Ecography

ECOG-02855
Gagne, R. B., Sprehn, C. G., Alda, F., McIntyre, P. B., Gilliam, J. F. and Blum, M. J. 2017. Invasion of the Hawaiian Islands by a parasite infecting imperiled stream fishes. – Ecography doi: 10.1111/ecog.02855

Supplementary material

Supplementary material

Appendix 1. Molecular methods: sequencing and genotyping

Appendix 2. Approximate Bayesian Computation analysis methodology

Appenix 3.

Table A1. Distribution of mitochondrial *cox1* haplotypes found in *Camallanus cotti* infecting *Awaous stamineus* in the Hawaiian Islands. Mitochondrial DNA sequence data were obtained for 31 of the 35 sampled watersheds.

Figure A1. Neighbor joining tree inferred from *Camallanus cotti* mitochondiral *cox1* sequences. The tree represents samples from the native range (China) and the introduced range (Hawaiian Islands). Nodal bootstrap support values ≥50 are shown above adjacent branches.

Figure A2. (A) Posterior probabilities [lnP(D)] from Bayesian cluster analysis of multi-locus genotypes with varying number of genetic clusters (K). (B) Rate of change of the posterior probabilities (ΔK) with varying number of genetic clusters (K).

APPENDIX 1.

Mitochondrial DNA Sequencing

A 427 base pair (bp) fragment of the mitochondrial *cytochrome oxidase I* gene (*cox1*) was amplified from 241 parasites collected in 2009. Polymerase chain reactions (PCRs) were performed using the primers FCOX1A (5'-ATY GG Y GGT TTT GGT AAT TG-3') and RCOX1A (5' -GAA GTA TTW AAA TTA CGA TC-3') (Wu *et al.*, 2009). A 15 µL reaction mixture was used containing: 1X PCR *Hotaq* buffer which includes 2 mM $MgCl_2$ (MCLab, Inc. San Francisco, CA), 0.83 mM $MgCl_2$, 0.083mM of each dNTP, 0.5 units *Hotaq* polymerase (MCLab, Inc. San Francisco, CA), 0.5 µM of each primer, and 2 µL DNA template at 10ng/µl concentration. Mastercycler pro thermocyclers (Eppendorf AG, Hamburg) were run with the following cycling parameters: a 10 minute 95°C hot start activation; 32 cycles of a 1 minute 94°C denaturing step, a 1 minute 51°C annealing step, and a 2 minute 72°C elongation; and a 10 minute 72°C final elongation period. PCR products were purified in 5 µL ExoSAP-It (USB, Affymetrix, Cleveland, OH) reactions diluted by a factor of 10 in ddH_2O. Cycle-sequencing reactions were then carried out using 2 µL Dyenamic (GE Healthcare, Buckinghamshire, UK), 2 µL 3.2 µM primer, 4 µL ddH_2O, and 2 µL cleaned PCR product. Cycle-sequencing products were purified using Sephadex (GE Healthcare Biosciences, Pittsburgh, PA) and sequenced in both directions on an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA). Sequences were assembled and edited using Sequencher v4.9 software (Gene Codes Corp., Ann Arbor, MI).

Microsatellite Development and Analysis

A suite of new microsatellite markers were developed for *C. cotti* using paired end Illumina sequencing (Castoe *et al.*, 2012). DNA from two individuals sampled from O`ahu and Moloka`i was sent to the Savannah River Ecology Laboratory in Georgia where Illumina runs were conducted to generate paired end 100 bp reads. Sequence reads were then analyzed with the program PAL_FINDER_v0.02.03 (Castoe *et al.*, 2012) to identify potential microsatellite loci. Loci were sorted by the number of appearances in the sequence data and by motif length. From this, 20 tetranucleotide repeats microsatellite loci were selected for screening. Candidate loci were run in Primer 3 (Rozen & Skaletsky, 1999) for primer design. The Qiagen multiplex kit (Qiagen, Inc., Valencia, CA) was used to conduct PCR trials for all twenty candidate loci. Of the 20 candidates, 14 loci successfully and consistently amplified. Information on the 14 loci and corresponding primers is provided in Table 2.

