Facultad de Ciencias
Universidad de Chile

The invasive process with low dispersal: the case of the California poppy *Eschscholzia californica*

Tesis

Entregada A La
Universidad De Chile
En Cumplimiento Parcial De Los Requisitos
Para Optar Al Grado De

Magíster en Ciencias Biológicas

Facultad De Ciencias

Por

María Loreto Castillo Castillo

Marzo, 2015

Director de Tesis
Dr. David Véliz

Co-Director de Tesis
Dr. Ramiro O. Bustamante

FACULTAD DE CIENCIAS

UNIVERSIDAD DE CHILE

INFORME DE APROBACION

TESIS DE MAGÍSTER

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magister presentada por la candidata.

María Loreto Castillo Castillo

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias Biológicas, en el examen de Defensa Privada de Tesis rendido el día 30 de Marzo de 2015.

Director de Tesis:
Dr. David Véliz ....

Co-Director de Tesis
Dr. Ramiro O. Bustamante ....

Comisión de Evaluación de la Tesis

Dra. Alejandra González ....

Dra. María Fernanda Pérez ....

A mis padres y hermanos

por creer en mi

y amarme incondicionalmente

Hoy me propongo opacar la luz del sol,
extraer fuerza de la naturaleza,
escalar el cielo,
tomarte de las manos
y huir al infinito

María Loreto Castillo Castillo, la Lore, como todos la llamamos cariñosamente, nació un mes de Julio en la hermosa ciudad de Punta Arenas. Llegó a Santiago cuando aún era muy chica, como dice ella, antes de que tuviera capacidad de decisión y pudiera negarse a dejar esa linda tierra. Siempre tuvo inclinación por la biología, y luego de decidir lo que quería, ingresó a estudiar Biología Ambiental en la Universidad de Chile. La Lore siempre ha tenido un gusto y amor por la naturaleza, pero por sobre todo, por el cuidado y la protección del medio ambiente. Es por eso que terminada su tesis de pregrado en el Laboratorio de Ecología Terrestre, siguió con sus estudios enfocados en especies invasoras e ingresó al programa de Magíster en Ciencias Biológicas. Hoy finaliza esta etapa, pero como sus ansias de conocimiento y de ser un aporte a la conservación de la naturaleza no terminan, continuará su camino hacia el Doctorado en una tierra aún más lejana de dónde nació, pero siempre con las ganas de volver a ella.

Seguramente lo que le espera será un gran desafío, pero la perseverancia y por sobre todo, la pasión que siente por lo que hace, le ayudarán a conseguir grandes cosas. Te deseo lo mejor del mundo porque lo mereces, estoy segura que donde sea que estés, siempre serás la mejor.

Rocío.

Agradecimientos

A mi familia, cuyo amor es la fuerza y el soporte fundamental de mi vida.

A Leonardo, por la magia que aportas a mi vida.

A mis amigos de la vida y a mis compañeros en esta aventura, porque que hicieron de esta una hermosa etapa, porque me instan a que nada es imposible.

A mis queridos tutores Ramiro Bustamante y David Véliz por los consejos, las enseñanzas, las risas, los retos, por compartir su sabiduría y conocimiento, por las largas conversaciones de la vida y por supuesto, de la ciencia. Por ser mis maestros.

A mis compañeros de laboratorio, por sus valiosos aportes a este trabajo y por los buenos ratos e historias compartidas.

A Ronny y Pancho, súper compañeros de terreno antes de que siquiera supiera que trabajaría con aquellos datos. Fueron hermosas las veladas y aventuras junto a ustedes.

A mis revisoras Alejandra González y Fernanda Pérez por sus acertados comentarios al desarrollo de este manuscrito.

Al proyecto FONDECYT 1100076, Proyecto ICM P05-002 IEB y a la Beca de Magíster Complementaria 2013; por el financiamiento otorgado para la realización de esta tesis.

Índice

Índice de Tablas

Índice de Figuras

Material Suplementario

Resumen

La dispersión tiene un importante rol en el proceso invasivo ya que determina la habilidad de las especies de diseminar hacia nuevas áreas no invadidas. Si la dispersión es moderada, patrones de aislamiento-por-distancia (IBD) pueden ser observados. Según este modelo, un incremento general en la diferenciación genética entre poblaciones estaría asociado con un incremento en la distancia física entre ellas. Sin embargo, frecuentes eventos de dispersión a larga distancia (LDD) pueden contrarrestar los efectos asociados con patrones de IBD. Adicionalmente, como la información genética refleja sólo aquellos movimientos de dispersión que resultan en una exitosa reproducción, bajo la perspectiva de la genética del paisaje, la dispersión es uno de los mecanismos clave que relacionan procesos microevolutivos con patrones del paisaje. Por último, durante el proceso de colonización y expansión, procesos demográficos tales como cuellos de botella, eventos fundadores y múltiples introducciones también pueden moldear la estructura genética en especies invasoras. Este estudio tiene como objetivo proporcionar información sobre los procesos que subyacen la invasión en plantas incluyendo ambas aproximaciones, genética del paisaje y procesos demográficos. La especie modelo en este estudio fue *Eschscholzia californica* Cham. (Papaveraceae), una planta invasora que habita la zona central de Chile. Un total de 506 individuos de 24 poblaciones fueron genotipados usando 8 microsatélites. Debido a un débil patrón de IBD, el nulo efecto de variables ambientales sobre la estructura genética y el bajo flujo génico entre poblaciones invasoras, los resultados sugieren un patrón de aislamiento-por-colonización, donde la estructura genética de las poblaciones estaria fuertemente determinada por efectos fundadores. Nuestros resultados también señalaron la existencia de múltiples introducciones de *E. californica* en Chile y eventos de LDD que no pudieron ser descartados. En conclusión, la evidencia presentada en este estudio sugiere un bajo efecto de la capacidad de dispersión y los efectos fundadores inferidos para *E. californica* en su proceso de invasión en Chile central.

Palabras claves: dispersión, eventos fundadores, múltiples introducciones, expansión geográfica, genética del paisaje.

Abstract

Dispersal has an important role in the invasive spread because determines the ability of the species to disseminate to new non-invaded areas. If dispersal is moderated, patterns of isolation-by-distance (IBD) could be observed. Under this model, a general increase in genetic differentiation among populations will be associated with increasing inter-population physical distance. However, frequency long-distance dispersal events (LDD) may counterbalance the effect associated with IBD patterns. In addition, as genetic data reflect only those dispersal movements that result in successful breeding, under a landscape genetic perspective, dispersal is one of the key mechanisms linking microevolutionary processes to landscape patterns. Lastly, during the colonization and expansion, demographic process such as bottlenecks, founder events and multiple introductions can also shaped the genetic structure in invasive species. This study aims to provide information about the process underlying invasion in plants including both landscape genetic and demographic process approach. The model species in this study was *Eschscholzia californica* Cham. (Papaveraceae), an invasive plant inhabiting Central Chile. A total of 506 individuals from 24 populations were genotyped for 8 microsatellites. Due a weak pattern of IBD, the null effect of the environmental variables on genetic structure and low gene flow between invasive populations, the results suggested a pattern of isolation-by-colonization were the populations genetic structure would be strongly determined by founder effects. Our result also pointed out the existence of multiple introductions of *E. californica* in Chile and LDD events that cannot be rejected. Overall, the evidence presented in this study suggest a low effect of the dispersal potential and founder effect inferred for *E. californica* in their invasive process in Central Chile.

Keywords: dispersal, founder events, multiples introductions, range expansion, landscape genetic.

Introduction

The biological invasions are defined as the arrival and successful spread of species into a new geographic range (*sensu* Richardson *et al.* 2000) a major main concern in conservation biology (Mack *et al.* 2000). Colonization and subsequent invasive spread implies consequences in the genetic structure of invasive species, therefore studies that enhance our understanding about this issue may highlight onto the microevolutionary process related with the invasiveness, constituting an important issue in order to elaborate management strategies to control invasion.

The dispersal has an important role in the invasive spread since it determines the ability of the species to disseminate to new non-invaded areas (With 2001). In this context, invasive species showed different degree of dispersion. In one hand, the majority of plants disperse over short distance, showed Short Distance Dispersal (SDD) and a low likelihood to propagule dispersal over long distances is assumed. If dispersal is moderated, patterns of isolation-by-distance (IBD) could be observed (Wright 1943). Under this model, a general increase in genetic differentiation among populations will be associated with increasing inter-population physical distance. On the other hand, other species showed a large dispersal of propagule. For example, Pergl *et al* (2011) showed that, for the invasive plant *Heracleum mantegazzianum,* the long-distance dispersal (LDD) is an important component of the population dynamics. Frequency LDD events could preserve the genetic diversity along the invaded range and may counterbalance the effect associated with IBD patterns (Fayard *et al.* 2009; Berthouly-Salazar *et al.* 2013).

Dispersal is one of the key mechanisms linking microevolutionary processes to landscape patterns since genetic data reflect only those dispersal movements that result in successful breeding (Broquet & Petit 2009). In this context landscape genetics approach is particularly well suited for assessing landscape influences on effective dispersal in invasive populations (e.g. Zalewski *et al.* 2009; Berthouly-Salazar *et al.* 2013). Lastly, during the colonization and expansion across their new habitat demographic process such as bottlenecks, founder events and multiple introductions (Chakraborty & Nei 1977; Clegg *et al.* 2002) can also shaped the genetic structure in invasive species.

Based on these evidences, it is possible to identify the key drivers of the genetic structure in invasive species: dispersal strategies, demographic process and landscape variables. In the last years, studies addressed questions about the mechanism that underline the biological invasions (e.g. Facon *et al.* 2006; Suarez & Tsutsui 2008) neglecting the importance of the landscape genetic approach in the plant invasion process.

Currently, the success of invasive species have been attributed to the high genetic diversity at in the invasive range, high propagule pressure, high genetic connectivity among invasive populations, good dispersal abilities and positive populations growth rates (With 2001; Ward 2008) among others. Phenotypic plasticity has also been invoked to explain plant invasive success (Bossdorf *et al.* 2005, Richards *et al.* 2006). Specific mechanisms involve the change from sexual to asexual reproduction (Haaga & Ebert 2004), or physiological changes in relation to environmental gradients (Paker *et al.* 2003; Lafuma 2003). However, research so far reveals no single pattern or unique combination of invasive attributes explaining or predicting the plant invasion, highlighting the importance of studies species by species.

