

# A novel approach to parasite population genetics: Experimental infection reveals geographic differentiation, recombination and host-mediated population structure in *Pasteuria ramosa*, a bacterial parasite of *Daphnia*

J. P. ANDRAS and D. EBERT

Zoologisches Institut, Evolutionsbiologie, Universität Basel, Vesalgasse 1, CH-4051, Basel, Switzerland

## Abstract

The population structure of parasites is central to the ecology and evolution of host-parasite systems. Here, we investigate the population genetics of *Pasteuria ramosa*, a bacterial parasite of *Daphnia*. We used natural *P. ramosa* spore banks from the sediments of two geographically well-separated ponds to experimentally infect a panel of *Daphnia magna* host clones whose resistance phenotypes were previously known. In this way, we were able to assess the population structure of *P. ramosa* based on geography, host resistance phenotype and host genotype. Overall, genetic diversity of *P. ramosa* was high, and nearly all infected *D. magna* hosted more than one parasite haplotype. On the basis of the observation of recombinant haplotypes and relatively low levels of linkage disequilibrium, we conclude that *P. ramosa* engages in substantial recombination. Isolates were strongly differentiated by pond, indicating that gene flow is spatially restricted. *Pasteuria ramosa* isolates within one pond were segregated completely based on the resistance phenotype of the host—a result that, to our knowledge, has not been previously reported for a nonhuman parasite. To assess the comparability of experimental infections with natural *P. ramosa* isolates, we examined the population structure of naturally infected *D. magna* native to one of the two source ponds. We found that experimental and natural infections of the same host resistance phenotype from the same source pond were indistinguishable, indicating that experimental infections provide a means to representatively sample the diversity of *P. ramosa* while reducing the sampling bias often associated with studies of parasite epidemics. These results expand our knowledge of this model parasite, provide important context for the large existing body of research on this system and will guide the design of future studies of this host-parasite system.

*Keywords:* bacteria, coevolution, *Daphnia*, Parasite, *Pasteuria*, population genetics

Received 28 July 2012; revision received 15 October 2012; accepted 21 October 2012

## Introduction

The population structure of parasites is of fundamental importance to the ecological and evolutionary dynamics of host-parasite systems. The degree, distribution and flow of genetic diversity, and the extent to which that diversity is recombined among genomes will directly affect the evolutionary potential of the parasite, which

will in turn influence the strength, tempo and spatial scale of host-parasite coevolutionary dynamics (Thompson 1994; Kaltz & Shykoff 1998; Lively 1999). For example, high levels of parasite gene flow relative to the host are expected to facilitate stronger patterns of local coadaptation, as the influx of novel genetic variation allows the parasite to counter the recombinatorial advantages enjoyed by sexual hosts (Gandon 2002). Consistent with this, strong patterns of parasite local adaptation were observed in two separate snail-helminth systems, where host populations were much more highly structured

Correspondence: Jason Andras, Fax: 0041 (0)61 267 03 62;  
E-mail: jpa24@cornell.edu

than parasite populations (Dybdahl & Lively 1996; Davies *et al.* 1999).

Parasite population structure is largely determined by life history traits such as life cycle complexity, transmission mode and dispersal mechanism (Criscione & Blouin 2004; Barrett *et al.* 2008). One of the most important life history features governing the population structure of microbial parasites is the degree to which they engage in some form of recombination (Tibayrenc *et al.* 1991). For purely clonal species, populations consist of assemblages of independently evolving lineages, and mutations are confined to the lineage in which they arose. Clonal species may benefit from a higher intrinsic growth rate and lower levels of intragenomic conflict relative to their sexual counterparts (Hickey 1982). However, the lack of horizontal genetic exchange precludes the generation of recombinatorial novelty and limits the evolutionary potential of clones (reviewed in Butlin 2002). Clonal species will be characterized by strong linkage disequilibrium across the genome, meaning that the evolutionary history of the lineage can be accurately described based on just a few polymorphic markers. Moreover, there will be linkage between neutral genetic diversity and functionally significant biological diversity, and traits of interest (e.g. infectivity and virulence) can be mapped directly on a bifurcating population tree (e.g. *Salmonella enterica*, Spratt & Maiden 1999).

Conversely, species undergoing recombination, even rarely, can exchange genetic material among genomes, thereby increasing their evolutionary potential. This source of genomic novelty could be important for the maintenance of parasite populations, particularly when gene flow is limited. In parasite populations engaging in substantial recombination, linkage disequilibrium will be weak or absent, and heritable phenotypic traits will not segregate among population genetic clusters (e.g. *Helicobacter pylori*, Spratt & Maiden 1999).

For parasites in general and obligate parasites in particular, spatial population structure will be influenced by the life history, distribution and vagility of hosts, which essentially serve as habitat islands (Nadler 1995). In cases of intraspecific host-parasite specificity, the genetic structure of host populations will also be of central importance. Resistance phenotypes of the host may effectively serve as filters, determining which and how many parasite strains can exist in a population, and which parasite genotypes may come into contact within the host. The most extreme case of such a scenario is a matching-genotype model (Seger & Hamilton 1988), where a given parasite genotype can only infect a matching host genotype. In such cases, different parasite genotypes with mutually exclusive host genotype ranges will not come into contact and, consequently, cannot recombine.

In this study, we explore the population structure of *Pasteuria ramosa*, an ecologically important and well-studied obligate bacterial parasite of freshwater crustaceans of the genus *Daphnia*. *Pasteuria ramosa* is an endospore-forming, Gram-positive bacterium of the *Bacillus-Clostridium* clade (Ebert *et al.* 1996), and has been reported in North America and throughout Eurasia. Infectivity of *P. ramosa* has been shown to be highly specific both among and within species of *Daphnia* (Carius *et al.* 2001; Luijckx *et al.* 2011; P. Luijckx, unpublished data), and prevalence can reach 100% in natural populations (Duncan *et al.* 2006). *Pasteuria ramosa* is transmitted exclusively horizontally among *Daphnia* individuals when endospores are stirred up from the spore bank in the sediment and ingested by the host during filter feeding. If host and parasite genotypes are compatible, spores of *P. ramosa* will adhere to the *Daphnia*, penetrate the body wall, and proliferate vegetatively in the haemolymph of the host. During the course of infection, *P. ramosa* castrates its host and induces gigantism and a distinct reddish-brown coloration, making infected *Daphnia* easy to identify. *Pasteuria ramosa* continues to grow and sporulate until the death of the host, at which point several million endospores are released to the sediment, where they can remain competent for decades and can be transmitted to other hosts (Ebert *et al.* 1996).

Because of the high specificity and strong fitness effects of infection, the *P. ramosa/Daphnia* system is commonly studied as a model of host-parasite coevolution (e.g. Carius *et al.* 2001; Decaestecker *et al.* 2007; Luijckx *et al.* 2012). Yet, there is currently very little known regarding the population structure of *P. ramosa*, representing a significant gap in our understanding of this system. The only population genetic study of *P. ramosa* to date examined variation at six polymorphic microsatellite loci among isolates sampled from 18 individual *Daphnia* of two species across four geographically distant localities (Mouton & Ebert 2008). This study observed differentiation at all levels (among localities, among host species, within host species) and reported that individual infected *Daphnia* may contain multiple genotypes of *P. ramosa*. However, the sample size of this study was small, and the sampling scheme was not designed to adequately resolve how population genetic variance was partitioned among hosts and localities. In addition, because the *P. ramosa* isolates were obtained from naturally infected *Daphnia* collected at a single time point, it is likely that they were not representative, as parasite samples from infected hosts during an epidemic are often highly biased (Maynard Smith *et al.* 1993; Spratt & Maiden 1999).

