, N.R., 2002. Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 8324–8329.

Deagle, B.E., Thomas, A.C., Shaffer, A.K., Trites, A.W., Jarman, S.N., 2013. Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: Which counts count? *Mol. Ecol. Resour.* **13**, 620–633.

Decaestecker, E., Gaba, S., Raeymaekers, J.A.M., Stoks, R., Van Kerckhoven, L., Ebert, D., De Meester, L., 2007. Host–parasite 'Red Queen' dynamics archived in pond sediment. *Nature*

450, 870–873.

Degnan, P.H., Ochman, H., 2012. Illumina-based analysis of microbial community diversity. *ISME J.* **6**, 183–194.

del Campo, J., Ruiz-Trillo, I., 2013. Environmental survey meta-analysis reveals hidden diversity among unicellular opisthokonts. *Mol. Biol. Evol.* **30**, 802–805.

Desjardins, C.A., Sanscrainte, N.D., Goldberg, J.M., Heiman, D., Young, S., Zeng, Q., Madhani, H.D., Becnel, J.J., Cuomo, C.A., 2015. Contrasting host–pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. *Nat. Commun.* **6**, 7121.

de Vienne, D.M., Refrégier, G., López-Villavicencio, M., Tellier, A., Hood, M.E., Giraud, T., 2013. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *New Phytol.* **198**, 347–385.

Diarra, K., Toguebaye, B.S., 1991. On the development cycle and ultrastructure of *Vavraia culicis* Weiser, 1947 (microsporida, pleistophoridae) with comments on the taxonomy of the genus *Vavraia* Weiser, 1977. *Eur. J. Protistol.* **27**, 134–140.

Djimde, A., Doumbo, O.K., Cortese, J.F., Kayentao, K., Doumbo, S., Diourte, Y., Dicko, A., Su, X.-Z., Nomura, T., Fidock, D.A., Wellems, T.E., Plowe, C. V., 2001. A molecular marker for chloroquine-resistant falciparum malaria. *N. Engl. J. Med.* **344**, 257–263.

Downie, S., Ramanath, S., Katz-Downie, D., Llanas, E., 1998. Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid RPO C1 intron sequences. *Am. J. Bot.* **85**, 563–591.

Duffy, M.A., Sivars-Becker, L., 2007. Rapid evolution and ecological host-parasite dynamics. *Ecol. Lett.* **10**, 44–53.

Dunn, A.M., Hogg, J.C., Hatcher, M.J., 2006. Transmission and burden and the impact of temperature on two species of vertically transmitted microsporidia. *Int. J. Parasitol.* **36**, 409–414.

Dybdahl, M.F., Jenkins, C.E., Nuismer, S.L., 2014. Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. *Am. Nat.* **184**, 1–13.

Dybdahl, M.F., Lively, C.M., 1995. Diverse, endemic and polyphyletic clones in mixed populations of a freshwater snail (*Potamopyrgus antipodarum*). *J. Evol. Biol.* **8**, 385–398.

Dyer, P.S., 2008. Evolutionary biology: microsporidia sex—a missing link to fungi. *Curr. Biol.* **18**, R1012–R1014.

Ebert, D., 2008. Host-parasite coevolution: Insights from the *Daphnia*-parasite model system. *Curr. Opin. Microbiol.* **11**, 290–301.

Ebert, D., 2005. Some Parasites of *Daphnia*. URL: http://www.ncbi.nlm.nih.gov/books/NBK2043/

Edgar, R.C., 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* **10**, 996–998.

Fahrenholz, H., 1912. Ectoparasiten und Abstammungslehre. *Zool. Anz.* **41**, 371–374.

Fenton, A., Antonovics, J., Brockhurst, M.A., 2009. Inverse-gene-for-gene infection genetics and coevolutionary dynamics. *Am. Nat.* **174**, E230–E242.

Fichot, E.B., Norman, R.S., 2013. Microbial phylogenetic profiling with the Pacific Biosciences sequencing platform. *Microbiome* **1**, 10.

Fischer, P., Schmetz, C., Bandi, C., Bonow, I., Mand, S., Fischer, K., Büttner, D.W., 2002. *Tunga penetrans*: Molecular identification of *Wolbachia* endobacteria and their recognition by antibodies against proteins of endobacteria from filarial parasites. *Exp. Parasitol.* **102**, 201–211.

Flor, H.H., 1971. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275–296.

Forde, S.E., Thompson, J.N., Bohannan, B.J.M., 2004. Adaptation varies through space and time in a coevolving host-parasitoid interaction. *Nature* **431**, 841–844.

Frank, S.A., 1996. Models of parasite virulence. *Q. Rev. Biol.* **71**, 37–78.

Frank, S.A., 1991. Spatial variation in coevolutionary dynamics. *Evol. Ecol.* **5**, 193–217.

Freeland, J.R., Noble, L.R., Okamura, B., 2000. Genetic consequences of the metapopulation biology of a facultatively sexual freshwater invertebrate. *J. Evol. Biol.* **13**, 383–395.

Freeman, M.A., Sommerville, C., 2009. *Desmozoon lepeophtherii* n. gen., n. sp., (Microsporidia: Enterocytozoonidae) infecting the salmon louse *Lepeophtheirus salmonis* (Copepoda: Caligidae). *Parasit. Vectors* **2**, 58.

French, S.S., DeNardo, D.F., Moore, M.C., 2007. Trade-offs between the reproductive and immune systems: facultative responses to resources or obligate responses to reproduction? *Am. Nat.* **170**, 79–89.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925.

Gaba, S., Ebert, D., 2009. Time-shift experiments as a tool to study antagonistic coevolution. *Trends Ecol. Evol.* **24**, 226–232.

Gandon, S., Buckling, A., Decaestecker, E., Day, T., 2008. Host-parasite coevolution and patterns of adaptation across time and space. *J. Evol. Biol.* **21**, 1861–1866.

Gaspar da Silva, D., Barton, E., Bunbury, N., Lunness, P., Bell, D.J., Tyler, K.M., 2007. Molecular identity and heterogeneity of trichomonad parasites in a closed avian population. *Infect. Genet. Evol.* **7**, 433–440.

Gatehouse, H.S., Malone, L.A., 1998. The ribosomal RNA gene region of *Nosema apis* (Microspora): DNA sequence for small and large subunit rRNA genes and evidence of a large tandem repeat unit size. *J. Invertebr. Pathol.* **71**, 97–105.

Giessler, S., Wolinska, J., 2013. Capturing the population structure of microparasites: Using ITS-sequence data and a pooled DNA approach. *Mol. Ecol. Resour.* **13**, 918–928.

Goldman, A.S.H., 1996. The efficiency of meiotic recombination between dispersed sequences in *Saccharomyces cerevisiae* depends upon their chromosomal location. *Genetics* **144**, 43–55.

Gómez-Moracho, T., Bartolomé, C., Bello, X., Martín-Hernández, R., Higes, M., Maside, X., 2015. Recent worldwide expansion of *Nosema ceranae* (Microsporidia) in *Apis mellifera* populations inferred from multilocus patterns of genetic variation. *Infect. Genet. Evol.* **31**, 87–94.

Gómez-Moracho, T., Maside, X., Martín-Hernández, R., Higes, M., Bartolomé, C., 2014. High levels of genetic diversity in *Nosema ceranae* within *Apis mellifera* colonies. *Parasitology* **141**, 475–481.

