THE ROYAL SOCIETY

Aggregation and distribution of strains in microparasites

C. C. Lord[1,2*], B. Barnard[3], K. Day[1], J. W. Hargrove[4], J. J. McNamara[5], R. E. L. Paul[1]†, K. Trenholme[6]‡ and M. E. J. Woolhouse[1,7]

[1]*Wellcome Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, Oxford UK OX1 3PS, UK*
[2]*Florida Medical Entomology Laboratory, University of Florida, 200 9th Street SE, Vero Beach, FL 32962, USA*
[3]*Onderstepoort Veterinary Institute, Onderstepoort 0110, Republic of South Africa*
[4]*ODA Insect Pest Management Initiative, Box CY52, Causeway, Harare, Zimbabwe*
[5]*MRC Trypanosomiasis Research Group, Department of Clinical Veterinary Science, University of Bristol, Langford, Bristol BS18 7DU, UK*
[6]*Walter & Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia*
[7]*Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK*

Recent research has shown that many parasite populations are made up of a number of epidemiologically distinct strains or genotypes. The implications of strain structure or genetic diversity for parasite population dynamics are still uncertain, partly because there is no coherent framework for the interpretation of field data. Here, we present an analysis of four published data sets for vector-borne microparasite infections where strains or genotypes have been distinguished: serotypes of African horse sickness (AHS) in zebra; types of *Nannomonas* trypanosomes in tsetse flies; parasite-induced erythrocyte surface antigen (PIESA) based isolates of *Plasmodium falciparum* malaria in humans, and the merozoite surface protein 2 gene (*MSP-2*) alleles of *P. falciparum* in humans and in anopheline mosquitoes. For each data set we consider the distribution of strains or types among hosts and any pairwise associations between strains or types. Where host age data are available we also compare age–prevalence relationships and estimates of the force of infection. Multiple infections of hosts are common and for most data sets infections have an aggregated distribution among hosts with a tendency towards positive associations between certain strains or types. These patterns could result from interactions (facilitation) between strains or types, or they could reflect patterns of contact between hosts and vectors. We use a mathematical model to illustrate the impact of host–vector contact patterns, finding that even if contact is random there may still be significant aggregation in parasite distributions. This effect is enhanced if there is non-random contact or other heterogeneities between hosts, vectors or parasites. In practice, different strains or types also have different forces of infection. We anticipate that aggregated distributions and positive associations between microparasite strains or types will be extremely common.

Keywords: strain structure; microparasite; basic reproduction number; African horse sickness; *Plasmodium falciparum*; *Trypanosoma*

1. INTRODUCTION

The mechanisms generating aggregation in macroparasites and the implications for control programmes have been of considerable interest since Crofton's seminal paper (Crofton 1971; May 1977; Anderson & May 1985, 1991; Bundy 1988; Guyatt *et al*. 1990). Aggregation in microparasites, however, has not been considered in detail. Multiple strains have been found in a number of microparasites, and this number is increasing rapidly due to recent advances in molecular biology. The distribution of strains among hosts in a microparasite is analogous to the distribution of individuals in a macroparasite, and potentially impacts epidemiology in similar ways. This paper is intended to address some of the questions about aggregation in microparasites and consider possible mechanisms which may generate observed distributions of strains between hosts.

While there is some debate about precisely how to define a strain within a parasite species (Gupta & Day 1994; Hastings & Wedgwood-Oppenheim 1997), for the purposes of this paper we define a strain as a genetically distinct and identifiable subpopulation of a parasite or pathogen which has distinct epidemiological, immunological or pathological characteristics. This type of structure is well-known in viruses, which frequently have multiple serotypes identified by the immunological response of a vertebrate host (e.g. foot and mouth: Domingo *et al*. 1990; African horse sickness: Coetzer & Erasmus 1994; bluetongue: Verwoerd *et al*. 1994; dengue: Kettle 1990). A similar structure has also been found in protozoa, such as malaria (Forsyth *et al*.

*Author and address for correspondence.
†Present address: Laboratoire Biochemie et Biologie Moleculaire des Insectes, Institut Pasteur Paris, France.
‡Present address: Menzies School of Health Research, Casuarina, NT 0081, Australia.

1989; Gupta *et al*. 1994*a*,*b*; Gupta & Day 1994; but see Babiker *et al*. 1995) and trypanosomiasis (Woolhouse *et al*. 1996). Many bacteria also have multiple strains, some of which have been stable for considerable lengths of time (Maynard Smith *et al*. 1991; Gupta *et al*. 1996).

The implications of multiple strains for control strategies are not yet fully understood, in part because there is no coherent framework for the analysis of field data. As has been suggested for human malaria (Gupta *et al*. 1994*b*), considering strains to be transmitted independently does dramatically lower estimates of the basic reproduction number (R_0) and consequent estimates of the intensity of control efforts. Alternatively, strains may interact (e.g. cross-immunity, competition or facilitation), thus affecting epidemiological patterns. Negative associations between strains (each strain observed alone more often than expected) could be generated by cross-immunity, where hosts previously exposed to one strain are wholly or partially immune to other strains. Competition between strains could lead to the complete exclusion of some strains (Gupta *et al*. 1996), resulting in different groups of strains in host populations. Facilitation between strains would result in positive associations between the strains; strains would be found more often in combinations than alone.

An aggregation of strains between hosts will also have an impact on control programmes, regardless of the mechanisms generating the aggregation. The degree of aggregation will affect whether targeted vaccination or intervention programmes are more efficient than mass treatment (Woolhouse *et al*. 1997). Aggregation may also affect the prevalence of a parasite in either the host or vector population, with consequent effects on control programmes.

The importance of these types of interactions in practice is not fully understood and one aim of this paper is to consider how much can be inferred from observed epidemiological patterns. Continued application of new techniques seems likely to continue the rapid increase in the number of data sets describing the epidemiology of parasite and pathogen strains. These data will provide an opportunity to explore the epidemiological significance of strain variation, and this paper is intended to provide some preliminary discussion of the analysis and interpretation of such data.

Here, we analyse four previously published data sets on the distribution of parasite and pathogen strains or types among hosts or vectors of different ages: PIESA isolates of *Plasmodium falciparum* malaria in humans; types of *Nannomonas* trypanosomes in tsetse flies; serotypes of African horse sickness virus in zebra; and *MSP-2* alleles of *P. falciparum* in humans and mosquitoes. We compare these data to a mathematical model which allows us to explore the relationship between various epidemiological processes and the resulting distribution of strains among hosts.