All individuals (2009, n = 278; 2011, n = 566) were subsequently genotyped at 14 loci in 10 µl PCR mixtures containing: 5 µl Qiagen master mix, 1 µl primer mixture containing 0.2 µM of each primer, 3 µl of RNase-free water, and 1 µl of template DNA at 10 ng/µl concentration. Thermocycler parameters were as follows: an initial heat activation of 15 minutes at 95°C; followed by 35 cycles of 30 seconds denaturing at 94°C, 90 seconds annealing at 59°C, 90 seconds extension at 72°C; and a final extension of 10 minutes at 72°C. Forward primers were labeled with HEX, FAM, or TAMRA fluorescent dyes to produce labeled PCR amplicons that were ran on an ABI 3730xl automated DNA sequencer (Applied Biosystems, Foster City, CA) with a 600 LIZ size standard (Applied Biosystems®, Waltham, MA). Electropherograms were scored and

binned with GeneMarker v9.0 software (Softgenetics, State College, PA). To evaluate potential genotyping errors all individuals with private alleles and a random subset of 75 individuals were genotyped twice. Data from three loci with less than three alleles were discarded; the remaining eleven loci were used in all subsequent analyses.

REFERENCES

Castoe, T.A., Poole, A.W., de Koning, A.P., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N. & Pollock, D.D. (2012) Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS One*, **7**, e30953.

Rozen, S. & Skaletsky, H. (1999) Primer3 on the WWW for general users and for biologist programmers. *Bioinformatics methods and protocols*, pp. 365-386. Springer.

Wu, S.G., Wang, G.T., Xi, B.W., Xiong, F., Liu, T. & Nie, P. (2009) Population genetic structure of the parasitic nematode *Camallanus cotti* inferred from DNA sequences of ITS1 rDNA and the mitochondrial COI gene. *Veterinary Parasitology*, **164**, 248-56.

APPENDIX 2.

Testing models of invasion history

All scenarios were run twice: in one set of runs, the prior distribution of demographic parameters considered all colonization events to have occurred over the last 200 years (i.e., the maximum time since poeciliid introductions), and the second set of runs considering colonization events to have occurred over 1,000 years (i.e., to account for potential introductions by Polynesian colonists). The timing for colonization events was set following the information available from historical records. Also, we assumed that the effective population size was the same for all populations and used a uniform distribution between 10-100,000 individuals, except for the ancestral population which was bounded between 10-10,000,000 individuals. Genetic variation within and between populations was characterized using a set of summary statistics. For the microsatellite data in each population and each population pair, we used the mean number of alleles per locus, the mean expected heterozygosity, the mean allele size variance, mean Garza and Williamson's *M* statistic for every population, pairwise F_{ST}, and shared allele distance values. For the mitochondrial data we used the number of haplotypes, the number of segregating sites and the mean number of pairwise differences for each population and population pair. We also used pairwise F_{ST} values and Tajima's *D* for the one sample summary statistics. We simulated 10^6 datasets per scenario to build reference tables. Posterior probabilities of the competing invasion scenarios were calculated using 1% of the simulated datasets to estimate the relative posterior probability of each scenario with a logistic regression over a linear discriminant analysis of the summary statistics (Estoup *et al.*, 2012). Posterior predictive error of the competing invasion scenarios (i.e., the

number of wrongly identified scenarios out of simulated datasets) was calculated using 500 simulated datasets closest to the observed dataset for the direct approach, as well as using 10^3 simulated datasets for the logistic approach (Estoup *et al.*, 2012; Cornuet *et al.* 2014).

