

# Spatial and host related genomic variation in partially sympatric cactophagous moth species

Daniel Poveda-Martínez<sup>1</sup>, Laura Varone<sup>2</sup>, Malena Fuentes Corona<sup>2</sup>, Stephen Hight<sup>3</sup>, Guillermo Logarzo<sup>2</sup>, and Esteban Hasson<sup>4</sup>

<sup>1</sup>Instituto de Ecología, Genética y Evolución de Buenos Aires

<sup>2</sup>Fundación para el Estudio de Especies Invasivas

<sup>3</sup>U.S. Department of Agriculture-ARS

<sup>4</sup>CONICET-UNIVERSIDAD DE BUENOS AIRES

March 15, 2021

## Abstract

Surveys of patterns of genetic variation in natural sympatric and allopatric populations of recently diverged species are necessary to understand the processes driving intra and interspecific diversification. The South American moths *Cactoblastis cactorum*, *Cactoblastis doddi* and *Cactoblastis bucyrus* are specialized in the use of cacti as host plants. These species have different distribution ranges and differ in patterns of host plant use. However, there are areas in which their ranges overlap, as in northwestern Argentina, where they are largely sympatric. Using a combination of genome-wide SNPs and mitochondrial data we investigated the phylogeographic patterns of these cactophilic moths and searched for footprints of hybridization. Additionally, we evaluated a moth population feeding on *Cleistocactus baumannii*, a plant never reported as a host for the genus. We identified three well delimited species and detected signs of historical gene flow. Our survey also revealed intraspecific geographic structure in both *C. doddi* and *C. cactorum* and showed that the moth population feeding on *C. baumannii* may be considered as conspecific to *C. bucyrus*. Overall, our results indicated historical events of genetic interchange occurred in *Cactoblastis* cactophagous moths, but host plants likely played an important role during divergence limiting gene flow across species.

## Introduction

Patterns of intra-and interspecific genetic diversity in closely related species are of special interest in evolutionary biology. Furthermore, the study of population genetic structure patterns in sympatric and allopatric populations of recently divergent species can provide information on the processes that drive diversification at the intra- and interspecific levels and the mechanisms that maintain species boundaries.

Closely related species of herbivorous insects with overlapping distribution ranges can be considered good candidates of ecological speciation by means of divergent selection (Nosil, 2012). In this context, the roles of mating timing (Santos et al., 2007), endosymbiotic infections (Fрати et al., 2004), and host plant associations (Poveda-Martinez et al., 2020; Prowell, McMichael, & Silvain, 2004) have been reported and referred to as processes shaping the fate of co-occurring species. Other processes, such as introgression following interspecific hybridization, may also play an important role in young sympatric species (Martin et al., 2013; Tonzo, Papadopoulou, & Ortego, 2020; Valencia-Montoya et al., 2020). Moreover, introgressive hybridization is a potential source of genetic variation that may prompt adaptive changes in the recipient species, thus playing an important role in adaptive radiations (Pardo-Diaz et al., 2012). Eventually, such adaptive changes may give rise to new behaviors or characteristics in the adapted species that could create an unprecedented pest.

Phytophagous insects are excellent models for the study of the genetic and ecological basis of intra and interspecific divergence (Schoonhoven, Van Loon, van Loon, & Dicke, 2005). Host shifts can lead to ecological specialization and diversification since the acquisition of new host plants often involves the challenge of facing chemically diverse environments, new food sources, new mating environments, and modified parasitoid interactions; factors that enable the coexistence of multiple species in sympatry (Forbes et al., 2017; Funk, Nosil, & Etges, 2006). Moreover, evidence supporting the hypothesis of a relationship between herbivory and insect diversity is abundant, such as the greater species richness in specialist herbivores compared to non-herbivore-related taxa, and the link between host shifts and speciation in several herbivorous groups (Funk et al., 2006; Simon et al., 2015; Wiens, Lapoint, & Whiteman, 2015).

The genus *Cactoblastis* consists of herbivorous moths native to southern South America specialized in the use of Cactaceae as breeding and feeding resources. The degree of adaptation of these moths to cactus hosts is striking since eggs are stacked one on top of the other (eggstick), mimicking the appearance of a cactus spine (Dodd, 1940). A feature that may play a role in the avoidance of natural enemies. Eggsticks are attached to the areola or the spine of a cactus plant and once the eggs hatch, the larvae burrow inside the stems of prickly pears (*Opuntia* spp.) or columnar cacti and start to feed (Varone et al., 2019). At least three *Cactoblastis* species, *Cactoblastis cactorum* Berg, *Cactoblastis bucyrus* Dyar, and *Cactoblastis doddi* Heinrich, can be found in several biogeographic areas of Argentina (McFadyen, 1985). These species have different geographic ranges and patterns of host plant use. However, there are areas of northwestern Argentina (NWA) in which the ranges of *Cactoblastis* spp. overlap (McFadyen, 1985). These areas consist of arid forests in the limit between the Puna, Prepuna, Chacoan and Monte biogeographic provinces (Morrone, 2014). In this area, the eastern slopes of pre-Andean mountain ranges of southern South America, two native cactus species are the most predominant in the landscape: *Opuntia sulphurea* Gillies ex Salm-Dyck and the emblematic giant cactus “cardón” *Trichocereus atacamensis* var *pasacana* Weber ex Rümpler (*T. atacamensis* from hereafter). However, patches of the exotic semi cultivated *Opuntia ficus-indica* Miller can also be found around some households in the area (Varone, Logarzo, Briano, Hight, & Carpenter, 2014).

*Cactoblastis cactorum* has a wider diet breadth and a more extended geographic distribution than the other two *Cactoblastis* species. Its diet includes several native species of the genus *Opuntia* spp. and the exotic *O. ficus-indica* (Varone et al., 2012; Varone et al., 2014). *Cactoblastis cactorum* can be found in Chacoan, Pampean and Monte biogeographic provinces. The moth is considered an important pest for prickly pear fruit production in its native distribution range where *O. ficus-indica* crops are an essential component of familiar subsistence (Zapata et al., 2005). Moreover, *C. cactorum* is an invasive species in both the United States and the Caribbean where it is also an important pest threatening *Opuntia* spp. biodiversity and prickly pear fruit production (Hight & Carpenter, 2009). In Australia and South Africa, *C. cactorum* was introduced as a biological control agent of invasive *Opuntia* spp. that were imported in the mid-1700s for use as natural agricultural fencing and a failed attempt to establish a cochineal dye industry (Raghu & Walton, 2007). *Cactoblastis doddi* has a more limited geographic distribution and diet breadth. It has been reported feeding on the native prickly pear, *O. sulphurea*, a predominant cactus species in the Monte, Puna and Prepuna biogeographic provinces from central western to NWA (McFadyen, 1985; Zimmermann, Erb, & McFadyen, 1979). *Cactoblastis doddi* was reported feeding on *O. ficus-indica*, but the records are regarded as uncertain (McFadyen, 1985). The third species of interest is *C. bucyrus*, a species distributed from central western Argentina to NWA. Recent surveys confirmed the use of *T. atacamensis* as a common host of *C. bucyrus* in NWA (Barbarich, Otegui, Esteven, Soto, & Varone, 2021) but also has been reported feeding on species of the genera *Echinopsis* and *Denmoza* (Heinrich, 1939; Mann, 1969). Larvae of a cactophilic moth that appeared different from *C. doddi*, *C. cactorum* or *C. bucyrus* were recently collected on the columnar cactus *Cleistocactus baumannii* Lem. Although *Cleistocactus* sp. was reported by Mann (1969) as a host of *C. cactorum*, it is doubtful that the insect was *C. cactorum* since all other hosts of *C. cactorum* are in the genus *Opuntia* in this native range, and morphological differences are present regarding the eggstick and markings on adults and larvae (Arce & Neder, 1999; Mann, 1969; McFadyen, 1985; Varone et al., 2014). This new findings could expand the diet breadth of the already described species and/or the number of *Cactoblastis* moth species, since changes in diet have been reported as a factor that drives diversification in

Lepidoptera (Funk et al., 2006).