2. METHODS

(a) *Data sets*

(i) *PIESA isolates of malaria*

Sera collected in Madang, Papua New Guinea were tested using agglutination assays for exposure to five *P. falciparum* isolates (Gupta *et al*. 1994*b*). The isolates differ in parasite-induced erythrocyte surface antigens (PIESAs). Four of the five isolates tested (1935, 1776, 1934, and 1917) were collected in PNG in 1987 (Forsyth *et al*. 1989), while the fifth (HB3) was derived from a Honduran isolate (Walliker *et al*. 1987). Data from individuals $<$ two years old were not included in this analysis, as there were too few samples in that age range. Individuals were grouped into the following age classes: 2–4, 5–9, 10–14, 15–20, 20–30 and $>$ 30 years. The midpoint of each interval was used for the force-of-infection analysis (λ_i, see below). Sample sizes in each age class are given in table 1.

(ii) *Trypanosomiasis*

Woolhouse *et al*. (1996) collected tsetse flies (*Glossina pallidipes* and *G. morsitans morsitans*) in the Zambezi Valley of Zimbabwe during June 1993. Flies were dissected and aged by ovarian category on a scale of 0–7. Mouthpart dissection was used to determine if flies were infected with trypanosomes of the subgenus *Nannomonas*, and midguts of mouthpart-positive flies were screened with PCR for parasite identification. DNA probes specific for three types of *Trypanozoon congolense* (savannah, riverine-forest and kilifi), *T. simiae* and *T. godfreyi* were used.

For the calculation of the force of infection (λ_i, see below), the full age range was used for analysis; ovarian categories were assigned ages based on ranges given by Challier (1965; category 1, 12.5 days (d); category 2, 21.5 d, category 3, 30.5 d; category 4, 39.5 d; category 5, 48.5 d; category 6, 57.5 d; category 7, 66.5 d). Because of low numbers of infected flies, flies were grouped into categories 0–3, 4–5 and 6–7 for the rest of the analysis.

(iii) *African horse sickness*

Serum was collected from free-living zebra foals in the Kruger National Park and analysed as described by Barnard (1993). Foals were captured, age estimated based on tooth eruption and wear, and serum collected. Animals were assigned to monthly age classes except for those aged seven and eight months, which could not be distinguished and so were combined into one class. Microneutralization assays were used to determine antibody titres specific to each of the nine serotypes. Because cross-reactions between serotypes may influence the statistical tests used, the analyses were carried out with known cross-reacting serotypes combined. If an animal was positive for either of the two members of a cross-reacting pair, it was considered positive for the combination (e.g. if positive for either serotype 1 or 2, it was positive for the combination 1–2). This resulted in five combinations of serotypes: 1–2, 3–7, 4, 5–8 and 6–9.

To prevent maternal antibodies from confounding the analyses, all foals less than five months of age and four foals aged five months which tested positive to eight or more serotypes were excluded from the analysis (Lord *et al*. 1997). Final sample sizes for each age class are given in table 1.

(iv) MSP-2 *alleles of malaria*

Paul *et al*. (1995) collected *Anopheles* mosquitoes and blood samples from humans in several villages in Papua New Guinea. Extracted parasite DNA from oocysts and sera were amplified for the genes for three proteins. Here, we restrict our analysis to villages in the coastal region, alleles of the gene for merozoite surface protein 2, and *An. koliensis* mosquitoes. Although 16 *MSP-2* alleles were recognized based on family and size, only the two families (IC and FC27) were used in this analysis, as the data were insufficient for analysis with 16 types. These data were analysed only for aggregation and associations, as there were no age data for mosquitoes. Humans were grouped into age classes of 0–4, 5–9, 10–14, 15–19 and $>$ 20 years; sample

Table 1. *Association, pairwise and λ statistics for three of the data sets used*

age class	sample size[a]	mean ± s.d.[b]	V/M[c]	Schluter's W[d]	negative PV_{ij} (no. calculated)	significant pairwise associations[e]	λ
PIESA isolates of malaria							
all	74	2.65 ± 1.77	1.18	< 0.001	0(10)	1776 & HB3, 1935 HB3 & 1934 1934 & 1917, 1935 1917 & 1935	1776, HB3 > 1917
2–4[f]	16	0.25 ± 0.58	1.33	n.s.	2(3)	n.d.	—
5–9	12	2.50 ± 1.0	0.40	n.s.	2(6)	n.d.	—
10–14	9	2.00 ± 1.22	0.75	n.s.	4(10)	n.d.	—
15–19	9	3.33 ± 1.32	0.53	n.s.	5(10)	n.d.	—
20–29	15	3.73 ± 1.16	0.36	n.s.	4(10)	n.d.	—
30 +	13	4.46 ± 0.78	0.14	n.s.	2(3)	n.d.	—
trypanosomiasis							
all	1146	0.08 ± 0.43	2.20*	< 0.001	1(10)	sav & riv, ken[g] riv & ken, godf godf & sim	sav > sim, godf kil > sim
0–3[h]	455	0.03 ± 0.25	2.31*	< 0.001	3(10)	sav & riv	—
4–5	466	0.10 ± 0.47	2.22*	< 0.001	2(10)	sav & riv, kil	—
6–7	225	0.16 ± 58	2.07*	< 0.001	3(10)	sav & riv, kil riv & kil	—
AHS							
all	107	2.79 ± 1.85	1.22	< 0.001	0(10)	all possible except 3–8 & 4	1–2, 3–7, 5–8, 6–9 > 4
5[i]	11	0.91 ± 1.04	1.20	n.s.	5(10)	n.d.	—
6	36	1.64 ± 1.40	1.19	< 0.005	1(10)	none	—
7–8	30	3.03 ± 1.71	0.97	< 0.005	0(10)	4 & 5–8 5–8 & 6–9	—
9	19	4.53 ± 0.51	0.06	n.s.	n.d.	n.d.	—
12	11	1.82 ± 0.40	0.03	n.s.	n.d.	n.d.	—

[a] Total number of samples in age class.
[b] Number of strains per host or vector.
[c] Variance to mean ratio for number of strains per host. *, significantly > 1 at $p < 0.0005$ (experiment-wise, $\alpha < 0.01$).
[d] p-value for test; n.s., not significant.
[e] Pairs significant at $p < 0.05$; all significant associations are positive; n.d., not done.
[f] Years.
[g] Type of *Nannomonas* trypanosomes: sim, *T. simiae*; godf, *T. godfreyi*; *T. congolense* types: sav, savannah; riv, riverine-forest; kil, Kilifi.
[h] Ovarian categories.
[i] Months.

