"Parasite turnover zone" at secondary contact: a new pattern in host-parasite population genetics.

Jana Martinu[1], Jan Stefka[2], Anbu Poosakkannu[1], and Václav Hypša[1]

[1]University of South Bohemia Faculty of Science
[2]University of South Bohemia

July 2, 2020

Abstract

We introduce a new pattern of population genetic structure in a host-parasite system that can arise after secondary contact of previously isolated populations. Due to different generation time and therefore different tempo of molecular evolution the host and parasite populations reach different degrees of genetic differentiation during their separation (e.g. in refugia). Consequently, during the secondary contact the host populations are able to re-establish a single panmictic population across the area of contact, while the parasite populations stop their dispersal at the secondary contact zone and create a narrow hybrid zone. From the host's perspective, the parasite's hybrid zone functions on a microevolutionary scale as a "parasite turnover zone": while the hosts are passing from area A to area B, their parasites turn genetically from the area A genotypes to the area B genotypes. We demonstrate this novel pattern on a model composed of Apodemus mice and Polyplax lice by comparing maternally inherited markers (complete mitochondrial genomes, and complete genomes of vertically transmitted symbiont Legionella polyplacis) with SNPs derived from the louse genomic data. We discuss circumstances that may lead to this pattern and possible reasons why it has been overlooked in the studies on host-parasite population genetics.

Keywords

host-parasite, secondary contact, genomics, *Polyplax serrata* ,

Apodemus flavicollis

Introduction

Population genetics of host-parasite associations are complex and dependent on many ecological features of both counterparts (Criscione, Poulin, & Blouin, 2005; Barrett, Thrall, Burdon, & Linde, 2008; Sweet & Johnson, 2018). Within hybrid zones and secondary contact zones this picture is likely to become even more complicated, possibly giving rise to new unexpected patterns. Unfortunately, very few studies have been devoted to this aspect of host-parasite interactions. From the most general point of view, it is assumed that since parasite is dependent on the host, its genetic structure will tend to mirror the host. In this respect, two assumptions are frequently expressed. First, the degree of congruence with the host is dependent on traits connected to the parasitic life-style, such as the degree of host-specificity, transmission mode, presence of dispersal stages, etc. (Maze-Guilmo, Blanchet, McCoy, & Loot, 2016). Generally, the more intimate the association, the higher the degree of congruence. However, this general view may be distorted by many specific traits of the particular host-parasite association. For example, the population structure of heteroxenous parasites (parasites with more than one host in their life cycles) is likely to reflect the least structured host, since any potential structure is erased by the more motile host (Jarne & Theron, 2001; Louhi, Karvonen, Rellstab, & Jokela, 2010). Similarly, with longer free living stage(s), the genetic structures of the host and the parasite become more incongruent (Jarne & Theron, 2001). However, phylogenetic incongruency was demonstrated even in homoxenous highly specific ectoparasites with direct life cycle, for example chewing

lice, due to species replacement (Hafner et al., 2019) or sucking lice, due to duplications, sortings, and host switches (du Toit et al., 2013). The second assumption is about the speed of diversification: since the parasites have a shorter generation time, they undergo faster genetic diversification, which may eventually lead to the parasite's duplication (i.e. formation of two sister species living on single host species; Page, Lee, Becher, Griffiths, & Clayton, 1998; scenario *a* in Figure 1). The assumption about the higher mutation rate in parasites was demonstrated in several studies (Nieberding, Morand, Libois, & Michaux, 2004; for ectoparasites: McCoy et al., 2005; Whiteman, Kimball, & Parker, 2007; Štefka et al., 2011; Johnson et al., 2014, but see Gómez-Díaz, González-Solís, Peinado, & Page, 2007; Jones & Britten, 2010 for the opposite results).

Although many studies have been devoted to comparing phylogenies and population structures of host-parasite associations, only a few analyzed these processes in connection to secondary contact zone and hybrid zone of the hosts (reviewed by Theodosopoulos, Hund, & Taylor, 2019), and only recently, de Bellocq et al. (2018) focused on detecting hybrid zone in parasite populations. Using two parasites of the house mouse *Mus musculus* , the nematode *Syphacia obvelata* and the fungus *Pneumocystis murina* , they found that within the host's hybrid zone both parasites create their own hybrid zone. They also demonstrated that the parasites (reaching higher genetic divergence) created significantly narrower hybrid zones than the host (scenario *b* in the Figure 1).

From a theoretical point of view, the assumptions and empirical evidence discussed above lead to a third possible scenario: during secondary contact the host does not create a hybrid zone but rather re-establishes a panmictic population, while the parasite accumulates a degree of genetic differences which prevents re-establishment of panmixia but does not lead to a complete speciation with prezygotic barriers (scenario *c* in the Figure 1). A paradoxical result of such an event would be the establishment of a parasite's hybrid zone within the host's panmictic population, which on a microevolutionary scale would function as a "parasite turnover zone": while the hosts are passing through this zone from area A to area B (Figure 1d), their parasites turn from the area A genotypes to the area B genotypes. To our knowledge, such "filter" has never been observed in nature. In fact, the presence of a parasite's hybrid zone in the scenario described here is difficult to guess *a priori* , as it is not indicated by the host's hybrid zone. However, in our previous work (Martinů, Hypša, & Štefka, 2018) we presented the genetic structure of postglacial Europe recolonization by the mice of the genus *Apodemus* and their ectoparasite, the louse *Polyplax serrata* , which corresponds to such scenario (Figure 2; see below for details).

Similar to all sucking lice, *P. serrata* is a permanent homoxenous ectoparasite with strict host specificity, which is transmitted almost exclusively during physical contact of its hosts. As such it falls into the category of highly intimate parasites displaying a high degree of congruence with their hosts. In Figure 2, we summarize the main features of the population genetic pattern obtained by the analysis of 379 bp mitochondrial haplotypes (Martinů et al., 2018). It shows that *P. serrata* is composed of several genetic lineages (Figure 2d) with different host-specificities and geographic distributions. This indicates that even such traits as the degree of host specificity may be very flexible and change rapidly at a shallow phylogenetic level. For example, the so-called *specific* (S) and *non-specific* (N) lineages, although closely related (sister lineages) and living in sympatry, differ in degree of their specificities, one being exclusive to *Apodemus flavicollis,* while the other can also live on *A. sylvaticus* . However, the most intriguing part of the pattern was detected within the S lineage. On the mtDNA based phylogenetic trees, the host (*A. flavicollis*) and the parasite (S-lineage of *P. serrata*), display the same basic structure. Their samples collected across all of Europe form two genetically distant clusters, suggesting recolonization from two different refugia (the taxa designated by red and blue colours in Figure 2; see Martinů et al., 2018 for discussion). However, while the two host's clusters have already spread across the entirety of Europe, their lice did not follow the same process. Instead, their two sub-lineages, designated as *specific east* (SE) and *specific west* (SW), ceased their dispersion after reaching the secondary contact zone in the middle of Europe (Figure 2). This disparity is surprising since the high intimacy of lice should predetermine them to mirror genetic structure of the host (e.g. Harper, Spradling, Demastes, & Calhoun, 2015 Lack of strong structure in *A. flavicollis* populations in the area of secondary contact is suggested also by recent SNP based studies. Martin Cerezo et al. (2020) found negligible pop-

2

ulation structure (pairwise F_{ST}<0.086) between three populations located up to 500 km apart in northern Poland. No suture in population structure in the west-east direction was found also in our RAD-seq dataset examining the population history of *A. flavicollis* across Europe (MS in preparation).