González-Tortuero, E., Rusek, J., Petrusek, A., Gießler, S., Lyras, D., Grath, S., Castro-Monzón, F., Wolinska, J., 2015. The Quantification of Representative Sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*. *Mol. Ecol. Resour.* **15**, 1385–1395.

Gosiewski, T., Jurkiewicz-Badacz, D., Sroka, A., Brzychczy-Włoch, M., Bulanda, M., 2014. A novel, nested, multiplex, real-time PCR for detection of bacteria and fungi in blood. *BMC Microbiol.* **14**, 144.

Grabner, D.S., Schertzinger, G., Sures, B., 2014. Effect of multiple microsporidian infections and temperature stress on the heat shock protein 70 (*hsp70*) response of the amphipod *Gammarus pulex*. *Parasit. Vectors* **7**, 170.

Grada, A., Weinbrecht, K., 2013. Next-generation sequencing: Methodology and application. *J. Invest. Dermatol.* **133**, e11.

Green, J., 1974. Parasites and epibionts of Cladocera. *Trans. Zool. Soc. London* **32**, 417–515.

Greischar, M.A., Koskella, B., 2007. A synthesis of experimental work on parasite local adaptation. *Ecol. Lett.* **10**, 418–434.

Greninger, A.L., Naccache, S.N., Federman, S., Yu, G., Mbala, P., Bres, V., Stryke, D., Bouquet, J., Somasekar, S., Linnen, J.M., Dodd, R., Mulembakani, P., Schneider, B.S., Muyembe-Tamfum, J.-J., Stramer, S.L., Chiu, C.Y., 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med.* **7**, 99.

Grosberg, R.K., Hart, M.W., 2000. Mate selection and the evolution of highly polymorphic self/nonself recognition genes. *Science* **289**, 2111–2114.

Gulland, F.M.D., 1992. The role of nematode parasites in Soay sheep (*Ovis aries* L.) mortality during a population crash. *Parasitology* **105**, 493–503.

Gulland, F.M.D., Fox, M., 1992. Epidemiology of nematode infections of Soay sheep (*Ovis aries* L.) on St Kilda. *Parasitology* **105**, 481–492.

Haag, K.L., James, T.Y., Pombert, J.-F., Larsson, R., Schaer, T.M.M., Refardt, D., Ebert, D., 2014. Evolution of a morphological novelty occurred before genome compaction in a lineage of extreme parasites. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15480–15485.

Haag, K.L., Sheikh-Jabbari, E., Ben-Ami, F., Ebert, D., 2013a. Microsatellite and single-nucleotide polymorphisms indicate recurrent transitions to asexuality in a microsporidian parasite. *J. Evol. Biol.* **26**, 1117–1128.

Haag, K.L., Traunecker, E., Ebert, D., 2013b. Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. *Mol. Ecol.* **22**, 314–326.

Hamilton, W.D., 1980. Sex versus non-sex versus parasite. *Oikos* **35**, 282–290.

Hanelt, B., Schmidt-Rhaesa, A., Bolek, M.G., 2015. Cryptic species of hairworm parasites revealed by molecular data and crowdsourcing of specimen collections. *Mol. Phylogenet. Evol.* **82**, 211–218.

Hatton-Ellis, T.W., Noble, L.R., Okamura, B., 1998. Genetic variation in a freshwater bryozoan. I: Populations in the Thames basin, UK. *Mol. Ecol.* **7**, 1575–1585.

He, Y., Caporaso, J.G., Jiang, X.-T., Sheng, H.-F., Huse, S.M., Rideout, J.R., Edgar, R.C., Kopylova, E., Walters, W.A., Knight, R., Zhou, H.-W., 2015. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. *Microbiome* **3**, 20.

Henriques-Gil, N., Haro, M., Izquierdo, F., Fenoy, S., Águilat, C. Del, 2010. Phylogenetic approach to the variability of the microsporidian *Enterocytozoon bieneusi* and its implications for inter- and intrahost transmission. *Appl. Environ. Microbiol.* **76**, 3333–3342.

Homan, W.L., Limper, L., Verlaan, M., Borst, A., Vercammen, M., van Knapen, F., 1997. Comparison of the internal transcribed spacer, ITS 1, from *Toxoplasma gondii* isolates and *Neospora caninum*. *Parasitol. Res.* **83**, 285–289.

Huang, W.-F., Tsai, S.-J., Lo, C.-F., Soichi, Y., Wang, C.-H., 2004. The novel organization and complete sequence of the ribosomal RNA gene of *Nosema bombycis*. *Fungal Genet. Biol.* **41**, 473–481.

Hudson, P.J., Dobson, A.P., 1997. Transmission dynamics and host-parasite interactions of *Trichostrongylus tenuis* in red grouse (*Lagopus lagopus scoticus*). *J. Parasitol.* **83**, 194–202.

Hurst, L.D., Pomiankowski, A., 1991. Causes of sex ratio bias may account for unisexual sterility in hybrids: a new explanation of Haldane's rule and related phenomena. *Genetics* **128**, 841–858.

Ibáñez-Escribano, A., Nogal-Ruiz, J.J., Arán, V.J., Escario, J.A., Gómez-Barrio, A., Alderete, J.F., 2014. Determination of internal transcribed spacer regions (ITS) in *Trichomonas vaginalis* isolates and differentiation among *Trichomonas* species. *Parasitol. Int.* **63**, 427–431.

Iiyama, K., Chieda, Y., Yasunaga-Aoki, C., Hayasaka, S., Shimizu, S., 2004. Analyses of the ribosomal DNA region in *Nosema bombycis* NIS 001. *J. Eukaryot. Microbiol.* **51**, 598–604.

Ironside, J.E., 2013. Diversity and recombination of dispersed ribosomal DNA and protein coding genes in Microsporidia. *PLoS One* **8**, e55878.

Ironside, J.E., 2007. Multiple losses of sex within a single genus of Microsporidia. *BMC Evol. Biol.* **7**, 48.

Iwen, P.C., Hinrichs, S.H., Rupp, M.E., 2002. Utilization of the internal transcribed spacer regions as molecular targets to detect and identify human fungal pathogens. *Med. Mycol.* **40**, 87–109.

Jaenike, J., 1978. An hypothesis to account for the maintenance of sex within populations. *Evol. Theory* **94**, 191–194.

Jarosz, A.M., Burdon, J.J., 1991. Host-pathogen interactions in natural populations of *Linum marginale* and *Melampsora lini*. II: Local and regional variation in patterns of resistance and racial structure. *Evolution* **45**, 1618–1627.

Jokela, J., Dybdahl, M.F., Lively, C.M., 2009. The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *Am. Nat.* **174**, S43–S53.

Jokela, J., Lively, C.M., 1995. Spatial variation in infection by digenetic trematodes in a population of freshwater snails (*Potamopyrgus antipodarum*). *Oecologia* **103**, 509–517.

Jünemann, S., Prior, K., Szczepanowski, R., Harks, I., Ehmke, B., Goesmann, A., Stoye, J., Harmsen, D., 2012. Bacterial community shift in treated periodontitis patients revealed by ion torrent 16S rRNA gene amplicon sequencing. *PLoS One* **7**, e41606.

Katinka, M.D., Duprat, S., Cornillot, E., Méténier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretaillade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., Peyret, P., Saurin, W., Gouy, M., Weissenbach, J., Vivarès, C.P., 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453.

