MOLECULAR ECOLOGY

Molecular Ecology (2011) **20**, 2510–2524 doi: 10.1111/j.1365-294X.2011.05113.x

More than meets the eye: detecting cryptic microgeographic population structure in a parasite with a complex life cycle

CHARLES D. CRISCIONE,*[1] ROMÁN VILAS,†[1] ESPERANZA PANIAGUA‡ and MICHAEL S. BLOUIN§

*Department of Biology, Texas A&M University, 3258 TAMU, College Station, TX 77843, USA, †Departamento de Genética, Universidad de Santiago de Compostela, 27002 Lugo, Galicia, Spain, ‡Laboratorio de Parasitología, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Galicia, Spain, §Department of Zoology, Oregon State University, 3029 Cordley Hall, Corvallis, OR 97331, USA

Abstract

Nonrandom recruitment of parasites among hosts can lead to genetic differentiation among hosts and mating dynamics that promote inbreeding. It has been hypothesized that strictly aquatic parasites with intermediate hosts will behave as panmictic populations among hosts because ample opportunity exists for random mixing of unrelated individuals during transmission to the definitive host. A previous allozyme study on the marine trematode *Lecithochirium fusiforme* did not support this hypothesis; in that, there was genetic differentiation among, and significant heterozygote deficiencies within, definitive hosts. We revisit this system and use microsatellites to obtain multilocus genotypes. Our goal was to determine whether cryptic subgroups and/or the presence of clones could account for the apparent deviation from 'panmixia'. We find strong evidence for cryptic subdivision (three genetic clusters) that causes the Wahlund effect and differentiation among definitive hosts. After accounting for these cryptic groups, we see panmictic genetic structure among definitive hosts that is consistent with the 'high mixing in aquatic habitats' hypothesis. We see evidence for cotransmission of clones in all three clusters, but this level of clonal structure did not have a major impact in causing deviations from Hardy–Weinberg equilibrium, and only affected genetic differentiation among hosts in one cluster. A cursory examination of the data may have led to incorrect conclusions about nonrandom transmission. However, it is obvious in this system that there is more than meets the eye in relation to the actual make-up of parasite populations. In general, the methods we employ will be useful for elucidating hidden patterns in other organisms where cryptic structure may be common (e.g. those with limited morphology or complex life histories).

Keywords: cryptic structure, population genetics, transmission, Trematoda, Wahlund effect

Received 3 December 2010; revision received 9 March 2011; accepted 21 March 2011

Introduction

A priori delimitation of a deme (i.e. a cohesive genetic population) can be hampered in organisms with limited morphology or complex life histories. For example, demographic populations of adult parasites can be isolated geographically and isolated among definitive hosts (where there is parasite sexual reproduction) within geographical locations. Thus, the question of what constitutes a deme for macroparasites has been raised repeatedly (Lydeard *et al.* 1989; Nadler 1995; Bush *et al.* 2001; Jarne & Theron 2001; Sire *et al.* 2001;

Correspondence: Charles D. Criscione, Fax: 979 845 2891; E-mail: ccriscione@mail.bio.tamu.edu

[1]The first two authors share primary authorship of the paper.

Criscione *et al.* 2005; Criscione & Blouin 2006). In terms of the distribution of parasites among hosts, the component population refers to all the parasites in a population of hosts, while the infrapopulation refers to the parasites in an individual host (Bush *et al.* 1997). The extent to which parasite populations are genetically subdivided at some level below that of the component population (e.g. into infrapopulations or groups of infrapopulations) has enormous practical consequences for their evolution. This is because (i) subdivision will influence the component population effective size, and thus overall levels of genetic diversity (Criscione & Blouin 2005; Prugnolle *et al.* 2005a), and (ii) subdivision changes the frequency distribution of single and multilocus genotypes (MLGs), thus altering opportunities for response to selection and adaptive evolution. For example, cotransmission of related individuals between hosts can greatly increase the rate of response to selection for rare recessive alleles in the component population, such as in the evolution of drug resistance (Cornell *et al.* 2003; Schwab *et al.* 2006).

We know very little about the extent to which adult parasite component populations are typically subdivided at lower levels. These levels could include individual infrapopulations, groups of infrapopulations (e.g. social or family groups of hosts, or hosts in particular microenvironments), temporal groups or even groups determined by intermediate host specificity, but that lack individual host affiliation in the definitive host. The answer depends on the life cycle and ecology of transmission of the parasite. For example, if offspring from the same definitive host infrapopulation frequently transmit together (i.e. clumped transmission) to new definitive hosts or reinfect their natal host, there will be substructuring among hosts. Alternatively, if there is substantial mixing of offspring from different infrapopulations before transmission to the next definitive host, there will be no demic structure below the component population level (Criscione *et al.* 2005; Prugnolle *et al.* 2005a; Criscione & Blouin 2006).

Among digenean trematodes, the asexual reproductive phase that occurs in mollusc first intermediate hosts adds another dynamic that influences the within- and among-host parasite population structure. In particular, the theoretical work by Prugnolle *et al.* (2005a) shows that a high variance in clonal reproductive success can decrease the effective size of a parasite component population and increase differentiation among definitive hosts. Furthermore, transmission that leads to clones in the same definitive host can increase the potential for inbreeding as mating between clones is equivalent to selfing in a hermaphroditic species (which most digeneans are). It is important to note that the clones cannot persist over generations because of an obligate sexual phase in the definitive host (Fig. 1).

Little is known about the population genetic structure of digeneans in fully aquatic systems. It has been predicted that strictly aquatic parasites with several intermediate hosts will behave as large, panmictic populations (at the component population level) because ample opportunity exists for mixing of unrelated individuals or clones during transmission to the definitive host (Criscione & Blouin 2006). However, there have been few explicit empirical tests of this hypothesis at the level of the definitive host (e.g. Criscione & Blouin 2006) with most work being conducted in second intermediate hosts (Rauch *et al.* 2005; Keeney *et al.* 2007a,b; Leung *et al.* 2009). Thus far, these studies, which utilized multilocus microsatellite genotypes, support the above hypothesis as few clones have been detected (>97% of genotyped worms are unique genotypes), and no deviations from Hardy–Weinberg equilibrium were observed across infrapopulations (except for null alleles suspected by Leung *et al.* 2009). One apparent exception to the 'high mixing in aquatic habitats' pattern was reported from an

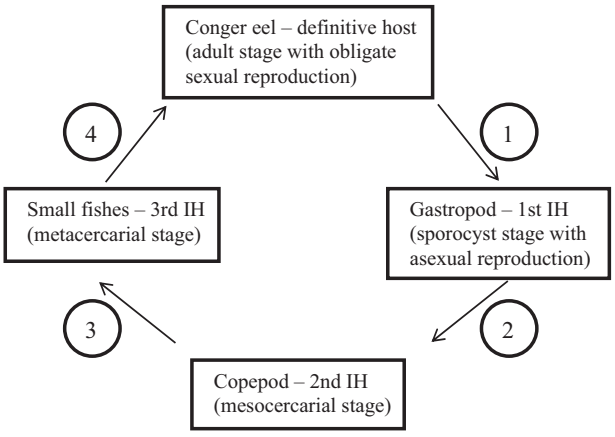

Fig. 1 Generalized life cycle of *Lecithochirium* spp. (Matthews & Matthews 1988). Adult flukes (with obligate sexual reproduction) infect the stomachs of conger eels. Eggs pass into the water via host faeces (1). Eggs are eaten by a gastropod (*Gibbula* spp.) where a miracidium becomes a mother sporocyst, which generates several daughter sporocysts (asexual reproduction). The daughter sporocycsts in turn produce many cercariae (asexual reproduction), which leave the snail (2) and penetrate a copepod. Small fish serve as third intermediate hosts (IH) upon eating an infected copepod (3). The conger eel definitive host becomes infected (4) upon ingesting an infected third intermediate host. Notice that steps 1, 2, 3 and 4 provide chances for mixing of parasite offspring before recruitment back into a definitive host. However, only steps 2, 3 and 4 allow mixing of clonal flukes produced in the snail before recruitment into the definitive host. Clones do not persist across generations because of the obligate sexual phase in the definitive host.

allozyme study where the marine digenean *Lecithochirium fusiforme* was collected from a population of conger eels (*Conger conger*) in a single bay in Spain (Vilas *et al.* 2003).

