

Katharina Dittmar de la Cruz · Michael F. Whiting

## Genetic and phylogeographic structure of populations of *Pulex simulans* (Siphonaptera) in Peru inferred from two genes (*CytB* and *Coll*)

Received: 12 March 2003 / Accepted: 1 April 2003 / Published online: 23 July 2003  
© Springer-Verlag 2003

**Abstract** In this paper we discuss the potential usefulness of determining the phylogeographic and phylogenetic patterns of a vector for understanding the spread of pathogens or insecticide resistance. We do so using the example of *Pulex simulans* in Peru. Six populations from six different localities were investigated. Mitochondrial DNA sequences were obtained and branching patterns were inferred using phylogenetic reconstruction methods and nested clade analyses. Ten different haplotypes were discovered. Phylogenetic analysis revealed *P. simulans* in Peru as a monophyletic group, containing clades that were generally not geographically correlated. The data suggest that *P. simulans* is not a single genetic entity but rather that this species shows a high degree of intraspecific variation. Restricted gene flow with long distance dispersal coupled with range expansion and long distance colonization are likely to have contributed to the observed patterns of variation.

### Introduction

An important factor in the association between a pathogen and its vector, as well as in the development and spread of insecticide resistance, is the population genetic structure of the vector. Genetic structure is largely controlled by effective population size, subdivision into multiple demes, and the rates of gene flow among these. Knowledge of the genetic structure not only gives insight into the rate and geographical scale over which both

adaptation and random differentiation in the particular vector populations occur, but also enables the researcher to accurately characterize and discriminate the vector in question and specifically associate a population or certain haplotype of vectors to a pathogen. Additionally, this knowledge is useful for controlling the development of insecticide resistance. High migration rates for instance, help to spread resistant haplotypes from the population of origin. Furthermore, changing effective population sizes might cause the fixation of certain favorable alleles in a population.

The goal of this study is to decipher the phylogeographic structure of a vector using the example of *Pulex simulans* and to demonstrate how this information might usefully complement field studies on the nature of pathogen-vector interaction and drug resistance.

*P. simulans* is a flea with a cosmopolitan distribution that infests a variety of animal hosts, including guinea pigs. In Peru, *P. simulans* readily infests human habitations because of the close association between the flea's host (in this case guinea pigs) and humans. Flea infestations can affect both human and animal health. Apart from the immediate discomfort inflicted through the immune response to the blood feeding of the flea, this species is also a possible vector for a variety of pathogenic protozoans and bacteria, including *Yersinia pestis* (plague), *Rickettsia* spp. (*Rickettsia typhi*, *Rickettsia rickettsi*, ELB agent (named after EL Laboratory in Soquel, Calif.), and *Trypanosoma* spp. (Acha and Szyfres 1994; Didier and La Scala 2001). Because this species is a major disease vector, flea populations are controlled through insecticides to halt the transmission of pathogens. Intensified treatment efforts usually correlate with the development of resistance to the chemicals used (Campos and Andrade 2002; Ziam and Guillet 2002).

### Materials and methods

A total of 127 fleas were obtained from guinea pig hosts from six locations throughout Peru (Table 1, Fig. 1). The localities sampled represent the known distribution of the species on domesticated

K. D. de la Cruz  
Institute of Parasitology, Veterinary Faculty,  
University of Leipzig, An den Tierkliniken 33,  
04103 Leipzig, Germany

K. D. de la Cruz (✉) · M. F. Whiting  
Department of Integrative Biology,  
Brigham Young University,  
401 WIDB, Provo, UT 84602, USA  
E-mail: kd98@email.byu.edu  
Fax: +1-801-4220090

**Table 1** Summary by state of the localities from which samples of *Pulex simulans* and outgroup taxa were obtained

State	Locality	No.	Coordinates
Lima	Lima/Cieneguilla	1	12°5'S 76°55'W
Huancayo	Huancayo	2	12°15'S 75°7'W
Ayacucho	Ayacucho	3	13°1'S 74°2'W
Moquegua	Moquegua	4	17°7'S 71°3'W
Ancash	Caraz	5	8°56'S 7742'W
Lambayeque	Salas	6	6°8'S 79°8'W

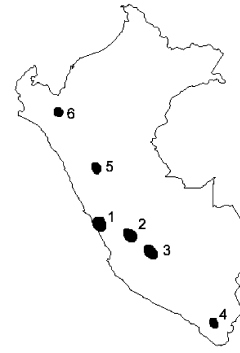
guinea pigs (Dittmar 2001). The species was identified as *P. simulans* based on morphological characters. The population size of *P. simulans* on all hosts was consistent with a high level of infestation, resulting in more than 20 fleas per animal. All host populations showed a monoinfestation. *Ctenocephalides felis felis* was chosen as an outgroup taxon because molecular phylogenetic studies suggest that it is a close relative to *P. simulans* (Whiting, unpublished data).

