Original Article (Primary Research Paper)

Title

Demographic history shapes genomic variation in an intracellular parasite with
a wide geographic distribution

Running Head

Determinants of genomic variation in *H. tvaer*

Authors

Pascal Angst*, Dieter Ebert, and Peter D. Fields

Department of Environmental Sciences, Zoology, University of Basel, Vesalgasse 1, 4051 Basel,
Switzerland

* Corresponding author: pascal.angst@unibas.ch

OrcIDs

Pascal Angst: 0000-0002-8654-2251

Peter D. Fields: 0000-0003-2959-2524

Dieter Ebert: 0000-0003-2653-3772

Abstract

Analyzing variation in a species' genomic diversity can provide insights into its historical demography, biogeography and population structure, and thus, its ecology and evolution. Although such studies are rarely undertaken for parasites, they can be highly revealing because of the parasite's coevolutionary relationships with hosts. Modes of reproduction and transmission are thought to be strong determinants of genomic diversity for parasites and vary widely among microsporidia (fungal-related intracellular parasites), which are known to have high intraspecific genetic diversity and interspecific variation in genome architecture. Here we explore genomic variation in the microsporidium *Hamiltosporidium*, a parasite of the freshwater crustacean *Daphnia magna*, looking especially at which factors contribute to nucleotide variation. Genomic samples from 18 Eurasian populations and a new, long-read based reference genome were used to determine the roles that reproduction mode, transmission mode and geography play in determining population structure and demographic history. We demonstrate two main *H. tvaerminnensis* lineages and a pattern of isolation-by-distance, but note an absence of congruence between these two parasite lineages and the two Eurasian host lineages. We suggest a comparatively recent parasite spread through Northern Eurasian host populations after a change from vertical to mixed-mode transmission and the loss of sexual reproduction. While gaining knowledge about the ecology and evolution of this focal parasite, we also identify common features that shape variation in genomic diversity for many parasites, e.g., distinct modes of reproduction and the intertwining of host–parasite demographies.

Keywords

Population genomics, range expansion, microsporidia, fungi, *Daphnia*, Cladocera

Introduction

Uncovering the ecological and evolutionary forces that drive variation in genetic diversity within and among populations across a species' range is important for understanding biological processes of basic and applied interest. Factors such as historical conditions (e.g., phylogeography and population structure), environmental conditions (e.g., adaptation to local climate), and stochastic processes (e.g., genetic drift and founder events) may contribute to patterns of genetic variation in a species (Bellis et al., 2020; Levicoy et al., 2021; Sendell-Price et al., 2021). For parasites, such studies take on an additional level of complexity, as parasite evolution is influenced by its dependence on the host and by issues such as host demography, parasite epidemiology, and coevolutionary interactions between hosts and parasites (Ebert & Fields, 2020). While a great deal of research has examined the effect of parasitism on host evolution, we know little about the factors that drive changes in the distribution of parasite genomic diversity, such as parasite life history, host shifts, mode of transmission and mode of reproduction (Janssen et al., 2004; Nieberding et al., 2008; Sandrock et al., 2011). The phylogeography of human parasites, which has been successfully studied to illuminate genetic epidemiology and variation in several systems, has proved to be a powerful tool in understanding, for example, *Mycobacterium tuberculosis* and viruses like the recently emerged coronavirus SARS-CoV-2 (Dellicour et al., 2021; Gagneux & Small, 2007). For the Gram-negative bacterium *Helicobacter pylori*, a known risk factor for stomach cancer in humans, the link between host and parasite biogeography and the parasite's mode of transmission was shown to be an important factor in the evolution of the system (Montano et al., 2015): similarly, *H. pylori*, like its human host, shows a continuous loss of genetic diversity with increasing geographic distance from East Africa. Its vertical mode of transmission most likely contributed to the geographical distribution of genetic diversity shared between the parasite and its host. The interplay between co-phylogeography and mode of transmission here highlights the importance of studying interactions between factors that drive genetic diversity when trying to understand host and parasite evolution.

In this article, we present a population genomic study of a parasitic microsporidium, focusing on the distribution of genomic diversity and relating it to key factors of parasite biology.

Genetic variation in parasites and their hosts may be influenced by the same or different factors; moreover they might influence each other's evolution in both specific and general ways (Langerhans, 2008). In *H. pylori*, for example, the parasite's genomic signature was so strong that it could be used to infer the history of the human host across the Pacific islands (Moodley et al., 2009). However, investigating interrelationships between multiple factors such as mode of transmission and biogeography in *H. pylori* is challenging, as both host and parasite need to be considered; therefore, such studies are still rare in parasite evolutionary biology.

One taxon that has recently gained attention as a potential model for parasite evolution is the microsporidia, due to its high variation in many aspects relevant to evolutionary and ecological theory (Murareanu et al., 2021; Wadi & Reinke, 2020). Microsporidia are obligate intracellular parasites with a phylogenetic position close to the fungi (James et al., 2013), although they show many highly derived features: for example, they have no mitochondria, have fewer genes and a very small genome size compared to other fungi, including the smallest known eukaryote genomes (Corradi, 2015). However, the latter two features are highly variable among microsporidia (Wadi & Reinke, 2020; Williams et al., 2008). Not only do microsporidia genomes vary tremendously in length and in gene numbers, but de-novo genome comparisons of microsporidia also found high levels of intraspecific genomic diversity (Pelin et al., 2015; Pombert et al., 2013). Moreover, microsporidia show high variation in their modes of reproduction and transmission, with ~18 % of species being thought to transmit vertically (Murareanu et al., 2021). Because disentangling and understanding the factors that drive genomic diversity may be easier in such a fast-evolving clade, microsporidia present a powerful model clade for evolutionary biology research.

The freshwater planktonic crustacean *Daphnia magna* is a well-established model system in ecology and evolution, including studies of host–parasite interactions (Altermatt & Ebert, 2008; Ebert, 2008; Orlansky & Ben-Ami, 2019). *D. magna* is parasitized by different species of microsporidia (Ebert, 2005), including the microsporidian genus *Hamiltosporidium* (Ebert, 2005; Haag et al., 2020; Pombert et al., 2015), of which two species, *H. magnivora* and *H. tvaerminnensis*, infect *D. magna* exclusively (Haag et al., 2011). Host and parasites are widespread across Eurasia. Both *Hamiltosporidium* species have, as compared to other microsporidia, relatively large, gene-sparse genomes, a feature that has been linked to a high number of repetitive, transposable elements (Parisot et al., 2014) that may have spread in the *Hamiltosporidium* genome due to the parasite’s ability to transmit vertically, which leads to bottlenecks and thus a decreased effective population size, N_e , and reduced effectiveness of selection (de Albuquerque et al., 2020; Haag et al., 2020). Previous studies have found that *H. tvaerminnensis* reproduces asexually and has a mixed-mode transmission (i.e., horizontal and vertical), whereas its sister species, *H. magnivora*, reproduces sexually and transmits among *D. magna* hosts only vertically (Haag et al., 2011; Haag et al., 2013b). *H. magnivora* is believed to have a second, as yet unknown host, to which it is transmitted horizontally (Mangin et al., 1995). While diversity in aspects like modes of reproduction and transmission makes the *Hamiltosporidium* genus a good model for studying parasite evolution in general, it might represent a more unique system in terms of understanding the consequences of gaining or losing a second host and of changing the amount of recombination following a switch in reproductive mode.

Several studies on the distribution of genetic variation in microsporidia in diverse hosts have shown that host demography contributes to parasite phylogeny, population structure, and demographic history (e.g., Gómez-Moracho et al., 2015; Shafer et al., 2009; Wang et al., 2019). Existing knowledge about *D. magna* demographic history (Fields et al., 2018; Orsini et al., 2012, 2013; Stollewerk, 2010) allows us to go one step further and investigate the parasite’s genetic demography in relation to that

of its host, looking at host–parasite co-phylogeny, co-biogeography, and coevolution. Previous studies on the *D. magna*–microsporidia host–parasite system have focused on mode of reproduction, demographic history, and transmission mode (Haag et al., 2013b; Haag et al., 2020) in an effort to understand the system’s colonization history and its transition from sexual to clonal reproduction. Here, we go a step further, reevaluating these questions using entire genomes sampled from a much larger swath of the species’ range, including Europe and Asia, and combining various approaches to investigate the evolutionary history and underlying mechanisms shaping genomic variation in this system. First, we ask how much intraspecific genomic variation exists in the microsporidian species *H. tvaerminnensis* using a new, highly improved long-read based reference genome. Based on a genomic study that estimated genomic variation in the microsporidium *Nosema ceranae* (Pelin et al., 2015), a close relative of *H. tvaerminnensis* with the same mode of transmission (Haag et al., 2020), a reasonable expectation was to find similar magnitudes of genetic diversity in *H. tvaerminnensis*. Second, using a multicontinental dataset, we ask how historical processes such as demography and phylogeography can explain the current distribution of genetic variation in *H. tvaerminnensis* across continental biogeographic scales? Knowing the phylogeography and demographic history of the host allowed us to test the co-phylogeography hypothesis. A finding that the parasite shares the same population structure and pattern of isolation-by-distance (IBD) as has been reported in the host (Andras et al., 2018; Fields et al., 2015) would support a model of phylogeographic co-cladogenesis. Since effects of demography and selection on genetic diversity are often hard to distinguish (Pavlidis et al., 2008), and nonadaptive processes may be an especially powerful force in the focal microsporidian genus (Haag et al., 2020), we also quantified the strength of positive selection in *H. tvaerminnensis*. Confirming weak selection strength would give us more confidence in our demography focused hypothesis and support previous findings (Haag et al., 2020).