To obtain a more complete picture of secondary contact in *P. serrata* , in this study we analyze three patterns derived from metagenomic data of 26 louse specimens collected across the secondary contact zone: nuclear SNPs, complete mitochondrial genomes, and complete genomes of the symbiotic bacterium *Legionella polyplacis* . We use these analyses to retrieve two kinds of information. First, we compare nuclear (SNP) and maternally inherited markers (mitochondrial genomes, symbiont genomes) to demonstrate a narrow hybrid zone between the SW and SE lineages of the lice. Second, we address two possible causes of the SE/SW incompatibility suggested in our previous work (Martinů et al., 2018). *P. serrata* carries the intracellular obligate symbiont*Legionella polyplacis* (Říhová, Nováková, Husník, & Hypša, 2017) which could be incompatible with the non-native genetic background. Similarly, since the *Polyplax* louse mitochondria are fragmented into 11 minichromosomes (Dong, Song, Jin, Guo, & Shao, 2014) a rearrangement of their genetic composition could theoretically lead to the SE/SW incompatibility. We, therefore, compare complete mitochondrial and symbiont genomes to assess the degree of their divergence.

Methods

Origin of the samples

Apodemus flavicollis mice were captured across the hybrid zone of*Polyplax serrata* -specific lineages in 2018 and 2019 in the north-west of the Czech Republic, using wooden snap traps. Field studies were carried out with permits approved by the Committee on the Ethics of Animal Experiments of the University of South Bohemia, by the Ministry of the Environment of the Czech Republic, and by the Ministry of the Agriculture of the Czech Republic (No. 51304/ENV/14-2981/630/14, MZP/2017/630/854, 22395/2014-MZE-17214). Mice were searched for lice by visual checking and combing. Lice were removed and stored in 100% ethanol in the -20°C. Genomic DNA from whole louse specimens was individually extracted using the Qiagen QIAamp DNA Micro Kit (Qiagen, Valencia, CA, USA). Membership to individual louse lineages (SE, SW, or N) was assigned by sequencing a fragment of the mitochondrial cytochrome oxidase subunit I gene (COI, 379 bp) as in Martinů et al (2018). All sequences of the S lineage available from previous work together with the newly gained specimens (Table S1) were collapsed into haplotypes using ALTER (*http://www.sing-group.org/ALTER/*). Then the specimens were assigned to individual lineages by phylogenetic reconstruction, using maximum likelihood method with 1000 bootstraps (Figure S1). Model GTR+I+G was used as the best-fit model, selected according to a corrected Akaike information criterion using jModelTest2 v2.1.10 (Darriba, Taboada, Doallo, & Posada, 2012; Guindon & Gascuel, 2003). Sample DBab5 (PolyN) from the N lineage was used as an outgroup. The S louse lineage was found on 24 *A. flavicollis* captured across the contact zone. Twenty-six lice (15 SE and 11 SW), 1 to maximum 3 from each parasitized host, were selected for genomic re-sequencing (Figure 3, Table 1). DNA concentrations were quantified with a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) using High Sensitivity reagents.

Generating and assembling genomic data

To obtain a reference genome sequence (of expected size approx. 110 Mbp), one *P. serrata* specimen (98c_-Pro_SE) was sequenced by Oxford Nanopore technology (ONT) on one MinION II flowcell, producing 16.7 Gbp of data (3.3 million reads of average size of 5 kbp). Library preparation and sequencing were performed at the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign, USA). The ONT reads were trimmed with filtlong (v0.2.0; https://github.com/rrwick/Filtlong) to retain sequences of at least 4000 bp and phred score of at least 20. Flye assembler (v2.5; Kolmogorov, Yuan, Lin, & Pevzner, 2019) was used to perform de novo genome assembly using ONT reads (with/without quality trimming) with the expected genome size of 110 Mbp as a parameter. The resulting assemblies were polished once with Racon (v1.3.3; https://github.com/isovic/racon) and twice with Medaka (v0.6.5; https://nanoporetech.github.io/medaka/) using the ONT reads. The final polishing was performed again with Racon, this time using Illumina reads. Quality checks of the final polished assemblies were performed with BUSCO (v3; Waterhouse et al., 2017).

Biological origin of the assembled contigs was verified by blastx (Altschul et al., 1997) of the polished assembly (98 contigs) against the *Pediculus humanus corporis* genome (GCA_000006295.1 JCVI_LOUSE_1.0). The contigs for which the blastx did not yield matches with *P. humanus corporis* genes were further analysed by blastx against nr/nt NCBI database. By this procedure, we identified non-louse origin in 68 contigs (0,46% of the assembly length), which were therefore excluded from the following analyses.

Whole genome re-sequencing was performed to generate metagenomic data used to map the SNPs, and to assemble genomes of the *L. polyplacis* symbionts and mitochondrial minichromosomes. gDNA libraries for 26 louse specimens were constructed for paired-end Illumina sequencing with the NovaSeq6000 instrument. All samples were sequenced on one Illumina Novaseq lane producing on average 59.5 million 150 bp paired-end reads (PE) per sample, resulting in coverage 18.9x - 34x for the louse contigs (see results). Libraries were constructed with an average insert size of 450bp. Fastq files were generated from the sequence data using Casava v.1.8.2 or bcltofastq v.1.8.4 with Illumina 1.9 quality score encoding. All sequencing and fastq file generation were carried out at the W. M. Keck Center (University of Illinois, Urbana, IL, USA). Illumina reads of the re-sequenced specimens were checked for quality and filtered in BBtools (https://jgi.doe.gov/data-andtools/bbtools/), then assembled using the SPAdes assembler v 3.10 (Bankevich et al., 2012), under default settings with the parameter careful, decreasing the number of mismatches and indels.