As effective dispersal of many individuals over time is reflected as gene flow, it includes a successful movement, survival and reproduction (Zeller *et al.* 2012). Thus, different environmental conditions between the source and the receiving habitat can produce a differential establishment success of immigrants (Orsini *et al.* 2013b). For instance, climatic variables such as precipitations and temperature influence the effective dispersal raising as a key factor to determinate the location of suitable habitat to migrate, this is particularly important for the primary production and therefore the resource availability (Berhouly-Salazar *et al.* 2013). In case of a population covering an environmental gradient, such the latitudinal climatic gradient present in Chile (Mooney 1977, Di Castri 1991), individuals may present differential responses to local conditions (Spurgin *et al.* 2014). For example, elevation gradients or topography can influence effective dispersal between population (Murphy *et al.* 2010, Storfer *et al.* 2010). Lastly, for several invasive species inhabiting disturbed places (areas where native habitat has been cleared) may enhance the risk of invasive spread (Hobbs & Huenneke 1992), moreover, these areas can favors the movement and establishment of individuals (With 2004) and therefore, may improve effective dispersal among populations.

This study focused on the invasion of *Eschscholzia californica* Cham. (Papaveraceae) in Chile. *E. californica* is an herbaceous native from California and invasive in other Mediterranean areas of the world (Stebbins 1965 It has been reported in New Zealand, Tasmania, and mainland Australia (Cook 1962).Their arrival in Central Chile occurred between 1890s and early 1900s, presumably for ornamental purposes; later was naturally established along railways lines (Frias *et al.* 1975; Arroyo *et al.* 2000). It also probably accidental introduced together with alfalfa seeds from California (Hillman & Henry 1928). By 1975, this species had extended 240 km North of Santiago de Chile to Los Vilos and 520 km to the south of Santiago de Chile. Fifteen years ago, this species was present in central-southern Chile, covering a wide latitudinal (30° S - 38° S) as well as altitudinal range, from 0 to 2000 m.a.s.l. (Arroyo *et al.* 2000). *E. californica* is a diploid plant (Becker *et al.* 2005) and present sexual reproduction, their flowers are pollinated by insects and their seeds are apparently gravity-dispersed. Whereas it has been described as self-incompatible in their native area (Cook 1961), in Chile this species presented both self-incompatible and self-compatible reproductive system. Arredondo-Núñez (2011) showed a partial compatibility in the invaded range (0.2<ISI<1; mean compatibility index at individual level = 0.32) with high intra-population variability. Previous studies in Chile have also showed no significant differences in fecundity between reproductive phenotype, suggesting that both reproductive strategies would be adequate to face the environmental heterogeneity in *E. californica* (Arredondo-Ñúñez 2011). In addition, not variation in their fecundity under different altitudes in Central Chile has been observed (Arredondo-Núñez 2011; Peña-Gómez & Bustamante 2012), however, at higher altitude seed recruitment was significant lower (Peña-Gómez & Bustamante 2012). Under this, pre-dispersal process (i.e., production of seed) would be not limited by stressing conditions such as altitudinal gradient, in contrast, these studies have proposed the importance that might have the post-dispersal process (i.e., seed recruitment) in a successful establishment.

Given the life-history traits for this species, this study aims to provide information about the process underlying invasion of *Eschscholzia californica* by addressing the following objectives: Firstly, we used molecular markers to determine spatial structure of populations across the whole distribution of *E. californica* in Central Chile in order to provide evidence to infer the place of introduction and the direction of the expansion in the invaded range. For this, populations genetic, Bayesian approach and historical information were used in the model analysis. Secondly, in order to assess the dispersal abilities of *E. californica* in Chile we used population genetic analysis and estimated recent migration rates between pairs of populations. Thirdly, in order to evaluate the effect of the environmental and physical variables on both the effective dispersal and demographic process, we tested several landscape features such as climatic condition, topography or disturbed places and correlated them with population genetic structure.

Methods

Study area and data collection

In Central Chile, rainfall increase and temperature decrease from north to south (Mooney 1977; Di Castri 1991); topography is complex; there is a central valley is flanked by coastal and Andean mountain ranges (Mooney 1977). In order to perform the analysis, 24 sites (named as population hereafter) were randomly sampled between 30° to 38° S (Table1). In total, 506 individuals of *E. californica* were sampled for the genetic analysis ranging from 10 to 30 samples per local population. The geographic location of each local population was recorded using Garmin eTrex30 GPS.

Microsatellite typing

Genomic DNA was extracted from silica gel-dried leaf tissue using CTAB protocol described by Tel-zur *et al.* (1999). Eight polymorphic microsatellite loci described for the species were amplified follow Véliz *et al.* (2012) namely *Ecalifdi1, Ecalifdi3, Ecalifdi5, Ecalifdi9, Ecalifdi11, Ecalifdi16, Ecalifdi22* y *Ecaliftet1*. PCR products were genotyped in MacrogenInc (http://dna.macrogen.com/eng/). Allele scoring was performed with PeakScanner (Applied Biosystems) and GeneMarker (Softgenetics) software.

Genetic analysis

Within population analyses. Departures to the Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were estimated for each microsatellite locus in each sampling site with GENEPOP 4.2 (Rousset 2012). A permutation test (5000 permutations) was used to test the significance. Standardized allelic richness (Ae) and heterozigocity (H_O and H_E) were estimated using FSTAT 2.9.3.2 software (Goudet 2002). A series of consecutive founder effects during expansion will create a gradient in genetic diversity decreasing along the spatial axis of expansion (Austerlitz *et al.* 1997). To identify these gradients in diversity Ae and heterozigocity were used. Considering that the first introduction of *E. californica* was described in the coastal range, central Chile, specifically in Valparaíso and Viña del Mar (Arroyo *et al.* 2000), the core of invasion was estimated at 33° Lat. close to these localities, with a posterior expansion to northward and southward. Under this scenario, a Spearman correlation and permutations test (999 permutations) were used to evaluate the fit of Ae and heterozygocity of each population and their latitude into each expansion range, as we assumed that these populations are distributed along a 1-dimensional axis of expansion which coincides with their latitude. Lastly, detection of genetic bottleneck in invasive populations was performed through a graphic approach accorded to Luikart *et al.* (1998). It is important to note that methods commonly used to detect bottlenecks (Piry *et al.* 1999; Garza & Williamson 2001) were not conducted since they are based on the heterozygote excess, and our populations presented heterozygote deficit.

Among population analysis. In order to determine differences among population, both classical and Bayesian analysis were performed. First, F_{ST} values (Weir & Cockerham 1984) were estimated for each pair of sites analyzed using GENETIX (Belkhir *et al.* 1996). Additionally, conditional genetic distance (cGD; Dyer & Nason 2004) was estimated among populations using Rstudio package of R software (R Core Team 2014). This measure simultaneously considers the genetic covariance of all populations into the analysis including both, the direct and indirect genetic connectivity among populations, thus providing a more sensitive measure than F_{ST}; previous studies have provide insight about their effective use in landscape genetic analysis (Dyer *et al.* 2011; Garroway *et al.* 2011). Secondly, to identify the number of genetic clusters (K) the GENELAND v. 1.0.7 was conducted (Guillot *et al.* 2005) taking into account the spatial localization of the sampling sites. This software uses Bayesian inference with Markov Chain Monte Carlo (MCMC) algorithms to infer genetic spatial discontinuities (Guillot *et al.* 2005). For this study, five replicates with $1x10^5$ MCMC iterations was used, the maximum rate of the Poisson process was fixed at 1,000 for each run and the number of nuclei in the Poisson-Voronoi tessellation fixed at 450 for a better representation of the number of individuals in the data set. Also, to visualize the level of genetic structure, a Correspondence Analysis was performed using the Adegenet package (Jombart 2008) implemented in R software (R Core Team 2014). As a complementary analysis, a non-Bayesian partitioning method, using the program FLOCK 3.0 (Duchesne & Turgeon 2012) was run. This method does not require the accomplishment to the HWE, thus it was useful for *E. californica* populations in the invaded area in Chile. This analysis is based on the iterated reallocation into K sub-samples (which have the highest likelihood score) of all genotypes pooled, following the multilocus maximum likelihood method (Duchesne & Turgeon 2012). In this case, 50 runs (each with 20 iterations) were conducted for different number of k ranging from 2 to 24. The results obtained for the complete sequence of k values were used as input in the TRACE software, a complementary method used with FLOCK. The TRACE software computes pairs of distance statistic, based on the Jaccard similarity coefficients and then it clusters the populations samples for each K maximizing nearness between them (Duchesne & Turgeon 2012). To generate pairs of random distance, 25 runs for each k value were chosen. Finally, a tree based on distance of Nei (Nei *et al.* 1983), representing the relationships among the sampled sites was reconstructed using a neighbour-joining (NJ) algorithm with the ape R package (R Core Team 2014). To determine the nesting scene that maximized the among-variance component, an analysis of molecular variance (AMOVA; Excoffier *et al.* 1992) was realized on each genetic clusters obtained from the analysis described above.

Gene flow. Recent migration rates between pairs of populations were calculated with BAYES-ASS software (Wilson & Rannala 2003). This software use a Bayesian criteria for estimating recent migration using multilocus genotype information, thus without assumptions of HWE of populations. In order to perform this analysis, ten million iterations were run with a burn-in of nine million.

Scenario explaining introduction and geographical expansion of E californica. In order to describe the past demography and the most probably scenario of introduction of *E. californica* in Chile, the DIYABC software was used (Cornuet *et al.* 2008). Previous studies have verified that this method is adequate to modelling the introduction history of non-native species using complex scenarios (Beamont 2010; Brouat *et al.* 2014). Considering the historical information that proposes multiples introductions of *E. californica* in Chile and the scenario of expansion previously described, the analyses with DIYABC were performed using the previous two clusters obtained from among population comparisons and the four most possible well-founded colonization hypotheses (Fig. S1; Supporting information). Scenario 1 propose done non-sampled source population with a single introduction event into the cluster 1, which populations are present into the area historically considered of first introduction, and posterior expansion (Scenario1). Scenario 2 proposed a single introduction event into the cluster 2 and posterior expansion (Scenario 2). Scenario 3 proposed two introduction events into the each cluster from one unsampled source population (Scenario 3). Scenario 4 includes two independent introductions from two unsampled source populations into each one cluster (scenario 4). In all scenarios migrations between clusters after expansion or introduction event were not included and each cluster was considered as isolated group. Considering that *E. californica* has been described as perennial in the invaded region (Frias *et al.* 1975) with more than one reproductive event by year (personal observation), we estimated a maximum of two generation by year, hence the prior introductions and divergence times were defined to be within the last 240 generations. Each locus was estimated to have allele number ranging from 34 to 64. The set mutation model was the default parameters given by the DIYABC software. To discriminate among the scenarios proposed, we used the mean number of alleles/loci and mean gene diversity/loci (Nei 1987) to characterize the genetic diversity within population. On the other hand, F_{ST} between two samples (Weir & Cockerham 1984) and distance between two samples (δμ2; Golstein *et al.* 1995) were used as summary statistic to characterize the genetic differentiation between populations. These statistics were used previously in the analysis of invasion routes (e.g., Estoup & Guillemaud 2010; Brouat *et al.* 2014). The prior distributions were uniform and defined as follow: $10 < N < 5000$; $1 < Nib < 500$; $1 < t_{db} < 30$; $1 < t1 < t2 < t3 < 240$; were N is the effective population size, Nib is the effective population size under founder effect, t_{db} is the duration time of bottleneck and, t is the number in generations. For each scenario, 1×10^6 simulated data sets was computed. Scenarios were pre-evaluated by performing a PCA as implemented in DIYABC software and by ranking the observed data set against simulated data set. Scenario comparison were realized using a linear discriminant analysis (LDA) on summary statistics taking into account the default values and then using a logistic regression with original summary statistics realized with the 1% of the closest data set. Results were similar when using more or less closest data sets (e.g. 2%, 0.8%, 0.6%, data not shown). The best scenario was selected using posterior probability value, with the 95% confidence interval (CI) not overlapping among scenarios. We evaluated confidence in our choice of scenario by calculating type I and II error rates using the standard procedures available in DIYABC (Cornuet *et al.* 2008). Posterior distributions of each parameter was also obtained and their performance was evaluated by computing the relative bias (see the DIYABC notice).