Lecithochirium fusiforme is a common stomach parasite of conger eels (definitive host) in the Atlantic Ocean off the coast of Europe (Gibson & Bray 1986). Infrapopulations can vary in intensity from the teens to the low hundreds, and prevalence typically averages about 70% throughout the year, although some seasonality has been suggested (Vilas & Paniagua 2004). The exact life cycle is not known, but sympatric congeners cycle through four hosts where gastropods (*Gibbula* spp.), copepods, and small fishes are the first, second and third intermediate hosts, respectively before reaching the definitive host (Fig. 1). Adult lifespan in conger eels is unknown, but likely a few months. The longevity (viable for at least 12–15 months) of the encapsulated metacercariae in fish, third intermediate hosts (Matthews & Matthews 1988), may facilitate the accumulation of specimens from several generations of the parasite. Thus, for *L. fusiforme*, one might expect high mixing of parasite offspring and low cotransmission of clones to definitive hosts. Therefore, we expect little subdivision among infrapopulations, and genotypic frequencies in agreement with Hardy–Weinberg expectations in both individual infrapopulations and the component population as a whole. Nevertheless, Vilas *et al.* (2003) found (i) moderate genetic differentiation among infrapopulations and (ii) significant heterozygote deficiencies within infrapopulations that varied in magnitude among loci and among infrapopulations. Taken at face value, these results do not support the hypothesis of high mixing in aquatic habitats. Rather, the results suggest nonrandom recruitment and mating of parasites among infrapopulations. However, the authors ruled out inbreeding as an explanation because three loci were in Hardy–Weinberg equilibrium. They go on to hypothesize that the heterozygote deficits could be caused by the Wahlund effect resulting from the recruitment of parasites from genetically differentiated groups into infrapopulations. This hypothesis, if true, does not necessarily negate the aquatic mixing hypothesis. For example, there could be random mating and no among-infrapopulation structure within each of the genetically differentiated groups. Unfortunately, Vilas *et al.* (2003) could analyse only one or two loci per individual because the parasites are small. Without MLGs, there is little ability to delineate cryptic structure (e.g. reproductively isolated groups) in a sample of individuals. Thus, it was not possible for Vilas *et al.* (2003) to test their hypothesis of cryptic subgroups. The lack of MLGs also precluded Vilas *et al.* (2003) from assessing the role of clonal structure (proportion and cotransmission of clones) in influencing their results. In addition, allozyme artefacts on certain loci, such as null alleles or zymogram misscoring, and the effect of natural selection cannot be ruled out in the data set of Vilas *et al.* (2003).

In this study, we resample *L. fusiforme* from the same population of eels, but this time using microsatellite loci to obtain MLGs. Our goal was to test whether cryptic subgroups and/or the presence of clones could explain the possible causes of the Hardy–Weinberg disequilibrium and deviation from panmixia among infrapopulations observed in the allozyme study of Vilas *et al.* (2003). If microsatellite loci show the same disequilibria, then we can rule out allozyme artefacts. With MLGs, we can test for clonal structure and use patterns of Hardy–Weinberg and linkage disequilibrium to test for cryptic subdivision within the component population. We find that cryptic subdivision (into three genetic clusters) does indeed explain the variable deviations from Hardy–Weinberg equilibrium and the differentiation among infrapopulations. After accounting for the three cryptic groups of parasites, we see a panmictic genetic structure among infrapopulations which is consistent with the 'high mixing in aquatic habitats' after all. We detected a small to moderate number of clones in all three clusters, with only one cluster showing higher clonal structure than that reported in other aquatic systems. We show that the presence of clones in Cluster II decreases F_{IS} within and increases F_{ST} among infrapopulations, as predicted by theory (Prugnolle *et al.* 2005a).

Materials and methods

Sampling and genotyping

Between January and April 2005, specimens identified as *Lecithochirium fusiforme* were collected from 12 conger eels. The fish were caught in an area of approximately 50 km^2 within Ría de Arousa (42°34′N; 8°56′W), a coastal embayment in Galicia, of northwestern Spain. Fish were caught by hook and immediately eviscerated. Transport of abdominal viscera to the laboratory, and collection and identification of parasites were as in the study by Vilas *et al.* (2003) and Vilas & Paniagua (2004). Flukes were stored in 70% ethanol. A total of 520 worms were genotyped at seven microsatellites (loci 18B10, 18A5, 20D1, 14F11, 18B8, 18E2, 14A5A; GenBank accessions DQ413187–DQ413192) as in the study by Vilas *et al.* (2006). Controls with fish DNA were negative. We threw out 13 specimens from the data set because of missing data or ambiguity in scoring. The remaining 507 worms retained in the study were typed at all seven loci (i.e. no missing data). The number of analysed worms per host, and per host within parasite

Table 1 Distribution of MLGs that had significant P_{sex} values. Unless noted, all P_{sex} values (i.e. the probability of getting at least n identical genotypes in the total sample size across all hosts of a given parasite cluster) were significant at $n = 2$

	Cluster I clones†					Cluster II clones												Cluster III clones					Unassigned MLG	
Host ID	*N*	*7*	*30*‡	*40*‡	*42*	*N*	*8*	*24*	*27*	*40*	*47*	*53*	*57*	*69*	*97*	*117*	*142*	*N*	*2*	*9*	*30*	*62*	N§	*1*
A	7					20	3		1						7			3					3	2
B	10					11					1		2	4		1		6					0	
C	1					52				2		1					1	0					1	
D	0					15			2									16			1		0	
E	49					0												2					1	
F	11					42					1	1	2	4		1		1					1	
G	9		1			20											1	6					0	
H	33					7												6				2	1	
I	3					11												28		2	1		2	1
J	14			1	2	7		3								1		7	2				2	
K	44		3			1												2					0	
L	54	2		2		0												0					0	
Sum	235	2	4	3	2	186	3	3	3	2	2	2	2	4	7	2	2	77	2	2	2	2	9	3

†For each cluster, we show *N*, the number of individuals per host, and the distribution of Multilocus genotypes (MLGs) identified as being clonemates (see Materials and methods for criteria). Clone IDs are the italicized numbers, which were numbered consecutively within each cluster, thus MLG *40* in Cluster I is not that same as MLG *40* in Cluster II.
‡MLGs *30* and *40* from Cluster I had significant P_{sex} values at $n = 4$ and 3, respectively. Thus, there may be up to three or two individuals that are the product of a sexual event within these two MLGs, respectively (i.e. not all four or three individuals may be the same clone).
§Nine worms were not assigned unequivocally to any cluster (*Q*-value < 70% to a single cluster). Three of these individuals had the same genotype; however, it is not possible to test whether these individuals are clonemates without population allele frequencies.

clusters (see Results), is given in Table 1 (additional sampling details are in Tables S1 and S2, Supporting information).

Tests for cryptic structure

We observed many individuals that shared the same MLG. Therefore, we first reduced the full data set ($N = 507$) to only one representative of each MLG (reduced data set $N = 332$). We explain the reasoning for this initial reduction in the Discussion. Two methods were used to determine whether there was cryptic structure in this set of unique MLGs. First, we used STRUCTURE v2.1 (Falush *et al.* 2003) to identify the most likely number of groups within the entire dataset. We ran values of $K = 1$–6 with five replications of each run (50 000 burn-in, 100 000 replications) using the correlated allele frequencies and admixture models. As STRUCTURE assumes Hardy–Weinberg and linkage equilibrium among loci, we also used multivariate methods, which do not carry these assumptions. We used the following multivariate methods to visualize potential cryptic structure: PCA in ADEGENET (Jombart 2008), PCoA (Principle Coordinate Analysis) in GENALEX v6.41 (Peakall & Smouse 2006) and AFC in GENETIX v4.05 (Belkhir *et al.* 2004). Qualitatively, all multivariate methods gave very similar results; thus, we only show the PCoA results.

STRUCTURE and the PCoA plot were concordant in identifying three very distinct clusters to which we could confidently assign all but seven MLGs (Fig. 2; details in Results). Thus, we treated each cluster as a separate population in subsequent analyses, as described in the next paragraph. The seven unassigned MLGs (nine individuals) were excluded from further analyses (see Results and discussion).

We estimated F_{ST} (Weir & Cockerham 1984) among the three clusters (i.e. ignoring host boundaries) using FSTAT v2.9.3 (Goudet 1995). F_{ST} values were standardized following Meirmans (2006) by recoding the data set (for details see Criscione & Blouin 2007).

