

# Drift effects on the multivariate floral phenotype of *Calceolaria polyrhiza* during a post-glacial expansion in Patagonia

C. C. MAUBECIN\*, A. COSACOV\*, A. N. SÉRSIC\*, J. FORNONI\*† & S. BENITEZ-VIEYRA\*

\*Laboratorio de Ecología Evolutiva y Biología Floral, Instituto Multidisciplinario de Biología Vegetal (IMBIV), CONICET and Universidad Nacional de Córdoba, Córdoba, Argentina

†Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México, México

## Keywords:

common principal component analysis;  
genetic drift;  
Patagonia;  
phenotypic integration;  
Pleistocene glaciations;  
**P**-matrix;  
Random Skewers;  
variance–covariance structure.

## Abstract

Quaternary environmental changes substantially impacted the landscape and promoted rapid evolutionary changes in many species; however, analyses of adaptive phenotypic variation in plants have usually neglected the underlying historical context. Here, we associate phylogeography and phenotypic evolution by analysing the divergence of *Calceolaria polyrhiza* multivariate floral phenotype after a Pleistocene post-glacial expansion in Patagonia. Phenotypic matrix (**P**) properties (size, shape, orientation and phenotypic integration) of six refugium and six recent populations from two different phylogroups were compared following different approaches. We found that **P**-matrix shape and orientation remained stable despite the strong phylogeographic footprint of post-glacial expansion. However, average proportional reductions in matrix size supported the expectation that drift had a significant effect on the floral phenotype in the northern phylogroup. When phylogeographic history was not included in the analyses, the results overestimated phenotypic differences, whereas under explicit phylogeographic control, drift appeared as the best explanation for matrix differences. In general, recent populations showed a larger phenotypic divergence among them, but a lower overall phenotypic variation than refugium populations. Random Skewers analyses indicated a lower potential response to selection in recently colonized populations than in refugium populations. We discuss that the combination of phylogeographic analyses with geographical distribution of functional phenotypic (genotypic) variation is critical not only to understand how historical effects influence adaptive evolution, but also to improve field comparisons in evolutionary ecology studies.

## Introduction

Quantitative genetic theory predicts that genetic drift will promote random phenotypic divergence among populations (Lande, 1979), together with proportional changes, *on average*, in the variance–covariance matrix describing the phenotype (i.e. changes in all traits at

the same time, affecting matrix size, but not shape and orientation, Roff, 2000; Phillips *et al.*, 2001; Jones *et al.*, 2003). Furthermore, drift is expected to constrain the population response to selection due to a random reduction in the amount of standing genetic variation and increased levels of inbreeding (Phillips *et al.*, 2001). However, determining the role of drift and selection on the evolution of variance–covariance matrices is still an unresolved issue (Roff, 2000). In addition, this is also a central aspect to understand the stability of the variance–covariance matrix, a necessary condition to predict the response to selection (Lande, 1979).

Correspondence: Constanza C. Maubecin, Laboratorio de Ecología Evolutiva y Biología Floral, Instituto Multidisciplinario de Biología Vegetal (IMBIV), CONICET and Universidad Nacional de Córdoba, CC 495, X5000ZAA, Córdoba, Argentina. Tel./fax: +54 351 5353800; e-mail: cmaubecin@gmail.com

Empirical studies examining geographical variation in variance–covariance matrices within species suggest that genetic drift plays a more important role than previously considered (Roff, 2000; Phillips *et al.*, 2001). However, when populations show differences in both the amount of variation (i.e. size) and internal organization (i.e. shape) of matrices, it is not possible to distinguish between drift and selection effects. Additionally, without information on the timescale since populations diverged, the absence of differences should not be used to support the stability expectation (Phillips & McGuigan, 2006; Arnold *et al.*, 2008). Phylogeography can provide both spatial and temporal information to select groups of populations that share the same history of drift for sampling design. Taking into account phylogeographic backgrounds allows a more reliable estimation of average similarities (or differences) on variance–covariance matrices in the field.

Here, we integrated the timescale information provided by previous phylogeographic analyses and replicated sets of populations with similar phylogeographic history to better understand how drift may still condition geographical patterns of adaptive evolution. In particular, we examined the geographical multivariate phenotypic variation in morphofunctional floral traits after a post-glacial population expansion.