Kay, K.M., Whittall, J.B., Hodges, S.A., 2006. A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. *BMC Evol. Biol.* **6**, 36.

Keeling, P.J., Fast, N.M., Corradi, N., 2014. Microsporidian genome structure and function, in: Weiss, L.M., Becnel, J.J. (Eds.), *Microsporidia: Pathogens of opportunity.* John Wiley & Sons, Inc., pp. 221–229.

King, K.C., Delph, L.F., Jokela, J., Lively, C.M., 2009. The geographic mosaic of sex and the Red Queen. *Curr. Biol*. **19**, 1438–1441.

Knight, M., Miller, A.N., Patterson, C.N., Rowe, C.G., Michaels, G., Carr, D., Richards, C.S., Lewis, F.A., 1999. The identification of markers segregating with resistance to *Schistosoma mansoni* infection in the snail *Biomphalaria glabrata*. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1510–1515.

Köhsler, M., Leitner, B., Blaschitz, M., Michel, R., Aspöck, H., Walochnik, J., 2006. ITS1 sequence variabilities correlate with 18S rDNA sequence types in the genus *Acanthamoeba* (Protozoa: Amoebozoa). *Parasitol. Res.* **98**, 86–93.

Koop, J.A.H., DeMatteo, K.E., Parker, P.G., Whiteman, N.K., 2014. Birds are islands for parasites. *Biol. Lett.* **10**.

Koskella, B., Lively, C.M., 2009. Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution* **63**, 2213–2221.

Koskinen, K., Auvinen, P., Björkroth, K.J., Hultman, J., 2014. Inconsistent denoising and clustering algorithms for amplicon sequence data. *J. Comput. Biol.* **22**, 743–751.

Krebes, L., Zeidler, L., Frankowski, J., Bastrop, R., 2014. (Cryptic) sex in the microsporidian *Nosema granulosis* – Evidence from parasite rDNA and host mitochondrial DNA. *Infect. Genet. Evol.* **21**, 259–268.

Kwiatkowski, M., Engelstädter, J., Vorburger, C., 2012. On genetic specificity in symbiont-mediated host-parasite coevolution. *PLoS Comput. Biol.* **8**, e1002633.

Lafontaine, D.L., Tollervey, D., 2001. The function and synthesis of ribosomes. *Nat. Rev. Mol. Cell Biol.* **2**, 514–520.

Lawley, B., Ripley, S., Bridge, P., Convey, P., 2004. Molecular analysis of geographic patterns of eukaryotic diversity in Antarctic soils. *Appl. Environ. Microbiol.* **70**, 5963–5972.

Lee, S.C., Corradi, N., Byrnes, E.J., Torres-Martinez, S., Dietrich, F.S., Keeling, P.J., Heitman, J., 2008. Microsporidia evolved from ancestral sexual fungi. *Curr. Biol.* **18**, 1675–1679.

Lee, S.C., Corradi, N., Doan, S., Dietrich, F.S., Keeling, P.J., Heitman, J., 2010. Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. *PLoS One* **5**.

Lee, S.C., Heitman, J., Ironside, J.E., 2014. Sex and the Microsporidia, in: Weiss, L.M., Becnel, J.J. (Eds.), *Microsporidia: Pathogens of opportunity.* John Wiley & Sons, Inc., pp. 231–243.

Leiro, J., Ortega, M., Iglesias, R., Estévez, J., Sanmartín, M.L., 1996. *Pleistophora finisterrensis* n. sp., a microsporidian parasite of blue whiting *Micromesistius poutassou*. *Syst. Parasitol.* **34**, 163–170.

Li, W., Cama, V., Akinbo, F.O., Ganguly, S., Kiulia, N.M., Zhang, X., Xiao, L., 2013. Multilocus sequence typing of *Enterocytozoon bieneusi*: Lack of geographic segregation and existence of genetically isolated sub-populations. *Infect. Genet. Evol.* **14**, 111–119.

Li, W., Cama, V., Feng, Y., Gilman, R.H., Bern, C., Zhang, X., Xiao, L., 2012. Population genetic analysis of *Enterocytozoon bieneusi* in humans. *Int. J. Parasitol.* **42**, 287–293.

Liao, D., 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64**, 24–30.

Little, T.J., Ebert, D., 1999. Associations between parasitism and host genotype in natural populations of *Daphnia* (Crustacea: Cladocera). *J. Anim. Ecol.* **68**, 134–149.

Liu, H., Pan, G., Song, S., Xu, J., Li, T., Deng, Y., Zhou, Z., 2008. Multiple rDNA units distributed on all chromosomes of *Nosema bombycis*. *J. Invertebr. Pathol.* **99**, 235–238.

Lively, C.M., 2010. A review of red queen models for the persistence of obligate sexual reproduction. *J. Hered.* **101**, S13–S20.

Lloyd-MacGilp, S.A., Chambers, S.M., Dodd, J.C., Fitter, A.H., Walker, C., Young, J.P.W., 1996. Diversity of the ribosomal internal transcribed spacers within and among isolates of *Glomus mosseae* and related mycorrhizal fungi. *New Phytol.* **133**, 103–111.

Lohr, J.N., Laforsch, C., Koerner, H., Wolinska, J., 2010a. A *Daphnia* parasite (*Caullerya mesnili*) constitutes a new member of the ichthyosporea, a group of protists near the animal-fungi divergence. *J. Eukaryot. Microbiol.* **57**, 328–336.

Lohr, J.N., Yin, M., Wolinska, J., 2010b. Prior residency does not always pay off—co-infections in *Daphnia*. *Parasitology* **137**, 1493–1500.

Lollis, L., Gerhold, R., McDougald, L., Beckstead, R., 2011. Molecular characterization of *Histomonas meleagridis* and other parabasalids in the United States using the 5.8S, ITS-1, and ITS-2 rRNA regions. *J. Parasitol.* **97**, 610–615.

Lom, J., Dyková, I., Körting, W., Klinger, H., 1989. *Heterosporis schuberti* n.sp., a new microsporidian parasite of aquarium fish. *Eur. J. Protistol.* **25**, 129–135.

Margulies, M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Martín-Hernández, R., Meana, A., García-Palencia, P., Marín, P., Botías, C., Garrido-Bailón, E., Barrios, L., Higes, M., 2009. Effect of temperature on the biotic potential of honeybee microsporidia. *Appl. Environ. Microbiol.* **75**, 2554–2557.

Maynard Smith, J., 1978. *The Evolution of Sex.* Cambridge University Press. ISBN-13:978-0521293020

McManus, D.P., Bowles, J., 1996. Molecular genetic approaches to parasite identification: their value in diagnostic parasitology and systematics. *Int. J. Parasitol.* **26**, 687–704.

Mee, P.T., Weeks, A.R., Walker, P.J., Hoffmann, A.A., Duchemin, J.-B., 2015. Detection of low-level *Cardinium* and *Wolbachia* infections in *Culicoides*. *Appl. Environ. Microbiol.* **81**, 6177–6188.

Mendoza, L., Taylor, J.W., Ajello, L., 2002. The class mesomycetozoea: a heterogeneous group of microorganisms at the animal-fungal boundary. *Annu. Rev. Microbiol.* **56**, 315–344.

Mikheyev, A.S., Tin, M.M.Y., 2014. A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* **14**, 1097–1102.