To analyze genetic structure in the area of secondary contact of the *Apodemus* -*Polyplax* populations (and the presumed hybrid zone of the parasite), we compared patterns obtained from three different sources: nuclear SNPs, mitochondrial genomes, and genomes of symbiotic bacterium *L. polyplacis* . Our hypothesis (visualized in the Figure1 c, d) predicts that for both maternally inherited markers (mitochondrial and symbiotic genomes), the inter-lineage divergence (i.e. SE vs. SW) will be much higher than any divergence within the lineages. This reflects the hypothetical scenario of two lineages isolated during their stay in two different refugia and also during most of the recolonization journey. On the contrary, since the hypothesis presumes hybridization, no such distinction should be found in nuclear SNPs across the presumed area of secondary contact; instead part of the samples will be expected to form an intermediate group representing hybrids. To visualize and quantify the diversification of maternally inherited markers, we reconstructed genealogical trees and calculated distance matrices for mitochondrial and symbiotic genomes. To visualize nuclear diversification using SNPs, we analyzed population structure using Principal Component Analysis (PCA) and Admixture software (Alexander et al., 2009). Lastly, we reconstructed genealogical relationships by building a maximum likelihood tree. Below, we provide details on processing the three sources of the data.

Nuclear SNPs

Reads from 26 louse genomes were individually mapped with Bowtie2 (Langmead & Salzberg, 2012) to the reference genome described above. Before mapping, the reference sequence was indexed using Samtools (Li et al., 2009) and a dictionary file was made with CreateSequenceDictionary in Picard v.2.0.1 (https://broadinstitute.github.io/picard/). After mapping with Bowtie2, resulting SAM files were sorted to the BAM files and indexed using Samtools. Duplicated sequences were removed from sorted BAM files with Picard v.2.0.1 and quality of mapping was verified with QualiMap (http://qualimap.bioinfo.cipf.es/). SNPs for population-level analysis were called using the GATK Genome Analysis Toolkit following the "Best Practices" guide from the Broad Institute (Van der Auwera et al., 2013). SNPs were jointly called for all 26 *P. serrata* samples and filtered with QD (quality by depth) <2.0, FS (Fisher strand test) >60.0, MQ (mapping quality) <40.0, and MQRankSum (mapping quality rank sum test) <−12.5. The SNPs from GATK were filtered with minor allele frequency (MAF) threshold 0.05 in PLINK 1.9 (https://www.cog-genomics.org/plink/1.9/). SNPs in linkage disequilibrium (LD) were pruned with the squared coefficient of correlation (r2) 0.2 and missing data threshold 0.2.

Resulting 47,595 variants passed filters and quality control in 26 *P. serrata* samples and used in population structure estimation. PCA was performed in R package SNPRelate (DOI: 10.18129/B9.bioc.SNPRelate) and phylogenetic tree based on Maximum likelihood analysis was reconstructed using SNPhylo pipeline (Lee, Guo, Wang, Kim, & Paterson, 2014) with 1000 bootstraps. Ten independent runs of admixture

4

analysis were performed to assess the proportion of individual ancestry with different numbers of hypothetical populations, using the ADMIXTURE software v.1.3.0 (Alexander et al., 2009). An unsupervised model approach based on maximum likelihood was applied to the genotype matrix for 1-10 populations (K=1-10). Most probable estimate of K was calculated with cross-validation procedure. Summarization and graphical output of independent runs for each K was obtained with CLUMPAK (Kopelman et al., 2015). Based on the obtained results samples were divided into three groups (SW, SE, and hybrids), and their diversity measures (heterozygosity and pairwise F_{ST}) were estimated in VCFtools (Danecek et al., 2011) and SNPRelate, respectively..

Legionella polyplacis *genomes reconstruction and comparison*

The contigs corresponding to *Legionella polyplacis* were visually identified using ORF prediction done in the Geneious package (the prokaryotic gene arrangement could be readily recognized by the density of predicted ORFs). In most samples, the genome of *L. polyplacis* was assembled into a single complete contig. In three assemblies (DPH41,19JA1_SW, 46MAN_SW) the symbiont genome was highly fragmented or could not be found, despite the good assembly quality of the louse genome (these samples were removed from the *L. polyplacis* analysis). Complete *L. polyplacis* genomes were annotated in RAST (Aziz et al., 2008) and aligned using Mafft algorithm implemented in Geneious. In three samples (Ne125b_Kot_SW, 99b_Pro_SE, 98c_-Pro_SE) the rRNA region did not assemble correctly and was only represented by short fragments. To extend these fragments into a full length, we used the program aTRAM 2.0 (Allen, LaFrance, Folk, Johnson, & Guralnick, 2018). Phylogenetic analysis of the resulting 530,063 bp long matrix was performed in Phyml (Guindon et al., 2010). The evolutionary model GTR+I+G was determined for the whole concatenated matrix (considering the strong phylogenetic signal in the matrix, the model selection is a purely formalistic step in the presented analysis) by a corrected Akaike information criterion in jModelTest2 (Darriba et al., 2012) based on the AIC. A comparison of gene content was done manually using the annotated alignment. We were taking into consideration the following criteria: presence of the annotations across all genomes; presence of the corresponding sequence (in case of missing annotation); distribution of differences (i.e. does the difference reflect the SE/SW split).

Minichromosome genomes reconstruction and comparison

Due to the conserved noncoding region shared by all minichromosomes, the Spades based assembly was not able to separate the different minichromosomes reliably. We, therefore, took an alternative approach. Using a single gene from each minichromosome as a reference (preliminary assignment of the genes to individual minichromosomes was based on the GenBank data available for *P. spinulosa* ; acc. nos. KF647762-KF647771) we mapped the reads and extended the sequences in the program aTRAM 2.0 (Allen at al., 2018). To obtain annotations of the resulting sequences, we combined two methods. The first method utilized the web based server Mitos (Bernt et al., 2013). Since this method missannotated or entirely missed some of the genes, we corrected the results by a blast based approach, taking advantage of the available annotated mitochondrial genome from *P. spinulosa* . We combined assembled minichromosome sequences into a custom database of *P. serrata* and used the 37 genes of *P. spinulosa* as blast queries (we run discontinous megablast and tblastx, both with E-value set to 10). We then aligned the *P. serrata* minichromosome sequences together with *P. asiatica* and *P. spinulosa* using Mauve (Darling, Mau, Blattner, & Perna, 2004). These alignments were used as a background for combining and correcting the annotations. To prepare a concatenated matrix, we trimmed all minichromosomes to equal lengths. Phylogenetic tree and genetic distances were retrieved from concatenated alignments by the same approach as for the *L. polyplacis* genomes.