REFERENCES

Estoup, A., Lombaert, E., MARIN, J.M., Guillemaud, T., Pudlo, P., Robert, C.P. & CORNUET, J. (2012) Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular ecology resources*, **12**, 846-855.

Table A1. Distribution of mitochondrial cytochrome *c* oxidase I haplotypes found in *Camallanus cotti* infecting *Awaous stamineus* in the Hawaiian islands. Mitochondrial sequence data was obtained for 31 of the 35 sampled watersheds.

Island	Watershed	H1	H2	H3	H4	H5	H6	Total
Hawai`i	Honoli`i	7						7
	Wailoa R	11						11
	Wailoa	5						5
	Nanue	1						1
	Ka`ie`ie	3						3
	All	**27**	**0**	**0**	**0**	**0**	**0**	**27**
Kaua`i	Hanalei	2						2
	Anahola	2						2
	Lawa`i	17						17
	Kapa`a	8						8
	Moloa`a	2						2
	Wainiha	17						17
	Waimea	8						8
	All	**56**	**0**	**0**	**0**	**0**	**0**	**56**
Maui	Iao			8				8
	Waihe`e			7				7
	Honokohau		3	1				4
	Pi`ina`au			8				8
	All	**0**	**3**	**24**	**0**	**0**	**0**	**27**
Moloka`i	Halawa	8		12	2			22
	Honouli Wai	7		5			1	13
	Pelekunu	12		1				13
	Wailau	20						20
	Waikolu	8		5				13
	All	**55**	**0**	**23**	**2**	**0**	**1**	**81**
O`ahu	Kea'ahala	7						7
	Walkane	1						1
	Waimanalo	6	1					7
	Waimea	7						7
	Alawai	1						1
	Kahana	3				1		4
	Kaluanui	1						1
	Ki`iki`i	10	1	1				12
	Kahaluu	4						4
	Wailupe			1				1
	All	**40**	**2**	**2**	**0**	**1**	**0**	**45**
Archipelago		**178**	**5**	**49**	**2**	**1**	**1**	**236**

Figure A1.

