Exploring species boundaries to understand the invasive impact of an introduced species on endemic species in the Galápagos Archipelago

Sarah Pangburn
Advisor: Andrea Sequeira

Biological Sciences
Wellesley College

April 2016

This material is copyrighted by Sarah Pangburn and Andrea Sequeira, 2016
Acknowledgements

I would like to start by thanking my fellow lab members, both past and present. Thank you to Courtney Stepien and Fei Mok for all of their work and Adrienne Cheng for taking me under her wing and being an amazing teacher and friend. A huge thank you to Flavia Mendonca de Sousa, Sara Eslami, Mary Kate Dornon, and Eleanor Zagoren for their hard work; this thesis would not have been possible without them. I would also like to thank my thesis committee members Professor Adam Matthews and Professor Jocelyne Dolce for their continual support. They have offered me advice on a multitude of topics, from my thesis to my life after Wellesley. I am greatly indebted to my thesis advisor, Professor Andrea Sequeira, who has provided me with constant support and encouragement, and has given me the opportunity to pursue all of my academic and research goals. Finally, I would like to thank my friends and family members, who have supported me in all of my academic endeavors.
# Table of Contents

List of Figures and Tables  iv  
Abstract  1  
Introduction  2  
Materials and Methods  16  
Results  27  
Discussion  43  
Appendix  55  
References  58
List of Figures and Tables

FIGURE 1--Map of Galápagos Archipelago. 10
FIGURE 2--Maximum parsimony strict consensus derived from a concatenated mitochondrial and nuclear dataset. 12
FIGURE 3--Weevil images of one introduced and three endemic Galapaganus species. 13
FIGURE 4--Map of sampled localities in Santa Cruz (above) and San Cristóbal (below). 17
FIGURE 5--Graphic representation of the variables estimated through the Isolation with Migration model. 22
FIGURE 6--Comparison of estimates of mitochondrial unidirectional gene flow (m). 28
FIGURE 7—Density distribution plots of mitochondrial m1 and m2 values for all species pairs. 29
FIGURE 8--Comparison of estimates of nuclear unidirectional gene flow (m). 30
FIGURE 9--Density distribution plots of nuclear m1 and m2 values for all species pairs. 31
FIGURE 10--Comparison of estimates of mitochondrial Theta (θ) for all species. 32
FIGURE 11--Density distribution plots of mitochondrial Theta (θ) estimates for Santa Cruz species. 33
FIGURE 12--Comparison of estimates of nuclear Theta (θ) for all species. 34
FIGURE 13--Density distribution plots of nuclear Theta (θ) values for Santa Cruz species. 35
FIGURE 14--Comparison of estimates of mitochondrial unidirectional effective gene flow (2Nm) in Santa Cruz Island. 36
FIGURE 15--Density distribution plots of nuclear Theta (θ) values for Santa Cruz species. 37
FIGURE 16--Visual summary generated with IMFig of the parameters estimated with IMa2 for each of the mitochondrial pair-wise comparisons. 38
FIGURE 17--Comparison of estimates of nuclear unidirectional effective gene flow (2Nm) in Santa Cruz Island. 40
FIGURE 18--Density distribution plots of nuclear unidirectional effective gene flow (2Nm) for Santa Cruz species pairs. 41
FIGURE 19--Visual summary generated with IMFig of the nuclear parameters estimated with IMa2 for each of the pair-wise comparisons. 42
TABLE 1. Sampled localities 18
TABLE 2. Number of populations and individuals per population for each species and sequence length per gene region. 21
APPENDIX 1: Run and prior details. 55
Abstract

Introduced species can threaten endemic species through competitive exclusion, niche displacement, introgression, predation, and hybridization. As a result of hybridization, endemic species can lose important genetic adaptations. Small island populations of endemic species may be especially in danger of extinction due to hybridization. The *Galapaganus* weevil system colonized the Galápagos Archipelago between 8.6 and 11.5 million years ago from continental Ecuador. However, genetic analyses indicate that *G. h. howdenae* was only just recently introduced to Santa Cruz Island via an accidental human-mediated introduction during the human colonization period (1832-1959). The focus of this study is on six species of the *Galapaganus* radiation, two endemics and one introduced on Santa Cruz and three endemics on San Cristóbal. An analysis of nuclear and mitochondrial DNA was completed using the Isolation with Migration model in order to explore the possibility of interspecies gene flow, specifically between the introduced *G. h. howdenae* and the highland specialist and single island endemic *G. ashlocki*, and the extent of the genetic impact of this recent introduction. Our analyses recover higher and significant mitochondrial and nuclear population migration estimates (2Nm) between *G. ashlocki* and *G. h. howdenae* as well as significant nuclear gene flow estimates between highland and lowland endemics in Santa Cruz. Estimates of population size indicate modest values for the introduced populations compared to its endemic counterparts, suggesting that the genetic footprint of the introduced species could be independent of its population size. Future studies will focus on estimating the timing of these gene-exchanges in an effort to understand determinants or precursors of increased interspecies gene flow.
Introduction

Species Introductions and Challenges to Biodiversity

Aside from habitat loss, invasive species pose one of the greatest threats to species and habitat diversity (Wilcove et al. 1998). Species introductions may cause large impacts at the community and ecosystem levels, as well as at the genetic level, to closely related species and to locally adapted or endemic populations of the same species. According to the U.S. Fish & Wildlife Service, invasive species are a leading cause of population decline and extinction in animals and can cost the United States billions of dollars in damages every year. More than 400 of the over 1,300 species currently protected under the Endangered Species Act are at risk due to displacement by, competition with, and predation by invasive species. An additional 180 species not currently protected are also at risk to become part of the endangered list due to invasive species (Service 2012).

Examples of the devastating effects of invasive species on native flora and fauna are abundant worldwide, both on continents and islands. From a conservation point of view, species introductions can be problematic because they can cause the extinction of these native populations. For example, the eastern grey squirrel (*Sciurus carolinensis*) is an invasive species from America in Europe and is causing a contraction in range of the European red squirrel. The introduction of the grey squirrel has proved problematic for the red squirrel due to resource competition and the spread of a poxvirus, which is deadly for the red squirrel (Sandro 2008). Sandro (2008) predicted that the spread of the grey squirrel from Italy to France and Switzerland could occur within 20 to 30 years, which would be a serious threat for the survival of the red squirrel. Invasive species, such as the zebra mussel (*Dreissena polymorpha*), have also been introduced to North America. The zebra mussel threatens North American freshwater mussels, especially in the Mississippi River basin, which has the highest concentration of native freshwater mussels. Due to high levels of competition,
native mussel populations normally go extinct within 4-8 years after the invasion of the zebra mussel (Ricciardi et al. 1998).

Because of the great potential for the loss of endemics, invasive species can be even more destructive on island systems such as the Galápagos Archipelago, which is the island system of focus in this study. Popular examples of destructive species introductions on islands include feral goats (Campbell and Donlan 2005), feral cats (Nogales et al. 2004), and feral mallards (Engilis and Pratt 1993). Feral goats have been known to overgraze on multiple islands, which can lead to primary and secondary biodiversity loss. So far, feral goats have been eradicated from 120 islands, including Santiago Island and Isabela Island in the Galápagos, Guadalupe Island in Mexico, and the Great Barrier Island in New Zealand, via a range of methods from helicopter culling to hunting dogs (Campbell and Donlan 2005). Nogales and coauthors (2004) pointed out that feral cats (*Felis catus*) are responsible for the extinction and endangerment of small mammals and bird species, especially on islands, and are labeled as one of the 100 worst invasive species. Feral cats have been removed from 48 islands in areas such as Baja California, New Zealand, Australia, the Pacific Ocean, Seychelles, the sub-Antartic, Macaronesia, Mauritius, and the Caribbean (Nogales et al. 2004). The Koloa Hawaiian duck is in danger due to habitat loss and predation by introduced species. The introduction of feral mallards to the Hawaiian Islands, due to accidental breeding and commercial use, has caused issues for the native Koloa populations due to competition and hybridization. The interbreeding between the mallards and Koloa could cause the loss of a pure Koloa population, especially because the domesticated mallard has adopted a multiple-copulation reproductive technique (Engilis and Pratt 1993). Insect introductions can cause devastating effects in island systems as well. As of 1998, there were a total of 2,621 species of introduced insects in the Hawaiian Islands, 400 of which were introduced intentionally (Nishida
The introductions of alien insect species to the Galápagos Archipelago total 292 species in 16 insect orders and are most often associated with horticulture, agriculture, and livestock transfer to the islands (Peck et al. 1998). These species have been introduced both intentionally and accidentally. Introductions also occurred via whaling ships during the first 300 years after Europeans discovered the Galápagos; however, most introductions occurred during the colonization period from the 1830s to the 1890s and from the 1920s to the 1970s, in which they were transported via immigrant settlers and wartime material during World War II (Peck et al. 1998).

**Island System and island biodiversity**

Island systems can have both wide-spread (cosmopolitan) and endemic species (species that are only found on the islands). Islands that are larger or closer to the mainland contain more species and have a greater diversity of species than smaller or more remote islands (Macarthur and Wilson 1963). There are two types of endemism that can result depending on the place of origin of the species divergences: paleoendemics and neoendemics. Paleoendemics are formed as a result of the narrowing of phenotypic and ecological ranges. These restricted ranges lead to even greater range restriction, which can lead to endemism. In other words, previously widespread species become locally extinct from portions of the distribution range, leaving smaller and isolated ranges as the only habitat available for that newly endemic species. Neoendemics, on the other hand, form when *in-situ* evolution occurs in an isolated area, such as an island, leading to new species that are endemic to that area and therefore have phenotypic adaptations to that area (Hermant et al. 2013). Island inhabitants are more susceptible to extirpation and extinction than their continental counterparts due to their geographic isolation and the general species poverty of island systems (Whittaker 1998). This may be due to island specific evolutionary trends, such as the loss of
long distance dispersal and “unnecessary” competitive and defensive features, which results from reduced numbers of predators, parasites and competitors. This phenomenon is known as ecological release. When new, introduced species then colonize an island system, they can, therefore, be a potential threat to the locally adapted and less competitively suited endemic island species. The introduced species threaten the endemic species through competitive exclusion, niche displacement, introgression (the backcross breeding of hybrids with one of the parental populations), predation, and hybridization, all of which can end in extinction (Mooney and Cleland 2001).