Total genomic DNA was obtained using a Quiagen extraction kit (protocol for animal tissues). In order to estimate the quality and the amount of genomic DNA, 4 µl of extraction product were electrophoresed on a 1% TBE gel. All samples were sequenced for a fragment of the cytochrome B gene (*cytB*) and the cytochrome oxidase II gene (*CoII*). A fragment of 569 bp of the *cytB* gene was amplified using the overlapping primer pairs A6 (forward: 5'-CAT ATT CAA CCA GAA TGA TAT-3'), B2 (reverse: 5'-TGA TGA AAY YTT GGA TCA TTA-3') and A5 (forward: 5'-AGG RCA AAT ATC ATT TTG AG-3'), B 1.1 (reverse: 5'-AAA TAT CAT TCT GGT TGA ATA TG-3'). A total of 627 bp of the *CoII* gene were amplified by using the primers AtLeu (forward) and BtLys (reverse) (Maekawa et al. 1999). For the *cytB* gene region, *Taq* polymerase Gold was used in the PCR with the following cycle conditions: 95°C for 12 min; 30 cycles of 30 s at 95°C, 30 s at 40°C, 2 min at 68°C, plus an additional 7 min at 68°C. For the *CoII* fragment regular *Taq* polymerase was used with the following amplification profile: 4 min at 94°C, 33 cycles of 1 min at 94°C, 1 min at 47°C and 3 min at 72°C and a final extension of 5 min at 72°C. Negative controls were included to assess possible contamination.

A total of 6 µl of each PCR product was examined on ethidium bromide stained 1% agarose-TBE gels. PCR fragments were purified on Sephadex/SOPE-resin (Edge BioSystems) columns. Both strands of all PCR products were cycle sequenced using FS terminators (Big Dye version 3.0, Applied Biosystems). Following cycle sequencing, the fragments were cleaned on Sephadex columns and subsequently sequenced using the Applied Biosystems 3100 Genetic Analyzer.

Sequences were aligned and edited using Sequencher 4.1. Sequence alignment was straightforward for both gene regions due to relatively low levels of genetic divergence among samples and conservation of the codon reading frame.

The selective neutrality of the markers was tested using Tajima's D-test statistic as implemented in MEGA 2.1 (Tajima 1989; Kumar et al. 2001). The Nei-Gojobori method (MEGA 2.1) was used to compute the number of non-synonymous versus synonymous changes per site at the different branches of the MP tree (Zhang et al. 1998). Because the data set for the phylogenetic analysis consisted of two different genes, an incongruence length difference test (Mikevich and Farris 1981), as implemented in PAUP 4.0b10 (Swofford 1998), was used to assess data set heterogeneity. Maximum parsimony (MP), maximum likelihood (ML) (PAUP 4.0b10; Swofford 1998) and Bayesian inference (MrBayes 2.01; Huelsenbeck et al. 2001) was used for the phylogenetic reconstruction. Under MP, all characters were considered unordered and equally weighted. Under both MP and ML, a heuristic search strategy employing 100 random stepwise additions and TBR branch swapping was implemented. The nodal support was measured using the non-parametric bootstrap method (Felsenstein

**Fig. 1** Map of collection sites of *Pulex simulans* from domesticated guinea pigs in Peru. The numbers are explained in Table 1

1995; 1,000 pseudoreplications) and partitioned Bremer support values were calculated with TreeRotv2b (Sorenson 1999). Mr Bayes was run twice with the same dataset to test the congruence of results. One million trees were generated.

The program MODELTEST (Posada and Crandall 1998) selected the HKY 85 model of evolution for the Bayesian and the maximum likelihood analyses.