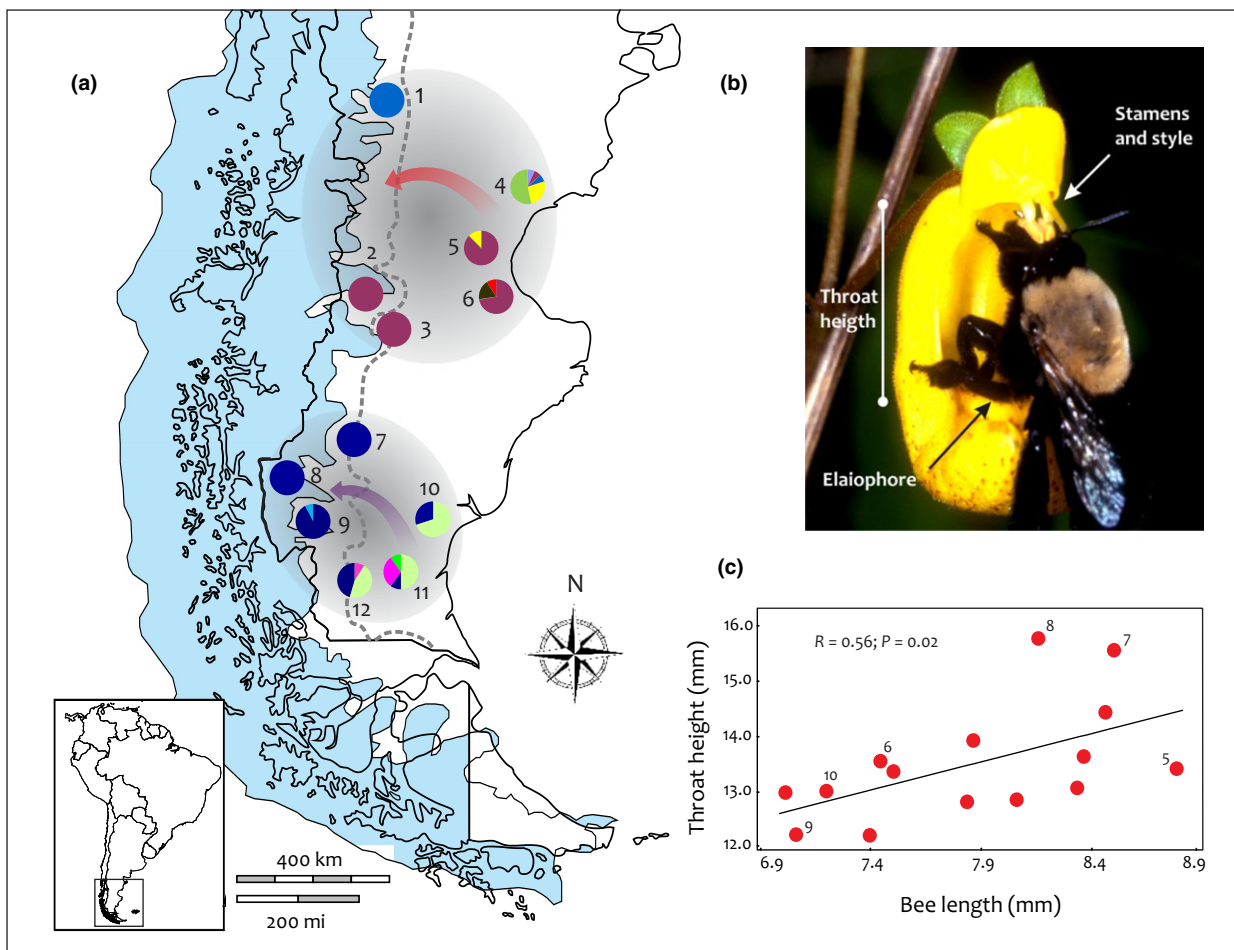
Many studies conducted in the South American Patagonian steppe revealed a large influence of Pleistocene glaciations on the displacement and fragmentation of species distribution ranges (reviewed in Sársic *et al.*, 2011). The perennial herb *Calceolaria polyrhiza* Cav. is a representative species of the impact of climatic fluctuations during glaciations in this region (Sársic *et al.*, 2011; Fig. 1a). In this species, phylogeographic patterns are consistent with a scenario of multiple Pleistocene refugia (i.e. places where populations persisted *in situ* during the glacial events) located in the eastern Patagonian steppe. In addition, genetic analyses revealed a subsequent post-glacial colonization after ice retreat from eastern Patagonia to the Andean flanks, resulting in a range expansion from refugia to previously ice-covered places, accompanied by genetic bottlenecks and drift (Cosacov *et al.*, 2010). The study revealed two phylogroups in northern and southern Patagonia, exhibiting a phylogeographic footprint consistent with a rapid post-glacial expansion and colonization from eastern to western areas of Patagonia. In addition, past demographic analyses indicated that the southern phylogroup has experienced much more intense effects of drift than the northern one (Cosacov *et al.*, 2010).

Across most of its present distribution range, *C. polyrhiza* is exclusively pollinated by an oil-collecting bee, *Centris cineraria* Smith (Tribe Centridini, Fig. 1b), which shows a strong correlation with mechanical fit-related traits of the flower (Cosacov *et al.*, 2014;

Fig. 1c). Pollinator's paleodistribution modelling showed that it likely persisted in the same areas where *C. polyrhiza* did during Pleistocene glaciations (Baranzelli *et al.*, 2013; Cosacov *et al.*, 2014). Hence, former ice-covered areas recolonized by *C. polyrhiza* populations would also have been recolonized by its ancestral pollinator, because the interaction is observed in present western populations. Given a strong effect of drift on recently colonized populations (Cosacov *et al.*, 2010), it is expected that historical contingencies probably disrupted these populations from their ancestral phenotypic optima, whereas plants that persisted *in situ* within glacial refugia would have remained closer to the adaptive peak.

Empirical evidence has shown that pollinators can impose selection pressure upon individual floral traits (Harder & Johnson, 2009), pairs of correlated traits (Campbell, 1992; Delph *et al.*, 2002; Benítez-Vieyra *et al.*, 2006, 2012; Baranzelli *et al.*, 2014; Fenster *et al.*, 2015) and higher interaction levels among traits like phenotypic integration (Ordano *et al.*, 2008). In addition, recent correlative evidence suggests that pollinator-mediated selection can increase the floral integration (González *et al.*, 2015). Therefore, given that populations of *C. polyrhiza* are mainly visited by the same pollinator (Cosacov, 2010; Cosacov *et al.*, 2014) and thus assumed to share the same adaptive peak, if drift overcame selection in recently colonized western populations we expect that glacial refugia show much more similarity among populations in the floral phenotype due to stabilizing selection (Lande, 1979) and higher phenotypic integration due to correlational selection (Roff & Fairbairn, 2012; Baranzelli *et al.*, 2014) than recently colonized western populations. From a multivariate perspective, the random effect of drift is expected to reduce the overall variation (i.e. matrix size) without drastically altering its internal organization (i.e. matrix shape) (Roff, 2000; Stepan *et al.*, 2002). This can be recognized as a proportional reduction in the size of floral matrices of re-colonized western populations relative to those of refugia. In addition, drift is expected to increase phenotypic divergence of the floral matrix among re-colonized populations.

The present study examined the patterns of variation in individual traits and phenotypic matrices of morphofunctional floral traits of *C. polyrhiza* in several refugia and re-colonized populations in northern and southern Patagonia. Specifically, we (1) characterize the size, shape, orientation and the levels of phenotypic integration of variance–covariance matrices of floral traits and (2) compare floral phenotypic matrices using hierarchical Flury's approximation and Random Skewers selection analyses to examine the differences among populations in multivariate properties of the floral phenotype.



**Fig. 1** (a) Geographical distribution of the 12 sampled *Calceolaria polyrhiza* populations in Patagonia (see details in Table S1). The inset depicts a map of South America with the location of the study area. The light blue shaded area and the dashed line indicate the limits of the Last Glacial Maximum and the Greatest Patagonian Glaciation, respectively. Pies indicate the number and relative abundance of haplotypes at the population level, taken from Cosacov *et al.* (2010). Refugium populations are those with high genetic variability and the presence of private haplotypes, located in the Patagonian steppe outside the glacial limits (eastern Patagonia). Recently colonized populations located in previously glaciated areas (western Patagonia) are monomorphic for one of the haplotypes found in putative source populations. Arrows represent the direction of independent colonization events in northern and southern Patagonia, where two different phylogroups are shown in grey shaded areas. (b) Flower of *Calceolaria polyrhiza* and its native pollinator, the oil-collecting bee *Centris cineraria*. Note that stigma and stamens make contact with the dorsal part of the bee's head. Mechanical fit traits (throat height and filament and style lengths) and elaiophore are indicated. (c) Across-populations correlation between throat height (mm) and *Centris cineraria* body length (mm) modified from Cosacov *et al.* (2014); six of these populations are included in the present study (numbered according to a).

