Supporting Information for online publication only

Manuscript:

HUMAN FACILITATED METAPOPULATION DYNAMICS IN AN EMERGING PEST SPECIES, *CIMEX LECTULARIUS*

Toby Fountain, Gavin Horsburgh, Ludovic Duvaux, Klaus Reinhardt, Roger K. Butlin

Microsatellite isolation

Microsatellite sequences were isolated using two methods, the creation of a microsatellite-enriched library and through analysis of a recently available transcriptome assembly (Otti and Reinhardt, in prep). First, genomic DNA was extracted using the ammonium acetate precipitation method described by Nicholls *et al.* (2000). A microsatellite-enriched library was constructed using the method of Armour *et al.* (1994). The library was enriched for the following di- and tetranucleotide microsatellite motifs: (GT)n, (CT)n, (GTAA)n, (CTAA)n, (TTTC)n and (GATA)n, which had been bound to magnetic beads by following the procedure described by Glenn & Schable (2005). Transformant colonies were not screened for the presence of a repeat but were directly sequenced by the NERC Biomolecular Analysis Facility at the University of Edinburgh. For the transcriptome approach, transcripts were screened for microsatellite sequences using the program MSATcommander (Faircloth 2008).

In total 436 sequences containing unique microsatellite regions were isolated (329 from the genomic library and 107 from the transcriptome). However, the majority of the sequences contained only short flanking regions or had compound repeats, leaving only 62 sequences that were useful for primer design. Primer sets were designed from these sequences using the program PRIMER3 (Rozen & Skaletsky 2000). These primers were tested for amplification using four individual bed bugs from a lab stock over a gradient of 12 different annealing temperatures (56 – 64°C) using a DNA Engine 2 thermal cycler (MJ Research, Bio-Rad, Hemel Hempstead, Herts, UK). A further 24 individuals were then genotyped at the polymorphic

loci using the temperature that produced the clearest and most consistent product when observed on an ABI3730 48-well capillary DNA analyser (Applied Biosystems, California). Each PCR contained approximately 10ng of genomic DNA, 0.2µM of each primer and 1µl QIAGEN multiplex PCR mix (QIAGEN Inc; (Kenta *et al.* 2008)). PCR amplification was performed using a DNA Engine Tetrad Thermal Cycler (MJ Research, Bio-Rad, Hemel Hempstead, Herts, UK) with the following touchdown program: 95°C for 15 minutes, followed by 11 cycles of 94°C for 30 seconds, annealing temperature of 65°C (which decreases by 1°C each cycle) for 1 minute 30 seconds, and 72°C for 1 minute. Then 26 cycles of 94°C for 30 seconds, 55°C for 1 minute 30 seconds and 72°C for 1 minute, then finally 72°C for 10 minutes. Amplified product was then analysed using an ABI3730 48-well capillary DNA analyser and allele sizes assigned using GENEMAPPER v.3.7 (Applied Biosystems, California).

Microsatellite characterisation

From the 39 genomic-library designed primers 15 loci did not amplify and a further 8 were monomorphic or yielded no specific product and from the 20 transcriptome designed primers 10 markers failed to amplify and five were monomorphic. For the 21 remaining markers the number of alleles per locus ranged between three and 12 with an average of 7.4 alleles per locus across the 18 infestations sampled in this study (see Table 4). All markers consistently amplified and yielded unambiguous product suitable for analysis. Null allele estimates were high (>0.2) for five markers but this may be an artefact of general low observed heterozygosity across loci caused by recent severe bottlenecks in the sampled infestations (see '*Within infestation diversity*' in the main text). No marker was found to be

sex linked and after Bonferroni correction no significant linkage disequilibrium was detected between markers.

References

Armour JA, Neumann R, Gobert S, Jeffreys AJ (1994) Isolation of human simple repeat loci by hybridisation selection. *Hum. Mol. Genet.* **3**(4), 599-605.

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A (accepted in Bioinformatics) DIYABC v2.0: a software to make Approximate Bayesian Computation inferences about population history using Single Nucleotide Polymorphism, DNA sequence and microsatellite data.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus specific primer design. *Molecular Ecology Resources*, **8**(1), 92-94.

Glenn TC , Schable NA (2005) Isolating microsatellite DNA loci. *Meth. Enzymol.* **395**, 202-222.