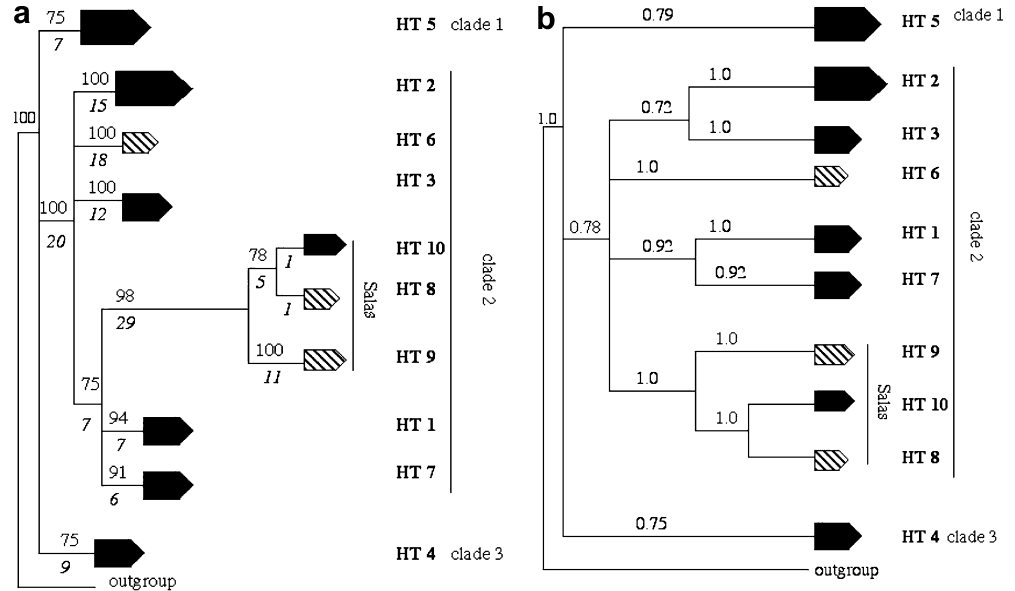
To complement the analysis, the nested pattern among the observed haplotypes was estimated. The networking algorithm developed by Templeton et al. (1995) and Templeton (1998) was used as implemented in TCS. The permutation tests were performed with GEODIS 2.0 (Posada et al. 2000). The inference key provided in Templeton and Sing (1993) was used to make statements about the processes that may be responsible for the observed clade structure.

Genetic polymorphism was estimated for the entire dataset and within each clade using a ML estimator implemented in FLUCTUATE (Kuhner et al. 1995). This program relaxes the assumption of constant population size and was deemed appropriate here as flea populations tend to undergo significant changes in population size.

