

Lineage fusion in Galápagos giant tortoises

RYAN C. GARRICK,* EDGAR BENAVIDES,† MICHAEL A. RUSSELLO,‡ CHAZ HYSENI,*† DANIELLE L. EDWARDS,† JAMES P. GIBBS,§ WASHINGTON TAPIA,¶** CLAUDIO CIOFI†† and ADALGISA CACCONE†

*Department of Biology, University of Mississippi, Oxford, MS 38677, USA, †Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA, ‡Department of Biology, University of British Columbia, Okanagan Campus, Kelowna, BC V1V 1V7, Canada, §College of Environmental Science & Forestry, State University of New York, Syracuse, NY 13210, USA, ¶Department of Applied Research, Galápagos National Park Service, Puerto Ayora, Isla Santa Cruz, Galápagos, Ecuador, **Biodiver S.A. Consultores, Km 5 Vía a Baltra, Isla Santa Cruz, Galápagos, Ecuador, ††Department of Evolutionary Biology, University of Florence, Via Romana 17, 50125 Florence, Italy

Abstract

Although many classic radiations on islands are thought to be the result of repeated lineage splitting, the role of past fusion is rarely known because during these events, purebreds are rapidly replaced by a swarm of admixed individuals. Here, we capture lineage fusion in action in a Galápagos giant tortoise species, *Chelonoidis becki*, from Wolf Volcano (Isabela Island). The long generation time of Galápagos tortoises and dense sampling (841 individuals) of genetic and demographic data were integral in detecting and characterizing this phenomenon. In *C. becki*, we identified two genetically distinct, morphologically cryptic lineages. Historical reconstructions show that they colonized Wolf Volcano from Santiago Island in two temporally separated events, the first estimated to have occurred ~199 000 years ago. Following arrival of the second wave of colonists, both lineages coexisted for approximately ~53 000 years. Within that time, they began fusing back together, as microsatellite data reveal widespread introgressive hybridization. Interestingly, greater mate selectivity seems to be exhibited by purebred females of one of the lineages. Forward-in-time simulations predict rapid extinction of the early arriving lineage. This study provides a rare example of reticulate evolution in action and underscores the power of population genetics for understanding the past, present and future consequences of evolutionary phenomena associated with lineage fusion.

Keywords: biodiversity, conservation, evolutionary history, Galápagos, introgressive hybridization

Received 30 May 2014; revision received 4 September 2014; accepted 10 September 2014

Introduction

Evolutionary radiations on islands have served as cornerstones of research on speciation. In several classic study systems with linearly arranged and sequentially aged islands, the progression-rule hypothesis – stepwise colonization from old to young islands – has received some empirical support (e.g. Galápagos: Parent *et al.* 2008; Benavides *et al.* 2009; Steinfartz *et al.* 2009; Hawaii:

Fleischer *et al.* 1998; Cowie & Holland 2008). Under this model, colonization of newly available habitats via rare long-distance dispersal initiates allopatric divergence of lineages. In these cases, regardless of whether stochastic processes such as drift (Wessel *et al.* 2013), repeatable processes such as natural selection (Losos *et al.* 1998) or a combination of both (Templeton 2008) promotes differentiation among diverging lineages, it is common for most members of a clade to be single-island endemics (Cowie & Holland 2008; Parent *et al.* 2008).

While the role of allopatric divergence is undisputed in island radiations, the notion that the number of

Correspondence: Ryan C. Garrick, Fax: +1 662 915 5144; E-mail: rgarrick@olemiss.edu

different lineages tends to accumulate over time at a rate determined by the counterbalancing forces of speciation and extinction is contentious (Wiens 2011). This may be partly due to observational bias. For example, in the case of young lineages, fusion may also be a dominant force, particularly when mechanisms driving reproductive isolation are weakly formed (Seehausen *et al.* 2008), as often occurs during rapid radiations. Yet, the importance of fusion events in evolutionary radiations is likely underestimated because incipient lineages tend to fuse so rapidly that the underlying processes are seldom caught in the act, and so empirical evidence appears sparse (Fitzpatrick *et al.* 2009).

Recent studies have provided empirical support for speciation in reverse (Behm *et al.* 2010; Webb *et al.* 2011; Vonlanthen *et al.* 2012), a process driven by a breakdown in extrinsic reproductive barrier(s) (Seehausen 2006) and often mediated by environmental change (Seehausen *et al.* 2008). Here, we expand upon these findings by showing that similar processes are occurring in Galápagos giant tortoises (*Chelonoidis* spp.), a group with life history traits (particularly long generation time) that are conducive to detecting fusion events as they are occurring.

From both an evolutionary and conservation perspective, *Chelonoidis becki* – the giant tortoise species endemic to Wolf Volcano, Isabela Island – is of considerable interest. Previous work has revealed two genetically distinct yet morphologically cryptic lineages, for which genetic substructuring generally corresponds with northern vs. western slopes of the volcano (Fig. 1). Furthermore, Santiago Island tortoises (*Chelonoidis darwini*) were revealed to be the likely source of Wolf Volcano colonists (original founding estimated at <280 000 years ago; Caccone *et al.* 2002; Ciofi *et al.* 2002; Russello *et al.* 2007; Poulakakis *et al.* 2008, 2012). However, until now, there was insufficient sampling and data to determine whether establishment of the two lineages involved a single vs. two temporally separate colonization events. The ramifications of these alternatives are significant: the former scenario implies *in situ* differentiation and thus a direct role for ecologically divergent selection, whereas the latter scenario requires only the operation of genetic drift (or by chance, genetically different founders), such that present-day allele frequency differences may be quite labile and reversible (Nosil *et al.* 2009).

Here, we use data from intensive sampling and screening of nuclear microsatellite and mitochondrial DNA (mtDNA) sequence variation (Garrick *et al.* 2012; Edwards *et al.* 2013) to evaluate the degree of differentiation between the two genetically distinct lineages of *C. becki*, and to reconstruct their colonization history, which is revealed to be most consistent with

two temporally separate colonization events. Next, we present evidence for recent and widespread introgressive hybridization between them and provide insights into the dynamics of this process. Specifically, we show that purebred females of the two lineages differ in their tendency to produce F₁ hybrid offspring, suggesting inherent differences in mate selectivity. Finally, we use forward-in-time simulations to anticipate the fate of these two lineages and show that one of them is likely to soon go extinct due to gene pool swamping.

Methods

We used previously generated data on genetic variation at 12 nuclear microsatellite loci and a 705-bp fragment of the mtDNA control region. Briefly, blood was sampled as in Caccone *et al.* (1999) and Ciofi *et al.* (2002), screening of microsatellite allele size polymorphisms followed Benavides *et al.* (2012) and sequencing of mtDNA used the methods of Russello *et al.* (2007). A recent re-examination of the genetic composition of *Chelonoidis becki* tortoises (Garrick *et al.* 2012; Edwards *et al.* 2013) was based on microsatellite genotypes and mtDNA haplotypes assayed from ~1700 individuals sampled within the geographic range of *C. becki* (i.e. from Wolf Volcano, Isabela Island), analysed together with corresponding data from an archipelago-wide reference data set ($N = 354$ individuals) that included representatives of all extant and most extinct Galápagos giant tortoise species (Russello *et al.* 2007, 2010; Poulakakis *et al.* 2008, 2012). Collectively, these studies provided the following: (i) strong support for 12 distinct genetic clusters within the archipelago-wide reference database, (ii) confirmation of the existence of two distinct groups of *C. becki* [herein referred to as the Piedras Blancas (PBL) lineage predominantly from the northern slope of Wolf Volcano and the Puerto Bravo (PBR) lineage from the western slope; Fig. 1] and (iii) identification of 841 tortoises that had ancestry only in the PBL and/or PBR clusters (see Appendix S1, Supporting information) – these individuals are the focus of the present study.