Table 2. *Associations between alleles for the* MSP-2 *locus in malaria*

age class	sample size[a]	prevalence (%)[b]	mean ± s.d.[c]	V/M[d]	Schluter's W[e]
human					
all	210	41	0.67 ± 0.86	1.11	< 0.001
0–4	38	45	0.71 ± 0.87	1.05	n.s.
5–9	37	68	1.08 ± 0.86	0.69	n.s.
10–14	27	44	0.70 ± 0.87	1.07	n.s.
15–19	17	47	0.82 ± 0.95	1.10	n.s.
20 +	91	26	0.44 ± 0.78	1.38	< 0.001
mosquito					
all	6731	0.3	0.0002 ± 0.01	0.16	< 0.001

[a] Total number of samples in age class.
[b] Prevalence of infection with either allele family (as determined by PCR probes).
[c] Number of strains per host or vector.
[d] Variance to mean ratio for number of strains per host. *, significantly > 1 at $p < 0.0005$ ($\alpha < 0.01$).
[e] p-value for test; n.s., not significant.

sizes for each class are given in table 2. We include this data set primarily for the comparison between vectors and hosts, as none of the other studies examined both in the same location.

(b) *Data analysis*

(i) *Aggregation*

The variance to mean ratio in the number of parasites per host can be used as an indication of aggregation (Pielou 1977; Anderson & May 1991). Here, we use this ratio based on the number of strains per host as a measure of aggregation of strains. Values greater than one indicate aggregation; the distribution of strains among hosts is not random. Frequently, age data are not available; it is of interest to determine if apparently aggregated distributions are the result of summing distributions with different parameters (e.g. each age class) (Grafen & Woolhouse 1993). We examine this by calculating the variance to mean ratio for each age class and with all age classes summed for each data set.

The variance to mean ratio may not be appropriate if prevalences vary between strains. If the distribution of parasite strains among hosts is independent of all other strains, there should be no covariance between the strains. Schluter's variance test (Schluter 1984; McCulloch 1985) compares the observed variance in the number of strains per host with that expected under the hypothesis of no covariance, and is equivalent to the summed binomial (Lotz & Font 1991) and Cochran's Q (McCulloch 1985). Schluter's W can be interpreted as comparing the observed distribution of the number of strains per host with that expected from the sum of independent binomial distributions and is calculated as

$$W = \frac{\sum_j (T - t_j)^2}{\sum_i p_i(1 - p_i)}, \tag{1}$$

where T is the mean number of strains per host, t_j the number of strains present in host j (j is the number of hosts in the class or system), and p_i the prevalence of strain i (with i strains in the system overall). $W \approx \chi^2_{aj}$; a significant result indicates covariance between the strains and non-independent distributions. The variance to mean ratio and Schluter's W are calculated for each data set, with age classes individually and combined, and the results compared.

(ii) *Associations*

When covariance between strains exists, pairwise associations are tested using Fisher's exact test (two-tailed; FREQ procedure (SAS 1988)). The significance level was adjusted for multiple tests using the Bonferroni correction (Rice 1989).

The direction of pairwise associations (positive: strains are found together; negative: strains are not found together) was examined using the index PV_{ij}.

$$PV_{ij} = \frac{ab - cd}{\sqrt{(a+d)(a+b)(a+c)(c+d)}}. \tag{2}$$

Here, a is the number of hosts uninfected with either strain, b is the number infected with i but not j, c is the number infected with j but not i, and d is the number infected with both strains (i and j). This index ranges between -1 (complete negative association) and $+1$ (complete positive association), and is undefined if all hosts are uninfected or infected with either strain (Pielou 1977). Therefore, we report the number of indices which are negative out of the number which could be calculated. PV_{ij} was calculated for each age class and for all age classes combined for three of the data sets. It is not informative for *MSP-2* alleles of malaria, as with two strains there is only one possible pairwise comparison.

(iii) *Force of infection*

For an endemic parasite, the force of infection can be calculated using age-prevalence data (Anderson & May 1991). We assume that hosts (whether insects or vertebrates) acquire infection with parasites of strain i with an age-independent constant force of infection λ_i. The proportion infected with strain i at age a, $p_i(a)$, is, therefore,

$$p_i(a) = 1 - e^{-\lambda_i(a-M)}. \tag{3}$$

For vertebrate hosts, M is the age at which maternal antibodies become undetectable. For trypanosomes in tsetse flies, M can be interpreted as the latent period, during which an infection is undetectable and not transmissible. Using maximum-likelihood methods (Williams & Dye 1994), λ_i can then be estimated as the value which maximizes the log likelihood, L,

$$L = \sum_a -x_{ia}\lambda_i(a - M) + y_{ia}\ln(1 - e^{-\lambda_i(a-M)}), \tag{4}$$

where x_{ia} is the number negative for parasite strain i at age a and y_{ia} the number positive at age a. For the PIESA malaria and AHS data, M was chosen to maximize the sum (over the five parasite strains) of the log likelihood. For trypanosomiasis, independent estimates of the latent period in tsetse flies (Woolhouse *et al*. 1993, 1994) were averaged, giving an estimate of 10.75 days. The variance of λ can be estimated as

$$\mathrm{var}(\lambda_i) = -\left[\sum_a -y_{ia}(a - M)^2\left[\frac{e^{-\lambda_i(a-M)}}{1 - e^{-\lambda_i(a-M)}} + \frac{e^{-2\lambda_i(a-M)}}{(1 - e^{-\lambda_i(a-M)})^2}\right]\right]^{-1}. \tag{5}$$

The goodness-of-fit of λ_i for each parasite strain can be tested by likelihood ratios, comparing the maximum log likelihood (ML) with the likelihood from the fully saturated model (SM), calculated using observed values; $2 \times (\mathrm{SM} - \mathrm{ML}) \approx \chi^2_{\alpha,n-3}$ where n is the number of age classes. For each system, a common λ was fitted using the data for all five strains, and tested by likelihood ratios (as above) to the fit obtained with individual values of λ_i for each strain. A significant lack of fit in this test indicates that the values of λ_I are different. 95% confidence intervals were calculated for pairwise comparisons (mean $\pm\, z_{0.025} \times$ s.d.); strains with non-overlapping confidence intervals are significantly different at the 5% level. For comparison, λ based on exposure to any strain was also calculated for each system; these are designated as 'any'.