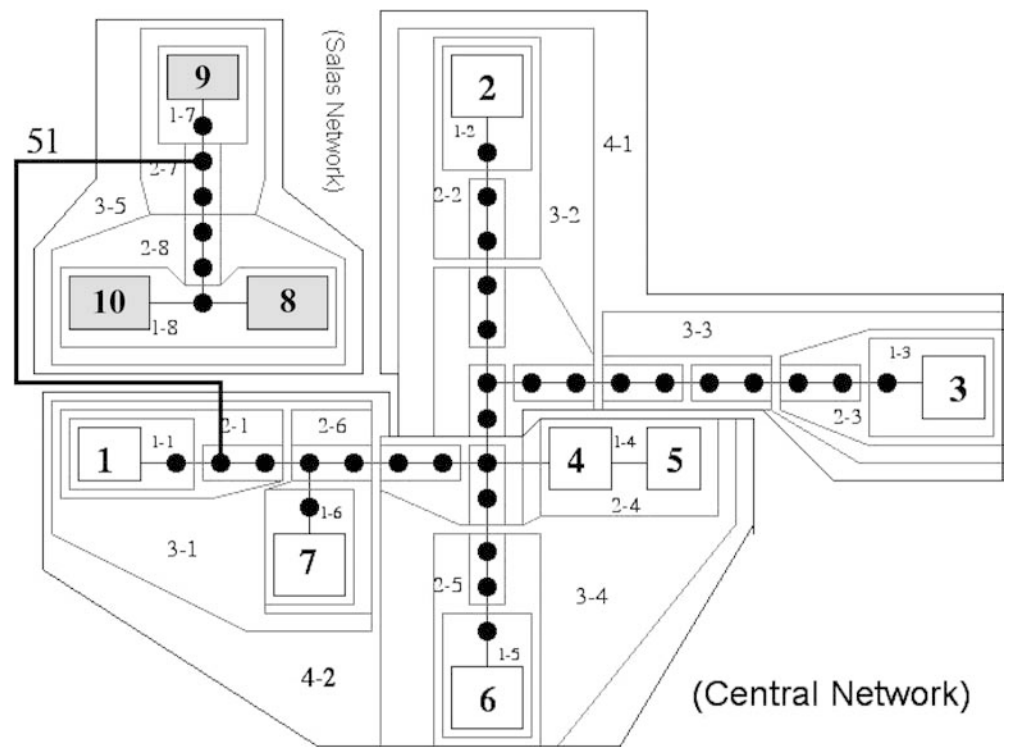
## Results

Non-parametric tests of the data set partition homogeneity (1,000 replicates) showed no evidence for heterogeneity ( $P=0.36$ ). Therefore the phylogenetic analysis was performed with the combined data set. However, the nested clade analysis was only performed with the *CoII* data. This is because the separate analysis of *cytB* with the nested clade approach provided identical, but less resolved results relative to the *CoII* data. Equally weighted MP analysis of the combined data set yielded four most parsimonious trees ( $L=269$ ;  $CI=0.94$ ;  $RI=0.999$ ) the strict consensus of which is given in Fig. 2a. All trees recovered *P. simulans* as three major clades. The clades correspond to haplotype 4 (clade 3), haplotype 5 (clade 1) and haplotypes 1, 2, 3, 6, 7, 8, 9, 10 (clade 2) (Fig. 2a). All Salas haplotypes (8, 9, 10) were recovered as a strongly supported group within clade 2, forming a monophyletic sister group to haplotypes 1 and 7. Interestingly, this branch is much longer than all other subclades (Fig. 2a), which is consistent with a higher substitution/fixation rate. The  $K_A/K_S$  ratio is 1.25, which suggests an event of positive selection for this branch.

**Fig. 2 a** Strict consensus phylogram of four equally weighted MP trees; bootstrap values (1,000 pseudoreplications) are shown if > 50%. **b** 50% majority rule consensus MCMC tree (MrBayes) of 20,000 trees; posterior probabilities are shown. *HT* corresponds to haplotypes and the size of the *pentagonal shape* indicates the number of individuals within this haplotype. The *light grey striped pentagonals* indicate that all included individuals correspond to one geographic region



**Fig. 3** Haplotype networks and nested clade design for *CoII* sequences of all *Pulex simulans* haplotypes. *Closed circles* represent extinct or unsampled haplotypes. The *numbers* inside the rectangles indicate the haplotype number. The *bold line* indicates the number of intermediate steps joining the two networks and their sites of connection



The Salas group forms the only subclade the haplotypes of which cluster mostly to a specific geographical location. However, haplotype 10 also contains individuals collected in the Moquegua area. The two Bayesian analyses produced congruent results. The consensus tree was entirely congruent with the MP tree, though less resolved. The tree generated under ML criteria was in all aspects congruent to the Bayesian tree and is not shown here.

Two networks were resolved using the *CoII* data set (Fig. 3). Parsimonious connections were statistically

justified ( $P_j \geq 0.95$ ) for between haplotype divergences of up to 15 mutational differences. Individual networks are not joined because divergence between the networks exceeded the 95% confidence limits of parsimonious connections derived from the estimation procedure (Templeton et al. 1995). By computing the pairwise base differences with PAUP 4.0b10 the number of intermediate steps joining the different networks was 51 (Fig. 3).

The nested design (Fig. 3) revealed that the central network (Fig. 3) is composed of non-geographically clustered haplotypes. The major haplotypes of the cen-

tral network are 2 (27) and 5 (24), both comprising multiple individuals from several locations (Huancayo, Lima and Caraz). All other haplotypes also correspond to at least two different localities.

The forming of the separate Salas network is consistent with a higher level of intraspecific divergence. Haplotype 10 includes individuals from two locations (Moquegua and Salas). GEODIS analysis (Posada et al. 2000) revealed significant non-random associations of some clades and sampling locations. For the central network as well as clade 3–4 of this network, the inference key of Templeton and Sing (1993) suggests a history of restricted gene flow and some long distance dispersal. Interior to these, the inference chain suggests a contiguous range expansion in association with 1–4, which comprises haplotypes 4 and 5, both containing individuals from Lima, Huancayo and Caraz. No firm conclusion could be drawn for the rest of the network, probably due to panmixia. The association of haplotypes 8 and 10 in clade 1–8 of the Salas network is best explained by some long distance colonization.

---

## Discussion

It is important to understand the underlying genetic structure of a vector because it provides insights into the threat of a parasitic infestation. For instance, these data suggest that there may not be a simple relationship between genetic similarity (haplotypes) and geographic regions. The results indicate that there is genetic structuring among fleas, although in this case it generally does not relate to a geographic pattern. In terms of a potential vector function or insecticide resistance this would mean that there will likely be no geographical barriers preventing its spread. However, by calculating the network connections in addition to a thorough phylogenetic analysis, one may be able to determine the responsible events for the lack of geographic patterning and predict possible vector-pathogen or host-insecticide resistance associations in certain locations. The occurrence of *P. simulans* and insect vectors in general is closely tied to the presence of their hosts. In our case the hosts (domesticated guinea pigs) are not continuously distributed throughout Peru. Rather, they are clumped in certain areas which are separated by areas of complete absence. However, the diagnosis of a contiguous range expansion between haplotypes 4 and 5 would suggest a gradual expansion over short distances in a certain timeframe. Haplotypes 4 and 5 are present in Lima, Huancayo and Caraz, which are separated by at least 200 km or more. Hence, the best explanations for a contiguous range expansion would be either a sampling error or the shift of *P. simulans* to other hosts, such as synanthrope rodents. These rodents could easily bridge the gaps between the different populations of guinea pigs and are known to occur quite frequently in their presence (Dittmar 2001). Thus, they could also easily spread

pathogens or act as an indirect reservoir for insecticide resistance genes.