Influence of landscape on effective dispersal

As already described, bioclimatic gradient and topography may affect the effective dispersal through populations of *E. californica* in Central Chile. Additionally, this species is well established in open areas (Cook 1962, Frias *et al.* 1975) and it has been described that both fecundity and size increased under open spaces, with low vegetative cover (Leger & Rice 2003). Considering this information, the hypothesis that environmental variables (climatic condition, altitude and land cover) and spatial physical distance, either separately or in combination, affect the effective dispersal among populations of *E. californica* in Central Chile was tested. For this, the relation of these landscape features and the genetic distance between pairs of populations was evaluated.

Climatic variables were represented by nineteen bioclimatic data layers obtained from the WorldClim dataset (Hijmans *et al.* 2005) at ~1-km spatial resolution, including monthly temperature and precipitation records from 1950 to 2000. Altitude was also obtained form WorldClim dataset (Hijmans *et al.* 2005). Land cover map is a remotely sensed variable (MODIS Land Cover), this was obtained from Global Land Cover Facility (http://www.landcover.org) and contains 22 categories since tree cover to non-forest/natural or developed land.

The effect of each landscape variables on effective dispersal of *E. californica* in Chile was evaluated using landscape genetic approach and SDMs (species distribution models). SDMs have been used in previous studies as adequate methods to generate information where species can live, and correlate this information with genetic differentiation of their populations (Wang *et al.* 2008; Velo-Anton *et al.*

2013). These models were performed with MaxEnt 3.3.3k (Phillip *et al.* 2006). This software predicts probability of presence of species based on their niche requirements. It uses different environmental variables (continuous, categorical) as predictors of the presence of species MaxEnt randomly selects 10,000 background points from the landscape and provides an estimate of probability of presence of a species. Presence records were obtained from geo-referenced points in Chile from museum specimens and our own collections (24 unique localities in this study). One SDMs for each landscape variable was generated into the invaded region of *E. californica* in Chile using default settings in MaxEnt. Current model accuracy of the models were checked using a random sample over the 25% dataset as training data and the remaining 75% were used in testing data for model validation. The area under the receiver operating characteristic (ROC), curve known as AUC (Swets 1988), was the metrics used to evaluate the performance of the models. The AUC ranges from 0.5 (random accuracy) to a maximum value of 1.0 (perfect discrimination).

Each raster maps generated by MaxEnt was converted into a 'conductance' surface that presumably constraints effective dispersal using the 'transition' function available in g distance package of R software (R Core Team 2014). Then, a connections scheme among the eight nearest cells was applied and a conductance grid was generated in which a pixel with a high probability of suitable habitat represent a low cost to dispersal. This function uses this grid and generates a matrix of movement probability between cells. Each cell in the matrix represented a cell in the original raster object and contained a value that can be interpreted as "conductance" for movement due to a determinate landscape feature. One "conductance" surface for movement (a spatial model of connectivity) was constructed for each landscape variable. Then, a pairwise resistance to dispersal was calculated among populations based on the least cost patch using gdistance package (R Core Team 2014). Optimal routes for dispersal were identified between populations following areas with high movement probability (McRae 2007), in this case, areas with high probability of suitable habitat. The cost of moving through the 'least cost path' was then calculated between all pairs of focal points or regions generating 276 pairwise ecological (cost) distances for each landscape variable.

To test the effect of each landscape feature on population genetic structure, Multiple Regression Matrices (Legendre *et al.* 1994) were used to examine the association of cGD and standardized F_{ST} ($F_{ST}/1-F_{ST}$) with ecological distances and physical distance. Physical distances between populations were calculated using the "rdist.earth" function from "fields" package (R Core Team 2014).

Results

Genetic analysis

Within population analysis. Pooling all samples, the number of alleles per locus was 17.4 in average, ranging from 9 for ecalifdi3 to 27 for ecaliftet9. Few loci present significant departures of Hardy–Weinberg Equilibrium, however, this deviation was not observed for all population for the same locus. Overall, a relative high inbreeding index value was found throughout invaded area of Chile ($F_{IS} = 0.33$; Table S1). At populations level, the rarefied allelic richness presented highest value at Puchuncaví (Ae = 3.92) and the lowest at Monte Zorro and Trehuaco (both with Ae = 1.86; Table 1). It is important to note that Monte Zorro and Vichuquén exhibited heterozygote excess pooling all loci and all other populations showed heterozygote deficit (Table 1).

A non-significant gradient in genetic diversity was recorded along the hypothesized axis of expansions (Fig. 1).For the north expansion range Ae and heterozigocity increased at lower latitudes (Ae: $R^2 = -0.33$; $P = 0.39$; H_E: $R^2 = -0.11$; $P = 0.78$; H_O: $R^2 = -0.33$; $P = 0.38$) and for the south expansion range Ae and H_E decrease at higher altitudes (Ae: $R^2 = -0.15$; $P = 0.59$; H_E: $R^2 = -0.06$; $P = 0.84$; H_O: $R^2 = 0.19$; $P = 0.51$). Following the graphical approach described by Luikart *et al.* (1998) a longer tail for populations was observed for *E. californica* (Fig. S2), suggesting bottleneck for all populations independently.

Among population analysis. Population differentiation, as measured by pairwise F_{ST} values (Tabla S2, Supporting Information) showed statistical differences for most pair of population. Few comparisons showed no statistical differences, for example two sites located northward (Recoleta and Illapel) and southward (Vichuquén and Hualqui). It is important to note that some populations geographically distant showed not differences in allele frequencies (for instance, Vichuquén and Recoleta; Fst = 0.09, P = 0.06).

Geneland software assigned the 24 population samples into 20 clusters. From the Correspondence Analysis, the first axis produces a cluster 1 with the following populations: Quinteros, Tunquén, Navidad, Coya, Las Chinchillas, Melipilla and Curicó and the cluster 2 with all other populations. The second axis placed the sites Coya, Rengo and Buchucuro separate from the other populations (Fig. 2a). The next axes were unable to define a higher genetic structure and separe one single sample population at time from the rest of cluster (result not showed).

Assignment analysis performed with FLOCK software was unable to determine an optimal k value; however, at k = 2 and k= 3 genetic clusters containing the following populations: Quinteros, Tunquén, Navidad and Coya (Fig. 2b for k = 2 and Tabla S3, Supporting Information). The sample populations Las Chinchillas and Puchuncaví were no allocated into any cluster.

The metric based on Nei's distance among samples identified Quinteros, Tunquén, Navidad, Coya, and Curicó as one cluster (Fig. 2c). Considering these results, two hypothetical nesting scenes were considered, one of them related the 20 genetic clusters identified with Geneland, and the other nesting scene identified 2 genetic clusters: cluster 1 that included Quinteros, Tunquén, Navidad and Coya and cluster 2 all other populations. These clusters were heterogeneous in size and both including population samples that are not neighbor. For the AMOVA analysis the partitioning populations suggested by Geneland software showed a not significant estimate among-groups differentiation (σ^2 $P = 0.99$), however, the partition that include 2 genetic clusters yielded a significant estimate among-groups differentiation (σ^2 $P < 0.001$) and explain the 18.15% of variance between groups.

Gene flow. Recent migration rates among populations of *E. californica* were extremely low with an average value of 0.011 and with a high standard deviation of ±0.010 for almost all pairs of populations (see Table S2, standard deviation for each estimate not showed).

Scenario explaining introduction and geographical expansion of E californica. Scenario3 resulted the most probable ($P = 0.48$, 95% CI = [0.47; 0.49]) which include two introduction events into the each cluster from one single unsampled ancestral population (Fig. 3). For this scenario, the probability of type I error was 0.83 and type II error was 0.10. The second most probably scenario was the Scenario 4 ($P = 0.40$, 95% CI = [0.39; 0.42]) which consider two introductions to each cluster originated from two un-sampled population source. The other two scenarios were poorly supported by the analysis (Fig. 3).

The posterior distributions of demographic parameters inferred under Scenario 3 showed to be robust for most parameters, as the relative bias calculated for each parameter was close to zero (Table S4, Supporting information). The effective population sizes was of 2,230 for cluster1 [*N1,* 95% CI = (692; 4,777)] and 4,130 for cluster2 [*N2,* 95% CI = (2,340; 4,970)]. The effective number of founder was of 114 for cluster1 [*N1b,* 95% CI = (12.8; 457)] and 289 for cluster2 [*N2b,* 95% CI = (56.6; 489)]. This represent a bottleneck with a reduction of more than 95% for both cluster, presumably with a duration of 21 generations [t_{db}, 95% CI = (1.5; 29.8)]. Assuming two generation by year for *E. californica,* we estimated the occurrence of 198 generations after the introduction [*t3*, 95% CI = (67; 240)] or approximately 100 years ago.

Influence of landscape on effective dispersal

The distribution of suitable habitat of *E. californica* obtained with MaxEnt showed with an average test AUC (±SD) value of performance of 0.73 (±0.07). F_{ST} was not related to physical distance between populations ($R = 0.017$; $P = 0.420$; Fig. 4a), however, our result showed a significant and low association between cGD and physical distance between populations ($R = 0.202$; $P < 0.01$; Fig. 4b) indicating a weak isolation-by-distance pattern. The landscape genetic approach indicated no significant relationship between both genetic distances and any ecological distances (climatic variables, altitude nor land cover; Table S5, Supporting information).

Discussion

Founder effect as the main demographic process

The main results of this study were i) the absence of isolation-by-distance pattern when F_{ST} and physical distance were used (as a measure of direct connectivity) and a weak pattern of isolation-by-distance when cGD and physical distance (as a measure of indirect connectivity) were considered; ii) the null effect of the environmental variables on genetic structure and iii) low gene flow among the sites studied. All of these results are in agreement with a pattern of isolation-by-colonization (IBC) where the genetic structure is expected to reflect the colonization (founder events) of the populations (Orsini *et al.* 2013b). Under this the genetic structure, the invasive process of *E. californica* seems to be strongly determined by founder effects and thus by colonization history rather than contemporary patterns of dispersal.