(c) *Mathematical model*

We developed a stochastic simulation model appropriate for a vector-borne microparasite infection. The model is an individual-based, age-structured expansion of the Ross–Macdonald model (Ross 1911; Macdonald 1957), with the parasite consisting of five strains. In the homogeneous model, vectors contact hosts randomly and acquire or transmit infections with a probability of 0.1 (arbitrarily set for demonstration). In the heterogeneous model, the transmission probabilities are the same, but biting probabilities vary between hosts as a log uniform distribution with a standardized variance (SV = 1.079), and are constant with host age. This is a discrete-time model with a time-step of one week. The life span of the hosts and vectors are 100 and five weeks, respectively. There are no latent periods; infectious hosts are infectious for up to 20 weeks and

infected vectors for life. Populations are maintained at 2000 vectors and 100 hosts by adding the appropriate number of individuals to the youngest age class. Each model is run 100 times for 1000 time-steps, which performance curves showed, to give stable estimates of the mean and variance of model outputs. The output of the models is the proportion of hosts and vectors ever infected and the number of strains in each age class, averaged over model runs. In addition, Schluter's W is calculated for each age class in each run. The proportion of runs with a significant W (non-independence in the distribution of strains) is also reported. Variance to mean ratios were not calculated, as we report the average results from the 100 runs. Turbo Pascal was used for all programming; a copy of the program can be obtained from the authors.

3. RESULTS

(a) *Prevalence*

Figure 1 shows the prevalence of strains for the PIESA isolates of malaria, types of trypanosomes, and serotypes of AHS, broken down by host age. For PIESA isolates of malaria and serotypes of AHS, the prevalence of all strains increased with host age to very high levels. For trypanosomes, the prevalence of most types increased with age; however, the maximum prevalence was low. The distribution of the number of strains infecting each host shifted right with age in these three systems (table 1, figure 2); however, the lower prevalence of trypanosomes in tsetse flies was apparent. In *MSP-2* alleles in malaria, the highest average number of strains per host occurs in the 5–9 year age class, in part due to a higher overall prevalence in that age group (table 2).

(b) *Aggregation*

Variance-to-mean ratios generally decreased with age (table 1), indicating that the distribution of strains per host became less aggregated. This is largely due to the overall increase in infection with age, and is least apparent in the tsetse flies where prevalences remain low even in the highest age class. However, in *MSP-2* alleles in malaria, the distribution was least aggregated in the 5–9 year age class and the ratio tended to increase with age (table 2).

(i) *PIESA isolates of malaria*

The variance-to-mean ratio was not significantly greater than one in any age class or in the data combined. Schluter's variance test was significant with all age classes combined, indicating that the distributions of the five PIESA isolates were different from that expected under the hypothesis of independence (table 1). However, it was not significant for any of the age classes separately.

(ii) *Trypanosomiasis*

The variance-to-mean ratio was significantly greater than one and Schluter's variance test was significant for all three age classes and age classes combined (table 1).

(iii) *African horse sickness*

Much of the analysis has already been reported (Lord *et al*. 1997); the full details can be found there. Table 1 summarizes the results. The variance-to-mean ratio was not significantly greater than one in any age class or the age classes combined. Schluter's variance test was significant in the data overall and in two age classes separately.

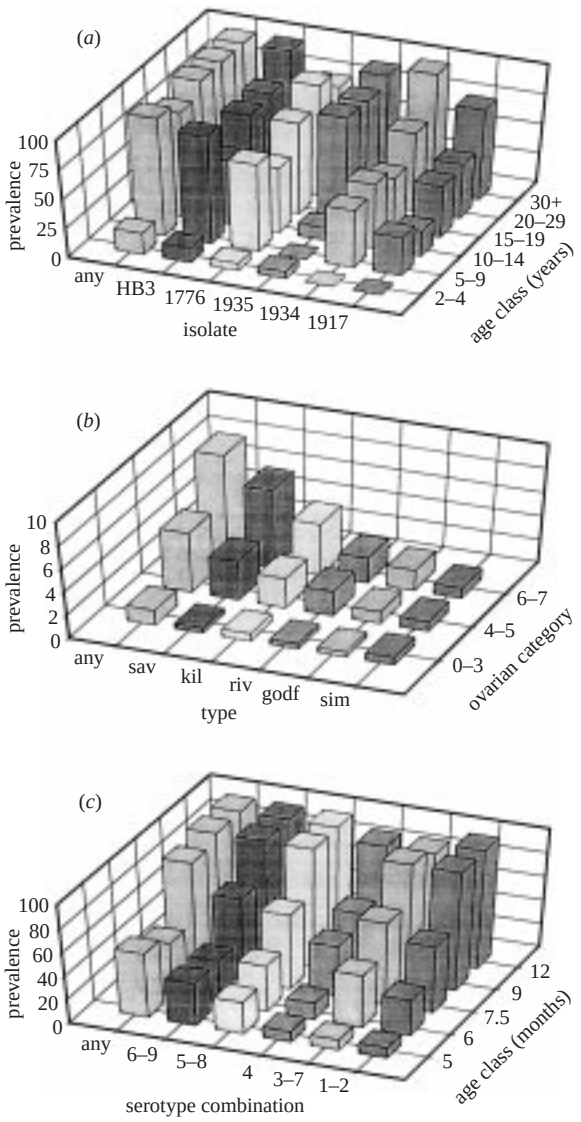

Figure 1. Prevalence by age for each of the three parasites. (*a*) *P. falciparum* malaria in humans. (*b*) *Nannomonas* trypanosomes in *Glossina* flies; note changed axis scaling. (*c*) AHS serotypes in zebra. 'Any' is the prevalence of individuals testing positive for exposure to any of the five strains. Sample sizes for each age class are given in table 1.