Results

Gene content and arrangement of the maternally inherited markers

Mitochondrial minichromosomes . Neither the Spades nor the aTRAM method was successful in reconstructing complete sequences of the eleven mitochondrial minichromosomes. However, the aTRAM assemblies contained whole coding regions and were used for both the phylogenetic reconstruction and the comparison of gene content between the SE and SW lineages. Although yielding considerable genetic differences (Table

S2), the mitochondrial minichromosomes show an identical arrangement of the genes (shared synteny) in both the SW and SE lineages (Table S3). This arrangement is also very similar to that in the related louse species, *P. spinulosa* . Concatenation of the minichromosomes produced a 15,693 bp long matrix. When phylogenetically analyzed, it yielded a tree with two very distant clusters corresponding to the SE and SW lineages (Figures 4b, S2). Within both clusters, the distances were significantly lower than the distance between the clusters. However, the overall range of distance was higher within the SW than within the SE (Table S2, Figure 4), likely reflecting the broader geographic sampling range of the SW lineage.

Legionella. Genomes of the symbiont *L. polyplacis* revealed phylogenomic structure parallel to the mtDNA (Figures 4a, S3), with a deep genetic split between the SW and SE lineages. The complete genomes displayed a high degree of similarities with all pairwise comparisons exceeding 99% identity across the 530,063 bp matrix. The contrast between the intra- and inter-cluster comparisons is better illustrated by the counts of the observed differences, which were 215-213 within the SW cluster and 0-113 within the SE cluster, compared to 3,702 - 3,727 between the clusters (Table S2). When comparing the genome sequences, we did not find any clear instance of missing genes. The majority of the gaps introduced by genomic alignment span just one or two nucleotides and were placed in intergenic regions (only one deletion span across 26 nucleotides, also located between the gene coding sequences). The annotations provided by RAST contained several differences between the two clusters, indicating that a gene present in one lineage is shortened or missing in the other cluster. In all of these cases, however, the differences were not caused by a convincing absence of the gene sequence but rather by failure of the algorithm to recognize the sequence as coding a gene, most likely due to the aberrant nature of highly derived symbiont genomes.

Diversity and population pattern of nuclear SNPs

Illumina reads from 26 genomes were successfully mapped on the reference with mean coverage ranging from 18.9x (30VOJ_SW) to 34x (29DZ_SW). None of the assemblies had to be filtered out due to low coverage and/or low Phred scores.

Patterns of population structure obtained using nuclear SNPs differed from genealogies based on maternally inherited genomes (Figure 4). In contrast to mtDNA, the structure derived by PCA from nuclear SNPs did not recognize only two mutually exclusive clusters corresponding to SW and SE mitochondrial lineages but also a third admixed cluster containing hybrid individuals. The pattern fitted very well geographical sampling along the anticipated contact zone in a transect running from north-west (SW population 2HBC_HorBl_-SW) to the south-east (SE population 39RAD_SE). The geographical outlier, sample NE125b_Kot_SW from north-east Germany, was positioned within the hybrid samples on the main axis, whilst being separated from the rest of the samples along the other axis. This pattern may reflect a large geographical distance from the Czech localities combined with possible occurrence of a contact zone in NE Germany (Figure 4, map Figure 3).

Bayesian analysis in Admixture found the highest support for the occurrence of two clusters (see Figure S4 for cross-validation values). Each of the two clusters corresponded to one of the mtDNA lineages, with samples from the centre of the contact zone and the German specimen Ne125b_Kot showing admixed origin (Figure 4d).

ML tree-building method yielded a similar pattern to PCA (Figure S5). It revealed the two parental populations as monophyletic clusters and the hybrids forming an intermediate paraphyletic assemblage (with the long-branched Ne125b_Kot sample clustering among the hybrids; Figure S5).

Measures of genetic diversity corresponded very well with the previous methods. When divided into three populations (SW, SE, and hybrids), the highest pairwise F_{ST} value was detected between the SW and SE clusters (0.59), compared to the lower values between hybrids and SW or SE clusters (0.21 and 0.36, respectively, Table S4). Similarly, hybrids showed approximately 2 fold increase in heterozygosity compared to either of the SW and SE clusters (Table S4), as expected for individuals admixed between two lineages.

Discussion

In this study, we show a new pattern of population structure that can arise at the secondary contact zone due to different courses of evolution in hosts and parasites. The main signature of this pattern is a conflicting arrangement of mitochondrial markers in the host and the parasite at the secondary contact zone (Figure 2). The host's mitochondrial lineages, coming from different refugia, mix across the area of the secondary contact and re-establish a panmictic population. In contrast, the parasite's mitochondrial lineages stop their dispersal at the secondary contact zone. In our model, the re-established panmixia of *A. flavicollis* across the secondary contac zone is strongly suggested by a previous study on mitochondrial and microsatellite markers (Martinů et al., 2018) and further corroborated by our recent RAD-seq analysis which did not find in *A. flavicollis* any genetic disruption analogous to the SW/SE split (ms in prep). For the louse *P. serrata* , we detected a sharp geographic division between the SE and SW lineages using short (379 bp) cytochrome oxidase I (COI) haplotypes sampled across Europe (Martinů et al., 2018). To obtain a more informative comparison of genetic distance within and between the SE/SW clusters, in the present study we demonstrate this split on near-complete mitochondrial genomes from 26 samples collected across the secondary contact zone (Figures. 3, 4). From a strictly theoretical point of view, the pattern produced by the mitochondrial data can be explained by several scenarios. The first explanation is based on the strong presumption that the louse population structure will be determined entirely by the hosts' migrations, given that the lice are highly host-specific and intimate parasites. Consequently, the discrepancy shown in Figure 2 would be a sampling or methodological artifact. However, considering the geographic extent and the number of samples in our previous study (Martinů et al., 2018), we believe that a methodological artifact is a highly implausible explanation. This view is further supported by the present study of the complete mitochondrial sequences and the same genealogical pattern obtained for 23 complete genomes of the maternally inherited symbiont *L. polyplacis* (Figure 4).

A second theoretical possibility assumes that the lice speciated during their separation in refugia before secondary contact of their hosts, due to their shorter generation time. A similar case was reported by Hafner et al. (2019) for a recent secondary contact of two subspecies of pocket gophers and their lice. While the gophers established a hybrid zone, their lice had already speciated and their contact resulted in "competitive parapatry", with one louse species replacing the other. The authors also pointed out that the distribution data on the pocket gophers and their chewing lice indicate many instances of range overlap, potentially representing zones of competitive parapatry or species replacements. There are two strong arguments against applying similar scenarios to our system. A theoretical objection is that since the two *A. flavicollis* mtDNA lineages do not create a secondary contact zone or hybrid zone, but intermix across Europe, it is difficult to envisage a mechanism that would prevent dispersion of the two new louse species across the secondary contact zone. Since both louse mtDNA lineages share the same host species and live in identical ecological environments (as evidenced by sampling both lineages even from the same host individuals), their mutually exclusive distribution is obviously not due to different adaptations (i.e. different host/environment specificities). Also, competitive exclusion is a very unlikely cause as demonstrated by the frequent coexistence of the S-lineage and N-lineage (Martinů et al. 2018). An empirical argument rests on the comparison between the mtDNA and SNP data. If the two mitochondrial lineages were fully isolated non-interbreeding species, we would expect to see the same pattern (i.e. two clearly separated and distant clusters) for both the mtDNA and the SNP sets. However, the comparison in Figure 4 shows that the two sets of data provide very different pictures. In contrast to the two distant mtDNA clusters, the SNPs sets create three distinct clusters corresponding to the two pure SW/SE lineages and an interspersed hybrid cluster containing samples from both mtDNA lineages (Figures. 4, S4, S5).