Invasive species are often introduced into novel areas in relatively small numbers of individuals resulting in founder effects (see Barrett *et al.* 2008). Added to this, persistent long-lasting founder effects are expected to be common in plants, as they exhibit rapid population growth (Orsini *et al.* 2013b). The founder effect inferred with ABC suggested introductions of *E. californica* by few individuals, the absence of gene-flow-drift equilibrium detected in populations as well as a recent genetic bottleneck evidenced by the graphic analysis (Fig. S2) indicate that founder effects in the colonization process are still running.

In cases where populations inhabit in an environmental gradient, individuals are expected to be adapted to local conditions (Spurgin *et al.* 2014) and populations may exhibit difficult signals to separate to the founder effect. Here, the variation of the abiotic conditions but low gene flow between populations of *E. californica* in Central Chile would indicate that these populations could experiment local adaptation. In this case, literature has sought to provide contrasting results. Leger and Rice (2007) described adaptive changes in growth patterns and flowering times under abiotic conditions in introduced Chilean populations of California poppy that resemble those seen in native Californian populations. In contrast, others studies have described that invasive populations are highly plastic in their physiological traits (Sanfuentes, 2014). Moreover, studies of reciprocal transplanting with populations of Central Chile have showed no local adaptation (Gallegos, in prep). Nevertheless, founder effect can persist despite adaptive differentiation in recently separated populations (Kolbe *et al.* 2012) and their influence can be evidenced even over evolutionary time scale (Spurgin *et al.* 2014).

Multiples introductions of E. californica in Chile

The genetic analyses of this study suggested that the invasion of *E. californica* in Chile was the result of at least two independent introductions. Specifically, the best scenario selected by ABC was the one involving two independent introduction of *E. californica*, one of them pointed out the introduction into the Central Chile. The area and time of second introduction is still unknown. However, ABC approach showed a low potential to discriminate between one or more unsampled source populations.

Thus, these introductions suggested repeated exchanges between Chile and a unique or distinct sources populations of *E. californica* at the time of its introduction. According with the historic records, this or these source populations would be localized in the native range of *E. californica.* Different small introductions in different areas are in agreement with historic registers of accidental introductions through the import of alfalfa seed and the inclusion of seeds in private gardens (Frias *et al.* 1975; Arroyo *et al.* 2000). As was observed in this study, including unsampled populations in the inferred evolutionary scenarios is an approach that can to make robust inferences with respect to scenario choice, even without samples of potential origin areas or hypothesis about introduction routes (e.g. Lombaert *et al.* 2011; Brouat *et al.* 2014). Overall, this issue of the colonization origin and areas where this species was introduced is an open question that deserves further analysis.

Multiple introductions may cause an increased genetic diversity along invasive range, helping to overcome the negative effects of founder events so that newly established populations can, and often do, arriving to the similar levels of genetic diversity to that in native populations (Lavergne & Molofsky 2007). Even when assumptions about the loss of genetic diversity in introduced populations require a sufficient sampling in both native and invaded range, a number of authors have noted that founder effect associated with initial colonization can significantly reduce genetic diversity in invasive populations. For instance, Henry *et al.* (2009) found that despite the multiple independent introductions registered of the invasive *Heracleum mantegazzianum,* their populations exhibited an important founder effect from the invasion process. Under this premise, the founder effect detected in the beginning of invasion of *E. californica* in Chile seems to be present still now regardless the multiples introductions registered in this study.

Dispersal strategies: SDD and LDD

As was explained before, our result pointed out the low gene flow and high genetic distance values that suggest poor dispersal potential in the California poppy. In addition, the genetic structure detected along Central Chile (low association between cGD and physical distance between populations) was consistent with a weak pattern of IBD. Even when our proposal of reproductive events was conservative, the populations of this species probably arrived 198 generations ago in Chile (under our ABC results), sufficient time to express a stronger IBD pattern. In fact, since their introduction, two different populations groups have already emerged (Fig. 2) explaining the 18.15% of variance between groups. However, IBD occurs under the assumption of populations in gene-flow-drift equilibrium (Wright 1943). Certainly, range expansions through a series of colonization events can produce genetic consequences that result in a non-equilibrium condition (Slatking & Excoffier 2012). Our results are in good agreement with this last proposition, since some populations of *E. californica* geographically separated showed not differences in allele frequencies indicating departure from gene-flow-drift equilibrium (Table S 2).

During the invasion process is expected a process of serial colonization events that produce a gradient in genetic diversity decreasing along the route of expansion (Austerlitz *et al.* 1997); this process often results in a pattern of IBD. It is important to note that allelic diversity is a more sensitive indicator of changes in population size than heterozygosity, because rare alleles can be lost easily (Nei *et al.* 1987). In this study, opposite patterns in genetic diversity on range expansions of *E. californica* was observed. Whereas a decline in allelic richness and H_E along the south expansion range, an increase in allelic richness or heterozigocity was observed along the north expansion range. Even when these correlations were not significant, our finding here provides a striking example of how others factors may affect levels of diversity following species introduction and generate different allele frequency gradients. In particular studies on dispersal strategies have given new insights about their consequences on genetic diversity in natural populations. Szövényi *et al.* (2012) showed null or weak IBD patterns and a positive, but not significant, correlation between allelic richness or H_E and distance from the core in populations of peat mosses. Berthouly-Salazar *et al.* (2013) detected a weak IBD pattern in the invasive European starlings and a weak but significant decrease in allelic richness along the expansion range. In all these cases, frequent long-distance-dispersal (LDD) would allow species to conserve, or even increase their genetic diversity along the expansion range (Fayard *et al.* 2009) eroding the IBD pattern (Berthouly-Salazar *et al.* 2013). In the light of the literature, the pattern of distribution of allele frequencies founded in the invasive populations of *E. californica* would be in agreement with the pattern expected under LDD events than species propagating only by short-distance dispersal. Thus, LDD events cannot be rejected has a process that could in part influence the genetic structure of invasive populations of *E. californica* in Chile. As in Berthouly-Salazar *et al.* (2013), the founder effects proposed here is not in contradiction with LDD events as the cumulative effect of founding events is not entirely prevented by LDD (Austerlitz *et al.* 1997; Comps *et al.* 2001).

Landscape variables as predictors of the effective dispersal

Our result indicated a null effect of the environmental variables on genetic structure. This result rejects our hypothesis about the importance of the environmental variables as predictors of the effective dispersal between invasive populations of *E. californica*. Effective seed dispersal necessarily requires successful establishment, as a consequence, seed migration coupled with successful germination and survival of seedling. Under the SDMs approach, optimal routes for dispersal were identified where the environmental conditions would influence the survival of individuals and thus a successful establishment. Considering this, the effective dispersal among populations is not influenced by the presence of these optimal routes with a high probability of success establishment. However, effective dispersal also integrates a successful movement. In plants, the adult organism is typically sessile with dispersal occurring via two types of propagules: pollen and seed. Pollen- and seed-mediated gene flow are both dependent on dispersal mechanisms extrinsic to the plant, be they biotic or abiotic, such that the spatial dimension of effective dispersal is determined not only by geographical distance but by features of the intervening landscape that influence the movement of pollen and seed dispersal vectors (Dyer *et al.* 2010). As was described by Anic *et al.* (2014) and Arredondo-Nuñez (2011), the pollination in of the California poppy is not limited by pollinators emerging the wind as the main probably vector of pollen transfer. After pollination, the resulted fruit is a long, slender pod that dries and splits, shooting seeds in all directions. Under this, the movement of dispersal vectors would be random with respect to landscape variation which could explain the non-significant fit between genetic distance and ecological distance.

Implications for invasion process

As our result suggested founder effects may be the key demographic process of the genetic structure in invasive populations of *E. californica*. Under this assumption, invasive populations of *E. californica* in Chile would be founded by few individuals who grow rapidly in size and established an advantage purely numeric. Given the poor dispersal potential, the established populations exchange a low number of migrants. By this way, reproductive success tends to rely heavily on few individuals and genetic differentiation arising during founding in the last hundreds of generations.

Fisher (1930) proposed that the capacity to adapt to novel conditions is proportional to genetic diversity present. Under this, it has been speculated that the erosion of genetic diversity due to founding events may limit the invasion process (Berthouly-Salazar *et al.* 2013). However, genetic diversity is not always a prerequisite for successful plant invasion (Poulin *et al.* 2005; Meimberg *et al.* 2006). Genetically similar individuals in the invading populations can tolerate environmental heterogeneity and a range of associated stresses, with potentially variable phenotypic expression in response to different conditions (Ward *et al.* 2008). Theoretical works have also demonstrated that demographic rates could be more important than dispersal capability in the prediction of the population persistence (South 1999; With & King 1999*b*). Populations of California puppies in Chile present a positive population grown (Alves 2012; Peña-Gómez & Bustamante 2012), indicating a potential to continue spread. For one hand, these conditions may indicate that their invasive process does not rely uniquely on the genetic connectivity of their pulations and therefore, the poor dispersal potential and founder effect inferred for *E. californica* may have not limited their invasive spread in Central Chile. On the other hand, this analysis suggested the importance of a successful establishment and variable phenotypic expression in response to different conditions on the invasion process of *E. californica.* Additionally, a history of multiples introductions, but also rare LDD events that cannot be discarded, but may have favored the invasion, both probably promoted by anthropogenic activities that occur along the invaded range.

References

Alves L (2012) *Demografía geográfica de una especie invasora de amplia distribución: el caso de Eschscholzia californica en Chile Central.*, Universidad de Chile.

Anic V, Henríquez CA, Abades SR, Bustamante RO (2014) Number of conspecifics and reproduction in the invasive plant Eschscholzia californica (Papaveraceae): is there a pollinator-mediated Allee effect? *Plant Biology* doi: 10.1111/plb.12293

Arredondo A (2011) *Diferenciación de los sistemas reproductivos y despliegues florales en Eschscholzia californica (Papaveraceae) en un gradiente altitudinal*, Universidad de Chile.

Arroyo MTK, Marticorena C, Matthei OLC (2000) Plant invasions in Chile: present patterns and future predictions. In: *Invasive species in a changing world* (eds. A. MH, J. HR), pp. 385-421. Island Press, Washington , D.C.

Austerlitz F, Heyer E (1999) Impact of demographic distribution and population growth rate on haplotypic diversity linked to a disease gene and their consequences for the estimation of recombination rate: Example of a French Canadian population. *Genetic Epidemiology* **16**, 2-14.

Barrett SCH, Colautti RI, Eckert CG (2008) Plant reproductive systems and evolution during biological invasion. *Molecular Ecology* **17**, 373-383.

Beaumont MA (2010) Approximate bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**, 379-406.

Becker A, Gleissbergy S, Smyth DR (2005) Floral and vegetative morphogenesis in California poppy (*Eschscholzia californica* Cham.)