Since there are uncertain features regarding biogeographic distribution and patterns of plant host use of *Cactoblastis* spp., we investigated phylogeographic patterns and phylogenetic relationships in this trio of closely related cactophilic moths using a combination of genome-wide SNPs and mitochondrial data. We also evaluated the status of the moth population feeding on *C. baumannii*, and searched for footprints of historical and contemporary interspecific gene flow between populations. We included pairs of sympatric and geographically distant populations of each *Cactoblastis* spp. This sampling scheme allowed us to investigate patterns of intra and interspecific genetic diversity and genetic admixture within and between species. Our genomic data revealed historical gene flow occurring between species, however, our hypothesis of ongoing hybridization in sympatric populations was rejected, suggesting that host plants played an important role during divergence limiting gene flow across species even after secondary contact.

## Methods

### Taxon sampling and DNA extraction

We collected 70 individuals from at least two populations of each of three species feeding on the host plants reported as part of their diet in the native range. Samples of sympatric populations of the three species were obtained in NWA (Figure 1). In this region, *C. cactorum* was collected on *O. ficus-indica* and *Opuntia quimilo* Schum, *C. doddi* on *O. sulphurea*, and *C. bucyrus* on *T. atacamensis* (Table 1). In our experimental design we also included allopatric populations of *C. doddi* and *C. cactorum* from central Argentina feeding on *O. sulphurea* in Mendoza Province and *Opuntia megapotamica* Arechav. in La Pampa Province, respectively. The locations of these collecting sites were largely allopatric relative to the populations sampled in NWA. In addition, we included 10 individuals collected on equal numbers of plants identified as *C. baumannii*, a host plant that has not been reported for any species of the genus *Cactoblastis* in Santiago del Estero Province.

Due to the difficulty of distinguishing these species in the field, individuals collected were assigned to each species based on the host plant on which they were collected, following McFadyen (1985) and Varone et al., (2014). Finally, we collected two individuals of a related cactophilic moth, *Tucumania* sp. from Formosa province collected on *Opuntia anacantha* Spag. to be used as an outgroup.

Eggsticks and I-IV instar larvae were collected in the field. A single individual larva per plant was collected and included in our genetic survey. To avoid the inclusion of siblings, a maximum of five plants were sampled per location site. When eggsticks were collected they were maintained in the lab under constant temperature (25 °C) and humidity (70 %), and after hatching they were reared until they reached the fourth instar larval stage to obtain a high DNA yield. Larvae were maintained on the respective cactus plant species on which they were collected in the field. Once the larva reached the fourth instar in the lab it was individually transferred to a vial containing 100% ethanol and stored at -20° C until DNA extraction.

Before genomic DNA extraction, each larva was dissected to remove any food residues and <30 mg of clean tissue harvested for use in DNA extraction using Qiagen DNeasy Blood & Tissue Kit following manufacturer's instructions (Qiagen, Inc., Valencia, California, USA). DNA was quantified using Qubit 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA) and quality was assessed in a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA).

### ddRAD libraries, reference genome assembly and sequencing

Double digest Restriction-site Associated DNA (ddRAD) libraries were prepared following Peterson et al (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). *NspI* and *MboI* restriction enzymes were selected as the best combination of enzymes to obtain the maximum number of loci. Fragmented DNA was purified and ligated to barcoded adapters. Samples were pooled on multiplexing batches and bead purified. For each pool, targeted fragments distribution was collected on BluePippin instruments (Sage Instruments Inc., Freedom, CA, United States). The gel eluted fraction was amplified with oligo primers that introduce TruSeq indexes and subsequently bead purified. Libraries were processed with Illumina cBot for cluster generation on the

flow-cell following manufacturer’s instructions and sequenced in paired-end 125 bp mode on a HiSeq2500 instrument (Illumina, San Diego, CA).

A reference draft genome was assembled using DNA of an additional *C. cactorum* specimen. We used a single adult male to extract DNA. Since adults do not feed, this procedure allowed us to obtain clean DNA, and to use it to filter exogenous reads of food scraps from the output of the sequencing of the ddrad libraries.

To build the reference genome (CactoFuEDEI.v1.fa), 270 nanograms of genomic DNA were sheared using Bioruptor with 3 cycles 15” ON - 90” OFF and purified with 0.77 volumes of beads. Celero DNA-Seq kit (NuGEN, San Carlos, CA) was used for library preparation following manufacturer’s instructions. The library was then sequenced on NovaSeq 6000 in paired-end 150 bp mode. The *de novo* assembly was accomplished with CLC Genomics Workbench v7.1 using default parameters and minimal contig length of 200bp.

### ddRAD filtering and mitochondrial sequencing

Raw reads quality scores were checked using FastQC v0.11.15. Demultiplexing, filtering, assembly and variant calling were performed using ipyrad v0.9.59 (Eaton & Overcast, 2020). Demultiplexing was done using the unique barcode and adapter sequences. Samples’ reads were filtered using the stricter filter for Illumina adapters, and filtered to clean up the edges of poor quality reads. A *Reference* based assembly method was implemented using the reference draft genome CactoFuEDEI.v1.fa. We allowed 20% of SNPs per locus, and polymorphic sites occurring across a maximum of 25% of samples. Ipyrad was run to obtain four datasets, each one controlling 25%, 50%, 80% and 90% of the minimum number of samples per locus (*msl*), respectively.

A fragment of the mitochondrial gene encoding for the Cytochrome oxidase subunit I (*COI*) was amplified using primers C1-J-2183 (Simon et al., 1994) and PatII (Caterino & Sperling, 1999). The PCR reaction components and PCR cycling protocol followed Marsico et al (Marsico et al., 2011). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany) and sequenced using Sanger method at Macrogen Inc. (Seoul, South Korea).

### Intra- and interspecific genetic admixture and diversity

Clustering analyses including the three sympatric populations from NWA and allopatric populations of *C. cactorum*, *C. doddi*, and the moth population found feeding on *C. baumannii*, were implemented using the sparse nonnegative matrix factorization approach (sNMF) (Frichot, Mathieu, Trouillon, Bouchard, & Francois, 2014). This method estimates proportions of individual genomes originating from K ancestral populations (individual ancestry coefficients), and the corresponding ancestral genotype frequencies. This approach was chosen to assign individuals to the recognized species, to estimate proportions of intra-and interspecific admixture. sNMF was implemented in R using the LEA package (Frichot & Francois, 2015). Ancestry coefficients were calculated for K values ranging from 2 to 8 for sympatric and allopatric populations, with 100 repetitions for each value of K. The optimal K value was assessed using the cross-entropy criterion.

Patterns of nuclear diversity for each population were characterized using observed ( $H_o$ ) and expected heterozygosities ( $H_e$ ), and allele richness (Ar) for ddRAD data, calculated with R package *diveRsity* (Keenan, McGinnity, Cross, Crozier, & Prodohl, 2013), and *hierfstat* (Goudet, 2005) packages available in R ([www.r-project.org](http://www.r-project.org)). Pairwise  $F_{st}$  were calculated using Weir and Cockerham (Weir & Cockerham, 1984) method. Patterns of mtDNA variation were estimated using nucleotide diversity ( $\pi$ ), haplotype diversity ( $H_d$ ), number of haplotypes (H), and pairwise  $F_{st}$  between populations using DNAsp v6 (Rozas et al., 2017). A haplotype Network was constructed using the Median Joining algorithm implemented in PopArt v1.7.1 (Leigh & Bryant, 2015).

