Next-generation sequencing reveals cryptic mtDNA diversity of *Plasmodium relictum* in the Hawaiian Islands

S. I. JARVI[1]*, M. E. FARIAS[1], D. A. LAPOINTE[2], M. BELCAID[3] *and* C. T. ATKINSON[2]

[1] *Department of Pharmaceutical Sciences, Daniel K. Inouye College of Pharmacy, University of Hawaii at Hilo, 34 Rainbow Drive, Hilo, HI 96720, USA*
[2] *U.S. Geological Survey, Pacific Island Ecosystems Research Center, Kilauea Field Station, P.O. Box 44, Building 343, Hawaii National Park, HI 96718, USA*
[3] *Information and Computer Sciences Department, University of Hawaii at Manoa, Pacific Ocean Science and Technology Building, Room 317, 1680 East-West Road, Honolulu, HI 96822, USA*

SUMMARY

Next-generation 454 sequencing techniques were used to re-examine diversity of mitochondrial cytochrome *b* lineages of avian malaria (*Plasmodium relictum*) in Hawaii. We document a minimum of 23 variant lineages of the parasite based on single nucleotide transitional changes, in addition to the previously reported single lineage (GRW4). A new, publicly available portal (Integroomer) was developed for initial parsing of 454 datasets. Mean variant prevalence and frequency was higher in low elevation Hawaii Amakihi (*Hemignathus virens*) with *Avipoxvirus*-like lesions ($P = 0.001$), suggesting that the variants may be biologically distinct. By contrast, variant prevalence and frequency did not differ significantly among mid-elevation Apapane (*Himatione sanguinea*) with or without lesions ($P = 0.691$). The low frequency and the lack of detection of variants independent of GRW4 suggest that multiple independent introductions of *P. relictum* to Hawaii are unlikely. Multiple variants may have been introduced in heteroplasmy with GRW4 or exist within the tandem repeat structure of the mitochondrial genome. The discovery of multiple mitochondrial lineages of *P. relictum* in Hawaii provides a measure of genetic diversity within a geographically isolated population of this parasite and suggests the origins and evolution of parasite diversity may be more complicated than previously recognized.

Key words: mtDNA, avian malaria, Hawaii, *Plasmodium relictum*, next generation sequencing, Integroomer.

INTRODUCTION

Ecological studies of avian haemosporidian parasites based largely on PCR amplification and direct sequencing of short segments of parasite mitochondrial genes have grown exponentially in number over the past decade (Bensch *et al*. 2009). Single nucleotide substitutions have been used to identify unique lineages that are often used as discrete taxa for studies of geographical distribution, host specificity and parasite–vector associations. These studies have revealed an unprecedented amount of diversity among these organisms that is often difficult to reconcile with more traditional species concepts based on morphological and biological character-istics. As a rule of thumb, haemosporidians that differ by more than 5% in mitochondrial sequence appear to be distinguishable by morphological features of erythrocytic stages of the parasites (Hellgren *et al*. 2007; Valkiūnas *et al*. 2009), with most diversity captured by sequencing a 479 bp region of the cytochrome *b* gene (Hellgren *et al*. 2007). In reality though, we know little about how to define these parasites as biological species because so little is known about their basic life cycles, vectors and epizootiology. This problem is illustrated by the remarkable diversity of mtDNA lineages that have been associated with *Plasmodium relictum*, one of the most widely distributed species of avian *Plasmodium* both in terms of host range and geographical distribution (Beadell *et al*. 2006).

In an attempt to place isolates of *P. relictum* from the Hawaiian Islands into a global phylogenetic context, Beadell *et al*. (2006) identified 31 distinct lineages from around the world in two well-supported clades which included lineages identified as *P. relictum* by established morphological criteria. A single lineage of *Plasmodium* (lineage 15, previously identified as lineage GRW4, Bensch *et al*. 2000) was identified in 75 infected introduced and endemic forest birds from Hawaii, Maui, Oahu and Kauai. The authors concluded that a single lineage of the parasite had been introduced to the islands and that the recent expansion of native forest bird populations in some low-elevation habitats (Woodworth *et al*. 2005) had likely not been facilitated by cryptic introduction of different parasite lineages of lower virulence. However, their report relied on results of traditional direct sequencing, which fails to detect related lineages making up less than 5–25%