Material and Methods

Daphnia magna Diversity Panel

The parasites examined in this study were derived from material collected within the framework of a large-scale biogeographic study of the host species, *D. magna* (Fields et al., 2015, 2018; Seefeldt & Ebert, 2019). Animals collected from across the species range were brought to the laboratory, and one iso-female line was created from each population (these iso-female lines are termed ‘clones’ for the course of this study). Animals from each clone were checked for microsporidia infections with phase-contrast microscopy, using squash-preparations or samples of the gut. The *D. magna* clones were sequenced with Illumina paired-end (PE) sequencing using a HiSeq 2500, with the respective parasite genome simultaneously sequenced in several cases. Whole-genome sequences of three *H. magnivora* and 15 *H. tvaerminnensis*, each from a different host clone collected from a different population, were obtained and used in the present study (Table 1). *H. tvaerminnensis* and *H. magnivora* are very closely related. Their species designation is based on morphological traits (Haag et al., 2011), but a genome-wide assessment of their relatedness has not been conducted thus far. Here, one common reference genome for the two *Hamiltosporidium* species was used.

Reference genome assembly

To obtain *D. magna* free of microbiota, except microsporidia, we followed a procedure described by Fields et al. (2015), using Qiagen Genomic Tips to obtain high molecular weight DNA from the FI-OER-3-3 clone, which was infected with *H. tvaerminnensis*. We used a Blue Pippin to isolate DNA fragments greater than 15 Kbp. A standard PacBio library was prepared, and one SMRTCELL was sequenced in CLR mode on a Sequel I at the D-BSSE (Basel, Switzerland).

After receiving the raw data in a subread BAM file, we initially mapped all reads to the *D. magna* reference genome (Fields et al., in prep). In order to achieve the most complete assembly of *H. tvaerminnensis*, we used Canu v.2.1.1 (Koren et al., 2017) with high sensitivity parameters to effectively work out the *H. tvaerminnensis* genome as part of the larger metagenome of *D. magna*

and *H. tvaerminnensis* in the read set. We used BlobToolKit v.2 (Challis et al., 2020) to further isolate contigs specific to *H. tvaerminnensis*. By mapping Illumina data from an uninfected host genotype, we were able to further disentangle suspected *H. tvaerminnensis* contigs from contigs that likely belonged to the host. Afterwards, we used purge_haplotigs v.1.1.1 (Roach et al., 2018) to haplodize the assembly of contigs that presumably derived only from the parasite genome. Finally, we polished the resulting *H. tvaerminnensis* assembly using Genomic Consensus v.2.3.3 (Chin et al., 2013) and Pilon v.1.23 (Walker et al., 2014).

To annotate the new *H. tvaerminnensis* genome, we lifted the previously published annotation from Haag et al. (2020) over using RaGOO v.1.1 (Alonge et al., 2019) and evaluated the biological completeness of this new genome draft using BUSCO v.4.0.1 (Seppey et al., 2019) and its microsporidia_odb10 (Creation date: 2020-08-05).

Mapping and variant calling

Raw reads were assessed for quality with FastQC v.0.11.7 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and subsequently trimmed with Trimmomatic v.0.38 (Bolger et al., 2014) to remove low quality sequence and adapter contamination. Trimming success was assessed by doing a second run of FastQC. Quality trimmed reads were mapped to the reference genome with BWA MEM v.0.7.17 (H. Li, 2013). SAMtools v.1.9 (H. Li et al., 2009) was used to convert the SAM files to BAM files, coordinate sort individual BAM files, and remove unmapped reads. Read groups were added and duplicates marked for individual BAM files using Picard Toolkit v.2.18.16 (Broad Institute, 2019). The average read depths were computed using SAMtools function depth. GATK v.3.8 HaplotypeCaller was used to generate variant calls (McKenna et al., 2010; Van der Auwera et al., 2013), specifically GVCFs, which were first generated for individual BAM files. We used the GATK function GenotypeGVCF to combine GVCFs

into all-site VCFs. Variant calls were generated for *H. tvaerminnensis*–*H. magnivora* in the form of a VCF file, which was filtered to exclude INDEL variants with VCFtools v.0.1.16 (Danecek et al., 2011).

Coding sequences – single copy orthologs

Protein coding regions are of special interest for phylogenetic analyses and for identifying signals of selection (including its efficacy) across the genome. In order to characterize variation in these particular subsections of the genome, we needed to extract subsets of the genome-wide VCF.

Protein sequences of *H. magnivora* (BEOM2 v1; GenBank accession: PITI000000000.1) were downloaded from NCBI and protein datasets of *H. tvaerminnensis* and *H. magnivora* were used as input for OrthoMCL v.2.0.9 to find one-to-one orthologs between the two (L. Li et al., 2003, following specifically the automation pipeline described at the Github repository:

<https://github.com/apetkau/orthomcl-pipeline>). The orthologous sequences of *H. tvaerminnensis* and *H. magnivora* were aligned with PRANK v.170427 (Löytynoja, 2014) using a custom script adapted from Fields et al. (in review). After an initial survey of pairwise alignment quality, we implemented a masking step, in which excessively divergent or poorly aligned sequences were excluded from downstream analysis. We then used the R v.3.5.1 (R Core Team, 2018) package *seqinR* v.3.4-5 (Charif & Lobry, 2007) to import FASTA alignments and *PopGenome* v.2.7.1 (Pfeifer et al., 2014) to manipulate the VCF files. To produce multiple sequence alignments, pseudo-haplotypes (i.e., a random assignment of SNPs to one or the other haplotype) were generated by recoding the structure of the VCF file. One variant-imputed version of the reference per pseudo-haplotype was produced for each *H. tvaerminnensis* sample with the function *vcf2fasta* of *VcfLib* v.1.0.0-rc2 (Garrison, 2019), and then coding sequences of interest were cut out of the variant-imputed versions of the reference with the script *gff2fasta.pl*

(https://github.com/ISUGenomics/common_scripts). The resultant coding sequences were aligned to the initial *H. tvaerminnensis* and *H. magnivora* alignments using MAFFT v.7.407 (Katoh et al., 2002; Katoh & Standley, 2013) and its --add option. Next, all positions masked in the two species reference

alignments were masked in the multi sequence-alignments with `generate_masked_ranges.py` (<https://gist.github.com/danielecook>) and BEDtools v.2.27.1 (Quinlan, 2014) function `maskfasta`.

Sequence variation and population genetic analyses

Per-site nucleotide differences (π) were calculated along the genome for *H. tvaerminnensis* in 1 Kbp windows using `pixy` v.0.95 (Korunes & Samuk, 2021), and the difference between lineages was statistically tested using a Wilcoxon signed rank test in R. Alleles of less than half and more than double the average sample coverage were filtered out for the calculation of π . Similarly, π was calculated for the coding sequences of *H. tvaerminnensis* using the script `selectionStats.py` (<https://github.com/tatumdmortimer/popgen-stats>). Both methods consider invariant sites in their calculations, making calculations more consistent with theoretical expectations and more comparable among species (Korunes & Samuk, 2021). This analysis was not conducted with *H. magnivora*, as only three genomes were available.

Population structure and phylogenetic analyses

We conducted both a principle component analysis (PCA) and a cluster analysis with the *Hamiltosporidium* whole-genome polymorphism data using the R packages *SNPRelate* v.1.14.0 and *gdsfmt* v.1.16.0 (Zheng et al., 2012). Additionally, the pattern of IBD was tested for *H. tvaerminnensis* by comparing the pairwise genetic differentiation of the samples with their pairwise geographical distance. The R base *stats* function `dist` was used to calculate the pairwise Euclidean distance, while *geodist* v.0.0.3 (Padgham & Sumner, 2019) was used to calculate the geographical distance between the samples. *VCFR* and *adegenet* v.2.1.2 (Jombart, 2008) were used in R for file import and format conversions. The correlation between differentiation and geographical distance was tested with a dbMEM analysis by RDA. Specifically, we transformed the explanatory variable, geographic distance, into dbMEMs using the R package *adespatial* v.0.3-14 (Dray et al., 2021) and decomposed the response variable, genetic differentiation, into principal components using the R

base *stats* function *prcomp*. The RDA was done in R using the package *vegan* v.2.5-7 (Oksanen et al., 2020) with significance assessed with 1,000 permutations. In addition to looking at whole-genome diversity, we also examined the phylogenetic signal in protein-coding regions of the genome directly. As the aforementioned pseudo-haplotypes are inappropriate for a number of phylogenetic methods that rely on haplotype information, we used ambiguity codes, an alternative method for unknown phase. We used the GATK-functions *SelectVariants* and *FastaAlternateReferenceMaker* to generate variant-imputed versions of the *H. tvaerminnensis* genome with ambiguity codes. Single copy ortholog genes were extracted from the alternative references using *gff2fasta.pl* and concatenated into a single sequence. The larger alignment of these individual samples was filtered for four-fold degenerate sites with *MEGA* v.7.16.0617 (Kumar et al., 2016). A Bayesian phylogenetic analysis was conducted using *BEAST2* v.2.5.1 (Bouckaert et al., 2019; Drummond et al., 2005) with the ambiguity code option enabled. The obtained tree was combined with each sample's geographical coordinates using the R package *phytools* v.0.7-20 (Revell, 2012). Similarly, for the host–parasite co-phylogeny, we applied the same methodology using the host genome (V3.0; Daphnia Genome Consortium).

Demographic history

We attempted to reconstruct the demographic history of the *H. tvaerminnensis* and *H. magnivora* samples by characterizing the unfolded site frequency spectrum (uSFS) of single copy orthologs, distinguishing the uSFS for synonymous and nonsynonymous sites. The script *siteFrequencySpectrum.R* (<https://github.com/tatumdmortimer/popgen-stats>) was used to interface with *PopGenome* and modified to compute the uSFS for both site classes. In the uSFS of *H. tvaerminnensis*, we used *H. magnivora* as an outgroup to determine the state of the variants (ancestral or derived), and vice versa. We used the function *gap.barplot* from the R package *plotrix* v.3.7-7 (Lemon, 2006) to plot the uSFS.