**Hybridization**

There are multiple documented cases of interspecies hybridization across both non-island (Adams et al. 2003, Hailer and Leonard 2008, Olave et al. 2011) and island species (Francisco-Ortega et al. 1996, Warren et al. 2012) due to overlapping habitat ranges and species introductions. Hybridization can continue to occur between species for millions of years after two species diverge due to the slow evolutionary process of post-zygotic isolation (Grant and Grant 1996), in which species maintain their identity in the face of gene-flow (Nosil 2008). Alternatively, previously geographically and reproductively isolated species can start to interbreed when contact is established due to range expansion. This secondary contact could be a result of human aided introductions as well. As a result, many endemic species are at risk by their new neighbors and ultimately face extinction.

For example, the red wolf (*Canis rufus*) in North America has been hybridizing with coyote (*Canis latrans*) populations since the 1960s (Adams et al. 2003). All wild red wolves were captured for their protection and released in 1984 in the Alligator River National Wild-life Refuge in North Carolina. However, the Adams et al. (2003) study of the refuge area indicated the presence of hybrids, including one previously unknown hybrid. They suggest a yearly non-invasive genetic
scanning technique using fecal samples in order to control hybridization and maintain the red wolf population (Adams et al. 2003). Other examples of hybridization between grey wolves and coyotes in the Great Lakes area have been detected, which threatens the species with the smaller population size (Hailer & Leonard 2008). Hailer & Leonard (2008) also found evidence for hybridization between three Canis populations: Mexican wolves (C. lupus baileyi), red wolves (C. rufus) and coyotes (C. latrans), which has impacted the wolf populations the most and the coyote populations the least. The South American Liolameus gracilis and Liolameus bibronii lizard species have overlapping geographic ranges in Argentina, and nuclear markers have proved that they are hybridizing (Olave et al. 2011). Warren et al. (2012) confirmed hybridization between female Fouida omissa and male F. madagascariensis birds in the western Indian Ocean islands via parameter estimates for secondary gene flow after divergence. Small populations on islands are especially in danger of extinction due to hybridization because they are less genetically diverse than mainland species and have weak crossing barriers (Mooney & Cleland 2001). In contrast, some examples of hybridization on islands, such as the interspecific hybridization of Argyranthemum flora in Macaronesia, have a positive impact and are linked to the diversification of the genus (Francisco-Ortega et al. 1996).

In order to estimate hybridization, one can perform phylogenetic analyses to understand whether morphologically divergent species are reciprocally monophyletic (in which they form discrete and separate evolutionary lineages). If the phylogenies are constructed with molecular data (mitochondrial and nuclear sequences), then finding discrepancies between evolutionary trees can signal the occurrence of hybridization between divergent, but closely related species. This has been seen in the case of the Neodiprion sawflies (Linnen and Farrell 2007) as well as in other animal systems (Toews and Brelsford 2012). Further exploration focusing on the timing of these gene exchanges can help determine whether gene flow occurred during species divergence or through
recent hybridization (Niemiller et al. 2008). Such estimations involve statistical tests using specific genes to estimate migration rates and other related population parameters including the timing of gene exchange (Hey and Nielsen 2007).

*Estimating Migration Rates Through IMa2*

One of the challenges associated with studying hybridization between two species is finding a software program that will create both migration rate estimates between populations and a time estimate for when populations originally diverged from the ancestral population. Through the use of the Hey lab isolation with migration software, IMa2, one can obtain these time and gene exchange estimates (Hey 2010). In addition, performing separate analysis on mitochondrial and nuclear gene regions might also offer clues on the timing of gene exchange.

When analyzing two populations using the IM program, there are parameters that are tested for both of the populations as well as an ancestral population from which the two populations derived. These parameter estimates include two unidirectional migration rates from population 1 to population 2 (m12) and from population 2 to population 1 (m21); a time estimate for when the ancestral population diverged into the two populations (t); and three mutation rate estimates (4Nµ, N=population size): Θ1 (population 1 mutation rate), Θ2 (population 2 mutation rate), and Θa (ancestral population mutation rate) (Sethuraman and Hey 2016). Within the IMa2 program itself, the Θ1, Θ 2, Θa, m12, and m21 parameters are labeled as q0, q1, q2, m0>1, and m1>0, respectively.

The program generates these estimates by running a Markov chain Monte Carlo simulation (Sethuraman and Hey 2016) and is operated through command prompt using command lines. Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. In this case, the Monte Carlo simulation uses either a method where the probability for a hypothesis gets updated as
more evidence becomes available (Bayesian method) or estimating the parameters of the statistical model given the data (likelihood method) to estimate the migration rate for the populations and the time for their divergence from each other (Nielsen and Wakeley 2001). As explained by Hey (2011a), each command line can include a number of terms preceding integers. The terms included in the command lines were: i (the input file), o (the output file), q (maximum population size prior value), b (number of burnin steps), l (length of the run), m (migration prior value), t (maximum splitting time prior value), and p (output options). Prior values are the values that one expects before the simulation is run (McCarthy 2007). Those prior values are starting points and do not influence the results of individual runs. The number of burnin steps is the number of steps that are initially discarded. These steps are discarded because the Markov chain can take multiple steps before it gets close to stable values converging with the posterior distributions (McCarthy 2007). Each command line and run will result in a new output file with different results.

In order to determine if the run was successful and produced reliable data, one should look at the Effective Sample Size Estimates (ESS values) that are generated. ESS values are estimates of the number of independently sampled points for each parameter and they indicate how well the Markov Chain is mixing (Hey 2009). Higher ESS values, therefore, indicate better mixing. Hey (2007) states that low ESS values are less stable and reliable with shorter run times, and one should generally have a run length of over 1 million in order to obtain more reliable ESS values (an acceptable ESS value is around 50). Many studies have explored the utility of the Isolation with Migration model in order to understand the nature of gene exchange in groups of organisms as varied as insects, apes and marine mammals (Pollard et al. 2006, Becquet and Przeworski 2007, Hellberg 2009). Niemiller et al. (2008), for example, used the IM software to study the divergence between spring and cave salamanders. By analyzing two mitochondrial genes and one nuclear gene in the IM program, they
determined that the divergence with the salamanders occurred with gene flow. In other words, the exchange in the salamander study had occurred early during the divergence of the two species, rather than more recently through post-divergence hybridization. This supports the adaptive-shift hypothesis for the salamanders, which states that speciation occurred without geographic barriers (and with gene flow). Divergence was prompted by selection preferences for cave vs. surface habitats (Niemiller et al. 2008).

**Galapaganus Study System and the Islands they Inhabit**

Our study system of choice to explore the occurrence and timing of gene exchange between closely related species is a genus of broad nosed weevils, in which many members are endemic to the Galápagos archipelago.

The Galápagos archipelago is famous for its expansive wildlife, and scientists have been intrigued by the island system since Charles Darwin brought the islands and its finches into the spotlight (Darwin 1859). It is also geologically intriguing. The archipelago consists of ten main volcanic islands and other smaller volcanoes. Thirteen of the volcanoes have been active (White et al. 1993). The islands were formed due to volcanic eruptions and the movement of a tectonic plate over a hotspot (Morgan 1971, Christie et al. 1992, White et al. 1993). Unlike other island chains, such as the Hawaiian volcanoes, the volcanoes in this archipelago do not form a linear chain (White et al. 1993). The oldest Galápagos volcanoes are positioned in the southeast of the archipelago (White et al. 1993), and the age of the islands increases eastward (Geist et al. 1985, Hickman and Lipps 1985). Hickman and Lipps (1985) measured the age of the islands using subtidal to supratidal fossiliferous marine deposits. They determined that all of the fossils were less than 2 million years old, which indicates that the islands emerged from the sea recently, and the evolution of the terrestrial biota on the islands occurred within the past 3-4 million years. Through other geological estimation
methods such as plate movement models, a maximum age of emergence for the islands has been determined to be between 4.5 and 6.3 million years (Geist 1996).

Figure 1. Map of Galápagos Archipelago. Insets illustrate legends, distance from the mainland, and topographic detail. Image taken from en.wikipedia.org/wiki/Galápagos_Islands.

There are six main ecological zones within the biggest Galápagos islands. The zones inhabited by the *Galapaganus* weevils are the lowlands, which include littoral and arid plants, the lowlands and mid-elevations, which include littoral, arid, and transition plants, and the moist highlands, which include *Scalesia, Miconia*, and Fern-Sedge (Pampa) plants (Sequeira et al. 2008a). Within Santa Cruz, an additional distinction of importance is the human created Agricultural Zone and the protected National Park (Mok et al. 2014).
The *Galapaganus* weevil genus contains fifteen species. Thirteen of these species are flightless, ten of which are endemic to the Galápagos Archipelago (Sequeira et al. 2008b). The *Galapaganus* genus did not branch out across the islands according to the island progression rule (Sequeira et al. 2008a). In fact, the *Galapaganus* ancestor colonized the archipelago between 8.6 and 11.5 million years ago from Continental Ecuador, which is earlier than the oldest geological age estimates of the islands (Sequeira et al. 2008a). This discrepancy between the introduction time estimate and the island age estimate might be due to the existence of submerged seamounts: 7 million years ago there were other emerged islands, which sunk below sea level south east of the archipelago before the younger islands had appeared (Sequeira et al. 2000) and were used as colonization platforms. The original colonizers reached the submerged islands and then the younger islands that exist today (Sequeira et al. 2000, Sequeira et al. 2008b, Figure 2). One member of the *Galapaganus* genus, *G. h. howdenae*, has a very different colonization history into the islands: it was accidentally introduced to Santa Cruz Island in the Galápagos from mainland Ecuador via humans during the colonization period (1832-1959). This occurred prior to the spurt in human population growth on the island (Mok et al. 2014).
Figure 2. Maximum parsimony strict consensus derived from a concatenated mitochondrial and nuclear dataset. Topological differences with the Bayesian majority rule consensus tree are marked with gray bars. Numbers above the branches indicate bootstrap values and numbers below the branches indicate Bayesian posterior probabilities expressed as percentages. The insets illustrate the hypothesized order of island colonization suggested by the MP (A) and Bayesian topologies (B). Dotted arrows incorporate equally parsimonious scenarios in each case. Numbers below the branches in the topology correspond to the arrows in the inset for two alternative colonization scenarios: MP: normal font, BI: italics, IC: Initial colonization and * indicates intra-island speciation. Taken from Sequeira et al. (2008b).
Questions and predictions

The focus of this study will be on six species of the *Galapaganus* weevils in the Galápagos, five of which are endemics and one, which is an introduced species: *G. h. howdenae*, *G. conwayensis*, *G. ashlocki*, *G. galapagoensis*, *G. collaris*, and *G. vandykei* (Figure 3).