The third hypothesis assumes that during their separation, the two parasite lineages reached a high degree of genetic differentiation resulting in a strong but not absolute postzygotic barrier, whilst lacking an efficient prezygotic barrier preventing them from mating. As a consequence, upon encountering each other they formed a narrow hybrid zone in which the majority of the inter-lineage matings fail or produce hybrids with lowered fitness. In this case, we would expect a sharp geographic division between the SW and SE populations with occasional hybrids identified by nuclear markers around the secondary contact zone. Based

on the data presented in this study and the previous extensive analysis of mtDNA (Martinů et al., 2018), we consider this hypothesis to be the best explanation of the observed patterns. A decoupled genetic structure of a host and its parasite(s) is not exceptional. It has been reported in various host-parasite associations and caused by different biological and/or environmental circumstances (e.g du Toit, van Vuuren, Matthee, & Matthee, 2013; Hafner et al., 2019). However, to our knowledge, the *Apodemus -Polyplax* association presented here is the first known example of genetic structuring caused by a parasite's hybrid zone created in the absence of the reciprocal host's hybrid zone. There are several possible factors behind the lack of evidence for similar patterns in nature. Firstly, only a few studies have dealt with hybrid zones in parasites, and they were usually approached in relation to their hosts' hybrid zone (e.g. Theodosopoulos et al. 2019). This is understandable considering the prevailing view of parasites' evolution being predominantly determined by their hosts. Secondly, it is likely that this pattern will emerge during secondary contact only at a specific ratio (or narrow range of ratios) of genetic diversification between host and parasite populations. If the diversification is too strong, it may either result in speciation of both counterparts (i.e. classical cospeciation Page, 2003), in speciation of the parasite and emergence of a hybrid zone in the host (Čížková et al. 2018; Hafner et al. 2019) or in hybrid zone for both counterparts (e.g. de Bellocq et al. 2018). On the contrary, if the diversification is too weak, both counterparts will re-establish panmictic populations. This only leaves a narrow window of time for hosts' panmixia vs. parasite's hybrid zone. Yet, such cases do not necessarily have to be rare in nature, they may just be understudied or unnoticed due to the *a priori* view that evolution in host-specific parasites is linked to their hosts. The case we present here shows that one possible indication of a decoupled pattern is a strong mtDNA structure in a highly panmictic host population.

Genetic incompatibility between two populations at the secondary contact zone can be caused by various mechanisms. Apart from the differences accumulated in the nuclear genetic information, interbreeding can also be prevented by the incompatibility of mitochondrial and nuclear genetic information (Wolff, Ladoukakis, Enríquez, & Dowling, 2014; Hill, 2019). In our system, the lice are known to have their mitochondrial DNA split into several circular minichromosomes (Cameron, Yoshizawa, Mizukoshi, Whiting, & Johnson, 2011; Song et al., 2019). The distribution of mitochondrial genes among the minichromosomes is not entirely conserved - there are several differences in the gene arrangements when comparing the species *Polyplax spinulosa* and *P. asiatica* (Dong et al., 2014). To address the theoretical possibility that the barrier between the SW and SE lineages is caused by the failure of nucleus-mitochondrion interaction due to different distributions of their mitochondrial genes on the minichromosomes, we reconstructed full minichromosomes (their coding part) from all sequenced samples. In all cases, we found the same gene arrangement. This observation does not rule out the nuclear-mitochondrial incompatibility as a cause of the barrier, but it shows that it would have to be due to point mutations rather than structural differences (Table S3). In a similar way, we were not able to detect any significant difference between the *L. polyplacis* genomes from the SE and SW lineages, indicating that neither differences in the symbionts' metabolic capacities are causing the gene flow barrier.

It would be speculative to infer other genetic sources of incompatibility between SE and SW lineages without a detailed study of the louse nuclear genome and more extensive sampling in the secondary contact zone, which is beyond the scope of the current study. Nevertheless, based on evidence collected from three genetic resources, the two maternally inherited markers (*Legionella* and mtDNA) and nuclear SNP diversity, we were able to unambiguously distinguish between the three possible scenarios of host-parasite incongruence. We propose a new mechanism in host-parasite co-evolution, where a narrow hybrid zone is present in the parasite without a corresponding break in the genetic structure of its host. In this way, the panmictic population of the host is rid of the parasite lineage present on one side of the parasite's hybrid zone, which is gradually replaced by a different parasite lineage on the other side. Given that this evolutionary scenario can easily pass unnoticed (due to the lack of structure in the host) we hypothesize that "parasite turnover zones" may be more common than is currently known, particularly in highly host-specific parasites.