* Corresponding author. Department of Pharmaceutical Sciences, Daniel K. Inouye College of Pharmacy, University of Hawaii at Hilo, 34 Rainbow Drive, Hilo, HI 96720, USA. E-mail: jarvi@hawaii.edu
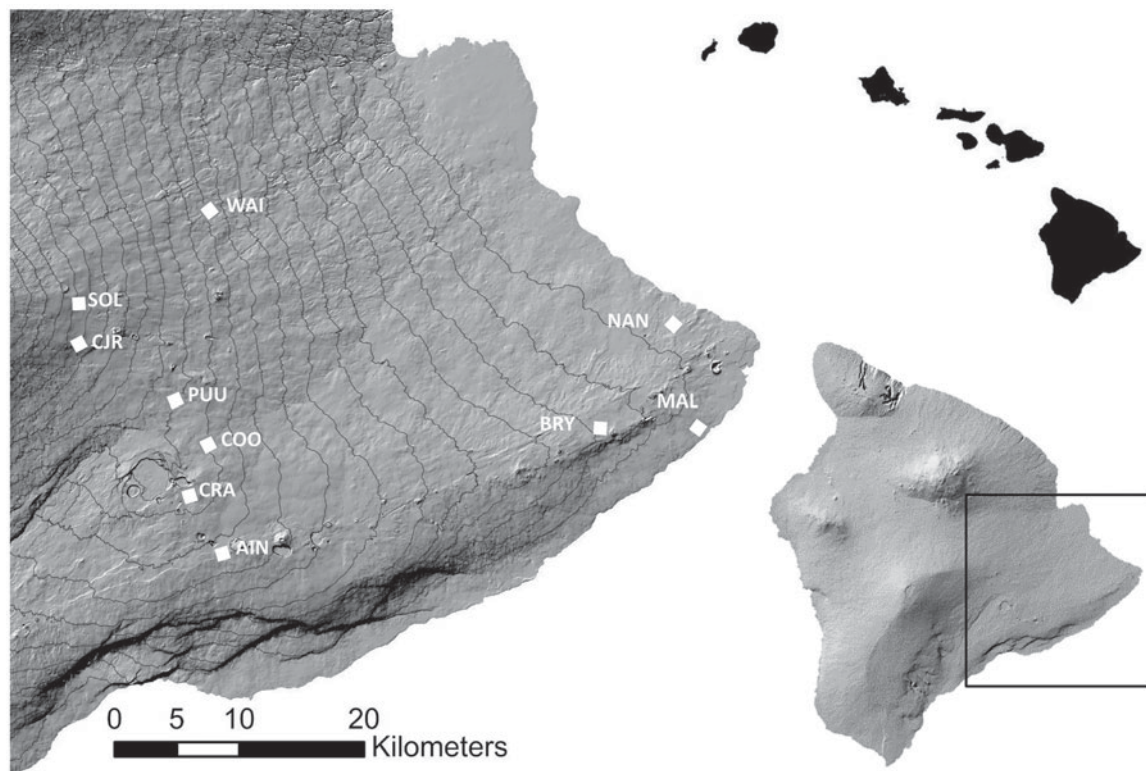
Fig. 1. Map of eastern Hawaii Island showing locations of 1 km^2 study sites. Solid lines indicate 250 m elevation contours. The high-elevation site (>1500 m above sea level) is CJ Ralph (CJR). The mid-elevation sites (1000 to 1300 m) are Ainahou Ranch (AIN), Crater Rim (CRA), Cooper's (COO), Puu Unit (PUU) and Waiakea (WAI). The low-elevation sites (<300 m) are Bryson's (BRY), Nanawale (NAN) and Malama Ki (MAL).

of genetically mixed infections (Kapoor *et al*. 2004; Hunt *et al*. 2005).

It has been suggested that traditional cloning and sequencing techniques be used when attempting to determine the presence of cryptic lineages in avian malarial infections (Perez-Tris and Bensch 2005). Such techniques are generally more efficient for the detection of minority variants in mixed infections than direct sequencing of PCR products (Lindner and Banik, 2009). Based on either cloning and sequencing or SNP analysis of partial nuclear *trap* (thrombospondin-related anonymous protein) genes of *P. relictum*, we have recently documented genetically diverse, mixed infections of *P. relictum* in the Hawaiian Islands (Jarvi *et al*. 2008; Farias *et al*. 2012). Based on these results, we hypothesized that mtDNA diversity in *P. relictum* in Hawaii may be higher than that detected by direct sequencing. This is of particular interest because co-evolutionary interactions among variants may affect both the virulence and transmission of the parasites across steep environmental gradients in the islands and within the diverse communities of native and non-native forest birds that occur along these gradients (Farias *et al*. 2012). In recent years, next-generation amplicon sequencing has developed into an approach that offers several advantages over both direct sequencing or cloning and sequencing. We therefore re-examined diversity of mitochondrial lineages

of *P. relictum* in Hawaii by using the Roche 454 platform for deep sequencing of a 346 bp fragment of the mitochondrial genome that encompasses a portion of the 479 bp region used by Hellgren *et al*. (2007) and other investigators. We illustrate here that while the predominant lineage appears to be GRW4, carefully confirmed rare lineages exist in Hawaii that are likely not due to multiple independent introductions.

MATERIALS AND METHODS

Sample collection and DNA extraction

Samples for this study were collected as part of a larger investigation of the ecology of avian malaria on the windward slopes of Mauna Loa and Kilauea Volcanoes (NSF DEB 0083944). Native and non-native forest birds were captured with mist nets on a monthly sampling schedule (minimum 2–4 days/month/site) at nine 1 km^2 study sites (Fig. 1) in 2002, 2003 and 2004 (Woodworth *et al*. 2005). Birds were measured, sexed and aged, and a blood sample was obtained by jugular venipuncture. Birds were examined for the presence or absence of *Avipoxvirus*-like lesions on the feet, legs, beak and around the eyes at the time of capture and infection status was recorded. Samples were screened for infection with *P. relictum* using a combination of serology and microscopy

as previously described (Woodworth *et al.* 2005). Genomic DNA was extracted from packed red blood cells stored in lysis buffer at −80 °C using a Qiagen DNeasy Tissue Extraction Kit following manufacturer's protocols. DNA concentrations were determined using a NanoDrop ND-1000 Spectrophotometer.