Mooi, F.R., He, Q., van Oirschot, H., Mertsola, J., 1999. Variation in the *Bordetella pertussis* virulence factors pertussis toxin and pertactin in vaccine strains and clinical isolates in Finland. *Infect. Immun.* **67**, 3133–3134.

Morgulis, A., Gertz, E.M., Schäffer, A.A., Agarwala, R., 2006. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040.

Mosher, J.J., Bernberg, E.L., Shevchenko, O., Kan, J., Kaplan, L.A., 2013. Efficacy of a 3rd generation high-throughput sequencing platform for analyses of 16S rRNA genes from environmental samples. *J. Microbiol. Methods* **95**, 175–181.

Mosher, J.J., Bowman, B., Bernberg, E.L., Shevchenko, O., Kan, J., Korlach, J., Kaplan, L.A., 2014. Improved performance of the PacBio SMRT technology for 16S rDNA sequencing. *J. Microbiol. Methods* **104**, 59–60.

Nakamura, K., Oshima, T., Morimoto, T., Ikeda, S., Yoshikawa, H., Shiwa, Y., Ishikawa, S., Linak, M.C., Hirai, A., Takahashi, H., Altaf-Ul-Amin, M., Ogasawara, N., Kanaya, S., 2011. Se-

quence-specific error profile of Illumina sequencers. *Nucleic Acids Res.* **39**, e90.

Nei, M., Gu, X., Sitnikova, T., 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7799–7806.

Noë, R., Hammerstein, P., 1995. Biological markets. *Trends Ecol. Evol.* **10**, 336–339.

Ocaña-Mayorga, S., Llewellyn, M.S., Costales, J.A., Miles, M.A., Grijalva, M.J., 2010. Sex, subdivision, and domestic dispersal of *Trypanosoma cruzi* lineage I in southern Ecuador. *PLoS Negl. Trop. Dis.* **4**, e915.

Okamura, B., 1996. Occurrence, prevalence, and effects of the myxozoan *Tetracapsula bryozoides* parasitic in the freshwater bryozoan *Cristatella mucedo* (Bryozoa: Phylactolaemata). *Folia Parasitol.* **43**, 262–266.

Olival, K.J., Dick, C.W., Simmons, N.B., Morales, J.C., Melnick, D.J., Dittmar, K., Perkins, S.L., Daszak, P., Desalle, R., 2013. Lack of population genetic structure and host specificity in the bat fly, *Cyclopodia horsfieldi*, across species of *Pteropus* bats in Southeast Asia. *Parasit. Vectors* **6**, 231.

O'Mahony, E.M., Tay, W.T., Paxton, R.J., 2007. Multiple rRNA variants in a single spore of the microsporidian *Nosema bombi*. *J. Eukaryot. Microbiol.* **54**, 103–109.

Parniske, M., Hammond-Kosack, K.E., Golstein, C., Thomas, C.M., Jones, D.A., Harrison, K., Wulff, B.B., Jones, J.D., 1997. Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the *Cf-4/9* locus of tomato. *Cell* **91**, 821–832.

Pedersen, A.B., Altizer, S., Poss, M., Cunningham, A.A., Nunn, C.L., 2005. Patterns of host specificity and transmission among parasites of wild primates. *Int. J. Parasitol.* **35**, 647–657.

Pelin, A., Selman, M., Aris-Brosou, S., Farinelli, L., Corradi, N., 2015. Genome analyses suggest the presence of polyploidy and recent human-driven expansions in eight global populations of the honeybee pathogen *Nosema ceranae*. *Environ. Microbiol.* **17**, 4443–4458.

Penczykowski, R.M., Laine, A.-L., Koskella, B., 2015. Understanding the ecology and evolution of host-parasite interactions across scales. *Evol. Appl.* **9**, 37–52.

Perkins, S.L., 2000. Species concepts and malaria parasites: detecting a cryptic species of *Plasmodium*. *Proc. R. Soc. B Biol. Sci.* **267**, 2345–2350.

Peyretaillade, E., Biderre, C., Peyret, P., Duffieux, F., Méténier, G., Gouy, M., Michot, B., Vivarès, C.P., 1998. Microsporidian *Encephalitozoon cuniculi*, a unicellular eukaryote with an unusual chromosomal dispersion of ribosomal genes and a LSU rRNA reduced to the universal core. *Nucleic Acids Res.* **26**, 3513–3520.

Pillet, L., Fontaine, D., Pawlowski, J., 2012. Intra-genomic ribosomal RNA polymorphism and morphological variation in *Elphidium macellum* suggests inter-specific hybridization in foraminifera. *PLoS One* **7**, e32373.

Piry, S., Guivier, E., Realini, A., Martin, J.F., 2012. |SE|S|AM|E| Barcode: NGS-oriented software for amplicon characterization - application to species and environmental barcoding. *Mol. Ecol. Resour.* **12**, 1151–1157.

Pombert, J.F., Xu, J., Smith, D.R., Heiman, D., Young, S., Cuomo, C.A., Weiss, L.M., Keeling, P.J., 2013. Complete genome sequences from three genetically distinct strains reveal high intraspecies genetic diversity in the microsporidian *Encephalitozoon cuniculi*. *Eukaryot. Cell* **12**, 503–511.

Poulin, R., Randhawa, H.S., 2015. Evolution of parasitism along convergent lines: from ecology to genomics. *Parasitology* **142**, S6–S15.

Preheim, S.P., Perrotta, A.R., Martin-Platero, A.M., Gupta, A., Alm, E.J., 2013. Distribution-based clustering: using ecology to refine the operational taxonomic unit. *Appl. Environ. Microbiol.* **79**, 6593–6603.

Puente-Sánchez, F., Aguirre, J., Parro, V., 2015. A novel conceptual approach to read-filtering in high-throughput amplicon sequencing studies. *Nucleic Acids Res.* **44**, e40.

Quail, M., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics* **13**, 341.

Ramos-Onsins, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* **19**, 2092–2100.

Raso, G., Vounatsou, P., Singer, B.H., N'Goran, E.K., Tanner, M., Utzinger, J., 2006. An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 6934–6939.

Refardt, D., Canning, E.U., Mathis, A., Cheney, S.A., Lafranchi-Tristem, N.J., Ebert, D., 2002. Small subunit ribosomal DNA phylogeny of microsporidia that infect *Daphnia* (Crustacea: Cladocera). *Parasitology* **124**, 381–389.

Refardt, D., Decaestecker, E., Johnson, P.T.J., Vávra, J., 2008. Morphology, molecular phylogeny, and ecology of *Binucleata daphniae* n. g., n. sp. (Fungi: Microsporidia), a parasite of *Daphnia magna* Straus, 1820 (Crustacea: Branchiopoda). *J. Eukaryot. Microbiol.* **55**, 393–408.

Reuter, M., Lehmann, L., Guillaume, F., 2008. The spread of incompatibility-inducing parasites in sub-divided host populations. *BMC Evol. Biol.* **8**, 134.

Robideau, G.P., De Cock, A.W.A.M., Coffey, M.D., Voglmayr, H., Brouwer, H., Bala, K., Chitty, D.W., Désaulniers, N., Eggertson, Q.A., Gachon, C.M.M., Hu, C.-H., Küpper, F.C., Rintoul, T.L., Sarhan, E., Verstappen, E.C.P., Zhang, Y., Bonants, P.J.M., Ristaino, J.B., Lévesque, C.A., 2011. DNA barcoding of oomycetes with cytochrome *c* oxidase subunit I and internal transcribed spacer. *Mol. Ecol. Resour.* **11**, 1002–1011.