Here, we investigate the population structure of *P. ramosa* from two geographically distant ponds in

Eurasia, infecting the most commonly parasitized host species, *D. magna*. To minimize sampling bias and to test the importance of host characteristics in governing parasite population structure, we sampled *P. ramosa* isolates directly from the sediment spore bank of the two ponds by experimentally infecting a panel of non-native *D. magna* hosts known to have different genotypes and resistance phenotypes. We employed a hierarchical sampling scheme designed to resolve the population genetic structure of *P. ramosa* at the levels of geography, host phenotype, host genotype and host individual. To determine whether isolates from experimental infections were comparable to natural infections, we also sampled naturally infected hosts native to one of the same source ponds. We genotyped all parasite isolates using nine previously published polymorphic microsatellite markers (Mouton *et al.* 2007), plus three new microsatellite markers developed from the partially assembled genome of *P. ramosa*. On the basis of previously published results, we expected *P. ramosa* would be differentiated between geographic localities. Due to the high specificity of infection, we also expected that populations of *P. ramosa* would be structured according to host phenotype and genotype. Finally, as identical genotypes of *P. ramosa* have been previously isolated from geographically distant localities (Mouton *et al.* 2007; J. Andras, unpublished data), we hypothesized that parasite population structure would be concordant with clonal reproduction.

## Methods

### Sample collection/experimental infection

Sediment samples were collected from two ponds, 2527 km apart, where both *D. magna* and *P. ramosa* are known to co-occur: Volgograd, Russia (site code = RU-BN; coordinates: N 50.157222°, E 43.391944°; sampled in June, 2008) and Hohliberg, Switzerland (Site code = CH-H; coordinates = N 47.557769° E 8.862783°, sampled in October, 2010). The surface area of pond RU-BN is approximately 12.2 ha, and pond CH-H is approximately 1.7 ha. Both ponds dry partially during the summer, and both are covered completely with ice in the winter. Sediment samples were collected from the upper 3 cm, drained of standing water and stored in the dark at 4 °C until use. The age range of *P. ramosa* spores contained in a given vertical section of sediment will depend on the sedimentation rate and the degree of sediment perturbation, and may therefore vary among and within ponds over both space and time. As we did not measure sediment age in the present study (e.g. by radiodating), we cannot estimate the age range of *P. ramosa* sampled. However, based on average sedi-

mentation rates observed in such ponds, it is likely that a 3-cm sample contains between 2 and 6 years of sediment.

To isolate and augment samples of *P. ramosa*, experimental infections were generated by exposing nonnative *D. magna* from eight distinct clonal lines to sediment from the ponds. *Daphnia magna* host clones were chosen based on their known resistance or susceptibility to two previously described clones of *P. ramosa*, (clones C1 & C19, Luijckx *et al.* 2011). Two *D. magna* clones were selected from each of the four possible composite resistance phenotypes (hereafter referred to as *resistotypes*, Table 1). Prior to experimental infection, *D. magna* clones were maintained under controlled conditions for at least three generations (eight females per 400 mL jar filled with ADaM—Kluettgen *et al.* 1994; 16:8 h light: dark; 20 °C; fed  $7.5 \times 10^7$  cells of chemostat-cultured *Scenedesmus obliquus* daily). Each of the 16 experimental infection trials (*D. magna* clone/sediment combination) was performed in three replicate jars with 10 *D. magna* individuals in each. Sixty juvenile females ( $7 \pm 1$  days old) from each of the eight *D. magna* clone cultures were distributed evenly across six 100-mL jars containing 50 mL of ADaM and a 15 mm magnetic stir bar. From each of the two pond samples, 120 cm<sup>3</sup> of sediment was suspended in 600 mL of ADaM and passed through a 200 µm filter. The filtrate was stirred vigorously to maintain a homogeneous suspension, and 25 mL aliquots were distributed to each jar.

*Daphnia magna* were exposed to sediment for 6 days, and the sediment was resuspended with brief magnetic stirring every second day. After exposure, *D. magna* individuals were removed from the sediment jars and transferred to 400-mL jars filled with ADaM. Exposed individuals were kept for 6 weeks under standard

**Table 1** *Daphnia magna* host clones used for experimental infections, along with their origin and resistance phenotype with respect to two laboratory clones of *Pasteuria ramosa*

<i>D. magna</i> host clone	Geographic origin	Resistotype (C1/C19 susceptibility)
HU-HO2	Bogarzo-to, Hungary	Susceptible/Susceptible
BE-M10	Leuven, Belgium	Susceptible/Susceptible
FI-AL1-4-4	Tvärminne, Finland	Susceptible/Resistant
FI-KELA-39-9	Tvärminne, Finland	Susceptible/Resistant
FI-XINB3	Tvärminne, Finland	Resistant/Susceptible
FI-KELA-18-10	Tvärminne, Finland	Resistant/Susceptible
DE-IINB1	Munich, Germany	Resistant/Resistant
IL-M1-1	Jerusalem, Israel	Resistant/Resistant

The four clones from Tvärminne, Finland originate from rock pool populations on four different islands.

conditions and transferred to fresh jars twice weekly to refresh the medium and remove offspring. After 6 weeks, individuals were visually screened for infection. Infected individuals were maintained under standard conditions until death, and were then collected and frozen until subsequent analysis. All *D. magna* clones were successfully infected by *P. ramosa* from both sediment samples.

We also sampled *D. magna* infected with *P. ramosa* directly from the CH-H pond (collected in June, 2011) to compare experimental isolates with naturally occurring isolates. Occasionally, infected *D. magna* will produce one or more clutches of clonal offspring prior to being completely castrated by the parasite. In such cases, it is possible to establish a clonal line of the host and also to isolate *P. ramosa* native to that host. The clonal offspring can then be used to determine the resistotype of the host, following the protocol of Luijckx *et al.* (2011). Using this approach, we collected 10 naturally infected *D. magna* clones—five of which were susceptible to both C1 & C19 laboratory clones of *P. ramosa*, and five of which were resistant to both C1 & C19. These are the only two *D. magna* resistotypes (of the four possible) known to occur in the CH-H pond (J. Andras, unpublished data). These infected hosts were kept in the laboratory under standard conditions until death and then frozen until subsequent analysis.

#### DNA Extraction, amplification and microsatellite genotyping

*Pasteuria ramosa* spores were isolated from three infected *D. magna* individuals of each host clone/sediment combination, one individual from each of the three replicate jars (two pond sediment samples  $\times$  eight *D. magna* clones  $\times$  three infected individuals = 48 experimental *P. ramosa* isolates), as well as the ten native CH-H infected hosts. DNA was extracted from each isolate using the NucleoSpin Tissue XS kit (Macherey Nagel), following a modified version of the manufacturer's protocol. Infected *D. magna* individuals were placed in a 1.5-mL plastic sample tube with 1 mL of deionized H<sub>2</sub>O and crushed with a plastic pestle. *Pasteuria ramosa* spores were pelleted via centrifugation for 3 min at 16 RCF, and the supernatant was removed. Spores were washed via resuspension/centrifugation twice more with 1 mL deionized H<sub>2</sub>O. After the final wash, the pellet was resuspended in 500  $\mu$ L of Lysis Buffer T1 (Macherey Nagel) and digested with lysozyme (5  $\mu$ L of 250 U/ $\mu$ L Ready-Lyse lysozyme, Epicentre; digested for 15 min at room temperature) followed by proteinase K (8  $\mu$ L of 20 mg/mL proteinase K, Macherey Nagel, digested for 24 h at 56 °C). Samples

were transferred to a 2-mL screw-top tube containing 0.3 cm<sup>3</sup> of 750- $\mu$ m-diameter ceramic beads (Macherey Nagel) and 0.1 cm<sup>3</sup> of 100- $\mu$ m-diameter zirconia beads (Biospec) and beat on a bead mill at 6.5 m/s for 1 min. After bead beating, 500  $\mu$ L of Lysis Buffer B3 was added, and samples were digested for 5 min at 70 °C. 500  $\mu$ L of ethanol was added, and samples were centrifuged for 5 min at 16 RCF. The supernatant was removed and passed through a NucleoSpin Tissue XS DNA extraction column. Thereafter, DNA extraction followed the manufacturer's protocol.

All samples were genotyped using primers for nine previously described microsatellite loci (Mouton *et al.* 2007), plus three new microsatellite loci developed from an unpublished genome assembly of *P. ramosa* (Table 2). Prior to settling on the use of microsatellites for this study, we surveyed a number of housekeeping genes and other loci commonly used for multilocus sequence typing and found insufficient variation to resolve *P. ramosa* clones that we knew to be phenotypically distinct. We also considered shotgun genome sequencing, but most natural *P. ramosa* isolates contain multiple haplotypes, and there are currently no computational techniques that can reconstruct individual multilocus haplotypes from such mixed sequences. In contrast, the microsatellite markers we chose were variable enough to reliably resolve different strains, and we were able to use them to estimate individual haplotypes, allowing for more rigorous population genetic analyses.