### Phylogenetic inference and introgression tests

Bayesian inference (BI) was used to estimate phylogenetic relationships among populations based on ddRAD and mtDNA datasets. For ddRAD we used the concatenated dataset including non-variant sites in order

to avoid acquisition bias resulting in overestimation of branch lengths and potentially wrong topological inference as recommended by Leaché, Banbury, Felsenstein, De Oca, & Stamatakis (2015). MrBayes v3.2.6 (Ronquist et al, 2012) was used for BI using the HKY as the best-fit model. Two independent runs were performed with four Markov Chain Monte Carlo, each run of 10 million generations with samplings every 1000th generation. Additionally, we estimated divergence times based on mtDNA, using a substitution rate of 0.017 substitutions per site per million years according to Papadopoulou, Anastasiou, & Vogler (2010) with the Yule Process as a tree prior and the strict clock model using BEAST2 (Bouckaert et al., 2014). Additionally, a species tree was estimated using the coalescent-based methods, SNAPP (Bryant, Bouckaert, Felsenstein, Rosenberg, & RoyChoudhury, 2012). Monomorphic sites and sites with no data were filtered resulting in a new matrix of 1,688 SNPs. We used the default model parameters in SNAPP for U and V equal to one, and we ran the analysis for 10,000,000 MCMC generations, sampling every 1,000 generations (Bouckaert et al., 2014; Drummond, Suchard, Xie, & Rambaut, 2012).

Since *C. cactorum*, *C. doddi* and *C. bucyrus* had partially overlapping geographic ranges, introgressive hybridization may have occurred between these closely related species in areas of sympatry. We used the four-taxon ABBA/BABA tests based on the D-statistic (Durand, Patterson, Reich, & Slatkin, 2011) to test for introgressive hybridization between species. For this method we used the largest dataset (msl= 25%) which maximizes the amount of data available as recommended by Eaton (Eaton & Overcast, 2020). The D statistic was used to test the null hypothesis of no introgression ( $D = 0$ ) between P3 and P1 or P2. D values significantly different from zero (Z-score  $[?] \geq 3$ ) indicated gene flow between P1 and P3 ( $D < 0$ ) or between P2 and P3 ( $D > 0$ ). We performed ABBA/BABA testing according to the scheme delineated in Supplementary Table 4. We assigned to P1 and P3 the sympatric populations of the species tested (A-B: *C. bucyrus* and *C. doddi* ; C-D: *C. cactorum* and *C. doddi* ; E-F: *C. bucyrus* and *C. cactorum* ), and as P2 the conspecific allopatric population of P1 following Yamasaki et al (2020) and Tonzo et al (2020). As P4, we used the allopatric population of the third species not considered in the test. ABBA/BABA tests were implemented using ipyrad analysis tools with 1,000 bootstrap replicates to obtain the standard deviation of the D-statistic (Duran et al., 2011, Eaton & Overcast, 2020).

### Demographic inference of gene flow by coalescent simulations

Interspecific scenarios of gene flow occurring before and after intraspecific divergence, as well as other relevant parameters such as divergence times, were tested using a coalescent-based modelling approach implemented in Fastsimcoal2 (Excoffier, Dupanloup, Huerta-Sanchez, Sousa, & Foll, 2013), based on the site frequency spectrum (SFS) of ddRAD data. The folded joint SFS was calculated considering a single SNP per locus using the Python script written by Isaac Overcast and available at GitHub (<https://github.com/isaacovercast/easySFS>) resulting in a matrix of 1688 unlinked SNPs. The effective population size ( $N_e$ ) for the population of Santiago del Estero was estimated to be fixed in the models.  $N_e$  was calculated according to the equation  $N_e = \pi/4\mu$  (Lynch & Conery, 2003). Nucleotide diversity ( $\pi = 0.0012$ ) was estimated using variant and invariant sites with DNAsp v12.03 (Rozas et al., 2017), and considering a mutation rate ( $\mu$ ) of  $2.90 \times 10^{-9}$  estimated for the butterfly *Heliconius melpomene* (Keightley et al., 2015) (Lepidoptera: Nymphalidae). Alternative models of gene flow were tested considering the sympatric and allopatric population of the three moth species. All models tested considered that the first interspecific splitting event involved *C. cactorum* followed by a second split separating *C. doddi* and *C. bucyrus* , and three independent intraspecific splits separating both populations of each species (Supplementary Figure 2). First, model 0 considered a scenario without gene flow; models 1-4 considered interspecific gene flow before intraspecific divergence; models 5-8 considered historical interspecific gene flow before intraspecific divergence, involving sympatric populations by pairs, and a single model considered the full scenario of interspecific gene flow. Scenarios considering ongoing gene flow were not tested because clustering and genetic population analyses did not show evidence supporting these patterns. From the models, we estimated demographic parameters as  $N_e$ , the divergence time and admixture time as well as different rates of gene flow (Supplementary Figure 2). Models were run independently using 50 replicates for each model, each model with 40 cycles of the Brent algorithm and 100,000 simulations. The best demographic model was identified using Akaike’s information criterion (AIC). To better compare between models, the AIC values were rescaled in terms of AIC differences ( $\Delta_i$ ) according to the formula:

$\Delta_i = AIC_i - AIC_{min}$ . A model with a  $\Delta AIC$  value of 0 and the highest AIC weight  $\omega_i$  served as the best model. Finally, a parametric bootstrapping approach was used to construct 95% confidence intervals of the estimated parameters running 100 bootstrap replicates using initialized values from the best model (Excoffier et al., 2013).

## Results

### Draft genome assembly, ddRAD and mtDNA data

To investigate intra and interspecific genetic structure and diversity, patterns of gene flow, and to untangle the phylogenetic relationships in *Cactoblastis* moths, a total of 70 specimens from 12 locations were genotyped for ddRAD (available at NCBI as BioProject PRJNA666743). The average number of paired-end reads per individual was  $1.7 \times 10^6$ . Sixteen individuals were excluded from the analysis due to low read numbers ( $< 3.0 \times 10^4$  reads). Also, 67 individuals from the same 12 locations were sequenced for a fragment of the mitochondrial Cytochrome oxidase I gene (*COI*) (available at Genbank MW064275-MW064313) (Supplementary Table 1). Samples were collected feeding on the main host plant species from six biogeographic provinces (Figure 1, Table 1).

We also assembled the first draft genome of *C. cactorum* using a total of  $2.8 \times 10^5$  reads (average read length = 1,453 bp) to use as reference mapping of the ddRAD reads. The assembly process yielded a total genome size of 411.6 Mbp (available as CactoFuEDEL.fa at the NCBI under accession number: JADGIL010000000).

Assemblies results configured with different percentages of minimum number of samples per locus (msl) using Ipyrad (Eaton & Overcast, 2020) varied in the number of loci recovered (Supplementary Table 2). However, it was worth mentioning that all datasets yielded similar results (available in Figshare: <https://doi.org/10.jhnmn6084/m9.figshare.13118249.v1>). Thus, we presented the results obtained with the numbers of loci recovered using 25% of msl, that consisted in a matrix of 37,243 SNPs. The analysis of the mitochondrial dataset revealed 39 haplotypes (length = 790 pb), 162 variable sites, 137 of which were parsimony-informative sample wide.

### Intra and interspecific genetic admixture and diversity

Analyses of genetic admixture based on the ddRAD dataset using sNMF revealed  $K = 3$  as the most likely value of clusters, consistent with the three nominal moth species: *C. doddi*, *C. cactorum* and *C. bucyrus* (Figure 2a). This method assigned the moth population feeding on *C. baumannii* to *C. bucyrus* cluster, therefore, we referred to it as allopatric *C. bucyrus* from hereafter, despite it is contradictory with the *a priori* presumed monophagy of

*C. bucyrus*.

Clustering analysis also revealed low but noticeable signals of admixture in sympatric populations of the three species (Figure 2b). In effect, *C. bucyrus* individuals from NWA showed genetic admixture with both *C. doddi* and *C. cactorum*. Also, *C. cactorum* individuals from NWA shared variants with sympatric *C. doddi* and *C. bucyrus*; and *C. doddi* individuals shared variants with sympatric *C. cactorum* and *C. bucyrus*. Overall, only allopatric *C. cactorum* population, i.e., not in evident current contact with closely related moths, showed footprints of admixture, since at least three individuals carried genetic elements of the other two species (Figure 2).

When we considered values of  $K$  greater than 3, sNMF distinguished allopatric from sympatric populations within each species (Supplementary Figure 1). With  $K = 4$ , both populations of *C. cactorum* appeared separated in two distinct clusters with a high degree of admixture among them. Similarly, clustering analysis based on  $K = 5$  split moths collected on *C. baumannii* from NWA *C. bucyrus* into two separated clusters with slight signals of admixture between them. Finally, *C. doddi* populations were recovered as separate clusters using a value of  $K = 6$ . These results suggested a certain degree of differentiation within each species. Finally, clustering analysis allowed us to identify five individuals initially considered as *C. cactorum*

because they were collected on *O. ficus-indica*, that clustered with *C. doddi* individuals, casting doubt on the presumptive specialization of the latter on *O. sulphurea*.

The Neighbor Joining network obtained with the mtDNA dataset revealed three clearly defined haplogroups representing each one of the three species, separated by a large number of mutational steps from each other (Figure 2c). Within each species, allopatric populations appeared differentiated from sympatric ones, also suggesting significant population structuring in *C. cactorum*, *C. doddi* and *C. bucyrus*. We did not find shared haplotypes between allopatric populations of each species, nor introgressed haplotypes between species.

Basic statistics quantifying intrapopulation genetic variation were presented in Table 2. Considering the ddRAD data, sympatric populations of *C. cactorum* and *C. doddi* were genetically more diverse than the respective allopatric populations, while moths feeding on *C. baumannii* were more diverse than NWA *C. bucyrus*. Regarding mtDNA, our survey showed that *C. doddi* had the highest haplotype diversity as compared to both *C. cactorum* and *C. bucyrus*.

Pairwise  $F_{st}$  values calculated using ddRAD data and mtDNA data also unveiled, as expected, strong genetic differentiation between species, and considerably, though lower, differentiation within species (Supplementary Table 3). Overall,  $F_{st}$  values obtained in interspecific comparisons were, on average, lower between pairs of sympatric populations than between allopatric ones. The lowest  $F_{st}$  was observed between sympatric populations of *C. cactorum* and *C. doddi* ( $F_{st} = 0.772$  and  $0.859$ , for ddRAD and mtDNA respectively), while the highest  $F_{st}$  was detected between *C. doddi* and *C. bucyrus* ( $F_{st} = 0.939$  for both ddRAD and mtDNA). In within species comparisons,  $F_{st}$  ranged from 0.244 (ddRAD) and 0.577 (mtDNA) in *C. cactorum*, to 0.642 (ddRAD) and 0.702 (mtDNA) in *C. bucyrus* from NWA and *C. bucyrus* from Santiago del Estero.

### Phylogenetic hypotheses and introgression footprints

Both ddRAD and mtDNA recovered three major clades, each one encompassing conspecific populations of each species (Figure 3). *Tucumania* sp. could not be used as an outgroup, as *a priori* planned, because of the long divergence of this cactophilic moth precluded its inclusion in further analyses of ddRAD dataset. However, it could be used to obtain the mtDNA tree, in which the columnar cactus feeder *C. bucyrus* appeared as the sister species of a clade formed by *C. doddi* and *C. cactorum* (Figure 3a). On the other hand, *C. cactorum* appeared as the sister species of a clade formed by *C. bucyrus* and *C. doddi* in the midpoint rooted tree inferred with ddRAD data (Figure 3b). The species tree inferred with SNAPP, also supported *C. cactorum* as the sister species of the clade formed by *C. doddi* and *C. bucyrus* (Figure 3c). Interestingly, two fairly differentiated subclades were apparent within each species in both ddRAD and mtDNA trees (Figure 3a-b respectively). In *C. cactorum*, one subclade consisted of specimens collected in NWA and the other subclade of individuals collected in La Pampa. Likewise, *C. doddi* populations appeared as differentiated subclades; the Mendoza subclade encompassed individuals feeding exclusively on *O. sulphurea*, and the NWA subclade included specimens collected on *O. sulphurea* and the exotic *O. ficus-indica*. Also, the *C. bucyrus* clade included two fairly differentiated subclades, one involving specimens feeding on *T. atacamensis* (NWA) and the other subclade of moths collected on *C. baumannii* from Santiago del Estero. Estimates of divergence times based on mtDNA data suggested that the genus *Cactoblastis* originated 3.41 million years ago (Mya) (95% of credibility interval = 2.78 - 4.08 Mya), that the *Opuntia* feeders *C. cactorum* and *C. doddi* shared their last common ancestor with *C. bucyrus* ca 2.33 Mya (95% of credibility interval = 1.84 - 2.84 Mya), and that the split between *C. cactorum* and *C. doddi* occurred 1.79 Mya (95% of credibility interval = 1.35 - 2.24 Mya) (Figure 3a).

ddRAD data, using an average of 12,897 unlinked ddRAD SNPs, was used to search for footprints of introgressive hybridization with the four-taxon ABBA/BABA test (Supplementary Table 4). These tests gave support to the hypothesis of introgressive hybridization between sympatric populations of *C. bucyrus* and *C. doddi* (D-stats = -0.271; Z-score = 6.141,  $p < 0.001$ ) and between allopatric populations of *C. cactorum* and *C. bucyrus* (D-stats = 0.279; Z-score = 6.427,  $p < 0.001$ ).

### Demographic modelling

Species and population demographic history of the three *Cactoblastis* spp. was assessed using coalescent-based simulations using Fastsimcoal2. These simulations tested models depicting alternative scenarios of gene flow, before and after intraspecific divergence (Supplementary Figure 2). When we tested scenarios of interspecific gene flow before intraspecific divergence, the most likely model was one assuming historical gene flow involving the three species, as compared to models considering only historical gene flow between pairs of species (Figure 4; Supplementary Table 5). These simulations (assuming gene flow before divergence of allopatric and sympatric populations within species) revealed that the highest rate of gene flow occurred between *C. bucyrus* and *C. doddi* (migration rate per generation:  $m = 0.0049$ ) as compared with the other pairs of sympatric populations (*C. cactorum* and *C. doddi* or *C. cactorum* and *C. bucyrus*) (Table 3). Assuming two generations per year in native species (Varone et al., 2014; Varone et al., 2019), we estimated that the last common ancestor shared by *C. cactorum* and the clade *C. doddi* + *C. bucyrus* occurred about 2.22 Mya, while *C. doddi* and *C. bucyrus* diverged more recently, 700,339 years. Similar analyses showed that within the presumptive conspecific populations of *C. bucyrus* (NWA and Santiago del Estero) divergence occurred 41,641 years ago.

When we considered scenarios of interspecific gene flow after divergence within species the most likely model was model 5. This model assumed historical gene flow involving the three sympatric populations in which the highest rate of gene flow was obtained between *C. cactorum* and *C. bucyrus* ( $m = 0.00356$ ), followed by *C. doddi* and *C. bucyrus* ( $m = 0.00108$ ). Coalescent simulations also revealed that the most ancient interspecific gene flow occurred between *C. bucyrus* and *C. doddi* during a short period of time from 12.9 thousand to 16.1 thousand years ago (Figure 4) (see Table 3 for confidence intervals). The most recent episode of interspecific gene flow occurred between sympatric populations of *C. cactorum* and *C. doddi*, and *C. cactorum* - *C. bucyrus* during a period of time between 1.8 thousand to 13.2 thousand years ago.

## Discussion

The survey of intra and interspecific genetic variation in the South American cactophilic moths of the genus *Cactoblastis* revealed a complex evolutionary history involving historical episodes of genetic interchange before and after intraspecific divergence. Our study combining a population genetic approach including nuclear (ddRAD) and mtDNA (*COI*) markers with surveys of host plant use allowed us to propose a phylogeographic hypothesis that shed light on the evolutionary history of a group of moths that included two pest species that threaten cactus biodiversity, negatively affecting local agricultural-based economies.

Clustering and phylogenetic analyses revealed that these closely related species were clearly differentiated despite the evidence of admixture in areas of sympatry and between presently allopatric populations. These results suggested that adaptation to alternative host plants played an important role as an isolating barrier among *Cactoblastis* species in the face of interspecific gene flow. Similar insect/host plant associations have been documented as an effective barrier to gene flow, even in pairs of closely related species with overlapping biogeographic distributions (Clarkson et al., 2014; Poveda-Martinez et al., 2019).

The footprint of historical admixture after and before intraspecific divergence supported by coalescent-based simulations using Fastsimcoal2 are concordant with the results of ABBA-BABA tests and clustering analysis using sNMF. These results suggested introgressive hybridization events between *C. doddi* and *C. bucyrus* in NWA, and between *C. cactorum* and *C. bucyrus* from Santiago del Estero. Scenarios of historical gene flow have been reported in several insect species, generally occurring in sympatry (Clarkson et al., 2014; Tonzo et al., 2020), and less frequently between allopatric and distantly related species (Ward & Simon, 2018; Zhang, Dasmahapatra, Mallet, Moreira, & Kronforst, 2016). Evidence of introgressive hybridization has also been reported in the grasshoppers *Omocestus minutissimus* Brullé and *Omocestus uhagonii* Bolívar suggesting past introgression in the area where their distributions overlap (Tonzo et al., 2020). This study points to a scenario of historical gene flow after secondary contact followed by the evolution of reproductive isolation that currently prevents hybridization among sympatric populations. Likewise, genetic studies suggest that *Sepsis cynipsea* L. and *Sepsis necynipsea* Melander & Spuler (Diptera: Sepsidae), two related and genetically differentiated flies, may occasionally hybridize in nature in sympatry (Giesen et al., 2020). Similarly, a pioneering genetic study of butterflies of the genus *Heliconius* Latreille including pairs of sympatric and

allopatric populations, revealed a genome-wide trend of increased shared variation in sympatry, indicating pervasive interspecific gene flow between these closely related species (Martin et al., 2013). In contrast, the moth *Plutella xylostella* L., an introduced pest in Australia, and the native *Plutella australiana* Landry & Hebert, have been sympatric for over 125 years. However, complementary approaches testing for genetic admixture supported reproductive isolation of *Plutella* species in Australia, despite their ability to hybridize in experimental laboratory crosses (Ward & Simon, 2018).

Even though admixture in *Cactoblastis* spp. were detected between presently sympatric and also between allopatric populations, overall genomic differentiation between species was lower in areas of sympatry than between pairs of allopatric populations, either for ddRAD and mtDNA. These footprints of gene flow seemed to be the product of ancient rather than recent or ongoing gene flow that probably occurred during early stages of host plant shift in which reproductive isolation was not completed. A common pattern of recent gene flow should result in reduced differentiation and an excess of shared variants between sympatric populations as compared to allopatric populations. Similar results were reported by Tonzo et al (2020) for the pairs of sympatric species of grasshoppers, where interspecific gene flow was suggested as a product of ancient introgression and not of recent gene flow that probably occurred during the last glacial maximum.

Introgression analyses revealed interspecific gene flow between distant populations of *C. cactorum* (NWA) and *C. bucyrus* (Santiago del Estero). However, it is important to note that in the southern edge of the distribution of the putative *C. bucyrus* population living on *C. baumannii*, coexisting populations of *C. cactorum* (not included in this study) have been reported (Varone et al., 2014). In this region, *C. cactorum* is an important pest of prickly pear crops and may be a contact zone favoring interspecific gene flow between these related species, as revealed by introgression tests, particularly after the introduction of large *O. ficus-indica* plantations. Thus, we may ask, whether interspecific gene flow between species may have played a role in the evolution of the capabilities of *C. cactorum* and *C. bucyrus* to become important pests of the cultivated *O. ficus-indica* and the emblematic national monument *T. atacamensis*, respectively. However, whole genome comparative analyses are necessary to determine the direction and extent of introgression, the distribution, and function of introgressed segments to test these hypotheses.

The cactophagous moths collected on the columnar cactus *C. baumannii* have apparent morphological differences and ecological preferences when compared to the three recognized *Cactoblastis* species (Barbarich et al., 2021; Heinrich 1939; Mann, 1969; McFadyen, 1985; Varone et al., 2014). The *C. baumannii* feeding *Cactoblastis* sp. was assigned to *C. bucyrus* by both clustering and phylogenetic analyses. However, both genetic markers revealed greater differentiation within *C. bucyrus* than between allopatric populations of either *C. cactorum* or *C. doddi*. In fact,  $F_{st}$  values between *C. bucyrus* and the *C. baumannii* feeding moths were 2.5 and 1.5 times greater for nuclear markers and 1.2 and 1.1 times greater for mtDNA than the values estimated within *C. cactorum* or *C. doddi*, respectively (Supplementary Table 3). Both *C. bucyrus* sensu stricto (NWA) and the allopatric moth collected on *C. baumannii* fed on columnar cacti of the subfamily Cactoideae, although of different genera (Barbarich et al., 2021; McFadyen, 1985). We hypothesized that the host plant shift from *C. baumannii* to *T. atacamensis*, or vice versa, was an important driver of genetic divergence between moth populations. However, these populations were far apart (~500 km), suggesting that isolation by distance may have played a role in differentiation (Forbes et al., 2017; Hood et al., 2020; Poveda-Martínez et al., 2020). Thus, to determine if our present observations represented merely a niche expansion to a new host or a species *in statu nascendi* requires additional collections of allopatric populations feeding on either *T. atacamensis* or *C. baumannii* (or even other columnar cacti) and a larger representation of the nuclear genome. Moreover, mate choice experiments aimed to evaluate prezygotic isolation, host plant preference and performance experiments, and breeding experiments testing for postzygotic isolation would help to confirm the taxonomic status of these moths breeding on columnar cacti. The available evidence was not sufficient to define whether the pattern of genomic differentiation corresponded to intraspecific variation, as occurred in *C. doddi* and *C. cactorum*. Nevertheless, despite our survey revealed significant population structure within *C. cactorum* with both nuclear and mitochondrial markers, differentiation was lower than within *C. bucyrus* and *C. doddi* despite the long distance (~1,350 km) separating these populations. This relatively low genetic differentiation suggested a pattern of extensive gene flow in the distribution range of *C.*

*cactorum*, reported to include several biogeographic provinces in Argentina (Varone et al., 2014). Moreover, a previous survey using the mitochondrial *COI* gene including a set of populations that did not comprise the southern edge of the species' distribution identified high mitochondrial diversity (Marsico et al., 2011). Our sampling scheme extended the sampling area to include a population representative of the southernmost area inhabited by *C. cactorum* in Argentina. Interbreeding experiments involving moth populations separated by more than 1,000 km did not show any evidence of sexual or postzygotic isolation (Fuentes Corona & Varone personal comm). Further studies including more *C. cactorum* populations associated with additional native and exotic host plants over a wider distribution of the species are necessary to understand the demographic history and dynamics of gene flow in this cactus pest.

Previous reports documented that *C. doddi* is restricted to *O. sulphurea* (Mann, 1969; McFadyen, 1985; Zimmermann et al., 1979). However, our results confirmed previous observations indicating that *C. doddi* also includes the exotic cactus *O. ficus-indica* in its diet (McFyden, 1980; McFadyen, 1985). We found two different *O. ficus-indica* patches attacked by *C. doddi* in urban locations where *O. ficus-indica* was scarce and uncommon and mainly used as ornamental plantings in NWA. The confirmation of the presence of *C. doddi* feeding on *O. ficus-indica* derived from the use of genetic markers proposes that this insect-plant association could be more common than what has been registered up to now.

Both nuclear (ddRAD) and mitochondrial (mtDNA) based phylogenetic inferences for the studied *Cactoblastis* populations recovered the three expected clades according to the population genetic survey (Figure 2-C); one for *C. cactorum*, another for *C. doddi* and the third included *C. bucyrus* collected on *T. atacamensis* and moths feeding on *C. baumannii*. However, the mitochondrial and the nuclear trees were not entirely congruent. In the ddRAD, and the species trees, *C. cactorum* appeared as the most external clade, while in the mtDNA tree, *C. bucyrus* appeared as the external clade. These species involved in the mito-nuclear incongruence were also identified in both introgression tests, ABBA-BABA and coalescent-based analyses, as having experienced interspecific gene flow. Such mito-nuclear incongruences are typically considered as a consequence of either past hybridization or incomplete lineage sorting (ILS), as it has been frequently in recently diverged insect taxa (Moreyra et al., 2019). Despite the incongruence mentioned, estimations of the time of divergence between species based on both genetic markers suggested that the origin of this trio of species occurred between the Upper Pliocene and the early Pleistocene. This time period coincident with the occurrence of glacial refugia resulted in the divergence of many taxa, including several Lepidoptera (Fagua et al., 2019; Hewitt, 2020).

The evidence of interspecific gene flow in the trio of *Cactoblastis* species demonstrated the capability of *C. cactorum* to hybridize with other species in its native range. We found evidence of ancient introgressive hybridization in closely related species, but patterns of introgression have also been reported in distantly related Lepidoptera (Zhang et al., 2016). *Cactoblastis cactorum* has been attacking native cacti in North America since 1989, threatening *Opuntia* spp. biodiversity and prickly pear agricultural production (Hight & Carpenter, 2009). Whether the acquisition of genetic variants via interspecific hybridization between *C. cactorum* and closely related species has played a role in fueling adaptation and adaptive radiation of this invasive species in its invaded area is still an open question. In this sense, gene flow between the invasive *Helicoverpa armigera* Hubner and the native *Helicoverpa zea* Boddie is an example of two moths species with evidence of hybridization and widespread introgression acquired for pyrethroid insecticide resistance (Valencia-Montoya et al., 2020).

## Acknowledgements

We appreciate Andres Sanchez Restrepo for technical assistance, and Mariel Guala for fieldwork support. We also appreciate the assistance of Vanina Tonzo with data analyses, and express our gratitude to members of Evolution Lab for insightful comments. We thank the Instituto de Ciencias Polares, Ambiente y Recursos Naturales of the Universidad Nacional de Tierra del Fuego, Biocódices SA, and the Centro de Cómputos de Alto Rendimiento (CeCAR) for granting use of computational resources. Funding was obtained from USDA APHIS-PPQ, Farm Bill 10201, and Fondo para la Investigación Científica y Tecnológica (PICT1447/2016 awarded to G.L.). D.P.M. and L.M. are recipients of PhD scholarships awarded by CONICET and ANPCyT,

respectively. L.V. and E.H. are members of Carrera del Investigador CONICET.

## References

- Arce de Hamity, M. G., & Neder de Román, L. E. (1999). Bioecología de *Cactoblastis bucyrus* (Lepidoptera: Phycitidae), especie dañina al cardón *Trichocereus pasacana* en la Prepuna jujeña (Argentina). *Revista de la Sociedad Entomológica Argentina* , 58 (3-4), 23-32.
- Barbarich, M. F., Otegui, F., Saint Esteven, A., Soto, I. M., & Varone, L. (2021). A protocol for health assessment of cacti populations: a case study from northwestern Argentina. *Ecological Indicators* , 121, 107174. <https://doi.org/10.1016/j.ecolind.2020.107174>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* , 10 (4), e1003537.
- Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* , 29 (8), 1917-1932. <https://doi.org/10.1093/molbev/mss086>
- Caterino, M. S., & Sperling, F. A. (1999). Papilion phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Molecular Phylogenetics and Evolution* , 11 (1), 122-137.
- Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., ... & Donnelly, M. J. (2014). Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature Communications* , 5 (1), 1-10. <https://doi.org/10.1038/ncomms5248>
- Dodd, A. P. (1940). The biological campaign against prickly-pear. Commonwealth Prickly Pear Board Bulletin, Brisbane, Australia, 77
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* , 29 (8), 1969-1973. <https://doi.org/10.1093/molbev/mss075>
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution* , 28 (8), 2239-2252.
- Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* , 36 (8), 2592-2594.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics* , 9 (10), e1003905.
- Fagua, G., Condamine, F. L., Dombroskie, J. J., Byun, B. K., De Prins, J., Simonsen, T. J., ... & Sperling, F. A. (2019). Genus delimitation, biogeography and diversification of *Choristoneura Lederer* (Lepidoptera: Tortricidae) based on molecular evidence. *Systematic Entomology* , 44 (1), 19-38.
- Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K., Widmayer, H. A., & Wilson, C. J. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* , 71 (5), 1126-1137.
- Fрати, F., Negri, I., Fanciulli, P. P., Pellecchia, M., De Paola, V., Scali, V., & Dallai, R. (2004). High levels of genetic differentiation between Wolbachia-infected and non-infected populations of *Folsomia candida* (Collembola, Isotomidae). *Pedobiologia* , 48 (5-6), 461-468.
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution* , 6 (8), 925-929. <https://doi.org/10.1111/2041>
- Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics* , 196 (4), 973-983. <https://doi.org/10.1534/genetics.113.160572>

- Funk, D. J., Nosil, P., & Etges, W. J. (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences* , 103 (9), 3209-3213. <https://doi.org/10.1073/pnas.0508653103>
- Giesen, A., Blanckenhorn, W. U., Schaefer, M. A., Shimizu, K. K., Shimizu-Inatsugi, R., Misof, B., ... & Kapun, M. (2020). Genomic signals of admixture and reinforcement between two closely related species of European sepsid flies. bioRxiv. Preprint at <https://doi.org/10.1101/2020.03.11.985903>
- Goudet, J. (2005). Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5 (1), 184-186. <https://doi.org/10.1111/j.1471>
- Heinrich, C. (1939). The cactus-feeding Phycitinae: a contribution toward a revision of the American pyralidoid moths of the family Phycitidae. *Proceedings of the United States National Museum* .
- Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405 (6789), 907-913. <https://doi.org/10.1038/35016000>
- Hight, S. D., & Carpenter, J. E. (2009). Flight phenology of male *Cactoblastis cactorum* (Lepidoptera: Pyralidae) at different latitudes in the southeastern United States. *Florida Entomologist* , 92 (2), 208-216.
- Hood, G. R., Powell, T. H., Doellman, M. M., Sim, S. B., Glover, M., Yee, W. L., ... & Feder, J. L. (2020). Rapid and repeatable host plant shifts drive reproductive isolation following a recent human-mediated introduction of the apple maggot fly, *Rhagoletis pomonella* . *Evolution* , 74 (1), 156-168.
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodohl, P. A. (2013). diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution* , 4 (8), 782-788.
- Keightley, P. D., Pinharanda, A., Ness, R. W., Simpson, F., Dasmahapatra, K. K., Mallet, J., ... & Jiggins, C. D. (2015). Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology and Evolution*, 32 (1), 239-243.
- Leache, A. D., Banbury, B. L., Felsenstein, J., De Oca, A. N. M., & Stamatakis, A. (2015). Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systematic biology*, 64 (6), 1032-1047.
- Leigh, J. W., & Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6 (9), 1110-1116.
- Lynch, M., & Conery, J. S. (2003). The origins of genome complexity. *Science* , 302 (5649), 1401-1404.
- Mann, J. (1969). Cactus-feeding insects and mites. *Bulletin of the United States National Museum*. 1-158 <https://doi.org/10.5479/si.03629236.256.1>
- Marsico, T. D., Wallace, L. E., Ervin, G. N., Brooks, C. P., McClure, J. E., & Welch, M. E. (2011). Geographic patterns of genetic diversity from the native range of *Cactoblastis cactorum* (Berg) support the documented history of invasion and multiple introductions for invasive populations. *Biological Invasions* , 13 (4), 857-868.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* , 23 (11), 1817-1828. <https://doi.org/10.1101/gr.159426.113>
- McFadyen, R.E. (1980). *Cactoblastis* [Lep.: Phycitidae] for the biological control of *Eriocereus martinii* [Cactaceae] in Australia. *Entomophaga*, 25 (1), 37-42. <https://doi.org/10.1007/BF02377520>
- McFadyen, R. E. (1985). Larval characteristics of *Cactoblastis* spp. (Lepidoptera: Pyralidae) and the selection of species for biological control of prickly pears (*Opuntia* spp.). *Bulletin of Entomological Research*, 75 (1), 159-168.

- Moreyra, N. N., Mensch, J., Hurtado, J., Almeida, F., Laprida, C., & Hasson, E. (2019). What does mitogenomics tell us about the evolutionary history of the *Drosophila buzzatii* cluster (repleta group)? *PLoS ONE*, *14* (11), e0220676.
- Morrone, J. J. (2014). Biogeographical regionalisation of the Neotropical region. *Zootaxa*, *3782*(1), 1-110. <http://dx.doi.org/10.11646/zootaxa.3782.1.1>
- Nosil, P. (2012). Ecological Speciation. Oxford Series in Ecology and Evolution. New York: Oxford University Press.
- Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27* (7), 1659-1672.
- Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., ... & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, *8* (6), e1002752.
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, *7* (5), e37135.
- Poveda-Martinez, D., Aguirre, M. B., Logarzo, G., Calderon, L., de la Colina, A., Hight, S., ... & Hasson, E. (2019). Untangling the *Hypogeococcus pungens* species complex (Hemiptera: Pseudococcidae) for Argentina, Australia, and Puerto Rico based on host plant associations and genetic evidence. *PLoS ONE*, *14* (7), e0220366. <https://doi.org/10.1371/journal.pone.0220366>
- Poveda-Martinez, D., Aguirre, M. B., Logarzo, G., Hight, S. D., Triapitsyn, S., Diaz-Sotero, H., ... & Hasson, E. (2020). Species complex diversification by host plant use in an herbivorous insect: The source of Puerto Rican cactus mealybug pest and implications for biological control. *Ecology and Evolution*, *10* (19), 10463-10480. <https://doi.org/10.1002/ece3.6702>
- Powell, D. P., McMichael, M., & Silvain, J. F. (2004). Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America*, *97*(5), 1034-1044.
- Raghu, S., & Walton, C. (2007). Understanding the ghost of *Cactoblastis* past: historical clarifications on a poster child of classical biological control. *BioScience*, *57* (8), 699-705.
- Roe, A. D., & Sperling, F. A. (2007). Population structure and species boundary delimitation of cryptic *Dioryctria* moths: an integrative approach. *Molecular Ecology*, *16*(17), 3617-3633.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61* (3), 539-542.
- Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34* (12), 3299-3302.
- Santos, H., Rousselet, J., Magnoux, E., Paiva, M. R., Branco, M., & Kerdelhue, C. (2007). Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences*, *274* (1612), 935-941. <https://doi.org/10.1098/rspb.2006.3767>
- Schoonhoven, L. M., Van Loon, B., van Loon, J. J., Dicke, M. (2005). Insect-plant biology. New York: Oxford University Press.
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction

primers. *Annals of the entomological Society of America* ,87 (6), 651-701.

Simon, J. C., d'Alençon, E., Guy, E., Jacquin-Joly, E., Jaquier, J., Nouhaud, P., ... & Streiff, R. (2015). Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, 14 (6), 413-423.

Tonzo, V., Papadopoulou, A., & Ortego, J. (2020). Genomic footprints of an old affair: Single nucleotide polymorphism data reveal historical hybridization and the subsequent evolution of reproductive barriers in two recently diverged grasshoppers with partly overlapping distributions. *Molecular Ecology* , 29 (12), 2254-2268.

Valencia-Montoya, W. A., Elfekih, S., North, H. L., Meier, J. I., Warren, I. A., Tay, W. T., ... & Jiggins, C. D. (2020). Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. *Molecular Biology and Evolution* , 37 (9), 2568-2583.

Varone, L., Acosta, M. M., Logarzo, G. A., Briano, J. A., Hight, S. D., & Carpenter, J. E. (2012). Laboratory Performance of *Cactoblastis Cactorum* (Lepidoptera: Pyralidae) on South and North American *Opuntia* species occurring in Argentina. *Florida Entomologist* , 1163-1173.

Varone, L., Aguirre, M. B., Lobos, E., Perez, D. R., Hight, S. D., Palottini, F., ... & Logarzo, G. A. (2019). Causes of mortality at different stages of *Cactoblastis cactorum* in the native range. *BioControl*, 64 (3), 249-261.

Varone, L., Logarzo, G. A., Briano, J. A., Hight, S. D., & Carpenter, J. E. (2014). *Cactoblastis cactorum* (Berg)(Lepidoptera: Pyralidae) use of *Opuntia* host species in Argentina. *Biological Invasions* , 16 (11), 2367-2380. <https://doi.org/10.1007/s10530-014-0670-9>

Ward, C. M., & Baxter, S. W. (2018). Assessing genomic admixture between cryptic *Plutella* moth species following secondary contact. *Genome Biology and Evolution* , 10 (11), 2973-2985. <https://doi.org/10.1093/gbe/evy224>

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* , 1358-1370.

Wiens, J. J., Lapoint, R. T., & Whiteman, N. K. (2015). Herbivory increases diversification across insect clades. *Nature Communications* , 6 (1), 1-7.

Yamasaki, Y. Y., Kakioka, R., Takahashi, H., Toyoda, A., Nagano, A. J., Machida, Y., ... & Kitano, J. (2020). Genome-wide patterns of divergence and introgression after secondary contact between *Pungitius* sticklebacks. *Philosophical Transactions of the Royal Society B*, 375 (1806), 20190548.

Zapata, R. M., Karlin, U. O., Lucero, F., Coirini, R. O., & Karlin, M. (2005). Manejo sustentable del ecosistema Salinas Grandes, Manejo de los tunales, Proyecto Desarrollo de un polo productivo integral en el norte de Córdoba: cultivo y procesamiento de la tuna (*Opuntia ficus-indica*) y especies aromáticas bajo certificación orgánica. Proyectos Federales de Innovación Productiva del Ministerio de Ciencia, Tecnología e Innovación Productiva, Cartilla Ndeg2.

Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R., & Kronforst, M. R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology* ,17 (1), 1-15.

Zimmermann, H. G., Erb, H. E., & McFadyen, R. E. (1979). Annotated list of some cactus-feeding insects of South America. *Acta Zoologica Lilloana* , 33 (2), 101-112.

## Data Availability

The datasets generated during the current study, input files, and demographic models tested with Fastsimcoal2 are available in Figshare: [https://doi.org/10.jh\\_nmn6084/m9.figshare.13118249.v1](https://doi.org/10.jh_nmn6084/m9.figshare.13118249.v1)

Raw reads are available at NCBI as BioProject PRJNA666743.

The reference draft genome is available as CactoFuEDEI\_v1.fa at the NCBI under accession number JADGIL010000000.