Kenta T, Gratten J, Haig NS *et al.* (2008) Multiplex SNP-SCALE: a cost-effective medium-throughput single nucleotide polymorphism genotyping method. *Molecular Ecology Resources*, **8**, 1230-1238.

Nicholls J, Double M, Rowell D, Magrath R (2000) The evolution of cooperative and pair breeding in thornbills Acanthiza (Pardalotidae). *Journal of Avian Biology*, **31**, 165–176.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods. Mol. Biol.*, **132**, 365-386.

Tables and Figures

Table S1. Within infestation diversity sample information

Infestation	Country	Locality	No. Individuals	No. Refugia
AUS	Australia	Rozelle	41	4
BIR1	UK	Birmingham	8	2
BIR2	UK	Birmingham	9	2
LON1	UK	London	52	12
LON2	UK	London	46	11

Table S2. Description of newly designed primers arranged into 5 multiplex panels. T_a = annealing temperature, origin of the markers (Microsatellite library, transcriptome) and total number of observed alleles across 19 infestations. Markers Cle001 and Cle006 (in bold) were not used in the within-infestation analysis due to problems with an early version of the multiplexing.

Locus	Accession No.	Repeat Motif	Primer Sequence 5'-3'	T_a	Multiplex/ Fluro Label	Expected/ Observed Size (bp)	Library/ Transcriptome	N_A
Cle001	HF969864	**$(TG)_{22}$**	**F TTCGTGTCACAGAAAGGTAAAGAG** **R AAGCCGAAATGTGATTCATTG**	**60**	**3/VIC**	**235** **227-338**	**Library**	**6**
Cle002	HF969865	$(CA)_{23}$	F CATGAAATTGGGAGTTTCTATGTTTC R TTACCGCCCATGTAAACGAG	60	2/6FAM	257 225-242	Library	12
Cle003	HF969866	$(TG)_{29}$	F TTCGTTTGTGTAGAACCTTGG R TACGTCCCTACAAGCTCACC	60	4/6FAM	269 214-227	Library	7
Cle004	HF969867	$(AC)_{19}$ $(CG)_1$ $(AC)_{14}$	F AACAGACAGTTGTCAGATTGGTATG R TGGAGCACCTTCTTAAACTTCAC	60	1/6FAM	286 219-278	Library	7
Cle005	HF969868	$(CT)_2$ $(CA)_1$ $(CT)_{17}$	F AGATAGGGCAACCTTTCAGAG R TTGGTGATAGTGAACGAACG	60	5/NED	315 306-316	Library	4
Cle006	HF969869	**$(GT)_{41}$**	**F AGTTTCGGGTGGTGATGTATG** **R TTGAATCGCATTCTTTGTGC**	**60**	**5/VIC**	**353** **307-354**	**Library**	**7**
Cle007	HF969870	$(GA)_{15}$	F TTGTGAGTGTGTCTCTCTCTACTGTG R CAGGTTCACAGGCCAAATG	60	4/6FAM	144 139-162	Library	8
Cle008	HF969871	$(AC)_{24}$	F TTACCGTCATTTATACTATCTGGTTGG R GGGAAGAAGCCCAGTCAC	60	4/VIC	161 130-230	Library	8
Cle009	HF969872	$(AG)_{34}$	F GACACCATGTATATATGTCGTAGATTGAG R GCGACGGAGTAAGAAACCAC	60	5/PET	198 172-232	Library	8
Cle010	HF969873	$(GA)_{24}$	F CCATTGACGGGTTGCTTC R CACTTTCTTGTAACCATCACCATC	60	3/NED	164 138-217	Library	12
Cle011	HF969874	$(GA)_{26}$	F AAATAATTTAGCTGCAAACAATAGG R TTGCCAGATTTCTCAATCG	60	4/NED	182 160-180	Library	11
Cle102	HF969875	$(TC)_{27}$ $(TTCATT)_1$ $(CT)_3$	F AGACCGTATAGTGCCAGGAGTC R AAGTCGACCGTCCACGAC	60	2/NED	197 187-242	Library	10
Cle013	HF969876	$(GT)_{21}$ $(TAT)_1$ $(GT)_9$	F TTCACAGATTTAAGCCTAACTGGTC R CAAATAACCTCGAATTCATACGC	60	1/VIC	233 201-269	Library	8
Cle014	HF969877	$(GA)_{38}$ $(GT)_5$ $(GAGA)_2$ $(TA)_1$ $(GA)_6$	F TGGGTGGGTGCTATTGTTTC R AGCAGTGTTGGCCGAGATAC	60	2/VIC	283 215-250	Library	6
Cle015	HF969878	$(TC)_6$	F TCATATGGGCGGATTAGAGC R TAACAATCTGGAGGCGGAAC	60	1/NED	302 290-304	Library	7
Cle016	HF969879	$(AC)_{34}$	F GTGTGACACATCGAGCAACC R ACGCCAGACGTGAATAAAGC	60	4/PET	352 181-357	Library	10
Cle017	HF969880	$(CG)_4$ $(GT)_{12}$	F TCGTGTCCACGCCTTAAAC R ATTCGTATACTCTCTCGAATTCTGC	60	2/PET	288 281-300	Transcriptome	7
Cle018	HF969881	$(AG)_{13}$ $(AT)_1$ $(AG)_3$	F ACGCAAAGTTAGGTCTCTTCAAC R TCCATCTATCTTTCCCTCTCTACC	60	4/6FAM	300 295-309	Transcriptome	3
Cle019	HF969882	$(CT)_5$ $(CA)_{11}$	F GATGGCGAGTGACGACTTG R TTGAAAGGACCGCTGCTC	60	1/6FAM	151 148-162	Transcriptome	5
Cle020	HF969883	$(ACA)_{12}$	F GGGCGGGTTGTCTAATTG R AGTAAGGGACCTCTCGAATCC	60	5/VIC	169 147-165	Transcriptome	4
Cle021	HF969884	$(TC)_4$ $(ATT)_1$ $(TC)_{12}$	F GCAACCCTGGACTTCTCAAC R TCAGCTCTCCATTAGAACGAAAC	60	3/6FAM	188 237-270	Transcriptome	6