PCR amplification and 454 sequencing

We designed and optimized 454 fusion primers (Roche, 2009) for preparation of amplicon sequencing of the cytochrome *b* gene of *P. relictum*. A total of 14 forward and 14 reverse primers each incorporated one of 14 unique 10 nt-long multiplex identifier (MID) sequences. The cytochrome *b* forward primer basic sequence is: 5′-CGTATCGCCTC-CCTCGCGCCATCAG(MID)AAAGCACATTT-AATTCATTATCC-3′, and the reverse primer: 5′-CTATGCGCCTTGCCAGCCCGCTCAG(MID)-TAGCACCCCAGAAACTCAT-3′. The sequencing primer (A or B) is upstream of the MID, and the cytochrome *b* specific sequence is downstream of the MID. A unique combination of the forward and reverse primers was used for amplification from each individual to allow for sequence assignment. The target-specific portion of the reverse fusion primers fall within the region of cytochrome *b* analysed by Hellgren *et al.* (2007), so that the 3′ ends of the sequences reported here overlap the 5′ ends of sequences reported in the MalAvi database by 161 nucleotides. To produce tagged amplicons, 1–7 μL gDNA were used as template in a 25 μL reaction containing 1X reaction buffer, 1 mM MgCl$_2$, 200 μM each dNTP, 0·8 μM each forward and reverse fusion primers, and 1·25 units FastStart High-Fidelity Enzyme Blend (Roche). Successful amplicons were isolated on a 2% low-melt agarose gel (Fisher), excised, and purified using the NucleoSpin Extract II Kit (Machery-Nagel) according to manufacturer's protocols. Purified amplicons were quantified either by BioSpec-Nano (Shimadzu) or PicoGreen (Life Technologies) analysis with signal detection on a StepOne Plus Real Time PCR system (Life Technologies). Equimolar amounts of each amplicon were mixed for sequencing with a maximum possible of 196 uniquely tagged individuals per 1/4 sequencing run. Sequences were run at ASGPB, University of Hawaii at Manoa on a Roche Next Generation Sequencer GS FLX system.

Data management and analysis

Initial parsing of sequence data was completed using a new, user-friendly and publicly available portal called Integroomer. The Integroomer portal was developed specifically to handle the requirements of the current project but was designed and implemented with the purpose of flexibly demultiplexing and deconvoluting any DNA construct. The pipeline was designed with performance in mind and can easily adapt to run in a grid or cloud environment or be integrated into the Galaxy platform (Blankenberg *et al.* 2010). Integroomer was developed in Python using the Django framework (www.djangoproject.com). A free instance of the Integroomer pipeline can be used at http://courge.ics.hawaii.edu/integroomer. For this study, the Roche primers A and B were trimmed using the sffInfo script in the SFF Tools suite (Roche). The reads were then demultiplexed and parsed using Integroomer.

In Integroomer, a construct is defined as a series of hyphen-delimited labels representing the adapters and a specific region of interest, termed GENSEQ, to be extracted. The labels comprising a construct are either user- or keyword-defined. User-defined labels represent physical adapters whose sequences are fixed such as MIDs and target specific priming sequences, while keyword-defined labels represent subsequences of unknown length and/or value. The keyword GENSEQ denotes the region of interest, while the REGEX and UNKNOWN keywords are used to describe a class of subsequences or to simply refer to sequences of unknown length or value. As an example, all cytochrome *b* amplicon sequences originating from a single individual and amplified using forward MID01 and reverse MID01 would be parsed with the following construct:

KEY-MID_01_FOR-CYTB_A_FOR-GENSEQ-CYTB_B_REV-MID_01_REV-UNKNOWN where CYTB_A_FOR and CYTB_B_REV refer to the target-specific portion of the forward and reverse fusion primers given above, and the label KEY refers to the Roche sequencing key (TCAG). The postfix substring is encoded using the reserved label UNKNOWN, allowing for reads which continue beyond the reverse primer for a varying number of nucleotides. The output of running this construct in Integroomer is an .sff file containing all reads matching the specified adapters and criteria.

To allow for the highest level of flexibility in the parsing process, adapters are assigned attributes to describe how they can be matched to the sequencing reads. These attributes are: (1) the direction, used to determine whether the adapter should be matched in the forward or reverse/complement orientation; (2) a Boolean value describing whether the occurrence of the adapter is required; (3) minimum overlap length; (4) the maximum number of allowed errors; and (5) whether the adapter is required to align to the end of the sequence (blunt alignment), or whether the sequence is allowed to continue beyond the adapter. Given that users specify the stringency and minimum quality required for every hit, Integroomer automatically filters the reads whose lengths are inferior

to the total sum of the minimal required adapter overlap lengths.

Integroomer output files containing parsed sequences were converted to fastq files using the Python script SFF_extract (http://bioinf.comav.upv.es/) and quality filtered using PRINSEQ (Schmieder and Edwards, 2011) and Sequencher 5.0 (GeneCodes). Sequences containing errors in the priming sites, ambiguous bases, or indels were discarded, as were sequences with a mean Q score of less than 30 following trimming of primers and sequences in which more than 10% of nucleotides had a Q score of 20 or less. To determine the probability that the observed polymorphisms were due to sequencing artefacts, each nucleotide, i, was assigned a *phred* quality score, Q, which was used to infer its probability $P(i)$ of being erroneously called. The probability $P(i)$ was computed as follows (Ewing and Green, 1998): $P(i) = 10^{-Q/10}$. Considering that the *phred* quality scores of n bases belonging to the same column of a multiple sequence alignment are independent, we can infer $P(variant)$, the probability of variant being erroneous as:

$$P(variant) = \prod_{i=1}^{n} P(i) \qquad (1)$$

Data were analysed using Sequencher 5.0 for alignment, and Minitab® 16.2.1 for statistical analysis. Phylogenetic and molecular evolutionary analyses such as construction of a neighbour-joining tree, average nucleotide divergence and nucleotide content were conducted using *MEGA* version 4 (Tamura *et al.* 2007) with the gamma parameter estimated in *MEGA* version 5 (Tamura *et al.* 2011).

Non-random distribution of variants

Since *Avipoxviruses* are known to be immunosuppressive (Smith and Kotwal, 2002), we evaluated mean variant frequency (# variants/total number of hosts) as well as variant prevalence (# hosts with the variant/total # hosts) in Apapane and Amakihi with and without *Avipoxvirus*-like lesions and tested for mean differences with a two-sample *t*-test (Minitab® 16.2.1). If variants are real rather than artefactual, we hypothesized that infection with *Avipoxvirus* would lead to significant changes in variant frequency and prevalence. We classified lesions as *Avipoxvirus*-like based on morphological criteria used in previous studies to identify active lesions (van Riper *et al.* 2002).