## Materials and methods

### Study species

*Calceolaria polyrhiza* (Calceolariaceae) is a predominantly xenogamous perennial herb endemic to Patagonia (Cosacov, 2010). It has hermaphroditic flowers (Fig. 1b) with two stamens and a two-lipped corolla with a distinctive inflated lower lip bearing a median lobe, the appendage or lap (Sérsic, 2004).

This lip is folded inwards and carries the elaiophore, a patch of oil-secreting trichomes. As many of its relative species, it produces nonvolatile oils as floral rewards that attract specialized oil-collecting solitary bees (Molau, 1988; Sérsic, 2004). *Calceolaria polyrhiza* is exclusively pollinated by *Centris cineraria* (Fig. 1b) across its distribution range in the Patagonian steppe, with bee abundance decreasing southwards (Cosacov, 2010).

### Sampling sites

*Calceolaria polyrhiza* is found in Patagonian areas that were ice-covered during the Pleistocene period (1.0 Ma–10 ka) and areas beyond the limits of ice sheets, where previous studies indicated that it persisted *in situ*, probably interacting with its main pollinator, the oil-collecting bee *Centris cineraria* (Cosacov *et al.*, 2010, 2014). A previous phylogeographic study of *C. polyrhiza* showed a pattern consistent with a rapid post-glacial expansion and colonization of the Andean flanks, achieved independently by different lineages from eastern Patagonian steppe populations (Cosacov *et al.*, 2010; Fig. 1a). The study provided several lines of evidence for identifying the potential Pleistocene refugia and post-glacial expansion areas for *C. polyrhiza*. It also revealed two geographically structured phylogroups in the north and south of Patagonia, suggesting independent post-glacial expansion and colonization from eastern Patagonia. Based on the previous phylogeographic study (Cosacov *et al.*, 2010), we selected three ‘refugium’ and three ‘recent’ populations within northern and southern Patagonian groups. Populations were selected by considering them either ‘recent’ or ‘refugium’: recent populations were those that are monomorphic (null genetic variability) or have lower genetic variability than refugium populations, which in turn have higher genetic variability containing different and exclusive haplotypes, including the one present in the recent populations (see Table S1). Based on this criterion, a total of 12 populations were

sampled covering representative areas of the distribution range of *C. polyrhiza* (Fig. 1a, Table S1).

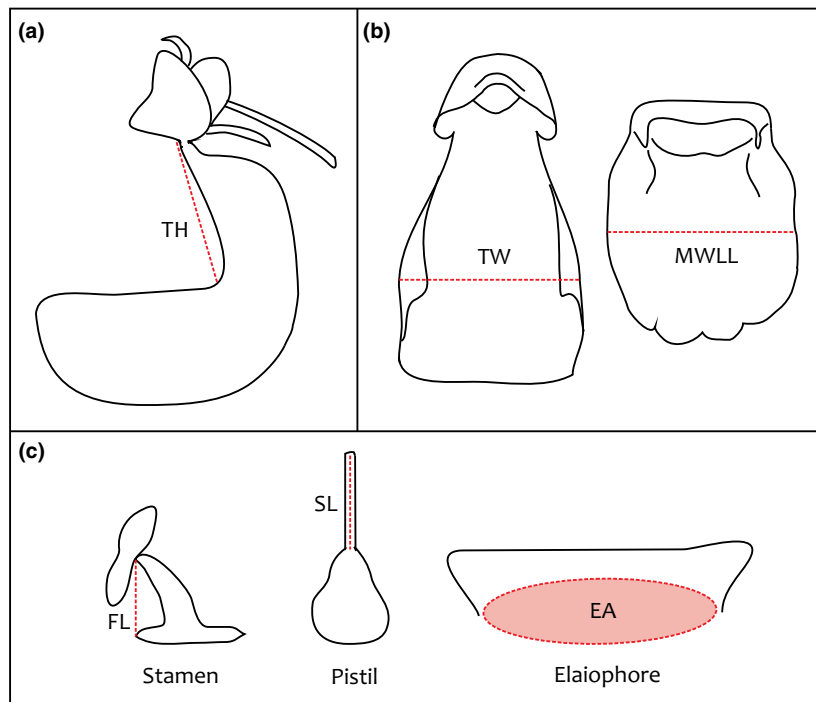
### Floral trait measurements

Phenotypic measurements were performed on digital images of a single flower (floral trait variance among individuals was proven to be larger than within-individual variance, Cosacov *et al.*, 2014) collected from 15 to 32 plants in each of the 12 populations ( $n = 325$  flowers measured), using Sigma Scan© software (Systat Software, Richmond, CA, USA). To obtain these images, the flowers were photographed with a scale using a Nikon Coolpix 5400 camera, in frontal and lateral views, and dissected as indicated in Fig. 2. Six floral characters were measured (Fig. 2): throat width (TW), throat height (TH), maximum width of lower lip (MWLL), filament length (FL), style length (SL) and elaiophore area (EA). Throat height is the ‘operative distance’, that is the distance between floral reward (i.e. the elaiophore) and fertile parts; therefore, this is the critical distance for the pollinator to fit the flower and, therefore, to efficiently transfer pollen from the anthers and to deposit it onto the stigmas (Fig. 1b).

### Statistical analysis

#### *Individual traits and multivariate phenotypic variation*

We explored among-populations differences in phenotypic means using one-way ANOVAS. Then, a nested



**Fig. 2** Schematic diagram of a *Calceolaria polyrhiza* flower showing linear morphometric measurements. (a) Flower lateral view showing throat height (TH), that is the distance between floral reward (elaiophore) and fertile parts. (b) Flower frontal view showing throat width (TW) and maximum width of lower lip (MWLL). (c) Flower dissected to measure filament and style lengths (FL and SL, respectively) and the elaiophore area (EA).



Multivariate analysis of variance (MANOVA) was performed to examine the effects of phylogroup (north vs. south) and population type (recent vs. refugium) nested in phylogroup on the geographical variation in individual traits ( $n = 325$ ). When a significant effect was detected, we performed additional univariate nested ANOVAs to determine which traits accounted for significant effect in the MANOVA. We also obtained the coefficient of variation in each trait for the studied populations and calculated an average value for refugium and recent populations in northern and southern phylogroups, which were then compared with a  $t$ -test. A descriptive statistical summary of measured traits for each population is provided in Table 1.

A principal component analysis (PCA) was performed on population means to examine whether recently colonized populations showed greater dispersion in the multivariate floral phenotype (morphological disparity) than refugium populations. In addition, we used a measure of disparity: the area of the convex hull enclosing data points from the same group, which quantifies the portion of morphological space occupied by populations (Drake & Klingenberg, 2010). In order to test whether disparity was significantly larger in recently colonized than in refugium populations, we compared the observed difference with the distribution of 10 000 differences obtained by randomly reshuffling the labels 'recent' or 'refugium' among populations.