T_a = Annealing temperature (°C)
N_A = Total number of alleles across 18 infestations

Table S3. Pairwise F_{ST} estimated using Weir and Cockerhams's ϑ (1984) between 11 refugia in the LON2 infestation

	LON2_A	*LON2_B*	*LON2_C*	*LON2_D*	*LON2_E*	*LON2_F*	*LON2_G*	*LON2_H*	*LON2_I*	*LON2_J*
LON2_B	-0.100									
LON2_C	-0.086	0.019								
LON2_D	0.081	0.140	0.067							
LON2_E	-0.138	0.103	0.151	0.100						
LON2_F	-0.055	0.381**	0.156	0.328**	-0.038					
LON2_G	-0.101	0.021	-0.015	0.316*	0.058	-0.130				
LON2_H	-0.021	0.287*	0.102	0.222*	-0.061	-0.010	-0.062			
LON2_I	-0.167	0.156	-0.073	0.014	-0.087	-0.111	-0.088	0.013		
LON2_J	-0.040	0.081	0.268	0.223	-0.102	-0.097	0.031	-0.054	-0.117	
LON2_K	-0.140	-0.052	-0.082	0.222	-0.219	-0.003	-0.112	-0.157	-0.091	-0.039

*significant (p<0.05), **highly significant (p<0.01)

Table S4. Pairwise F_{ST} estimated using Weir and Cockerhams's ϑ (1984) between five *C. lectularius* infestations

	AUS	*BIR1*	*BIR2*	*LON1*
BIR1	0.602***			
BIR2	0.551***	0.492*		
LON1	0.760***	0.834***	0.717***	
LON2	0.706***	0.705***	0.612***	0.793***

*Significant (P<0.05)

***Highly Signifcant (P<0.001)

Table S5: Deviation of test statistics generated from predictive posterior distribution for each scenario from observed data (attached in a supplementary excel table)

Table S6 Prior and posterior distributions of model parameters for additional simulations using DIYABC v2 (Cornuet et al. accepted; http://www1.montpellier.inra.fr/CBGP/diyabc/). DIYABC v2 allowed the extension of the prior on mutation rate to $1 \times 10^{-6} - 1 \times 10^{-3}$. Priors for all other parameters remained the same as in the original analysis. 1×10^{6} simulations were performed for each scenario. The parameter estimates for scenario 2 are shown.