Classification of *Chelonoidis becki* individuals

Admixture between PBL and PBR lineages was apparent from previous analyses (Garrick *et al.* 2012; Edwards *et al.* 2013). To infer the types and frequencies of hybridization events that had occurred in the wild, we assigned the focal 841 tortoises to one of the following five classes using an approach based on membership coefficients (Q -values) estimated in STRUCTURE v2.3.3 (Pritchard *et al.* 2000): (i) PBR purebreds, (ii) PBL purebreds, (iii) F₁

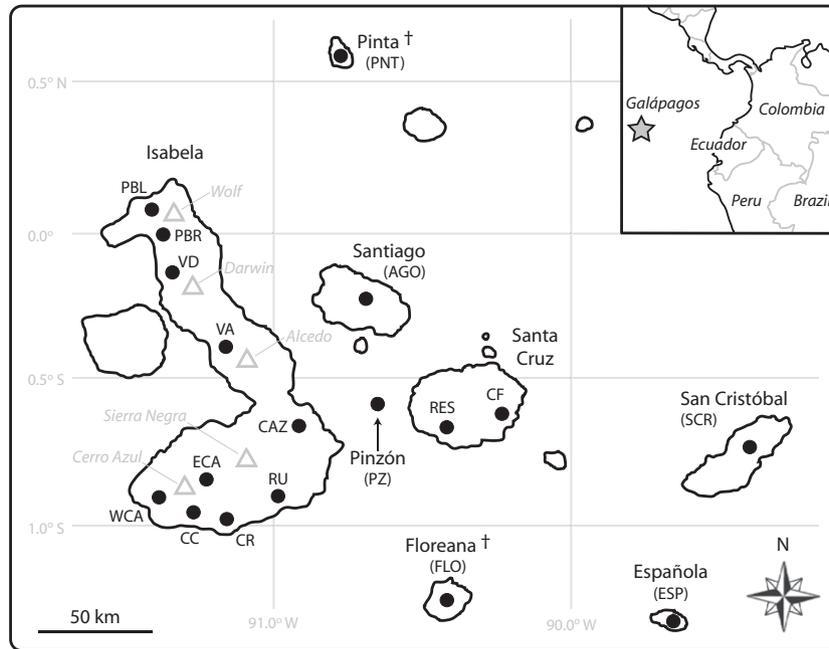


Fig. 1 Map of Galápagos Archipelago showing major islands. On Isabela Island, five volcanos are labelled and indicated by grey triangles. Solid black circles and associated abbreviations mark the locations where *Chelonoidis becki* lineages predominantly occur [i.e. Piedras Blancas (PBL) and Puerto Bravo (PBR)], as well as the locations of the samples representing most extinct (†) and all extant Galápagos giant tortoise species included in the archipelago-wide reference database [*C. microphyes*: Darwin Volcano (VD); *C. vandenburghi*: Alcedo Volcano (VA); *C. guntheri*: Cazuela (CAZ), Roca Unión (RU) and Cabo Rosa (CR); *C. vicina*: Eastern Cerro Azul (ECA) and Western Cerro Azul (WCA); *Chelonoidis darwini*: Santiago (AGO); *C. ephippium*: Pinzón (PZ); *C. porteri*: La Reserva (RES); *C. sp. nov.*: Cerro Fatal (CF); *C. hoodensis*: Española (ESP); and *C. chathamensis*: San Cristóbal (SCR)]. *Inset*: Location of Galápagos (star) relative to continental South America.

hybrids, (iv) PBR \times F₁ backcrosses or (v) PBL \times F₁ backcrosses. Run settings were as follows: correlated allele frequency and admixture models, 1×10^5 burn-in and 5×10^5 MCMC iterations with $K = 12$ as a fixed parameter, with five replicate runs combined using CLUMPP v1.1.2b (Jakobsson & Rosenberg 2007).

To establish Q -value thresholds that account for the stochasticity associated with genotypic recombination, we used HYBRIDLAB v1.0 (Nielsen *et al.* 2006) to simulate crosses within and between purebred lineages and backcrosses. These *in silico* data sets (100 individuals \times 12 loci) had similar characteristics to the empirical genetic data and were analysed in STRUCTURE together with the archipelago-wide reference data set (run settings as above) to generate Q -value distributions for the five classes of tortoises. Following Garrick *et al.* (2012), we assigned each of the 841 *C. becki* individuals to one of these classes by jointly considering two measures that were established from HYBRIDLAB simulations: Q_R , the expected range of Q -values within each parental cluster (i.e. PBL and PBR) and Q_D , the expected difference in Q -values between the two parental clusters (Table S1; Appendix S1, Supporting information).

Genetic diversity and differentiation

Mitochondrial genetic polymorphism among PBR purebreds, PBL purebreds, and among F₁ and backcross tortoises, was quantified using DNA sequence summary statistics [i.e. number of haplotypes (N_{hap}), segregating sites (S) and haplotypic diversity (H_d)] calculated in DNASP v5.10 (Librado & Rozas 2009). For the same groups, microsatellite polymorphism summary statistics [i.e. rarefaction-corrected allelic richness (A_R) based on subsamples of 72 diploid individuals, and expected heterozygosity (H_E)] were calculated using HP-RARE v1.0 (Kalinowski 2005). For microsatellites, multilocus averages were used to summarize these diversity metrics. We also assessed levels of genetic differentiation between PBR and PBL purebreds using two metrics that, when compared to one another, can potentially distinguish whether divergences registered by microsatellites occurred over relatively short vs. long timescales [i.e. F_{ST} vs. R_{ST} calculated in GENEPOP v4.0 (Rousset 2008) and $R_{ST}CALC$ v2.2 (Goodman 1997), respectively]. Additionally, we calculated divergences based on mtDNA sequence data using two alternative measures [i.e. uncorrected p -distance and maximum-likelihood-

corrected distances calculated in PAUP* v4.0b10 (Swoford 2002)]. To evaluate whether levels of differentiation between the two *C. becki* lineages are similar to those seen between other Galápagos giant tortoise taxa, divergence estimates were compared to the distribution of values calculated between pairs of recognized species, plus one cryptic species (see Appendix S1, Supporting information).

Mitochondrial DNA haplogroups

Evolutionary relationships among mtDNA haplotypes carried by purebred and hybrid *C. becki* tortoises were estimated using statistical parsimony (Templeton *et al.* 1992) with a 95% confidence criterion, implemented in TCS v1.21 (Clement *et al.* 2000). Haplotypes were first compared to those in the archipelago-wide reference database to identify mtDNA sequences that were 'native' (i.e. found exclusively in *C. becki* from Wolf Volcano) vs. 'non-native' (i.e. derived from another species). Next, focusing only on *C. becki* mtDNA sequences, haplotypes that occurred at high frequency in purebreds of one lineage (e.g. PBR purebreds) but were *not* found in the other group (i.e. PBL purebreds, and vice versa) were identified. These haplotypes were then grouped with close relatives in the parsimony network to establish 'haplogroups'.

Wolf Volcano colonization history

Previous work has shown that the Galápagos giant tortoise species endemic to Santiago Island (*Chelonoidis darwini*) is the sister taxon of *C. becki* and the likely source of colonization of Wolf Volcano (see Introduction). However, divergence between the two *C. becki* lineages could be explained by three competing scenarios, which generate different predictions with respect to population tree topology: (1) a single colonization of Wolf Volcano, followed by the formation of two lineages that diverged *in situ* (Fig. 2a,b) or (2) two temporally separate colonizations of Wolf Volcano, with the first colonists on an independent evolutionary trajectory prior to arrival of the second ones. This *ex situ* divergence hypothesis can be further subdivided into two scenarios in which (2a) the PBL lineage arrived first (Fig. 2c,d) or (2b) the PBR lineage did (Fig. 2e,f). To statistically distinguish among these alternatives, we used approximate Bayesian computation (ABC; Beaumont *et al.* 2002), implemented in DIYABC v2.0.4 (Cornuet *et al.* 2014). This method uses summary statistics to characterize an empirical data set and then calculates the same summary statistics from numerous data sets simulated under each of the three alternative colonization scenarios, which are specified via user-defined priors on

model parameters. The posterior probability of each scenario, given the empirical data, is then used to identify the best-fit model. Finally, point estimates (and confidence intervals) for parameters included in the best-fit model are calculated from a subset of the simulated data sets that were used for the preceding model selection step (Bertorelle *et al.* 2010).

Data sets. ABC analyses focused on microsatellite and mtDNA sequence data from purebred individuals. To assess robustness of inferences while also achieving computational tractability, we ran ABC on three subsets of the available data. Each was comprised of 50 multilocus microsatellite genotypes and 50 mtDNA sequences randomly drawn from a larger pool of purebred PBL individuals [where necessary, 'non-native' PBL haplotypes (Fig. S1, Supporting information) were discarded and replaced by another random draw]. The same subsampling was applied to PBR purebreds. Together, this represents the equivalent of 100 *C. becki* individuals with no missing data. A total of 35 purebred *C. darwini* individuals, previously genotyped by Russello *et al.* (2007), were also included. Owing to some missing data, this gene pool was represented by 25 multilocus microsatellite genotypes and 21 mtDNA sequences. Consequently, subsampling was not applied in the case of *C. darwini*.