To estimate changes in the effective population size history of *Hamiltosporidium* samples, we used PSMC v.0.6.5 (H. Li & Durbin, 2011). Therefore, all BAM files were downsampled to an average whole-genome coverage of 50× with the SAMtools function view. Afterwards, one consensus sequence per sample was produced using SAMtools option mpileup, BCFtools v.1.9 option call (H. Li, 2011) and vcutils.pl vcf2fq, which is included in BCFtools. SNPs with a coverage of less than a third or more than double of the average whole-genome coverage were filtered as suggested by the developer (<https://github.com/lh3/psmc>). The script plotPsmc.r (Liu & Hansen, 2017) was used for plotting the inferred historical dynamics of N_e . Similarly, “pseudo-diploid” samples were analyzed. Therefore, downsampled BAM files of two samples were jointly input to SAMtools to produce two-sample VCF files using BCFtools. WhatsHap v.0.18 (Martin et al., 2016) was then used for read-based phasing. Unphased sites were masked and haplotypes were combined in all four possible ways before converting them to four FASTQ files using vcf2fq.

Rate of adaptive nucleotide substitutions

The magnitude of positive selection, or the rate of adaptive substitution (α), was inferred for *H. tvaerminnensis* by using both the classical McDonald–Kreitman test (MKT; McDonald & Kreitman, 1991) and more recently derived extensions. Thereby, intraspecific diversity and divergence counts as compared to the outgroup (i.e., counts of polymorphisms segregating within the species and counts of fixed differences between the ingroup and the outgroup) were created from a multi-sequence alignment of single copy orthologs. These alignments were first concatenated and then used as input for iMKT v.0.2 (Murga-Moreno et al., 2019), a web-based application (accessed on 05-21-2021) to perform standard MKTs (McDonald & Kreitman, 1991). We also used iMKT to estimate α based on the asymptotic MKT approach of Haller and Messer (2017). This approach differs from a traditional MKT in that it makes an explicit attempt to control for the confounding effects of low frequency, deleterious allele classes. Finally, we estimated the distribution of fitness effects to acquire a model-based estimate of α and ω_A (an estimate of adaptation that may sometimes be

more useful than α) using dfe-alpha v.2.16 (Eyre-Walker & Keightley, 2009; Keightley & Eyre-Walker, 2007), after masking internal stop codons with MACSE v.2.03 (Ranwez et al., 2011). We ran dfe-alpha using a wrapper script adapted from Fields et al. (in review).

Results

A reference genome of H. tvaerminnensis based on long reads

As the tissue of the host and the intracellular parasite are difficult to separate, we sequenced host and *H. tvaerminnensis* together using long-read PacBio technology. Our initial assembly of the host and parasite genomes resulted in an assembly length of 123 megabase pair (Mbp). While the host genome had an average coverage < 50 $\times$, the parasite genome had a coverage mostly > 100 $\times$ (Fig. S1A). The *H. tvaerminnensis* genome had an assembly length of 28.5 Mbp and showed nearly 0 $\times$ coverage when we analyzed reads from a sequenced *D. magna* host known to be uninfected by *H. tvaerminnensis* (Fig. S1B). We used purge_haplotigs to compile our final, haploid assembly of *H. tvaerminnensis*.

This new reference genome significantly improved the contiguity of the previous *H. tvaerminnensis* genome, which was based solely on short-read Illumina sequencing (Haag et al., 2020) and which had a total length of 18.3 Mbp, a maximum contig length of 0.06 Mbp, an N50 of 0.01 Mbp, and was composed of 2915 contigs. In contrast, our new long-read based assembly had a total length of 21.6 Mbp, a max contig length of 2.7 Mbp, an N50 of 1.03 Mbp, and 27 contigs. Along with the improved contiguity, the biological completeness of the new reference genome improved, with a ~15 % increase in both genome length and BUSCO completeness score (Table S1). However, without cytological data, this improved assembly may still not represent complete chromosomal contigs.

Samples, mapping, and sequence variation

Infected clonal host lines were sequenced together with the parasites, resulting in 15 genomic samples, each from a different Eurasian population with sufficient representation of *H. tvaerminnensis* reads to describe its genomic variation (Table 1). Genomic samples from three populations of our outgroup *H. magnivora* were also included to extend the number of analytical approaches for characterizing genome-wide variation in the *H. tvaerminnensis* sample. The percentage of sequencing reads mapped to the parasite's new reference genome ranged from 6.28 % to 63.70 %; most of the remaining reads were from the *D. magna* host. The average whole-genome coverage was greater than 94× in all cases of *Hamiltosporidium* infections. The overall SNP density in *H. tvaerminnensis* was 25 SNPs per kilobase pair (Kbp), with the total number of SNPs being 554,935. The average whole-genome π (Nei & Li, 1979) value was 0.009, or approximately 1 %.

Population structure and phylogenetic analyses

We assessed the amount of genomic variation that differed or was shared between samples by applying a PCA to the whole-genome SNP data (Table S2). The first two eigenvectors of the PCA (Fig. S2) divided the *Hamiltosporidium* samples into three clusters: PC1 (explaining 36.09 % of the variance) separated *H. tvaerminnensis* samples from those of *H. magnivora*; and PC2 (22.35 %) split *H. tvaerminnensis* samples into two clusters, one Northern (N = 9) encompassing Scandinavian and East Asian samples, and one Southern (N = 6), encompassing all Mediterranean samples. A cluster analysis applied to the whole-genome SNP data resulted in the same groupings (Fig. S2), as did a Bayesian phylogenetic tree estimation based on four-fold degenerate sites of the 2,229 single copy orthologs (Fig. 1). After finding these groupings consistently in all of our population structure analyses, we refer to them as Northern and Southern *H. tvaerminnensis* lineages.

We inferred the demographic history for the two *H. tvaerminnensis* lineages separately. The SNP density and the π value showed that the Northern lineage (11 SNPs per Kbp) was considerably less SNP-dense and less diverse ($\pi = 0.005$) than the Southern lineage (19 SNPs per Kbp; $\pi = 0.008$;

Wilcoxon signed rank test: $p < 0.001$). We found a positive correlation between pairwise differentiation and geographical distance for the overall *H. tvaerminnensis* data and the species' lineages using distance-based Moran's eigenvector maps (dbMEM) analysis by redundancy analysis (RDA) (overall: $R^2 = 0.15$, $p = 0.197$; Northern: $R^2 = 0.43$, $p = 0.001$; Southern: $R^2 = 0.82$, $p = 0.003$), which has been suggested as preferable over the formerly used Mantel tests (Legendre et al., 2015). The much higher correlation coefficients within the lineages, as compared to the total dataset, strongly underline the importance of considering the phylogeography within the lineages.

Demographic history

We examined the demographic history of *H. tvaerminnensis* looking at both the full dataset and each the Northern and Southern lineages individually. The unfolded site frequency spectra (uSFS) of *H. tvaerminnensis* suggest strong deviation from what one would expect in a Wright-Fisher population (Crawford & Lazzaro, 2012), i.e., a panmictic population in mutation-drift balance (Fig. 2A, C, and D). Specifically, we do not observe any left-skewed distribution in *H. tvaerminnensis*, although it was seen in *H. magnivora* (Fig. 2B). The Northern *H. tvaerminnensis* lineage had a clear peak around intermediate frequencies consistent with the earlier observation of fixed heterozygosity at many northern European sample sites (Haag et al., 2013b). Such a clear peak was not observed in the Southern *H. tvaerminnensis* lineage, however, which showed a more even spread than expected from a Wright-Fisher population.

To further understand the evolution of these lineages, we analyzed the changes of N_e across time using software based on a Pairwise Sequentially Markovian Coalescent (PSMC) model to infer population history. Additionally, we did a very rough time calibration using publicly available estimates of mutation rates from the well-studied 'red yeast' fungi *Rhodotorula toruloides* (Long et al., 2016) and previously reported generation times of microsporidia (Kramer, 1965). PSMC revealed a relatively consistent pattern across samples from the same species and lineage (Fig. 3). Ancient N_e ,

(i.e., around 100 thousand (K) to 1 million (M) years ago) was similar among all species and lineages but diverged in more recent times, with N_e for *H. magnivora* being less dynamic across time than N_e for all *H. tvaerminnensis* samples. The biggest difference between the species was the recent, strong reduction in N_e a few thousand years ago, seen only in the *H. tvaerminnensis* samples. There were also differences among the *H. tvaerminnensis* samples. Specifically, a more ancient decrease in N_e , between around 15 K and 50 K years ago was seen in all *H. tvaerminnensis* samples except those from Southeast Europe/Middle East, suggesting distinct population histories for the eastern and western samples of the Southern lineage. Apart from that, the consistency within the Northern lineage and the two groups of the Southern lineage gives us confidence about the quality of this population history measure.

According to the PSMC results, neither *Hamiltosporidium* species showed extreme upward vertical trajectories in N_e , which one would expect in a diploid species that reproduces asexually over a long time period and has non-recombining chromosomes that separately accumulate mutations (i.e., the Meselson effect; Cepitis, 2003; Weir et al., 2016). This result could suggest that the supposedly asexual *H. tvaerminnensis* experiences some recombination, as is believed to be true for the sexual *H. magnivora*. To further understand how mode of reproduction evolved, we ran the PSMC method with *in silico*, “pseudo-diploid” samples, i.e., single phased haplotypes of our diploid samples that we combined in a pairwise manner. While this method is frequently used to estimate the relative divergence times of species (Cahill et al., 2016; Mattle-Greminger et al., 2018), we used it to investigate the degree of separate chromosomal evolution within and between *Hamiltosporidium* species. The PSMC results using “pseudo-diploids” differed strongly from the results using the original samples (compare Fig. 3A and B), especially in recent times. N_e in ancient times was similar in both analyses, with none of the intraspecies “pseudo-diploid” samples’ PSMC results showing extreme upward vertical trajectories; however, in recent times, N_e showed extreme upward vertical trajectories in all the intraspecies “pseudo-diploid” samples (Fig. 3B). This confutes long-time

asexuality and represents spatially based population differentiation. Of the intraspecies combinations, *H. magnivora* showed the deepest upward vertical trajectories in time (orange in Fig. 3B), suggesting the highest differentiation (see also tree in Fig. 1); however, upward vertical trajectories in inter-lineage and interspecies “pseudo-diploid” samples are further in the past (green lines in Fig. 3B), which is expected for more diverged haplotype combinations. Interestingly, there may have been lineage-specific events, including hybridization with *H. magnivora* (light green in Fig. 3B): in contrast to the other “pseudo-diploid” haplotype combinations, the PSMC results of Northern *H. tvaerminnensis*–*H. magnivora* “pseudo-diploid” haplotype combinations show upward vertical trajectories in N_e followed by a return to lower values, which we interpret as putative gene flow between the species.