Parameters	Prior range	Mean	Median	Mode	HPD90 low	HPD90 high
Ns	Loguniform [100 - 50000]	6160	3940	2510	1490	19400
Ne	Loguniform [10 - 100]	28.5	27.4	25.8	12.5	49.0
Nb	Uniform [2 - 14]	6.64	6.11	2.00	2.06	12.8
t2	Loguniform [2 - 10]	3.95	3.20	2.00	2.00	8.48
t3	Uniform [11 - 100]	58.5	57.8	50.8	23.2	95.2
Mean μ	Uniform [10^{-6} - 10^{-3}]	3.32×10^{-4}	3.41×10^{-4}	8.80×10^{-5}	5.55×10^{-5}	8.36×10^{-4}
Mean P	Uniform [0.1 - 0.3]	0.120	0.110	0.100	0.100	0.178

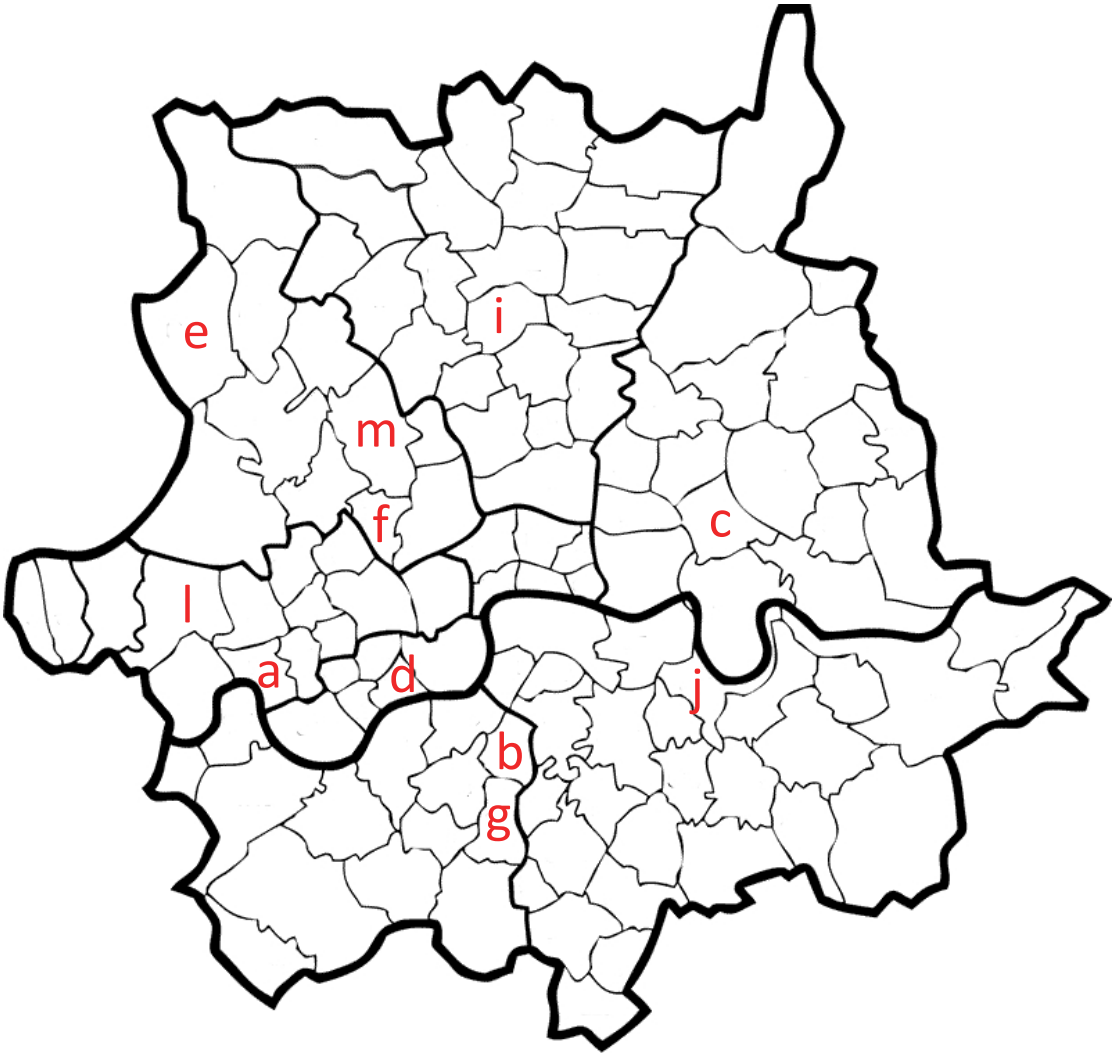

Figure S1: A map of the sampling locations in the city of London, UK used to test between infestation diversity. Each region corresponds to a postcode area and each letter an individual infestation. h and k are located outside the city and at co-ordinates 51.6905N, -0.0338W and 51.3843N, -0.4207W, respectively.