Colonization model parameters and priors. All competing scenarios had the same three sampled populations (i.e. *C. becki*'s PBR and PBL lineages from Wolf Volcano, Isabela Island, and *C. darwini* from Santiago Island), with broad uniform priors applied to contemporary and ancestral effective population sizes (i.e. $N_e = 10$ to 10 000 individuals). All competing scenarios also had two lineage splitting times: t_1 represented either *in situ* divergence between the PBR and PBL lineages (scenario 1) or the most recent colonization of Wolf Volcano (scenarios 2a, b), whereas t_2 represented the earliest colonization of Wolf Volcano (Fig. 2). Recently, Poulakakis *et al.* (2012) estimated the age of the *becki*+*darwini* stem clade as ~400 000 years and timing of divergence between *C. becki* and *C. darwini* as ~280 000 years ago. We conservatively used 400 000 years ago as an upper bound on the t_2 uniform prior and set the lower bound at 125 000 years ago, which approximately corresponds with the penultimate glacial period. The t_1 uniform prior (12 500 years ago to 288 000 years ago) spanned the same overall duration (i.e. 275 000 years) and included considerable temporal overlap with t_2 , and so a conditional prior ($t_1 < t_2$) was used. In all cases, we assumed a 25-year generation time for Galápagos giant tortoises (Throp 1975).

Although the absence of strong bottlenecks associated with lineage divergence is plausible given that past sea

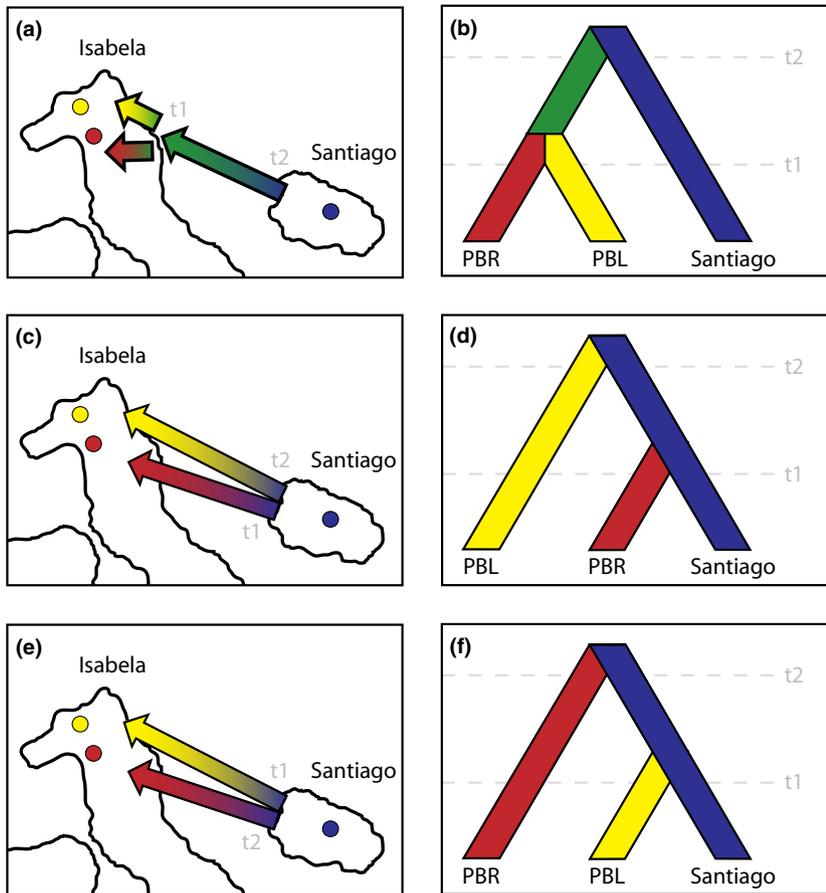


Fig. 2 Historical divergence between the Puerto Bravo (PBR) and Piedras Blancas (PBL) lineages of *Chelonoidis becki* assessed using approximate Bayesian computation. The tortoise species from Santiago Island (*Chelonoidis darwini*) has previously been identified as the most likely source of founders. Three alternative scenarios are shown together with their respective tree topologies and relative splitting times (t_1 and t_2 represent youngest vs. oldest splits, respectively): (1) a single colonization of Wolf Volcano followed by *in situ* divergence of the two *C. becki* lineages (scenario 1, panels a–b); (2) two temporally separate colonizations of Wolf Volcano with the order of arrival being PBL followed by PBR (scenario 2a, panels c–d) or PBR followed by PBL (scenario 2b, panels e–f).

level fluctuations may have repeatedly created temporary land bridges between the two focal Galápagos Islands over the past ~700 000 years (Ali & Aitchison 2014), we also considered the possibility that bottleneck events were associated with long-distance overwater colonization. To do this, we ran the same three data subsets (described above) in models with or without bottlenecks. As the goal was to add biological realism rather than to make inferences about the existence or nature of the bottlenecks themselves, only a narrow set of conditions relating to bottleneck severity (uniform prior: $N_e = 2$ –20 individuals) and duration (uniform prior: 1–5 generations) were considered. As before, given some overlap between prebottleneck vs. bottleneck N_e , conditional priors were used to ensure that bottlenecks always caused a population size reduction.

Mutation model parameters and priors. Microsatellite data were simulated using a generalized stepwise-mutation model. This assumes that alleles evolve via replication slippage, where each new mutation increases or decreases allele length by a number of repeated motifs drawn from a geometric distribution (Estoup *et al.* 2002). Microsatellite mutation rates used here (uniform

prior, $\mu = 1.5 \times 10^{-4}$ to 1.5×10^{-3}) are consistent with values previously reported for herpetofauna and long-lived vertebrates (Zhang & Hewitt 2003). The parameter of the geometric distribution (uniform prior, $P = 0.05$ to 0.15) allowed for rare multistep mutations. MtDNA sequence data were simulated using a Tamura & Nei (1993) mutation model, with values for the proportion of invariant sites (pinvar = 0.7113) and gamma shape ($\gamma = 0.8937$) determined using MODELTEST v3.0 (Posada & Crandall 1998) and set as fixed parameters. Mutation rates used here (uniform prior, $\mu = 2 \times 10^{-7}$ to 2×10^{-6}) encompass a previous point estimate for Galápagos tortoise mtDNA control region (i.e. 8.5×10^{-7} , Beheregaray *et al.* 2004; assuming 25-year generation time, Throp 1975).

Summary statistics. From the empirical and simulated microsatellite data, the following summary statistics were calculated: mean number of alleles across loci (single sample and two samples), mean gene diversity across loci (Nei 1987; single sample and two samples), F_{ST} between two samples (Weir & Cockerham 1984) and $(\delta\mu)^2$ distance between two samples (Goldstein *et al.* 1995). For mtDNA sequences, summary statistics

were as follows: number of distinct haplotypes (single sample), number of segregating sites (single sample), mean pairwise difference (single sample), Tajima's (1989) D (single sample), mean of between sample pairwise differences (two sample) and F_{ST} between two samples (Hudson *et al.* 1992). Given three populations in each scenario, this set generated 36 summary statistics.

Scenario choice, error assessment and parameter estimates. For each DIYABC run, 1.5×10^6 multilocus genetic data sets were simulated. Principal component analysis was used for preliminary assessment of whether the three colonization scenarios (and their priors) generated data sets with broadly similar characteristics to the empirical data. To select the best-fit scenario, posterior probabilities (PP) were computed via logistic regression (Fagundes *et al.* 2007; Table S2, Supporting information) on the 1% of simulated data sets closest to the empirical data. We also calculated type I error (proportion of times the chosen scenario does *not* have the highest PP when it is true) and type II error (proportion of times the chosen scenario has the highest PP when it is *not* true; Table S2, Supporting information). Finally, we estimated the posterior distribution of values of each free parameter in the best-fit model using the 1% of simulated data sets closest to the empirical data and the 'logit' transformation. For each free parameter, the median was taken as the point estimate, and the confidence interval (CI) was bounded by 5% and 95% quantiles (Table S2, Supporting information).

Spatial distributions of purebred and admixed individuals

To examine whether geographic separation between the two *C. becki* lineages exists, we characterized the extent of spatial clustering of PBR and PBL purebreds and their hybrids. To do this, bivariate kernel density estimation was used to identify areas on Wolf Volcano with the highest probability of containing each of the five classes of *C. becki* tortoises. This is a nonparametric technique in which a known density function (the kernel) is averaged across the observed data points (GPS coordinates of individuals) to create a smooth approximation. Kernel density was estimated using the `kde2d` function of the MASS package (Venables & Ripley 2002) in R (R Development Core Team 2012), and a normal distribution was used to choose bandwidth or neighbourhood size for the approximation. The spatial projection used a Gaussian kernel with a grid of 1200 (longitude) \times 1440 (latitude) points.