Berthouly-Salazar C, Hui C, Blackburn TM, *et al.* (2013) Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Molecular Ecology* **22**, 5793-5804.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bossdorf O, Auge H, Lafuma L, Rogers W, Siemann E, Pratti D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* **144**, 1-11.

Broquet T, Petit E (2009) Molecular estimation of dispersal for ecology and population genetics. *Annual Review of Ecology, Evolution, and Systematics* **40**, 193-216.

Brouat C, Tollenaere C, Estoup A, *et al.* (2014) Invasion genetics of a human commensal rodent: the black rat *Rattus rattus* in Madagascar. *Molecular Ecology* **23**, 4153-4167.

Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with stepwise mutation model. *Evolution* **31**, 347-356.

Clegg SM, Degnan SM, Kikkawa J, *et al.* (2002) Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences* **99**, 8127-8132.

Comps B, Gömöry D, Letouzey J, Thiébaut B, Petit RJ (2001) Diverging trends between heterozygosity and allelic richness during postglacial colonization in the european beech. *Genetics* **157**, 389-397.

Cornuet JM, Santos F, Beaumont MA, *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713-2719.

Di Castri F (1991) An ecological overview of the five regions of the world with a Mediterranean climate. In: *Biogeography of mediterranean invasions* (eds. Groves RH, Di Castri F), pp. 3-16. Cambridge University Press, Cambridge, UK.

Cook SA (1962) Genetic system, variation, and adaptation in *Eschscholzia californica. Evolution* **16**, 278-299.

Duchesne P, Turgeon J (2012) FLOCK provides reliable solutions to the "number of populations" problem. *Journal of Heredity.*

Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology* **13**, 1713-1727.

Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology* **19**, 3746-3759.

Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why, how and so what? *Molecular Ecology* **19**, 4113-4130.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Facon B, Genton BJ, Shykoff J, *et al.* (2006) A general eco-evolutionary framework for understanding bioinvasions. *Trends in Ecology and Evolution* **21**, 130-135.

Fayard J, Klein EK, Lefevre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of evolutionary biology* **22**, 2171-2182.

Frías L D, Godoy R, Iturra P, *et al.* (1975) Polymorphism and geographic variation of flower color in chilean populations of *Eschscholzia californica. Plant Systematics and Evolution* **123**, 185-198.

Fisher RA (1930) *The genetical theory of natural selection*, 2nd edn. Clarendon Press, Oxford, New York.

Garroway CJ, Bowman J, Wilson PJ (2011) Using a genetic network to parameterize a landscape resistance surface for fishers, *Martes pennanti*. *Molecular Ecology* **20**, 3978-3988.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**, 463-471.

Goudet J (2002) *FSTAT 2.9.3.2: A program to estimate and test gene diversities and fixation indices.* http://www2.unil.ch/popgen/softwares/fstat.htm

Guillot G, Mortier F, Estoup A (2005) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes* **5**, 712-715.

Haag CR, Ebert D (2004) A new hypothesis to explain geographic parthenogenesis. *Annals Zoologica Fennici.* **41**, 539-544.

Henry P, Le Lay G, Goudet J, *et al.* (2009) Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Molecular Ecology* **18**, 2819-2831.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965-1978.

Hillman FH, Henry HH (1928) The incidental seeds found in commercial seed of alfalfa and red clover. *Proceedings of the international seed testing association* **6**, 1-20.

Hobbs RJ, Huenneke LF (1992) Disturbance, diversity, and invasion: implications for conservation. *Conservation Biology* **6**, 324-337.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-1405.

Kolbe JJ, Leal M, Schoener TW, Spiller DA, Losos JB (2012) Founder effects persist despite adaptive differentiation: a field experiment with lizards. *Science* **335**, 1086-1089.

Lafuma L, Balkwill K, Imbert E, Verlasque R, Maurice S (2003) Ploidy level and origin of the European invasive weed *Senecio inaequidens* (Asteraceae). *Plant Systematics and Evolution* **243**, 59-72.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the national academy of sciences of the united states of america* **104**, 3883-3888.

Leger EA, Rice KJ (2003) Invasive California poppies (*Eschscholzia californica* Cham.) grow larger than native individuals under reduced competition. *Ecology Letters* **6**, 257-264.

Leger EA, Rice KJ (2007) Assessing the speed and predictability of local adaptation in invasive California poppies (*Eschscholzia californica*). *Journal of Evolutionary Biology* **20**, 1090-1103.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. *Evolution* **48**, 1487-1499.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of heredity* **89**, 238-247.

Mack RN, Simberloff D, Mark Lonsdale W, *et al.* (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* **10**, 689-710.

McRae BH (2006) Isolation by resistance. *Evolution* **60**, 1551-1561.

Mooney HA (1977) *Convergent evolution in Chile and California: Mediterranean climate ecosystems*, Pensilvania.

Meimberg H, Hammond JI, Jorgensen CM, *et al.* (2006) Molecular evidence for an extreme genetic bottleneck during introduction of an invading grass to California. *Biological Invasions* **8**, 1355-1366.

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. *Ecology* **91**, 252-261.

Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of molecular evolution* **19**, 153-170.

Orsini L, Vanoverbeke J, Swillen I, Mergeay J, De Meester L (2013b) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology* **22**, 5983-5999.

Parker IM, Rodriguez J, Loik ME (2003) An evolutionary approach to understanding the biology of invasions: local adaptation and general-purpose genotypes in the weed *Verbascum thapsus*. *Conservation Biology* **17**, 59-72.

Peña-Gómez FT, Bustamante RO (2012) Life history variation and demography of the invasive plant Eschscholzia californica Cham. (Papaveraceae), in two altitudinal extremes, Central Chile. *Gayana Botánica* **69**, 113-122.

Pergl J, Müllerová J, Perglová I, Herben T, Pyšek P (2011) The role of long-distance seed dispersal in the local population dynamics of an invasive plant species. *Diversity and Distributions* **17**, 725-738.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231-259.

Poulin J, Weller SG, Sakai AK (2005) Genetic diversity does not affect the invasiveness of fountain grass (*Pennisetum setaceum*) in Arizona, California and Hawaii. *Diversity and Distributions* **11**, 241-247.

Piry S, Luikart G, Gournet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity* **90**, 502-503.

R Development Core Team (2014) R: A Language and Environment for Statistical Computing. R version 3.0.3. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Richards CL, Bossdorf O, Muth NZ, Gurevic J, Pigliucci M (2006) Jack of all trades, master of some? On the role of phenotypic plasticity in plant invasions. *Ecology Letters* **9**, 981-993.

Richardson DM, Pyšek P, Rejmánek M, *et al.* (2000) Naturalization and invasion of alien plants: concepts and definitions. *Diversity and Distributions* **6**, 93-107.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular ecology resources* **8**, 103-106.

Sanfuentes C (2014) *Tolerancia de Eschscholzia californica en ambientes saturados por cobre: ¿Adaptación local como una respuesta evolutiva?*, Universidad de Chile.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* **191**, 171-181.

Shigesada N, Kawasaki K, Takeda Y (1995) Modeling stratified diffusion in biological invasions. *American Naturalist* **146**, 229-251.

South A (1999) Dispersal in spatially explicit population models. *Conservation*

Biology **13**, 1039-1046.

Spurgin LG, Illera JC, Jorgensen TH, Dawson DA, Richardson DS (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology* **23**, 1028-1039.

Stebbins GL (1965) Colonizing species of the native California flora. In: *The Genetics of Colonizing Species* (eds. Baker HG, Stebbins GL). Academic Press, New York.

Storfer A, Murphy MA, Spear SF, Holderegger R, Waits LP (2010) Landscape genetics: where are we now? *Molecular Ecology* **19**, 3496-3514.

Suarez AV, Tsutsui ND (2008) The evolutionary consequences of biological invasions. *Molecular Ecology* **17**, 351-360.

Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* **240**, 1285-1293.

Szövényi P, Sundberg S, Shaw AJ (2012) Long-distance dispersal and genetic structure of natural populations: an assessment of the inverse isolation hypothesis in peat mosses. *Molecular Ecology* **21**, 5461-5472.

Véliz D, Gauci R, Bustamante RO (2012) Characterization of novel microsatellite markers for *Eschscholzia californica* (Papaveraceae), an invasive species in central Chile. *American Journal of Botany* **99**, e366-e368.

Velo-Antón G, Parra JL, Parra-Olea G, Zamudio KR (2013) Tracking climate change in a dispersal-limited species: reduced spatial and genetic connectivity in a montane salamander. *Molecular Ecology* **22**, 3261-3278.

Ward SM, Gaskin JF, Wilson LM (2008) Ecological genetics of plant invasion: what do we know? *Invasive Plant Science and Management* **1**, 98-109.

Weir B, Cockerham C (1984) Estimating F-Statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177-1191.

With KA (2002) The landscape ecology of invasive spread. *Conservation Biology* **16**, 1192-1203.

With KA (2004) Assessing the risk of invasive spread in fragmented landscapes. *Risk Analysis* **24**, 803-815.

With KA, King AW (1999a) Dispersal thresholds in fractal landscapes: a consequence of lacunarity thresholds. *Landscape Ecology* **14**, 73-82.

Wright S (1943) Isolation by distance. *Genetics* **28**, 114-138.

Yu-Huang W, Kuoh-Cheng Y, Cara Lin B, Liang-Kong L (2008) Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology* **23**, 989-1000.

Zeller K, McGarigal K, Whiteley A (2012) Estimating landscape resistance to movement: a review. *Landscape Ecology* **27**, 777-797.

Table 1 Summary of sampling sites, spatial coordinates, abbreviation and genetic values for each site/population. n: number of individuals sampled, A:mean number of alleles; Ae: richness rarefied to n=4; H_E: expected heterozygosity; H_O: observed heterozyosity. List of population follow a north–south gradient.