RESULTS

Cytochrome *b* sequences from parasite populations from a total of 97 malaria-positive Hawaii Amakihi and 42 malaria-positive Apapane were evaluated in this study. Samples were collected across

an altitudinal gradient of 3 low-elevation (< 300 m), 3 mid-elevation (1200 m) and 1 high-elevation (1500 m) study sites (Fig. 1, Table 1). Because abundance of both native species varies with elevation, we were not able to obtain balanced sample sizes of each species at each elevational zone. Blood samples were collected from Hawaii Amakihi at 3 low-elevation and 3 mid-elevation study sites with most (84%, 81/97) of the samples originating from 3 low-elevation sites. Apapane samples were collected from 2 low-, 3 mid- and 1 high-elevation study sites with 90% (38/42) of the samples originating from mid-elevation sites. A total of 26 238 reads were obtained and after quality filtering a total of 16 636 sequences were further evaluated. Sequences that appeared only once (singlets) were removed, as were sequences from individuals that had <10 usable sequences/individual. Also removed from the final dataset were loci that did not meet the 95% criterion rule for defining polymorphic loci in populations described in Hartl and Clark (1997), thus only variants that were detected in ⩾7 individual hosts (⩾5% of the host population of 139) were included in this study.

Our final dataset consisted of 15 953 reads of which 15 742 were GRW4 (Table 1). A total of 211 variant sequences were further analysed. Sequences were imported into MEGA 4·0 and in a neighbour-joining tree (Kimura 2-parameter) 23 variant clusters each containing identical sequences originating from 8–15 individuals per cluster were delineated (not shown). The numbers of sequences found at each geographical site and in each species for the 23 variants and GRW4 are described in Table 1. All variants are single nucleotide transition substitutions.

For each variant, the quality scores of the bases belonging to its multiple sequence alignment were used in equation (1) (see methods) to compute the probability that the variant is the product of sequencing error (right column, Table 1). The extremely low probabilities obtained are clear evidence that the observed polymorphisms are highly likely not sequencing artefacts.

All of the 139 avian hosts possessed GRW4. A total of 49 hosts possessed only GRW4 (35%), GRW4 plus 1 additional variant were detected in 34 hosts (24%), 2 additional variants in 27 hosts (19%), 3 additional variants in 14 hosts (10%), 4 additional variants in nine hosts (6·5%), 5 additional variants in 5 hosts (3·6%), and 6 additional variants were detected in 1 host (0·7%). A total of 132 variant sequences were obtained from 121 Hawaii Amakihi and 79 variant sequences from 72 Apapane. The hypothesized amino acid sequences show that nucleotide substitutions in 9 of the 23 variants are non-synonymous resulting in amino acid changes, 11 are synonymous substitutions. Three substitutions resulted in stop codons and these sequences were eliminated from further analysis. A summary of nucleotide and amino acid substitutions in all variants as compared

Table 1. Total number of GRW04 and variant sequences detailed by species and geographic site of origin, and calculated probability that the variant sequences are due to sequencing error. Site name abbreviations are: BRY (Brysons), MAL (Malama Ki), NAN (Nanewale), PUU (Pu`u unit), AIN (Ainahou), CRA (Crater Rim), COO (Coopers), WAI (Waikea) and CJR (CJ Ralph)

Species	Amakihi								Apapane								Probability due to sequencing error
	Low elevation			Mid elevation			Totals		Low elevation		Mid elevation			High elevation	Totals		
Site n	BRY 27	MAL 21	NAN 33	PUU 1	AIN 3	CRA 12	# Reads	nv[a] 97	MAL 2	NAN 1	COO 25	CRA 1	WAI 12	CJR 1	# Reads	nv[a]	
GRW04	2445	2042	3701	179	498	819	9684	97	357	20	3736	564	1218	163	6058	42	
HIHOCR001	1		6			1	8	7			3		1	1	5	5	5·012E-56
HIHOCR002	2	3	3				8	7			5				5	5	7·943E-64
HIHOCR003		2	2				4	4		1	4	1			6	5	1·259E-50
HIHOCR004	2	3	1				6	6			4	1			5	4	1·995E-39
HIHOCR005	1		1		1	1	4	5			3			1	4	4	1·585E-40
HIHOCR006	3	1	3			1	8	8					1		1	1	6·310E-43
HIHOCR007	2	3	3				8	7			2				2	2	1·000E-47
HIHOCR008	1	1	2				4	4			4		2		6	5	5·012E-51
HIHOCR009	2	1	3				6	6					2		2	2	1·000E-32
HIHOCR010		4	1				5	4			3			1	4	4	1·000E-36
HIHOCR011	1	2	1			2	6	6	1		2		2		5	4	5·012E-39
HIHOCR013		1	4			1	6	6		1	1		1		3	3	1·585E-32
HIHOCR014			3		1	1	5	6			2	1			3	3	5·012E-39
HIHOCR015	1	3	2			1	7	6	1				1		2	2	3·981E-31
HIHOCR016		1				1	2	2	1		3	1	1		6	5	5·012E-34
HIHOCR017	1		6				7	6			1		1		2	2	7·943E-30
HIHOCR018	2	1	5		1		9	9					2		2	2	1·259E-34
HIHOCR019	1	1	2				4	4			2	1			3	3	1·000E-34
HIHOCR020	1	1	3				5	5			3				3	2	6·310E-31
HIHOCR022	1	1	1			1	4	4			2		2		4	3	1·995E-29
HIHOCR023	1		5				6	4			1		2		3	3	6·310E-51
HIHOCR026	2	3	1				6	6			1				1	1	1·000E-28
HIHOCR029			4		1		5	5			1		1		2	2	2·512E-35
Total usable sequences	2470	2074	3763	179	502	829	9817		360	22	3783	569	1235	168	6137		

[a] nv, Total number of individuals with variant.