Hybridization dynamics and forward-in-time simulations

The existence of distinct mtDNA haplogroups specifically associated with the PBR vs. PBL lineages (see Results; Fig. S1, Supporting information) permitted investigation of whether hybridization among them is occurring at random. Random hybridization predicts bidirectional introgression of mtDNA, assuming equal fitness of individuals from all types of crosses. However, the magnitude of introgression in each direction also depends on sex ratios, which may not be equal in one or both parental groups. To first establish the expectations of the null hypothesis of random mating, we used field-based sex identifications of individuals (Garrick *et al.* 2012), together with genetic assignments, to estimate sex ratios for different classes of *C. becki* tortoises. Based on empirical frequencies of purebred males and females, we calculated the proportion of F_1 hybrids expected to carry PBR mtDNA haplogroup sequences when the null hypothesis of random hybridization is true (given only two lineages, this focus on the PBR mtDNA haplogroup was arbitrary; Table S3, Supporting information). Expected vs. observed values were then compared using χ^2 tests, with significance assessed at the 0.05 level (one-tailed test). The same approach was used to assess the null hypothesis of random mating between purebred \times F_1 individuals (i.e. χ^2 tests of expected vs. observed frequencies of backcrosses carrying PBR mtDNA haplogroup sequences). Together, these comparisons help characterize hybridization dynamics over the past two generations.

To explore the consequences of continued introgressive hybridization among *C. becki* tortoises, characteristics of the present generation of purebred plus admixed individuals (G_0) vs. outcomes after a single generation of random mating (G_1) were compared. For these analyses, we used a subset of individuals ($N = 502$) for which information on sex, mtDNA sequence and microsatellite genotypes was available. Observed frequencies of tortoises with each combination of characteristics (i.e. male vs. female, PBL vs. PBR mtDNA haplogroup and purebred vs. hybrid microsatellite-based classification) were used to calculate the probability of all possible mate pairings (Table S4, Supporting information) and thus to obtain estimates of projected changes in frequencies of mtDNA haplogroup sequences, and purebred tortoises, on Wolf Volcano.

We simulated crosses using HYBRIDLAB to model stochasticity associated with the parental pairs that may form during a single cycle of reproduction and random segregation of microsatellite alleles during gamete formation. To seed the simulations, we selected random subsets of 200 and 400 adults to represent the current

generation's breeders. Given that Wolf Volcano's current census population size (N_c) is estimated at ~8000 tortoises and approximately 50% of those sampled to date have ancestry only with *C. becki* (the remaining 50% are either admixed with other species or unclassified; Garrick *et al.* 2012), these values correspond to a $N_e : N_c$ ratio of 0.05 and 0.10, respectively. This is consistent with empirical data for numerous wild species (Frankham 1995). Next, random male–female crosses generated 200 offspring representing the next generation. To compare the trajectory and magnitude of changes in levels of genetic diversity between G_0 and G_1 gene pools, observed heterozygosity (H_O) and A_R were quantified for each. To assess the extent of decay of pre-existing genetic structure, linkage disequilibrium (LD) among alleles at microsatellite loci was also compared across the two generations. These statistics were calculated using GENEPOP and HP-RARE.

Results

Classification of *Chelonoidis becki* individuals

Using previously established criteria (Garrick *et al.* 2012), the majority (81%) of the 841 *C. becki* tortoises were unambiguously assigned to one of the five classes. Classification of the remaining individuals was carried out by determining which of the two (or more) alternative classifications minimized the disparity between empirical Q -values and the mode of the distribution of simulated Q -values (i.e. the highest probability solution). Of the 841 *C. becki* tortoises, 220 were sampled from locations on the northern slope of Wolf Volcano and the other 621 from locations on the western slope. Within the set of 279 individuals classified as purebreds, PBR purebreds were more abundant than PBL purebreds (71% vs. 29%, respectively). Furthermore, across the total sample, admixed individuals as a group (i.e. F_1 hybrids plus both backcrosses) were more common than both types of purebreds combined (67% vs. 33%, respectively; Table 1, Fig. 3a).

Genetic diversity and differentiation

Based on summary statistics standardized to account for differences in sample size (i.e. A_R and H_E for microsatellites and H_d for mtDNA sequences), genetic polymorphism among PBL purebreds was consistently greater than that among PBR purebreds (Table 1). Across the total sample, admixed individuals as a group show the highest polymorphism at microsatellite loci. Measures of mtDNA polymorphism were also highest in admixed individuals, but with one exception that is consistent with asymmetrical mtDNA introgression: H_d sequentially

decreased in the following order: purebred PBL > F_1 hybrids > backcrosses (Table 1).

Microsatellite data indicated that PBL and PBR lineages exhibit species-level genetic differences (Fig. 4). Their pairwise F_{ST} value matched the modal F_{ST} value in the distribution generated by pairwise comparisons of recognized Galápagos giant tortoise species. Similarly, the PBR vs. PBL R_{ST} value fell well within the distribution that represents species-level differences, albeit slightly closer to the left tail (Fig. 4, left). Both metrics of mtDNA sequence-based differentiation (i.e. uncorrected p -distances and maximum-likelihood-corrected distances), however, provided a less compelling case for species-level differences (Fig. 4, right).

Mitochondrial DNA haplogroups

We identified three haplogroups within the focal purebred, hybrid and backcross *C. becki* individuals: one characteristic of the PBR lineage and two characteristic of the PBL lineage. As a group, PBR purebreds carried two closely related mtDNA haplotypes (labelled R-2 and R-4); both are considered native to *C. becki* (Fig. S1, Supporting information). In contrast, PBL purebreds had a larger array of haplotypes (L-2 to L-5), including two divergent mtDNA haplotypes that appear 'non-native' as they are mostly found in a tortoise species (*C. vandenberghi*) endemic to Alcedo Volcano, Isabela Island. Considering only the native *C. becki* haplotypes, PBL and PBR purebreds were genetically distinct as their mtDNA sequences formed reciprocally monophyletic clades. However, the level of divergence (i.e. differences of ≥ 6 mutations) was not unusual compared to values obtained from the same genetic marker within other *Chelonoidis* species (i.e. ~5 mutational differences are typically seen at the intraspecific level). While additional haplotypes occurred in F_1 hybrids and backcross individuals, the clear phylogenetic separation between PBR and PBL mtDNA haplogroups was maintained (Fig. S1, Supporting information).

Wolf Volcano colonization history

ABC analyses identified the scenario with two temporally separate colonization events, in which PBL was the first lineage to arrive in Wolf Volcano (i.e. *ex situ* scenario 2a, Methods; Fig. 2c,d), as the best-fit model (PP = 0.87; competing scenarios 1 and 2b: PP = 0.02 and 0.11, respectively). Confidence in scenario choice was acceptable (i.e. type I and type II error ~0.13 each), and analyses of replicate data subsets produced very similar results (Table S2, Supporting information). Parameter estimates for the best-fit scenario are reported here as averages across runs that did not include bottlenecks, but point estimates and

Table 1 Genetic diversity at 12 microsatellite loci and the mitochondrial DNA (mtDNA) control region in *Chelonoidis becki* tortoises classified as purebreds, F₁ hybrids and backcrosses. Abbreviations are as follows: *N*, number of individuals (not all have mtDNA sequence data); *A_R*, allelic richness; *H_E*, expected heterozygosity; *Nhap_{native}*, number of haplotypes that are found almost exclusively in *C. becki*; *Nhap_{non-native}*, number of haplotypes that appear to be derived from *C. vandenberghi*; *H_d*, haplotypic diversity; and *S*, number of segregating sites

Classification	Microsatellites			Mitochondrial DNA				
	<i>N</i>	<i>A_R</i>	<i>H_E</i>	<i>N</i>	<i>N hap_{native}</i>	<i>N hap_{non-native}</i>	<i>H_d</i> *	<i>S</i> *
Purebreds								
PBL	80	8.60	0.76	72	2	2	0.040	1
PBR	199	6.82	0.67	194	2	0	0.010	1
Hybrids								
F ₁	117	11.26	0.80	113	3	2	0.036	5
Backcrosses								
PBL backcross	152	11.95	0.81	147	3	5	0.283	8
PBR backcross	293	11.68	0.78	288	4	1	0.063	6

*Calculations were based on native *C. becki* haplotypes only (see Fig. S1, Supporting information).

their confidence intervals remained largely unchanged when the additional complexity of brief, but severe bottlenecks associated with long-distance overwater dispersal was incorporated (Figs S2 and S3, Supporting information, respectively). The first set of colonists from Santiago Island are estimated to have arrived ~199 000 years ago (90% CI = 136 000–349 000 years ago), followed by a second colonization event estimated at ~53 000 years ago (90% CI = 22 000–151 000 years ago). These results suggest that although the PBL lineage has independently evolved from *Chelonoidis darwini* for an estimated ~146 000 years longer than the PBR lineage, both have co-inhabited Wolf Volcano for a substantial amount of time.