Roth, A., Fischer, M., Hamid, M.E., Michalke, S., Ludwig, W., Mauch, H., 1998. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. *J. Clin. Microbiol.* **36**, 139–147.

Roudel, M., Aufauvre, J., Corbara, B., Delbac, F., Blot, N., 2013. New insights on the genetic diversity of the honeybee parasite *Nosema ceranae* based on multilocus sequence analysis. *Parasitology* **140**, 1346–1356.

Rushworth, C.A., Song, B.-H., Lee, C.-R., Mitchell-Olds, T., 2011. *Boechera*, a model system for ecological genomics. *Mol. Ecol.* **20**, 4843–4857.

Sachs, J.L., Simms, E.L., 2006. Pathways to mutualism breakdown. *Trends Ecol. Evol.* **21**, 585–592.

Sadd, B.M., Schmid-Hempel, P., 2009. Principles of ecological immunology. *Evol. Appl.* **2**, 113–121.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Sasal, P., Trouve, S., Muller-Graf, C., Morand, S., 1999. Specificity and host predictability: a comparative analysis among monogenean parasites of fish. *J. Anim. Ecol.* **68**, 437–444.

Schatz, D.G., 2004. V(D)J recombination. *Immunol. Rev.* **200**, 5–11.

Schloss, P.D., Westcott, S.L., Jenior, M.L., Highlander, S.K., 2016. Sequencing 16S rRNA gene fragments using the PacBio SMRT DNA sequencing system. *PeerJ PrePrints* **4**, e778v2.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing *mothur*: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541.

Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., Schmitt, I., 2013. Illumina metabarcoding of a soil fungal community. *Soil Biol. Biochem.* **65**, 128–132.

Schneider, D.I., Klasson, L., Lind, A.E., Miller, W.J., 2014. More than fishing in the dark: PCR of a dispersed sequence produces simple but ultrasensitive *Wolbachia* detection. *BMC Microbiol.* **14**, 121.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Bolchacova, E., Voigt, K., Crous, P.W., Miller, A.N., Wingfield, M.J., Aime, M.C., An, K.-D., Bai, F.-Y., Barreto, R.W., Begerow, D., Bergeron, M.-J., Blackwell, M., Boekhout, T., Bogale, M., Boonyuen, N., Burgaz, A.R., Buyck, B., Cai, L., Cai, Q., Cardinali, G., Chaverri, P., Coppins, B.J., Crespo, A., Cubas, P., Cummings, C., Damm, U., de Beer, Z.W., de Hoog, G.S., Del-Prado, R., Dentinger, B., Dieguez-Uribeondo, J., Divakar, P.K., Douglas, B., Duenas, M., Duong, T.A., Eberhardt, U., Edwards, J.E., Elshahed, M.S., Fliegerova, K., Furtado, M., Garcia, M.A., Ge, Z.-W., Griffith, G.W., Griffiths, K., Groenewald, J.Z., Groenewald, M., Grube, M., Gryzenhout, M., Guo, L.-D., Hagen, F., Hambleton, S., Hamelin, R.C., Hansen, K., Harrold, P., Heller, G., Herrera, C., Hirayama, K., Hirooka, Y., Ho, H.-M., Hoffmann, K., Hofstetter, V., Hognabba, F., Hollingsworth, P.M., Hong, S.-B., Hosaka, K., Houbraken, J., Hughes, K., Huhtinen, S., Hyde, K.D., James, T., Johnson, E.M., Johnson, J.E., Johnston, P.R., Jones, E.B.G., Kelly, L.J., Kirk, P.M., Knapp, D.G., Koljalg, U., Kovacs, G.M., Kurtzman, C.P., Landvik, S., Leavitt, S.D., Liggenstoffer, A.S., Liimatainen, K., Lombard, L., Luangsa-ard, J.J., Lumbsch, H.T., Maganti, H., Maharachchikumbura, S.S.N., Martin, M.P., May, T.W., McTaggart, A.R., Methven, A.S., Meyer, W., Moncalvo, J.-M., Mongkolsamrit, S., Nagy, L.G., Nilsson, R.H., Niskanen, T., Nyilasi, I., Okada, G., Okane, I., Olariaga, I., Otte, J., Papp, T., Park, D., Petkovits, T., Pino-Bodas, R., Quaedvlieg, W., Raja, H.A., Redecker, D., Rintoul,

T.L., Ruibal, C., Sarmiento-Ramirez, J.M., Schmitt, I., Schussler, A., Shearer, C., Sotome, K., Stefani, F.O.P., Stenroos, S., Stielow, B., Stockinger, H., Suetrong, S., Suh, S.-O., Sung, G.-H., Suzuki, M., Tanaka, K., Tedersoo, L., Telleria, M.T., Tretter, E., Untereiner, W.A., Urbina, H., Vagvolgyi, C., Vialle, A., Vu, T.D., Walther, G., Wang, Q.-M., Wang, Y., Weir, B.S., Weiss, M., White, M.M., Xu, J., Yahr, R., Yang, Z.L., Yurkov, A., Zamora, J.-C., Zhang, N., Zhuang, W.-Y., Schindel, D., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 6241–6246.

Schoebel, C.N., Tellenbach, C., Spaak, P., Wolinska, J., 2011. Temperature effects on parasite prevalence in a natural hybrid complex. *Biol. Lett.* **7**, 108–111.

Schoebel, C.N., Wolinska, J., Spaak, P., 2010. Higher parasite resistance in *Daphnia* populations with recent epidemics. *J. Evol. Biol.* **23**, 2370–2376.

Schulte, R.D., Makus, C., Hasert, B., Michiels, N.K., Schulenburg, H., 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7359–7364.

Scorzetti, G., Fell, J.W., Fonseca, A., Statzell-Tallman, A., 2002. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. *FEMS Yeast Res.* **2**, 495–517.

Selman, M., Sak, B., Kváč, M., Farinelli, L., Weiss, L.M., Corradi, N., 2013. Extremely reduced levels of heterozygosity in the vertebrate pathogen *Encephalitozoon cuniculi*. *Eukaryot. Cell* **12**, 496–502.

Siemens, D.H., Roy, B.A., 2005. Tests for parasite-mediated frequency-dependent selection in natural populations of an asexual plant species. *Evol. Ecol.* **19**, 321–338.

Simakova, A. V, Pankova, T.F., Tokarev, Y.S., Issi, I. V, 2005. *Senoma* gen. n., a new genus of microsporidia, with the type species *Senoma globulifera* comb. n.(syn. *Issia globulifera* Issi et Pankova 1983) from the malaria mosquito *Anopheles messeae* Fall. *Protistology* **4**, 135–144.

Slamovits, C.H., Fast, N.M., Law, J.S., Keeling, P.J., 2004. Genome compaction and stability in microsporidian intracellular parasites. *Curr. Biol.* **14**, 891–896.

Slatkin, M., 1987. Gene flow and the geographic structure of natural populations. *Science* **236**, 787–792.

Sommer, S., Courtiol, A., Mazzoni, C.J., 2013. MHC genotyping of non-model organisms using next-generation sequencing: a new methodology to deal with artefacts and allelic dropout. *BMC Genomics* **14**, 542.