Acknowledgements

References

Alexander, D. H., Novembre, J., Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*,;19,1655–64 Allen, J. M., LaFrance, R., Folk, R. A., Johnson, K. P., & Guralnick, R. P. (2018). aTRAM 2.0: An Improved, Flexible Locus Assembler for NGS Data. *Evolutionary Bioinformatics*, 14, 0–3.*https://doi.org/10.1177/1176934318774546* Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. (2008). The RAST Server: rapid annotations using subsystems technology. BMC *Genomics*, 9, 75. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev & M. A., Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19, 455–477. 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 and Evolution*, 23, 678–685.*https://doi.org/10.1016/j.tree.2008.06.017* Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J. M., Middendorf, J. & Stadler, P. F. (2013). MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation.*Molecular Phylogenetics and Evolution*, 69, 313–319. Cameron, S. L., Yoshizawa,K., Mizukoshi, A., Whiting, M. F., & Johnson, K. P. (2011). Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). *BMC Genomics*, 12, 394.*https://doi.org/10.1186/1471-2164-12-394* Cerezo, M. L., Kucka, M., Zub, K., Chan, Y. F., & Bryk, J. (2020). Population structure of Apodemus flavicollis and comparison to*Apodemus sylvaticus* in northern Poland based on RAD-seq.*BMC Genomics*, 21(1), 1–14. Criscione, C., Poulin, R. & Blouin, M. (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes.*Molecular Ecology*, 14, 2247–2257. Čížková, D., Baird, S. J. E., Těšíková, J., Voigt, S., Ľudovít, Ď., Piálek, J., & Goüy de Bellocq, J. (2018). Host subspecific viral strains in European house mice: Murine cytomegalovirus in the Eastern (*Mus musculus musculus*) and Western house mouse (*Mus musculus domesticus*). *Virology*, 521, 92–98.*https://doi.org/10.1016/j.virol.2018.05.023* 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. & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 Darling, A.C.E., Mau, B., Blattner, F.R. & Perna, N.T. (2004). Mauve: Multiple alignment of conserved genomic sequence with rearrangements.*Genome Research*, 14, 1394–1403. Darriba, D., Taboada, G., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9, 772–772. de Bellocq, J., Wasimuddin, Ribas, A., Bryja, J., Pialek, J. & Baird, S. (2018). Holobiont suture zones: Parasite evidence across the European house mouse hybrid zone. *Molecular Ecology*, 27, 5214–5227. Dong, W.G., Song, S., Jin, D.-C., Guo, X.-G., & Shao, R. (2014). Fragmented mitochondrial genomes of the rat lice, *Polyplax asiatica* and *Polyplax spinulosa*: intra-genus variation in fragmentation pattern and a possible link between the extent of fragmentation and the length of life cycle. *BMC Genomics*, 15, 44.*https://doi.org/10.1186/1471-2164-15-44* du Toit, N., van Vuuren, B., Matthee, S. & Matthee, C. (2013). Biogeography and host-related factors trump parasite life history: limited congruence among the genetic structures of specific ectoparasitic lice and their rodent hosts. *Molecular Ecology*, 22, 5185–5204. Gómez-Díaz, E., González-Solís, J., Peinado, J. A. & Page, R. D. M (2007). Lack of host-dependent genetic structure in ectoparasites of*Calonectris* shearwaters. *Molecular Ecology*, 16, 5204–5215. Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W. & Gascuel O. (2010).

New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59, 307–21. Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.*Systematic Biology*, 52, 696–704. Hafner, D., Hafner, M., Spradling, T., Light, J. & Demastes, J. (2019). Temporal and spatial dynamics of competitive parapatry in chewing lice.*Ecology and Evolution*, 9, 7410–7424. Harper, S., Spradling, T., Demastes, J. & Calhoun, C. (2015). Host behaviour drives parasite genetics at multiple geographic scales: population genetics of the chewing louse, *Thomomydoecus minor.Molecular Ecology*, 24, 4129–4144. Hill, G.E. (2019). Mitonuclear Ecology. *Oxford Series in Ecology and Evolution.* 302 pp. Jarne, P. & Theron, A. (2001). Genetic structure in natural populations of flukes and snails: a practical approach and review. *Parasitology*, 123, 27–40. Johnson, K. P., Allen, J. M., Olds, B. P., Mugisha, L., Reed, D. L., Paige, K. N., & Pittendrigh, B. R. (2014). Rates of genomic divergence in humans, chimpanzees and their lice. Proceedings of Royal Society B, 281, 20132174. doi.org/10.1098/rspb.2013.2174 Jones, P. H. & Britten, H. B. (2010). The absence of concordant population genetic structurein the black-tailed prairie dog and the flea, *Oropsylla hirsuta*, with implications for the spread of*Yersinia pestis. Molecular Ecology*, 19, 2038–2049. Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*, 37, 540. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N, A, Mayrose, I. (2015). CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15, 1179–1191.*https://doi:10.1111/1755-0998.12387*Langmead, B., Salzberg, S. (2012). Fast gapped-read alignment with Bowtie 2.*Nature Methods*, 9, 357–359. Lee, T. H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data.*BMC Genomics*, 15. 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, 2078–2079. Louhi, K. J., Karvonen, A., Rellstab, Ch. & Jokela, J. (2010). Is the population genetic structure of complex life cycle parasites determined by the geographic range of the most motile host?*Infection, Genetics and Evolution*, 10, 1271–1277. Martinů, J., Hypša, V. & Štefka, J. (2018). Host specificity driving genetic structure and diversity in ectoparasite populations: Coevolutionary patterns in *Apodemus* mice and their lice.*Ecology and Evolution*, 8, 10008–10022. Maze-Guilmo, E., Blanchet, S., McCoy, K. & Loot, G. (2016). Host dispersal as the driver of parasite genetic structure: a paradigm lost?*Ecology Letters*, 19, 336-347. McCoy, K. D., Chapuis, E., Tirard, C., Boulinier, T., Michalakis, Y., Le Bohec, C., Le Maho, Y. & Gauthier-Clerc M. (2005). Recurrent evolution of host-specialized races in a globally distributed parasite.*Proceedings of the Royal Society B -Biological Sciences*, 272, 2389–2395. Nieberding, C., Morand, S., Libois, R. & Michaux, J. (2004). A parasite reveals cryptic phylogeographic history of its host. *Proceedings of the Royal Society B-Biological Sciences*, 271, 2559–2568. Page, R. D. M. (2003). *Tangled Trees: phylogeny, cospeciation and coevolution.* (Roderic D.M Page, Ed.). Chicago: University of Chicago Press,. Page, R. D. M., Lee, P. L. M., Becher, S. A., Griffiths, R. & Clayton, D. H. (1998). A Different Tempo of Mitochondrial DNA Evolution in Birds and Their Parasitic Lice.*Molecular Phylogenetics and Evolution*, 9, 276–293. Říhová, J., Nováková, E., Husník, F. & Hypša, V. (2017). *Legionella*Becoming a Mutualist: Adaptive Processes Shaping the Genome of Symbiont in the Louse *Polyplax serrata. Genome Biology and Evolution*, 9, 2946–2957. Song, F., Li, H., Liu, G. H., Wang, W., James, P., Colwell, D. D. & Shao, R. (2019). Mitochondrial genome fragmentation unites the parasitic lice of Eutherian mammals. *Systematic Biology*, 68, 430–440.*https://doi.org/10.1093/sysbio/syy062*Štefka, J., Hoeck, P. E. a, Keller, L. F., & Smith, V. S. (2011). A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. *BMC Evolutionary Biology*, 11, 284. Sweet, A. D., & Johnson, K. P. (2018). The role of parasite dispersal in shaping a host-parasite system at multiple evolutionary scales.*Molecular Ecology*, 2724, 5104–5119.*https://doi.org/10.1111/mec.14937*Theodosopoulos, A. N., Hund, A. K., & Taylor, S. A. (2019). Parasites and Host Species Barriers in Animal Hybrid Zones. *Trends in Ecology and Evolution*, 34, 19–30.*https://doi.org/10.1016/j.tree.2018.09.011*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*, 43, 11.10.1–33. Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G.

& Zdobnov, E. M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics.*Molecular Biology and Evolution*, 35, 543–548. Whiteman, N. K., Kimball, R. T. & Parker, P.G. (2007). Co-phylogeography and comparative population genetics of the threatened Galápagos hawk and three ectoparasite species: ecology shapes population histories within parasite communities. *Molecular Ecology*, 16, 4759–4773. Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A., & Dowling, D. K. (2014). Mitonuclear interactions: Evolutionary consequences over multiple biological scales. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369, 20130443.

Data Accessibility Statement

DNA sequences of COI obtained in the frame of the study will be made available in GenBank upon acceptance of the MS. DNA alignments of 26 minichromosomes and alignment of concatenated genes of *Legionella polyplacis* from 23 specimens are submitted to Dryad database. Fastq files of Illumina pair-end sequences of 26 *P. serrata* genomes will be made available in SRA archive upon acceptance of the MS.

Author Contributions

J.M. performed laboratory work, J.M. and A.P made data analyses under the supervision of J.Š. and V.H., with V.H. and J.Š. conceiving the study of *P. serrata* S lineage secondary contact zone. All four authors contributed toward the design of the study, and drafted the manuscript.

Figure legends

Figure 1. Different scenarios of the secondary contact outcomes in a host-parasite association. **(a)** the host re-establishes a panmictic population while the parasite differentiates into two separated species, which keep spreading and eventually live in sympatry.**(b)** both host and the parasite create hybrid zone; parasite's hybrid zone is narrower. **(c)** the host re-establishes panmictic population while the parasite creates a narrow hybrid zone. **(d)**as a consequence of the scenario c, the host moves freely across the whole area but it experiences a "parasite turnover" when crossing the hybrid zone.

Figure 2 . Genetic background of *Apodemus spp.* and*Polyplax serrata* adopted from Martinů et al., 2018. **(a)**The main genetic pattern of *Apodemus* hosts relevant to this study. Inner mtDNA structure of *A.sylvaticus* is ignored since it is not relevant to the discussion of the secondary contact zone (see the text). **(b)** Geographic distribution of the hosts: both spp. co-occur across Europe with their intraspecific mtDNA clades randomly dispersed. Both spp. are parasitized by the non-specific lineage (N) of the lice shown on the d. **(c)** Two mtDNA lineages of *A. flavicollis* which in addition to the N lineage are parasitized by the specific lineage (S; depicted on d). SCZ: schematic representation of the presumed secondary contact zone of the two mtDNA lineages of both the *A. flavicollis* and the S lineage of the parasite. While the host's lineages dispersed across the whole Europe after glaciation, the parasite's lineages remain confined to the two exclusive areas (red and blue) with a narrow hybrid zone (the violet line). **(d)** The main genetic pattern of *Polyplax serrata* showing ecological complexity of this parasite: the N and S lineages live in sympatry but differ in the degree of host specificity (single host vs two hosts); the SW (specific west) and SE (specific east) lineages create a hybrid zone (the „parasite turnover zone") which transverses panmictic population of the host.

Figure 3. **(a)** Distribution of the *P. serrata* S lineage samples analyzed in Martinů et al., (2018). **(b)** Detailed distribution of the samples used in this study. Blue and red designate the SW and SE sublineages, respectively.
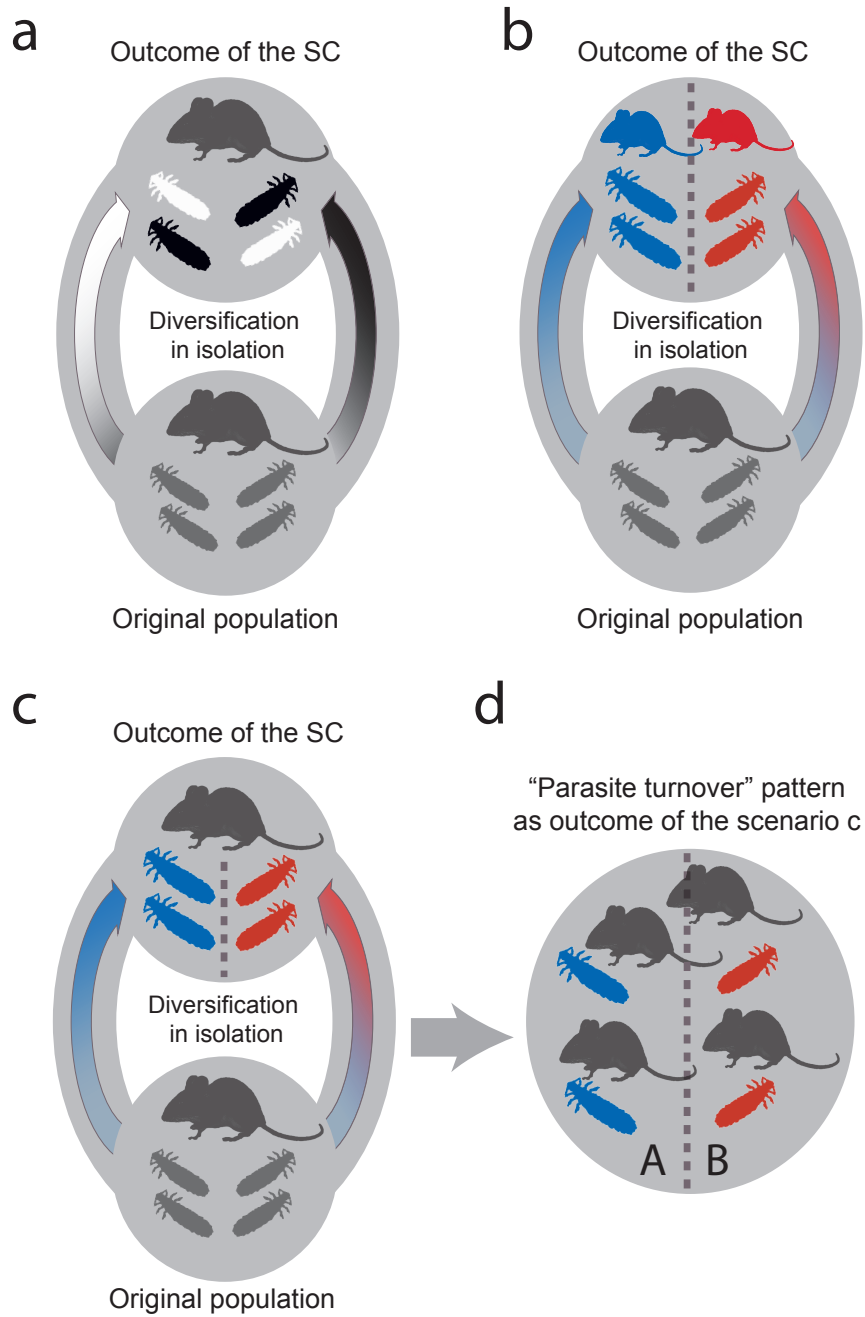
Figure 4. Comparison of maternally inherited markers and nuclear SNPs. In all parts the two mitochondrial lineages are designated by different colours: blue=SW, red=SE, and the nuclear hybrids are identified by purple background. **(a)** phylogenetic tree based on complete genomes of *L. polyplacis* from 23 samples of *P. serrata* . **(b)** phylogenetic tree based on concatenated 11 minichromosomes of *P. serrata* . **(c)** PCA analysis of nuclear SNPs showing the pure parental clusters SW/SE and the hybrid cluster by the blue, red and purple background, respectively.**(d)** Admixture analysis of SNP data under K2 parameter recognizing the parental populations and hybrids.
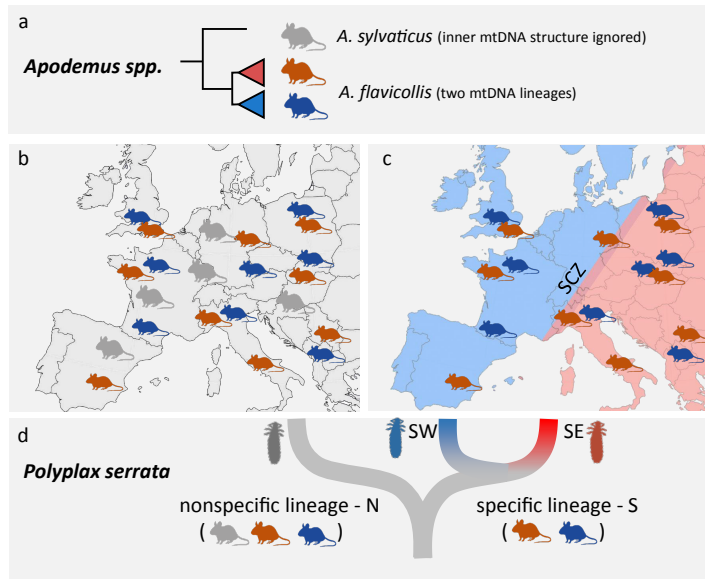
Tables