Spatial distributions of purebred and admixed individuals

Based on presence/absence distribution data alone, the PBL and PBR lineages are sympatric. However, when considering their local abundances using bivariate kernel density estimation, distributions more closely resemble parapatry. Purebred PBL individuals were primarily sampled from the northern slope of Wolf Volcano, whereas PBR purebreds were most commonly sampled from the western slope (Fig. 3b,c). While F₁ hybrids and PBR × F₁ backcross individuals were concentrated on the volcano's western slope (89 of 117 and 254 of 293, respectively), the PBL × F₁ backcross individuals were more evenly distributed (66 northern vs. 86 western; Fig. 3d–f).

Hybridization dynamics and forward-in-time simulations

Empirical data on sex ratios showed that even under the null hypothesis of random mating between PBL and

PBR purebreds, mtDNA introgression between these two *C. becki* lineages is expected to be asymmetrical (Table S3, Supporting information). The PBL purebreds (*N* = 65 adults of known sex) deviated significantly from a 1:1 sex ratio ($\chi^2 = 12.94$ d.f. = 1, *P* < 0.001) due to a strong male bias ($\delta:\text{♀} = 2.61$), while PBR purebreds (*N* = 169 adults) showed no significant departure from an equal sex ratio ($\delta:\text{♀} = 0.88$; $\chi^2 = 0.72$ d.f. = 1, *P* = 0.397). Accounting for this sex ratio difference, empirical data based on genetic classifications showed that mate pairings between PBL and PBR purebreds were significantly different from random ($\chi^2 = 30.98$ d.f. = 1, *P* < 0.001), due to an overrepresentation of the PBR mtDNA haplogroup in F₁ hybrids. However, this was not true for the next generation of hybridization (i.e. the point at which introgression occurs), as we failed to reject the null hypothesis that backcrossed individuals were generated by random mating between F₁ hybrids and either purebred PBL ($\chi^2 = 1.88$ d.f. = 1, *P* = 0.170) or purebred PBR ($\chi^2 = 2.42$ d.f. = 1, *P* = 0.120) individuals.

Projected changes in mtDNA frequencies after a single generation of random mating [i.e. from the current (*G*₀) to future (*G*₁) generation] indicated an increase in frequency of the PBR haplogroup (0.88 to 0.93). However, this was coupled with a decrease in the proportion of PBR mtDNA haplogroup sequences carried by PBR purebreds (0.33 to 0.11). From *G*₀ to *G*₁, the PBL mtDNA haplogroup was predicted to decrease in overall frequency (0.12 to 0.08) and also in occurrence within purebred PBL tortoises (0.08 to 0.01; Fig. S4, Supporting information). Similar results were obtained when considering tortoise classifications based on microsatellite genotypic data alone, as the frequency of PBR purebreds was predicted to decrease by ~66% (0.33

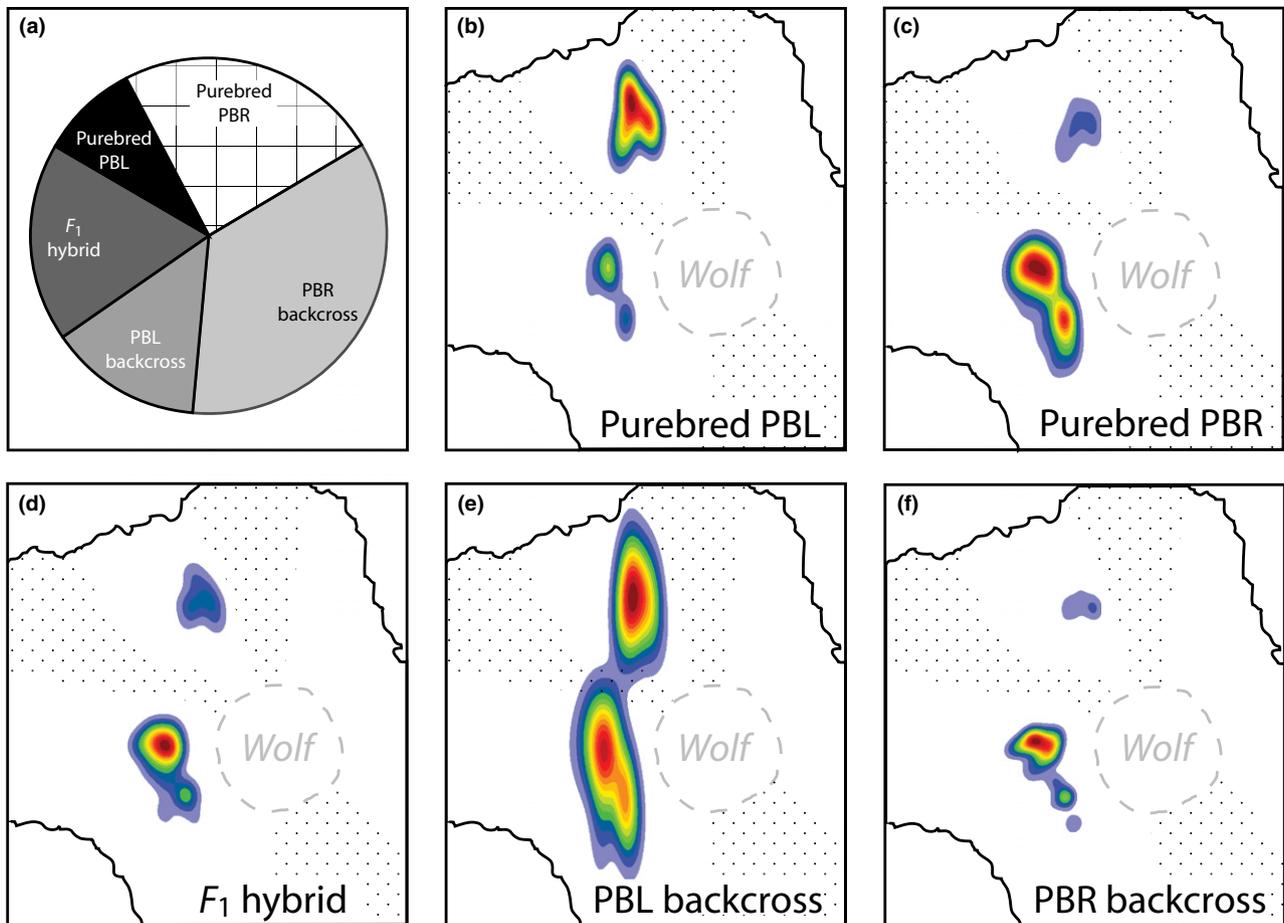


Fig. 3 Abundance and spatial distribution of *Chelonoidis becki* tortoises classified as purebreds, F_1 hybrids and backcrosses. Panel a: Relative abundance of five classes of individuals ($N = 841$). Panels b–f: Geographic areas with the highest probability of containing members of each of the five classes, determined via bivariate kernel density estimation, are indicated by hot (red) colours, whereas low probability of occurrence is indicated by cool (blue) colours. The area surrounded by a broken grey line is the location of the caldera on the summit of Wolf Volcano, and three zones of radially distributed volcanic vents are marked by dotted sections (modified from Geist *et al.* 2005).

to 0.11), and purebred PBL tortoises approached local extinction (0.08 to 0.01; ~94% reduction). Conversely, the frequency of hybrid tortoises was projected to increase by ~150% (0.59 to 0.89) after just a single generation of random mating. Within this group, F_1 hybrids and first-generation backcrosses decreased in frequency, whereas ~80% of individuals were projected to be F_2 or third-generation hybrids (e.g. double backcrosses; Fig. S5, Supporting information).

Forward-in-time simulated crosses showed that the gene pool of *C. becki*'s emerging hybrid swarm is expected to differ from the current generation's gene pool in several ways. First, we found that LD at microsatellite loci is likely to decay considerably (but not completely), independent of N_e (Fig. S6, Supporting information). Second, some metrics of within-group genetic diversity were projected to show a future

increase, whereas others exhibited little or no change. For example, A_R is expected to increase (median gain of 1.8 or 2.8 alleles per locus, depending on N_e), whereas H_O is likely to show little or no increase, irrespective of N_e (Figs S7 and S8, Supporting information, respectively).