Sampling site	Latitude	Longitude	Abbreviation	n	A	Ae	H_O	H_E
Recoleta	-30.511	-71.101	Rec	28	7.38	3.64	0.43	0.63
Ovalle	-30.614	-71.129	Ova	12	4.25	3.18	0.32	0.52
Las Chinchillas	-31.571	-71.109	LCh	10	3.75	2.97	0.23	0.51
Illapel	-31.652	-71.206	Ill	22	4.38	2.94	0.35	0.58
Caimanes	-31.933	-71.137	Cai	27	3.00	2.17	0.23	0.37
Palquico	-32.259	-71.138	Pal	24	4.25	2.59	0.25	0.45
Pedegua	-32.353	-71.069	Ped	29	4.75	2.63	0.29	0.46
Puchuncaví	-32.732	-71.442	Puch	19	5.63	3.92	0.27	0.69
Quinteros	-32.792	-71.479	Qui	24	4.63	2.95	0.29	0.52
Tiltil	-33.156	-70.896	Til	9	3.63	2.82	0.30	0.45
Tunquen	-33.274	-71.652	Tun	24	4.75	2.75	0.45	0.52
Melipilla	-33.385	-71.393	Mel	28	5.63	3.23	0.30	0.49
Navidad	-33.964	-71.865	Nav	24	4.25	2.88	0.32	0.57
San Francisco de Mostazal	-33.972	-70.707	SFco	29	6.23	3.77	0.27	0.68
Coya	-34.195	-70.563	Coy	24	6.25	3.68	0.34	0.66
Rengo	-34.402	-70.871	Ren	22	5.63	3.71	0.35	0.69
Vichuquén	-34.778	-72.045	Vich	10	4.25	3.41	0.50	0.44
Iloca	-34.953	-72.184	Ilo	10	4.50	2.53	0.37	0.58
Curicó	-34.993	-71.229	Cur	12	3.25	3.20	0.39	0.57
Peyuhue	-35.826	-72.612	Pey	12	3.13	2.64	0.41	0.55
Buchucuro	-36.078	-72.794	Buch	10	3.63	2.79	0.33	0.54
Monte Zorro	-36.329	-72.835	Mzo	10	2.13	1.86	0.33	0.28
Trehuaco	-36.426	-72.664	Tre	20	3.63	1.86	0.33	0.50
Hualqui	-36.893	-73.020	Hua	10	4.88	3.63	0.38	0.61

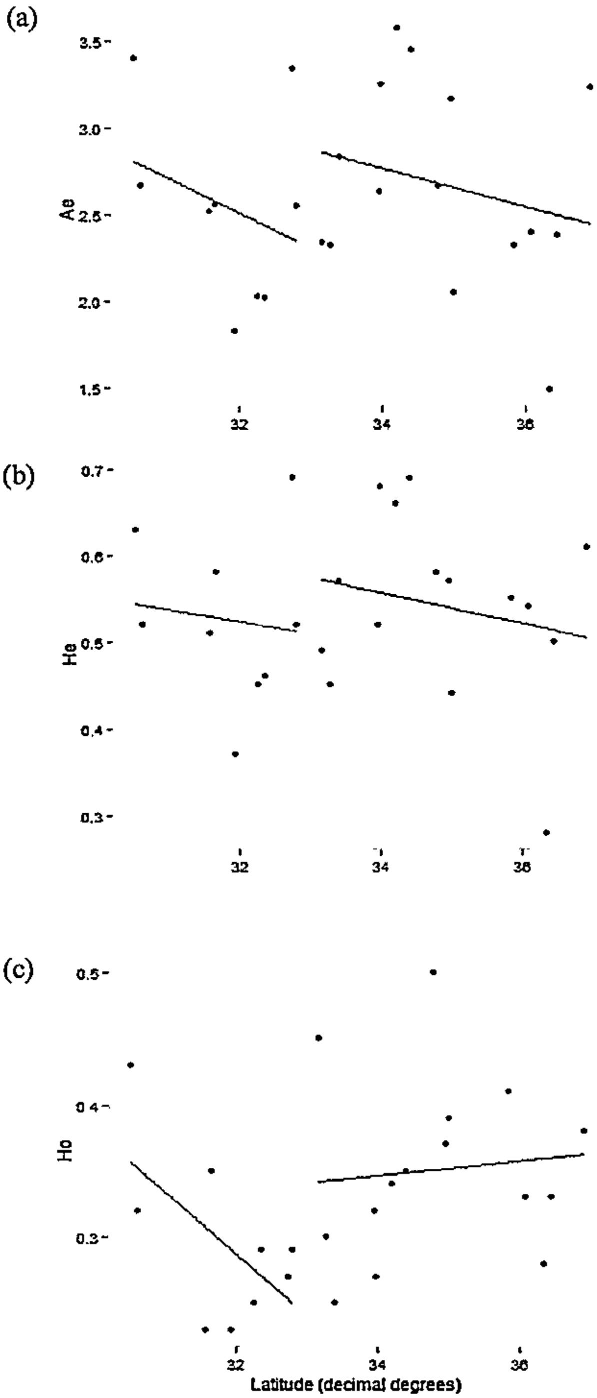


Fig. 1 Pattern of changes in genetic diversity along the hypothesized invasion pathway. The latitude of populations allocated along the north (in red) and south (in blue) invasion pathway (considering 33° S as invasion core).

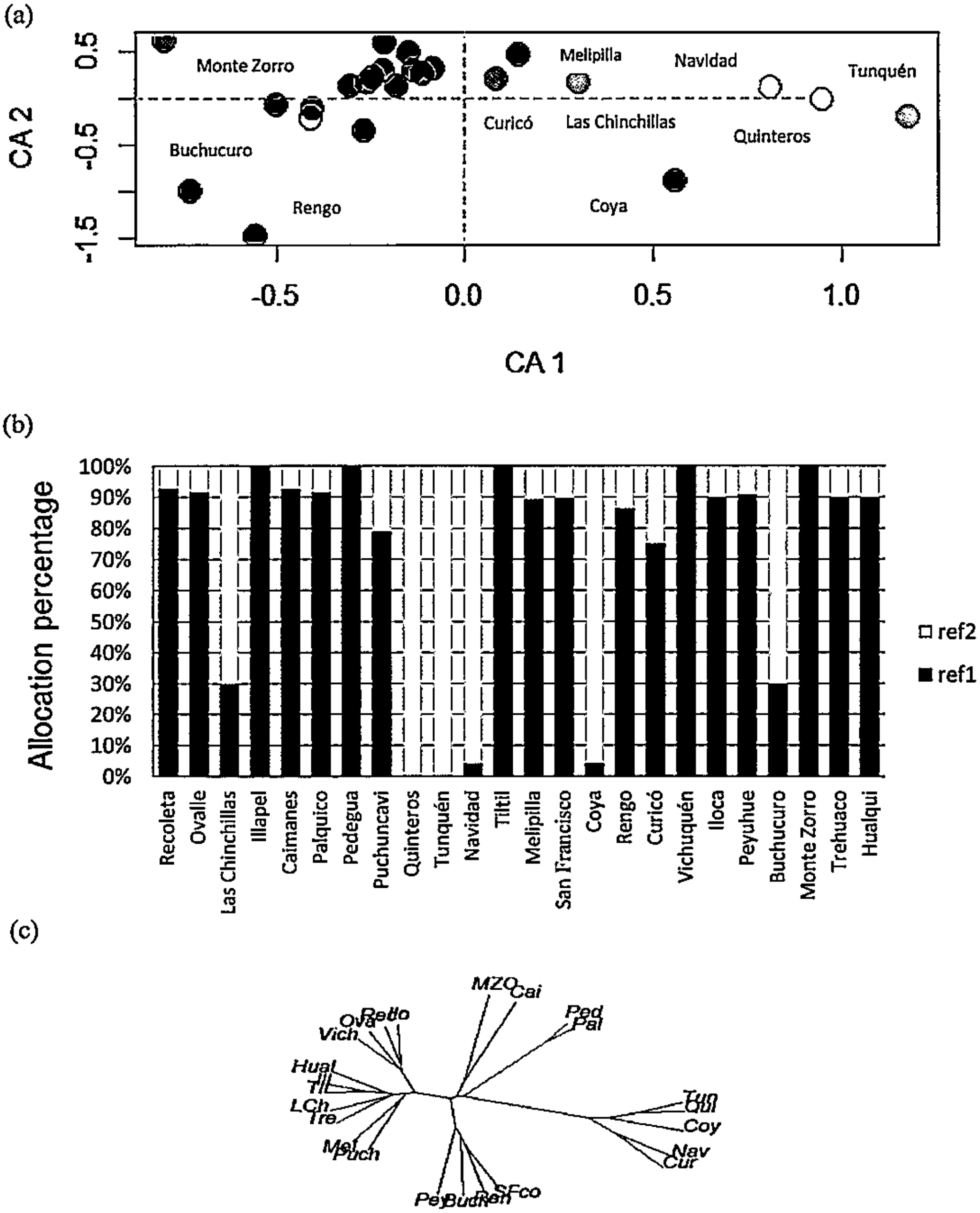


Fig. 2 Cluster analysis performed on the populations of *E. californica* in Chile. (a) Representation of the populations on a correspondence analysis (CA) plot. CA1 = 12.4% and CA2 = 10.8%. (b) Allocation percentage values per sampling site for K = 2 determined by FLOCK. (c) Tree based on the Nei's distance using neighbor-joining (NJ) algorithm. For abbreviations of samples sites, see Table 1.

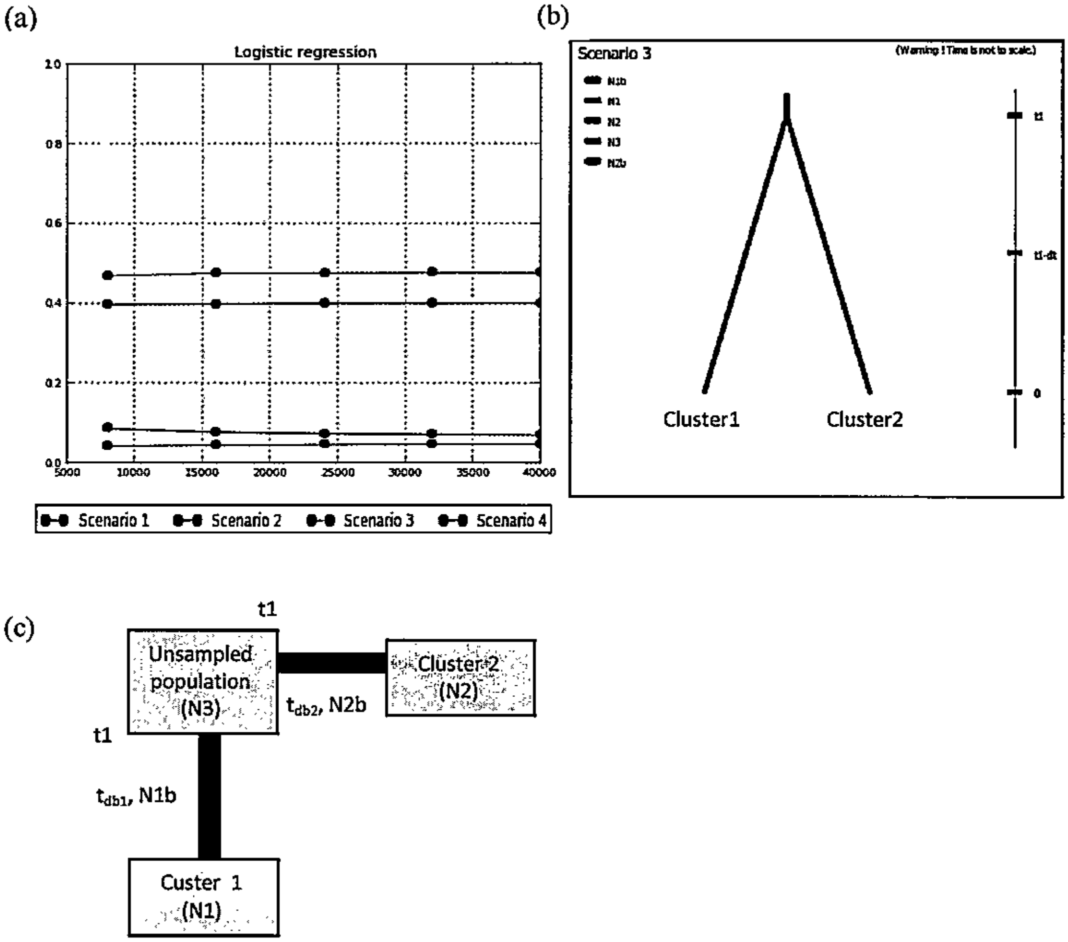
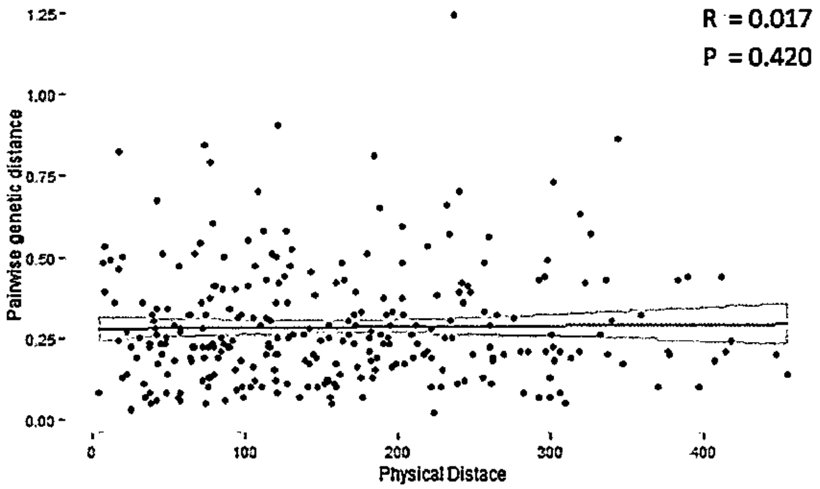