![Weevil images of one introduced and three endemic *Galapaganus* species](image)

**Figure 3. Weevil images of one introduced and three endemic *Galapaganus* species.** Dorsal and lateral views were taken using ten different focal points and compiled with Photomontage using the MCZ facility (Museum of Comparative Zoology, Harvard). From left to right: *G. h. howdenae*, *G. conwayensis*, *G. galapagoensis* and *G. vandykei*.

Previous analyses have shown clear phylogenetic species boundaries (Sequeira et al. 2008b) as well as location and habitat preference for four of these *Galapaganus* species: *G. h. howdenae* is found on mainland Ecuador and Santa Cruz island; *G. conwayensis* is found on Santa Cruz, Isabela, and Pinta and favors the lowland and mid-elevation areas but can be found in other ecological zones; *G. ashlocki* is found on Santa Cruz and favors the highland zones; *G. galapagoensis* is found on San Cristóbal and favors all ecological zones; *G. collaris* is found on San Cristóbal and Floreana and favors the highland zones; *G. vandykei* is found on Floreana, San Cristóbal, and Española islands (Sequeira et al. 2008a).

Although these species boundaries have been explored through phylogenetic means (Sequeira et al. 2008b), it has not been tested if recent hybridization is occurring between
populations. More specifically, is interspecies hybridization occurring between the introduced *G. h. howdenae* and endemic *G. conwayensis* and *G. ashlocki* populations in the highlands of Santa Cruz? In order to answer this question, six species comparisons will be made using sequences of mitochondrial and nuclear genes. The specific contrasts that will be performed are for species in Santa Cruz: *G. h. howdenae, G. conwayensis* and *G. ashlocki*, and in San Cristóbal: *G. galapagoensis, G. vandykei* and *G. collaris*, which act as a control.

The focus of these comparisons will be the estimation of migration rates between pairs of species measured using the IMa2 software. These migration rates will show if there is significant and recent gene flow, which would indicate if hybridization is occurring or not. If hybridization is recent, we predict that there will be larger estimates of genetic exchange between species pairs that include the introduced *G. h. howdenae* and Santa Cruz endemics *G. conwayensis* and *G. ashlocki*. In other words, when compared to the San Cristóbal species pairs, the species comparison for Santa Cruz would result in a higher rate of recent gene flow.

Another aspect to consider is the nature of the molecules under study and their inheritance patterns. Mitochondrial DNA is inherited maternally as a single non-recombining unit, therefore, hybrids, despite having equal contributions from both parental species, will carry mitochondrial DNA identical to the female parent. The signal of cross species hybridization therefore remains undiluted in this molecule in comparison with nuclear markers. The tendency of mitochondrial DNA to cross interspecific barriers is somewhat counter-intuitive considering the key function of the enzymes that it encodes, which could give rise to hybrid dysfunction (Boratynski et al. 2014) or hybrids with reduced fitness. However, mitochondrial introgression is still prevalent (Darras and Aron 2015). Given the method of inheritance of mitochondrial DNA, we expect the effects of hybridization to be observed more prominently in the mitochondrial genome than in the nuclear
Why do we need to know if *G. h. howdenae* is currently hybridizing with endemics in general, and in particular, with highland endemics (i.e. *G. ashlocki*)? Single island endemics have restricted distributions and their survival depends on effective reproductive isolation from close relatives. In addition to its status as a single island endemic, *G. ashlocki* is also a highland specialist, which means it only feeds and lives on plant hosts at higher elevations of Santa Cruz. These restrictions make *G. ashlocki* endangered (IUCN designation). The results of this study can help answer questions about the conservation of *G. ashlocki*, as well as help guide broader questions about the future evolution of endemic species in the face of introductions. Relevant broader questions are: Can useful adaptations be lost due to hybridization? Could this create an issue for the survival of endemic species?
Materials and Methods

Sampling

Samples were collected from different areas within Santa Cruz and San Cristóbal islands: *G. h. howdenae*, *G. conwayensis*, and *G. ashlocki* samples were collected from Santa Cruz whereas *G. galapagoensis*, *G. vandykei*, and *G. collaris* samples were collected from San Cristóbal (Table 1, Figure 4). Adult weevils were collected by beating on known host plants. Samples were preserved in 100% ethanol until DNA extraction.
Figure 4. Map of sampled localities in Santa Cruz (above) and San Cristóbal (below). Maps adapted from Mok et al. (2014) and Sequeira et al. (2008b). Insets represent location of Santa Cruz in the archipelago and of the archipelago in Western South America. Location codes on both panels follow Table 1, and within the Santa Cruz map, locality codes in normal font indicate *G. h. howdenae*, italics indicate *G. conwayensis*, and underlined codes indicate *G. ashlocki* populations.
Table 1. Sample localities. Sample localities listed for each species by island including locality code and location name, GPS coordinates, and altitude in meters.

<table>
<thead>
<tr>
<th>Area</th>
<th>Population code and locality</th>
<th>Coordinates</th>
<th>Altitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>South East</td>
<td>SR16, Tortuga Bay</td>
<td>S 00°45.584'; W 90°20.031'</td>
<td>0m</td>
</tr>
<tr>
<td></td>
<td>SR03, CDRS</td>
<td>S 00°44.483'; W 90°18.1'</td>
<td>0m</td>
</tr>
<tr>
<td></td>
<td>SR21, El Chato</td>
<td>S 00°39.946'; W 90°20.383'</td>
<td>106m</td>
</tr>
<tr>
<td></td>
<td>SR15, Caamaño</td>
<td>S 00°45.348'; W 90°18.536'</td>
<td>0m</td>
</tr>
<tr>
<td></td>
<td>SR19, Garapatero</td>
<td>S 00°40.849'; W 90°13.417'</td>
<td>45m</td>
</tr>
<tr>
<td></td>
<td>SR20, Between Garapatero and Bellavista</td>
<td>S 00°40.256'; W 90°15.014'</td>
<td>215m</td>
</tr>
<tr>
<td></td>
<td>SR34, West end agricultural region</td>
<td>S 00°37.855'; W 90°26.233'</td>
<td>277m</td>
</tr>
<tr>
<td>Los Gemelos</td>
<td>SR13/SR25/SR27, Los Gemelos</td>
<td>S 00°37.534'; W 90°23.098'</td>
<td>614m</td>
</tr>
<tr>
<td></td>
<td>SR23, Mina de Granillo Negro</td>
<td>S 00°34.308'; W 90°20.101'</td>
<td>260m</td>
</tr>
<tr>
<td></td>
<td>SR33, Mina de Granillo Rojo</td>
<td>S 00°36.917'; W 90°22.147'</td>
<td>572m</td>
</tr>
<tr>
<td>Conway Bay</td>
<td>SR30, At the Beach</td>
<td>S 00°33.543'; W 90°31.089'</td>
<td>5m</td>
</tr>
<tr>
<td></td>
<td>SR31, Up the trail CB</td>
<td>S 00°33.592'; W 90°30.897'</td>
<td>24m</td>
</tr>
<tr>
<td></td>
<td>SR32, At 50m CB</td>
<td>S 00°33.619'; W 90°30.244'</td>
<td>56m</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Agricultural Zone</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR10, Road to Cerro Croker</td>
<td>S 00°41.324, W 90°19.495'</td>
<td>249m</td>
</tr>
<tr>
<td></td>
<td>SR07, Finca Steve Devine</td>
<td>S 00°40.021, W 90°24.192'</td>
<td>351m</td>
</tr>
<tr>
<td></td>
<td>SR11, Above El Chato</td>
<td>S 00°38.364, W 90°25.605'</td>
<td>422m</td>
</tr>
<tr>
<td></td>
<td>SR22</td>
<td>S 00°39.946, W 90°26.383'</td>
<td>106m</td>
</tr>
</tbody>
</table>