Stern, R.F., Andersen, R.A., Jameson, I., Küpper, F.C., Coffroth, M.-A., Vaulot, D., Le Gall, F., Véron, B., Brand, J.J., Skelton, H., Kasai, F., Lilly, E.L., Keeling, P.J., 2012. Evaluating the ribosomal internal transcribed spacer (ITS) as a candidate dinoflagellate barcode marker. *PLoS One* **7**, e42780.

Stollewerk, A., 2010. The water flea *Daphnia*—a 'new' model system for ecology and evolution? *J. Biol.* **9**, 21.

Suga, H., Ruiz-Trillo, I., 2015. Filastereans and ichthyosporeans: Models to understand the origin of metazoan multicellularity, in: Ruiz-Trillo, I., Nedelcu, A.M. (Eds.), *Evolutionary transitions to multicellular life.* Springer Netherlands, Dordrecht, pp. 117–128.

Suh, Y., Thien, L.B., Reeve, H.E., Zimmer, E.A., 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *Am. J. Bot.* **80**, 1042–1055.

Sweeney, A.W., Doggett, S.L., Piper, R.G., 1990. Host specificity studies of *Amblyospora indicola* and *Amblyospora dyxenoides* (Microspora: Amblyosporidae) in mosquitoes and copepods. *J. Invertebr. Pathol.* **56**, 415–418.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tanabe, K., Jombart, T., Horibe, S., Palacpac, N.M.Q., Honma, H., Tachibana, S.-I., Nakamura, M., Horii, T., Kishino, H., Mita, T., 2013. *Plasmodium falciparum* mitochondrial genetic diversity exhibits isolation-by-distance patterns supporting a sub-Saharan African origin. *Mitochondrion* **13**, 630–636.

Tay, W.T., O'Mahony, E.M., Paxton, R.J., 2005. Complete rRNA gene sequences reveal that the microsporidium *Nosema bombi* infects diverse bumblebee (*Bombus* spp.) hosts and contains multiple polymorphic sites. *J. Eukaryot. Microbiol.* **52**, 505–513.

Taylor, M.J., Hoerauf, A., 1999. *Wolbachia* bacteria of filarial nematodes. *Parasitol. Today* **15**, 437–442.

Tellenbach, C., Tardent, N., Pomati, F., Keller, E., Hairston, N., Wolinska, J., Spaak, P. (2016) Cyanobacteria facilitate parasite epidemics in *Daphnia*. Submitted to *Ecology.*

Templeton, A.R., Crandall, K. a, Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**, 619–633.

Thompson, J.N., Burdon, J.J., 1992. Gene-for-gene coevolution between plants and parasites. *Nature* **360**, 121–125.

Thrall, P.H., Burdon, J.J., 2003. Evolution of virulence in a plant host-pathogen metapopulation. *Science* **299**, 1735–1737.

Thrall, P.H., Burdon, J.J., 1997. Host-pathogen dynamics in a metapopulation context: the ecological and evolutionary consequences of being spatial. *J. Ecol.* **85**, 743–753.

Tripet, F., Christe, P., Møller, A.P., 2002. The importance of host spatial distribution for parasite specialization and speciation: A comparative study of bird fleas (Siphonaptera: Ceratophyllidae). *J. Anim. Ecol.* **71**, 735–748.

Tsai, S.-J., Huang, W.-F., Wang, C.-H., 2005. Complete sequence and gene organization of the *Nosema spodopterae* rRNA gene. *J. Eukaryot. Microbiol.* **52**, 52–54.

Vagelli, A., Paramá, A., Sanmartín, M.L., Leiro, J., 2005. *Glugea vincentiae* n. sp. (Microsporidia: Glugeidae) infecting the Australian marine fish *Vincentia conspersa* (Teleostei: Apogonidae). *J. Parasitol.* **91**, 152–157.

van Hannen, E.J., Mooij, W., van Agterveld, M.P., Gons, H.J., Laanbroek, H.J., 1999. Detritus-dependent development of the microbial community in an experimental system: Qualitative analysis by denaturing gradient gel electrophoresis. *Appl. Envir. Microbiol.* **65**, 2478–2484.

Van Valen, L., 1973. A new evolutionary law. *Evol. Theory* **1**, 1–30.

Vávra, J., Hyliš, M., Oborník, M., Vossbrinck, C.R., 2005. Microsporidia in aquatic microcrustacea: The copepod microsporidium *Marssoniella elegans* Lemmermann, 1900 revisited. *Folia Parasitol.* **52**, 163–172.

Vávra, J., Larsson, J.I.R., 1994. *Berwaldia schaefernai* (Jírovec, 1937) comb. n. (protozoa, microsporida), fine structure, life cycle, and relationship to *Berwaldia singularis* Larsson, 1981. *Eur. J. Protistol.* **30**, 45–54.

Vernon, J.G., Okamura, B., Jones, C.S., Noble, L.R., 1996. Temporal patterns of clonality and parasitism in a population of freshwater bryozoans. *Proc. R. Soc. B Biol. Sci.* **263**, 1313–1318.

Vilas, R., Criscione, C.D., Blouin, M.S., 2005. A comparison between mitochondrial DNA and the ribosomal internal transcribed regions in prospecting for cryptic species of platyhelminth parasites. *Parasitology* **131**, 839–846.

Vossbrinck, C.R., Andreadis, T.G., Vavra, J., Becnel, J.J., 2004. Molecular phylogeny and evolution of mosquito parasitic microsporidia (Microsporidia: Amblyosporidae). *J. Eukaryot. Microbiol.* **51**, 88–95.

Vossbrinck, C.R., Debrunner-Vossbrinck, B.A., 2005. Molecular phylogeny of the Microsporidia: ecological, ultrastructural and taxonomic considerations. *Folia Parasitol.* **52**, 131–142.

Wang, L.-L., Chen, K.-P., Zhang, Z., Yao, Q., Gao, G.-T., Zhao, Y., 2006. Phylogenetic analysis of *Nosema antheraeae* (Microsporidia) isolated from Chinese oak silkworm, *Antheraea pernyi*. *J. Eukaryot. Microbiol.* **53**, 310–313.

Wang, X., Yao, J., Sun, Y., Mai, V., 2013. M-pick, a modularity-based method for OTU picking of 16S rRNA sequences. *BMC Bioinformatics* **14**, 43.

Weigl, S., Körner, H., Petrusek, A., Seda, J., Wolinska, J., 2012. Natural distribution and co-infection patterns of microsporidia parasites in the *Daphnia longispina* complex. *Parasitology* **139**, 870–880.

Werren, J.H., 1987. The coevolution of autosomal and cytoplasmic sex ratio factors. *J. Theor. Biol.* **124**, 317–334.

Werren, J.H., Windsor, D., Guo, L., 1995. Distribution of *Wolbachia* among neotropical arthropods. *Proc. R. Soc. London B-Biological Sci.* **262**, 197–204.

Wilkinson, T.J., Rock, J., Whiteley, N.M., Ovcharenko, M.O., Ironside, J.E., 2011. Genetic diversity of the feminising microsporidian parasite *Dictyocoela*: New insights into host-specificity, sex and phylogeography. *Int. J. Parasitol.* **41**, 959–966.

Wilson, K., Grenfell, B.T., Pilkington, J.G., Boyd, H.E.G., Gulland, F.M.D., 2004. Parasites and their impact., in: Clutton-Brock, T.H., Pemberton, J.M. (Eds.), *Soay Sheep: Dynamics and Selection in an Island Population.* Cambridge University Press, Cambridge, UK, pp. 113–165.

Wirawan, A., Harris, R.S., Liu, Y., Schmidt, B., Schröder, J., 2014. HECTOR: a parallel multistage homopolymer spectrum based error corrector for 454 sequencing data. *BMC Bioinformatics* **15**, 131.

Wolinska, J., Bittner, K., Ebert, D., Spaak, P., 2006. The coexistence of hybrid and parental *Daphnia*: the role of parasites. *Proc. R. Soc. B Biol. Sci.* **273**, 1977–1183.

Wolinska, J., Giessler, S., Koerner, H., 2009. Molecular identification and hidden diversity of novel *Daphnia* parasites from European lakes. *Appl. Environ. Microbiol.* **75**, 7051–7059.

Wolinska, J., Keller, B., Manca, M., Spaak, P., 2007. Parasite survey of a *Daphnia* hybrid complex: Host-specificity and environment determine infection. *J. Anim. Ecol.* **76**, 191–200.

Wolinska, J., King, K.C., 2009. Environment can alter selection in host-parasite interactions. *Trends Parasitol.* **25**, 236–244.

Wolinska, J., Petrusek, A., Yin, M., Koerner, H., Seda, J., Giessler, S., 2014. Population structure of a microparasite infecting *Daphnia*: spatio-temporal dynamics. *BMC Evol. Biol.* **14**, 247.

Wolinska, J., Seda, J., Koerner, H., Smilauer, P., Petrusek, A., 2011a. Spatial variation of *Daphnia* parasite load within individual water bodies. *J. Plankton Res.* **33**, 1284–1294.

Wolinska, J., Spaak, P., 2009. The cost of being common: Evidence from natural *Daphnia* populations. *Evolution* **63**, 1893–1901.

Wolinska, J., Spaak, P., Koerner, H., Petrusek, A., Seda, J., Giessler, S., 2011b. Transmission mode affects the population genetic structure of *Daphnia* parasites. *J. Evol. Biol.* **24**, 265–273.

Xiang, Z., Censini, S., Bayeli, P., Telford, J., Figura, N., Rappuoli, R., Covacci, A., 1995. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun.* **63**, 94–98.

Zhang, J.-R., Hardham, J.M., Barbour, A.G., Norris, S.J., 1997. Antigenic variation in Lyme disease borreliae by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**, 275–285.

Zuk, M., Stoehr, A.M., 2002. Immune defense and host life history. *Am. Nat.* **160**, S9–S22.

Authors contributions

Chapter 2:

J. Wolinska applied for funding and coordinated this project. J. Wolinska and A. Petrusek designed the study. J. Rusek designed sequencing primers and established the PCR protocol, while A. Petrusek prepared and performed the pyrosequencing reactions. E. Gonzalez-Tortuero developed the QRS pipeline. D. Lyras, S. Grath and F. Castro-Monzon contributed to the development of QRS pipeline in earlier stages. E. Gonzalez-Tortuero, with the help of S. Gießler, performed statistical analyses and, together with J. Wolinska, S. Gießler and A. Petrusek, wrote the manuscript. D. Lyras and S. Grath helped with writing the Supporting Information. All authors approved the final version of the manuscript and Supporting Information.

Chapter 3:

J. Wolinska and P. Spaak applied for funding and coordinated this project. J. Wolinska, P. Spaak and S. Gießler designed the study. C. Tellenbach contributed to sampling. P. Turko performed microsatellite analyses of *Daphnia*. J. Rusek, with the help of I. Maayan, prepared *Caullerya* samples for 454 pyrosequencing. L. Pialek and A. Petrusek designed, and L. Pialek performed the pyrosequencing reactions. E. Gonzalez-Tortuero performed all bioinformatics and statistical analyses. E. Gonzalez-Tortuero and J. Wolinska wrote the manuscript, with comments and editing by A. Petrusek, P. Spaak, C. Tellenbach and I. Maayan. All authors approved the final version of the manuscript and Supporting Information.

Chapter 4:

J. Wolinska applied for funding and coordinated this project. J. Wolinska and J. Rusek designed the study. A. Petrusek contributed to sampling. J. Rusek designed sequencing primers, and J. Rusek and I. Maayan established and carried out the PCR protocol. L. Pialek and A. Petrusek designed, and L. Pialek performed the pyrosequencing reactions. E. Gonzalez-Tortuero performed all bioinformatics and statistical analyses. S. Laurent contributed to the interpretation of the genetic tests' results. E. Gonzalez-Tortuero and J. Wolinska wrote the manuscript, with comments and editing by A. Petrusek and I. Maayan. All authors approved the final version of the manuscript and Supporting Information.

Curriculum vitae

Enrique González Tortuero, PhD candidate

Personal Data

Address: Motzstr. 43, 10777 Berlin, Germany
Email: gonzalez@igb-berlin.de or enriquegleztortuero@gmail.com

Date and Place of Birth: 08/12/1986, Guadalajara, Spain
Citizenship: Spanish

Education

01/2013 – Present: PhD (Candidate) in Biology
Ludwig-Maximilians-Universität (LMU), Munich, Germany
Thesis title: "Host-parasite dynamics in a natural system: revealing the evolutionary change in parasite populations infecting *Daphnia*"
Supervisor: Justyna Wolinska

10/2010 – 09/2011: MSc in Evolutionary Biology
Complutense University of Madrid (UCM), Madrid, Spain
Thesis title: "Transmission rate of *Wolbachia* (Rickettsiales) in Pyrenean populations of *Chorthippus parallelus* (Orthoptera: Acrididae)"
Supervisor: José L. Bella

09/2004 – 06/2010: Licenciatura en Biología (BSc in Biology)
University of Alcala (UAH), Alcalá de Henares, Madrid, Spain

Research experience

03/2014 – Present: Berlin Centre for Genomics in Biodiversity Research, Berlin, Germany
03/2014 – Present: Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
01/2013 – 02/2014: Ludwig-Maximilians-University, Munich, Germany
09/2011 – 12/2012: Astrobiology Centre (CAB) – National Institute for Aerospace Technology (INTA), Torrejón de Ardoz, Madrid, Spain
10/2010 – 08/2011: Department of Plant Biology, UAH, Alcalá de Henares, Madrid, Spain
10/2009 – 04/2010: Department of Ecology, UAH, Alcalá de Henares, Madrid, Spain
10/2009 – 01/2010: Department of Geology, UAH, Alcalá de Henares, Madrid, Spain
07/2009 – 09/2009: Museo Nacional de Ciencias Naturales – Spanish National Research Council (MNCN – CSIC), Madrid, Spain

Courses

- Course "Python from Scratch (for high-throughput sequencing applications)", Berlin Centre for Genomics in Biodiversity Research, Berlin, Germany, 05/2014

- Course "Python 101: by biologists, for biologists", University of Lisbon, Lisbon, Portugal, 05/2012.
- Course "Genomics of Prokaryotes", Miguel Hernandez University, Sant Joan d'Alacant, Alicante, Spain, 12/2011.
- Course of Bioinformatics: Introduction to sequence analysis, Technical University of Madrid, Madrid, Spain, 09/2011.
- Course on Molecular Evolution, Phylogenetics, Phylogenomics and Adaptation [MEPPA] (6th Edition), Prince Felipe Research Centre, Valencia, Spain, 05/2011 - 06/2011.
- Theoretical-practical course "Basic techniques on Molecular Genetics", UCM, Madrid, Spain, 05/2010.