(iv) MSP-2 *alleles of malaria*

Schluter's variance test was significant in both humans and mosquitoes when all ages were combined (table 2). Only the oldest human age class showed significant non-independence when the age classes were considered separately. The variance-to-mean ratio was not significantly greater than one in for any age class or in age classes combined, for either humans or mosquitoes.

(c) *Associations*

(i) *PIESA isolates of malaria*

Several significant pairwise associations were observed in the data overall (table 1). However, in the age classes separately Schluter's variance test was not significant; therefore, pairwise statistics were not calculated. The index of association (PV_{ij}) was uniformly positive in the data overall and was often positive in the age classes separately (table 1).

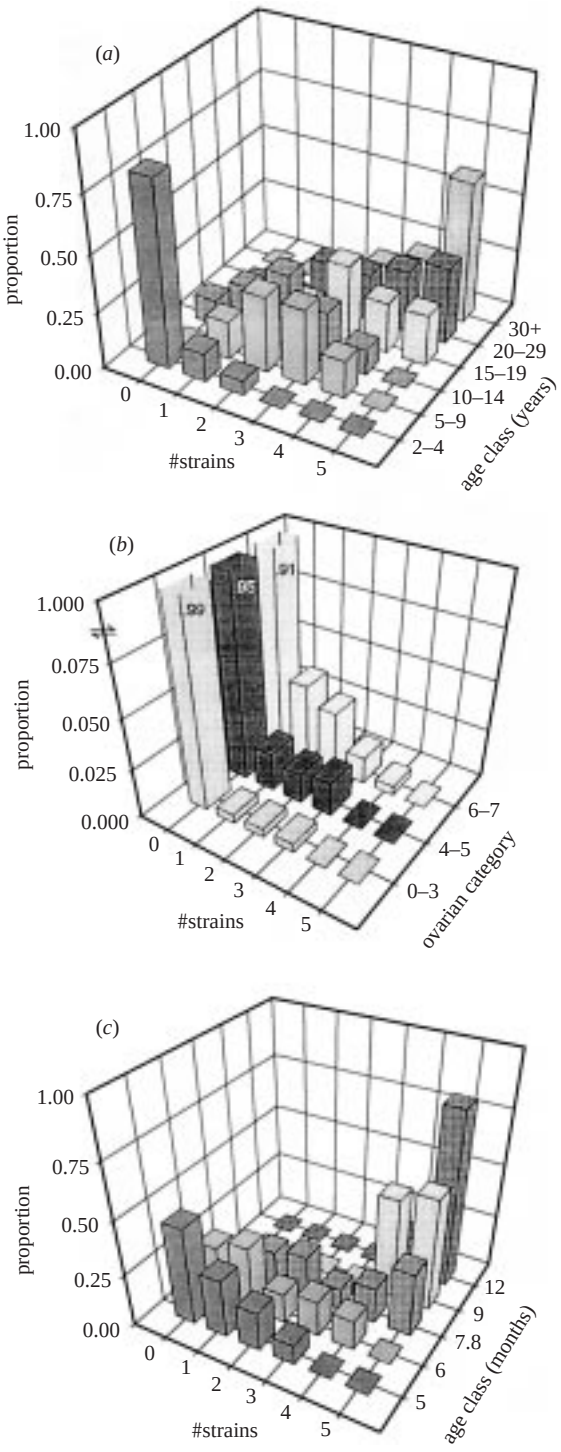

Figure 2. Distributions of the number of strains per host by age for each of the three parasites. (*a*) PIESA isolates of *P. falciparum* malaria in humans. (*b*) *Nannomonas* trypanosomes in *Glossina* flies; note changed axis scaling. (*c*) AHS serotypes in zebra.

(ii) *Trypanosomiasis*

Significant pairwise associations were found in all three age classes and age classes combined (table 1). PV_{ij} was positive in all significant comparisons and in most non-significant ones; however, there were a few negative indices in each age class.

(iii) *African horse sickness*

In the age classes combined, nine out of the ten possible pairwise associations were significant. However, when studying the age classes separately, pairwise associations were detected in only one class, and there were only two significant associations (table 1). PV_{ij} was negative in only one comparison; all significant associations and most of the rest were positive.

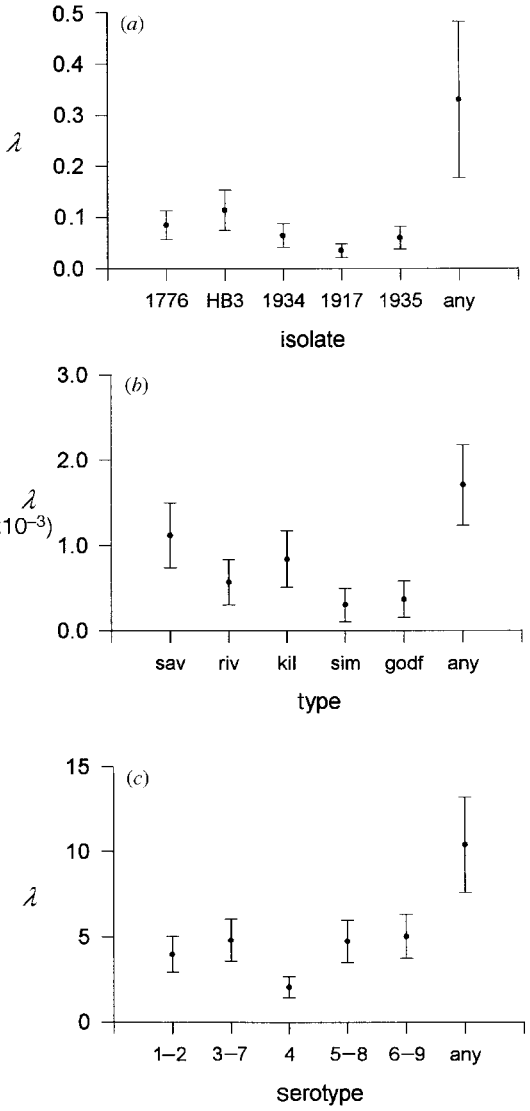

Figure 3. Range of λ values. Best fit value and 95% confidence intervals. Values are significantly different at $\alpha = 0.05$ if the confidence intervals do not overlap. (*a*) PIESA isolates of *P. falciparum* malaria in humans. (*b*) *Nannomonas* trypanosomes in *Glossina* flies. (*c*) AHS serotypes in zebra, 'any' as defined for figure 1.