Table 2. *Summary of nucleotide (nt) and amino acid (aa) substitutions in all variants as compared with GRW4*

Variant	N or S[a]	nt position	nt change from/to	aa position	aa change from/to
1	S	196	T/C		
2[b]	—	273	G/A		
3	S	4	T/C		
4	S	235	A/G		
5	N	158	A/G	53	S/G
6[b]	—	339	G/A		
7	N	35	G/A	12	G/R
8[b]	—	27	G/A		
9	N	197	T/C	66	S/P
10	S	49	G/A		
11	S	325	T/C		
13	N	281	G/A	94	G/R
14	S	271	A/G		
15	S	175	A/G		
16	S	316	T/C		
17	S	265	T/C		
18	S	169	T/C		
19	N	330	T/C	110	V/A
20	N	6	C/T	2	P/L
22	N	50	A/G	17	V/I
23	N	335	C/T	112	P/S
26	S	23	T/C		
29	N	106	C/T	35	P/L

[a] N designates non-synonymous substitutions, S designates synonymous substitutions.
[b] Variants resulting in stop codons.

with GRW4 is presented in Table 2. Average nucleotide divergence was estimated at 0.006 ± 0.001 SE (Kimura 2-parameter, Γ-corrected with the gamma parameter estimated in MEGA 5, standard error estimate obtained by bootstrap with 5000 replicates) and was based on the number of base substitutions per site ($n = 346$) over all sequence pairs ($n = 21$, i.e. 20 variants and GRW4). Total G–C content was 25·3%. All sequences are available in GenBank accession nos KF428723–KF428746, or as an alignment file upon request.

Variant frequency and variant prevalence comparisons were made among Apapane and Hawaii Amakihi with or without *Avipoxvirus*-like lesions. Variant frequency (#variants/#hosts) was higher ($P = 0.006$) in malaria-infected Amakihi with lesions ($n = 15$, mean = 2·07), than in Amakihi with no lesions ($n = 57$, mean = 0·88). However, no significant differences ($P = 0.691$) were detected between Apapane that were lesion positive ($n = 23$, mean 1·52) and those that were lesion negative ($n = 16$, mean 1·31). Similarly, variant prevalence (# of individual hosts harbouring a variant/the total number of hosts) in malaria-infected Amakihi with lesions (mean 0·1033) was significantly higher ($P = 0.001$) than in Amakihi with no lesions (mean 0·0456) (Fig. 2). However, no significant differences ($P = 0.497$) were detected between Apapane that were lesion positive (mean 0·0761) and those that were lesion negative (mean 0·0656) (data not shown).

DISCUSSION

We document previously undetected levels of variation of *P. relictum* in Hawaii based on next-generation sequencing of a 346 bp fragment of the cytochrome *b* gene from 97 Hawaii Amakihi and 42 Apapane. Until now, direct sequencing of *Plasmodium* mtDNA fragments (cytochrome *b*) from 75 Hawaiian forest birds identified only the most common lineage (GRW4) (Beadell *et al.* 2006). The application of new sequencing methodologies that do not require cloning and allow for increased depth of coverage provide a viable alternative and, as expected, allow for more thorough evaluation.

We detected 23 additional lineages based on single nucleotide transitional changes at 23 distinct sites. Single nucleotide polymorphisms (SNPs) appearing in only one or two individuals might be attributable to PCR or sequencing error, but each variant reported here was detected in 7–15 different individuals. These SNPs were not concentrated in any one region, nor did they occur in regions of high G–C or A–T content. For additional verification, we calculated the probabilities that the SNPs defining each variant were due to sequencing artefacts. The extremely low probabilities (in Table 1) provide clear evidence that these polymorphisms are highly unlikely to be due to sequencing error. Small differences introduced by singlet polymorphisms in the cytochrome *b* gene have been shown to have
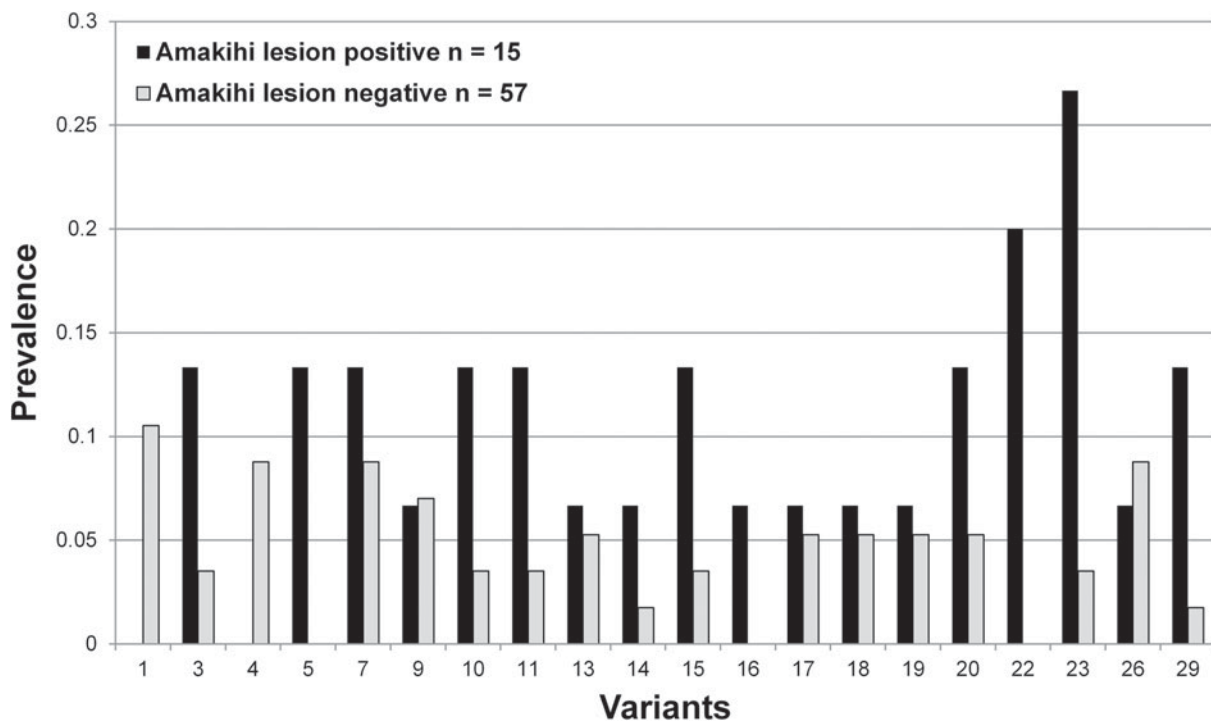
Fig. 2. Average variant prevalence in Amakihi with *Avipoxvirus*-like lesions appears greater than in Amakihi without lesions ($P = 0.001$). Prevalence was calculated as the number of individual hosts harbouring that variant/the total number of hosts.