For subsequent analyses, allometric effects were controlled by standardizing the complete sample of floral traits to zero mean and unity standard deviation. All traits were standardized to have the same total variance in the entire sample, but the differences in variance among populations were preserved. Then, each of the six standardized traits was regressed against ovary size (calculated as the average between ovary height and width). Ovary size was taken as a surrogate of flower size, in order to remove the size effect (Berner, 2011; Torices & Méndez, 2014). Ovary size positively correlated with all the other flower traits ( $0.15 > r > 0.6$  in all cases  $P < 0.01$ ). Regression residuals obtained for each population separately were used to construct phenotypic variance-covariance matrices (**P**) (Appendix S1). Although inferences about matrix evolution crucially depend on the intensity of the relationship between the phenotypic and genotypic variation, **P**-matrices have been consistently used as an indirect estimator of the **G**-matrix (Cheverud, 1988; Steppan *et al.*, 2002; Kolbe *et al.*, 2011). However, caution should be taken when using **P**-matrices to infer future responses to selection.

### Comparison of phenotypic matrices

Phenotypic matrices were compared using two statistically complementary analyses (Cheverud & Marroig, 2007): common principal components analysis (CPCA, Phillips & Arnold, 1999) and Random Skewers analyses

**Table 1** Population means and coefficient of variation values (CV) of floral traits from 12 populations of *Calceolaria polyrhiza*. Measured floral traits (abbreviations): throat width (TW), throat height (TH), maximum width of lower lip (MWLL), filament length (FL), style length (SL) and elaiophore area (EA). Sample sizes are provided in parentheses for each population. One-way ANOVAs were performed among populations;  $F_{(d.f., N)}$  and significance  $P$  values are provided. Means followed by different letter within each column are significantly different according to a *post hoc* Tukey's test ( $P < 0.05$ ).

	Populations	Reference number	Traits (mm)					
			TW	TH	MWLL	FL	SL	EA
Northern recent populations	Lago Vintter (31)	1	8.754 <sup>a</sup> (8.169)	7.489 <sup>a</sup> (10.390)	15.487 <sup>a</sup> (12.730)	1.378 <sup>a</sup> (13.381)	1.751 <sup>a</sup> (9.743)	5.972 <sup>a</sup> (20.454)
	Los Antiguos (30)	2	10.652 <sup>cd</sup> (13.749)	10.391 <sup>f</sup> (9.970)	11.225 <sup>bc</sup> (15.410)	1.841 <sup>bcd</sup> (16.481)	2.438 <sup>c</sup> (17.672)	11.082 <sup>cd</sup> (31.136)
	Sumich (29)	3	9.387 <sup>ab</sup> (8.740)	9.247 <sup>cde</sup> (13.234)	11.818 <sup>bcd</sup> (16.190)	2.082 <sup>cde</sup> (13.566)	2.480 <sup>c</sup> (9.106)	8.108 <sup>ab</sup> (21.781)
Northern refugium populations	La Begonia (22)	4	10.890 <sup>cd</sup> (10.500)	10.179 <sup>def</sup> (15.117)	10.936 <sup>abc</sup> (16.060)	1.856 <sup>bc</sup> (21.174)	2.402 <sup>c</sup> (12.372)	8.459 <sup>ab</sup> (26.946)
	Los Manantiales (15)	5	11.331 <sup>de</sup> (13.525)	10.268 <sup>d</sup> (14.309)	12.981 <sup>d</sup> (14.950)	1.629 <sup>ab</sup> (24.904)	2.367 <sup>c</sup> (10.534)	11.017 <sup>cd</sup> (21.396)
	Caleña Olivia (27)	6	8.794 <sup>a</sup> (11.26)	8.902 <sup>cd</sup> (12.035)	12.137 <sup>bcd</sup> (14.790)	1.389 <sup>a</sup> (23.613)	2.029 <sup>b</sup> (17.518)	7.363 <sup>ab</sup> (26.350)
Southern recent populations	Tres Lagos (27)	7	9.965 <sup>bc</sup> (10.989)	8.715 <sup>bc</sup> (14.738)	11.777 <sup>bcd</sup> (13.720)	2.107 <sup>de</sup> (15.562)	2.539 <sup>c</sup> (12.289)	11.380 <sup>d</sup> (25.074)
	Helisingfors (29)	8	12.132 <sup>e</sup> (11.348)	10.407 <sup>ef</sup> (13.725)	15.381 <sup>e</sup> (17.370)	2.162 <sup>e</sup> (12.138)	2.898 <sup>d</sup> (12.554)	15.202 <sup>e</sup> (20.026)
	PN Los Glaciares (32)	9	8.705 <sup>a</sup> (12.817)	8.843 <sup>c</sup> (12.351)	9.302 <sup>a</sup> (17.870)	1.977 <sup>cde</sup> (16.671)	2.488 <sup>c</sup> (11.508)	8.572 <sup>b</sup> (32.533)
Southern refugium populations	Monte León (29)	10	9.655 <sup>abc</sup> (10.621)	8.323 <sup>abc</sup> (10.906)	10.485 <sup>ab</sup> (15.110)	2.157 <sup>e</sup> (16.782)	2.400 <sup>c</sup> (10.983)	8.889 <sup>bc</sup> (28.102)
	La Vanguardia (30)	11	8.922 <sup>ab</sup> (9.916)	7.557 <sup>a</sup> (13.735)	10.533 <sup>ab</sup> (17.410)	2.026 <sup>cde</sup> (6.312)	2.305 <sup>c</sup> (10.922)	7.957 <sup>ab</sup> (25.347)
	Tapi-Alke (24)	12	11.106 <sup>de</sup> (15.305)	7.674 <sup>ab</sup> (12.860)	12.431 <sup>cd</sup> (22.180)	2.502 <sup>d</sup> (11.676)	2.419 <sup>c</sup> (9.408)	9.197 <sup>bcd</sup> (18.407)
$F_{(d.f., N)}$			26.18 <sup>(11,335)</sup> , $< 0.001$	23.87 <sup>(11,335)</sup> , $< 0.001$	27 <sup>(11,334)</sup> , $< 0.001$	33.04 <sup>(11,337)</sup> , $< 0.001$	26.07 <sup>(11,335)</sup> , $< 0.001$	25.75 <sup>(11,337)</sup> , $< 0.001$
$P$								