Tests for clones and clonal structure

The low diversity at the loci suggested that we would observe multiple individuals having the same MLG (hereafter 'multicopy MLGs') by chance alone. Thus, we determined whether the seven loci had enough power to identify all genetically unique individuals within each cluster by using GENCLONE v2.0 (Arnaud-Haond & Belkhir 2007) to resample loci and recalculate the number of unique MLGs for all locus combinations from 1 to 7. That the mean number of unique MLGs never reached an asymptote over all combinations of loci from 1 to 7 (data not shown) indicated that we had low

power to distinguish genetically unique individuals in all three clusters (i.e. many unrelated individuals will have identical 7-locus genotypes by chance alone). Therefore, we conducted three analyses to test for the presence of true clones in each of the clusters.

1 We tested whether the total number of unique MLGs observed in each cluster was significantly less than the number expected in a same-sized sample of unrelated individuals drawn from the same allele frequency distribution. We performed this overall test for clonal structure by randomizing alleles among individuals within each cluster and recalculating the number of unique MLGs expected (9 999 randomizations). If clonal structure exists, then the mean number of unique MLGs simulated will be greater than the number of unique MLGs observed. These tests, which implicitly assume random mating, were carried out with GENODIVE (beta version by P. Meirmans; http://www.bentleydrummer.nl/software/software/GenoDive.html).
2 To identify truly clonal individuals within each cluster, we used GENCLONE to calculate P_{sex} values for each multicopy MLG (significance determined at $P_{sex} < 0.05$). P_{sex} is the probability of observing n copies of a MLG in a sample of size N. If the P_{sex} of a multicopy MLG at $n = 2$ is significant, then all copies of that MLG can be considered to be the product of clonal reproduction (Gregorius 2005). If P_{sex} is not significant at $n = 2$, but is significant for some $n > 2$, then one can only reject the hypothesis that more than n-1 clones exist within n copies of that MLG (Gregorius 2005). For example, if a MLG occurs as four copies and P_{sex} is significant at $n = 3$ but not at $n = 2$, then one can conclude that at most three of the four genetically identical individuals can be clonemates. GENCLONE can calculate P_{sex} based on random mating or accounting for inbreeding; however, either calculation yielded the same result in our study in part because of the fact that each cluster had no significant level of inbreeding (see Results).
3 We tested whether those significant multicopy MLGs ($P_{sex} < 0.05$ at $n = 2$) tended to co-occur in the same hosts more often than expected by chance alone. Because of the many hosts involved in the life cycle, evidence that the significant multicopy MLGs also tend to co-occur in the same eel would be strong evidence that they really represent clones. For significant multicopy MLGs at $n = 2$, we tested the null hypothesis of no association between members of a putative clone and the infrapopulation in which each member was collected. If there is clumped transmission of clones, there will be a significant association. Substantial mixing of clones before reaching the definitive host would lead to no association. This analysis requires a contingency table of each putative clone × infrapopulation. For example, in Cluster III, there are 77 individuals (spread over 10 hosts) of which four MLGs with two copies each were significant (Table 1). This creates a 73 (MLGs) × 10 (infrapopulations) contingency table. Because these contingency tables contain many cells with 0 or 1, we used the program RXC (by M. Miller; http://www.marksgeneticsoftware.net/). RXC employs the metropolis algorithm to obtain an unbiased estimate of the exact P-value (i.e. Fisher's exact test) for any sized contingency table. The following Markov chain parameters were used to test significance: 5000 dememorizations, 5000 batches, 5000 permutations per batch. Results of clonal analyses are given in Table 2.

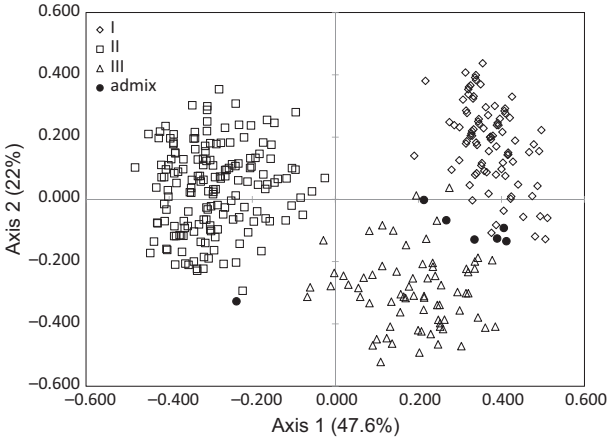

Fig. 2 Cluster assignments based on the STRUCTURE analysis superimposed on the individual-based PCoA analysis in GENALEX v6.41. Both analyses were conducted after collapsing all repeated multilocus genotypes (MLGs) into single individuals. The three clusters are represented by the different shapes. Solid circles correspond to the seven putatively hybrid (admixed) MLGs (i.e. Q-value < 0.7). Notice there is near complete agreement between the two analyses in the assignment of individuals to three clusters.

Tests of the effects of cryptic subdivision and of clones on genetic structure

All estimates and significance tests of F_{IS} and F_{ST} (Weir & Cockerham 1984) were generated using FSTAT (Goudet 1995). For all analyses using F_{ST}, we also report standardized F_{ST} values following the recoding scheme suggested by Meirmans (2006).

We examined the effects of cryptic structure on Hardy–Weinberg and genotypic equilibrium by comparing component populations (i.e. all parasites among all hosts sampled) that were composed of each cluster sepa-

rately to the component population that did not distinguish among clusters. Hardy–Weinberg and genotypic equilibrium were tested with 1000 randomizations of alleles among individuals or of genotypic associations among pairs of loci, respectively. We quantified genotypic disequilibrium within each relevant sample as the number of pairwise combinations of loci that were statistically significant ($P < 0.05$) of the total number of pairwise comparisons. We also used the *Z*-transform test (Whitlock 2005) to combine probabilities across pairwise tests to determine whether there was overall significant genotypic disequilibrium within each relevant sample. To examine the effects of clonal structure at the component population level, the previous tests were analysed with and without all repeated copies of MLGs. In the without tests, all multicopy MLGs (whether significant or not) were reduced to one unique copy over the whole data set. We did not just remove significant MLGs because we wanted to contrast the results from each cluster component population to the component population data set that does not distinguish among clusters, where testing clone significance cannot be done owing to lack of a suitable baseline allele frequency distribution (see Discussion). These results are presented in Table 3.

We also compared among-infrapopulation hierarchical *F*-statistics between a data set that did not distinguish among clusters to that in data sets that were composed of individuals from their respective clusters. Significance of average F_{IS} within hosts and of F_{IT} was tested with 10 000 randomizations of alleles among individuals within infrapopulations and overall, respectively. Differentiation among infrapopulations was tested with 10 000 randomizations of genotypes among infrapopulations. To examine the effects of clonal

Table 2 Summary of the number of MLGs and evidence for clonal structure in the three clusters

	N†	*N*-MLGs‡	*N*-MLGs simulated (*P*-value)§	Observed MLG > 1 copy	*N*-MLGs with P_{sex} < 0.05 at $n = 2$	*P*-value of random association of clones among hosts
Cluster I	235	89	88.5 (0.618)	42	2	0.025
Cluster II	186	165	180.2 (0.0001)	11	11	0.000
Cluster III	77	71	74.1 (0.059)	6	4	0.024

MLGs, multilocus genotypes.
†The total number of individuals genotyped that were assigned to each cluster.
‡The number of unique MLGs observed within each cluster.
§The mean number of unique MLGs simulated after 9 999 randomizations of alleles among individuals (in parentheses is the *P*-value, the proportion of simulated values that were less than or equal to the observed value).

Table 3 F_{IS}, genotypic disequilibrium (GD) among loci, gene diversity (H_s) and allelic richness (*A*) for the component population as a whole (i.e. treating all parasites within all hosts as one population) and for the component population of each cluster separately

	Sample size†		F_{IS}		GD‡		H_s§		*A*¶	
	w††	w/out	w	w/out	w	w/out	w	w/out	w	w/out
All	498	325	0.350***	0.311***	21 (21)***	21 (21)***	—	—	—	—
Cluster I	235	89	0.025	0.079*	1 (15)	0 (15)	0.233	0.298	2.51	2.90
Cluster II	186	165	0.015	0.028	6 (21)**	1 (21)	0.441	0.444	3.99	4.02
Cluster III	77	71	0.028	0.033	0 (15)	0 (15)	0.391	0.397	3.71	3.71

MLGs, multilocus genotypes.
†The nine individuals that could not be unequivocally assigned to one of the three clusters were excluded from these analyses.
‡Lists the number of pairwise combinations of loci that were in significant genotypic disequilibrium at $P < 0.05$ (the number of tested combinations are in parentheses; one locus was monomorphic in each of the Clusters I and III, hence only 15 pairwise comparisons). The asterisks in the GD columns denote significance for a test of overall genotypic disequilibrium based on the *Z*-transform test.
§H_s is the mean gene diversity of seven loci.
¶*A* is mean allelic richness over all seven loci rarefied to the smallest sample size of 77 for w or 71 w/out.
††With (w) and without (w/out) all copies of the MLGs. In the w/out tests, all multicopy MLGs (whether significant or not) were reduced to one unique copy over the whole data set (see main text for reasoning). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

structure in the hierarchical analyses, all multicopy MLGs (whether significant or not) were reduced to one unique copy per host. These results are presented in Table 4.