For each species, the altitude is given in meters.
<table>
<thead>
<tr>
<th>Location</th>
<th>Coordinates</th>
<th>Elevation</th>
</tr>
</thead>
<tbody>
<tr>
<td>El Chato</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR06,</td>
<td>S 00°39.3735, W 90°17.257'</td>
<td>240m</td>
</tr>
<tr>
<td>2km north of El Casajo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR29,</td>
<td>S 00°37.962, W 90°26.134'</td>
<td>369m</td>
</tr>
<tr>
<td>West End Agricultural Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>National park</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR08/ SR18, Media Luna</td>
<td>S 00°40.11, W 90°19.438'</td>
<td>459m</td>
</tr>
<tr>
<td>SR09,</td>
<td>S 00°39.9685, W 90°19.503'</td>
<td>505m</td>
</tr>
<tr>
<td>Miconia Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SR12/ SR17, La Caseta</td>
<td>S 00°39.499, W 90°19.659'</td>
<td>614m</td>
</tr>
<tr>
<td>SR01, Cerro Croker trail</td>
<td>S 00°36.000', W 90°21.000'</td>
<td>400m</td>
</tr>
<tr>
<td><strong>G. ashloki</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Los Gemelos</td>
<td>S 00°37.534, W 90°23.098'</td>
<td>614m</td>
</tr>
<tr>
<td><strong>San Cristóbal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. collaris</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highlands: El Junco</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC01, El junco</td>
<td>S 00°36.148', W 90°20.757'</td>
<td>620m</td>
</tr>
<tr>
<td><strong>G. galapagoensis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lowlands South West</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC02, SE Wreck Bay</td>
<td>S 00°56.45700', W 89°35.10217'</td>
<td>0m</td>
</tr>
<tr>
<td>SC04, E Wreck Bay</td>
<td>S 00°55.38950', W 89°34.90383'</td>
<td>100m</td>
</tr>
<tr>
<td>SC06, N Wreck Bay</td>
<td>S 00°53.551', W 89°36.524'</td>
<td>0m</td>
</tr>
<tr>
<td><strong>G. vandykei</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North East</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC05, Rosa Blanca Bay</td>
<td>S 00°49.148', W 89°20.757'</td>
<td>0m</td>
</tr>
</tbody>
</table>

**DNA Extraction and Sequencing**

DNA was extracted from three legs for each specimen using the DNeasy Tissue Kit (Qiagen, Valencia, CA). PCR reaction conditions differed when amplifying nuclear genes (argK, ITS, EF1alpha) versus mitochondrial genes (COI, COII, 12S). Amplification conditions and primers followed the protocol described in Sequeira et al. (2008a) and Sequeira et al. (2008b).
The conditions and temperature profiles to amplify each gene region differed mainly in MgCl2 concentration and annealing temperature. Aliquots of PCR reactions were run in agarose gels (1.5%) to determine correct amplification size. The positive samples were purified with MiniElute PCR Purification Kit (Qiagen, Valencia, CA) and sent to University of Chicago Comprehensive Cancer Center DNA Sequencing and Genotyping Facility for sequencing.

**DNA Analysis: Sequence Editing and Alignment**

After receiving the sequencing data, the sequence fragments were renamed and edited in Sequencher (version 3.5, Gene Codes Corporation). A consensus was created from the forward and reverse primer sequences for each individual and gene region that were successful, and the sequence was edited. The protein coding sequences (COI, COII, ArgK, Eflalpha) were aligned in Sequencher, whereas the non-protein coding sequences (ITS, 12S, Eflalpha intron) were aligned in Clustal X Version 2.0 (Larkin et al. 2007).

**Estimation of Population Parameters**

**Preparation of input files for IMa2**

These alignments were then pasted into text files in order to create the input files for the IMa2 software. Twelve input files were created overall: six comparisons (G. h. howdenae & G. conwayensis, G. h. howdenae & G. ashlocki, G. conwayensis & G. ashlocki, G. galapagoensis & G. vandykei, G. galapagoensis & G. collaris, and G. vandykei & G. collaris) for both the nuclear and mitochondrial genes. For the nuclear gene comparisons, each gene region was analyzed separately within each file in order to allow for different mutation rates between loci. The program produces one combined estimate for each parameter. Information contained in the infile indicated how many specimens corresponded to each population and species as well as the length
of each gene region in base pairs. Table 2 summarizes the number of individuals and sequence length for each population and gene region analyzed.

**Table 2. Number of populations and individuals per population for each species and sequence length per gene region.** “Label” indicates ecological preference or origin (endemic or introduced) for each species, and “population codes” correspond to Figure 1 and Table 1 and provide information on the geographic coverage of the sample.

<table>
<thead>
<tr>
<th>Island/Species</th>
<th>Label</th>
<th>Population codes</th>
<th>COI, COII, 12S (1686 bp)</th>
<th>ITS (815 bp)</th>
<th>ArgK (728 bp)</th>
<th>Efl (815 bp)</th>
<th>Efl intron (120 bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Santa Cruz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G. h. howdenae</em></td>
<td>Introduced</td>
<td>SRO1, 06, 07, 08, 18, 17, 22, 28, 29</td>
<td>94</td>
<td>80</td>
<td>56</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td><em>G. ashlocki</em></td>
<td>Highland endemic</td>
<td>SR05</td>
<td>11</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td><em>G. conwayensis</em></td>
<td>Lowland endemic</td>
<td>SR02,03,13,19,21,32,33 (23, 27, 30, 31, 34)</td>
<td>49</td>
<td>46</td>
<td>43</td>
<td>56</td>
<td>52</td>
</tr>
<tr>
<td>San Cristóbal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G. galapagoensis</em></td>
<td>Lowland endemic</td>
<td>SC02,04,06</td>
<td>13</td>
<td>5</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>G. collaris</em></td>
<td>Highland endemic</td>
<td>SC01</td>
<td>6</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>G. vandykei</em></td>
<td>Lowland endemic</td>
<td>SC05</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**Running IMa2**

**What is an IM run?**

Output files are generated through a Markov Chain Monte Carlo simulation, which forms genealogies for each locus from a random sample. Posterior probability values (probability values that also consider the available information) are generated by using the priors set by the user in the command line. The posterior probability values from these simulations are then averaged in a final mean value for each parameter. The number of simulations is determined by the user. Usually the first 10% of the estimates of each run are discarded since they will be far off the desired convergent values. The number of iterations discarded is indicated by the user
through a burn-in command.

*What is IM calculating in each iteration?*

In each iteration, IMa2 calculates estimates for each parameter (q0, q1, q2, t, m0>1, and m1>0), and each estimate calculated is measured against the previous estimate to determine if it will be preserved as a better fit to the model or if it is rejected and the previous value is retained. The model includes the gene genealogy (evolutionary tree of individual gene sequences) and the mutation rates plus the parameters against the data. The fit is calculated through a likelihood equation. The likelihood value features in the decision of retaining or rejecting the values of each iteration. For each run, the ESS values (a measure of the scanning of the parameter space) are calculated from the acceptance rate of each parameter.

![Diagram](image)

**Figure 5. Graphic representation of the variables estimated through the Isolation with Migration model.** A, 1, and 2 labels represent the ancestral and the two derived populations respectively. The m1 and m2 labels indicate migration rates when the mutation rate is provided by the user while m1/µ and m2/µ are the migration rates relative to the mutation rates; the same applies for the theta value (θ). Figure taken from Hey and Nielsen (2004).
How does IM figure out migration rates?

As introduced in the previous section, the parameters of the two-population Isolation with Migration model include a population mutation rate for each population (q1, q2 and qa.), the scaled time at which the ancestral population gave rise to the two descendant populations, t, and the two scaled migration rates, m1 and m2. All parameters are calculated relative to the mutation rate. If the user provides a mutation rate then the actual demographic parameters can be calculated. We did not provide a mutation rate, therefore, our values are comparative.

Within the context of common origin for all the individuals sampled and for some common origin for the two species under study, the migration rates are estimated from fitting gene exchange into the model that would better match the patterns observed in the data.

How does IM figure out population size? And why does the population size value include the mutation rate?

Population size is also included in the model due to the relationship between mutation rate and population size. With bigger population sizes, there is more opportunity for more mutations to arise whereas with smaller population sizes, the probability of mutation is lower; however, if it occurs, the effect of mutations may be more extreme because they will spread through the population more quickly. The model takes all of these factors into account when estimating each parameter. The parameter used as an indirect estimate of population size (also scaled by the mutation rate) is \( \Theta = 4N\mu \) where N and \( \mu \) are the effective population size and the mutation rate per nucleotide site per generation, respectively.

How is 2Nm different from only m and which one is more informative?

Two Nm (2Nm) is the number of haploid genomes times the migration rate, which creates an estimate of migration effect depending on population size of the donor population. Due
to the relationship between population size and gene migration rate, this value indicates the total impact of migration rate for each comparison. In terms of our biological analysis, this is the estimate that provides the most information.

*How do we anticipate that sample sizes will affect our results?*

Smaller sample sizes and fewer populations from particular species may affect the accuracy of the migration rate estimates. With fewer specimens, there are fewer representatives of the gene pool of the entire population, which increases the chance of a skewed representation of the amount of genetic variation present. In turn, the absent genotypes may impact the construction of the gene genealogies.

*Sets of runs per species pair*

Each file was run at least ten times in order to obtain ten optimal output files (which were assessed and determined optimal by their ESS values). Appendix 1 lists all of the comparisons and details the ESS values, the command line priors, and average run times for each run per comparison. As detailed in the table, even though the command line was common to all runs, the runtime for each file varied depending upon the gene comparison. The one exception of command line variation is the run length in cases when the ESS values were too low and runs needed to be extended: mitochondrial comparison for *G. galapagoensis* & *G. collaris* and *G. vandykei* & *G. collaris* and nuclear comparison for *G. conwayensis* & *G. ashlocki*.

An example command line was: q1 100 m1 10 m2 10 t450 b1000 l200000 p45. Four of these terms act as priors for parameters to be estimated: “q1 [100]” is the maximum for the initial population size, “m1 [10]” and “m2 [10]” act as the maximum migration values in both directions, and “t [450]” is the maximum splitting time for ancestral divergence from the two populations in question (450 is large enough to encompass the time period of divergence for the *Galapaganus* genus).
of these terms are run and output related: “b” indicates the burn-in in number of steps (and each step is multiplied by 100, meaning that one record is saved every 100 iterations), “l” indicates run length in number of saved records (multiplied by 100), and “p” indicates the program to print the estimates of population migration rate and histograms. A successful run will produce good ESS values independent of the starting prior values.

Summary Statistics

The values within each output file that were used for analysis and the visualization of the data were the mean values for: q0, q1, q2, t, m0>1, and m1>0 parameters. For each set of 10 runs, the mean values of each run were compiled into a box plot using the Tyers Boxplot R software (http://boxplot.tyerslab.com). Median values were contrasted between species pairs and non-overlapping 95% confidence intervals indicated differing medians (Chambers et al. 1983). Density distributions were also plotted in Excel for three of the runs for each species pair in order to show continuous histograms for each parameter. The density of a continuous random variable is a function that describes the relative likelihood for this random variable to take on a given value. In our case, the variable is each of the parameters. These distributions include all of the calculated value points and the likelihood (comparable with the frequency) of each calculated value.