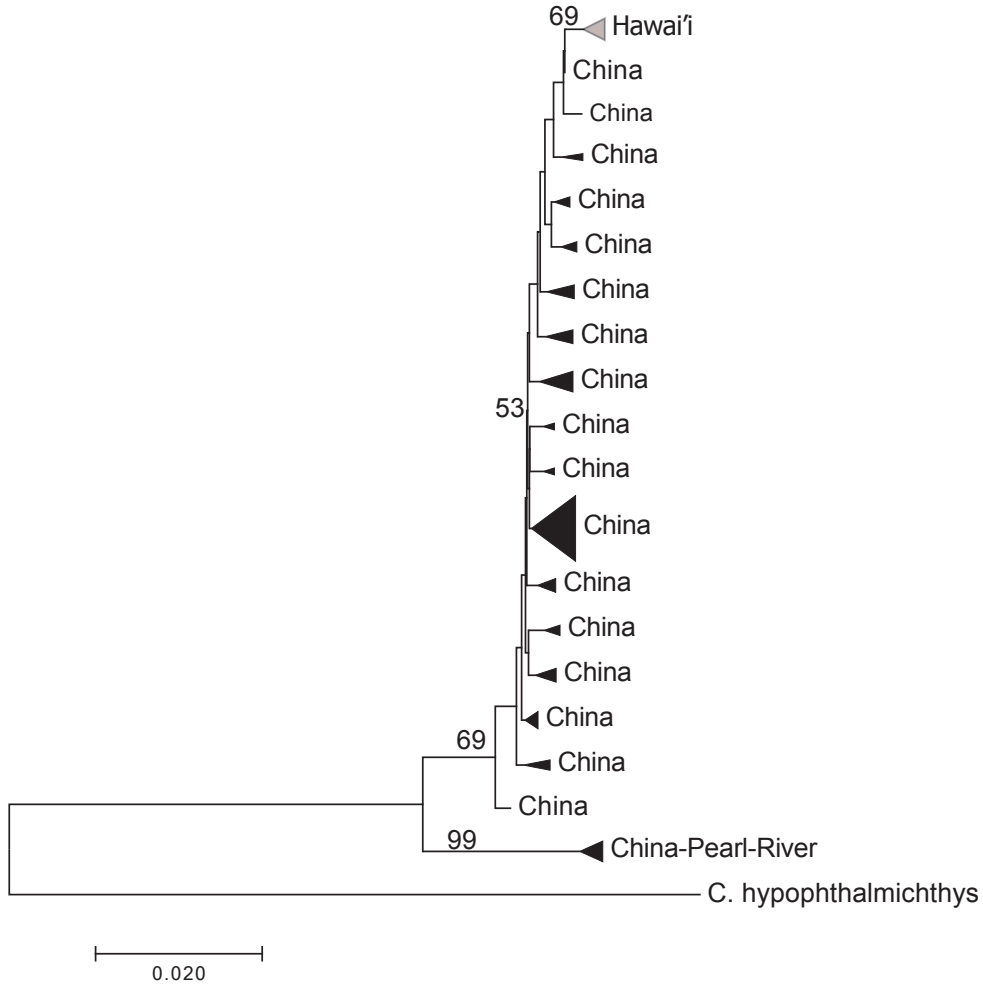
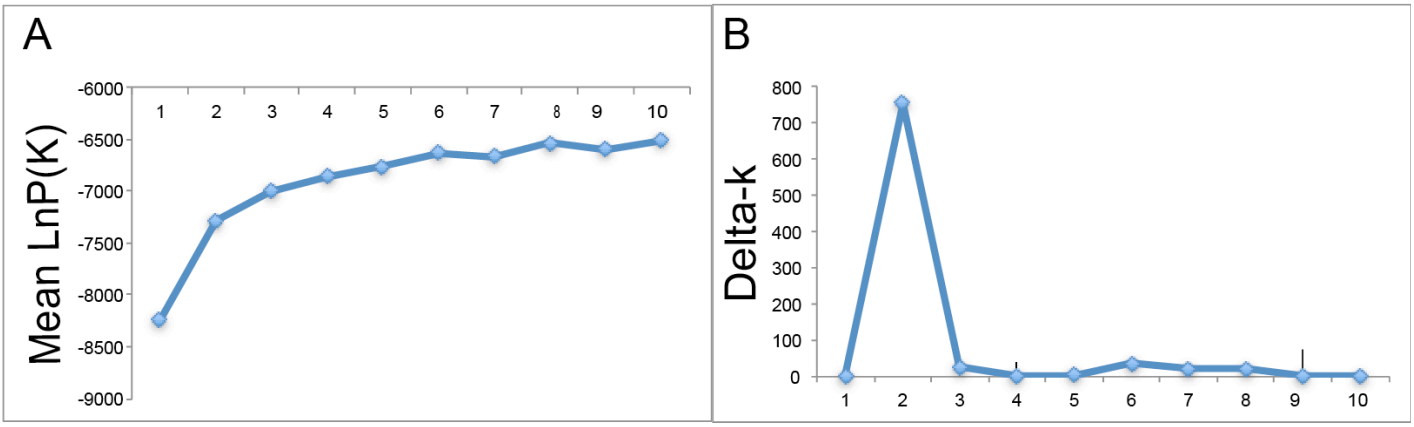

69
Hawai'i
China
China
China
China
China
China
China
China
53
China
China
China
China
China
China
69
China
China
99
China-Pearl-River
C. hypophthalmichthys
0.020

Figure A2


A
Mean LnP(K)
-6000
-6500
-7000
-7500
-8000
-8500
-9000
1
2
3
4
5
6
7
8
9
10
B
Delta-k
800
700
600
500
400
300
200
100
0
1
2
3
4
5
6
7
8
9
10