(iv) MSP-2 *alleles in malaria*

For both humans and mosquitoes, the pairwise association between the two allele families was significantly positive when all ages were combined. Only the oldest human age class showed non-independence and was considered separately; the pairwise association between families IC and FC27 was significantly positive (table 2).

(d) *Force of infection*

In the three systems tested, the force of infection varied between strains (figure 3, table 1; $p < 0.005$ for all systems). The best fit was obtained with $M = 1.83$ years for malaria in humans and 4.15 months for AHS in zebra.

Table 3. *Results of model simulations*

	homogeneous		heterogeneous	
age class	mean ± s.d.[a]	proportion significant[b]	mean ± s.d.[a]	proportion significant[b]
host				
all	3.74 ± 1.47	0.00	3.00 ± 1.86	0.38
0–19	1.59 ± 1.25	0.19	1.47 ± 1.62	0.94
20–39	3.39 ± 0.97	0.01	2.67 ± 1.79	0.97
40–59	4.21 ± 0.74	0.01	3.27 ± 1.72	0.98
60–79	4.66 ± 0.51	0.02	3.70 ± 1.56	1.00
80–99	4.86 ± 0.36	0.02	3.89 ± 1.45	0.91
vector				
all	0.12 ± 0.36	0.61	0.15 ± 0.42	0.98
0	0.00 ± 0.00	0.00	0.00 ± 0.00	0.00
1	0.06 ± 0.24	0.09	0.08 ± 0.30	0.51
2	0.12 ± 0.35	0.18	0.16 ± 0.42	0.51
3	0.17 ± 0.41	0.08	0.23 ± 0.50	0.50
4	0.23 ± 0.48	0.15	0.28 ± 0.56	0.37

[a] Number of strains per host or vector.
[b] Proportion of runs with Schluter's W significant at $p < 0.01$.

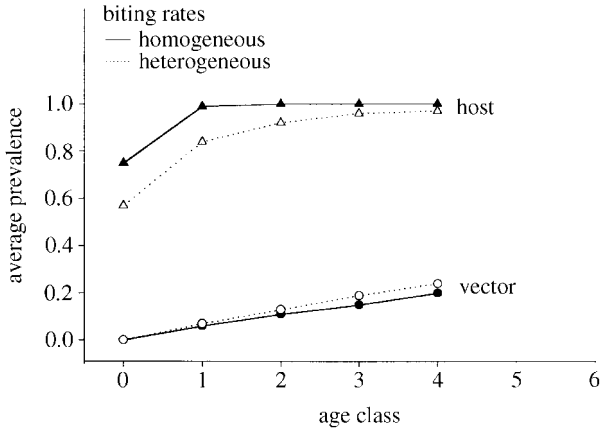

Figure 4. Equilibrium prevalence in models. Average of 100 runs for both host and vector. Age classes as labelled for vector; for hosts: 0, 0–19 weeks; 1, 20–39; 2, 40–59; 3, 60–79; 4, 80–99.

Although these values differ from those previously suggested for the decline in maternal antibody (four to six months for malaria in Nigeria (Achidi *et al*. 1995) and 4.8 months (Lord *et al*. 1997) for AHS), changing the value of M affected only the absolute values of λ and not the relative rankings or significant differences (data not shown). Pairwise comparisons showed malaria isolates 1776 and HB3 had a significantly higher λ than isolate 1917. In tsetse flies, λ for *T. congolense* savannah and kilifi types were significantly higher than that for *T. simiae*; *T. congolense* savannah type also had a significantly higher λ than *T. godfreyi*. AHS serotype 4 had a significantly lower force of infection than any other type. The force of infection calculated based on exposure to any strain was significantly higher than λ for any individual strain in malaria and AHS; for trypanosomes λ_{any} any was significantly higher than all but one of the individual types.

(e) *Mathematical model*

The results of the model are given in table 3 and figure 4. It is clear that non-independent distributions of strains can be generated without any heterogeneity in the model, but this is greatly increased by the addition of heterogeneity in biting rates. Schluter's W was significant in a large proportion of the runs for most age classes in the heterogeneous case and in more runs than would be expected by chance in the homogeneous case. The prevalence of the parasite overall is much higher in the hosts than in the vectors, reaching a plateau at a relatively young age in hosts but not in the vectors. Adding heterogeneity decreases the prevalence in the hosts but increases prevalence in the vector (figure 4). In the vector, combining the data across age classes accentuates the non-independent distribution of strains because this involves summing distributions with different means; this effect was not observed for the hosts.

4. DISCUSSION

There are a number of common features to the epidemiological patterns shown by the parasites included in this study. Multiple infections are common in both hosts and vectors. Taken over the whole host population the distribution of strains or types tends to be aggregated. This aggregation is poorly characterized by the variance to mean ratio of the number of strains per host because different strains typically have different prevalences. Schluter's test is a preferable method of demonstrating non-random distributions for such data.

Interpretation of these patterns and their implications depends first on the nature of the assay used to determine the presence of an infection. In this study, the data for trypanosomes in tsetse and *MSP-2* malaria alleles in mosquitoes and humans are based on PCR amplification of parasite DNA and are therefore measures of current infection status. On the other hand, the data for AHS in zebra and PIESA malaria isolates in humans are based on the presence of antibodies and are therefore measures of cumulative infections (depending on the duration of the antibody response once an infection is cleared). Measures of current infection will generally show a lower number of

strains per host, and may more accurately reflect the potential for biological interactions between strains.

However, in either case, aggregated distributions over the whole host population may be artefactual due to variation in prevalences between host age classes. Summing distributions with different means inevitably produces aggregation (Grafen & Woolhouse 1993). This point is illustrated by the output of a simple mathematical model of age-structured but otherwise homogeneous populations; the model tends to produce aggregated distributions in vectors where prevalences are low and increase over the whole age range, but not in hosts where prevalences rapidly reach high levels and are similar across a wide age range. All our data sets show heterogeneities in strain prevalences with age. Thus, although aggregation over the whole host population is interesting in terms of its consequences (e.g. increased opportunities for outcrossing), it is not informative about the mechanisms that generate it.