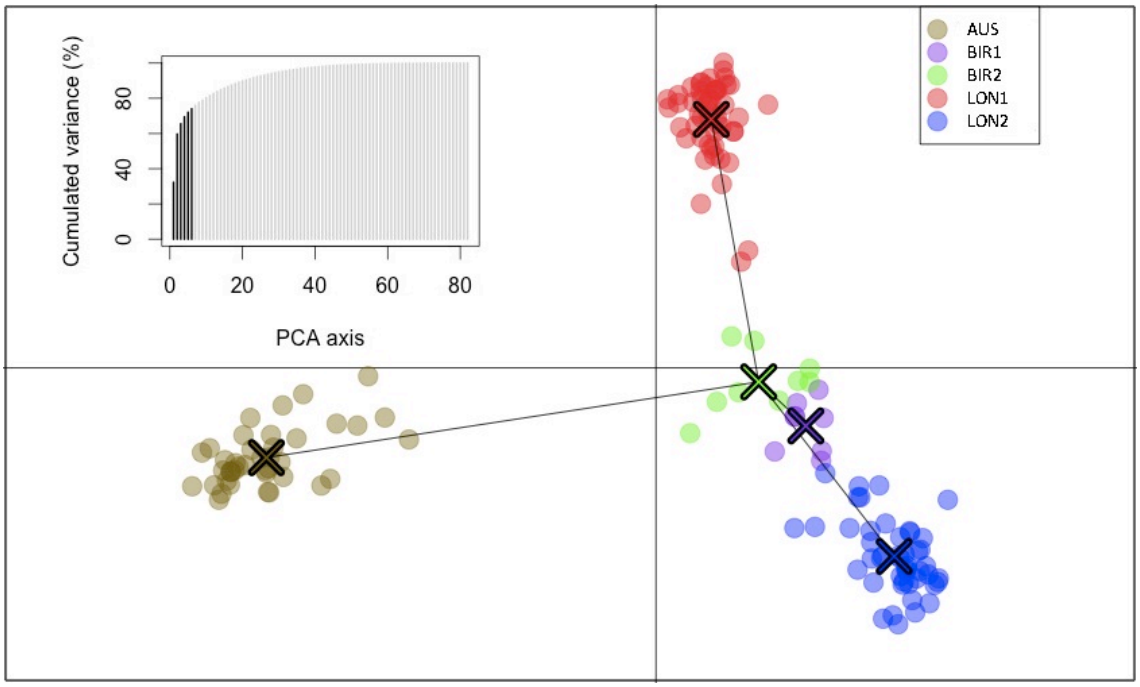

Figure S2 Ordination plot of DAPC genetic clusters in the within-infestation dataset, including a minimum spanning tree based on the squared distances between infestations within the entire space. Centre of each cluster is designated with an X. Cumulative variance plot shows the number of retained principal components in the analysis, all discriminant functions were retained.