Both cryptic subdivision (e.g. the three clusters) and the presence of clones should increase F_{ST} among hosts and increase genotypic disequilibrium among loci. In contrast, a large amount of clonal reproduction should decrease within host F_{IS}, (Prugnolle *et al.* 2005a), counteracting the traditional Wahlund effect of cryptic subdivision.

Vilas *et al.* (2003) observed a surprisingly large variance in F_{IS} among loci at the component and infrapopulation levels and among infrapopulations at individual loci. Our microsatellite data showed a similar pattern when we do not distinguish among the three genetic clusters (Table 5; see Discussion). To determine whether the variation in F_{IS} among infrapopulations could be explained by the Wahlund effect caused by the three genetic groups identified by the STRUCTURE analysis, we tested for a correlation between infrapopulation multilocus F_{IS} (not distinguishing cluster membership) and the unbiased Simpson's index of diversity for a finite population (Krebs 1999). Here the diversity index quantifies the extent to which each infrapopulation is a mix of individuals from different clusters. The STRUCTURE-identified clusters were treated as three taxonomic groups for the calculation of Simpson's index for each infrapopulation. A positive correlation would support our hypothesis of the Wahlund effect. For example, infrapopulations that consist mostly of worms from one cluster (low Simpson's index) would be in Hardy–Weinberg equilibrium, while infrapopulations that were a mix of worms from two or three clusters would show large, positive F_{IS}. Results from this analysis are shown in Fig. 3. We also tested whether dissimilarity of cluster composition between infrapopulations explained the variance in pairwise F_{ST} between infrapopulations by plotting pairwise F_{ST} against cluster composition dissimilarity as quantified by the Horn–Morisita index (Krebs 1999). Again, the STRUCTURE-identified clusters were treated as three taxonomic groups for the calculation of pairwise dissimilarity among infrapopulations. A Mantel test (10 000 permutations) as implemented in FSTAT was used to test for a significant correlation (results in Fig. 4).

Vilas *et al.* (2003) also remarked that among allozyme loci, there was a positive correlation between average within infrapopulation F_{IS} and F_{ST} among infrapopulations. Again, the presence of three clusters could explain this observation. Here loci showing the largest allele frequency differences among the three clusters would also show the highest average within-infrapopulation F_{IS} values and the highest among-infrapopulation F_{ST} values. We tested this hypothesis by testing for correlations between, the average F_{IS} within infrapopulations per locus against the F_{ST} among infrapopulations per locus, and each of those two values against F_{ST} for that locus among the three clusters. These results are shown in Fig. 5.

Table 5 F_{IS} among loci calculated at the component population level (i.e. treating all parasites within all hosts as one population) from Table 1 of Vilas *et al.* (2003) (allozymes) and this study (microsatellites). We only compare our data set where multicopy MLGs are not reduced ($N = 498$) to compare with Vilas *et al.* (2003), who could not generate MLGs

Allozyme	F_{IS}	Microsatellite	F_{IS}
Aco	0.306	18B10	−0.065
Ada	0.649	18A5	0.289
Gpi	−0.051	20D1	0.385
Idh	−0.006	14F11	0.778
Pgm-1	0.471	18B8	0.657
Pgm-2	−0.008	18E2	0.161
		14A5A	0.104

MLGs, multilocus genotypes.

Table 4 Within and among infrapopulation hierarchical *F*-statistics with (w) and without (w/out) all copies of MLGs. In the w/out tests, all multicopy MLGs (whether significant or not) were reduced to one unique copy per host (see main text for reasoning). In the F_{ST} columns, standardized F_{ST} (Meirmans 2006) is shown in parentheses. Results are for all worms disregarding cluster membership (All) and for each cluster separately

		Sample size		F_{IT}		F_{ST}		F_{IS}	
	# hosts†	w	w/out	w	w/out	w	w/out	w	w/out
All	12	498	410	0.360***	0.348***	0.182 (0.314)***	0.152 (0.278)***	0.218***	0.231***
Cluster I	11	235	167	0.026	0.039	0.011 (0.015)	−0.003 (0)	0.016	0.042
Cluster II	10	186	170	0.018	0.026	0.016 (0.028)**	0.004 (0.007)	0.002	0.022
Cluster III	10	77	73	0.026	0.025	−0.011 (0)	−0.022 (0)	0.036	0.045

MLGs, multilocus genotypes.
†Representatives of each cluster were not found in all 12 hosts, hence only 10 or 11 hosts per cluster.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

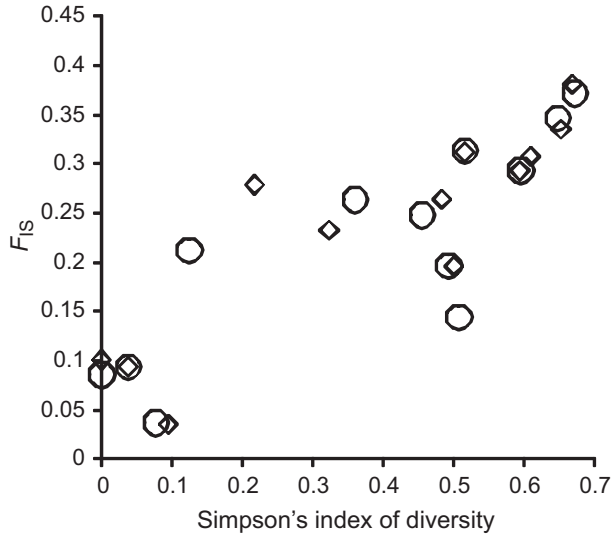

Fig. 3 Plot of multilocus F_{IS} for each infrapopulation vs. Simpson's index of diversity based on the number of individual parasites from each cluster within each infrapopulation. This analysis excluded the nine admixed individuals. Shown are the relationships when all individuals (circles) are included and when the data set is reduced to a single representative of each multilocus genotype within an infrapopulation (diamonds). (r, Pearson's correlation, is 0.84 ($P = 0.0007$) and 0.88 ($P = 0.0002$), respectively). The correlation demonstrates the Wahlund effect (i.e. as an infrapopulation contains a greater diversity of the three parasite clusters, F_{IS} increases).

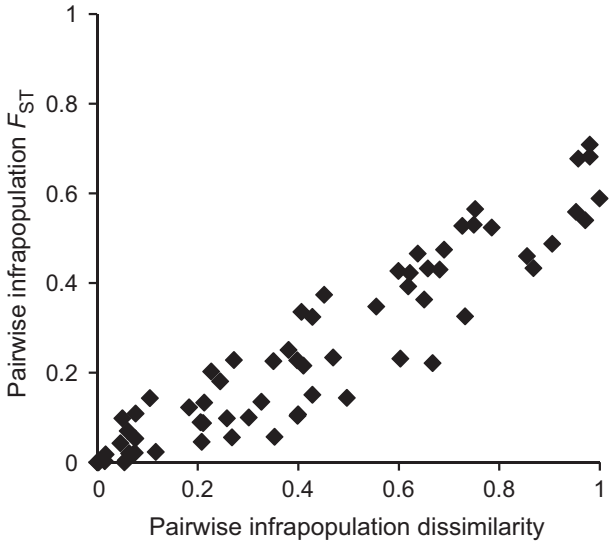

Fig. 4 Plot of pairwise infrapopulation F_{ST} (standardized) vs. a measure of the difference in cluster composition between infrapopulations (Horn-Morisita dissimilarity index). This analysis excluded the nine admixed individuals, but included all copies of each repeated multilocus genotype ($N = 498$). The Mantel test showed a significant correlation ($r = 0.934$; $P < 0.0001$). The same results are obtained with nonstandardized values (data not shown). The correlation demonstrates that the variable pairwise F_{ST} among infrapopulations is driven by the community makeup of the clusters within infrapopulations.