significant biological consequences. For example, in *Plasmodium falciparum*, resistance to atovaquone has been associated with specific point mutations in the cytochrome *b* gene in the region spanning codons 271–284 (Korsinczky *et al*. 2000; Schwobel *et al*. 2003; Ekala *et al*. 2007). Similarly minor changes have been documented as multiple lineages that differ from GRW4 by one nucleotide (PADOM10, MYIALE02, PRBQ22, PRBQ21), two nucleotides (LINOL101, LINOL102) or five nucleotides (PRBQ20, Bensch *et al*. 2009). Within this framework, the sequences reported here may therefore be considered unique mitochondrial lineages.

We designed our primers to amplify a cytochrome *b* region that appeared more variable than the 479 bp fragment analysed by Hellgren *et al*. (2007). Based on previous cloning studies of a 1193 bp fragment from six *P. relictum* isolates from four native Hawaiian species on the islands of Kauai and Hawaii (54 sequences), a greater number of polymorphic sites was found outside of this region than within, especially upstream of the 5′ primer of this 479 bp fragment (Jarvi *et al*. unpublished results). The 3′ end (161 bp) of the 346 bp fragment we amplified overlaps with the 479 bp fragment. Thirteen of the 23 polymorphic sites were found within the overlap while we identified an additional 10 lineages, five of which resulted in non-synonymous substitutions that would have been missed by focusing on the 479 bp region alone.

The rare, distinct lineages reported here likely either evolved from the ancestral type (i.e. GRW4)

via mutation events or are the result of multiple introductions of parasites to the Hawaiian Islands. Multiple lineages may have been introduced to Hawaii either through introduction of multiple infected hosts or through introduction of a few individual hosts infected with multiple parasite lineages. In the case of multiple introductions, GRW4 clearly appears highly competitive. Based on currently available evidence, *P. relictum* has been in Hawaii for no more than 100 years (van Riper *et al*. 1986). Ricklefs and Outlaw (2010) report a rate for cytochrome *b* evolution in avian malaria parasites that corresponds to $\sim 1.2\%$ sequence divergence per million years. If one presumes a steady rate of evolution over time, one might then expect 0.00012% sequence divergence per 100 years. This is much lower than our estimate of average evolutionary divergence of 0.006 (0.001) or 0.6% among GRW4 and the 20 variants (excluding variants possessing stop codons, i.e. potential pseudogenes). If the variation observed in this study is presumed due to mutations of a single introduced lineage (i.e. GRW4), the observed level of divergence in this 346 bp fragment having occurred within this time period seems highly unlikely based on predicted divergence rates. It could be that the cytochrome *b* gene in *P. relictum* in this disease system is evolving more rapidly. This might be expected given the year-round transmission, the relatively high levels of parasitae-mias, and the lower selective pressure placed on the parasites by susceptible native Hawaiian forest birds. Also, co-infection with *Avipoxvirus* can result

in immunosuppression in avian hosts, thus potentially further reducing selective pressure on parasite populations.

We found higher mean variant frequency and prevalence in low-elevation Hawaii Amakihi co-infected with *Avipoxvirus*, which provides additional support that these variants are likely not artefacts and that they may have distinct biological characteristics. Poxviruses possess genes that especially modify the innate mechanisms of the host immune response, therefore sustaining viral persistence during early infections (Smith and Kotwal, 2002). A release of immunological regulation of *Plasmodium* parasites in chronic malaria infections due to *Avipoxvirus* co-infection may result in recrudescence of malaria with potential impacts on malaria diversity, which could provide an explanation for our finding of a higher mean variant prevalence in Hawaii Amakihi with *Avipoxvirus*-like lesions than in those without lesions. These data mirror earlier preliminary (unpublished) studies of *P. relictum* diversity in low-elevation Hawaii Amakihi and mid-elevation Apapane based on SNP analysis of the nuclear *trap* gene of *P. relictum*. The *trap* gene is a single copy gene in *Plasmodium* spp. (Robson *et al.* 1990), thus the detection of multiple *trap* sequences within a single host reflects multiple variants. We found the rate of multiple *trap* infection was much higher in malaria-infected Hawaii Amakihi with lesions (70%) as compared with malaria-infected Hawaii Amakihi without lesions (46%), while the rates of multiple *trap* infection were very similar in malaria-infected Apapane with lesions (39%) versus those infected with malaria only (35%). We know from experimental studies that birds exhibit premunition to *P. relictum*, i.e. exposure to one malarial isolate provides immunological protection against a second isolate (Atkinson *et al.* 2001). It is unlikely then, that chronically infected hosts would acquire new variants. Upon co-infection with *Avipoxviru*s, suppression of the host immune response may lead to recrudescence of existing infections, and/or possibly allow for acquisition of new variants through mosquito-borne transmission. If co-infection with *Avipoxvirus* results in increased malaria diversity through immunosuppression mechanisms, we would expect the effects to be most evident in populations or species exhibiting stronger immunity to malaria, i.e. more resistant population such as low-elevation Amakihi. Our preliminary data strongly suggest that this is the case, at least in low-elevation Hawaii Amakihi. We do not see the same effect upon co-infection of mid-elevation Apapane. Malaria and pox transmission levels vary between elevations, being year-round and endemic at low elevations and seasonal at mid-elevations (Atkinson and LaPointe, 2009). Malaria diversity based on *trap* is greater at low elevations than mid-elevations (Farias *et al.* 2012), which is consistent with expectations of greater parasite diversity in areas of higher transmission (Conway, 2007). Species differences may play a role in the increase in malaria diversity observed upon co-infection of low-elevation Hawaii Amakihi but not mid-elevation Apapane. However, in this comparison, species effects might be confounded by potential elevation-based transmission effects and cannot be determined at this time.