(Cheverud & Marroig, 2007). CPC software (available at <http://pages.uoregon.edu/pphil/software.html>) was used to perform CPCA. The CPCA compares two or more matrices considering their eigenvectors and eigenvalues, which describe the size, shape and orientation of matrices. This analysis uses the *Flury hierarchy* to test different hypotheses of complex and hierarchical relationships among matrices. The steps include *equality* (identical eigenvectors and eigenvalues, i.e. identical size, shape and orientation), *proportionality* (equal eigenvectors, but eigenvalues differing in a scalar amount, i.e. same shape and orientation, different but proportional size), *all principal components in common* (equal eigenvectors, but different eigenvalues, i.e. same shape and orientation, nonproportional differences in size) or *i principal components in common* with *i* ranging from  $n-2$  to 1, where  $n$  is the number of traits (i.e. progressive differences in shape, orientation and size) and finally *unrelated* structure (matrices have dissimilar eigenvectors and eigenvalues) (Phillips & Arnold, 1999; Steppan *et al.*, 2002). Therefore, with this analysis, we could assess whether *C. polyrhiza* populations share attributes such as the amount of overall phenotypic variation (i.e. matrix size) or internal structure of variation (i.e. matrix shape and orientation) and to explore how population matrices differed to each other (i.e. equality, proportionality).

Assuming no historical effects (i.e. no phylogeographical structure), the matrices of the 12 populations were compared simultaneously. In a second round of analyses, we made the comparisons between the average **P**-matrix of refugium and recently colonized populations within each phylogroup to control for the phylogeographical structure (Cosacov *et al.*, 2010). If genetic drift dominated during the process of population expansion after the last glaciation event, matrices of refugium and recent populations should share similar structure, but should differ proportionally in size. In this case, average **P**-matrices of recently colonized populations will have a smaller size than average **P**-matrices of refugium populations. Then, we compared northern and southern average **P**-matrices of refugium and recent populations. In all cases, we chose the best model according to the Akaike Information Criterion (AIC; Akaike, 1973; Burnham & Anderson, 2002). When the difference in AIC values between two models was below two units, both models were considered equally supported.

We also examined between- and within-phylogroup differences in average refugium and recent population **P**-matrices by comparing their response to selection vectors following the Random Skewers approach (Cheverud & Marroig, 2007), as a complementary analysis to CPCA. This method compares the evolutionary response to selection of two variance–covariance matrices as a measure of matrix similarity. This procedure is based on the multivariate breeder's equation:  $\Delta z = G\beta$

(Lande, 1979; Lande & Arnold, 1983), where  $\Delta z$  is the vector of mean responses, **G** is the matrix of genetic variances and covariances and  $\beta$  is the vector of selection gradients. A large number of random selection vectors ( $\beta$ ), called Random Skewers, is simultaneously applied to both matrices and their response (to selection) vectors ( $\Delta z$ ) are compared. Given any fixed selection vector, if two matrices are similar, the orientation and magnitude of the response vectors should also be similar (Cheverud & Marroig, 2007). In this study, we used phenotypic matrices (**P**) instead of **G**-matrices (Cheverud, 1988; Steppan *et al.*, 2002; Game & Caley, 2006), and 100 000 random selection vectors were used for each contrast. The routines to perform Random Skewers analyses were written in R software (R Core Team 2015) and are available from the authors as Supporting Information (Appendix S2). First, we compared the orientation of the response vectors using the average vector correlation that measures the colinearity of the response vectors in the multivariate morphometric space. The average vector correlation is equal to the cosine of the angle between vectors, with the values ranging from zero (completely different) to one (identical orientation). By comparing the orientation of the response vectors, we mainly evaluated the similarity in shape of the compared matrices. To statistically evaluate whether the response vectors had the same orientation (i.e. whether **P**-matrices have the same orientation, and thus, compared populations share the same internal structure of variation), we used a Monte Carlo procedure based on the distribution of 100 000 correlations between vectors of random  $k$  elements and unit length, where  $k$  is the number of studied traits (Cheverud & Marroig, 2007).

Secondly, we compared the magnitude of the response vectors by calculating the response vectors' length ratio as a measure of the relative levels of variance in the two matrices along this dimension of multivariate space. This ratio assesses the similarity in matrix size (i.e. the amount of variation) and was calculated within phylogroups, by dividing the response vector of the average refugium into the response vector of its corresponding average recent population, and between phylogroups, by dividing the response vector of the northern average refugium or recent population into the response vector of its corresponding southern average refugium or recent population. If the two matrices are equal, the magnitude of the response is expected to be the same for both matrices, and the ratio will be equal to one.

#### Phenotypic integration

The Wagner–Cheverud integration index (INT, Wagner, 1984; Cheverud *et al.*, 1989) was calculated for each population as the variance of the eigenvalues of the phenotypic correlation matrix, expressed as a fraction of its maximal value for a given number of traits (Armbruster *et al.*, 2009) and corrected for sample size

differences (Wagner, 1984). This index ranges from zero, when there is no correlation between traits, to 100% when traits are fully correlated and cannot vary independently. Phenotypic integration levels were estimated for the studied populations and were jackknifed to determine whether the floral matrix of each population of *C. polyrhiza* is significantly integrated. Estimation of integration values and jackknife procedure were performed with PHENIX package (Torices & Muñoz-Pajares, 2015) of R software. Levels of phenotypic integration were also calculated for average **P**-matrices of refugium and recent populations of northern and southern phylogroups.

## Results

### Variation in individual traits

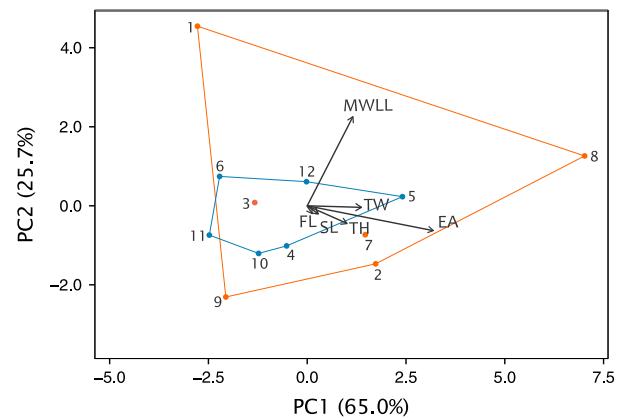
All phenotypic traits significantly differed among populations (Table 1). Nested MANOVA showed that northern and southern phylogroups, and refugium and recent populations within each phylogroup, differed in mean values of phenotypic floral traits (Wilks'  $\lambda = 0.559$  and  $0.641$  respectively,  $P < 0.05$ ), and univariate nested ANOVAS showed that the differences were accounted for by all traits ( $P < 0.05$ , Table S2). Average coefficients of variation in individual traits were slightly higher in refugium populations than in recently colonized northern populations, without showing statistical differences in either the north ( $t = 0.814$ , d.f. = 4,  $P = 0.434$ ) or south ( $t = -0.316$ , d.f. = 4,  $P = 0.758$ ).