Discussion

Although lineage fusion is rarely caught in the act, its consequences can be far-reaching. These events can prevent erosion of evolutionary potential by rapidly enhancing genetic and morphological variation (Grant & Grant 2014), but can also have negative impacts on individual fitness and lineage integrity. Thus, it is important to understand the historical factors that cause such events and the phenomena that drive subsequent

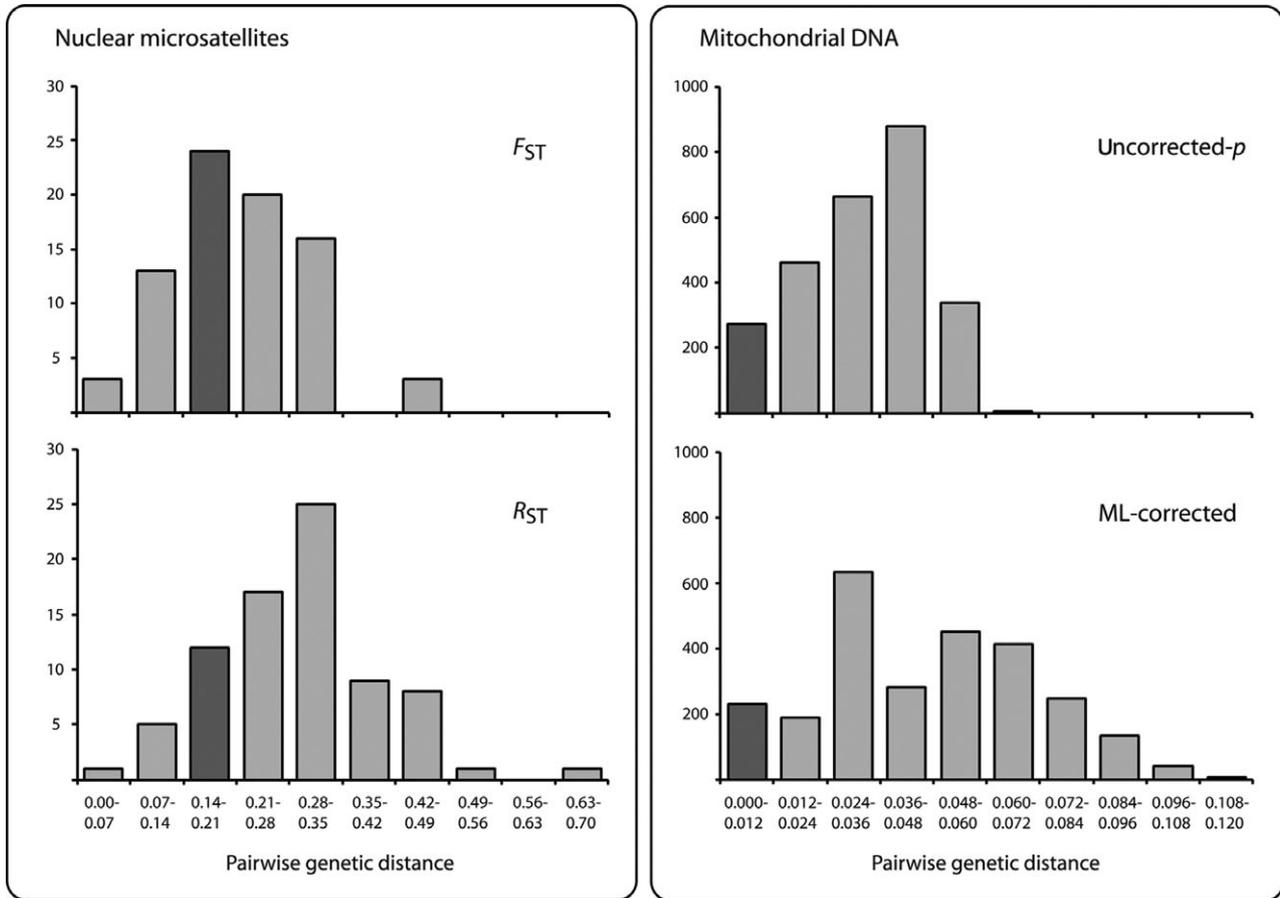


Fig. 4 Frequency distribution of genetic distances between different Galápagos giant tortoise species. On each histogram, a dark grey column indicates where the observed genetic distance between the two lineages of *Chelonoidis becki* falls. *Left*: Microsatellite genetic distances calculated from purebred individuals in the reference database (13 taxa, $N = 79$ pairwise interspecific comparisons) measured using F_{ST} (top) or R_{ST} (below). *Right*: DNA sequence genetic distances based on mtDNA haplotypes from purebred individuals in the reference database (78 species-specific haplotypes, $N = 2619$ pairwise interspecific comparisons), measured using uncorrected p -distances (top), or maximum-likelihood (ML)-corrected distances (TrN+I+G model; below).

genetic revolutions. Below, we examine the role of lineage fusion on islands and consider its taxonomic implications for recently formed species. We then hypothesize mechanisms underlying the fusion of *Chelonoidis becki*'s PBL and PBR lineages on Wolf Volcano and close by evaluating projected outcomes of continued introgressive hybridization.

Importance of lineage fusion in evolution

Fusion events provide an important counterpoint to the traditional view that radiations on islands are characterized by repeated allopatric and/or ecological divergence of lineages that maintain their integrity on secondary contact or in the face of ongoing gene flow (Schluter 2000). Fusion usually occurs when pre- and/or post-zygotic isolating mechanisms are absent or weak, which is typical of recently diverged lineages (Seehausen *et al.*

2008). Analytical advances are providing insights into the prevalence of lineage fusion (Gerard *et al.* 2011), allowing investigation of the processes that initiate or immediately follow from these events. As a result, recent divergence followed by widespread introgressive hybridization has now been documented for many groups, from invertebrates (Vergilino *et al.* 2011; Talavera *et al.* 2013) to vertebrates (Hendry *et al.* 2006; Webb *et al.* 2011; Cui *et al.* 2013). Collectively, these studies suggest that fusion might be operating more broadly than previously thought. In the Galápagos, two classic exemplars of rapid radiations [i.e. Darwin's finches (Grant & Grant 1992, 2008, 2014; Grant *et al.* 2005; Hendry *et al.* 2006; Kleindorfer *et al.* 2014) and giant tortoises (this study)] also show evidence for reticulate evolutionary histories. Indeed, in fluctuating environments such as those of isolated island archipelagos, fusion is likely a recurring process (Grant & Grant 2014).

Taxonomic implications of fusion in Galápagos giant tortoises

Here, we consider the idea that under certain circumstances, *intraspecific* hybridization has the potential to yield insights into the stability of higher-level species boundaries. Briefly, most named Galápagos giant tortoises species show differences consistent with phylogenetic species status (Cracraft 1983), as previously reported (Caccone *et al.* 1999, 2002; Ciofi *et al.* 2002; Beheregaray *et al.* 2004; Russello *et al.* 2007, 2010; Poulakakis *et al.* 2008, 2012; Chiari *et al.* 2009; Garrick *et al.* 2012; Edwards *et al.* 2013). Yet, many are single-island endemics and as such have largely precluded *in situ* tests of reproductive compatibility as a criterion under the biological species concept (Mayr 1942). A case in point is *Chelonoidis darwini* – the lineage recognized as a distinct species endemic to Santiago Island and the progenitor of the two *C. becki* lineages on Wolf Volcano. Our data show that *C. becki*'s PBL and PBR lineages fail to meet the requirements of the biological species concept, as they are fusing together and producing viable, fertile hybrids. Moreover, this fusion is taking place after an estimated ~146 000–199 000 years of separation, whereas the most recently arrived members of *C. becki* on Wolf Volcano (i.e. the PBR lineage) are estimated to have diverged from *C. darwini* only ~53 000 years ago. Thus, assuming divergence times are positively correlated with the levels of intrinsic reproductive isolation (e.g. Singhal & Moritz 2013), our inferences about intraspecific hybridization in *C. becki* illustrate how the species-level status of *C. darwini* could be scrutinized, if the biological (cf. phylogenetic) species concept was adopted.

Causes of lineage fusion in Galápagos giant tortoises

Much of our understanding about the short-term dynamics of hybridization comes from human-mediated events (Hendry *et al.* 2006; Seehausen 2006; Seehausen *et al.* 2008; Fitzpatrick *et al.* 2009). However, given that humans arrived in the Galápagos only recently (c. 1600s, Townsend 1925) relative to giant tortoise generation time (~25 years, Throp 1975), the *C. becki* lineage fusion likely represents a natural event. Our data are consistent with the notion that recently diverged lineages within evolutionary radiations on islands are prone to fusion and that the resulting hybrids may incur no fitness costs (Carson & Templeton 1984; Grant *et al.* 2005). The inferred phylogeographic scenario for *C. becki* includes a relatively recent (~199 000 years ago) estimated time of initial separation between the PBL and PBR lineages, which both independently diverged from the same *C. darwini* ancestor on Santiago Island (Figs S2 and S3,

Supporting information). Furthermore, given that the majority (445/841) of individuals assigned to one of the five identifiable classes were backcrosses (Table 1; Fig. 3a), F₁ hybrids are clearly fertile and likely carry little or no disadvantage. Interestingly, our analysis of hybridization dynamics showed that PBL females have a stronger propensity to avoid mating with purebred males of the alternative *C. becki* lineage than do PBR females, suggesting that differences in mate choice may be operating. Conversely, PBL and PBR females show no detectable avoidance of mating with F₁ males. Thus, hybrids and backcross individuals may have a particularly broad range of potential mates because they are less recognizable as being members of one or the other distinct lineage of *C. becki*. Furthermore, they will also be most commonly encountered owing to their high abundance, whereas purebred mates will become increasingly scarce (Grant & Grant 2014). Due to this swamping, the integrity of the PBL lineage is unlikely to be maintained over the long term.