A total of 35% of hosts tested possessed only GRW4 and 65% of hosts possessed GRW4 along with at least one other variant, but no hosts harboured variants in the absence of GRW4. If multiple, independent introductions were responsible for the observed variation, one might expect to find variants independent of GRW4. To date, the mtDNA genomes of *P. yoelii*, *P. falciparum*, *P. vivax*, *P. gallinaceum* and *P. juxtanucleare* appear to be ~6 kb, and exhibit a linear, tandem, repeated structure (Vaidya and Arasu, 1987; Feagin, 1992; Sharma *et al.* 1998; Omori *et al.* 2007). Given the tandemly repeated structure of the *Plasmodium* mitochondrial genome and the finding of three variants with stop codons, it is possible that some of these GRW4 concatenated repeats might contain cytochrome *b* variants. Another possibility is that these variants exist with GRW4 in heteroplasmy. Lastly, it is also possible that these variants are nuclear mtDNA (NUMTs) copies, however this is least likely since *Plasmodium* have low numbers of NUMTs (Richly and Leister, 2004) and only one mitochondrion per merozoite during early stages of the asexual erythrocytic cycle (Divo *et al.* 1985; Hopkins *et al.* 1999). These scenarios all offer viable explanations for the lack of detection of variants independent of GRW4.

While the source of variation has yet to be documented, multiple, independent introductions seem unlikely. Instead, we hypothesize that these variants, whether they exist within the concatenated repeat structure of the *Plasmodium* mitochondrial genome or exist in heteroplasmy, were either introduced to Hawaii with GRW4 or arose due to an exceptionally rapid mutation rate potentially occurring during replication of the mtDNA genome. Whole genome sequencing of these variants might further our understanding of the nature of mitochondrial diversity. A comparison of next generation sequencing of parasites in ancient versus modern-day populations as well as examination of the geographical range and diversity of these lineages across the length of the archipelago with continental populations may further our understanding of the origin and evolution of this parasite in the Hawaiian Islands.

REFERENCES

Atkinson, C. T. and Lapointe, D. A. (2009). Ecology and pathogenicity of avian malaria and pox. In *Conservation Biology of Hawaiian Forest Birds: Implications for Island Avifauna* (ed. Pratt, T. K., Atkinson, C. T., Banko P. C., Jacobi, J. D. and Woodworth, B. L.), pp. 234–252. Yale University Press, New Haven, CT, USA.

Atkinson, C. T., Dusek, R. J. and Lease, J. K. (2001). Serological responses and immunity to superinfection with avian malaria in experimentally-infected Hawai`i Amakihi. *Journal of Wildlife Diseases* **37**, 20–27.

Beadell, J. S., Ishtiaq, F., Covas, R., Melo, M., Warren, B. H., Atkinson, C. T., Bensch, S., Graves, G. R., Jhala, Y. V., Peirce, M. A., Rahmani, A. R., Fonseca, D. M. and Fleischer, R. C. (2006). Global phylogeographic limits of Hawai`i's avian malaria. *Proceedings of the Royal Society of London Series B* **273**, 2935–2944.

Bensch, S., Stjernman, M., Hasselquist, D., Ostman, O., Hansson, B., Westerdahl, H. and Torres-Pinheiro, R. (2000). Host specificity in avian blood parasites: a study of *Plasmodium* and *Haemoproteus* mitochondrial DNA amplified from birds. *Proceedings of the Royal Society of London Series B* **267**, 1583–1589.

Bensch, S., Hellgren, O. and Pérez-Tris, J. (2009). MalAvi: a public database of malaria parasites and related haemosporidians in avian hosts based on mitochondrial cytochrome b lineages. *Molecular Ecology Resources* **9**, 1353–1358.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology* **19.10**, 1–21.

Conway, D. J. (2007). Molecular epidemiology of malaria. *Clinical Microbiology Reviews* **20**, 188–204.

Divo, A. A., Geary, T. G., Jenson, J. B. and Ginsburg, H. (1985). The mitochondrion of *Plasmodium falciparum* visualized by rhodamine 123 fluorescence. *Journal of Protozoology* **32**, 442–446.

Ekala, M., Khim, N., Legrand, E., Randrianarivelojosia, M., Jambou, R., Fandeur, T., Menard, D., Assi, A., Henry, M., Rogier, C., Bouchier, C. and Mercereau-Puijalon, O. (2007). Sequence analysis of *Plasmodium falciparum* cytochrome b in multiple geographic sites. *Malaria Journal* **6**, 164. doi:10.1186/1475-2875-6-164.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research* **8**, 186–194. PMID 9521922.