Aggregation also occurs within age classes (although, for the reasons just given, caution is required when age classes are wide and prevalences are changing). This aggregation was most marked for trypanosomes in tsetse and apparent in some age classes for AHS in zebra, but did not occur for malaria in humans (except in the oldest, and widest, age class in Papua New Guinea). The mathematical model suggests that aggregation within age classes will frequently be apparent in vectors, but may not be apparent in hosts if the component populations are homogeneous. Conversely, heterogeneity in biting rates per host may lead to highly aggregated parasite distributions, especially in the host.

Heterogeneity in biting rate is one type of host heterogeneity. Equally, other forms of heterogeneity in the host or the vector population will also lead to aggregation—a well-studied example would be genetic heterogeneities in susceptibility to infection. It is not possible to infer which of many possible heterogeneities is operating from parasite distribution data alone. It should also be mentioned that the detection of aggregation is dependent on a number of factors, especially sample size (small sample sizes give low statistical power) and parasite prevalence (very low or very high prevalences also give low statistical power). This may explain why the aggregation of AHS serotypes could not be detected in the youngest or oldest hosts (with low and high prevalences, respectively). Care should also be taken to ensure strains can be clearly separated, as artefactual positive associations can result from incomplete specificity.

A more detailed examination of parasite distributions is provided by testing for pairwise associations. Clearly, aggregation in itself is likely to result in some significant pairwise associations and a number were revealed in this study. Our data sets were all horizontal studies of single populations. The significance of pairwise associations will, however, be determined by studies of different populations or by the same population at a number of points in time. If pairwise associations are purely a consequence of aggregation then no particular association is expected to appear consistently (with the proviso that differences in prevalences will affect statistical power to detect associations). Persistent or widespread associations would therefore be evidence for interactions between or common host requirements of certain parasite strains.

Our data also show heterogeneities in the force of infection of different parasite strains for AHS, trypanosomes and human malaria. This type of heterogeneity is especially important in the context of the basic reproduction number, R_0. As shown by Gupta & Day (1994), if strains or types can be considered epidemiologically independent then estimates of R_0 based on the entire parasite population will overestimate the R_0 values for individual strains. The data analysed here suggest that this effect could reduce R_0 estimates by at least a factor of two (figure 3). Clearly, this would have a significant impact on the expected impact of disease control programmes (Anderson & May 1991).

Parasite strains may vary in other ways, including pathology (malaria: Gupta *et al*. 1994*a*; AHS: Laegreid *et al*. 1993; *Toxoplasma*: Sibley & Boothroyd 1992). Pathology may also be affected by the distribution of parasites; for example, dengue haemorrhagic fever may result from infection with more than one dengue virus serotype (PAHO 1994). This type of heterogeneity may have significant and complex effects on the impacts and design of disease control programmes. However, the relationships between markers used for strain identification and pathology remain poorly understood (Hastings & Wedgewood-Oppenheim 1997).

Indeed the genetic basis of strain variation in general introduces a number of complications. Strains may be defined by one (e.g. AHS: Grubman & Lewis 1992) or several (e.g. influenza: Gupta & Day 1994) loci. In either case, other loci may influence transmissibility or strain interactions and so will affect epidemiological patterns. Both recombination and the number of strain-determining loci will affect strain structure (Hastings & Wedgewood-Oppenheim 1997) in sexual organisms such as malaria and also in viruses such as AHS which have segmented genomes. Gupta *et al*. (1996) suggested that competition (mediated by host immune responses) can structure parasite populations into distinct strains despite recombination. However, some studies of malaria show considerable gene flow and no evidence of strain or spatial structure (Babiker *et al*. 1995). Interaction with the genetic structure of the host immune response may also affect strain structure, as has been suggested for malaria (Gilbert *et al*. 1998). Further information on the markers used to distinguish strains and the genetics of characteristics such as transmissibility and immunogenicity will be required before the genetics of strain determination can be fully understood.

In practice, heterogeneity in host–vector contact (and no doubt in many other aspects of the host–parasite system) is the rule rather than the exception (Woolhouse *et al*. 1997) and this implies that there is an expectation of aggregated distributions of vector-borne microparasites among hosts. This may swamp the effects of any positive interactions between strains or types (but these may be detected in different ways, as described above) and may override the effects of any negative interactions. Crofton's (1971) paper led to a large number of attempts to explain aggregated distributions of macroparasites, and that experience cautions that equally careful analyses will be needed to identify the factors contributing to the aggregated distributions of microparasites among hosts.

We thank Sunetra Gupta and Dan Haydon for discussions of strain variation. C.C.L. was supported in part by a BBSRC Link

grant and M.E.J.W. by the Royal Society. Florida Agricultural Experiment Station journal series no. R-06227.

REFERENCES

Achidi, E. A., Perlmann, H., Salimonu, L. S., Perlmann, P., Walker, O. & Asuzu, M. C. 1995 A longitudinal study of seroreactivities to *Plasmodium falciparum* antigens in Nigerian infants during their first year of life. *Acta Trop.* **59**, 173–183.

Anderson, R. M. & May, R. M. 1985 Herd immunity to helminth infection and implications for parasite control. *Nature* **315**, 493–496.

Anderson, R. M. & May, R. M. 1991 *Infectious diseases of humans: dynamics and control*. Oxford University Press.

Babiker, H. A., Charlwood, J. D., Smith, T. & Walliker, D. 1995 Gene flow and cross-mating in *Plasmodium falciparum* in households in a Tanzanian village. *Parasitology* **111**, 433–442.

Barnard, B. J. H. 1993 Circulation of African horsesickness virus in zebra (*Equus burchelli*) in the Kruger National Park, South Africa, as measured by the prevalence of type specific antibodies. *Onderstepoort J. Vet. Res.* **60**, 111–117.

Bundy, D. A. P. 1988 Population ecology of intestinal helminth infections in human communities. *Phil. Trans. R. Soc. Lond.* B **321**, 405–420.

Challier, A. 1965 Amélioration de la méthode de détermination de l'âge physiologique des glossines. Etudes faites sur *Glossina palpalis palpalis* Vanderplank, 1949. *Bull. Soc. Pathol. Exotique* **58**, 250–259.

Coetzer, J. A. W. & Erasmus, B. J. 1994 African horse sickness. In *Infectious diseases of livestock with special reference to southern Africa* (ed. J. A. W. Coetzer, G. R. Thomson & R. C. Tustin), pp. 460–475. Oxford University Press.

Crofton, H. D. 1971 A quantitative approach to parasitism. *Parasitology* **62**, 343–364.