### Multivariate phenotypic variation

The first two principal components of the PCA explained 90% of the total variation recorded among populations, revealing an underlying correlation structure among floral traits (Fig. 3). Score distribution in the phenotypic space from the first and second principal components revealed much more phenotypic divergence among recently colonized populations than among refugium populations (Fig. 3). The area enclosed by the points of the recently colonized populations in the space defined by the first two principal axes was significantly larger than the area enclosed by refugium populations (convex hull = 24.353 and 0.364, respectively;  $P = 0.0267$ ).

### Structure of phenotypic matrices

The CPCA performed with the 12 populations simultaneously, without considering their historical context, revealed that they had all principal components in common (i.e. all five eigenvectors were shared among matrices, but with different eigenvalues; AIC values are provided in Table 2; Fig. 4), indicating that the 12 populations shared matrix shape and orientation (covariation



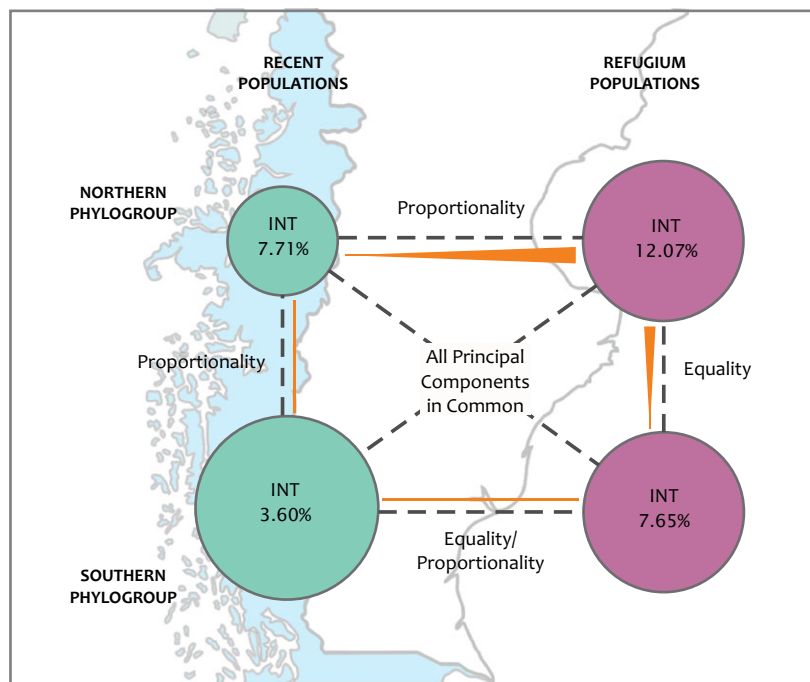
**Fig. 3** Biplot of the first two axes of the PCA ordination for the whole set of *C. polyrhiza* populations. Red dots correspond to recently colonized populations and blue dots, to refugium populations. Populations are numbered according to Fig. 1 and Table S1. The amount of explained variance associated with each axis is provided in parentheses. Convex hull that encloses refugium and recent populations is drawn. Outer convex hull corresponds to recently colonized populations indicating much more phenotypic divergence than refugium populations (inner convex hull). Measured floral traits (abbreviations): throat height (TH), throat width (TW), the maximum width of lower lip (MWLL), filament length (FL), style length (SL) and elaiophore area (EA).

structure) but showed nonproportional differences in size (i.e. populations differed in the amount and distribution of variation in the multivariate space). The first two eigenvalues accounted for 43 to 77% of the total variance. Throat width and the maximum width of lower lip (associated with corolla size) were the traits that most contributed to the first eigenvector, whereas filament length and throat height (associated with flower–pollinator mechanical fit) were the traits that most contributed to the second eigenvector (see Table 3).

When considering different historical contexts, the comparison of average **P**-matrices between refugium and recent populations within northern and southern phylogroups revealed a proportional relationship (i.e. eigenvectors – orientation of axes – were the same, whereas all eigenvalues – variances along each axis – differed by a scalar amount, Table 2, Fig. 4). Thus, refugia and recent populations shared the same internal structure of variation, but differed in the overall amount of variation in a proportional way. As expected, less phenotypic variation (i.e. smaller size) was found in recent populations than in refugia only within the northern phylogroup (proportionality constant = 0.68). In contrast, recent populations for the southern phylogroup had more phenotypic variation than refugium populations (Table 2, proportionality constant = 1.30), even though the equality hypothesis attained a similar support than the proportionality hypothesis (Table 2,  $\Delta AIC = 0.84$ ).

**Table 2** Akaike Information Criterion values from the common principal component analysis, comparing the pooled sample of 12 *C. polyrhiza* populations simultaneously, average **P**-matrix of refugium vs. recent populations (R vs. C) within northern and southern phylogroups and average **P**-matrix of refugium and recent populations between phylogroups (N vs. S). Data in boldface correspond to the best model according to the AIC score. When the difference in AIC values between two models was below two units, both models were considered equally supported.

Model	Pooled population sample	Northern Phylogroup (R vs. C)	Southern Phylogroup (R vs. C)	Refugium populations (N vs. S)	Recent populations (N vs. S)
Equality	507.993	16.431	<b>11.033</b>	<b>15.855</b>	21.292
Proportionality	458.063	<b>13.546</b>	<b>10.189</b>	<b>16.749</b>	<b>14.628</b>
CPC	<b>433.601</b>	20.754	14.787	22.753	17.512
CPC (4)	441.590	22.569	16.592	23.840	19.391
CPC (3)	452.630	24.817	20.241	26.052	23.174
CPC (2)	481.222	28.079	25.291	29.420	26.923
CPC (1)	467.412	34.649	32.737	36.248	33.405
Unrelated	462.000	42.000	42.000	42.000	42.000



**Fig. 4** Summary of the comparisons among average **P**-matrix of refugium (purple circles) and recent (green circles) populations from northern and southern Patagonia. Circle sizes are proportional to the matrix size. Dashed lines indicate common principal component analysis results. Comparisons through Random Skewers method are shown with orange lines, where the thickness of each extreme is proportional to the magnitude of the response to selection of the respective matrix (e.g. similar thickness in both extremes of the line indicates that the connected matrices had similar response to selection). Values inside the circles are the phenotypic integration values.