The IMFig program was used to generate figures representing population migration rate (2Nm) between two populations directly from the output file using a Python code. The boxes represent each population, horizontal lines represent splitting times, and the curved arrows represent population migration rates. The width of the arrow represents the intensity of the value, and arrows are marked with an asterisk if 2Nm values are statistically significant. Significance is determined through a likelihood ratio test where the likelihood or fit of the model is contrasted with and without migration (Hey 2011b). The Likelihood Ratio statistic $2[\text{LogHo (no migration)} - \text{LogHa(with migration)}]$ is
calculated and approximated to a Chi square distribution where the degrees of freedom are the difference in the number of parameters between the two models. The migration model is deemed to have a better fit if the value of the statistic is significant at 5% (Nielsen and Wakeley 2001).
Results

Migration parameter estimates derived from mitochondrial sequences

The IMa2 analyses of mitochondrial sequence data indicates that within Santa Cruz, the estimated gene flow or migration rate from G. h. howdenae to G. ashlocki was significantly larger than the other Santa Cruz migration estimates. The unidirectional migration rate estimates for G. ashlocki to G. h. howdenae, G. conwayensis to G. h. howdenae, G. h. howdenae to G. conwayensis, G. ashlocki to G. conwayensis, and G. conwayensis to G. ashlocki were all not significantly different from each other. The migration estimates from the mitochondrial data for the San Cristóbal comparisons show that the rate of gene flow from G. galapagoensis to G. collaris, G. collaris to G. galapagoensis, G. vandykei to G. galapagoensis, and G. galapagoensis to G. vandykei were all not significantly different from each other but were significantly different from the migration rates from G. vandykei to G. collaris and G. collaris to G. vandykei. These migration rates were also significantly larger and different from the Santa Cruz comparisons except for the G. h. howdenae to G. ashlocki comparison. The San Cristóbal comparisons, G. vandykei to G. collaris and G. collaris to G. vandykei, were significantly different from all of the other values but not significantly different from each other (Fig. 6). Although these last two values were the largest migration estimates, they also stem from the smallest sample size, which may be a contributing factor to the large estimate value. Smaller sample sizes can skew results because they are more prone to sampling errors, and, therefore, may not be accurate representations of the population. In summary, within Santa Cruz (and among the simulations with the largest sample sizes) the largest mitochondrial gene flow estimates were between introduced and endemic species.
Figure 6. Comparison of estimates of mitochondrial unidirectional gene flow (m). Means of each of the ten runs per species comparison were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).

Density distributions illustrating the values obtained from a sample of three runs per species pair were also plotted for the mitochondrial sequence data. Out of all the comparisons, only the migration data for the Santa Cruz comparisons *G. ashlocki* to *G. h. howdenae* and *G. h. howdenae* to *G. conwayensis* had well-defined frequency peaks not at zero (Fig. 7), indicating that the values with the largest likelihood, and most frequently found, during the run are in accordance to the trends of the standard summary statistics used. The agreement with the mean values illustrates that mean m1 and m2 values are an accurate representation of the parameters obtained during the simulation.
Figure 7. Density distribution plots of mitochondrial m1 and m2 values for all species pairs.
The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

Migration parameter estimates derived from nuclear sequences

For the Santa Cruz comparisons, the migration estimates derived from nuclear data for G. ashlocki to G. h. howdenae, G. h. howdenae to G. ashlocki, G. conwayensis to G. h. howdenae, and G. h. howdenae to G. conwayensis were very low and all not significantly different from each other. These comparisons were significantly different from the highest migration values for Santa Cruz, from G. ashlocki to G. conwayensis and from G. conwayensis to G. ashlocki, which were also significantly different from each other. For the San Cristóbal comparisons, the migration rate from G. galapagoensis to G. collaris had the highest value. This value was not significantly different from the Santa Cruz G. ashlocki to G. conwayensis value but was significantly different from all of the other San Cristóbal comparisons. G. collaris to G. galapagoensis and G. vandykei to G. galapagoensis were not significantly different form each other and G. vandykei to G. collaris and G. collaris to G. vandykei were not significantly different but they all were significantly different
from *G. galapagoensis* to *G. vandykei* (Fig. 8). These generally lower values for introduced to endemic comparisons were not unexpected given the lower mutation rate for nuclear genes and the recent sympatry between the introduced and endemic species. We will explore the relative impact of mitochondrial versus nuclear exchange further in the discussion.

![Figure 8. Comparison of estimates of nuclear unidirectional gene flow (m). Means of each of the ten runs per species comparison were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).](image)

The density distributions show the larger likelihood values, or a frequency peak away from zero, for both the migration rate from *G. ashlocki* to *G. conwayensis* and *G. conwayensis* to *G. ashlocki*. *G. conwayensis* to *G. h. howdenae* had a peak slightly above zero, whereas *G. ashlocki* to *G. h. howdenae* and *G. h. howdenae* to *G. conwayensis* had peaks at zero (Fig. 9). When the higher likelihood values found during the simulation are not away from zero, the simulation is favoring a model where the value of that parameter is estimated at zero, indicating no gene exchange for each of those pairwise comparisons.
Figure 9. Density distribution plots of nuclear m1 and m2 values for all species pairs. The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

Theta, or population size parameters, derived from mitochondrial sequences

The population size estimates using mitochondrial sequences were compiled from the two sets of runs performed for each species (a total of 20 runs). Estimates plotted in Fig. 10 indicate the *G. conwayensis* and *G. ashlocki* estimates as the largest population sizes and that those estimates were not significantly different from each other. Estimates of *G. conwayensis* theta were significantly different from all the other values, however, theta estimates for *G. ashlocki* were not significantly larger than those for *G. collaris* or *G. vandykei*, but were larger than those for *G. galapagoensis* and *G. h. howdenae*. The population size estimates for *G. galapagoensis*, *G. collaris*, and *G. vandykei* were all not significantly different from each other. *G. h. howdenae* had the smallest theta estimate and was significantly different from all the other values. Population size estimates are derived from the amount of variation present; they are not demographic values and they are scaled by
the mutation rate. Nonetheless, they provide a useful initial comparison. We observe that the introduced species appears to have a much smaller theta and effective population size than its endemic counterparts.

**Figure 10. Comparison of estimates of mitochondrial Theta (θ) for all species.** Means of each of the twenty runs per species (derived from the two pair-wise calculations each species is involved in) were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).

Figure 11 shows the density distributions for population size estimates for three comparisons. As with the migration rate estimates, this shows how the IMa2 software assigns larger likelihood to a particular population size estimate: the peak in the graphs correlate with these estimates. From these comparisons we again see that the population sizes estimated for *G. ashlocki* and *G. conwayensis*, in both pairwise comparisons, were larger than those estimated for *G. h. howdenae*. 
**Figure 11.** Density distribution plots of mitochondrial Theta (θ) estimates for Santa Cruz species. The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

**Theta or population size parameters derived from nuclear sequences**

The population size estimates for all six species using nuclear sequences indicate that *G. galapagoensis* had the greatest mean population size. This value was not significantly different from that of *G. ashlocki* but was different from all the other population size estimates. *G. h. howdenae* had the smallest population size according to this estimate and was significantly different from all populations except *G. conwayensis* and *G. ashlocki*. The *G. conwayensis* population estimate was only significantly different from *G. galapagoensis* and *G. vandykei*, which was significantly different from all except *G. ashlocki* and *G. collaris*. The *G. collaris* population size estimate was only significantly different from *G. h. howdenae* and *G. galapagoensis*. The *G. ashlocki* size estimate was not significantly different from any other population size estimate (Fig. 12). In summary, nuclear estimates of effective population size were larger for some San Cristóbal endemics than for the endemic and introduced species in Santa Cruz. In the discussion we will explore the large range of
the estimates for *G. ashlocki*.

*Figure 12. Comparison of estimates of nuclear Theta (θ) for all species.* Means of each of the twenty runs per species (derived from the two pair-wise calculations each species is involved in) were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).

Figure 13 shows the density distributions for population size estimates for three comparisons. All the simulations indicated higher likelihood values at low values of Theta or population size values.
Figure 13. Density distribution plots of nuclear Theta (θ) values for Santa Cruz species. The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

Population migration rate (2Nm) derived from mitochondrial sequences for Santa Cruz Island

The 2Nm values convey an overall estimate of migration effect because it factors in population size of the donor population with the migration rate estimates. Figure 14 shows this estimated effect using mitochondrial sequences for the Santa Cruz comparisons (San Cristóbal comparisons are included in the IMFig analyses and representations). The highest estimate was for *G. ashlocki* to *G. h. howdenae*, which was statistically different from all other estimates. The lower values reported were from *G. ashlocki* to *G. conwayensis*, the reciprocal from *G. conwayensis* to *G. ashlocki*, and a slightly larger value from *G. h. howdenae* to *G. conwayensis*. These estimates were all not statistically different from each other but were statistically different from the other estimates. The 2Nm estimate from *G. conwayensis* to *G. h. howdenae* and *G. h. howdenae* to *G. ashlocki* were not statistically different from each other but were from the other estimates. In summary, here we still see reported larger estimates of mitochondrial gene flow between the introduced and the Santa Cruz highland endemic,
however, in the opposite direction (which was probably influenced by the inclusion of population sizes in the estimate).

Figure 14. Comparison of estimates of mitochondrial unidirectional effective gene flow (2Nm) in Santa Cruz Island. Means of combined estimates of population size and migration rates per Santa Cruz species comparison were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).