Results

Cryptic structure

In STRUCTURE, the optimal K was determined to be 3. We obtained the same qualitative results if we run STRUCTURE with or without the correlated allele frequencies and admixture models or with the full data set (i.e. $N = 507$) (data not shown). The mean ln $P(D)$ reached its largest values at $K = 3$ (mean = −3427.32) and $K = 4$ (mean = −3426.62), and then declined (Fig. S1, Supporting information). However, at $K = 4$, there was a larger variance in the ln $P(D)$ among the 5 runs (Fig. S1, Supporting information). Furthermore, at $K \geq 4$, the Q-values of individuals in Cluster II become evenly split among each new successive K. Thus, K values ≥ 4 are not plausible. We assigned MLGs to one of the three clusters if their Q-value > 0.7. Overall, there was very strong assignment for most MLGs, as 301 (90.7%) of the MLGs had Q-values ≥90%, 19 (5.7%) ≥80%, and 5 (1.5%) ≥70%. Only seven MLGs had Q-values < 70% (actually nine individuals, as one MLG was present in three copies). These seven MLGs may represent hybrids between the clusters, although the low allelic polymorphism precludes definitive conclusions (Vaha & Primmer 2006). Because we considered these seven MLGs to be uncertain in their cluster assignment, we removed them from subsequent analyses.

We saw excellent agreement between the results of STRUCTURE and the PCoA analysis (Fig. 2); thus the finding of these three clusters is not based solely on the assumptions of the models in STRUCTURE. Based on the first (47.6%) and second axes (22%), the PCoA analysis shows 3, largely nonoverlapping clouds of points that correspond to the STRUCTURE cluster assignments. It is interesting to note that the seven MLGs that had Q-values < 70% fall out at the intersections of the clusters shown in Fig. 2 (six falling between clusters I and III, and one between clusters II and III). This result is congruent with the partitioning of their Q-values among the clusters (data not shown). Table 1 shows the sample sizes per cluster per host after cluster assignment of the full data set ($N = 507$). Allelic richness and gene diversities in the three clusters are given in Table 3. Although a cluster may have had a locus fixed for one allele, no locus showed a fixed difference among the three clusters (Table S3, Supporting information).

Clones and clonal structure

Despite the low power of our loci to distinguish among genetically distinct individuals, we still observed multi-

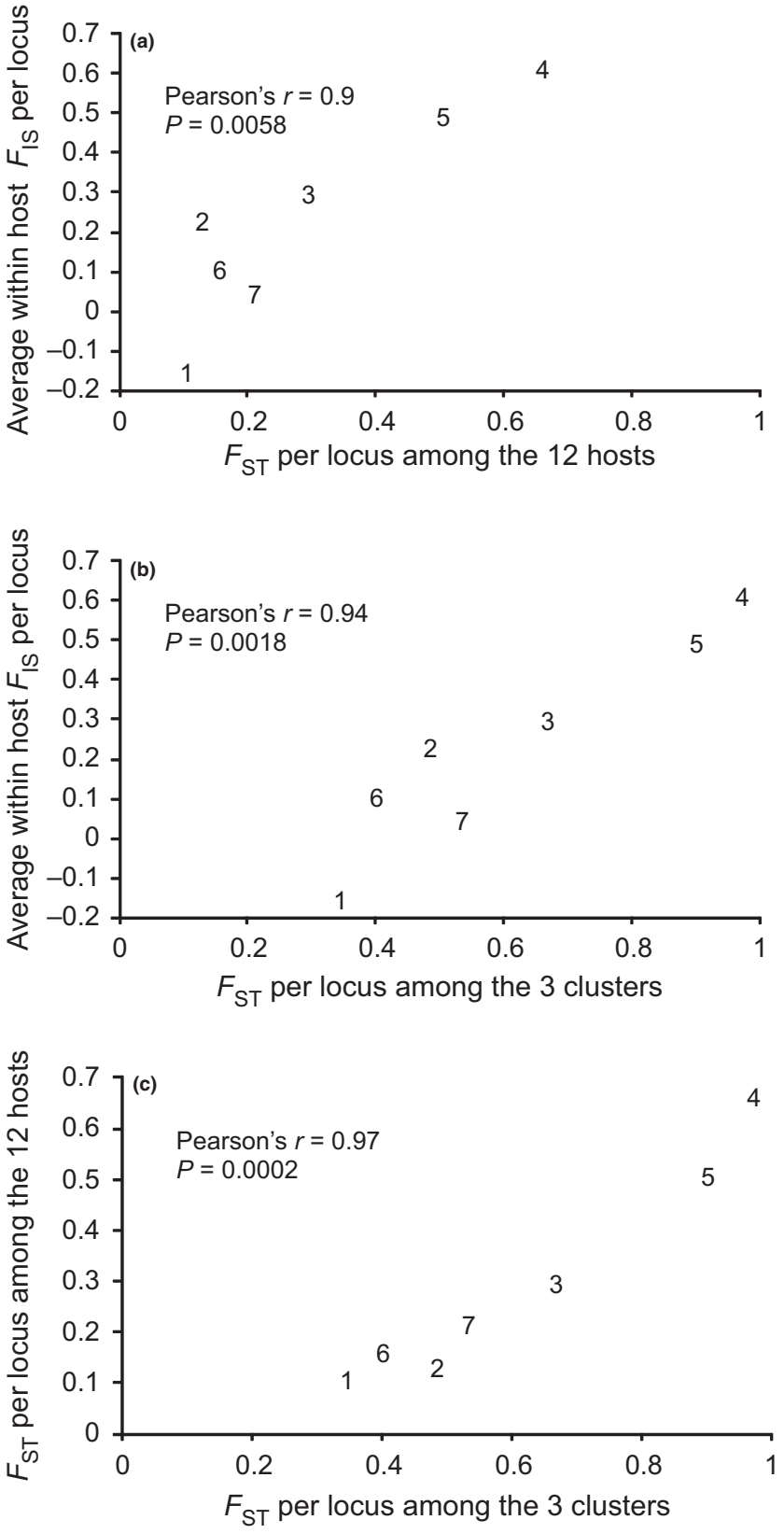

Fig. 5 Demonstration of how cryptic structure can generate an among-locus correlation in average within-infrapopulation F_{IS} and among-infrapopulation F_{ST}. (a) Scatterplot of average within-infrapopulation F_{IS} vs. among-infrapopulation F_{ST} for each locus. F_{IS} and F_{ST} values were generated under a hierarchical *F*-statistics design of infrapopulations within the component population (as in Table 4). The data set included all copies of each repeated multilocus genotype ($N = 498$), but excluded the nine admixed individuals. Vilas *et al.* (2003) noticed the same correlation using allozyme data. (b) Average F_{IS} within infrapopulations plotted against F_{ST} among the three clusters. (c) F_{ST} among infrapopulations plotted against F_{ST} among the three clusters. Correlations in B and C demonstrate the Wahlund effect in that loci having the greatest allele frequency variance among the three clusters also show the largest average F_{IS} within infrapopulations and F_{ST} among infrapopulations (see main text). Shown are standardized values of F_{ST}; the same results are obtained with nonstandardized values (data not shown). The numbers 1–7 in each graph correspond to the seven microsatellite loci and are provided so the reader can view the actual allele frequencies in the three clusters, which are given in Table S3 (Supporting information).

copy MLGs with significant P_{sex} values in all three clusters (Tables 1 and 2). Furthermore, members of multicopy MLGs having significant P_{sex} values also tended to co-occur nonrandomly in the same hosts (Tables 1 and 2). Thus, we believe these individuals are true clones. Table 1 shows the distribution of significant multicopy MLGs (i.e. those judged to be the products of clonal propagation via the P_{sex} tests) among hosts. The complete distribution of all multicopy MLGs (significant or otherwise) among hosts, by cluster, is in Table S1 (Supporting information).

The frequency of clones varied among clusters. In Cluster I, there was weak evidence for clonal structure, with as many unique MLGs being generated via simulation ($n = 88.5$) as were observed in the sample ($n = 89$) (Table 2). Thus, it is likely that members of most of the 42 multicopy MLGs observed in Cluster I represent individuals that are products of sexual reproduction and just happen to be identical at all seven loci by chance. Nevertheless, two of these 42 multicopy MLGs had $P_{sex} < 0.05$ at $n = 2$ (Table 1), and members of these two MLGs co-occurred in the same hosts more often than expected by chance alone (Tables 1 and 2). Two additional multicopy MLGs had $P_{sex} < 0.05$ at $n = 3$ or 4 (Table 1). Because there may be up to three or two individuals that are the product of a sexual event within these two MLGs, we did not test for their cotransmission. Cluster III showed slightly stronger clonal structure, with marginal nonsignificance via simulation ($P = 0.059$; Table 2), and with four of the six multicopy MLGs having $P_{sex} < 0.05$ at $n = 2$. These multicopy MLGs showed significant nonrandom co-occurrence within hosts; Tables 1 and 2. In contrast, Cluster II shows very strong evidence of clonal structure (only 165 MLGs observed vs. more than 180 expected via simulation; $P = 0.0001$). Furthermore, all 11 multicopy MLGs in this cluster have $P_{sex} < 0.05$ at $n = 2$, and again, members co-occur within hosts more often than expected by chance (Tables 1 and 2). Therefore, we see evidence for clones in all three clusters, although the amount of clonal structure varies substantially among them, with Cluster II having relatively many clones and Cluster I few.