Mitochondrial haplotypes are available at Genbank under accession numbers MW064275-MW064313.

### Author contributions

D.P.M., L.V., G.L., and E.H. designed the work. D.P.M. conducted molecular work and data analyses. D.P.M., L.V., M.F., E.H., S.H., and G.L. performed field collections. D.P.M., and E.H. drafted the manuscript. All coauthors made contributions to the final version of the manuscript.

### Additional Information

Supplementary File

The authors declare that they have no conflicts of interest.

### Tables

**Table 1.** Taxon sampling of *Cactoblastis* species in Argentina. NWA: northwestern Argentina. Numbers in parenthesis indicate the number of plants from which moths were collected.

Taxa	Province	Latitude	Longitude	Host plant species
<i>C. cactorum</i>	Jujuy (NWA)	-24.3938	-65.2713	<i>O. ficus-indica</i> (5)
<i>C. cactorum</i>	Salta (NWA)	-25.0543	-65.4970	<i>O. ficus-indica</i> (4)
<i>C. cactorum</i>	Salta (NWA)	-25.2077	-64.9460	<i>O. ficus-indica</i> (4)
<i>C. cactorum</i>	Salta (NWA)	-25.2077	-64.9460	<i>O. quimilo</i> (1)
<i>C. cactorum</i>	La Pampa	-36.2329	-66.9374	<i>O. megapotamica</i> (5)
<i>C. cactorum</i>	La Pampa	-36.2204	-65.5706	<i>O. megapotamica</i> (6)
<i>C. cactorum</i>	La Pampa	-36.4074	-67.1410	<i>O. megapotamica</i> (4)
<i>C. doddi</i>	Jujuy (NWA)	-24.4775	-65.0382	<i>O. ficus-indica</i> (1)
<i>C. doddi</i>	Jujuy (NWA)	-23.6014	-65.4058	<i>O. sulphurea</i> (5)
<i>C. doddi</i>	Jujuy (NWA)	-23.6590	-65.4334	<i>O. sulphurea</i> (6)
<i>C. doddi</i>	Salta (NWA)	-25.0543	-65.4970	<i>O. ficus-indica</i> (5)
<i>C. doddi</i>	Mendoza	-34.6639	-68.5389	<i>O. sulphurea</i> (5)
<i>C. bucyrus</i>	Jujuy (NWA)	-23.6014	-65.4058	<i>T. atacamensis</i> (10)
<i>Cactoblastis</i> on <i>Cleistocactus</i>	Santiago del Estero	-28.0242	-64.6124	<i>C. baumannii</i> (10)
<i>Tucumania</i> sp.	Formosa	-25.8424	-58.7924	<i>O. anacantha</i> (2)

**Table 2.** Measures of genetic diversity in sympatric (NWA) and allopatric populations sampled for each population of *Cactoblastis* sp. based on mtDNA ddRAD data.

Region	Host plants	mtDNA n	mtDNA $\pi$	mtDNA $H_d$
Cactoblastis cactorum	Cactoblastis cactorum	Cactoblastis cactorum	Cactoblastis cactorum	Cactoblastis cactorum
NWA	<i>O. ficus-indica</i> <i>O. quimilo</i>	15	0.007	0.914
La Pampa	<i>Opuntia megapotamica</i>	10	0.004	0.844
Cactoblastis doddi	Cactoblastis doddi	Cactoblastis doddi	Cactoblastis doddi	Cactoblastis doddi
NWA	<i>O. sulphurea</i> , <i>O. ficus-indica</i>	9	0.008	0.806
Mendoza	<i>Opuntia sulphurea</i>	6	0.006	1
Cactoblastis bucyrus	Cactoblastis bucyrus	Cactoblastis bucyrus	Cactoblastis bucyrus	Cactoblastis bucyrus
NWA	<i>Trichocereus atacamensis</i>	13	0.001	0.756
Santiago del Estero	<i>C. baumannii</i>	11	0.006	0.818

n: Sample size;  $\pi$ : Nucleotide diversity;  $H_d$ : Haplotype diversity; H: number of haplotypes;  $H_e$ : Expected heterozygosity;  $H_o$ : observed heterozygosity; Ar: Allelic richness; NWA: Northwestern Argentina.

**Table 3.** Point estimates and 95% confidence intervals of demographic parameters inferred from coalescent simulations using Fastsimcoal2.

Estimated parameters	Model 1. Before intraspecific divergence	Model 1. Before intraspecific divergence
	Point estimate	Lower bound
Ne BS	27,398	3,742
Ne DA	119082*	44,553
Ne DS	442,965	108,731
Ne CA	398676*	128,466
Ne CS	178,012	95,244
Tdiv 1	49,125	40,952
Tdiv 2	38,460	23,399
Tdiv 3	83,283	53,807
Tdiv 4	1,400,678	563,085
Tdiv 5	4,452,304	3,501,657
M C-D	0.00040	0.00002
M C-B	0.00153	0.00003
M D-B	0.00496	0.00003
T adm 1	-	-
T adm 2	-	-
T adm 3	-	-
T adm 4	-	-

Ne: population size; Tdiv: time of divergence per generation; Tadm: time of admixture; M: migration rate per generation. B: *Cactoblastis bucyrus* ; D: *Cactoblastis doddi* ; C: *Cactoblastis cactorum* ; S: sympatric; A: allopatric. Point estimates indicated by \* fall outside the confidence intervals.

### Figure legends

**Figure 1.** Geographic locations where *Cactoblastis* spp. moth populations were sampled. The zoomed area corresponds to northwestern Argentina (NWA) where the three species of *Cactoblastis* coexist. Locations of the allopatric populations, and the moth population feeding on a new host plant are shown outside of the zoomed area. Colored regions represent the biogeographic provinces according to the Morrone (2014) regionalisation.

**Figure 2.** Clustering analyses based on ddRAD, and neighbor joining network based on mtDNA. (a) Visual representation of the three clusters ( $k=3$ ) representing each *Cactoblastis* species. Vertical bars represent individuals along with membership probability of each individual to the corresponding cluster/species. Codes denote geographic origin of populations: NWA (northwestern) Argentina; SDE (Santiago del Estero); MDZ (Mendoza); LPP (La Pampa). Asterisk indicates individuals assigned to *C. doddi* feeding on *O. ficus-indica*. (b) shows pie charts of admixture proportions of all individuals in their respective geographic locations, including the sympatric region NWA as a zoomed panel. (c) The network shows the three major haplogroups corresponding to each species. Values between and within haplogroups are the number of mutational steps, and remarked squares denote the haplotypes from NWA in each species. Color codes in a, b, and c panels are the same: Light green: *C. bucyrus* ; Dark blue: *C. doddi* , and purple: *C. cactorum* .

**Figure 3.** Phylogenetic hypotheses and species tree inference. Phylogenetic hypotheses based on (a) mtDNA, and (b) ddRAD inferred by Bayesian inference, and (c) species tree reconstructed using coalescent method in SNAPP based on 1,688 SNPs, and an esquematic representation of main host plant associated with each clade. Posterior probability are displayed in nodes; divergence times estimated in millions of years are shown

in panel (a) in parenthesis. Clades remarked in (a) and (b) were involved in mito-nuclear incongruence. (Color codes correspond to colors assigned in Figure 2).

**Figure 4.** Visual representation of demographic history using coalescent simulations inferred by Fast-simcoal2. Size of dotted arrows is intended to denote the relative amount of gene flow between lineages before/after intraspecific divergence. Red discontinuous lines delimit the period of time during which gene flow between sympatric populations occurred. Admixture and divergence time are given in thousand years (Kyr) and million of years (Myr). Labels are constructed by the name of the species, and the geographic distribution. NWA northwestern Argentina; SDE (Santiago del Estero); MDZ (Mendoza); LPP (La Pampa).