Figure 15 shows the density distributions for these six unidirectional estimates derived for three Santa Cruz comparisons using mitochondrial data. All the simulations show higher likelihood values at low values of 2Nm except for the comparison from *G. ashlocki* to *G. h. howdenae*. Again, the density distributions correspond well with the results from the summary statistics.
Figure 15. Density distribution plots of mitochondrial Theta ($\theta$) values for Santa Cruz species. The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

The graph generated using the IMFig software (Fig. 16) for one outfile for each of the pairwise mitochondrial comparisons for Santa Cruz indicate significant population migration values (2Nm) from *G. ashlocki* to *G. conwayensis*, *G. ashlocki* to *G. h. howdenae*, *G. h. howdenae* to *G. ashlocki*, and from *G. h. howdenae* to *G. conwayensis*. For the data generated for San Cristóbal, none of the values were indicated as being significant. At first glance, this could seem to contradict the trends observed with raw migration estimates m1 and m2. However, these rates incorporate the population size estimates into the 2Nm estimate and were statistically tested with alternative models and each time the model accounting for migration had a significantly higher likelihood.
Figure 16. Visual summary generated with IMFig of the parameters estimated with IMa2 for each of the mitochondrial pair-wise comparisons. Arrows indicate values of 2Nm and appear on the figure only if significant.
Population migration rate (2Nm) derived from nuclear sequences for Santa Cruz Island

Figure 17 displays the estimated 2Nm values (population migration rate) for the Santa Cruz comparisons using nuclear sequences. The highest 2Nm estimate was for *G. ashlocki* to *G. conwayensis*, which was statistically different from all other estimates except *G. conwayensis* to *G. ashlocki*, which was the second highest estimate and was statistically different from all the other estimates as well. The estimates for *G. ashlocki* to *G. h. howdenae*, *G. h. howdenae* to *G. ashlocki*, *G. conwayensis* to *G. h. howdenae*, and *G. h. howdenae* to *G. conwayensis* were all very low and were not statistically different from each other. Low values of effective gene exchange derived from nuclear sequences are not surprising when one of the members of the compared pairs is the recently introduced *G. h. howdenae*; the relevance of higher values for nuclear gene exchange between endemic species will be explored in the discussion.
Figure 17. Comparison of estimates of nuclear unidirectional effective gene flow (2Nm) in Santa Cruz Island. Means of combined estimates of population size and migration rates per Santa Cruz species comparison were compiled into a box plot displaying medians (solid line), means (+), and maximum and minimum values (whiskers).

Figure 18 shows the density distributions for three Santa Cruz comparisons using the nuclear data. Most of the estimates had the highest likelihood peaks at zero except for the comparisons between G. ashlocki and G. conwayensis in both directions. The estimate of effective gene flow from G. ashlocki into G. conwayensis displayed a wider range and therefore incorporated larger values of 2Nm estimates than in the other direction, however, the shape of the distribution is flat and not as well delimited as other comparisons
Figure 18. Density distribution plots of nuclear unidirectional effective gene flow (2Nm) for Santa Cruz species pairs. The probability density function for three runs for each species pair comparison. The x-axis indicates the range for the estimation of the parameter while the y-axis indicates the likelihood of each value during the run.

The IMFig generated figures (Fig. 19) from one sample run for each pairwise comparison using nuclear data found only two population migration rates (2Nm) to be significant: the Santa Cruz island comparisons of G. conwayensis to G. ashlocki and G. h. howdenae to G. ashlocki. The rest of the Santa Cruz comparisons and the San Cristóbal comparisons were not significant in this test.
Figure 19. Visual summary generated with IMFig of the nuclear parameters estimated with IMa2 for each of the pair-wise comparisons. Arrows indicate values of 2Nm and appear on the figure only if significant.
Discussion

Species boundaries

Species boundaries are defined within three main species concepts: the phenetic species concept, the biological species concept, and the phylogenetic species concept. The phenetic species concept draws species boundaries around groupings of phenotypically similar populations (groups whose composite morphological or behavioral traits are similar) (Michener and Sokal 1957, Cain and Harrison 1960, Sokal and Sneath 1963). The biological species concept defines species boundaries by the presence of gene flow between populations of one species and the absence of gene flow with any other population (Mayr 1942); those species are said to be reproductively isolated. The phylogenetic species concept groups those organisms that are placed as each-other’s closest relative in a phylogenetic tree and form a group with a common ancestor, also known as a monophyletic group or natural group. The maintenance of species boundaries is dependent upon the lack of gene flow between populations. In this sense, the biological species concept is the one that best addresses the process of isolation between populations. Reproductively isolating mechanisms include prezygotic isolating mechanisms and postzygotic isolating mechanisms. Prezygotic isolating mechanisms prevent the formation of the zygote by either preventing mating or fertilization and include habitat isolation, temporal isolation, behavioral isolation, and gametic incompatibility. Postzygotic isolating mechanisms interfere with either the survival or the overall fitness of the zygote in the current or even the next generation and include zygote death, nonviable or sterile F1 hybrids, and nonviable or sterile F2 hybrids (Dobzhansky 1937).

If these isolating mechanisms are removed or fail to prevent the formation of hybrids, gene flow will occur and the species boundary is suspect to change. Maintaining species boundaries can be important for the maintenance of endemic species; the genetic combinations present in
endemics are possibly unique and could become diluted when those boundaries cease to exist. For example, if one population moves into the habitat of the other population, habitat isolation is removed and the species boundary is at risk of dissolving due to gene flow if no other boundaries are present. Gene flow can occur during various steps along the history of a species: when it is first diverging (isolation in the face of gene-flow and in the presence of strong diversifying selection), during secondary contact with a species after divergence (hybridization), or all along during the divergence process and afterwards.

**Hybridization**

Hybridization occurs when two genetically distinct populations interbreed (Rhymer and Simberloff 1996). This definition does not include the taxonomic status of the populations, meaning hybridization includes the interbreeding of two populations from two different species and the interbreeding between two populations that are not considered separate species taxonomically but are distinct genetically. Because gene pools mix during hybridization, there is potential for the loss of genotypically distinct populations, especially in rare populations. Endemic populations, such as those of endemic *G. ashlocki* on Santa Cruz, are rare populations that would be at risk due to hybridization. This is especially prominent for *G. ashlocki* since it is a single island endemic and a highland specialist; there is a single population known for *G. ashlocki* at Los Gemelos in Santa Cruz. Hybridization can occur as a result of the loss of environmental heterogeneity due to the increase in gene admixture and gene flow (Seehausen et al. 2008). Hybridization can also lead to introgression, which is the backcrossing of F1 hybrids to one or both parent populations (Rhymer and Simberloff 1996) and can be a threat to the rare parent population. Rhymer and Simberloff (1996) make a distinction between hybridization and introgression because hybridization can threaten small populations even if gene pools do not mix: because of the allocation of mating
resources with partners from another species, the rare species diminishes its chances to maintain healthy population sizes.

Although hybridization can pose threats to rare populations, it can also increase variation, such as in the *Geospiza fortis* finches (Lamichhaney et al. 2015). Lamichhaney et al. (2015) found that past introgressive hybridization produced an increase in variation in beak size due to variation in the *ALX1* gene in the hybrids. This increased the amount of available food resources and influenced the adaptive radiation that occurred in the finches. Another example of introgressive hybridization increasing variation and causing adaptive radiation is in the *Heliconius* butterflies (Consortium 2012). The authors found evidence for hybrid exchange between three co-mimic species, *Heliconius melpomene, Heliconius timareta, and Heliconius elevatus*, at two genomic regions involved with mimicry.

Grant and Grant (1992) studied the occurrence and effect of hybridization by compiling data regarding the frequency of hybridization in bird species. They found that as many as 895 species of birds have produced hybrid offspring (9.2% of bird species). Because hybridization is rarely detected, Grant and Grant (1992) hypothesize that it may be more frequent than 9.2%. They also noted that hybridization occurs more frequently in the temperate zone and in ducks and geese than other bird species. When evaluating the fitness consequences of hybridization, hybrids can be genetically disadvantaged but hybrid superiority can also occur, which can result in new polyploid species (Grant and Grant 1992). Hybridization has had a major impact in plant biodiversity: 40% of plant species have arisen due to the creation of polyploid species as a result of hybridization (Ehrlich and Wilson 1991).

Introgression can be a source of new variation as well as a potential source of adaptive variation, in which adaptive variants are gained from other (donor) species when introgressed into
the (recipient) species. This is termed adaptive introgression (Hedrick 2013). Adaptive variation from introgression might have higher initial frequency than new adaptive mutations but lower frequency than standing variation, making the impact of adaptive introgression variation potentially intermediate (Hedrick 2013).

The advent of techniques that allow for the discovery and mapping of thousands of species-diagnostic SNPs (single nucleotide polymorphisms), such as RAD sequencing, is allowing for the better understanding of the effects of hybridization on native populations (e.g., fitness consequences), including those of vertebrate species (Hand et al. 2015) where hybridization was previously believed to be extremely unlikely (Dowling and Secor 1997). Using these techniques, the detection of winged, interspecific hybrid individuals in the *Zelandoperla* stonefly group raises the intriguing possibility that a previously flightless lineage could reacquire flight via introgression (Dussex et al. 2016).

*What does Theta (or 4Nu) mean and why do we use it?*

When modeling and evaluating the possibility of interspecies gene flow, as we do in this study, we need to consider that the gain of alleles in a particular population can also come from intrinsic processes such as mutation. The impact of the rate of mutation hinges on the availability of alleles to modify into new alleles, which in turn depends on the population size. So, the 4Nu values estimate population mutation rates. This estimate takes into account population size and the mutation rate per generation, and we use it as an indicator of population size.

*Effective population size and different genomes*

The effective population size is the number of individuals that an idealised population would need to have in order to produce similar numbers of offspring as the real population of interest. In some simple cases, this effective population size is equal to the number of breeding individuals in the
real population (Wright 1938). Effective population size of mitochondrial DNA is four times smaller than that of nuclear DNA due to its uni-parental inheritance. In animals, nuclear DNA is usually diploid and bi-parentally inherited whereas mitochondrial DNA is haploid and, with a few exceptions, maternally inherited. Because of these differences in ploidy and inheritance, the mitochondrial effective population size (Ne(mt)) is four times smaller than for nuclear loci (Birky et al. 1989), and thus more susceptible to the effects of genetic drift (Hoarau et al. 2004). For these reasons in the context of our study, the theta (or population size) comparisons that will be more informative are those between species using the same marker.