Effects of cryptic clusters and clonal reproduction on genetic structure in the sample

The three clusters were well differentiated, with F_{ST} among them = 0.452 (0.684 standardized) or 0.388 (0.61) in the collapsed dataset of 332 individuals. The effect of that cryptic structure on genetic parameters of the entire component population that ignores cluster membership is clearly illustrated in Table 3. There is very high and significant F_{IS} and genotypic disequilibrium in the sample as a whole, but this mostly disappears when each cluster is examined separately. For example, F_{IS} drops from 0.35 in the sample as a whole, to between 0.015 and 0.028 within clusters. Similarly, all 21 pairwise combinations of loci (100%) show significant genotypic disequilibrium ($P < 0.05$) in the whole data set (Table 3). That drops to 7%, 29% and 0% of pairwise combinations being significant when analysed separately for Clusters I, II and III, respectively. Note that the 29% (6 of 21) is in Cluster II, the cluster having the strongest clonality (e.g. all 11 multicopy MLGs have $P_{sex} < 0.05$ at $n = 2$). In the combined probability *Z*-transform test, Cluster II shows overall significant genotypic disequilibrium (Table 3). If we account for clonal structure by using only one copy of each significant MLG, then only one of 21 comparisons, which is not significant overall, in Cluster II remains significant. The lack of genotypic disequilibria without collapsing multicopy MLGs in Clusters I and III further supports that there is little clonal structure in these two clusters.

Table 4 shows that one would come to very different conclusions about structure within and among infrapopulations if one did not account for the cryptic clusters or the clones. F_{ST} among hosts (infrapopulations) is large (0.18; 0.31 standardized) and significant for the data set as a whole, but drops to near zero if one looks at each cluster separately. F_{ST} remains statistically significant in Cluster II until one accounts for clones (collapsing the significant MLGs to one copy per infrapopulations), at which point F_{ST} goes to near zero. The effect of clones is also apparent in that we see that average F_{IS} within hosts increases in Cluster II (though not enough to deviate from Hardy–Weinberg equilibrium) when significant MLGs are collapsed (significant increase in F_{IS} across all loci, $P = 0.016$ Wilcoxon signed-rank test). This is as expected because the presence of clones tends to reduce average F_{IS} within infrapopulations (Prugnolle *et al.* 2005a). Thus, the presence of cryptic clusters and of clones causes apparent structure among infrapopulations that disappears when we account for those two phenomena.

We also show that the highly variable F_{IS} values that Vilas *et al.* (2003) observed among infrapopulations results from the cryptic clusters. There is a strong, positive correlation between multilocus F_{IS} and Simpson's index of cluster diversity in each infrapopulation that holds with or without the inclusion of repeated MLGs (Fig. 3). Here variable Wahlund effects result because infrapopulations differ in their cluster composition, ranging from mostly pure to a mix of individuals from different clusters (cluster compositions reported in Table 1; e.g. host L was pure, carrying individuals only from Cluster I, while host B was mixed, with 10, 11 and

six individuals from Clusters I, II and III, respectively). The difference in cluster composition also drives variable pairwise F_{ST} estimates among infrapopulations such that the greater the cluster-community dissimilarity between infrapopulations, the greater the genetic differentiation (Fig. 4).

Discussion

Cryptic structure

Using samples from the same host population, we found the same pattern of high variation in F_{IS} among loci as Vilas *et al.* (2003) found using allozymes (Table 5; two-sample test of variance $P = 0.97$; *t*-test of mean $P = 0.55$). The similar patterns between microsatellites and allozymes rule out that the allozymes themselves generated artifactual deviations from Hardy–Weinberg equilibrium (e.g. via null alleles or misscoring). Thus, the high variation in F_{IS} among loci appears to represent a real biological phenomenon. As Vilas *et al.* (2003) concluded and as we do here, an inbred mating system is unlikely to produce this pattern as there would not be so much variation in F_{IS} among loci. Vilas *et al.* (2003) also noticed a qualitative correlation where allozyme loci having high average within-infrapopulation F_{IS} also had the high average F_{ST} among infrapopulations. This correlation was mildly nonsignificant (data from Table 3 of Vilas *et al.* 2003; Pearson's $r = 0.75$, $P = 0.08$). We see the same correlation with microsatellite loci, and it is statistically significant (Fig. 5a). Vilas *et al.* (2003) argued this pattern is reflective of the Wahlund effect. The logic being that the stochasticity of genetic drift resulted in loci with varying degrees of allele frequency divergence among the groups that were found to be admixed. Thus, loci that have low divergence among groups will show the least amount of heterozygote deficit in admixed infrapopulations and have low F_{ST} among infrapopulations regardless of the makeup of the infrapopulations (i.e. the frequency of individuals of each cluster within each infrapopulation). In accord with their hypothesis, we observed significant correlations in these latter two relationships (Fig. 5b, c, respectively). This line of reasoning assumes two things: (i) there is random mating within each cluster and (ii) to generate high values of F_{ST} at diverged loci, individuals of the admixed groups need to occur in different frequencies among infrapopulations. Unfortunately, the amount of tissue needed for allozyme typing precluded the ability to obtain MLGs for individual worms. Thus, Vilas *et al.* (2003) could not explore the above assumptions in more detail.

Revisiting this system, but this time obtaining MLGs, we find strong evidence that cryptic population genetic structure is indeed the cause of the overall Wahlund effect within the component population of the presumed single fluke species, *Lecithochirium fusiforme*. Our analyses alone provide clear evidence for the Wahlund effect caused by the presence of three cryptic groups. We see that within each of these clusters, there is random mating once cluster composition is taken into account (discussed below; Tables 3 and 4). It is evident that the frequency of individuals from the three clusters varies among infrapopulations (Table 1) and that their distribution among infrapopulations (Table 1) provides a clear explanation for the variation in multilocus F_{IS} among hosts (Fig. 3) and in pairwise F_{ST} between hosts (Fig. 4).

Clonal structure

The life cycle of digeneans includes an obligate larval asexual phase in the mollusc first intermediate host and an obligate adult sexual phase in the definitive host. Thus, when sampling adults, it is critical to determine whether identical MLGs are the result of sexual or clonal reproduction. The reason is that analyses with and without clones yield different answers to questions about transmission or mating systems (Prugnolle *et al.* 2005a). For example, clonal reproduction in flukes does not reflect reproductive events in the previous parental generation; thus, they should be removed before making F_{IS} estimates that will be used to infer the mating system (Criscione & Blouin 2006). To identify clones, population allele frequencies are needed to calculate P_{sex} values (Arnaud-Haond *et al.* 2007). However, if one is dealing with an admixed sample of individuals (e.g. the Wahlund effect as shown here), then population allele frequencies will not be correct. The Wahlund effect could lead to an overestimation of the number of clones present among repeated MLGs because the expected frequency of any given genotype will be lower, especially if there are unique alleles among groups that are admixed. Thus, we are presented with a 'catch-22' as how does one test for clones if there is cryptic structure, but how does one determine structure in the presence of potential clones, which would violate the assumptions of the methods implemented in STRUCTURE? We chose to reduce to the data set to one representative of each MLG and then run the Bayesian clustering algorithm of STRUCTURE. Although this procedure may remove identical individuals that are the product of a sexual rather than clonal event, the effect on the clustering should be less severe than the effect of using allele frequencies from an admixed group to estimate P_{sex} values.

We detected multiple copies of clones in all three clusters; however, clonal structure varied among the

three clusters. Cluster II had the most pronounced clonal structure with all 11 repeated MLG testing as significant clones. This yields a percentage of truly unique MLGs to genotyped individuals of 89% (165/186). Cluster III had a percentage of 95% (73/77) as all nonsignificant repeated MLGs are considered the product of sexual reproduction and thus are unique clones. Cluster I had a percentage of 99% (233/235) or 97% (228/235 if all four MLGs in Table 1 are considered clones). Comparing our results with those from studies where clone significance was tested across the entire component population, Clusters I and III have values comparable to those in other trematodes having fully aquatic life cycles. For example, in the only other study in definitive hosts, Criscione & Blouin (2006) observed between 99% and 100% of unique MLGs to genotyped individuals in *Plagioporus shawi*. Other studies have been conducted at the second intermediate host level and reported values of 97% for *Diplostomum pseudospathaceum* in sticklebacks (Rauch *et al.* 2005), 97% and 98% for *Maritrema novaezealandensis* in crabs and amphipods, respectively (Keeney *et al.* 2007a,b), and 98% in *Gymnophallus* sp. in cockles (Leung *et al.* 2009).