We suggest that, as for Darwin's finches (Grant *et al.* 2005), hybridization among Galápagos giant tortoises has been a recurrent feature of their adaptive radiation. The presence of 'non-native' mtDNA haplotypes in *C. becki*'s PBL lineage that appear to be derived from *C. vandenberghi* native to Alcedo Volcano, Isabela Island (Fig. S1, Supporting information), generally suggests that species boundaries in the group may be somewhat porous. Furthermore, the spatial structuring of the PBL and PBR purebreds (Fig. 3b,c) indicates that while geographic isolation probably contributed to past gene flow limitation, in the absence of a physical barrier to dispersal, hybridization can become prolific. We hypothesize that recent changes in lava flows altered connectivity between vegetated areas on Wolf Volcano, driving the two *C. becki* lineages into secondary contact. Historically, radially distributed volcanic vents in three zones (north, northwest and southeast of the summit) produced frequent high-volume lava flows, and these generated rough top surfaces that are difficult or impossible for tortoises to negotiate (Geist *et al.* 2005). However, recent lava flows (<500–1800 years ago) have been limited to southern vents only (Geist *et al.* 2005), creating new opportunities for genetic mixing between PBL and PBR lineages. Other factors may also have inhibited free gene flow. These include strong philopatry to nesting areas as seen in other tortoises (Paquette *et al.* 2010), ecological differences between local habitats and/or the impacts of periodic historical sea level fluctuations (Ali & Aitchison 2014). Although our data do not suggest a major role for human-mediated disturbances on lineage fusion, the relative importance of natural vs. anthropogenic impacts does require further resolution.

Potential outcomes of continued hybridization

Our short-term projections showed that a single generation of random mating among adult *C. becki* tortoises could lead to purebred PBL individuals becoming extinct from the wild. Notably, the integrity of the PBR lineage is also expected to be threatened, owing to the formation of a hybrid swarm comprised mostly of F₂ and double backcross individuals (Fig. S5, Supporting information). When focusing on mtDNA, simulations indicated that the PBL haplogroup would decrease in frequency, whereas the PBR haplogroup could eventually sweep to fixation (Fig. S4, Supporting information). However, given that our projections are simplistic (i.e. assume random mating) and that the consequences of interbreeding beyond the first or second future generations are difficult to predict because mating patterns are complex and fitness outcomes are not known, these projections are tentative. Nonetheless, the rate of lineage fusion is expected to be rapid. For example, long-term empirical data on introgressive hybridization between two species of Darwin's finches revealed dramatic morphological and genetic convergence over only 30 years (i.e. six generations), with complete fusion expected in 43 years (Grant & Grant 2014).

One often overlooked benefit of fusion between lineages is the rapid acquisition of new genetic diversity by one or both lineages (Grant & Grant 2008). Hybrid individuals typically have increased heterozygosity, and some empirical studies have demonstrated strong heterozygosity–fitness correlations (Coltman *et al.* 1999). In the case of *C. becki*, simulations indicated that heterozygosity (H_D) would remain largely unchanged (Fig. S8, Supporting information). This can be explained by heterozygosity already being high. Consequently, novel genotypes generated by hybridization will tend to replace existing (but redundant) heterozygous genotypes. Thus, genetic rescue is an unlikely outcome of the present fusion event. Furthermore, given the projected rapid decay of LD (Fig. S6, Supporting information), if lineage-specific co-adapted mito-nuclear gene complexes (or nuclear epistatic interactions) currently exist, they may quickly break down. Nonetheless, we reiterate that lineage fusion is not necessarily always detrimental, as it has the potential to counteract the effects of inbreeding in small populations and may also rapidly enhance genetic and morphological variation, thereby increasing long-term viability (Grant & Grant 2014). Understanding the circumstances under which positive vs. negative ramifications of lineage fusion are most likely to emerge represents a challenging but important area of future research.

Acknowledgements

The Galápagos National Park Service and Charles Darwin Research Station were instrumental in supporting sample collection. L.B. Beheregaray, K.B. Dion, S. Glaberman, M.C. Milinkovitch, N. Poulakakis and J.R. Powell helped with sample and/or data collection, and D.J. Geist provided valuable discussions about the geological history of Wolf Volcano. Financial and other support came from the Galápagos National Park Service, Bay and Paul Foundation, Eppley Foundation, Galápagos Conservancy, National Geographic Society, Turtle Conservation Fund and Yale Institute for Biospheric Studies. This manuscript was improved by constructive comments from Brent Emerson and two anonymous reviewers.

References

- Ali JR, Aitchison JC (2014) Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. *Journal of Biogeography*, **41**, 1227–1241.
- Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.
- Beheregaray LB, Gibbs JP, Havill N *et al.* (2004) Giant tortoises are not so slow: rapid diversification and biogeographic consensus in the Galápagos. *Proceedings of the National Academy of Sciences USA*, **101**, 6514–6519.
- Behm JE, Ives AR, Boughman JW (2010) Breakdown in post-mating isolation and the collapse of a species pair through hybridization. *American Naturalist*, **175**, 11–26.
- Benavides E, Baum R, Snell HM, Snell HL, Sites JW Jr (2009) Island biogeography of Galápagos lava lizards (Tropiduridae: *Microlophus*): Species diversity and colonization of the archipelago. *Evolution*, **63**, 1606–1626.
- Benavides E, Russello M, Boyer D *et al.* (2012) Lineage identification and genealogical relationships among captive Galápagos tortoises. *Zoo Biology*, **31**, 107–120.
- Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.
- Caccone A, Gibbs JP, Ketmaier V, Suatoni E, Powell JR (1999) Origin and evolutionary relationships of giant Galápagos tortoises. *Proceedings of the National Academy of Sciences USA*, **96**, 13223–13228.
- Caccone A, Gentile G, Gibbs JP *et al.* (2002) Phylogeography and history of giant Galápagos tortoises. *Evolution*, **56**, 2052–2066.
- Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics*, **15**, 97–131.
- Chiari Y, Hyseni C, Fritts TH *et al.* (2009) Morphometrics parallel genetics in a newly discovered and endangered taxon of Galápagos tortoise. *PLoS ONE*, **4**, e6272.
- Ciofi C, Milinkovitch MC, Gibbs JP, Caccone A, Powell JR (2002) Microsatellite analysis of genetic divergence among populations of giant Galápagos tortoises. *Molecular Ecology*, **11**, 2265–2283.
- Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

- Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite mediated selection against inbred Soay sheep in a freeliving, island population. *Evolution*, **53**, 1259–1267.
- Cornuet J-M, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.
- Cowie RH, Holland BS (2008) Molecular biogeography and diversification of the endemic terrestrial fauna of the Hawaiian Islands. *Philosophical Transactions of the Royal Society B*, **363**, 3363–3376.
- Cracraft J (1983) Species concepts and speciation analysis. *Current Ornithology*, **1**, 159–187.
- Cui R, Schumer M, Kruesi K *et al.* (2013) Phylogenomics reveals extensive reticulate evolution in *Xiphophorus* fishes. *Evolution*, **67**, 2166–2179.
- Edwards DL, Benavides E, Garrick RC *et al.* (2013) The genetic legacy of Lonesome George survives: giant tortoises with Pinta Island ancestry identified in Galápagos. *Biological Conservation*, **157**, 225–228.
- Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, **11**, 1591–1604.
- Fagundes NJ, Ray N, Beaumont M *et al.* (2007) Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences USA*, **104**, 17614–17619.
- Fitzpatrick BM, Johnson JR, Kump DK *et al.* (2009) Rapid fixation of non-native alleles revealed by genome-wide SNP analysis of hybrid tiger salamanders. *BMC Evolutionary Biology*, **9**, 176.
- Fleischer RC, McIntosh CE, Tarr CL (1998) Evolution on a volcanic conveyor belt: using phylogeographic reconstructions and K-Ar-based ages of the Hawaiian Islands to estimate molecular evolutionary rates. *Molecular Ecology*, **7**, 533–545.
- Frankham R (1995) Effective population-size: adult-population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.
- Garrick RC, Benavides E, Russello MA *et al.* (2012) Genetic rediscovery of an 'extinct' Galápagos giant tortoise species. *Current Biology*, **22**, R10–R11.
- Geist DJ, Naumann TR, Standish JJ *et al.* (2005) Wolf volcano, Galápagos Archipelago: melting and magmatic evolution at the margins of mantle plume. *Journal of Petrology*, **46**, 2197–2224.
- Gerard D, Gibbs HL, Kubatko L (2011) Estimating hybridization in the presence of coalescence using phylogenetic intra-specific sampling. *BMC Evolutionary Biology*, **11**, 291.
- Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics*, **139**, 463–471.
- Goodman SJ (1997) R_{ST} -Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.
- Grant PR, Grant BR (1992) Hybridization of bird species. *Science*, **256**, 193–197.
- Grant BR, Grant PR (2008) Fission and fusion of Darwin's finches populations. *Philosophical Transactions of the Royal Society B*, **363**, 2821–2829.
- Grant PR, Grant BR (2014) *40 Years of Evolution. Darwin's finches on Daphne Major Island*. Princeton University Press, Princeton, New Jersey.
- Grant PR, Grant BR, Petren K (2005) Hybridization in the recent past. *American Naturalist*, **166**, 56–67.
- Hendry AP, Grant PR, Grant BR *et al.* (2006) Possible human impacts on adaptive radiation: beak size bimodality in Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1887–1894.
- Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–589.
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.
- Kalinowski ST (2005) HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.
- Kleindorfer S, O'Connor JA, Dudaniec RY *et al.* (2014) Species collapse via hybridization in Darwin's tree finches. *American Naturalist*, **183**, 325–341.
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
- Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, **279**, 2115–2118.
- Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (Version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, **6**, 971–973.
- Nosil P, Funk D, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Paquette SR, Louis EE Jr, Lapointe F-J (2010) Microsatellite analyses provide evidence of male-biased dispersal in the radiated tortoise *Astrochelys radiata* (Chelonia: Testudinidae). *Journal of Heredity*, **101**, 403–412.
- Parent CE, Caccione A, Petren K (2008) Colonization and diversification of Galápagos terrestrial fauna: a phylogenetic and biogeographical synthesis. *Philosophical Transactions of the Royal Society B*, **363**, 3347–3361.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Poulakakis N, Glaberman S, Russello M *et al.* (2008) Historical DNA analysis reveals living descendants of an extinct species of Galápagos tortoise. *Proceedings of the National Academy of Sciences USA*, **105**, 15464–15469.
- Poulakakis N, Russello M, Geist D, Caccione A (2012) Unraveling the peculiarities of island life: vicariance, dispersal and the diversification of the extinct and extant giant Galápagos tortoises. *Molecular Ecology*, **21**, 160–173.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

- R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Russello MA, Beheregaray LB, Gibbs JP *et al.* (2007) Lonesome George is not alone among Galápagos tortoises. *Current Biology*, **17**, R317–R318.
- Russello MA, Poulakakis N, Gibbs JP *et al.* (2010) DNA from the past informs *ex situ* conservation for the future: an “extinct” species of Galápagos tortoise identified in captivity. *PLoS ONE*, **5**, e8683.
- Schluter D (2000) *The Ecological Theory of Adaptive Radiation*. Oxford University Press, Oxford, UK.
- Seehausen O (2006) Conservation: losing biodiversity by reverse speciation. *Current Biology*, **16**, R334–R337.
- Seehausen O, Takimoto G, Roy D, Jokela J (2008) Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, **17**, 30–44.
- Singhal S, Moritz C (2013) Reproductive isolation between phylogeographic lineages scales with divergence. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20132246.
- Steinfartz S, Glaberman S, Lanterbecq D *et al.* (2009) Progressive colonization and restricted gene flow shape island-dependent population structure in Galápagos marine iguanas (*Amblyrhynchus cristatus*). *BMC Evolutionary Biology*, **9**, 297.
- Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer, Sunderland, Massachusetts.
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Talavera G, Lukhtanov VA, Rieppel L, Pierce NE, Vila R (2013) In the shadow of phylogenetic uncertainty: the recent diversification of *Lysandra* butterflies through chromosomal change. *Molecular Phylogenetics and Evolution*, **69**, 469–478.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Templeton AR (2008) The reality and importance of founder speciation in evolution. *BioEssays*, **30**, 470–479.
- Templeton AR, Crandall KA, Sing CF (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, **132**, 619–633.
- Throp JL (1975) Note on the management and reproduction of the Galápagos tortoise at the Honolulu Zoo. In: *Breeding Endangered Species in Captivity* (ed. Martin RD), pp. 39–42. Academic Press, New York.
- Townsend CH (1925) The Galápagos tortoises in their relation to the whaling industry. A study of old logbooks. *Zoologica*, **4**, 55–135.
- Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, 4th edn. Springer, New York.
- Vergilino R, Markova S, Ventura M, Manca M, Dufresne F (2011) Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. *Molecular Ecology*, **20**, 1191–1207.
- Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357–362.
- Webb WC, Marzluff JM, Omland KE (2011) Random interbreeding between cryptic lineages of the Common Raven: evidence for speciation in reverse. *Molecular Ecology*, **20**, 2390–2402.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wessel A, Hoch H, Asche M *et al.* (2013) Founder effects initiated rapid species radiation in Hawaiian cave planthoppers. *Proceedings of the National Academy of Sciences USA*, **110**, 9391–9396.
- Wiens JJ (2011) The causes of species richness patterns across space, time, and clades and the role of “ecological limits”. *Quarterly Review of Biology*, **86**, 75–96.
- Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

R.C.G. and G.C. conceived the study. M.A.R., C.H., J.P.G., W.T., C.C. and A.C. conducted the fieldwork. R.C.G., E.B., M.A.R. and C.H. performed the laboratory work. R.C.G., E.B. and C.H. analysed the data. R.C.G., E.B., M.A.R., D.L.E., J.P.G. and A.C. drafted the manuscript. All authors read and approved the final version.

Data accessibility

Microsatellite genotypes and aligned DNA sequences are available for download at <http://datadryad.org> under DRYAD Repository entry doi:10.5061/dryad.5t1q.

Supporting information

Additional supporting information may be found in the online version of this article.

Appendix S1 Methods.

Table S1 Criteria used to classify 841 Wolf Volcano tortoises as purebreds, F_1 hybrids or backcrosses, based on STRUCTURE *Q*-values derived from crosses simulated in HYBRID-LAB.

Table S2 Assessment of historical divergence between the Puerto Bravo (PBR) and Piedras Blancas (PBL) lineages of *C. becki*, estimated using approximate Bayesian computation.

Table S3 Proportions of offspring expected to have a PBR mtDNA haplogroup sequence if mating between the two lineages of *C. becki* on Wolf Volcano is random.

Table S4 Probability of random male (σ) \times female (ϕ) pairings, calculated for each of eight types of *C. becki* tortoises.

Fig. S1 Statistical parsimony network showing evolutionary relationships among mitochondrial DNA (mtDNA) sequences carried by Wolf Volcano tortoises.

Fig. S2 Best-fit model of historical divergence between the Puerto Bravo (PBR) and Piedras Blancas (PBL) lineages of *C. becki*, estimated using approximate Bayesian computation.

Fig. S3 Best-fit model of historical divergence between Puerto Bravo (PBR) and Piedras Blancas (PBL) lineages of *C. becki*, including hypothetical bottleneck events, estimated using approximate Bayesian computation.

Fig. S4 Histograms comparing the frequency of two *C. becki* mtDNA haplogroups in the present generation (G_0) vs. projected frequencies after one generation of random mating (G_1).

Fig. S5 Histograms comparing the frequency three classes of *C. becki* tortoises, as determined using nuclear microsatellite

data (i.e., purebred PBR, purebred PBL, and hybrids), in the present generation (G_0) vs. projected frequencies after one generation of random mating (G_1).

Fig. S6 Frequency distributions comparing the current level of linkage disequilibrium (LD) among microsatellite alleles of *C. becki* tortoises (G_0 ; solid lines, filled circles) vs. projected LD after one generation of random mating (G_1 ; dashed lines, open circles).

Fig. S7 Box-and-whisker plots showing projected change in allelic richness (A_R) at microsatellite loci of *C. becki* tortoises, after one generation of random mating.

Fig. S8 Box-and-whisker plots showing projected change in observed heterozygosity (H_O) at microsatellite loci of *C. becki* tortoises, after one generation of random mating.