Based on product size ranges, microsatellite primers were combined into multiplex PCR reactions, four loci per reaction and forward primers within each reaction were uniquely labelled with fluorescent dyes. The concentration of each primer pair was adjusted to achieve approximately equal product ratios. Multiplex PCRs were 10  $\mu$ L total volume: 1  $\mu$ L DNA extract, 5  $\mu$ L Multiplex PCR Master Mix (Qiagen), 3  $\mu$ L H<sub>2</sub>O and 1  $\mu$ L primer mix solution (see Table 2 for primer concentrations). Reactions were performed on a PTC-200 Peltier Thermal Cycler (MJ Research) with the following protocol: an initial denaturation step at 95 °C for 15 min, followed by 20 cycles of 94 °C for 30 s, 57 °C for 90 s, 72 °C for 90 s and a final extension at 72 °C for 10 min. PCR products were analysed on an ABI 3130 Automated Capillary DNA Sequencer using the GeneScan-500 LIZ size standard (Applied Biosystems). Allele sizes were scored using GENEMAPPER version 4.0 (Applied Biosystems) and validated by eye. All isolates in this study were monomorphic at one of the 12 loci (PR17), despite the fact that this locus is known to be polymorphic among *P. ramosa* isolates from other localities (J. Andras, unpublished data). Consequently, this locus was excluded from all analyses.

**Table 2** Primer sequences, PCR details, size range and diversity for 12 microsatellite loci of *Pasteuria ramosa*

Locus	Primer sequence (5'–3')	Multiplex reaction	Primer concentration in PCR (µM)	Size range (b.p.)	Number of alleles observed
PR1*	F: JOE-ACCTAAAGAACAGGAATATCTGGA R: GCATGGAATGATTTTTGCTG	2	0.2	222–372	10
PR2*	F: TAMRA-CTGCTGGATGGATGGACTACGTGA R: ACCGGTCCCGTAGGTATAGG	3	0.3	261–308	4
PR3*	F: FAM-GGACCAATCGAACCAGGTAT R: AACGGTTTCTTCGCTTGTTG	2	0.2	254–421	7
PR4*	F: TAMRA-GGTAACCCTGGATGTCCTGA R: ATCCCGTTACAAATGGGACA	2	0.3	325–396	8
PR8*	F: FAM-GCATCAAATACAAAACAAATGAAG R: TGTTTCTCTCGCGTTTCCTT	3	0.2	394–431	6
PR9*	F: FAM-ATACGACGAACGGAACAAGA R: AACCAAAGAATTAACGCCATT	1	0.1	329–527	20
PR11*	F: FAM-CAAGCCAAATAAACGCATCC R: TAGCGAAGAACACCAACGTG	1	0.1	166–187	6
PR12*	F: JOE-TCTTTAGTAGTTGCTTTGCTTGAA R: AACATCTTGGACCCCTTTA	1	0.1	224–271	9
PR16*	F: TAMRA-GGCAGGAACAAAATTAAGCA R: CGTTCCAAAGCGTTTTATGG	1	0.4	282–325	7
PR17†	F: ROX-CACACACTTGCTCCATGGTC R: AAACACTACCGGATAGCGAAAAA	3	0.2	205–207	2
PR18†	F: ROX-AAAGAAAAGCTTCGTTTAAACGTG R: CATTATCCACCCCAAATCA	2	0.2	154–188	10
PR19†	F: JOE-ACGACCCAATCCGTTGATAG R: CCAAGGCACGTTAGAAGAAA	3	0.2	218–245	9

\*Reported in Mouton *et al.* (2007).

†Developed from *P. ramosa* genome assembly.

Although *P. ramosa* is haploid, multiple alleles were observed at some loci in some samples. Since all primer pairs used in this study amplified only a single allele per locus from clonal laboratory strains of *P. ramosa*, it is assumed that the observation of multiple alleles per locus is indicative of multiple *P. ramosa* genotypes within a single infected host, and not due to paralogous copies of the locus. In samples containing multiple alleles per locus, discrete haplotypes could not be definitively assigned. However, the electropherogram peak height of these microsatellite loci has been shown to correlate well with relative abundance of *P. ramosa* spores with different haplotypes (Ben-Ami *et al.* 2008; J. Andras, unpublished data). On this basis, we were able to estimate the primary multilocus haplotype of all isolates according to the relative fluorescence intensity of the allele peaks observed at each locus. For each individual, we defined the 'primary' allele at each locus in each sample as the one that occurred individually or had the highest fluorescence peak and a peak height at least ten times greater than the second highest peak. By this method, the primary allele was clear in 99% of all allele calls. Using

these primary alleles, we were able to assign each sample a single multilocus haplotype (hereafter referred to as the *estimated haplotype dataset*), which is estimated to represent the most prevalent *P. ramosa* haplotype in the sample. This dataset allows individual haplotypes to be identified within and among populations and also makes possible the estimation of linkage disequilibrium among loci, which is necessary to detect recombination. In addition to the estimated haplotypes, we also constructed an alternative dataset containing all alleles at all loci, coded as dominant markers based on their presence/absence (hereafter referred to as the *comprehensive allelic dataset*). This approach is commonly used for AFLP, RFLP and DNA fingerprinting data and has the benefit of including the total observed allelic variation present in an isolate. In contrast to individually resolved haplotypes, the comprehensive allelic data of an isolate serves as a descriptor of the aggregate *P. ramosa* population within a single infected host (the infrapopulation, *sensu* Bush *et al.* 1997). The estimated haplotype dataset and the comprehensive allelic dataset were used for alternative analyses of genetic diversity and population structure.

### Genetic diversity

The genetic diversity of *P. ramosa* isolates was quantified by two alternative methods for each host clone, host resistotype and pond. Based on the comprehensive allelic dataset, mean allelic richness was calculated per locus per individual by counting the total number of unique alleles at each locus and averaging across all loci. Gene diversity (Nei 1987) was calculated based on the estimated haplotype dataset using the program GENODIVE 2.0b22 (Meirmans & van Tienderen 2004). Diversity measures were checked for normality and compared among localities and host resistotypes using a nested ANOVA run in JMP 9.0 (SAS Institute) [Fixed Effects = Source Pond, Host Resistotype, Source Pond\*Host Resistotype; Random Effect = Host Clone(Host Resistotype)]. *Post hoc* comparisons were made between resistotypes within each pond using the Tukey–Kramer HSD test.

### Linkage disequilibrium and recombination

To test for recombination in *P. ramosa*, several measures of linkage disequilibrium (LD) were calculated using the estimated haplotype dataset. Duplicate haplotypes (clonemates), which were only observed within ponds and not between them, were removed from the dataset for all subsequent analyses, as they inflate LD estimates. LD was measured between all pairs of loci within each pond using GENEPOP 4.1 (Raymond & Rousset 1995; —100 batches of 1000 iterations). The significance of multiple comparisons among locus pairs was assessed using the Benjamini–Hochberg correction for false discovery (Benjamini & Hochberg 1995)—a statistically more powerful analogue of the commonly used Bonferroni correction. LD was also measured over all loci within each population by calculating the index of association ( $I_A$ , Maynard Smith *et al.* 1993) and  $\bar{r}_d$ , a related measure that is insensitive to the number of loci used and ranges between 0 and 1, making it comparable across studies (using the program MULTILOCUS—Agapow & Burt 2001—1000 randomizations to assess significance). These indices quantify the ratio of the observed variance of pairwise allelic mismatches among haplotypes to the variance expected under random association of alleles at different loci. Values of  $I_A$  or  $\bar{r}_d$  that differ significantly from zero provide evidence of linkage disequilibrium. We also examined the data for direct evidence of recombination by testing for pairs of incompatible loci using MULTILOCUS. Pairs of loci are defined as incompatible if the observed genotypes at those loci could only have arisen by recombination or homoplasy (Estabrook & Landrum 1975). This test has relatively low statistical power but provides strong direct evidence for recombination when the results are positive.

### Population structure

To partition the genetic variance of *P. ramosa* according to source pond, host resistotype and host clone, we performed hierarchical analyses of molecular variance (AMOVA, Excoffier *et al.* 1992) on the comprehensive allelic dataset using the program ARELQUIN 3.5 (Excoffier *et al.* 2005). One AMOVA contained all isolates grouped according to host resistotype, nested within geographic pond. Isolates were then split by pond and two additional AMOVAs were performed separately on CH-H and RU-BN, with isolates grouped according to host clone, nested within host resistotype. The significance of each analysis was assessed relative to 1000 randomly permuted datasets. Natural isolates from pond CH-H were not included in the AMOVA, as they did not conform to the same balanced hierarchical sampling scheme of the experimental isolates.

Population structure among *P. ramosa* isolates was also inferred using two alternative clustering methods that make no a priori assumptions regarding group membership. First, using the comprehensive allelic dataset, genetic distances among individual isolates of *P. ramosa* were calculated using the program RESTDIST (Felsenstein 1993; —data modelled as restriction fragments). These genetic distances were used to construct a dendrogram based on the Fitch–Margoliash clustering method using the program FITCH (Felsenstein 1993—search for best tree, 10× randomized input order, global rearrangements) with 1000 bootstrap replicates to generate support values for each branch.

Next, population structure was inferred using the estimated haplotype dataset, based on a Bayesian clustering approach implemented by the program STRUCTURE 2.3.3 (Pritchard *et al.* 2000). We ran Markov chain Monte Carlo simulations with  $1 \times 10^6$  steps (burn-in  $1 \times 10^6$ , admixture model with no prior population information, correlated allele frequencies). Five independent runs were performed at each K from 1 to 8. To estimate the number of populations represented in our data, we employed the  $\Delta K$  method (Evanno *et al.* 2005) as implemented by STRUCTURE HARVESTER (Dent & vonHoldt 2012). For the optimal value of K, cluster membership coefficients for the five replicate runs were matched and averaged using the program CLUMPP (Jakobsson & Rosenberg 2007—Greedy algorithm,  $1 \times 10^5$  random input orders).

## Results

### Genetic diversity

Multiple alleles were observed for at least one locus in 98% (47/48) of experimental infections and 70% (7/10) of

the natural infections analysed, indicating that most infected *D. magna* individuals hosted more than one *P. ramosa* haplotype. The highest number of alleles observed at a single locus in a single isolate was five, and mean allelic richness per isolate ranged from 1 to  $4.25 \pm 0.25$ . A nested ANOVA of mean allelic richness per isolate identified significant effects of source pond ( $F_{1,516} = 138.59$ ,  $P < 0.0001$ ), host resistotype ( $F_{3,4} = 17.65$ ,  $P = 0.009$ ), and source pond\*host resistotype ( $F_{3,516} = 8.53$ ,  $P < 0.0001$ ). There was no significant effect of host clone nested within host resistotype ( $F_{4,516} = 0.76$ ,  $P = 0.55$ ). Pond RU-BN was substantially more diverse than CH-H, having more alleles overall (65 vs. 47) and higher mean richness per isolate ( $2.06 \pm 0.05$  vs.  $1.30 \pm 0.04$ , Fig. 1). *Post hoc* comparisons illustrate that mean allelic richness also differed significantly among host resistotypes within ponds. In both ponds, *D. magna* that were susceptible to both laboratory clones of *P. ramosa* had the highest allelic richness, while those that were resistant to both laboratory clones had the lowest richness (Fig. 1). *Post hoc* comparisons also illustrate no significant differences between the mean allelic richness of naturally infected hosts native to the CH-H pond and experimental infections of nonnative hosts of the same resistotype (Fig. 1). All diversity measures are summarized in Table 3.

#### Linkage disequilibrium and recombination

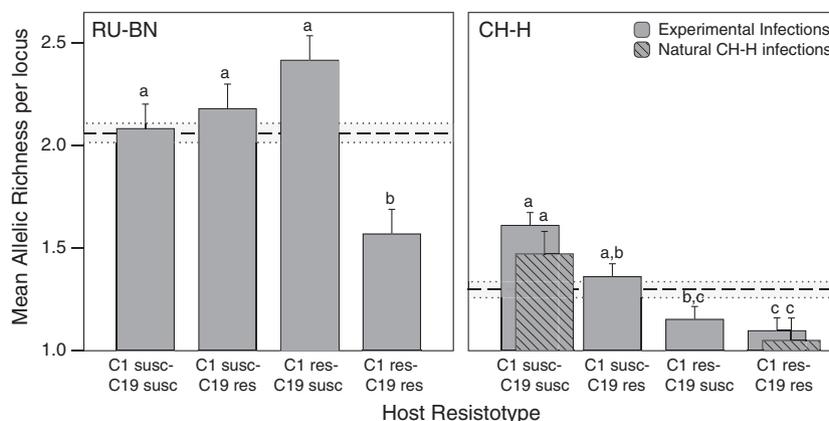
Based on the estimated haplotype dataset, there were 33 unique haplotypes among the 58 *P. ramosa* isolates analysed, and six of these haplotypes were observed in more than one isolate (clonemates). Clonemates were

generally more prevalent in the CH-H pond (four haplotypes observed 3–11 times each) than the RU-BN pond (two haplotypes observed twice each). There were no haplotypes shared between the two ponds.

Indices of association ( $I_A$  &  $\bar{r}_d$ ) were significantly different from zero across isolates from both RU-BN and CH-H, indicating an overall pattern of linkage disequilibrium (Table 4). However, fewer than half of the pairwise tests for LD among loci were significant within each pond (RU-BN: 21.8%, CH-H: 49.1%). In fact, even when isolates were not split by pond and no correction for multiple comparisons was applied, the proportion of pairwise comparisons showing significant LD was just 85%. Together, these results indicate that there is significant LD among loci within populations of *P. ramosa*, but that this linkage is much weaker than would be expected under purely clonal reproduction. This assessment is supported by the frequency of incompatible loci. Over the entire dataset, 72.3% of all possible locus pairs (40/55) were incompatible, meaning the combinations of haplotypes observed among those locus pairs could only have arisen through recombination or homoplasy.

#### Population structure

Based on an AMOVA of all samples in the comprehensive allelic dataset, the majority of genetic variance among *P. ramosa* experimental infections (40.75%) was partitioned between localities (Table 5). Within each pond, the observed structure was distinct. For CH-H isolates, approximately two-thirds of the genetic variance was partitioned among host resistotypes, and one-third was



**Fig. 1** Allelic richness per *Pasteuria ramosa* isolate, averaged across all microsatellite loci ( $\pm$ SE) for each host resistotype in pond RU-BN (left panel) and pond CH-H (right panel). Hatched bars in right panel signify naturally occurring *P. ramosa* infections of *Daphnia magna* native to pond CH-H. Richness values are not shown for native host resistotypes C1 susceptible-C19 resistant or C1 resistant-C19 susceptible, as these resistotypes do not occur in pond CH-H. Letters above bars denote results of Tukey-Kramer HSD *post hoc* comparisons within each pond. Groups not connected by the same letter are significantly different. Dashed horizontal lines with shaded region in the background of each panel represent the overall mean allelic richness for each pond  $\pm$ SE.

**Table 3** Genetic diversity of *Pasteuria ramosa* by source, host resistotype and host clone ( $\pm$ SE)

<i>P. ramosa</i> source	Allelic richness	Gene diversity	Host resistotype	Allelic richness	Gene diversity	Host clone	Allelic richness	Gene diversity	
RU-BN Experimental infections	2.06 $\pm$ 0.06	0.55 $\pm$ 0.07	C1 Susceptible-C19 Susceptible	2.08 $\pm$ 0.12	0.58 $\pm$ 0.06	HU-HO2	2.19 $\pm$ 0.21	0.33 $\pm$ 0.12	
						BE-M10	1.97 $\pm$ 0.12	0.30 $\pm$ 0.13	
			C1 Susceptible-C19 Resistant	2.18 $\pm$ 0.13	0.49 $\pm$ 0.07	FI-AL144	2.13 $\pm$ 0.17	0.61 $\pm$ 0.13	
						FI-KELA-39-9	2.22 $\pm$ 0.18	0.30 $\pm$ 0.12	
			C1 Resistant-C19 Susceptible	2.41 $\pm$ 0.13	0.59 $\pm$ 0.07	FI-XINB3	2.41 $\pm$ 0.17	0.82 $\pm$ 0.05	
						FI-KELA-1810	2.41 $\pm$ 0.18	0.30 $\pm$ 0.13	
			C1 Resistant-C19 Resistant	1.57 $\pm$ 0.09	0.38 $\pm$ 0.12	DE-IINB1	1.75 $\pm$ 0.15	0.49 $\pm$ 0.15	
						IL-M1-1	1.38 $\pm$ 0.09	0.30 $\pm$ 0.11	
		1.31 $\pm$ 0.03	0.54 $\pm$ 0.03	C1 Susceptible-C19 Susceptible	1.61 $\pm$ 0.08	0.39 $\pm$ 0.06	HU-HO2	1.55 $\pm$ 0.11	0.12 $\pm$ 0.08
						BE-M10	1.66 $\pm$ 0.10	0.67 $\pm$ 0.08	
CH-H Natural infections			C1 Susceptible-C19 Resistant	1.36 $\pm$ 0.07	0.09 $\pm$ 0.06	FI-AL144	1.47 $\pm$ 0.10	0.00	
						FI-KELA-39-9	1.25 $\pm$ 0.08	0.06 $\pm$ 0.06	
			C1 Susceptible-C19 Resistant	1.15 $\pm$ 0.06	0.07 $\pm$ 0.05	FI-XINB3	1.22 $\pm$ 0.09	0.00	
						FI-KELA-1810	1.08 $\pm$ 0.08	0.15 $\pm$ 0.10	
			C1 Resistant-C19 Resistant	1.09 $\pm$ 0.04	0.09 $\pm$ 0.06	DE-IINB1	1.11 $\pm$ 0.05	0.06 $\pm$ 0.06	
						IL-M1-1	1.08 $\pm$ 0.04	0.12 $\pm$ 0.08	
		1.28 $\pm$ 0.05	0.56 $\pm$ 0.03	C1 Susceptible-C19 Susceptible	1.47 $\pm$ 0.08	0.15 $\pm$ 0.08	CH-H-705	n/a	
						CH-H-146	1.33 $\pm$ 0.14	n/a	
						CH-H-148	1.08 $\pm$ 0.08	n/a	
						CH-H-309	1.91 $\pm$ 0.22	n/a	
			C1 Resistant-C19 Resistant	1.05 $\pm$ 0.03	0.04 $\pm$ 0.04	CH-H-445	n/a		
					CH-H-524	1.58 $\pm$ 0.14	n/a		
					CH-H-528	1.08 $\pm$ 0.08	n/a		
					CH-H-530	1.00 $\pm$ 0.00	n/a		
					CH-H-531	1.08 $\pm$ 0.08	n/a		
					CH-H-532	1.00 $\pm$ 0.00	n/a		
						1.08 $\pm$ 0.08	n/a		

Allelic richness values are averaged per locus per host individual and were calculated based on the comprehensive allelic dataset. Gene diversities were calculated based on the estimated haplotype dataset.

**Table 4** Linkage disequilibrium/recombination statistics for all isolates, RU-BN isolates and CH-H isolates. Statistics are described in detail in the text

	$I_A$	$P(I_A)$	$\bar{r}_d$	$P(\bar{r}_d)$	Incompatible locus pairs (%)	Locus pairs in significant linkage disequilibrium (%)
OVERALL	2.59	<0.001	0.26	<0.001	72.3	85.4
RU-BN	2.10	<0.001	0.21	<0.001	58.2	21.8
CH-H	9.11	<0.001	0.91	<0.001	1.8	49.1

**Table 5** Results of separate AMOVAS run for all *Pasteuria ramosa* experimental infections, experimental infections from locality RU-BN, and experimental infections from locality CH-H. Analyses were based on the comprehensive allelic dataset

	Source of variation	d.f.	Sum of squares	Variance component	Percent variation	Fixation indices	P-value
All Isolates	Among localities	1	181.75	$Va = 6.52$	40.75	$F_{CT} = 0.408$	0.0284
	Among resistotypes within localities	6	152.08	$Vb = 3.17$	19.85	$F_{SC} = 0.335$	0.0000
	Among isolates within resistotypes	40	252.00	$Vc = 6.30$	39.40	$F_{ST} = 0.606$	0.0000
RU-BN	Among resistotypes	3	45.50	$Va = 0.01$	0.13	$F_{CT} = 0.001$	0.5455
	Among host clones within resistotypes	4	60.33	$Vb = 2.15$	19.95	$F_{SC} = 0.201$	0.0029
	Among isolates within host clones	16	138.00	$Vc = 8.63$	79.92	$F_{ST} = 0.200$	0.0000
CH-H	Among resistotypes	3	106.58	$Va = 5.46$	66.97	$F_{CT} = 0.670$	0.0352
	Among host clones within resistotypes	4	11.00	$Vb = 0.03$	0.34	$F_{SC} = 0.010$	0.4614
	Among isolates within host clones	16	42.67	$Vc = 2.67$	32.69	$F_{ST} = 0.673$	0.0000

partitioned among individual isolates within host clones. None of the variance of CH-H isolates was explained by host clone. In contrast, host resistotype explained approximately 20% of the genetic variation among RU-BN isolates, with the rest of the variance partitioned among individual isolates (79.92%). None of the genetic variance of RU-BN isolates was explained by host resistotype.

Patterns observed in both genetic clustering analyses were consistent with AMOVA results. The dendrogram based on genetic distances among isolates illustrates two well-supported clusters that segregate perfectly based on geographic origin (Fig. 2). Within the CH-H cluster are two smaller clusters that segregate based on the resistotype of the host—one cluster containing isolates exclusively from C1-susceptible hosts, the other containing all isolates from C1-resistant hosts. None of the structure among *P. ramosa* isolates from the CH-H pond was explained by the C19 resistance phenotype of the host. In contrast to the CH-H pond, *P. ramosa* isolates from RU-BN showed no structure based on host phenotype but did exhibit some structure based on host clone (e.g. clones FI-AL1-4-4, FI-KELA-39-9, and IL-M1-1).

Bayesian clustering based on the estimated haplotype dataset shows a nearly identical pattern (Fig. 3). The  $\Delta K$  method identified a three-cluster solution as

best explaining the structure in the data. One cluster contained all haplotypes from RU-BN. The other two clusters contained all samples from CH-H and were segregated by host resistotype in exactly the same way as the dendrogram results discussed above. In both the dendrogram of the comprehensive allelic dataset and the Bayesian clustering analysis of the estimated haplotype dataset, natural *P. ramosa* isolates from the CH-H pond segregated based on host resistotype in exactly the same way as experimental infections (Figs 2 and 3).

## Discussion

This study substantially expands our understanding of the population structure and life history of *Pasteuria ramosa*, a model parasite of *Daphnia* species. We used *P. ramosa* spore banks from the sediments of two geographically distant ponds to experimentally infect a panel of *D. magna* host clones whose resistance phenotypes were previously known and found that experimental infections were comparable to natural infections from the same spore source. We observed relatively high allelic and haplotypic diversity that was structured between localities and among host resistance phenotypes. We also observed evidence for recombination,