Transmission and mating systems

When viewed collectively (not taking into account clones or cryptic clusters), the component population of *L. fusiforme* appears to reject the hypothesis that fully aquatic transmission promotes mixing of larval parasites before recruitment into definitive hosts. However, after parsing the sample into its three cryptic genetic clusters, we see that the apparent differentiation among infrapopulations and the heterozygote deficiencies within infrapopulations disappear. True multicopy clones were not abundant enough to affect those parameters in Clusters I and III, but did have a noticeable effect in Cluster II.

It is interesting that the three cryptic clusters differ in their clonal structure. There was evidence for only weak clonal structure in Cluster I based on the randomizations and that very few MLGs had significant P_{sex} values (Table 2). This could be because of lower power (low allelic variation), but the lack of genotypic disequilibrium (only ~7% of loci comparisons; Table 3) when all copies of MLGs are included suggests that the identical MLGs really are not clones, but are products of sexual reproduction. For Cluster I, there is no genetic subdivision among infrapopulations and no deviation from random mating expectations, whether or not you adjust for the few observed clones. Cluster III has slightly more clonal structure, but again not enough to substantially affect among-infrapopulation differentiation or F_{IS} within infrapopulations. Thus, for Clusters I and III, we see 'panmixia' across their component populations (i.e. support for the aquatic mixing hypothesis). In contrast, clonal structure does impact the genetic structure in Cluster II. In particular, when repeated copies of clones are not removed from the analysis, there is more genotypic disequilibria (~29% of loci comparisons; Table 3) and there is significant genetic structure among infrapopulations (Table 4). When clones are removed, there is reduced (and not significant overall) genotypic disequilibria (only ~4% of loci comparisons; Table 3), a slight increase in F_{IS} (but still nonsignificant), and no genetic structure among infrapopulations (Table 4). By removing clones, one can interpret the level of mixing prior to the clonal phase of the parasite, i.e. prior to infecting the mollusc first intermediate host (Prugnolle *et al.* 2005a) and can infer the mating system of the prior adult generation. Thus, in Cluster II we see random mating in the prior adult generation and high mixing of larval parasites prior to mollusc definitive hosts. But, subsequent cotransmission of clones into definitive hosts increased levels of genetic differentiation among infrapopulations. These dynamics are in agreement with the theoretical work by Prugnolle *et al.* (2005a).

In summary, all clusters show mixing of parasites before infecting the mollusc first intermediate host. There is some clonal aggregation from the mollusc to the definitive host (mostly in Cluster II, although all three clusters tested significant for coclonal transmission), but this level of clonal structure does not have a major impact on causing deviations from random mating as evidenced by the nonsignificant F_{IS} values. Therefore, these data support the hypothesis that parasites that cycle in fully aquatic habitats will have a panmictic component population structure. See Criscione & Blouin (2006) for a discussion that compares population genetic patterns among aquatic and terrestrial flatworm parasite systems.

What are the three cryptic clusters?

Why do we find three morphologically identical and apparently reproductively isolated groups in apparent sympatry? It is worth mentioning a few potential and not necessarily mutually exclusive, hypotheses as they provide future avenues of study. (1) The three clusters could obviously represent true cryptic species. Undeniably, the identification of digenean cryptic species is becoming more common as more studies utilize molecular markers (Criscione & Blouin 2004; Miura *et al.* 2005; Locke *et al.* 2010; Pérez-Ponce de León & Nadler 2010; Razo-Mendivil *et al.* 2010). For example, Criscione & Blouin (2004) found evidence for cryptic species of hemiurid digeneans infecting the same individual fish hosts, just as in this study. However, if these are cryptic

species in our '*L. fusiforme*' sample, then they must be very recently diverged because there are no fixed allelic differences among the clusters (Table S3), the loci amplified equally well in all three clusters, and the F_{ST} values between the three clusters were on the scale of what has been observed among highly geographically structured conspecifics [nonstandardized F_{ST} between I vs. II = 0.492 (0.733 standardized), I vs. III = 0.449 (0.627), and II vs. III = 0.347 (0.601) with repeated MLGs, and I vs. II = 0.435 (0.707), I vs. III = 0.338 (0.515), and II vs. III = 0.343 (0.598) using only unique MLGs]. For example, these values are in the range of those reported for geographically separated, conspecific populations of a digenean parasite of salmon in Oregon (*Plagioporus shawi;* Criscione & Blouin 2007) and for conspecific populations of vertebrates such as amphibians (e.g. Blouin *et al.* 2010). If they are good species, it is interesting to speculate on what keeps them genetically distinct in sympatry (after all, we did identify apparent hybrids). Perhaps each uses a unique intermediate host and disruptive selection maintains mating barriers.

It is worth noting that Bartoli & Gibson (2007) recently placed *L. fusiforme* as a junior synonym of *L. grandiporum* based on morphological comparisons of specimens from *C. conger* and the moray eel, *Muraena helena*, off Corsica in the Western Mediterranean. Yet we found three reproductively isolated groups in a single bay. Given the extensive morphological variation of many characters given by Bartoli & Gibson (2007), additional molecular data from specimens of different definitive host species and geographical locations should be obtained to help elucidate the potential for cryptic species and phenotypic plasticity.

The next two hypotheses invoke a single species, but argue for either temporal or spatial separation in transmission. (2) A mixture of flukes that are the product of different temporal breeding events could cause the Wahlund effect we observed. In this case, temporally isolated breeding units produce offspring cohorts with different allele frequencies. Different temporal cohorts may end up in the same infrapopulation because the parasites can live for extended periods of time in the fish, third intermediate hosts. A similar phenomenon is observed in Pacific salmon when individuals of different brood years unite in each year's spawning population, thereby creating a transient Wahlund effect that would disappear in the progeny (Waples 1990). We find it unlikely that just three well-defined clusters with such high levels of genetic differentiation would appear under these dynamics, but it is worth mentioning the possibility. (3) A third hypothesis is that there is one species but three geographically separate transmission cycles. Under this hypothesis, the hosts we sampled represent sinks. Plenty of 'hybrid' offspring are produced in those hosts, but these hybrids never transmit to the next generation. This could occur, for example, if three genetically differentiated source populations send offspring into the bay (say via the fish, third intermediate hosts), and the life cycle cannot be completed within the bay, perhaps owing to the absence of a critical intermediate host. One could even imagine a scenario in which this source-sink phenomenon occurs on a much smaller scale within the bay (e.g. if successful transmission depends on a particular habitat that may be patchy and if the eels are sedentary and territorial). These ideas are purely speculative, but given how frequently molecular ecologists stumble upon cryptic 'species' in molecular surveys of helminths, it is worth considering what features of parasite life cycles might predispose them to so much apparent differentiation in sympatry.

Concluding remarks

Although population genetic studies are increasing among parasitic taxa, there is still a paucity of such data given the taxonomic and life-history variation found among organisms with a parasitic life style. Our data serve to illustrate one of many future surprises that we expect from population genetic studies in parasites. In general, we suspect that cryptic structure could be commonplace among organisms with limited morphology or complex life histories. The methods we employ can be applied to a diversity of organisms to tease apart apparent deviations from equilibrium. As seen here, there is more than meets the eye in that cryptic structure can be found on a small geographical scale and be present within a given individual host, but that the causes may not be obvious. A naive analysis of this data set would have concluded that there is strong structure among infrapopulations sampled on a small geographical scale—a clear rejection of aquatic mixing hypothesis. But in fact, that structure is driven mostly by the cryptic clusters and to a lesser extent by the presence of clones. At the level of sexually reproducing parasites within clusters, there is random recruitment among hosts. Thus, this data set supports the hypothesis that parasites with complete aquatic transmission will have panmictic structure among hosts within a given host population. In addition, our data provide some of the first empirical support (see also Prugnolle *et al.* 2005b) for the theoretical expectations of how clonal variance impacts trematode population genetic structure (Prugnolle *et al.* 2005a).

Acknowledgements

RV was supported by a PD/MEC fellowship from the European Union and the Spanish Ministry of Education and Sci-

ence. We thank the OSU Center for Genome Research and Biocomputing for the genotyping.