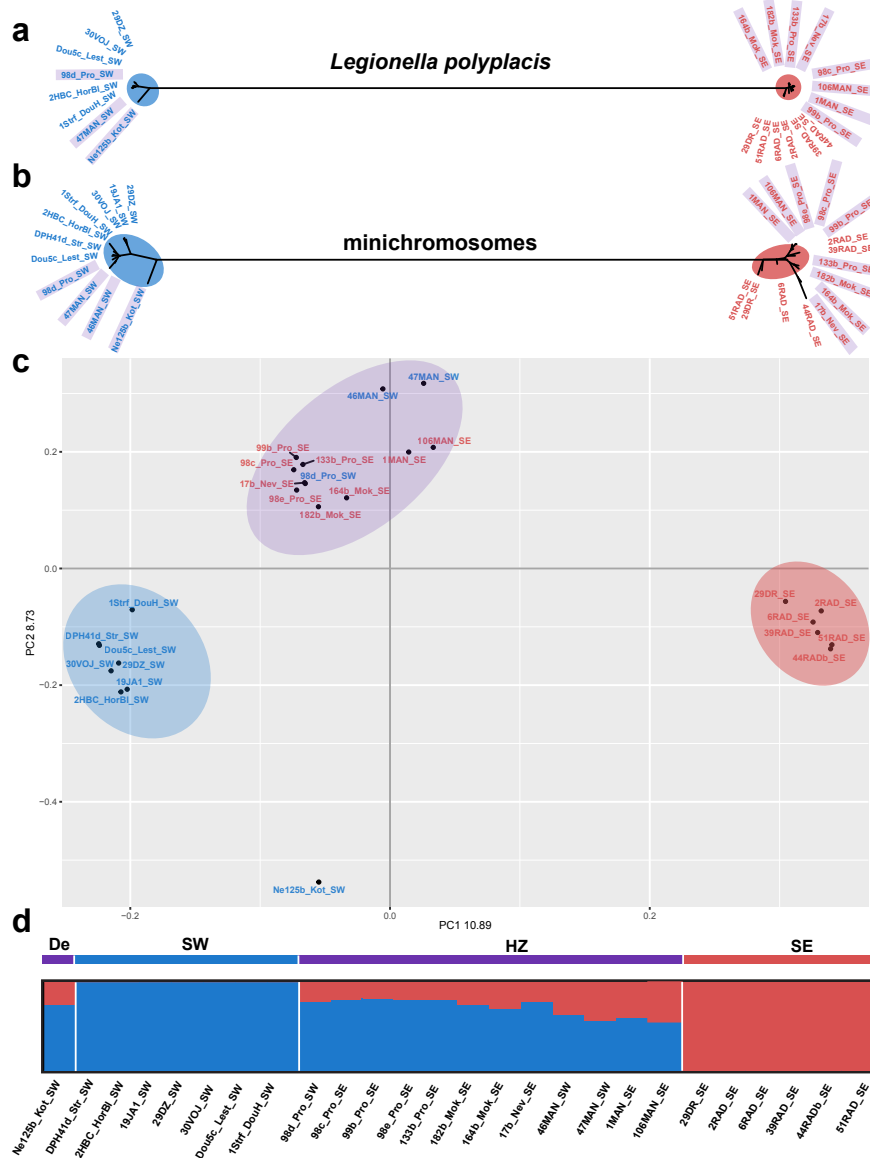**Table 1.** List of the specimens analyzed in this study.

Hosted file

Martinu_TurnoverZone_Revised_Tracked.docx available at https://authorea.com/users/297297/articles/465384--parasite-turnover-zone-at-secondary-contact-a-new-pattern-in-host-parasite-population-genetics

a

Apodemus spp.

A. sylvaticus (inner mtDNA structure ignored)

A. flavicollis (two mtDNA lineages)

b

c

SCZ

d

Polyplax serrata

SW   SE

nonspecific lineage - N

specific lineage - S

Hosted file

Figure 3.pdf available at https://authorea.com/users/297297/articles/465384--parasite-turnover-zone-at-secondary-contact-a-new-pattern-in-host-parasite-population-genetics

Hosted file

Table1.xlsx available at https://authorea.com/users/297297/articles/465384--parasite-turnover-zone-at-secondary-contact-a-new-pattern-in-host-parasite-population-genetics