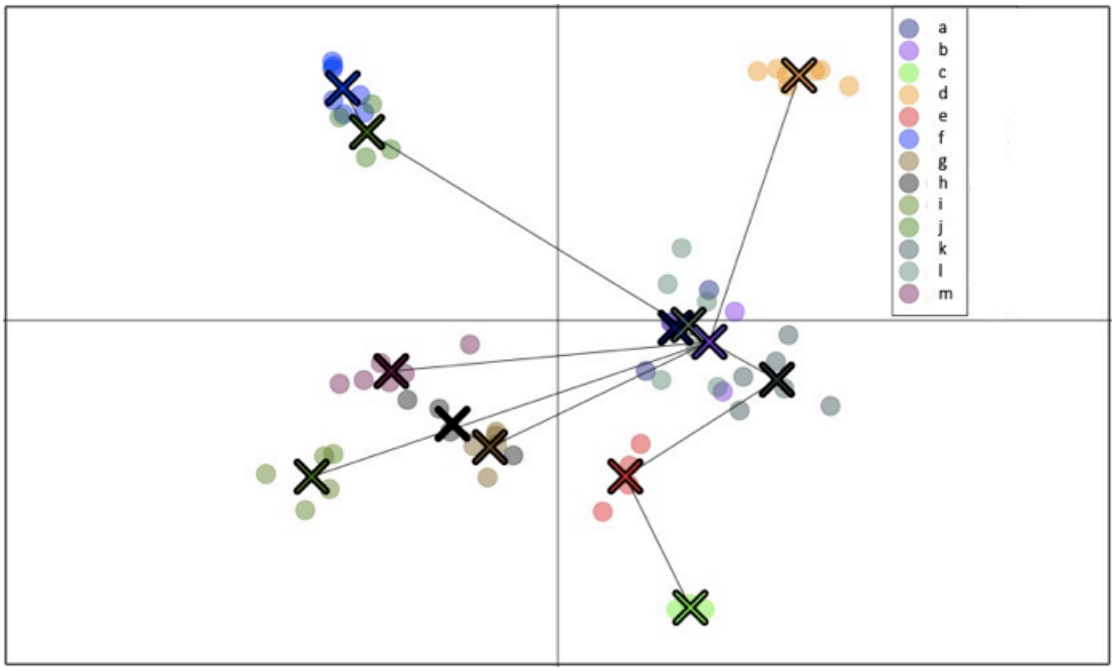

Figure S3 Ordination plot of DAPC genetic clusters in the metapopulation diversity data set, including a minimum spanning tree based on the squared distances between infestations within the entire space. Centre of each cluster is designated with an X.

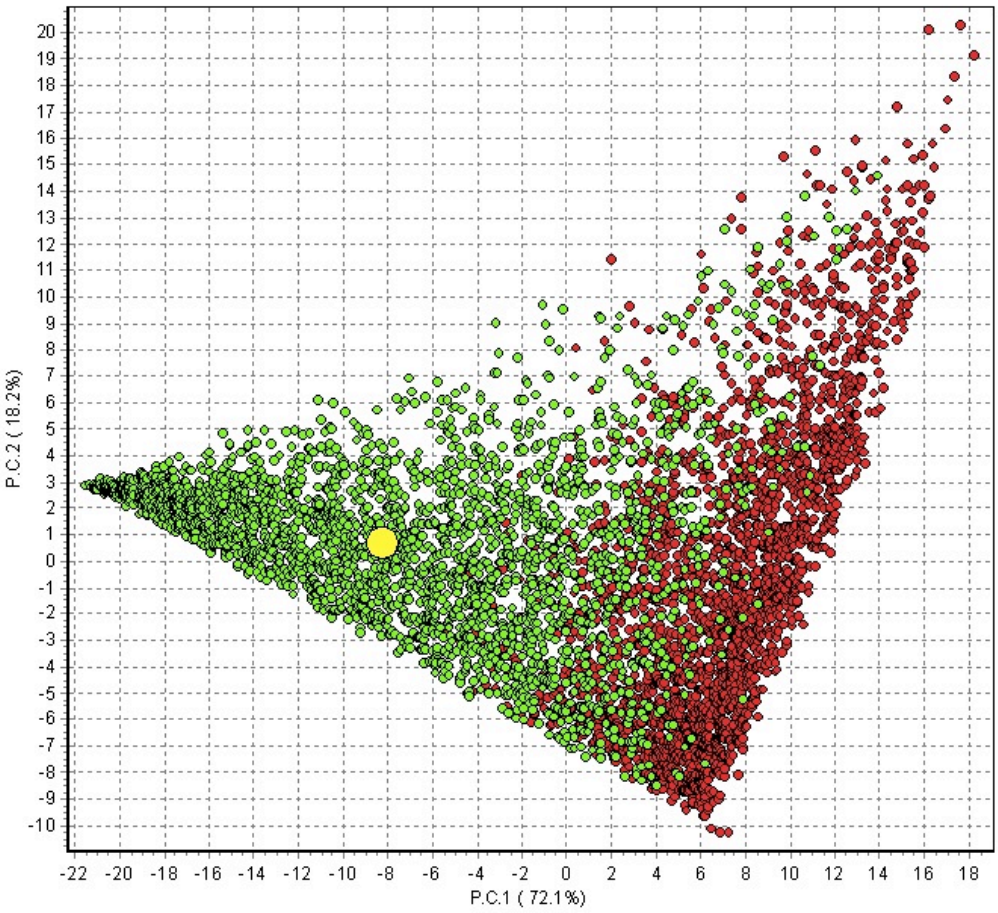

Figure S4 Principal components analysis of summary statistics of 100,000 simulated data sets generated with two demographic scenarios (red = scenario one, green = scenario two). Observed dataset (yellow dot) is projected on the plane formed by the first two principal components.