We suggested that the haplotypes of *P. simulans* included in the central network may be characterized mainly by sporadic bouts of gene flow, over short or long distances, owing to the non-random movement by the infested guinea pigs. Locations 1–5 (Fig. 1) are interconnected through an extensive trade in guinea pigs for breeding purposes, and fleas are passively transported with them. This means that in the case of the spread of a pathogen this analysis would not only prove that there are apparent connections between even very distantly located areas, but it would also provide a working hypothesis with regard to the spread of the pathogen to a sylvatic cycle. If, for instance, the same pathogen carrying haplotypes of fleas on domestic animals would turn up in an epidemiological survey on wild rodents, chances are that this pathogen could also soon appear there, without an apparent direct contact with the infected domestic animals.

The Salas population is clustered in a separate network in the nested clade analysis and is the only clade with a geographical correlation. Superficially, one could assume isolation with no recent contact. The higher number of non-synonymous substitutions as well as the higher genetic polymorphism in the sequence data could be induced by a lower effective population size leading to selective pressure and higher fixation rates (Johnson and Seger 2001). However, the simultaneous presence of individuals from Moquegua and Salas within haplotype 10 challenges this hypothesis. Since all of the representatives of haplotype 10 found in Moquegua are as distantly related to the other haplotypes present at this location, as are its sister-haplotypes 8 and 9, it is likely that a recent introduction of this haplotype from the Salas area occurred into the Moquegua area. In the case of an imminent threat of a vector-transmitted disease, this would call for immediate action.

In the case of insecticide resistance, it may be difficult to predict the rate of development in any given population, as this will depend on the fluctuation of population sizes of the vector and thus the intensity of selection in that population. However, understanding the underlying genetic and phylogeographic structure generally allows for greater predictability in preventive measures against both of these problems.

**Acknowledgements** We thank the Karl-Enigk-Foundation, Germany, and the NSF grant no. DEB 9983195 for support and Manuel Gonzales for the help obtaining the Salas specimen.

---

## References

- Acha P, Szyfres B (1994) Zoonoses and communicable diseases common to man and animals. PAHO, Washington
- Campos J, Andrade CF (2002) Insecticide resistance in *Simulium* populations (Diptera, Simuliidae). *Cad Saude Publ* 18:661–671
- Didier R, La Scola B (2001) A flea associated *Rickettsia* pathogen of humans. *Emerg Infect Dis* 7:23–29

- Dittmar K (2001) Untersuchungen zum Vorkommen von Ektoparasiten bei domestizierten und wildlebenden Meerschweinchen (*Cavia* spp.) sowie an präinkaischen Meerschweinchenmumien in Peru, Südamerika. Doctoral dissertation, University of Leipzig, Leipzig
- Felsenstein J (1995) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* 294:2310–2314
- Johnson KP, Seger J (2001) Elevated rates of nonsynonymous substitution in Island birds. *Mol Biol Evol* 18:874–881
- Kuhner MK, Yamato J, Felsenstein J (1995) Estimating effective population size and neutral mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* 140:1421–1430
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Maekawa K, Kitade O, Matsumoto T (1999) Molecular phylogeny of the orthopteroid insects based on the mitochondrial cytochrome oxidase II gene. *Zool Sci* 16:175–184
- Mikevich MF, Farris SJ (1981) The implications of congruence in *Menidia*. *Syst Zool* 30:351–370
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14:817–818
- Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol Ecol* 9:487–488
- Sorenson MD (1999) TreeRot, version 2. Boston University, Boston
- Swofford DL (1998) PAUP\* Phylogenetic analysis using parsimony (\*and other methods), version 4.0. Sinauer, Sunderland, Mass.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–596
- Templeton AR (1998) Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Mol Ecol* 7:381–397
- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analysis with cladogram uncertainty and recombination. *Genetics* 134:659–699
- Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of geographical distribution of mitochondrial DNA haplotypes in the tiger salamander *Ambystoma tigrinum*. *Genetics* 140:767–782
- Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* 95:7308–7313
- Ziam M, Guillet P (2002) Alternative insecticides: an urgent need. *Trends Parasitol* 10:161–163