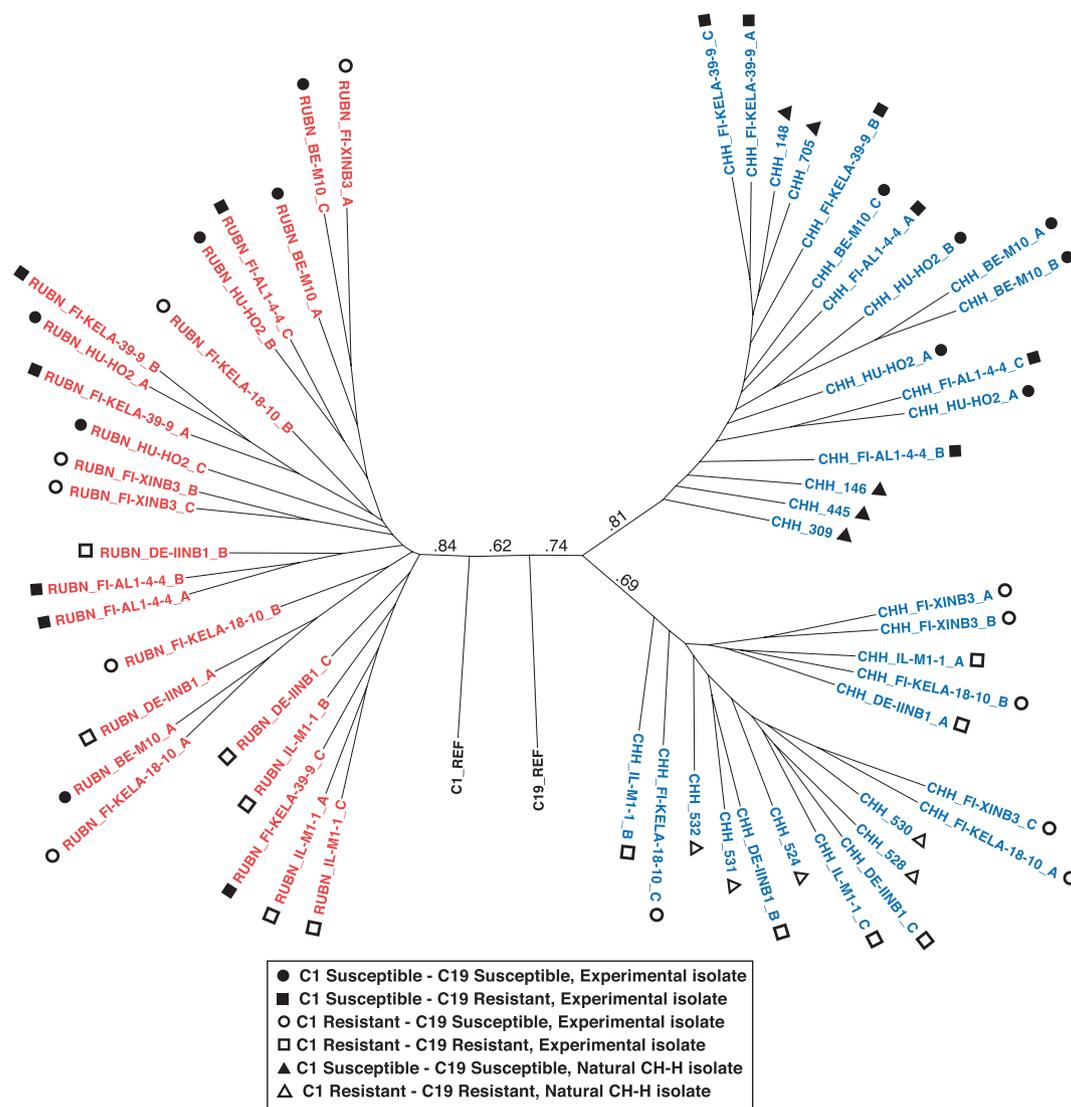


Fig. 2 Dendrogram of genetic distances among *Pasteuria ramosa* isolates based on the comprehensive allelic dataset. Branch tips are labelled with the host clone, colour-coded by pond (Red = RU-BN, Blue = CH-H). Host names are followed by an icon denoting the resistotype of the host, as described in the legend. Laboratory clones C1 and C19 are included for reference. Branch lengths are scaled to genetic distance, and select branches with bootstrap support greater than 600/1000 are labelled accordingly.

suggesting that *P. ramosa* is capable of generating evolutionary novelty by exchanging genetic material horizontally among genomes.

#### Genetic diversity

The large majority of infected *Daphnia magna* individuals included in this study contained multiple *P. ramosa* haplotypes, as evidenced by the presence of multiple alleles per locus. Isolates from pond RU-BN had higher overall allelic richness than those from CH-H, and a higher number of *P. ramosa* types per host. Moreover, *D. magna* hosts that were susceptible to *P. ramosa* clones C1 & C19 were infected by significantly more strains from the sediment

spore bank than hosts that were resistant to both clones. This was true for both localities surveyed, as well as for natural infections of native hosts from pond CH-H, suggesting that the host resistotype, as quantified by two laboratory clones of *P. ramosa*, may be generally representative of *D. magna* resistance to *P. ramosa* and may govern patterns of host/parasite interaction in natural populations.

Coinfection of multiple parasite strains has been reported previously in this system (Mouton *et al.* 2007) and is common in natural populations of parasitized hosts (Read & Taylor 2001). Nonetheless, the high prevalence of coinfection observed in this study is striking, and may have important consequences for the ecology

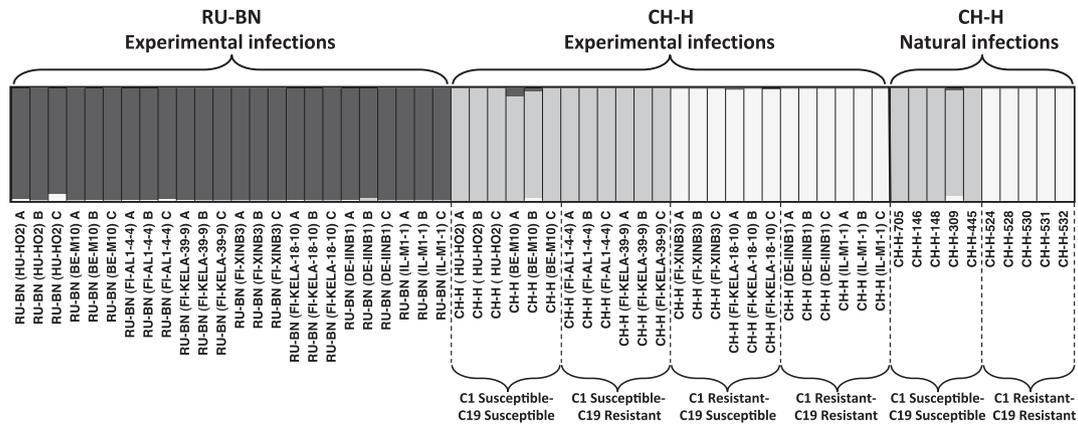


Fig. 3 Graphical summary of Bayesian clustering results based on the estimated haplotype dataset. Haplotypes were assigned among three genetic clusters represented by different shades of grey. Isolates are ordered based on their geographic origin (labels above) and the clone and resistotype of their host (labels below). Note: Bayesian clustering analysis was performed both with and without the inclusion of duplicate haplotypes. Since the results were qualitatively indistinguishable, the results plotted here include all samples, to illustrate the cluster assignments of isolates from each host.

and evolution of this system. In contrast to the predictions of numerous models and observations made in other host-parasite systems, Ben-Ami *et al.* (2008) found that *D. magna* experimentally infected by multiple strains of *P. ramosa* were more fecund than those infected by single strains of the parasite. The authors suggest that this may be evidence of scramble competition, whereby multiple parasite isolates hinder each other leaving more host resources to be channelled to fecundity. If these results are generally applicable, multiple infections may provide a means for *D. magna* populations to persist in the face of extremely high prevalence of parasitism by *P. ramosa*. Additionally, in populations where nearly all hosts become infected, the higher fecundity of multiply-infected hosts suggests that those which are susceptible to a broader variety of *P. ramosa* strains may actually have a fitness advantage over hosts with a narrower range of susceptibility. Expanding upon the results of the present study, such hypotheses can be explored in natural populations of *P. ramosa* using localities, hosts and isolates exhibiting differing degrees of intrahost diversity.