Farias, M. E. M., Atkinson, C. T., LaPointe, D. A. and Jarvi, S. I. (2012). Analysis of the *trap* gene provides evidence for the role of elevation and vector abundance in the genetic diversity of *Plasmodium relictum* in Hawai`i. *Malaria Journal* **11**, 305.

Feagin, J. E. (1992). The 6 kb element of *Plasmodium falciparum* encodes mitochondrial genes. *Molecular and Biochemical Parasitology* **52**, 145–148.

Hartl, D. L. and Clark, A. G. (1997). *Principles of Population Genetics*, 3rd Edn. Sinauer Associates, Sunderland, MA, USA.

Hellgren, O., Krizanauskiene, A., Valkiūnas, G. and Bensch, S. (2007). Diversity and phylogeny of mitochondrial cytochrome b lineages from six morphospecies of avian *Haemoproteus* (Haemosporida: Haemoprotidae). *Journal of Parasitology* **93**, 889–896.

Hopkins, J., Fowler, R., Krischna, S., Wison, I., Mitchell, G. and Bannister, L. (1999). The plastid in *Plasmodium falciparum* asexual blood stages: a three-dimensional ultrastructural analysis. *Protist* **150**, 283–295.

Hunt, P., Fawcett, R., Carter, R. and Walliker, D. (2005). Estimating SNP proportions in populations of malaria parasites by sequencing: validation and applications. *Molecular and Biochemical Parasitology* **143**, 173–182.

Jarvi, S. I., Farias, M. E. M. and Atkinson, C. T. (2008). Genetic characterization of Hawaiian isolates of *Plasmodium relictum* reveals mixed-genotype infections. *Biology Direct* **3**, 25.

Kapoor, A., Jones, M., Shafer, R. W., Rhee, S. Y., Kazanjian, P. and Delwart, E. L. (2004). Sequencing-based detection of low-frequency human immunodeficiency virus type 1 drug-resistance mutants by an RNA/DNA heteroduplex generator-tracking assay. *Journal of Virology* **78**, 7112–7123.

Korsinczky, M., Chen, N., Kotecka, B., Saul, A., Rieckmann, K. and Cheng, Q. (2000). Mutations in *Plasmodium falciparum* cytochrome b that are associated with atovaquone resistance are located at a putative drug-binding site. *Antimicrobial Agents and Chemotherapy* **44**, 2100–2108.

Lindner, D. L. and Banik, M. T. (2009). Effects of cloning and root-tip size on observations of fungal ITS sequences from *Picea glauca* roots. *Mycologia* **101**, 157–165.

Omori, S., Sata, Y., Isobe, T., Yukawa, M. and Murata, K. (2007). Complete nucleotide sequences of the mitochondrial genomes of two avian malaria protozoa, *Plasmodium gallinaceum* and *Plasmodium juxtanucleare*. *Parasitology Research* **100**, 661–664.

Perez-Tris, J. and Bensch, S. (2005). Diagnosing genetically diverse avian malaria infections using mixed-sequence analysis and TA-cloning. *Parasitology* **131**, 15–23.

Richly, E. and Leister, D. (2004). NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21**, 1081–1084.

Ricklefs, R. E. and Outlaw, D. C. (2010). A molecular clock for malaria parasites. *Science* **329**, 226.

Robson, K. J. H., Hall, J. R. S., Davies, L. C., Crisanti, A., Hill, A. V. S. and Wellems, T. E. (1990). Polymorphism of the TRAP gene of *Plasmodium falciparum*. *Proceedings of the Royal Society B: Biological Science* **242**, 205–216.

Roche Diagnostics GmbH (2009). Amplicon fusion primer design guidelines for GS FLX Titanium. *Series Lib-A Chemistry* TCB No. 013-2009.

Schmieder, R. and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864.

Schwobel, B., Alifrangis, M., Salanti, A. and Jelinek, T. (2003). Different mutation patterns of atovaquone resistance to *Plasmodium falciparum in vitro* and *in vivo*: rapid detection of codon 268 polymorphisms in the cytochrome b as potential *in vivo* resistance marker. *Malaria Journal* **2**, 5.

Sharma, I., Pasha, P. T. and Sharma, Y. D. (1998). Complete nucleotide sequence of the *Plasmodium vivax* 6 kb element. *Molecular and Biochemical Parasitology* **97**, 259–263.

Smith, S. A. and Kotwal, G. J. (2002). Immune response to poxvirus infections in various animals. *Critical Reviews in Microbiology* **28**, 149–185.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.

Vaidya, A. B. and Arasu, P. (1987). Tandemly arranged gene clusters of malarial parasites that are highly conserved and transcribed. *Molecular and Biochemical Parasitology* **22**, 249–257.

Valkiūnas, G., Iezhova, T. A., Loiseau, C., Smith, T. B. and Sehgal, R. N. M. (2009). New malaria parasites of the subgenus *Novyella* in African rainforest birds, with remarks on their high prevalence, classification and diagnostics. *Parasitology Research* **104**, 1061–1077.

van Riper, C., III, van Riper, S. G., Goff, M. L. and Laird, M. (1986). The epizootiology and ecological significance of malaria in Hawaiian land birds. *Ecological Monographs* **56**, 327–344.

van Riper, C., III, van Riper, S. G. and Hansen, W. R. (2002). Epizootiology and effect of avian pox on Hawaiian forest birds. *Auk* **119**, 929–942.

Woodworth, B., Atkinson, C. T., LaPointe, D. A., Hart, P. J., Spiegel, C. S., Tweed, E. J., Henneman, C., LeBrun, J., Denette, T., DeMots, R., Kozar, K. L., Triglia, D., Lease, D., Gregor, A., Smith, T. and Duffy, D. (2005). Host population persistence in the face of introduced vector-borne diseases: Hawai`i Amakihi and avian malaria. *Proceedings of the National Academy of Sciences USA* **102**, 1531–1536.