Fig. 3 Result of the DIYABC analysis. (a) Logistic regression representing the posterior probability of each scenario. (b) and (c) the most likely scenario (Scenario3) with Cluster 1 = Quinteros, Tunquén, Navidad and Coya, and Cluster 2 = all other populations. Description in Method section.

(a)



(b)

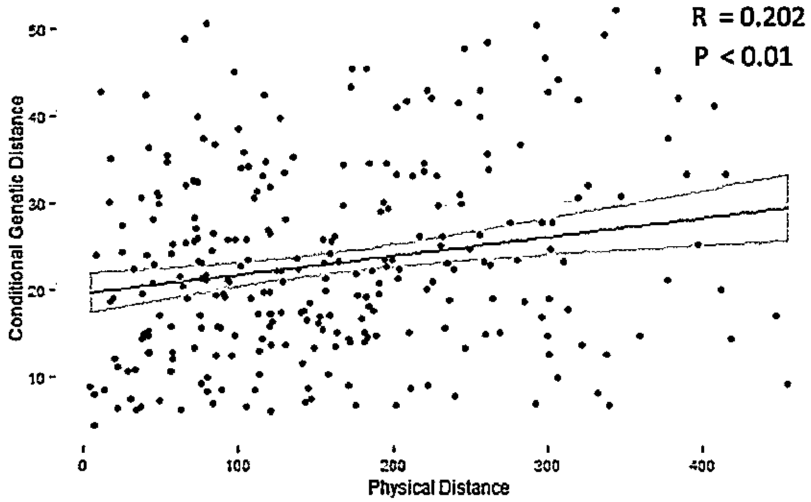


Fig. 4 Coefficient of regression (R) and P-value for regression on physical distance between populations fitting to (a) pairwise genetic distance and (b) conditional genetic distance for *E. californica.*

Table S1 Summary of parameters per loci. He: unbiaised expected heterozygosity; Ho: observed heterozyosity; A: number of alleles; and F_{IS} (Weir & Cockerham 1984).

				All populations
Locus	A	H_O	H_E	F_{IS}
ecalifdi5	10	0.21	0.77	0.614
ecalifdi22	23	0.33	0.86	0.507
ecalifdi11	24	0.52	0.80	0.246
ecaliftet1	16	0.23	0.78	0.559
ecalifdi1	13	0.09	0.31	0.653
ecalifdi3	9	0.65	0.65	-0.100
ecaliftet9	27	0.58	0.70	0.082
ecalifdi16	18	0.44	0.81	0.317
Over all loci	17.4	0.37	0.70	0.210

Table S2 Pairwise F_{ST} for 24 populations (above diagonal). Significant values were calculated with permutation test (10 000 permutations). *= $P \leq 0.05$; **= $P < 0.01$; bold value are not significant. Recent migration rates among populations of *E. californica* estimated with BAYES-ASS (below diagonal). For abbreviations of populations, see Table 1. List of population follow a north–south gradient.

	Rec	Ova	LCh	Ill	Cai	Pal	Ped	Puch	Qui	Til	Tun	Mel	Nav	SFco	Coy	Ren	Vich	Ilo	Cur	Pey	Buch	MZo	Tre	Hua
Rec	-	**0.06**	**0.10**	0.09**	0.23**	0.19**	0.21**	0.12**	0.20**	**0.06**	0.26**	0.16**	0.18**	0.08**	0.17**	0.13**	**0.09**	**0.05**	0.16**	0.18**	0.17**	0.23**	0.09**	**0.07**
Ova	0.04	-	**0.16**	0.11*	0.36**	0.27**	0.29**	**0.12**	0.24**	**0.06**	0.33**	0.19**	0.20**	0.14**	0.20**	0.21**	**0.06**	**0.06**	0.19**	**0.19**	0.17*	0.31**	0.11*	**0.07**
LCh	0.01	0.01	-	**0.07**	0.33**	**0.14**	0.18**	**0.10**	0.17**	**0.09**	0.24**	0.12*	0.16**	0.10*	0.19**	0.17**	**0.12**	**0.11**	0.21**	0.27**	0.22**	**0.38**	**0.10**	**0.07**
Ill	0.01	0.02	0.03	-	0.28**	0.15**	0.16**	0.10**	0.18**	**0.02**	0.24**	0.17**	0.18**	0.10**	0.18**	0.18**	**0.09**	0.11*	0.21*	0.16**	0.22**	0.31**	0.10**	**0.05**
Cai	0.01	0.01	0.01	0.01	-	0.33**	0.34**	0.32**	0.38**	0.30**	0.46**	0.29**	0.37**	0.23**	0.31**	0.32**	0.35**	0.30**	0.36**	0.37**	0.42**	0.46**	0.36**	0.30**
Pal	0.01	0.01	0.03	0.01	0.01	-	**0.02**	0.18**	0.26**	0.15*	0.33**	0.20**	0.24**	0.14**	0.28**	0.23**	0.25**	0.22**	0.29**	0.29**	0.33**	0.45**	0.22**	**0.12**
Ped	0.01	0.01	0.01	0.01	0.01	0.08	-	0.21**	0.30**	0.19**	0.37**	0.25**	0.29**	0.18**	0.30**	0.24**	0.28**	0.25**	0.33**	0.30**	0.34**	0.44**	0.22**	0.15*
Puch	0.01	0.01	0.01	0.01	0.01	0.04	0.01	-	0.16**	**0.10**	0.24**	0.14**	0.20**	0.09**	0.16**	0.16**	**0.07**	0.10*	0.23**	0.15*	0.17**	0.33**	0.13**	**0.06**
Qui	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.19**	0.07*	0.16**	0.13**	0.22**	0.15**	0.25**	0.24**	0.25**	0.24**	0.24**	0.29**	0.41**	0.28**	0.15*
Til	0.01	0.02	0.01	0.04	0.01	0.01	0.01	0.01	0.01	-	0.27**	0.15**	0.18*	**0.07**	0.18**	0.18**	**0.08**	**0.09**	0.20**	0.18**	0.22**	0.34**	**0.12**	**0.07**
Tun	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.08	0.01	-	0.23**	0.16**	0.28**	0.15**	0.30**	0.32**	0.32**	0.27**	0.29**	0.36**	0.47**	0.34**	0.22**
Mel	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.16**	0.15**	0.20**	0.19**	0.17*	0.17**	0.17**	0.27**	0.24**	0.34**	0.16**	0.11*
Nav	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.20**	0.19**	0.24**	0.23**	0.20**	0.14**	0.28**	0.30**	0.40**	0.24**	0.12**
SFco	0.02	0.01	0.01	0.02	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	-	0.17**	0.09**	0.09*	**0.07**	0.21**	0.15**	0.15*	0.27**	0.12**	**0.07**
Coy	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.18**	0.20**	0.20**	0.22**	0.20**	0.20**	0.35**	0.24**	0.16**
Ren	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.014	0.01	-	0.17**	0.18**	0.21**	0.20**	**0.10**	0.32**	0.22**	0.15**
Vich	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.05	0.01	0.01	0.01	0.01	0.01	0.01	-	**0.05**	0.22**	0.19**	0.18**	**0.30**	**0.13**	**0.11**
Ilo	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.21**	0.18*	0.18*	**0.29**	**0.09**	**0.06**
Cur	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.31**	0.30**	0.40**	0.25**	0.17**
Pey	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.20**	0.41**	0.26**	**0.11**
Buch	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.06	0.01	0.01	0.01	0.01	-	0.40**	0.23**	**0.18**
MZo	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	0.36**	**0.35**
Tre	0.01	0.01	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	-	**0.09**
Hua	0.01	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	0.01	0.02	-

Table S3 Allocation (ref) of populations performed with TRACE software (a complementary method used with FLOCK) for k = 2 and k = 3 values. For abbreviations of samples populations, see Table 1.

k = 2	
ref1	Quiteros Tunquén, Navidad, Coya
ref2	Recoleta, Ovalle, Illapel, Palquico, Pedegua, Tiltil, Melipilla, San Francisco, Rengo,Vichuquén, Iloca, Peyuhue, Monte Zorro, Trehuaco, Hualqui
k = 3	
ref1	Quiteros Tunquén, Navidad, Coya
ref2	Illpel, Palquico, Pedegua, Trehuaco, Hualqui
ref3	Recoleta, Caimanes, Rengo, Iloca, Buchucuro, Monte Zorro

Table S4 ABC estimations of posterior distributions of parameters under scenario 3 using the prior distribution (see Method section) and the confidence estimates for each parameter. The table shown the median of the posterior distributions for each parameter and the 95% CI and the relative bias (see the DIY-ABC notice).