Rate of adaptive nucleotide substitutions

As genomic signals of demography and selection are often hard to disentangle, we assessed the importance of selection in *H. tvaerminnensis* by estimating the rate of adaptive substitutions (α) using a standard McDonald–Kreitman test (MKT). The α values of *H. tvaerminnensis* ranged from 0.121 to 0.227 (Table 2). Asymptotic MKT derived α values, which are expected to be less affected by demographic factors, ranged from 0.222 to 0.279. However, the asymptotic methodology needed more data in some frequency categories than was available in our study (i.e., more neutral and selected diversity) to perform the asymptotic fit over the data (Jesús Murga-Moreno, pers. comm.). Therefore, the estimation of α using asymptotic MKT resulted in large confidence intervals and was unfeasible for the Northern lineage. The α values of *H. tvaerminnensis* estimated using dfe-alpha ranged from -0.026 to 0.203. Estimates of α have previously been found to be above 0.5 in many species and to reach up to 0.9 (Galtier, 2016). The consistently low estimates of α for *H. tvaerminnensis* support previous findings that genomic variation in *Hamiltosporidium* spp. might be predominantly driven by nonadaptive processes rather than by adaptive evolution (Haag et al., 2020).

Discussion

A core aim of population genetics is to understand genetic variation over space and time and link it to variation in the ecology of the species. To understand genetic variation in parasites, it is further helpful to know the host's biology and features that determine interactions, such as parasite transmission and host–parasite coevolution. Our population genomic study provides insights into the variation of genomic diversity, population structure, and demographic history in a highly specific parasite that exclusively infects the planktonic freshwater species *Daphnia magna*, a model system in host-parasite evolutionary ecology. This parasite, the microsporidium *Hamiltosporidium*, is widely spread and locally very common (Decaestecker et al., 2005; Ebert et al., 2001; Goren & Ben-Ami, 2013). Our analyses use samples spanning the Eurasian region of the host's Holarctic range where *D. magna* forms two clades, a Western Eurasian and an East Asian clade (Fields et al., 2018). We also find two clades for *H. tvaerminnensis*, but the host and parasite clades are not congruent to each other, thus are not the result of co-cladogenesis. Furthermore, we find relatively low levels of adaptive evolution, which is consistent with the hypothesis that a genome expansion may have resulted from high levels of genetic drift, corroborating a mechanism for the evolution of highly divergent genome architectures in microsporidia in general (Haag et al., 2020). Within the genus, we hypothesize that *H. tvaerminnensis* might be derived from *H. magnivora* (de Albuquerque et al., 2020), as *H. tvaerminnensis*, which presumably no longer depends on a second host, has been able to colonize geographic regions inaccessible to *H. magnivora*. This apparently became possible with the evolution of direct horizontal transmission among *D. magna* individuals (only vertical in *H. magnivora*), thus eliminating the need for a second host (*H. magnivora* is still believed to depend on a second, yet unknown, host for horizontal transmission). Living without a second host, also resulted in a loss of sexual reproduction (Fig. 4), which has been suggested to take place in the second host (Mangin et al., 1995), with profound consequences for the evolution of genetic diversity of *H. tvaerminnensis*.

No pattern of host-parasite (co-)phylogeography

Because microsporidian parasites and their hosts interact closely, it might be assumed that parasite and host phylogeography should be congruent—the within species equivalent of Fahrenholz rule (Fahrenholz, 1913), which states that host and parasite phylogenies are expected to be congruent to each other, i.e., show a pattern of co-cladogenesis. It is now known, however, that this is rarely the case (Page, 2003). A strict co-phylogeography may be expected if the parasite’s mode of transmission is uniquely vertical (Werren et al., 2008) and when host and parasite co-disperse (Page, 2003). A previous attempt by Pelin et al. (2015) to reconstruct a phylogeography based on whole-genome data for the microsporidium *N. ceranae* remained inconclusive due to high, human-driven migration and lack of isolation-by-distance (IBD). In a different parasite of *D. magna*, the horizontally transmitted bacterium *Pasteuria ramosa*, a pattern of IBD was observed (Andras et al., 2018; Fields et al., 2015), but the issue of co-phylogeography was not addressed in the *Pasteuria–Daphnia* system. In the *H. tvaerminnensis–D. magna* system studied here, vertical transmission is common, and natural dispersal as well as co-dispersal seems likely (Haag et al., 2013b), especially because the unstable pond habitat of *D. magna* requires a high propensity for migration. We therefore hypothesized that we would find a pattern of host–parasite co-phylogeography and co-cladogenesis. Although *D. magna* shows two main clades in Eurasia, a Western Eurasian and an East Asian clade (Fields et al., 2018), our phylogeographic analysis suggests that the split between the two main lineages of *H. tvaerminnensis* is clearly distinct from that of the host. Instead of a Western Eurasia/Eastern Asia split, we found a North–South split, a signal consistent across all the phylogeographic analyses we conducted (Fig. 1, Fig. S2). Thus, the phylogeography of *D. magna* and *H. tvaerminnensis* is non-overlapping and cannot be explained by vertical transmission and co-dispersal alone. Nevertheless, a pattern of IBD (as has been shown for *D. magna* by Fields et al. (2015) and Andras et al. (2018)) was observed for *H. tvaerminnensis*, supporting the assumption of

low or moderate gene flow between populations and reinforcing the intimate host–parasite association.

The divergence of Northern and Southern parasite lineages

Our phylogenetic analysis of *Hamiltosporidium* revealed a clear separation between a Northern and a Southern clade of *H. tvaerminnensis*. Not surprisingly, therefore, the correlation between pairwise genetic differentiation and geographical distance was clearer within these two parasite lineages than in the combined dataset. We speculate that the Northern *H. tvaerminnensis* lineage resulted from the parasite spreading after the host expanded northwards out of its glacial refugium about 10,000 years ago when the ice retreated from the North. This spread of the Northern lineage of *H. tvaerminnensis* might have accompanied a population bottleneck, which resulted in reduced genetic diversity. Furthermore, the uSFS of Northern *H. tvaerminnensis* indicates fixed heterozygosity at many positions (Fig. 2C), a finding consistent with Haag et al.'s (2013b) suggestions that the spread to the North coincided with the emergence of asexuality and a subsequent population expansion. The Southern lineage's samples, in comparison, are more diverged from each other, even though their currently known geographic spread is smaller than that of the Northern lineage. This could have resulted from an older transition to asexual reproduction or an additional glacial refugium other than the one in Southeast Europe/Middle-East for *D. magna* (Fields et al., 2018). However, this needs further investigation. Additionally, estimating divergence times between the parasite lineages and comparing them to the host clades' divergence time could indicate whether co-dispersal, i.e., range expansion of host and parasite at a similar time and in the same direction, played an essential role for the evolution of the host and *H. tvaerminnensis*.

Relative importance of selection

Models to determine the demographic history of a population are typically based on the assumption of neutrality, expecting that deviations from neutrality will lead to biased results (Johri et al., 2021).

Discussing demographic history with regard to selection is therefore important. In *Hamiltosporidium* spp. It has been proposed that nonadaptive processes reduce the efficiency of adaptive evolution (Haag et al., 2020). Briefly, horizontal transmission is thought to have been the transmission mode of the *Hamiltosporidium* ancestors until, at some point, an ancestral lineage developed the ability to transmit vertically among *D. magna* as well, which is now the case for all *Hamiltosporidium* species (Haag et al., 2020). A shift from horizontal transmission to mixed-mode transmission, especially if vertical transmission predominated and played an important role in co-dispersal, could have reduced effective population size, thus making it easier for slightly deleterious mutations to increase in frequency by genomic drift and making purifying and positive selection less efficient (Haag et al., 2020). Consequently, the number of nonsynonymous substitutions fixed by genomic drift increases while the proportion of substitutions fixed by positive selection (α) decreases.

To test this theory, we used our whole-genome data to determine the relative strength of selection in *Hamiltosporidium*. Of the different methods used to estimate α in our study, dfe-alpha relies on a specific model of demography, any deviation from which might confound this method. The SFS, as well as historical changes in N_e were inferred (see section below; Fig. 2 and Fig. 3) to determine how the demography of the *Hamiltosporidium* samples could violate dfe-alpha's underlying demographic model. Indeed, Northern *H. tvaerminnensis* samples exhibited a clear recent bottleneck in their SFS and demographic history, and the estimated α value of this lineage was negative (Table 2), suggesting that (a) positive selection is either weak or constrained, or (b) the underlying likelihood model of dfe-alpha was not appropriate for this data. Test and reference regions were similarly affected by demography, as the synonymous and nonsynonymous SFS have the same shape (Fig. 2). To ensure a demographically robust analysis for our estimation of α , we conducted both a standard MKT and a more recent approach called asymptotic MKT (Haller & Messer, 2017). It has been suggested that this latter approach can accommodate a wider range of demographic histories by accounting for the effects of low frequency, deleterious variants. α values estimated using MKTs were similar to those using dfe-alpha. Also, the overall dataset always had the

highest estimate, while the Northern lineage always had the lowest. Furthermore, among all the species for which α and the rate of adaptive divergence relative to neutral divergence (ω_A), has been estimated, *H. tvaerminnensis* is on the lower end of the selection efficacy distribution (e.g., Galtier, 2016; Rousselle et al., 2020). Taken together, our findings support the hypothesis of weak selection efficacy, increased genetic drift, and a resulting genome expansion in *Hamiltosporidium* spp. proposed by Haag et al. (2020).

Historical changes in N_e

Historical processes are a major contributor to present-day distribution of genetic variation, influencing both similarities and differences in the co-phylogeography that a parasite might share with its host. The effective population size history of *H. tvaerminnensis* constructed with the PSMC method showed a rather consistent pattern across samples from the same lineage but revealed differences among lineages (Fig. 3A). Since the mutation rates and generation times of *H. tvaerminnensis* and *H. magnivora* are unknown, we used available estimates for related species to estimate time intervals of interest. The more ancient decrease in N_e , which is observable in many species (Hewitt, 1996; Holder et al., 1999; Stewart et al., 2010), is thought to stem from the species' range contraction, which happened when the glaciers expanded at the beginning of the last glaciation about 110,000 years ago. It is followed by an increase in N_e , presumably after the glaciers retreated about 10,000 years ago, as also described for the host *D. magna* (Fields et al., 2018). Samples that originated from hosts in the Middle East, a region where glaciers had no direct impact and where the presumed glacial refugia of the Western Eurasian *D. magna* clade was located, consistently seem to be more stable during the whole period of glacial expansion and recession. Numerous other species have shown patterns consistent with southern refugia during this period (Stewart et al., 2010). Importantly, our Spanish samples of the Southern lineage show an N_e pattern of change that is more similar to non-Middle Eastern samples (Fig. 3A), coinciding with the absence of a glacial refugium for hosts in southwestern Eurasia (Fields et al., 2018). Thus, *Daphnia* hosts and

their parasites presumably recolonized Western Europe from the same refugium, including the Hispanic peninsula after the last glaciation.

A second decrease where N_e dips to very low estimates is observed in more recent times for all *H. tvaerminnensis* samples but not for *H. magnivora*. A number of potential factors can decrease the effective population size: specialization of the parasite to a single host species; a clonal mode of reproduction (i.e. a switch to asexuality); an epidemic dynamic of boom and bust; restricted and local dispersal; frequent population extinctions and recolonizations; a short-lived, annual or ephemeral host; and small, fragmented host populations (Barrett et al., 2008). Although several of these factors likely contributed to the N_e dynamics of *H. tvaerminnensis*, it is unclear which of them predominated. For example, *D. magna* varies strongly in susceptibility to *H. tvaerminnensis*, with the most susceptible hosts occurring in unstable habitats that have a high propensity to dry up in summer (like ephemeral rock or desert pools) (Cabalar et al., 2019; Lange et al., 2015). These unstable habitats are also associated with frequent population extinctions and recolonizations as well as a loss of diversity because of frequent population bottlenecks.

However, as previous analyses have suggested that *H. tvaerminnensis* transitioned from a sexual to an asexual reproduction mode at some point in its recent history (Haag et al., 2013a), we hypothesize that the more recent decrease in *H. tvaerminnensis*' N_e might be related mainly to the evolution of horizontal transmission directly from *D. magna* to *D. magna* (Fig. 4). The loss of sexual reproduction in the absence of a second host, which is believed to be necessary for sexual reproduction in *H. magnivora* (Mangin et al., 1995), reduced N_e . However, the transition from sexual to exclusively asexual reproduction would leave a very distinct historical pattern of change in N_e as inferred via PSMC. Specifically, because individual homologous chromosomes can no longer recombine, one would expect to see an upward, nearly vertical trajectory in N_e as mutations accumulate on separate chromosomal copies, which our reconstructed history of *H. tvaerminnensis*'

population size does not show, or at least not deep enough in time when such a transition would be easier to pin down with the PSMC method. This implies that exclusive asexuality is probably not ancient, coinciding with the SFS of the Northern lineage, which shows a distinct peak at 0.5 caused by fixed heterozygosity. Being exclusive asexual for a long period of time would provide sufficient time for individual haplotypes to accumulate their own distinct mutations, which is not what the SFS shows (Fig. 2). Alternatively, although *H. tvaerminnensis* may be a mostly asexual species, it could still show occasional recombination, for example, when it encounters its now presumably facultative second host. A loss of sexual reproduction with occasional recombination would explain the low estimates of N_e without extreme upward vertical trajectories of non-recombining chromosomes, using the PSMC method. However, any discussion about the second decrease in N_e should be undertaken with the mindfulness that estimates of PSMC from the recent past are generally more uncertain (H. Li & Durbin, 2011), and that PSMC has not often been used for understanding the demographic history of parasitic species (but see e.g., Hecht et al., 2018).

When comparing *H. magnivora* and *H. tvaerminnensis*, we found that their histories of N_e do not align precisely. Along with differences in their demographies, differences in mutation rate and generation time, both of which used here may not be optimal for each species, can also cause aberrations with the PSMC method. The PSMC results for the *in silico*, “pseudo-diploid” samples support the idea that hybridization events happened between *H. magnivora* and the Northern *H. tvaerminnensis* lineage. The combinations of *H. tvaerminnensis* and *H. magnivora* haplotypes show the expected, nearly vertical trajectories of no longer recombining lineages in deeper time (Fig. 3B). However, combinations of the Northern *H. tvaerminnensis* lineage and *H. magnivora* reveal a demographic signal consistent with gene flow between the two species. Overall, our analyses support the suggestion that an ancestral *H. tvaerminnensis* became able to transmit horizontally directly from *D. magna* to *D. magna* (*H. magnivora* is unable to do this) and thus was able to persist in habitat without its second host. This allowed *H. tvaerminnensis* to colonize Northern habitats, but

also made it reliant almost exclusively on asexual reproduction, with the latent ability to recombine if the second host were present.

Comparatively high genomic variation in Hamiltosporidium

Large-sized microsporidia genomes are generally difficult to assemble due to a high number of repetitive elements (Parisot et al., 2014). *H. tvaerminnensis* has been target of previous assembly trials using shotgun and Illumina sequencing (Corradi et al., 2009; Haag et al., 2020). Here, we were able to produce a very contiguous genome assembly of this microsporidium with a large genome size using long-read PacBio sequencing. *H. tvaerminnensis* has a whole-genome SNP density of approximately 25 SNPs per Kbp, which is more diverse than its close relative *N. ceranae* (12.7) and on the high end compared to other microsporidia and fungi (Pelin et al., 2015). A likely reason why the multi-continent *N. ceranae* sample has a lower density estimate is because the species was recently introduced into the sampled region and may have experienced a bottleneck during range expansion (Pelin et al., 2015). In contrast, *D. magna*, the host of *H. tvaerminnensis*, is native to Eurasia.

Conclusion

The evidence presented here of a non-overlapping phylogeography between two partners in an intimate host–parasite system with a wide geographic distribution has several implications. It suggests that the parasite, *H. tvaerminnensis*, might be a rather young parasite of its host, *D. magna*. Else, it could not be simultaneously wide-spread *and* have a phylogeography distinct from its host. Also, it implies that co-dispersal of host and parasite, while common, is not the only form of dispersal. Thus our research underscores the importance and potential of using samples from the whole-species range in population genomic studies. Regarding microsporidia with long genomes, we could strengthen the assumption of a reduced selection efficacy, likely as a result of high levels of genetic drift that ultimately caused genome expansion (Haag et al., 2020). By quantifying genomic variation and selection efficacy in other microsporidia, more evidence for this hypothesis and for the

evolution of different genome architectures in this taxon could be provided. However, while our study reveals some general principles on the evolution of genomic variation in a parasite, it also shows how specific factors shape genomic variation, some of these (or their combination) being possibly unique to this system (e.g., loss of the second host and the shift to exclusive asexuality in the absence of the second host). It therefore provides a case study for the genomic evolution of a microsporidium at a fine scale, which awaits comparison to other systems that will help paint a bigger picture of the evolution of specific obligate parasites.

Acknowledgements

We thank Jürgen Hottinger, Michelle Krebs and Andrea Cabalzar for help in the field and laboratory. Members of the Ebert group provided feedback on the study and the manuscript. This work was supported by the Swiss National Science Foundation (SNSF) (grant numbers 310030B_166677 and 310030_188887 to DE).

References

- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, 20(1), 224. <https://doi.org/10.1186/s13059-019-1829-6>
- Altermatt, F., & Ebert, D. (2008). Genetic diversity of *Daphnia magna* populations enhances resistance to parasites. *Ecology Letters*, 11(9), 918–928. <https://doi.org/10.1111/j.1461-0248.2008.01203.x>
- Andras, J. P., Fields, P. D., & Ebert, D. (2018). Spatial population genetic structure of a bacterial parasite in close coevolution with its host. *Molecular Ecology*, 27(6), 1371–1384. <https://doi.org/10.1111/mec.14545>
- Barrett, L. G., Thrall, P. H., Burdon, J. J., & Linde, C. C. (2008). Life history determines genetic structure and evolutionary potential of host–parasite interactions. *Trends in Ecology & Evolution*, 23(12), 678–685. <https://doi.org/10.1016/j.tree.2008.06.017>
- Bellis, E. S., Kelly, E. A., Lorts, C. M., Gao, H., DeLeo, V. L., Rouhan, G., Budden, A., Bhaskara, G. B., Hu, Z., Muscarella, R., Timko, M. P., Nebie, B., Runo, S. M., Chilcoat, N. D., Juenger, T. E., Morris, G. P., dePamphilis, C. W., & Lasky, J. R. (2020). Genomics of sorghum local adaptation to a parasitic plant. *Proceedings of the National Academy of Sciences*, 117(8), 4243–4251. <https://doi.org/10.1073/pnas.1908707117>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>

- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, 15(4), e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>
- Broad Institute. (2019). “Picard Toolkit.” Broad Institute, GitHub repository. <http://broadinstitute.github.io/picard/>
- Cabalzar, A. P., Fields, P. D., Kato, Y., Watanabe, H., & Ebert, D. (2019). Parasite-mediated selection in a natural metapopulation of *Daphnia magna*. *Molecular Ecology*, 28(21), 4770–4785. <https://doi.org/10.1111/mec.15260>
- Cahill, J. A., Soares, A. E. R., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1699), 20150138. <https://doi.org/10.1098/rstb.2015.0138>
- Ceplitis, A. (2003). Coalescence times and the Meselson effect in asexual eukaryotes. *Genetical Research*, 82(3), 183–190. <https://doi.org/10.1017/s0016672303006487>
- Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. *G3 Genes|Genomes|Genetics*, 10(4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In U. Bastolla, M. Porto, H. E. Roman, & M. Vendruscolo (Eds.), *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (pp. 207–232). Springer. https://doi.org/10.1007/978-3-540-35306-5_10
- Chin, C.-S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, 10(6), 563–569. <https://doi.org/10.1038/nmeth.2474>
- Corradi, N. (2015). Microsporidia: Eukaryotic Intracellular Parasites Shaped by Gene Loss and Horizontal Gene Transfers. *Annual Review of Microbiology*, 69(1), 167–183. <https://doi.org/10.1146/annurev-micro-091014-104136>
- Corradi, N., Haag, K. L., Pombert, J.-F., Ebert, D., & Keeling, P. J. (2009). Draft genome sequence of the *Daphnia* pathogen *Octospora bayeri*: Insights into the gene content of a large microsporidian genome and a model for host-parasite interactions. *Genome Biology*, 10(10), R106. <https://doi.org/10.1186/gb-2009-10-10-r106>
- Crawford, J. E., & Lazzaro, B. P. (2012). Assessing the Accuracy and Power of Population Genetic Inference from Low-Pass Next-Generation Sequencing Data. *Frontiers in Genetics*, 3. <https://doi.org/10.3389/fgene.2012.00066>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- de Albuquerque, N. R. M., Ebert, D., & Haag, K. L. (2020). Transposable element abundance correlates with mode of transmission in microsporidian parasites. *Mobile DNA*, 11, 19. <https://doi.org/10.1186/s13100-020-00218-8>
- Decaestecker, E., Declerck, S., De Meester, L., & Ebert, D. (2005). Ecological implications of parasites in natural *Daphnia* populations. *Oecologia*, 144(3), 382–390. <https://doi.org/10.1007/s00442-005-0083-7>

- Dellicour, S., Durkin, K., Hong, S. L., Vanmechelen, B., Martí-Carreras, J., Gill, M. S., Meex, C., Bontems, S., André, E., Gilbert, M., Walker, C., Maio, N. D., Faria, N. R., Hadfield, J., Hayette, M.-P., Bours, V., Wawina-Bokalanga, T., Artesi, M., Baele, G., & Maes, P. (2021). A Phyldynamic Workflow to Rapidly Gain Insights into the Dispersal History and Dynamics of SARS-CoV-2 Lineages. *Molecular Biology and Evolution*, 38(4), 1608–1613. <https://doi.org/10.1093/molbev/msaa284>
- Dray, S., Bauman, D., Blanchet, G., Borcard, D., Clappe, S., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi, N., & Wagner, H. H. (2021). *adespatial: Multivariate Multiscale Spatial Analysis* (R package version 0.3-14) [Computer software]. <https://CRAN.R-project.org/package=adespatial>
- Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Molecular Biology and Evolution*, 22(5), 1185–1192. <https://doi.org/10.1093/molbev/msi103>
- Ebert, D. (2005). *Ecology, epidemiology, and evolution of parasitism in Daphnia*. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information.
- Ebert, D. (2008). Host–parasite coevolution: Insights from the *Daphnia*–parasite model system. *Current Opinion in Microbiology*, 11(3), 290–301. <https://doi.org/10.1016/j.mib.2008.05.012>
- Ebert, D., & Fields, P. D. (2020). Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, 21, 754–768. <https://doi.org/10.1038/s41576-020-0269-1>
- Ebert, D., Hottinger, J. W., & Pajunen, V. I. (2001). Temporal and Spatial Dynamics of Parasite Richness in a *Daphnia* Metapopulation. *Ecology*, 82(12), 3417–3434. [https://doi.org/10.1890/0012-9658\(2001\)082\[3417:TASDOP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[3417:TASDOP]2.0.CO;2)
- Eyre-Walker, A., & Keightley, P. D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution*, 26(9), 2097–2108. <https://doi.org/10.1093/molbev/msp119>
- Fahrenholz, H. (1913). Ectoparasiten und Abstammungslehre. *Zoologischer Anzeiger*, 41, 371–374.
- Fields, P. D., Obbard, D. J., McTaggart, S. J., Galimov, Y., Little, T. J., & Ebert, D. (2018). Mitogenome phylogeographic analysis of a planktonic crustacean. *Molecular Phylogenetics and Evolution*, 129, 138–148. <https://doi.org/10.1016/j.ympev.2018.06.028>
- Fields, P. D., Reisser, C., Dukić, M., Haag, C. R., & Ebert, D. (2015). Genes mirror geography in *Daphnia magna*. *Molecular Ecology*, 24(17), 4521–4536. <https://doi.org/10.1111/mec.13324>
- Gagneux, S., & Small, P. M. (2007). Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *The Lancet Infectious Diseases*, 7(5), 328–337. [https://doi.org/10.1016/S1473-3099\(07\)70108-1](https://doi.org/10.1016/S1473-3099(07)70108-1)
- Galtier, N. (2016). Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLOS Genetics*, 12(1), e1005774. <https://doi.org/10.1371/journal.pgen.1005774>
- Garrison, E. (2019). *VcfliB* [C++]. vcfliB. <https://github.com/vcfliB/vcfliB>
- 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. *Infection, Genetics and Evolution*, 31, 87–94. <https://doi.org/10.1016/j.meegid.2015.01.002>
- Goren, L., & Ben-Ami, F. (2013). Ecological correlates between cladocerans and their endoparasites from permanent and rain pools: Patterns in community composition and diversity. *Hydrobiologia*, 701(1), 13–23. <https://doi.org/10.1007/s10750-012-1243-5>
- Haag, K. L., Larsson, J. I. R., Refardt, D., & Ebert, D. (2011). Cytological and molecular description of *Hamiltosporidium tvaerminnensis* gen. Et sp. Nov., a microsporidian parasite of *Daphnia magna*, and establishment of *Hamiltosporidium magnivora* comb. Nov. *Parasitology*, 138(4), 447–462. <https://doi.org/10.1017/S0031182010001393>

- Haag, K. L., Pombert, J.-F., Sun, Y., de Albuquerque, N. R. M., Batliner, B., Fields, P., Lopes, T. F., & Ebert, D. (2020). Microsporidia with Vertical Transmission Were Likely Shaped by Nonadaptive Processes. *Genome Biology and Evolution*, 12(1), 3599–3614. <https://doi.org/10.1093/gbe/evz270>
- Haag, K. L., Sheikh-Jabbari, E., Ben-Ami, F., & Ebert, D. (2013). Microsatellite and single-nucleotide polymorphisms indicate recurrent transitions to asexuality in a microsporidian parasite. *Journal of Evolutionary Biology*, 26(5), 1117–1128. <https://doi.org/10.1111/jeb.12125>
- Haag, K. L., Traunecker, E., & Ebert, D. (2013). Single-nucleotide polymorphisms of two closely related microsporidian parasites suggest a clonal population expansion after the last glaciation. *Molecular Ecology*, 22(2), 314–326. <https://doi.org/10.1111/mec.12126>
- Haller, B. C., & Messer, P. W. (2017). asymptoticMK: A Web-Based Tool for the Asymptotic McDonald–Kreitman Test. *G3: Genes, Genomes, Genetics*, 7(5), 1569–1575. <https://doi.org/10.1534/g3.117.039693>
- Hecht, L. B. B., Thompson, P. C., & Rosenthal, B. M. (2018). Comparative demography elucidates the longevity of parasitic and symbiotic relationships. *Proceedings of the Royal Society B: Biological Sciences*, 285(1888). <https://doi.org/10.1098/rspb.2018.1032>
- Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58(3), 247–276. <https://doi.org/10.1006/bijl.1996.0035>
- Holder, K., Montgomerie, R., & Friesen, V. L. (1999). A Test of the Glacial Refugium Hypothesis Using Patterns of Mitochondrial and Nuclear Dna Sequence Variation in Rock Ptarmigan (*Lagopus mutus*). *Evolution*, 53(6), 1936–1950. <https://doi.org/10.1111/j.1558-5646.1999.tb04574.x>
- James, T. Y., Pelin, A., Bonen, L., Ahrendt, S., Sain, D., Corradi, N., & Stajich, J. E. (2013). Shared signatures of parasitism and phylogenomics unite Cryptomycota and microsporidia. *Current Biology: CB*, 23(16), 1548–1553. <https://doi.org/10.1016/j.cub.2013.06.057>
- Janssen, C. S., Phillips, R. S., Turner, C. M. R., & Barrett, M. P. (2004). *Plasmodium* interspersed repeats: The major multigene superfamily of malaria parasites. *Nucleic Acids Research*, 32(19), 5712–5720. <https://doi.org/10.1093/nar/gkh907>
- Johri, P., Riall, K., Becher, H., Excoffier, L., Charlesworth, B., & Jensen, J. D. (2021). The Impact of Purifying and Background Selection on the Inference of Population History: Problems and Prospects. *Molecular Biology and Evolution*, msab050. <https://doi.org/10.1093/molbev/msab050>
- Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Keightley, P. D., & Eyre-Walker, A. (2007). Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*, 177(4), 2251–2261. <https://doi.org/10.1534/genetics.107.080663>
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research*, 27(5), 722–736. <https://doi.org/10.1101/gr.215087.116>
- Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and

- divergence in the presence of missing data. *Molecular Ecology Resources*, 21(4), 1359–1368.
<https://doi.org/10.1111/1755-0998.13326>
- Kramer, J. P. (1965). Generation time of the microsporidian *Octosporea muscaedomesticae* flu in adult *Phormia regina* (Meigen) (Diptera, Calliphoridae). *Zeitschrift Für Parasitenkunde*, 25, 309–313. <https://doi.org/10.1007/BF00329585>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Lange, B., Kaufmann, A. P., & Ebert, D. (2015). Genetic, ecological and geographic covariables explaining host range and specificity of a microsporidian parasite. *Journal of Animal Ecology*, 84(6), 1711–1719. <https://doi.org/10.1111/1365-2656.12421>
- Langerhans, R. B. (2008). Coevolution. In B. Fath (Ed.), *Encyclopedia of Ecology* (1st ed., pp. 644–648). Elsevier.
- Legendre, P., Fortin, M.-J., & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, 6(11), 1239–1247. <https://doi.org/10.1111/2041-210X.12425>
- Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-News*, 6, 8–12.
- Levicoy, D., Flores, K., Rosenfeld, S., & Cárdenas, L. (2021). Phylogeography and genetic diversity of the microbivalve *Kidderia subquadrata*, reveals new data from West Antarctic Peninsula. *Scientific Reports*, 11(1), 5705. <https://doi.org/10.1038/s41598-021-85042-7>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. <http://arxiv.org/abs/1303.3997>
- Li, H., & Durbin, R. (2011). Inference of Human Population History From Whole Genome Sequence of A Single Individual. *Nature*, 475(7357), 493–496. <https://doi.org/10.1038/nature10231>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, 13(9), 2178–2189. <https://doi.org/10.1101/gr.1224503>
- Liu, S., & Hansen, M. M. (2017). PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular Ecology Resources*, 17(4), 631–641. <https://doi.org/10.1111/1755-0998.12606>
- Long, H., Behringer, M. G., Williams, E., Te, R., & Lynch, M. (2016). Similar Mutation Rates but Highly Diverse Mutation Spectra in Ascomycete and Basidiomycete Yeasts. *Genome Biology and Evolution*, 8(12), 3815–3821. <https://doi.org/10.1093/gbe/evw286>
- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. In D. J. Russell (Ed.), *Multiple Sequence Alignment Methods* (pp. 155–170). Humana Press. https://doi.org/10.1007/978-1-62703-646-7_10
- Mangin, K. L., Lipsitch, M., & Ebert, D. (1995). Virulence and transmission modes of two microsporidia in *Daphnia magna*. *Parasitology*, 111(2), 133–142. <https://doi.org/10.1017/S0031182000064878>
- Martin, M., Patterson, M., Garg, S., Fischer, S. O., Pisanti, N., Klau, G. W., Schöenhuth, A., & Marschall, T. (2016). WhatsHap: Fast and accurate read-based phasing. *BioRxiv*, 085050. <https://doi.org/10.1101/085050>

- Mattle-Greminger, M. P., Bilgin Sonay, T., Nater, A., Pybus, M., Desai, T., de Valles, G., Casals, F., Scally, A., Bertranpetit, J., Marques-Bonet, T., van Schaik, C. P., Anisimova, M., & Krützen, M. (2018). Genomes reveal marked differences in the adaptive evolution between orangutan species. *Genome Biology*, 19(1), 193. <https://doi.org/10.1186/s13059-018-1562-6>
- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 351(6328), 652–654. <https://doi.org/10.1038/351652a0>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Montano, V., Didelot, X., Foll, M., Linz, B., Reinhardt, R., Suerbaum, S., Moodley, Y., & Jensen, J. D. (2015). Worldwide Population Structure, Long-Term Demography, and Local Adaptation of *Helicobacter pylori*. *Genetics*, 200(3), 947–963. <https://doi.org/10.1534/genetics.115.176404>
- Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H. M., Breurec, S., Wu, J.-Y., Maady, A., Bernhöft, S., Thiberge, J.-M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D. Y., Marshall, B. J., & Achtman, M. (2009). The peopling of the Pacific from a bacterial perspective. *Science (New York, N.Y.)*, 323(5913), 527–530. <https://doi.org/10.1126/science.1166083>
- Murareanu, B. M., Sukhdeo, R., Qu, R., Jiang, J., & Reinke, A. W. (2021). Generation of a Microsporidia Species Attribute Database and Analysis of the Extensive Ecological and Phenotypic Diversity of Microsporidia. *MBio*, 12(3), e01490-21. <https://doi.org/10.1128/mBio.01490-21>
- Murga-Moreno, J., Coronado-Zamora, M., Hervas, S., Casillas, S., & Barbadilla, A. (2019). iMKT: The integrative McDonald and Kreitman test. *Nucleic Acids Research*, 47(W1), W283–W288. <https://doi.org/10.1093/nar/gkz372>
- Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, 76(10), 5269–5273.
- Nieberding, C. M., Durette-Desset, M.-C., Vanderpoorten, A., Casanova, J. C., Ribas, A., Deffontaine, V., Feliu, C., Morand, S., Libois, R., & Michaux, J. R. (2008). Geography and host biogeography matter for understanding the phylogeography of a parasite. *Molecular Phylogenetics and Evolution*, 47(2), 538–554. <https://doi.org/10.1016/j.ympev.2008.01.028>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O’Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *vegan: Community Ecology Package* (R package version 2.5-7) [Computer software]. <https://CRAN.R-project.org/package=vegan>
- Orlansky, S., & Ben-Ami, F. (2019). Genetic resistance and specificity in sister taxa of *Daphnia*: Insights from the range of host susceptibilities. *Parasites & Vectors*, 12(1), 545. <https://doi.org/10.1186/s13071-019-3795-y>
- Orsini, L., Mergeay, J., Vanoverbeke, J., & De Meester, L. (2013). The role of selection in driving landscape genomic structure of the waterflea *Daphnia magna*. *Molecular Ecology*, 22(3), 583–601. <https://doi.org/10.1111/mec.12117>
- Orsini, L., Spanier, K. I., & DE Meester, L. (2012). Genomic signature of natural and anthropogenic stress in wild populations of the waterflea *Daphnia magna*: Validation in space, time and experimental evolution. *Molecular Ecology*, 21(9), 2160–2175. <https://doi.org/10.1111/j.1365-294X.2011.05429.x>
- Padgham, M., & Sumner, M. D. (2019). *geodist: Fast, Dependency-Free Geodesic Distance Calculations*. <https://CRAN.R-project.org/package=geodist>.
- Page, R. D. M. (2003). *Tangled Trees: Phylogeny, Cospeciation, and Coevolution*. University of

- Chicago Press.
- Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., El Alaoui, H., Biron, D. G., Brasset, E., Vauray, C., Peyret, P., Corradi, N., Peyrettaillade, É., & Lerat, E. (2014). Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biology and Evolution*, 6(9), 2289–2300. <https://doi.org/10.1093/gbe/evu178>
- Pavlidis, P., Hutter, S., & Stephan, W. (2008). A population genomic approach to map recent positive selection in model species. *Molecular Ecology*, 17(16), 3585–3598. <https://doi.org/10.1111/j.1365-294X.2008.03852.x>
- 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*. *Environmental Microbiology*, 17(11), 4443–4458. <https://doi.org/10.1111/1462-2920.12883>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An Efficient Swiss Army Knife for Population Genomic Analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Pombert, J.-F., Haag, K. L., Beidas, S., Ebert, D., & Keeling, P. J. (2015). The *Ordospora colligata* Genome: Evolution of Extreme Reduction in Microsporidia and Host-To-Parasite Horizontal Gene Transfer. *MBio*, 6(1), e02400-14. <https://doi.org/10.1128/mBio.02400-14>
- 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*. *Eukaryotic Cell*, 12(4), 503–511. <https://doi.org/10.1128/EC.00312-12>
- Quinlan, A. R. (2014). BEDTools: The Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics*, 47, 11.12.1-11.12.34. <https://doi.org/10.1002/0471250953.bi1112s47>
- R Core Team. (2018). *R: The R Project for Statistical Computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ranwez, V., Harispe, S., Delsuc, F., & Douzery, E. J. P. (2011). MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons. *PLOS ONE*, 6(9), e22594. <https://doi.org/10.1371/journal.pone.0022594>
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 460. <https://doi.org/10.1186/s12859-018-2485-7>
- Rousselle, M., Simion, P., Tilak, M.-K., Figuet, E., Nabholz, B., & Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLOS Genetics*, 16(4), e1008668. <https://doi.org/10.1371/journal.pgen.1008668>
- Sandrock, C., Schirrmeister, B. E., & Vorburger, C. (2011). Evolution of reproductive mode variation and host associations in a sexual-asexual complex of aphid parasitoids. *BMC Evolutionary Biology*, 11(1), 348. <https://doi.org/10.1186/1471-2148-11-348>
- Seefeldt, L., & Ebert, D. (2019). Temperature- versus precipitation-limitation shape local temperature tolerance in a Holarctic freshwater crustacean. *Proceedings of the Royal Society B*, 286(1907), 20190929. <https://doi.org/10.1098/rspb.2019.0929>
- Sendell-Price, A. T., Ruegg, K. C., Robertson, B. C., & Clegg, S. M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*, 30(11),