Domingo, E., Mateu, M. G., Martinez, M. A., Dopazo, J., Moya, A. & Sobrino, F. 1990 Genetic variability and antigenic diversity of foot-and-mouth disease virus. In *Applied virology research* (ed. E. Kurstak, R. G. Marusyk, F. A. Murphy & M. H. V. van Regenmortel), pp. 233–266. New York: Plenum Press.

Forsyth, K. P., Philip, G., Smith, T., Kum, E., Southwell, B. & Brown, G. V. 1989 Diversity of antigens expressed on the surface of erythrocytes infected with mature *Plasmodium falciparum* parasites in Papua New Guinea. *Am. J. Trop. Med. Hyg.* **41**, 259–265.

Gilbert, S. C., Plebanski, M., Gupta, S., Morris, J., Cox, M., Aidoo, M., Kwiatkowski, D., Greenwood, B. M., Whittle, H. C. & Hill, A. V. S. 1998 Association of malaria parasite population structure, HLA and immunological antagonism. *Science* **279**, 1173–1177.

Grafen, A. & Woolhouse, M. E. J. 1993 Does the negative binomial distribution add up? *Parasitol. Today* **9**, 475–477.

Grubman, M. J. & Lewis, S. A. 1992 Identification and characterization of the structural and non-structural proteins of African horsesickness and determination of the genome coding assignments. *Virology* **186**, 444–451.

Gupta, S. & Day, K. P. 1994 A strain theory of malaria transmission. *Parasitol. Today* **10**, 476–481.

Gupta, S., Hill, A. V., Kwiaitowski, D., Greenwood, A. M., Greenwood, B. M. & Day, K. P. 1994*a* Parasite virulence and disease patterns in *Plasmodium falciparum* malaria. *Proc. Natl Acad. Sci. USA* **91**, 3715–3719.

Gupta, S., Trenholme, K., Anderson, R. M. & Day, K. P. 1994*b* Antigenic diversity and the transmission dynamics of *Plasmodium falciparum*. *Science* **263**, 961–963.

Gupta, S., Maiden, M. C. J., Feavers, I. M., Nee, S., May, R. M. & Anderson, R. M. 1996 The maintenance of strain structure in populations of recombining infectious agents. *Nature Med.* **2**, 437–442.

Guyatt, H. L., Bundy, D. A. P., Medley, G. F. & Grenfell, B. T. 1990 The relationship between the frequency distribution of *Ascaris lumbricoides* and the prevalence and intensity of infection in human communities. *Parasitology* **101**, 139–143.

Hastings, I. M. & Wedgwood-Oppenheim, B. 1997 Sex, strains and virulence. *Parasitol. Today* **10**, 375–383.

Kettle, D. S. 1990 *Medical and veterinary entomology*. Wallingford, UK: CAB International.

Laegreid, W. W., Skowronek, A., Stone-Marschat, M. & Burrage, T. 1993 Characterization of virulence variants of African horse sickness virus. *Virology* **195**, 836–839.

Lord, C. C., Woolhouse, M. E. J. & Barnard, B. J. H. 1997 Transmission and distribution of virus serotypes: African horse sickness in zebra. *Epidemiol. Infect.* **118**, 43–50.

Lotz, J. M. & Font, W. F. 1991 The role of positive and negative interspecific associations in the organization of communities of intestinal helminths of bats. *Parasitology* **103**, 127–138.

McCulloch, C. E. 1985 Variance tests for species associations. *Ecology* **65**, 1676–1681.

Macdonald, G. 1957 *The epidemiology and control of malaria*. Oxford University Press.

May, R. M. 1977 Togetherness among schistosomes: its effects on the dynamics of the infection. *Math. Biosci.* **35**, 301–343.

Maynard Smith, J., Dowson, C. G. & Spratt, B. G. 1991 Localized sex in bacteria. *Nature* **349**, 29–31.

PAHO 1994 *Dengue and dengue hemorrhagic fever in the Americas: guidelines for prevention and control*. Washington, DC: Pan American Health Organization.

Paul, R. E. L., Packer, M. J., Walmsley, M., Lagog, M., Ranford-Cartwright, L. C., Paru, R. & Day, K. P. 1995 Mating patterns in malaria parasite populations of Papua New Guinea. *Science* **269**, 1709–1711.

Pielou, E. C. 1977 *Mathematical ecology*. New York: Wiley.

Rice, W. R. 1989 Analyzing tables of statistical tests. *Evolution* **43**, 223–225.

Ross, R. 1911 *The prevention of malaria*. London: Murray.

SAS 1988 *SAS user's guide*. Cary, NC: SAS Institute.

Schluter, D. 1984 A variance test for detecting species associations, with some example applications. *Ecology* **65**, 998–1005.

Sibley, L. D. & Boothroyd, J. C. 1992 Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature* **359**, 82–85.

Verwoerd, D. W. & Erasmus, B. J. 1994 Bluetongue. In *Infectious diseases of livestock with special reference to southern Africa* (ed. J. A. W. Coetzer, G. R. Thomson & R. C. Tustin), pp. 443–459. Oxford University Press.

Walliker, D., Quakyl, L. A., Wellems, T. E., McCutchan, T. F., Szarfman, A., London, W. T. & Corcoran, L. M. 1987 Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**, 1661–1666.

Williams, B. G. & Dye, C. 1994 Maximum likelihood for parasitologists. *Parasitol. Today* **10**, 489–493.

Woolhouse, M. E. J., Hargrove, J. W. & McNamara, J. J. 1993 Epidemiology of trypanosome infections of the tsetse fly *Glossina pallidipes* in the Zambezi Valley. *Parasitology* **106**, 479–485.

Woolhouse, M. E. J., Bealby, K., McNamara, J. J. & Silvtongwe, J. 1994 Trypanosome infections of the tsetse fly *Glossina pallidipes* in the Luangwa valley, Zambia. *Int. J. Parasitol.* **24**, 987–993.

Woolhouse, M. E. J., McNamara, J. J., Hargrove, J. W. & Bealby, K. A. 1996 Distribution and abundance of trypanosome (subgenus *Nannomonas*) infections of the tsetse fly *Glossina pallidipes* in southern Africa. *Mol. Ecol.* **5**, 11–18.

Woolhouse, M. E. J. (and 12 others) 1997 Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl Acad. Sci. USA* **94**, 338–342.