Memberships

2013 – Present: Present: European Society of Evolutionary Biology (ESEB)
2013 – Present: Sociedad de Científicos Españoles en la República Federal de Alemania (Society of Spanish Researchers in Germany; CERFA)
2012: Sociedad Española de Bioquímica y Biología Molecular (Molecular Biology and Biochemistry Spanish Society; SEBBM)
2010 – 2012: Colegio Oficial de Biólogos de la Comunidad de Madrid (COBCM)

Conferences

Oral talks

- **González-Tortuero E**, Rusek J, Turko P, Petrusek A, Maayan I, Piálek L, Tellenbach C, Gießler S, Spaak P, Wolinska J (2016) *The dynamics of* Daphnia *parasite genotypes are indicative for negative frequency-dependent selection, as observed across multiple epidemics in a Central European lake.* 5th Symposium of the DFG-SPP 1399. Münster, Germany, 17 – 20 May.
- **González-Tortuero E**, Rusek J, Petrusek A, Gießler S, Lyras D, Grath S, Castro-Monzón F, Wolinska J (2014) *Verification of quantitative results obtained by Next Generation Sequencing: case study on ITS1 sequence variation in a microparasite infecting* Daphnia*.* 4th Symposium of the DFG-SPP 1399. Noer, Kiel, Germany, 29 September – 2 October.
- **Gonzalez-Tortuero E** (2013) *Population structure of a microparasite infecting* Daphnia*: spatio-temporal dynamics.* 3rd Symposium of the DFG - SPP 1399. Blossin, Berlin, Germany, 29 August – 2 September.

Posters

- **Gonzalez-Tortuero E**, Rusek J, Petrusek A, Giessler S, Lyras D, Grath S, Castro-Monzon F, Wolinska J (2015) *The QRS (Quantification of Representative Sequences) pipeline for amplicon sequencing.* XV Congress of the European Society of Evolutionary Biology, Lausanne, Switzerland, 9 – 14 August.
- **Gonzalez-Tortuero E**, Rusek J, Giessler S, Petrusek A, Wolinska J (2013) *Bioinformatic analysis of amplicon sequencing data to study spatial and temporal variation in a* Daphnia *microparasite.* 106th Annual Meeting of the German Zoological Society (DZG), Munich, Germany, 13 – 16 September.

- **Gonzalez-Tortuero E**, Rusek J, Giessler S, Petrusek A, Wolinska J (2013) *Bioinformatic analysis of amplicon sequencing data to study spatial and temporal variation in a* Daphnia *microparasite.* XIV Congress of the European Society of Evolutionary Biology, Lisbon, Portugal, 19 – 24 August.
- **Gonzalez-Tortuero E**, Aranda-Martin M, Lopez de Saro FJ, Gomez MJ (2012) *Pfam-based survey of prokaryotic transposases in completely sequenced genomes.* 22nd International Union of Biochemistry and Molecular Biology (IUBMB) and 37th Federation of European Biochemistral Societies (FEBS) congresses, Seville, Spain, 4 – 9 September.
- **Gonzalez-Tortuero E**, Zabal-Aguirre M, Martinez-Rodriguez P, Bella JL (2011) Wolbachia *and the Pyrenean* Chorthippus parallelus *hybrid zone.* Course on Molecular Evolution, Phylogenetics, Phylogenomics and Adaptation (MEPPA; 6th edition), Valencia, Spain, 30 May – 3 June.
- Alonso A, **Gonzalez-Tortuero E**, Camargo JA, Castro Diez P (2010) *Desiccation tolerance of the exotic aquatic snail* Potamopyrgus antipodarum *(Hydrobiidae, Mollusca) as a spreading mechanism.* 6th NEOBIOTA conference, Copenhagen (Denmark), 14 – 17 September.
- Menor Salazar C, Ruiz Zapata B, **Gonzalez-Tortuero E**, Gil Garcia MJ, Herraez Sanchez de las Matas I (2010) *Análisis polínico de la Turbera de Cantos (Palencia).* XVII International Symposium APLE (Asociación de Palinologos de la Lengua Española), Ourense, Spain, 7 – 10 July.

Reviewer duties

Bioinformatics, PLOS Neglected Tropical Diseases

Peer-reviewed publications

- **González-Tortuero E**, Rusek J, Turko P, Petrusek A, Maayan I, Pialek L, Tellenbach C, Gießler S, Spaak P, Wolinska J (in press) *Daphnia* parasite dynamics across multiple *Caullerya* epidemics indicate selection against common parasite genotypes. *Zoology*.
- **González-Tortuero E**, Rusek J, Maayan I, Petrusek A, Pialek L, Laurent S, Wolinska J (2016) Genetic diversity of two *Daphnia*-infecting microsporidian parasites, based on sequence variation in the internal transcribed spacer region. *Parasit Vectors*, 9: 293.
- **González-Tortuero E**, Rusek J, Petrusek A, Gießler S, Lyras D, Grath S, Castro-Monzón F, Wolinska J (2015) The Quantification of Representative Sequences pipeline for amplicon sequencing: case study on within-population ITS1 sequence variation in a microparasite infecting *Daphnia*. *Mol Ecol Resour*, 15: 1385 – 1395.
- Gómez MJ, Díaz-Maldonado H, **González-Tortuero E**, López de Saro FJ (2014) Chromosomal replication dynamics and interaction with the β sliding clamp determine orientation of bacterial transposable elements. *Genome Biol Evol*, 6: 727 – 740.
- **Gonzalez-Tortuero E**, Martinez Perez FD (2010) Consecuencias evolutivas y biológicas causadas por bacterias del genero *Wolbachia* en artrópodos. *Bol SEA* 36: 189 – 202. (In Spanish)

Book chapters

- López de Saro FJ, Gomez-Rodriguez MJ, **Gonzalez-Tortuero E**, Parro-Garcia V (2013) The dynamic genomes of acidophiles. In Seckbach J, Oren A, Stan-Lotter H. *Polyextremophiles: life under multiple forms of stress.* Springer Dordrecht Heidelberg Verlag, NY (USA), 81 – 97 pp.

Papers in preparation

- **González-Tortuero E**, López de Saro FJ (in prep) Mobile elements and antibiotic resistance.
- **González-Tortuero E**, Rolff J, Rodriguez-Rojas A (in prep) On amino acid composition and protein oxidation proneness in *Escherichia coli*: an spatial gradient.
- **González-Tortuero E**, Wolinska J, Rolff J, Rodriguez-Rojas A (in prep) Triphosphate nucleotide transport by bacteria is constrained by oxidative environment.
- **González-Tortuero E**, Rodriguez-Rojas A (in prep) The ATP/ADP carrier equilibrated the relationship between the proto-eukaryotic cell and proto-mitochondrion.

Statutory declaration and statement

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 28.04.2016 Enrique González Tortuero

Erklärung

Hiermit erkläre ich, *

■ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.

■ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.

☐ dass ich mich mit Erfolg der Doktorprüfing im Hauptfach _____ und in den Nebenfächern _____ bei der Fäkultat für _____ der _____ (Hochschule/Universität) unterzogen habe.

☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den 28.04.2016 Enrique González Tortuero

*) Nichtzutreffendes streichen