The effective population size values are modeled and estimated starting from the amount of variation present in the population sample. It is, therefore, important to consider this when analyzing the results. A small sample could result in low variation in demographically large populations or reflect actual low demographic size. In other words, sample size could potentially affect population size estimates because a larger sample size could provide a better estimate of the variation present in comparison to a smaller sample rate. Regardless of this caveat, these estimates are useful because in conjunction with migration estimates, they can determine the effective population migration rate of a population and thus the overall impact of gene exchange. **Nuclear versus mitochondrial sequences**

Nuclear and mitochondrial DNA can yield different results because of their different methods of inheritance. As mentioned above, nuclear DNA recombines and is inherited both from the paternal and the maternal inputs. Mitochondrial DNA, on the other hand, is only inherited maternally and does not recombine. This pattern of exclusive inheritance is well known as “maternal inheritance.” How the paternal mitochondria and mitochondrial DNA are eliminated from the cytoplasm of gametes or zygotes remains an enigma. Recently, a variety of mechanisms, including specific nuclease-dependent systems, ubiquitin–proteasome system, and autophagy have been shown to
degrade the paternal mitochondrial DNA or the paternal mitochondria themselves in order to prevent paternal mitochondrial DNA transmission (Sato and Sato 2013). The uniparental transmission implies that if there is a signal for hybridization, for example, by the female parent, mitochondrial DNA will be passed on undiluted to the hybrid offspring while the nuclear DNA will be mixed and diluted. As a result, the evidence of hybridization will be stronger in mitochondrial DNA than in nuclear DNA.

**Implications and context of mitochondrial introgression and population size**

Using the mitochondrial sequences, the m1 and m2 values, representing rates of unidirectional gene flow, were largest from the introduced *G. h. howdenae* population to the highland endemic *G. ashlocki* population. The mitochondrial sequences also indicated large migration values between the two San Cristóbal endemic populations *G. vandykei* and *G. collaris*. This finding may have resulted due to the small sampling size or may indicate actual high amounts of gene flow between the two populations of endemics. Small sample sizes can skew results because they are more prone to sampling error and, therefore, do not act as a true representation of the population (Whitlock and Schluter 2015).

Mitochondrial introgression has been found in insects such as *Pterostichus* ground beetles and *Neodiprion* sawflies. In the ground beetle example, phylogenetic incongruence between mitochondrial lineages and morphological identifications prompted the authors to suggest that interspecific hybridization and subsequent mitochondrial introgression from *P. habui* to *P. thunbergi* have occurred (Kosuda et al. 2016). To identify likely cases of recent and ancient introgression in *Neodiprion*, the authors hypothesize that shared hosts and/or pheromones facilitate hybridization, whereas disparate abundances between hybridizing species promote mitochondrial introgression (Linnen and Farrell 2007).

The mitochondrial theta estimate results demonstrated smaller population sizes for introduced
G. h. howdenae population compared to all of the endemic populations. Because the population size estimate is derived from the amount of variation present, this result could indicate that G. h. howdenae population has low variation compared to all the endemic populations and has comparable census population size or it could reflect actual small census size. Other mitochondrial data from the Sequeira lab (Sequeira et al. in prep, Sequeira et al. 2012) indicate less variation in G. h. howdenae compared to the other endemic species as well as continental populations of this species. The microsatellite data, however, collected from multiple populations of this species, does not show this low variation. It is possible that microsatellites accumulate alleles at a higher rate than mitochondrial sequences despite their nuclear origin and that many new alleles have been produced since the introduction of G. h. howdenae to the islands (Sequeira et al. in prep). A true low census estimate for the G. h. howdenae population would offer information on the extent of the effect an introduced population can have on the endemic populations, even with low population size.

Understanding the population size of a recently introduced species is paramount to estimating its invasive potential since the larger the size of the invasive colony, the more likely it will eventually become invasive (Crooks and Soulé 1999). In the case of G. h. howdenae, given our results, its impact could be independent of population size, which in turn poses more challenges for the management of this recent introduction.

Finally, the mitochondrial 2Nm values, which give an overall estimate of migration effect, show the highest, significant results from the endemic G. ashlocki population to the introduced G. h. howdenae population. These results indicate recent gene exchange between the endemic and introduced populations but differ in direction from the migration rates because they are weighed by the relative population size of the donor. Since G. ashlocki population estimates were larger than those of
G. h. howdenae, the effective gene exchange appears to have changed direction. In any case, we see evidence of endemic, introduced mitochondrial exchange. This also provides genetic evidence of the effects of recent introduced population expansion into the highlands. The “El Niño” events and subsequent “La Niña” droughts that affect the islands approximately every ten years, most often cause natural range expansions and mixtures due to range overlap between previously allopatric populations and species. The expansion of the range of the introduced G. h. howdenae and the endemic G. conwayensis into the highlands is an example of one of those natural mixtures. This gene exchange has possible genetic impacts on future generations: possible loss or gain of variation and loss or gain of adaptation in the endemic species (Hedrick 2013) as well as blurred species boundaries, which could lead to the complete loss of the identity of the endemic species. Other implications could include biological behavioral and mating changes such as less choosy males.

Gene exchange between introduced and wild populations has been documented in species of Viola (Krahulcova et al. 1996), smooth cordgrass (Daehler and Strong 1997) and tiger salamanders (Fitzpatrick et al. 2010). In the case of the salamander, the speed of spread of introduced alleles into an endangered species underscores the importance of genetic impact of an introduction even in the absence of ecological dominance.

Implications and context of nuclear gene exchange and population size

In summary, the nuclear results for the m1 and m2 values illustrated a low gene exchange estimate between the introduced population and all of the endemic populations. The migration estimates were higher for the other comparisons, with the highest (and statistically significant) being from G. ashlocki to G. conwayensis.

This nuclear DNA result for the m1 and m2 estimates varies from the result using mitochondrial DNA, in which the highest gene flow estimate occurred from the introduced population to the endemic
*G. ashlocki* population. A higher migration estimate for the introduced to the endemic population using mitochondrial DNA is to be expected, however, due to the nature of inheritance (effects of hybridization will appear more prominently in the mitochondrial data and will take longer for these effects to show up in the nuclear data). The introduced *G. h. howdenae* population colonized Santa Cruz sometime during the human colonization period of 1832-1959 (Mok et al. 2014). This is fairly recent on a biological time scale, and the point of contact between the introduced population and the highland endemic occurred even more recently. We can therefore presume that the time of hybridization has not been long and may explain why it is not seen in the nuclear DNA. Analyzing the mitochondrial DNA and nuclear DNA results together tells us that hybridization is occurring between the introduced and endemic populations but is not showing up as greatly in the nuclear DNA yet, perhaps due to the recent timing of the contact and hybridization.

The nuclear theta estimate resulted in *G. galapagoensis*, the lowland endemic population from San Cristóbal, with the greatest mean population estimate. This high population estimate for *G. galapagoensis* is reasonable due to the fact that it is one of the older species and is found in more localities on the San Cristóbal island than the other species (Sequeira et al. 2008b). *G. ashlocki* had the second largest mean population estimate, however, this estimate was accompanied by a large standard deviation and is therefore not as remarkable. Like in the mitochondrial DNA result, *G. h. howdenae* had the lowest mean population estimate. The other populations had similar mean estimates to each other. It is logical that *G. h. howdenae* had the lowest mean estimate because it is the introduced population. As discussed with the mitochondrial data, a low census estimate coupled with a high impact estimate provides proof that an introduced population can still have a large potential effect on endemic populations despite low population size.
The highest, significant nuclear $2N_m$ estimates for the Santa Cruz populations were for $G. ashlocki$ to $G. conwayensis$ and from $G. conwayensis$ to $G. ashlocki$, as indicated by both the box plots and density distribution plots. The other population estimates were all very low and not statistically different from each other as well as not significantly different from zero according to the test derived by Nielsen and Wakeley (2001). These values estimate the overall effect of gene migration, which indicate that the highest migration effect is found between the two endemic populations $G. ashlocki$ and $G. conwayensis$. According to phylogenetic estimates (Sequeira et al. 2008b), $G. ashlocki$ evolved from $G. conwayensis$ in an intra-island speciation event, which means that this overall migration effect could be due to recent gene flow or past gene flow. Past gene flow could have occurred during divergence. When gene flow during divergence occurs, the two populations undergo divergence despite gene flow due to selection preferences or niche occupancy. On the other hand, recent gene flow would indicate that hybridization is occurring as a result of secondary contact after divergence.

Examples of gene exchange during divergence also explored using the IM model include cave salamanders (Niemiller et al. 2008), fresh-water fishes (Kotlik et al. 2008), and stickleback fishes (Berner et al. 2009). In these three cases it was possible to distinguish between ongoing gene flow and that occurring after secondary contact. Later studies warn against relying solely on the posterior distribution of mean migration time provided by IM to distinguish between the two scenarios (Niemiller et al. 2010) and instead suggest complementing with methods of approximate Bayesian computation (Lucas et al. 2016).

**Unanswered questions**

There are still questions that, if answered, would provide further insight on the magnitude, extent, and long-term consequences of this gene exchange between the introduced and
endemic populations, as well as that between the two endemic populations. Discovering more about the range expansion of the introduced population into the highlands would provide more information about the timing of this gene exchange as well as the impacts of it. Determining how much of the population has moved into the highlands and tracking the population’s movements could provide insight into correlations with determinants or precursors of increased gene flow. Looking at correlations between environmental data and population movement into the highlands could also provide insight into the timing and reason for this gene flow. Regarding the exchange between the Santa Cruz endemic populations, it would be beneficial to determine the timing of the event and if and how this exchange lines up with the divergence time of the two species. It is recommended that future studies follow these questions and directions in order to uncover more about this gene exchange.