References

Arnaud-Haond S, Belkhir K (2007) GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Molecular Ecology Notes*, **7**, 15–17.

Arnaud-Haond S, Duarte CM, Alberto F, Serrao EA (2007) Standardizing methods to address clonality in population studies. *Molecular Ecology*, **16**, 5115–5139.

Bartoli P, Gibson DI (2007) The status of *Lecithochirium grandiporum* (Rudolphi, 1819) (Digenea: Hemiuridae), a rarely reported and poorly known species from the Mediterranean moray eel *Muraena helena* L. in the Western Mediterranean. *Systematic Parasitology*, **68**, 183–194.

Belkhir K, Borsa P, Chikhi L, Rausfast N, Bonhomme F (2004) *GENETIX 4.05, Windows TM Software for Population Genetics*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université Montpellier II, Montpellier, France.

Blouin MS, Phillipsen IC, Monsen KJ (2010) Population structure and conservation genetics of the Oregon spotted frog, *Rana pretiosa*. *Conservation Genetics*, **11**, 2179–2194.

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.

Bush AO, Fernandez JC, Esch GW, Seed JR (2001) *Parasitism: The Diversity and Ecology of Animal Parasites*. Cambridge University Press, Cambridge.

Cornell SJ, Isham VS, Smith G, Grenfell BT (2003) Spatial parasite transmission, drug resistance, and the spread of rare genes. *Proceedings of the National Academy of Sciences*, **100**, 7401–7405.

Criscione CD, Blouin MS (2004) Life cycles shape parasite evolution: comparative population genetics of salmon trematodes. *Evolution*, **58**, 198–202.

Criscione CD, Blouin MS (2005) Effective sizes of macroparasite populations: a conceptual model. *Trends in Parasitology*, **21**, 212–217.

Criscione CD, Blouin MS (2006) Minimal selfing, few clones, and no among-host genetic structure in a hermaphroditic parasite with asexual larval propagation. *Evolution*, **60**, 553–562.

Criscione CD, Blouin MS (2007) Parasite phylogeographical congruence with salmon host evolutionarily significant units: implications for salmon conservation. *Molecular Ecology*, **16**, 993–1005.

Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular Ecology*, **14**, 2247–2257.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Gibson DI, Bray RA (1986) The Hemiuridae (Digenea) of fishes form the north-east Atlantic. *Bulletin of the British Museum (Natural History), Zoology Series*, **51**, 1–125.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Gregorius HR (2005) Testing for clonal propagation. *Heredity*, **94**, 173–179.

Jarne P, Theron A (2001) Genetic structure in natural populations of flukes and snails: a practical approach and review. *Parasitology*, **123**, S27–S40.

Jombart T (2008) ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Keeney DB, Waters JM, Poulin R (2007a) Clonal diversity of the marine trematode *Maritrema novaezealandensis* within intermediate hosts: the molecular ecology of parasite life cycles. *Molecular Ecology*, **16**, 431–439.

Keeney DB, Waters JM, Poulin R (2007b) Diversity of trematode genetic clones within amphipods and the timing of same-clone infections. *International Journal for Parasitology*, **37**, 351–357.

Krebs CJ (1999) *Ecological Methodology*. Benjamin/Cummings, Menlo Park, California.

Leung TLF, Poulin R, Keeney DB (2009) Accumulation of diverse parasite genotypes within the bivalve second intermediate host of the digenean *Gymnophallus* sp. *International Journal for Parasitology*, **39**, 327–331.

Locke SA, McLaughlin JD, Dayanandan S, Marcogliese DJ (2010) Diversity and specificity in *Diplostomum* spp. metacercariae in freshwater fishes revealed by cytochrome c oxidase I and internal transcribed spacer sequences. *International Journal for Parasitology*, **40**, 333–343.

Lydeard C, Mulvey M, Aho JM, Kennedy PK (1989) Genetic variability among natural populations of the liver fluke in *Fascioloides magna* in white-tailed deer, *Odocoileus virginianus*. *Canadian Journal of Zoology*, **67**, 2021–2025.

Matthews BF, Matthews RA (1988) The ecsoma in hemiuridae (Digenea, Hemiuroidea): tegumental structure and function in the mesocercaria and the metacercaria of *Lecithochirium furcolabiatum* (Jones, 1933) Dawes, 1947. *Journal of Helminthology*, **62**, 317–330.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution*, **60**, 2399–2402.

Miura O, Kuris AM, Torchin ME, Hechinger RF, Dunham EJ, Chiba S (2005) Molecular-genetic analyses reveal cryptic species of trematodes in the intertidal gastropod, *Batillaria cumingi* (Crosse). *International Journal for Parasitology*, **35**, 793–801.

Nadler SA (1995) Microevolution and the genetic structure of parasite populations. *Journal of Parasitology*, **81**, 395–403.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pérez-Ponce de León G, Nadler SA (2010) What we don't recognize can hurt us: a plea for awareness about cryptic species. *Journal of Parasitology*, **96**, 453–464.

Prugnolle F, Liu H, de Meeus T, Balloux F (2005a) Population genetics of complex life-cycle parasites: an illustration with trematodes. *International Journal of Parasitology*, **35**, 255–263.

Prugnolle F, Roze D, Theron A, de Meeus T (2005b) *F*-statistics under alternation of sexual and asexual reproduction: a model and data from schistosomes (platyhelminth parasites). *Molecular Ecology*, **14**, 1355–1365.

Rauch G, Kalbe M, Reusch TBH (2005) How a complex life cycle can improve a parasite's sex life. *Journal of Evolutionary Biology*, **18**, 1069–1075.

Razo-Mendivil U, Vázquez-Dominguez E, Rosas-Valdez R, de León GPP, Nadler SA (2010) Phylogenetic analysis of nuclear and mitochondrial DNA reveals a complex of cryptic species in *Crassicutis cichlasomae* (Digenea: Apocreadiidae), a parasite of Middle-American cichlids. *International Journal for Parasitology*, **40**, 471–486.

Schwab AE, Churcher TS, Schwab AJ, Basanez M-G, Prichard RK (2006) Population genetics of concurrent selection with albendazole and ivermectin or diethylcarbamazine on the possible spread of albendazole resistance in *Wuchereria bancrofti*. *Parasitology*, **133**, 589–601.

Sire C, Durand P, Pointier J-P, Theron A (2001) Genetic diversity of *Schistosoma mansoni* within and among individual hosts (*Rattus rattus*): infrapopulation differentiation at microspatial scale. *International Journal of Parasitology*, **31**, 1609–1616.

Vaha JP, Primmer CR (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*, **15**, 63–72.

Vilas R, Paniagua E (2004) Seasonal occurrence and diversity of hemiuridae (Trematoda : Digenea) in conger eels (*Conger conger*) from Spain. *Comparative Parasitology*, **71**, 233–237.

Vilas R, Paniagua E, Sanmartín ML (2003) Genetic variation within and among infrapopulations of the marine digenetic trematode *Lecithochirium fusiforme*. *Parasitology*, **126**, 465–472.

Vilas R, Criscione CD, Paniagua E, Blouin MS (2006) Microsatellite loci from the trematode *Lecithochirium fusiforme*, a parasite of the European conger eel. *Molecular Ecology Notes*, **6**, 1064–1066.

Waples RS (1990) Conservation genetics of Pacific salmon. II. Effective population size and the rate of loss of genetic variability. *Journal of Heredity*, **81**, 267–276.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Whitlock MC (2005) Combining probability from independent tests: the weighted *Z*-method is superior to Fisher's approach. *Journal of Evolutionary Biology*, **18**, 1368–1373.

C.C.'s laboratory focuses on the ecology and evolution of parasites. In particular, he uses a molecular ecology/population genetics approach to elucidate transmission dynamics, mating systems, and natural history of parasites. R.V.'s research interests are in the conservation genetics of economically important fishes and the molecular ecology of parasites. E.P.'s laboratory examines the population biology and molecular ecology of parasites. M.B.'s laboratory focuses on the causes and consequences of genetic structuring, and on applications of methods for pedigree reconstruction in natural populations. He works on a variety of taxa including fish, amphibians and parasites.

Data Accessibility

Microsatellite genotype data deposited at Dryad: doi:10.5061/dryad.8959.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Complete distribution of each multicopy MLG (whether significant or not) among the three clusters and putative hybrids.

Table S2 Summary of the number of worms sampled, genotyped, and analyzed from each host.

Table S3 Allele frequencies for the seven microsatellite loci among the three clusters.

Fig. S1 Determination of the optimal *K* using STRUCTURE.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.