#### Linkage disequilibrium and recombination

There was significant linkage disequilibrium observed across isolates within both localities, and nearly 20% of all observed haplotypes occurred independently in more than one isolate, indicating that *P. ramosa* engages in significant clonal reproduction. This is hardly surprising, as binary fission is the default reproductive mechanism for all prokaryotes, and only a subset of species are able to transmit genetic material horizontally via parasexual processes. More noteworthy is the fact that patterns of LD were much weaker than would be

expected under purely clonal reproduction. In addition, nearly 75% of all possible pairs of loci were incompatible, meaning that the combination of haplotypes observed at those loci could only have arisen through recombination or homoplasmy. These measures of LD and recombination are based on haplotypes that were, in most cases, estimated from mixed isolates. Although the microsatellite loci used in this study are highly quantitative (Ben-Ami *et al.* 2008), and our method of haplotype estimation was conservative, erroneous haplotype reconstruction could obscure patterns of LD and generate spurious incidents of recombination. However, the magnitude of the observed patterns is unlikely to be the result of error and homoplasmy alone. Rather, these results provide strong evidence for relatively frequent genetic recombination among *P. ramosa* genomes. A departure from LD and the presence of incompatible locus pairs are both highly conservative criteria for the detection of recombination, and there are numerous examples of bacterial species known to commonly engage in horizontal gene transfer yet still exhibit significant LD (Maynard Smith 1994; Spratt & Maiden 1999). It follows that if patterns of LD are weak, as is the case in this study, recombination must be quite common. This conclusion could hold significant implications for the ecology and evolution of this system. If *P. ramosa* is able to generate novel biological variation by shuffling pathogenicity genes among genomes, this could dramatically increase its evolutionary rate and shift the balance of host-parasite coevolution. Alternatively, if the mechanisms underlying infection are under simple genetic control (e.g. a single locus), then recombination among distinct lineages could decouple the infection phenotype (infectotype) from neutral genetic diversity without affecting the underlying phenotypes

or the associated ecological and evolutionary host-parasite dynamics. To determine the significance of recombination, future studies will need to ascertain how commonly it occurs between different infectotypes of *P. ramosa* and whether such recombination fundamentally alters the infectotypes of subsequent generations. The RU-BN population—with low levels of LD, a high proportion of incompatible loci and no structure by host resistotype—could provide evidence of recombination among different *P. ramosa* infectotypes and, correspondingly, may be a good candidate for quantitative genetic studies of parasite traits such as infectivity and virulence.

### Population structure

*Pasteuria ramosa* isolates were strongly differentiated between the two geographically distant ponds, as evidenced by AMOVA results and the complete segregation by pond observed with both genetic clustering methods. As only two distantly separated localities were included in this study, it is not possible to estimate the spatial scale of population structure. Nonetheless, the mutual exclusivity of isolates from these two localities conclusively indicates that gene flow of *P. ramosa* is spatially limited, and populations are not panmictic. All else being equal, spatially restricted gene flow will reduce the evolutionary potential of parasites. If gene flow is restricted, local adaptation of parasites to their hosts is expected to be weak or absent, and local extinction is more likely (Barrett *et al.* 2008). The negative consequences of spatially restricted gene flow in *P. ramosa* may be mitigated to some extent by the potential for recombination. However, there is only evidence for very weak local adaptation in the *D. magna*-*P. ramosa* system (Ebert *et al.* 1998), and this could be due in part to the spatial structure of the parasite population.

Dispersal of *P. ramosa* spores could occur via the same mechanisms thought to be important for the host (Figuerola & Green 2002). A phylogeographic study of *D. magna* across Europe found genetic differentiation among localities at both regional and local scales, but also observed haplotypes with very broad geographic distribution, indicating the potential for long-distance dispersal, most likely by birds (De Gelas & De Meester 2005). If parasitized *D. magna* were ingested by migratory waterfowl, it is quite probable that the highly resilient endospores of *P. ramosa* could survive passage through the gut and be excreted in another distant pond. Clonemates of *P. ramosa* have been sampled from geographically distant localities across Europe (Mouton *et al.* 2007; J. Andras, unpublished data). Combined with the results of this study, these observations are consistent with an overall population structure similar

to the host: potentially strong spatial differentiation, coupled with occasional long-distance dispersal events. In future studies, the addition of more localities distributed throughout the range will resolve spatial patterns of population structure and inform hypotheses regarding the mechanisms of dispersal and gene flow of *P. ramosa*.

In addition to geographic structure, isolates from pond CH-H were differentiated into two distinct groups based on the resistotype of the host. Although hosts were categorized into four resistotypes based on their resistance to two distinct laboratory clones of *P. ramosa* (C1 & C19—Table 1), experimental isolates clustered completely based on host resistance to C1 and not C19. Moreover, this same pattern was observed for natural infections of native hosts, indicating that it is representative of host-parasite interactions in the natural population. *Pasteuria ramosa* isolates from pond CH-H were not segregated among host clones within resistotypes, suggesting that the resistance phenotypes quantified by C1 and C19 characterize general attributes of resistance and infection. The mechanisms of infection by and resistance to C1 and C19 are of particular interest because these clones have been and continue to be used in a number of studies investigating the quantitative and molecular genetics of this host-pathogen system (Luijckx *et al.* 2011, 2012; Duneau *et al.* 2011; Hall & Ebert 2012). As these mechanisms become better understood, it will be important to investigate their variation in natural populations, along with other unique mechanisms not yet represented in laboratory clones. The clear clustering of isolates by host resistotype in pond CH-H suggests three hypotheses relating to the infection mechanisms used by *P. ramosa* in this population:

- 1 The C1 infection phenotype (infectotype) of *P. ramosa* is present and polymorphic in pond CH-H, as multiple distinct isolates segregate perfectly according to the resistance or susceptibility of the host to C1.
- 2 The C19 infectotype is not present in pond CH-H, as isolates from both genetic clusters are able to infect both C19-susceptible and C19-resistant hosts.
- 3 There is a third infectotype of *P. ramosa* present in pond CH-H that has an opposite infection pattern to C1.

The simple pattern observed in pond CH-H of two host resistotypes and two corresponding parasite infectotypes with mutually exclusive compatibility conforms to the matching-genotype model of host-parasite coevolution, which is predicted to support coevolutionary cycling via negative frequency-dependent selection (Agrawal & Lively 2002). As such, pond CH-H could

serve as an excellent natural system for future investigations of the coevolutionary dynamics of *D. magna* and *P. ramosa*.

In contrast to CH-H, pond RU-BN showed no structure whatsoever based on host resistotype, although 20% of the genetic variation was partitioned among host clones (Table 5). The disparity in host-mediated population structure between the two localities is striking, and there are several possible explanations for this pattern. It is possible that neutral marker diversity in the RU-BN population has been shuffled by recombination and is unlinked with the parasite's infection phenotype (infectotype), whereas the linkage between neutral diversity and biological diversity has not been broken by recombination in the CH-H population. The RU-BN pond is more than seven times larger in surface area than CH-H, and potentially older, since the CH-H pond was completely drained for an extended period of time prior to 1962. Thus, the effective population size of *P. ramosa* at RU-BN is likely larger, as evidenced by the higher genetic diversity observed there, and would provide more opportunities for different infectotypes to recombine.

Larger and older ponds also tend to have more genetically diverse populations of *D. magna* (Haag *et al.* 2005; Vanoverbeke *et al.* 2007). If there are host resistotypes present in the RU-BN pond that allow different infectotypes of *P. ramosa* to coinfect, this would facilitate the shuffling of neutral and phenotypic diversity, as parasexual processes must almost certainly occur during the vegetative stage within the host and not during the resting endospore stage. In the smaller CH-H pond, the strict segregation of *P. ramosa* between the two native *D. magna* resistotypes (C1 & C19-susceptible and C1 & C19-resistant) could prevent these parasite lineages from coming into contact, thereby precluding recombination.

Alternatively, it is possible that the RU-BN population consists of a diversified lineage of *P. ramosa* with one or more broad infectotypes, capable of infecting most or all host clones used in this study. However, despite the lack of differentiation among host resistotypes, the observed structure among host genotypes indicates that infection is not entirely random in the RU-BN pond. This suggests that, although the host resistotypes defined by the two *P. ramosa* laboratory clones explain the majority of the population structure observed in the CH-H pond, they do not capture the entire story. There are likely other resistance mechanisms that are polymorphic within the host resistotypes used in this study that account for the structure among host genotypes observed in the RU-BN population. These alternative hypotheses could be resolved by passaging or cloning multiple CH-H and RU-BN isolates

to reduce their diversity and characterizing their infectotypes relative to a panel of different host clones. This information would indicate whether the infectotype of *P. ramosa* is generally consistent among members of a population genetic cluster, or whether clusters can be composed of strains with different patterns of infectivity.

## Conclusions

By experimentally infecting a panel of nonnative hosts with spores from pond sediments, we were able to sample natural populations of *Pasteuria ramosa* with an unbiased design that resolved structure at the levels of geography, host resistance phenotype and host genotype. Isolates of *P. ramosa* were strongly differentiated between localities, indicating spatially restricted gene flow and limited dispersal. Yet, we also found evidence for substantial recombination among *P. ramosa* genomes, indicating that lineages are not purely clonal and that genes of functional significance might potentially be shuffled in natural populations. We also observed strong population structure and significantly different levels of intrahost parasite diversity that were correlated with the resistance phenotype of the host. This filtering effect by host resistotype is consistent with laboratory studies identifying strong specificity in this host-parasite system (Luijckx *et al.* 2011), and this is the first time, to our knowledge, such a pattern has been reported for a nonhuman parasite. Patterns of diversity and population structure observed in experimental infections were highly congruent with natural infections from the same pond. These results indicate that the resistance phenotypes quantified based on only two distinct laboratory clones are of general relevance in natural populations and highlight the utility of experimental infection from the spore bank as a means of investigating the population structure of parasites whose within-host stages may be nonrepresentative, ephemeral or difficult to sample. Combined with the substantial body of work investigating the ecology and evolution of *P. ramosa* and *D. magna*, the results of this study will help inform and refine coevolutionary models, provide guidance for the sampling design of future studies of natural populations and open a number of new interesting questions for research on this model host-parasite system.

## Acknowledgements

This work was funded by the Swiss National Science Foundation. The authors would like to thank Yan Galimov for help with sample collection and Jürgen Hottinger and Urs Stiefel for assistance in the laboratory.

## References

- Agapow PM, Burt A (2001) Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, **1**, 101–102.
- Agrawal A, Lively CM (2002) Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evolutionary Ecology Research*, **4**, 79–90.
- Barrett LG, Thrall PH, Burdon JJ, Linde CC (2008) Life history determines genetic structure and evolutionary potential of host-parasite interactions. *Trends in Ecology and Evolution*, **23**, 678–685.
- Ben-Ami F, Mouton L, Ebert D (2008) The effects of multiple infections on the expression and evolution of virulence in a *Daphnia*-endoparasite system. *Evolution*, **62**, 1700–1711.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate – A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.
- Bush AO, Lafferty KD, Lotz JM, Shostak AW (1997) Parasitology meets ecology on its own terms: Margolis et al. revisited. *Journal of Parasitology*, **83**, 575–583.
- Butlin R (2002) The costs and benefits of sex: new insights from old asexual lineages. *Nature Reviews Genetics*, **3**, 311–317.
- Carius HJ, Little TJ, Ebert D (2001) Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. *Evolution*, **55**, 1136–1145.
- Criscione CD, Blouin MS (2004) Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution*, **58**, 198–202.
- Davies CM, Webster JP, Kruger O, Munatsi A, Ndamba J, Woolhouse MEJ (1999) Host-parasite population genetics: a cross-sectional comparison of *Bulinus globosus* and *Schistosoma haematobium*. *Parasitology*, **119**, 295–302.
- De Gelas K, De Meester L (2005) Phylogeography of *Daphnia magna* in Europe. *Molecular Ecology*, **14**, 753–764.
- Decaestecker E, Gaba S, Raeymaekers JAM *et al.* (2007) Host-parasite 'Red Queen' dynamics archived in pond sediment. *Nature*, **450**, 870–U816.
- Dent EA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Duncan AB, Mitchell SE, Little TJ (2006) Parasite-mediated selection and the role of sex and diapause in *Daphnia*. *Journal of Evolutionary Biology*, **19**, 1183–1189.
- Duneau D, Luijckx P, Ben-Ami F, Laforsch C, Ebert D (2011) Resolving the infection process reveals striking differences in the contribution of environment, genetics and phylogeny to host-parasite interactions. *BMC Biology*, **9**, doi:10.1186/1741-7007-9-11.
- Dybdahl MF, Lively CM (1996) The geography of coevolution: comparative population structures for a snail and its trematode parasite. *Evolution*, **50**, 2264–2275.
- Ebert D, Rainey P, Embley TM, Scholz D (1996) Development, life cycle, ultrastructure and phylogenetic position of *Pasteuria ramosa* Metchnikoff 1888: rediscovery of an obligate endoparasite of *Daphnia magna* Straus. *Philosophical Transactions of the Royal Society of London. Series B-Biological Sciences*, **351**, 1689–1701.
- Ebert D, Zschokke-Rohringer CD, Carius HJ (1998) Within- and between-population variation for resistance of *Daphnia magna* to the bacterial endoparasite *Pasteuria ramosa*. *Proceedings of the Royal Society of London. Series B-Biological Sciences*, **265**, 2127–2134.
- Estabrook GF, Landrum L (1975) A simple test for the possible simultaneous evolutionary divergence of two amino acid positions. *Taxon*, **24**, 609–613.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes, Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
- Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.
- Felsenstein J (1993) PHYLIP (*Phylogeny Inference Package*) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, Washington.
- Figuerola J, Green AJ (2002) Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biology*, **47**, 483–494.
- Gandon S (2002) Local adaptation and the geometry of host-parasite coevolution. *Ecology Letters*, **5**, 246–256.
- Haag CR, Riek M, Hottinger JW, Pajunen VI, Ebert D (2005) Genetic diversity and genetic differentiation in *Daphnia* metapopulations with subpopulations of known age. *Genetics*, **170**, 1809–1820.
- Hall MD, Ebert D (2012) Disentangling the influence of parasite genotype, host genotype and maternal environment on different stages of bacterial infection in *Daphnia magna*. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3176–3183.
- Hickey DA (1982) Selfish DNA – a sexually-transmitted nuclear parasite. *Genetics*, **101**, 519–531.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford)*, **23**, 1801–1806.
- Kaltz O, Shykoff JA (1998) Local adaptation in host-parasite systems. *Heredity*, **81**, 361–370.
- Kluettgen B, Duellmer U, Engels M, Ratte HT (1994) ADaM, an artificial freshwater for the culture of zooplankton. *Water Research*, **28**, 743–746.
- Lively CM (1999) Migration, virulence, and the geographic mosaic of adaptation by parasites. *The American Naturalist*, **153**, S34–S47.
- Luijckx P, Ben-Ami F, Mouton L, Du Pasquier L, Ebert D (2011) Cloning of the unculturable parasite *Pasteuria ramosa* and its *Daphnia* host reveals extreme genotype-genotype interactions. *Ecology Letters*, **14**, 125–131.
- Luijckx P, Fienberg H, Duneau D, Ebert D (2012) Resistance to a bacterial parasite in the crustacean *Daphnia magna* shows Mendelian segregation with dominance. *Heredity*, **108**, 547–551.
- Maynard Smith J (1994) Estimating the minimum rate of genetic-transformation in bacteria. *Journal of Evolutionary Biology*, **7**, 525–534.
- Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? *Proceedings of the National Academy of Sciences of the United States of America*, **90**, 4384–4388.

- Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.
- Mouton L, Ebert D (2008) Variable-number-of-tandem-repeats analysis of genetic diversity in *Pasteuria ramosa*. *Current Microbiology*, **56**, 447–452.
- Mouton L, Nong G, Preston JF, Ebert D (2007) Variable-number tandem repeats as molecular markers for biotypes of *Pasteuria ramosa* in *Daphnia* spp. *Applied and Environmental Microbiology*, **73**, 3715–3718.
- Nadler SA (1995) Microevolution and the genetic-structure of parasite populations. *Journal of Parasitology*, **81**, 395–403.
- Nei M (1987) Molecular Evolutionary Genetics. In: *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.
- Read AF, Taylor LH (2001) The ecology of genetically diverse infections. *Science*, **292**, 1099–1102.
- Seger J, Hamilton WD (1988) Parasite and sex. In: *The Evolution of Sex* (ed. Levin REMBR), pp. 176–193. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Spratt BG, Maiden MCJ (1999) Bacterial population genetics, evolution and epidemiology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **354**, 701–710.
- Thompson JN (1994) *The Coevolutionary Process*. University of Chicago Press, Chicago, Illinois.
- Tibayrenc M, Kjellberg F, Arnaud J *et al.* (1991) Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. *Proceedings of the National Academy of Sciences of the USA*, **88**, 5129–5133.
- Vanoverbeke J, De Gelas K, De Meester L (2007) Habitat size and the genetic structure of a cyclical parthenogen, *Daphnia magna*. *Heredity*, **98**, 419–426.

---

The research presented here was conducted by J.P.A. as a component of his postdoctoral research in the laboratory of D.E.

---

#### Data accessibility

Estimated haplotypes and allele presence-absence data are available via DRYAD; doi:10.5061/dryad.r406f.