_Broader implications_

Hybridization changes and blurs species boundaries. It is, therefore, essential to study hybridization in order to understand speciation and extinction. Hybridization can cause the loss of valuable adaptations for rare or endemic populations, which can eventually lower population sizes to dangerously low levels. Hybridization and introgression can also increase variation and create new species as illustrated in the Darwin finches (Lamichhaney et al. 2015). Because hybridization is not detected frequently, it is essential to continue studies on it to understand how fluid boundaries can affect, either positively or negatively, the ability of populations to respond to historical landscape changes and other environmental changes. The *Galapaganus* weevil system, therefore, can act as a system to compare future island introductions and their impact on boundaries of island endemic species to. Within the Galapágos, *G. h. howdenae* has been introduced only in Santa Cruz to date. Even though the existing barriers to its dispersal to other islands are effective, they are not insurmountable
(Roque-Albelo et al. 2006). As a result, *G. h. howdenae* adults could disperse, via flight or transportation on plants that are exported from Santa Cruz, to other islands such as Isabela and potentially encounter other endemic *Galapaganus*.

This system also suggests that an introduced population can have a high impact despite low population size or low genetic variation. An introduced population, therefore, has the ability of making a large impact on endemic and rare populations independent from their competitive abilities.
### Appendix

**Appendix 1: Run and prior details.** Run conditions, prior values, ESS values, and run duration for all 12 comparisons (mitochondrial and nuclear).

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Genome</th>
<th>ESS Values (Log[P], τ0, tmrca0)</th>
<th>Command Line Priors</th>
<th>Example Run Time (x10)</th>
<th>Comparison</th>
<th>Genome</th>
<th>ESS Values (Log[P], τ0, tmrca0, 1, 2, 3)</th>
<th>Command Line Priors</th>
<th>Example Run Time (x10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. h. howdenae &amp; G. conwayensis</td>
<td>Mitochondrial (COI, COII, 12S)</td>
<td>12, 27, 39</td>
<td>q1 100</td>
<td>4 hours, 37 minutes, 51 seconds</td>
<td>G. h. howdenae &amp; G. conwayensis</td>
<td>Nuclear (ITS, Argk, EF1alpha)</td>
<td>23, 141524, 21, 17, 28, 21</td>
<td>q1 100</td>
<td>6 hours, 36 minutes, 16 seconds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 45, 38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23, 149065, 21, 16, 28, 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 37, 28</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td>23, 143760, 21, 17, 29, 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12, 335, 44</td>
<td>l200000</td>
<td></td>
<td></td>
<td></td>
<td>225, 1461532, 211, 163, 281, 196</td>
<td>b1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 56, 36</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td>22, 141305, 21, 16, 27, 20</td>
<td>m1 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12, 110, 47</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td>23, 150869, 21, 18, 27, 22</td>
<td>m2 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>13, 71, 44</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td>22, 144817, 20, 16, 27, 19</td>
<td>t450</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 63, 33</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td>22, 144798, 21, 17, 27, 21</td>
<td>p45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 834, 76</td>
<td>q1 100</td>
<td>5 hours, 46 minutes, 6 seconds</td>
<td>G. h. howdenae &amp; G. ashlocki</td>
<td>Nuclear (ITS, Argk, EF1alpha)</td>
<td>19, 28, 14, 18, 18, 21</td>
<td>q1 100</td>
<td>1 hour, 10 minutes, 22 seconds</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 172, 61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19, 28, 14, 20, 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 1007, 67</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td>19, 28, 14, 19, 21, 19, 21</td>
<td>b1000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 282, 92</td>
<td>l200000</td>
<td></td>
<td></td>
<td></td>
<td>20, 28, 14, 18, 19, 22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 653, 73</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td>20, 29, 14, 21, 18, 22</td>
<td>m1 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 187, 78</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td>19, 28, 14, 18, 21</td>
<td>m2 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>11, 723, 39</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td>20, 28, 15, 19, 19, 22</td>
<td>t450</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10, 381572, 87</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td>19, 28, 14, 17, 18, 21</td>
<td>p45</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19, 26, 14, 17, 19, 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitochondrial (COI, COII, 12S)</td>
<td>Nuclear (ITS, Argk, EF1alpha)</td>
<td>26 hours</td>
<td>51 minutes, 44 seconds</td>
<td>24 minutes,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>----------</td>
<td>------------------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. conwayensis &amp; G. ashlocki</td>
<td>21, 661490, 39617</td>
<td>11, 135500, 12, 12, 13, 14</td>
<td>q1 100</td>
<td>102, 749, 125, 122, 126, 127</td>
<td>q1 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19, 416646, 65169</td>
<td>56, 2129, 65, 70, 64, 65</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40, 156, 21</td>
<td>58, 1739, 66, 69, 64, 74</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20, 402161, 42078</td>
<td>57, 774830, 71, 69, 72, 75</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17, 662403, 43539</td>
<td>58, 228, 74, 85, 77, 77</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48, 163, 18</td>
<td>57, 756438, 65, 65, 66, 68</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17, 625512, 39971</td>
<td>58, 742, 71, 70, 72, 74</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>16, 735, 39081</td>
<td>57, 1113, 67, 64, 67, 68</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>23, 646502, 38164</td>
<td>78, 4662, 75, 61, 78, 90</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19, 417058, 41527</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. galapaganus &amp; G. vandykei</td>
<td>14, 139, 101</td>
<td>19, 295332, 29, 41</td>
<td>q1 100</td>
<td>19, 293941, 30, 44</td>
<td>q1 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 165, 52</td>
<td>19, 412554, 30, 44</td>
<td>b1000</td>
<td>18, 412554, 30, 44</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 84144, 98</td>
<td>19, 285797, 33, 42</td>
<td>b1000</td>
<td>19, 285797, 33, 42</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 217, 39</td>
<td>19, 41817, 27, 41</td>
<td>l200000</td>
<td>19, 41817, 27, 41</td>
<td>l200000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 67135, 306</td>
<td>18, 406568, 28, 46</td>
<td>m1 10</td>
<td>18, 406568, 28, 46</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13, 412920, 323</td>
<td>18, 415543, 30, 44</td>
<td>m2 10</td>
<td>18, 415543, 30, 44</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15, 416722, 72</td>
<td>19, 416222, 32, 49</td>
<td>t450</td>
<td>19, 416222, 32, 49</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 392619, 185</td>
<td>19, 9522, 30, 46</td>
<td>p45</td>
<td>19, 9522, 30, 46</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 62886, 184</td>
<td>19, 426115, 32, 46</td>
<td>p45</td>
<td>19, 426115, 32, 46</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14, 279, 142</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. galapaganus</td>
<td>12, 214, 227</td>
<td>17, 102, 36, 47</td>
<td>q1 100</td>
<td>17, 47, 32, 39</td>
<td>q1 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12, 146, 179</td>
<td>17, 47, 32, 39</td>
<td>q1 100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

56
<table>
<thead>
<tr>
<th>&amp; G. collaris</th>
<th>12S</th>
<th>12, 79604, 271</th>
<th>b1000</th>
<th>28 seconds</th>
<th>&amp; G. collaris</th>
<th>EF1alpha</th>
<th>17, 203, 35, 46</th>
<th>b1000</th>
<th>9 seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>12, 65309, 144</td>
<td>l200000 (4), l2000000 (6)</td>
<td></td>
<td></td>
<td></td>
<td>17, 414535, 39 50</td>
<td>l200000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119, 3091, 1879</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td>17, 120, 37, 49</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119, 2170, 1653</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td>15, 64, 29, 38</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119, 2705, 1525</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td>18, 417904, 44, 55</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119, 2415, 1687</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18, 91, 36, 47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>119, 3296, 1690</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td>17, 409955, 39, 56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120, 2592, 2123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17, 808, 39, 47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. vandykei &amp; G. collaris</td>
<td>Mitochondrial (COI, COII, 12S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 135, 36</td>
<td>q1 100</td>
<td></td>
<td></td>
<td></td>
<td>22, 222057, 43, 30, 28</td>
<td>q1 100</td>
<td></td>
<td>44 minutes, 22 seconds</td>
<td></td>
</tr>
<tr>
<td>19, 101, 37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24, 240964, 42, 34, 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 98, 41</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
<td>24, 235926, 65, 51, 26</td>
<td>b1000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20, 152, 37</td>
<td>l2000000 (5), l20000000 (5)</td>
<td></td>
<td></td>
<td></td>
<td>22, 228898, 30, 29, 24</td>
<td>l200000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19, 89, 33</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
<td>23, 233956, 39, 30, 27</td>
<td>m1 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>195, 1001, 370</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
<td>24, 234752, 31, 35, 24</td>
<td>m2 10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>197, 982, 369</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
<td>23, 224562, 11, 11, 23</td>
<td>t450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>196, 972, 363</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24, 230813, 30, 9, 26</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>194, 1051, 359</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
<td>24, 235944, 38, 30, 24</td>
<td>p45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>197, 991, 382</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25, 237149, 49, 40, 25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Daehler, C. C., and D. R. Strong. 1997. Hybridization between introduced smooth cordgrass (Spartina alterniflora; Poaceae) and native California cordgrass (S-foliosa) in San Francisco Bay, California, USA. American Journal of Botany 84:607-611.


Hey, J. 2007. Introduction to the IM and IMa computer programs. Department of Genetics, Rutgers University

Hey, J. 2009. Using the IMa Program. Department of Genetics, Rutgers University


Hey, J. 2011b. Using the IMfig program. Department of Genetics, Rutgers University.


Kosuda, S., K. Sasakawa, and H. Ikeda. 2016. Directional mitochondrial introgression and
character displacement due to reproductive interference in two closely related Pterostichus ground beetle species. J Evol Biol.


Seehauscn, O., G. Takimoto, D. Roy, and J. Jokela. 2008. Speciation reversal and biodiversity
dynamics with hybridization in changing environments. Mol Ecol 17:30-44.


Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance