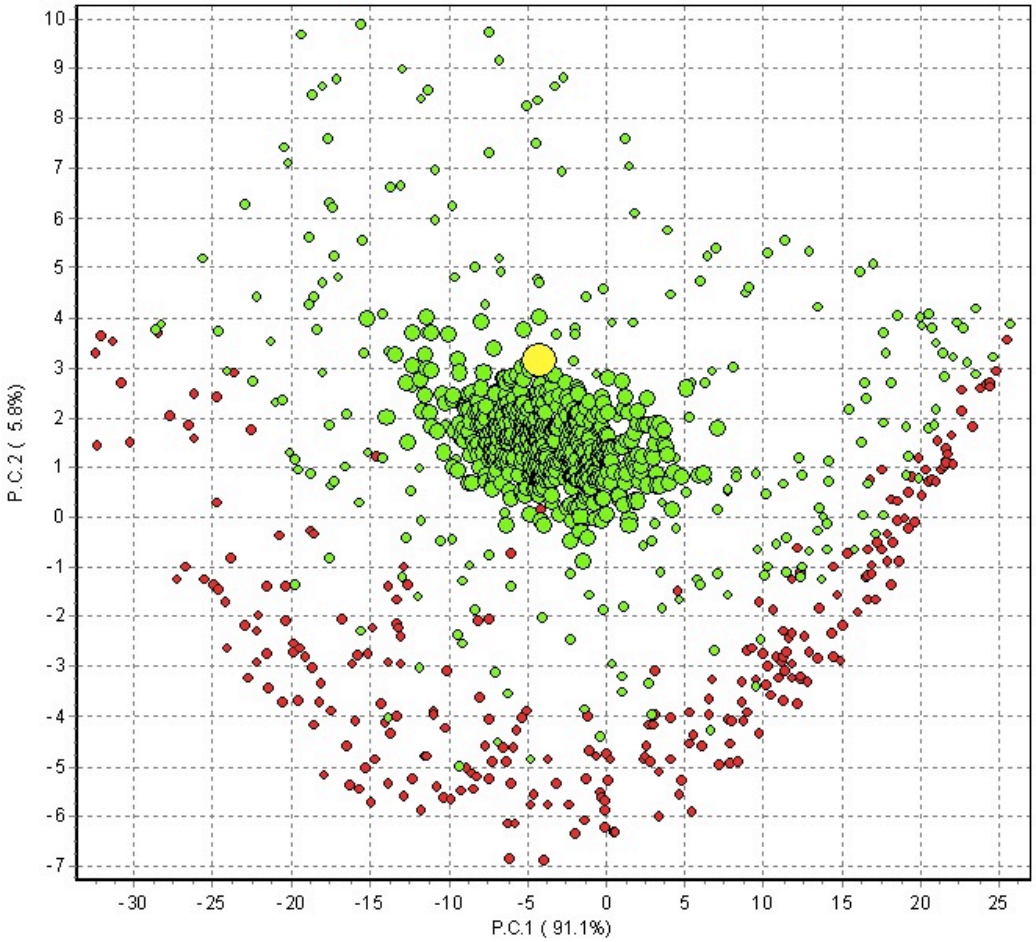

Figure S5 Principal components analysis of summary statistics from 500 pseudo-observed datasets generated from the posterior predictive distribution of scenario two (large green dots), plotted along with summary statistics of simulations generated from priors of scenario one and two (small red and green dots respectively). The observed dataset (yellow dot) is projected on the plane formed by the first two principal components.

Figure S6 Prior (Red) and Posterior (Green) distributions of parameters obtained under scenario 2 using the adjusted mutation rate (Table S6). The x-axis shows the range of parameter values, and the y-axis the probability density (attached as a supplementary PDF).