- 2495–2510. <https://doi.org/10.1111/mec.15898>
- Seppy, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing Genome Assembly and Annotation Completeness. In M. Kollmar (Ed.), *Gene Prediction: Methods and Protocols* (pp. 227–245). Springer. https://doi.org/10.1007/978-1-4939-9173-0_14
- Shafer, A. B. A., Williams, G. R., Shutler, D., Rogers, R. E. L., & Stewart, D. T. (2009). Cophylogeny of *Nosema* (Microsporidia: Nosematidae) and Bees (Hymenoptera: Apidae) Suggests Both Cospeciation and a Host-switch. *Journal of Parasitology*, 95(1), 198–203. <https://doi.org/10.1645/GE-1724.1>
- Stewart, J. R., Lister, A. M., Barnes, I., & Dalén, L. (2010). Refugia revisited: Individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, 277(1682), 661–671. <https://doi.org/10.1098/rspb.2009.1272>
- Stollewerk, A. (2010). The water flea *Daphnia*—A “new” model system for ecology and evolution? *Journal of Biology*, 9(2), 21. <https://doi.org/10.1186/jbiol212>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 11(1110), 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Wadi, L., & Reinke, A. W. (2020). Evolution of microsporidia: An extremely successful group of eukaryotic intracellular parasites. *PLOS Pathogens*, 16(2), e1008276. <https://doi.org/10.1371/journal.ppat.1008276>
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE*, 9(11), e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Wang, H.-Y., Qi, M., Sun, M.-F., Li, D.-F., Wang, R.-J., Zhang, S.-M., Zhao, J.-F., Li, J.-Q., Cui, Z.-H., Chen, Y.-C., Jian, F.-C., Xiang, R.-P., Ning, C.-S., & Zhang, L.-X. (2019). Prevalence and Population Genetics Analysis of *Enterocytozoon bieneusi* in Dairy Cattle in China. *Frontiers in Microbiology*, 10, 1399. <https://doi.org/10.3389/fmicb.2019.01399>
- Weir, W., Capewell, P., Foth, B., Clucas, C., Pountain, A., Steketee, P., Veitch, N., Koffi, M., De Meeüs, T., Kaboré, J., Camara, M., Cooper, A., Tait, A., Jamonneau, V., Bucheton, B., Berriman, M., & MacLeod, A. (2016). Population genomics reveals the origin and asexual evolution of human infective trypanosomes. *ELife*, 5, e11473. <https://doi.org/10.7554/eLife.11473>
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: Master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), 741–751. <https://doi.org/10.1038/nrmicro1969>
- Williams, B. A., Lee, R. C., Becnel, J. J., Weiss, L. M., Fast, N. M., & Keeling, P. J. (2008). Genome sequence surveys of *Brachiola algerae* and *Edhazardia aedis* reveal microsporidia with low gene densities. *BMC Genomics*, 9(1), 200. <https://doi.org/10.1186/1471-2164-9-200>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

Data accessibility

1011 All analysis scripts as well as raw and processed data are available at
 1012 https://github.com/pascalangst/Angst_etal_2021_MolEcol, NCBI SRA database (BioProject IDs
 1013 XXXX), and NCBI GenBank, Zenodo XXXX (see git repo for all accessions).

1014

1015 **Author contributions**

1016 All authors designed the study. PA analysed the data and wrote the manuscript. All authors reviewed
 1017 the manuscript.

1018

1019

Fig. 1: Geophylogeny of *Hamiltosporidium* spp. with co-phylogeny to the *D. magna* host. Both phylogenetic trees are based on four-fold degenerate sites but are not on the same scale. Colored lines indicate the main clusters derived from our population structure analyses: *H. magnivora* (orange), Northern *H. tvaerminnensis* (dark blue), Southern *H. tvaerminnensis* (cyan), and connect the tree tips of *Hamiltosporidium* spp. (above) and *D. magna* (below) to the sampling locality. Eurasia appears as an orthographic projection with numbers indicating latitudes and longitudes. Host and *H. tvaerminnensis* phylogenetic trees are mostly congruent, except that East Asian samples of the parasite cluster with Northern European samples rather than forming a separate clade as in the host.

Fig. 2: Unfolded site frequency spectra of (A) all *H. tvaerminnensis* samples, (B) all *H. magnivora* samples, (C) the Northern and (D) the Southern *H. tvaerminnensis* lineage. Distribution of the allele frequencies in the single copy orthologs is shown after identifying each position as either ancestral, i.e., invariant to the outgroup, or derived, i.e., different from the outgroup. Non-/synonymous substitutions are shown separately. Of particular interest is the peak at intermediate frequencies in the Northern lineage, as it represents mostly fixed heterozygote sites originating from the transition from sexual to asexual reproduction.

1041

Fig. 3: PSMC results of *Hamiltosporidium* samples. The effective population size history of *Hamiltosporidium* samples estimated using PSMC (A). Each line represents one sample, with colors indicating its cluster affiliation according to our population structure analyses. Samples from Spain (Southern *H. tvaerminnensis* cluster) show decreases in N_e compared with the other samples from the Southern *H. tvaerminnensis* cluster, which are stable over the same time period. A base-10 log scale is used for the x-axis, which is time calibrated using a rough approximation. No extreme upward vertical trajectories are observed in contrast to (B), which shows the PSMC results of *Hamiltosporidium* "pseudo-diploid" samples. In (B), each line represents one *in silico*, "pseudo-diploid" sample, i.e., single phased haplotypes of our diploid samples combined in a pairwise manner. Orange indicates an intraspecies combination of *H. magnivora* (*H.m.*) haplotypes; blue shows intraspecies combinations of *H. tvaerminnensis* (*H.t.*) haplotypes, and green depicts interspecies combinations of *H. tvaerminnensis* (*H.t.*) haplotypes and *H. magnivora* (*H.m.*) haplotypes. Here, the Northern *H. tvaerminnensis* lineage is abbreviated as *H.t.-N*, and the Southern *H. tvaerminnensis* lineage as *H.t.-S*. The *H. tvaerminnensis* intraspecies combinations do not show extreme upward vertical trajectories until very recently, which suggests a rather young species rather than a diploid species that has been reproducing asexually for a long time. Only one "pseudo-diploid" sample turns downwards after an extreme upward vertical trajectory, the interspecies haplotype combination of Northern *H. tvaerminnensis* and *H. magnivora* haplotypes, which suggests gene flow between the two. This is indicated by the dotted light green line, which is drawn out of scale for the ease of presentation. The axes cover the same ranges in both graphs.

1063

1064

1065

1066

1067

1068

1069

1070

1071

1072

1073

1074

Fig. 4: Putative evolutionary change in the *Hamiltosporidium* lifecycle. 1) The ancestral lifecycle of *Hamiltosporidium* might have involved asexual vertical transmission to offspring and transmission to a second host where sexual reproduction could take place, as is the case for *H. magnivora*. 2) However, a novel transmission strategy arose wherein horizontal transmission amongst the host *D. magna* without sexual reproduction 3) allowing this derived *Hamiltosporidium* group to colonize geographical regions where the second host is absent, as is the case for *H. tvaerminnensis*. Sexual reproduction would then only be possible in the presence of the now facultative second host.

Tables

Table 1: Sample information. Average whole-genome coverage for the parasites is denoted in the last column.

Sample ID	Country	Latitude	Longitude	Parasite species	Coverage
BE-OM-2	Belgium	50.863280	4.721370	<i>H. magnivora</i>	594×
BE-T1-2	Belgium	50.823050	4.593940	<i>H. magnivora</i>	145×
BE-WH2-11	Belgium	51.334780	3.348130	<i>H. magnivora</i>	149×
CY-PA3-2	Cyprus	35.034060	33.954930	<i>H. tvaerminnensis</i>	598×
ES-DO1-1	Spain	36.978360	-6.477640	<i>H. tvaerminnensis</i>	282×
ES-OM-5	Spain	37.254353	-6.969114	<i>H. tvaerminnensis</i>	348×
FI-BR1-39-6	Finland	59.845253	23.274526	<i>H. tvaerminnensis</i>	219×
FI-FHS2-11-3	Finland	60.273780	27.218750	<i>H. tvaerminnensis</i>	94×
FI-OER-3-3	Finland	59.788590	23.174140	<i>H. tvaerminnensis</i>	1097×
IL-G-3	Israel	32.228840	34.831140	<i>H. tvaerminnensis</i>	703×
IL-PS-6	Israel	32.255710	34.853270	<i>H. tvaerminnensis</i>	689×
IL-SH-4	Israel	31.701030	34.713760	<i>H. tvaerminnensis</i>	600×
RU-BAYA1-1	Russia	53.017833	106.886833	<i>H. tvaerminnensis</i>	756×
RU-TY1-3	Russia	50.255500	89.546830	<i>H. tvaerminnensis</i>	389×
RU-TY2-2	Russia	50.212170	89.717830	<i>H. tvaerminnensis</i>	523×
RU-ZB1-1	Russia	50.413170	114.724703	<i>H. tvaerminnensis</i>	374×
SE-G4-20	Sweden	60.253040	18.306090	<i>H. tvaerminnensis</i>	562×
SE-H1-4	Sweden	58.342330	11.218030	<i>H. tvaerminnensis</i>	336×

Table 2: Proportion of adaptive amino-acid substitutions (α) estimated for different *H. tvaerminnensis* lineages with McDonald–Kreitman tests (MKT) and dfe-alpha. α derived from Standard MKT presented with p-values (Fisher’s exact test), and α derived from dfe-alpha with a CI derived from 1,000 bootstrap samples. Rate of adaptive divergence relative to neutral divergence (ω_A), also with a CI derived from 1,000 bootstrap samples, was estimated using dfe-alpha.

Lineage	Standard MKT					
	α	P-value				
<i>H. tvaerminnensis</i> - all	0.227	<0.001				
<i>H. tvaerminnensis</i> - north	0.121	<0.001				
<i>H. tvaerminnensis</i> - south	0.202	<0.001				

Lineage	dfe-alpha					
	α	α low	α high	ω_A	ω_A low	ω_A high
<i>H. tvaerminnensis</i> - all	0.203	0.142	0.272	0.066	0.041	0.093
<i>H. tvaerminnensis</i> - north	-0.026	-0.104	0.061	-0.008	-0.030	0.021
<i>H. tvaerminnensis</i> - south	0.139	0.004	0.156	0.047	0.001	0.054