Parameter	Median	95%CI	Bias
N1	2230	[692; 4777]	0.074
N2	4130	[2340; 4970]	0.037
N3	3290	[1240; 4910]	0.020
t_1	198	[67; 240]	-0.015
t_{db1}	20.9	[1.5; 29.8]	0.246
N1b	114	[12.8; 457]	0.261
N2b	289	[56.6; 489]	0.235
$\mu_{microsat}$	8.6e-4	[5.3e-4; 9.9e-4]	-0.033
P	0.27	[0.15; 0.30]	0.017
SNI	2.0e-6	[1.9e-8; 1.0e-5]	-0.845

Table S5 Coefficient of multiple determination (R^2) and probabilities (P) for multiple regressions on distance matrices models fit to pairwise genetic distance (F_{ST}) and conditional genetic distance (cGD) for *E. californica*. Empty cells represent ecological variables whose inclusion in the model was not found to be significant (at $\alpha = 0.05$).

	F_{ST}		cGD	
Predictor	R^2	P	R^2	P
Annual mean temperature	0.0008	0.8256	0.0026	0.2911
Mean diurnal range	0.0001	0.9360	0.0015	0.5863
Isothermality	0.0001	0.9779	0.0012	0.6383
Temperature seasonality	0.0004	0.9530	0.0007	0.7062
Max temperature of warmest month	0.0002	0.9170	0.0022	0.5655
Min temperature of coldest month	0.0003	0.9529	0.0024	0.4097
Temperature annual range	0.0002	0.9657	0.0012	0.6567
Mean temperature of wettest quarter	0.0001	0.9235	0.0024	0.3948
Mean temperature of driest quarter	0.0003	0.9607	0.0024	0.3392
Mean temperature of warmest quarter	0.0003	0.8990	0.0043	0.3900
Mean temperature of coldest quarter	0.0002	0.9020	0.0040	0.3989
Annual precipitation	0.0047	0.6490	0.0077	0.3001
Precipitation of wettest month	0.0002	0.9170	0.0075	0.2757
Precipitation of driest month	0.8130	0.0009	0.0035	0.1407
Precipitation seasonality	0.0006	0.8457	0.0033	0.1687
Precipitation of wettest quarter	0.0022	0.7498	0.0022	0.2790
Precipitation of driest quarter	0.0040	0.9790	0.0031	0.1576
Precipitation of warmest quarter	0.0006	0.8610	0.0106	0.1949
Precipitation of coldest quarter	0.0023	0.7390	0.0021	0.2720
Land cover	0.0496	0.9115	0.0026	0.4996
Altitude	0.0001	0.9441	0.0022	0.5177

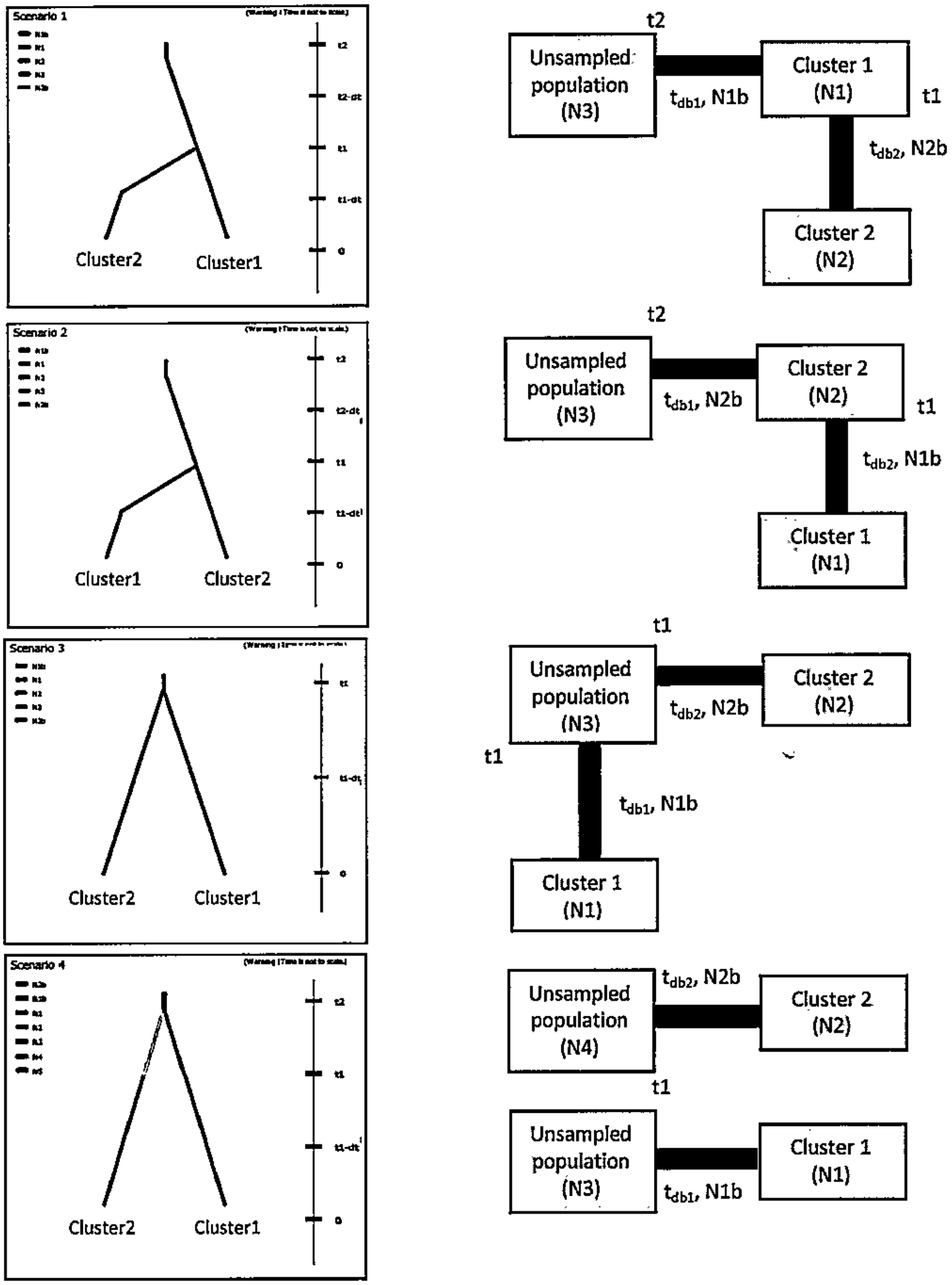


Fig. S1 Scenarios used in the Approximate Bayesian Computation (ABC) analyses. Cluster 1 = Quinteros, Tunquén, Navidad and Coya; Cluster 2 = all other populations). N1–N5: effective population sizes for populations 1–5. N1b and N2b: effective number of founder specimens for invasive populations 1 and 2. t_{db1} and t_{db2}: time duration of bottleneck for populations 1 and 2. t is the number in generations, with t1<t2. Detailed description in Method.

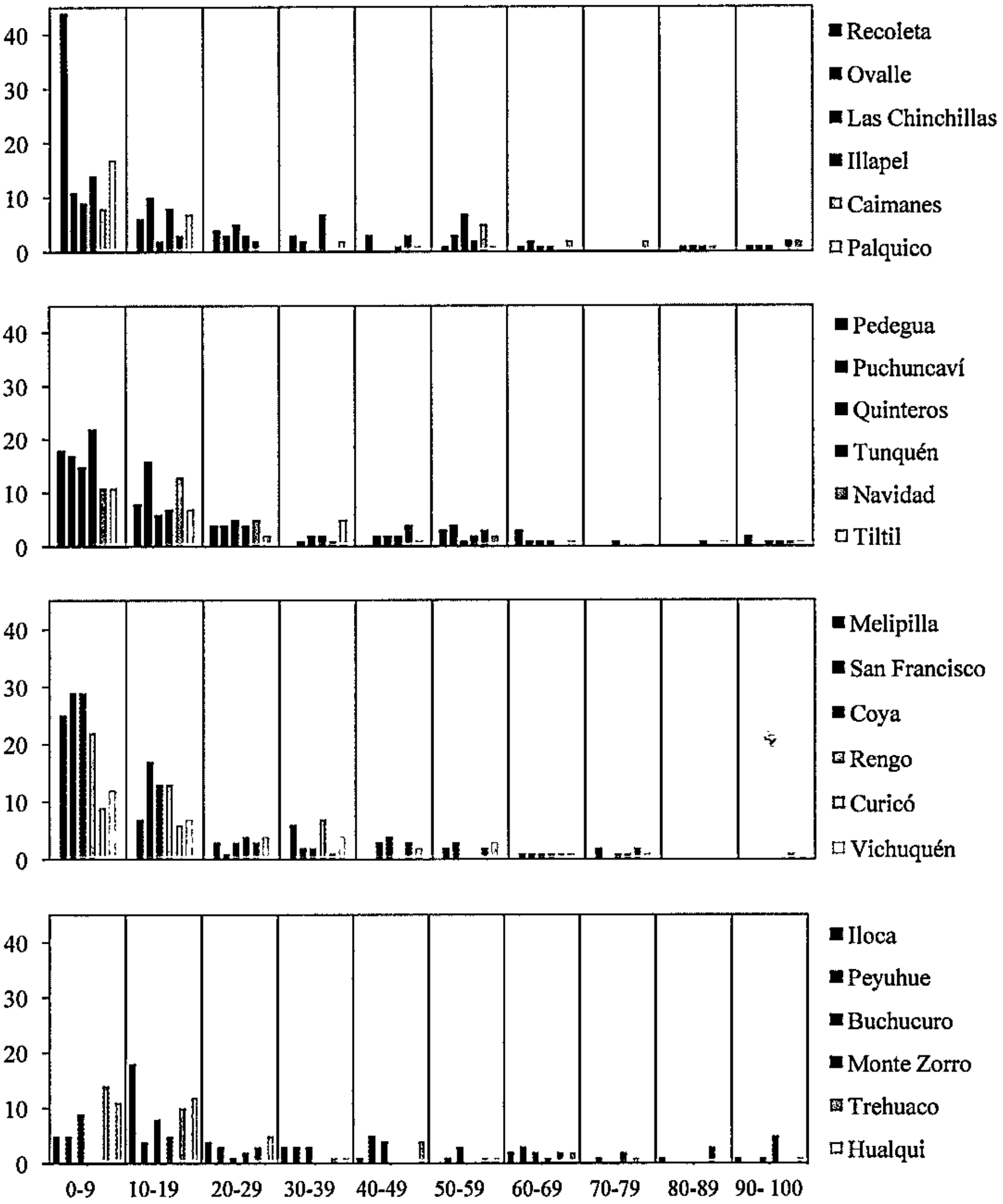


Fig. S2 Graphic approach according to Luikart *et al.* (1998) based on the count of occurrence of alleles (abscissa) in each frequency class (ordinate). This approach was used to detect recent genetic bottleneck considering all alleles in the 24 populations of *E. californica*. In non-bottlenecked populations, there are many alleles that occur at low frequencies and few alleles that occur with high frequency. List of population follow a north–south gradient.