The comparison of average **P**-matrix of refugium populations between phylogroups showed that these matrices were equal (i.e. identical orientation and size, Fig. 4), although the proportionality attained a similar support ( $\Delta AIC = 0.89$ ). The same procedure with average **P**-matrix of recent populations of both phylogroups revealed that these matrices were proportional, having the southern recent populations on average more phenotypic variation than the northern ones (Table 2, Fig. 4, proportionality constant for average **P**-matrix of southern recent populations = 1.57).

The pairwise matrix comparisons following the Random Skewers procedure indicated overall that the

direction of the response to selection within the multivariate phenotypic space was similar between refugium and recent populations. The average correlation between response vectors, representing the cosine of the angle between the response to selection vectors, ranged from 0.89 to 0.92 (corresponding to an angle of 23° to 27°). These correlations were in all cases larger than the value expected between random pair of vectors within the same multivariate space ( $P < 0.05$ ). However, differences in the magnitude of the response vectors were detected; the ratio between the lengths of the response vectors of average refugium and recent population showed that, within the northern



**Table 3** Trait contribution to the two first eigenvectors of the phenotypic variance–covariance floral matrix obtained from common principal component analysis performed with 12 populations. Values in boldface indicate floral traits with the highest loadings on each eigenvector. Measured floral traits (abbreviations): throat height (TH), throat width (TW), the maximum width of lower lip (MWLL), filament length (FL), style length (SL) and elaiophore area (EA).

TRAITS	Vector 1	Vector 2
TW	<b>0.544</b>	–0.408
MWLL	<b>0.562</b>	–0.193
TH	0.441	<b>0.529</b>
FL	0.126	<b>0.650</b>
SL	0.153	0.297
EA	0.392	–0.078

phylogroup, response to the selection of refugium populations would be 2.74 times more intense than that of recent populations, whereas within the southern phylogroup, refugium and recent populations would have similar responses (ratio = 0.98, Fig. 4). These results suggest that refugium and recent populations shared the covariation structure, but refugium populations within the northern phylogroup showed more overall phenotypic variation than recent populations, whereas within the southern phylogroup the amount of variation would be similar for both types of populations. On the other hand, the comparison of the average matrices of refugium between phylogroups showed that northern refugium populations would have a larger response than southern ones (1.81 times), being consistent with the higher variation observed in the former ones, whereas northern and southern recent populations would have similar responses (ratio = 0.91, Fig. 4), suggesting similar amounts of overall phenotypic variation.

### Phenotypic integration

The integration level in the studied populations differed significantly from zero, ranging from 2.04% to 18.26% (Table S3). Although average levels of floral integration in refugium populations (12.07% and 7.65% for northern and southern refugium populations, respectively, Fig. 4) were apparently higher than in recent populations (7.71% and 3.60% for northern and southern recent populations, respectively, Fig. 4), no statistical difference was detected within phylogroups ( $P > 0.05$ , Table S3).

### Discussion

Although Quaternary environmental changes substantially impacted the landscape and promoted fast evolutionary changes in numerous species (Hewitt, 2000; Tribsch & Schönschetter, 2003; Hosner *et al.*, 2014), the

connection between phylogeography and phenotypic evolution has been little examined in animal systems (e.g. Diniz-Filho *et al.*, 1999; Kuriyama *et al.*, 2011; Dennenmoser *et al.*, 2015) and almost unexplored in plants. To our knowledge, the present study is the first that explores the geographical variation in the variance–covariance matrices under an explicit phylogeographic scenario. The results revealed a significant effect of the phylogeographic background of *C. polyrhiza* populations on the present pattern of multivariate variation in the floral phenotype in Patagonia.

After at least ~15 000 years since the Last Maximum Glaciation, the effects of drift are still present on the phenotypic variance–covariance matrix of *C. polyrhiza* flowers and are likely to constrain future responses to selection on recently colonized populations. Overall differences between refugium and recent populations were more pronounced in the amount of variation than in the covariance structure and integration levels of the flowers. This pattern supports our expectation that drift affected the strength, but not the direction of the potential phenotypic evolution of recently expanded populations.

The northern and southern phylogroups showed differences in their potential evolutionary response to selection. In particular, northern refugium populations were found to have lower interpopulation differences, much more within-population phenotypic variation and a stronger potential response to selection with respect to recent populations than southern ones. This pattern is consistent with the previous phylogeographic evidence indicating that northern refugium populations have more genetic variation and larger effective population size than southern populations, which were historically much more impacted by drift (Cosacov *et al.*, 2010). Despite the differences in matrix size between northern and southern recent populations, both expressed a similar potential response to selection, a pattern that also appeared when comparing recent and refugium southern populations. This pattern suggests that in the south a response to selection was more constrained than in the north region.

Given that the same specialized pollinator is present throughout most of the distribution range of the species, it is expected that populations within refugia are likely to be closer to the nearest adaptive peak than those of recently colonized sites, where drift would have overcome selection. In turn, this hypothesis suggests that the structure of the floral matrix would have remained fairly stable since the post-glaciation event. Accordingly, the comparison of average variance–covariance matrices of refugium populations between phylogroups showed that these matrices were equal (i.e. identical orientation, shape and size), supporting the expectation that northern and southern refugia shared the same adaptive peak. Together, the strong similarity among refugium populations and proportional changes among recently colonized

populations support the expectation that drift affected variance–covariance matrices during the past demographic expansion towards western areas in Patagonia.

In the past, comparative studies attempted to test the stability of the variance–covariance matrix by comparing populations or species without available knowledge of the divergence timescale or information about the genetic history of populations or lineages (reviewed in Roff, 2000; Phillips *et al.*, 2001; Arnold *et al.*, 2008). For instance, if two populations recently diverged from a common ancestor, they are likely to be statistically equal, erroneously supporting the stability expectation. Thus, the stability should be tested among populations that were exposed for a period long enough to detect a response to selection in the phenotype. Over a period long enough since divergence, the absence of differences among matrices will support the hypothesis of stability. Phylogeography can provide not only that kind of timescale information but also information about the underlying historical process of the studied populations. The results from this study using two independent groups of populations exposed to a recent demographic expansion revealed a significant reduction in the rate, but no differences in the direction of evolution. Thus, our results reject the size stability expectation due to a significant reduction in variation during post-glacial population expansion; however, the results do support the stability expectation related to matrix shape and orientation. Hence, we advocate that examination of the stability hypothesis will benefit from the connection between phylogeography and phenotypic evolution improving field comparisons of variance–covariance matrices (Lande, 1979; Jones *et al.*, 2003).

When the hypothesis of stability is rejected, determining the major causes of differences still remains a central challenge. Despite previous attempts to understand population differences in variance–covariance matrices, only few cases have assessed the effects of natural selection and genetic drift (Roff & Mousseau, 2005; Berner *et al.*, 2008, 2010; Eroukhmanoff & Svensson, 2011). Proportional changes among matrices are usually ascribed to drift (Roff, 2000), but this expectation corresponds theoretically to the average expected change of a set of populations exposed to the same history of drift and selection (Lynch, 1988; Phillips *et al.*, 2001). To date, a few laboratory experiments have used a significant number of experimental lines to test the effects of drift, providing reliable estimates of the average effect of variation in population size on variance–covariance matrices (e.g. Phillips *et al.*, 2001). To what extent this effect is expressed under natural conditions is still unknown (Phillips & McGuigan, 2006). On the other hand, most field comparisons of variance–covariance matrices lack the critical historical information to contrast groups of matrices with similar historical backgrounds. In this study, analyses performed without

considering the phylogeographic history of populations overestimated matrix differences. In contrast, the same analyses comparing the average refugium and recent population matrices were not able to reject the hypothesis of proportionality. Our study illustrates a potential risk when comparing matrices under natural conditions and suggests that explicitly considering the phylogeographic history of selected populations provides desirable complementary information to examine the effects of drift on future responses to selection. At this point, it is worth reminding that inferences using **P**-matrix strongly depend on how good **P** estimate **G**; thus, evolutionary inferences should be considered with caution.

Moreover, considering phylogeographic information in phenotypic evolutionary studies provides the opportunity to hypothesize about the predominance of historical selection vs. drift in different populations. Differences in shape or orientation of matrices within groups of populations that share the same phylogeographic history will support the hypothesis that selection exerted a stronger effect than drift on the variance–covariance matrices. In contrast, proportional changes in matrix size among populations will support the hypothesis that drift was the main evolutionary process. Moreover, comparison of populations that underwent drift of contrasting intensity can help to build a hypothetical map of geographical areas where particular selection patterns would be expected. For example, in *C. polyrhiza*, a previous study suggested that throat height and filament and style lengths are the critical flower-bee fitting traits involved in the pollination mechanism, showing a strong plant-pollinator phenotypic matching likely mediated by pollinator selection (Cosacov *et al.*, 2014). Because populations within refugia were less affected by drift than those of recently colonized areas, the former are more likely closer to the nearest adaptive peak. Thus, one could expect that stabilizing selection on flower-bee fitting traits prevails in eastern areas of Patagonia and that directional selection prevails in western areas. However, major phenotypic changes would be observed in those traits with the highest loading on the first principal component of the floral matrix (i.e. throat width and the maximum width of lower lip, involved in floral signalling).

The number of studies that show the importance of historical processes vs. ecological factors has increased, in an attempt to explain the variation in some population attributes across the distribution range of species (e.g. Eckert *et al.*, 2008 and references therein). When historical process still affects the present geographical phenotypic patterns, mismatches can arise between present ecological conditions and phenotypic variation. Climatic changes occurring during the Pleistocene are particularly well documented and have been increasingly analysed in combination with genetic information to understand how past environmental changes

contribute to lineages distribution and diversification (e.g. Cosacov *et al.*, 2010, 2013). In the present study, we illustrate how the combination of phylogeographic information with across-population phenotypic variation can help to disentangle the relative magnitude of drift on the potential evolutionary response of populations. Thus, phylogeography represents an essential baseline to understand where and when drift disrupted the phenotype from its expected optimum.

## Acknowledgments

We thank Jorgelina Brasca for editing the English version of the manuscript. We thank Administración de Parques Nacionales (Argentina) for permitting sampling collection in protected areas. We also thank Drs John Pannel, Marcos Méndez, Rubén Torices and an anonymous reviewer for valuable suggestions on an earlier version of this manuscript. This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Ministerio de Ciencia, Técnica e Innovación Productiva and Agencia Nacional de Promoción Científica y Tecnológica (grants PIP 11220080101264, PID 201101-00245, FONCYT PICT-2011-0837, 2011-0709 and 2012-2603); and Myndel Pedersen Foundation. JF thanks DGAPA-UNAM for the support during a sabbatical stay at IMBIV (CONICET and Universidad Nacional de Córdoba, Argentina).

## References

- Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium of Information Theory* (B.N. Petrov & F. Csaki, eds), pp. 267–281. Akademiai Kiado, Budapest, Hungary.
- Armbruster, W.S., Hansen, T.F., Pélabon, C., Pérez-Barrales, R. & Maad, J. 2009. The adaptive accuracy of flowers: measurement and microevolutionary patterns. *Ann. Bot.* **103**: 1529–1545.
- Arnold, S.J., Bürger, R., Hohenlohe, P.A., Ajie, B.C. & Jones, A.G. 2008. Understanding the evolution and stability of the G-matrix. *Evolution* **62**: 2451–2461.
- Baranzelli, M.C., Cosacov, A., Cocucci, A.A. & Sérsic, A.N. 2013. Dinámica temporal de la interacción de *Calceolaria polyrhiza* y sus polinizadores: una aproximación utilizando modelos predictivos de nicho. VII Reunión Anual de la Sociedad Chilena de Evolución, Chile.
- Baranzelli, M.C., Sérsic, A.N. & Cocucci, A.A. 2014. The search for Pleiades in trait constellations: functional integration and phenotypic selection in the complex flowers of *Morrenia brachystephana* (Apocynaceae). *J. Evol. Biol.* **27**: 724–736.
- Benítez-Vieyra, S., Medina, A.M., Glinos, E. & Cocucci, A.A. 2006. Pollinator-mediated selection on floral traits and size of floral display in *Cyclopogon elatus*, a sweat bee-pollinated orchid. *Funct. Ecol.* **20**: 948–957.
- Benítez-Vieyra, S., Glinos, E., Medina, A.M. & Cocucci, A.A. 2012. Temporal variation in the selection on floral traits in *Cyclopogon elatus* (Orchidaceae). *Evol. Ecol.* **26**: 1451–1468.
- Berner, D. 2011. Size correction in biology: how reliable are approaches based on (common) principal component analysis? *Oecologia* **166**: 961–971.
- Berner, D., Adams, D.C., Grandchamp, A.C. & Hendry, A.P. 2008. Natural selection drives patterns of lake–stream divergence in stickleback foraging morphology. *J. Evol. Biol.* **21**: 1653–1665.
- Berner, D., Stutz, W.E. & Bolnick, D.I. 2010. Foraging trait (co) variances in stickleback evolve deterministically and do not predict trajectories of adaptive diversification. *Evolution* **64**: 2265–2277.
- Burnham, K.P. & Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York, NY.
- Campbell, D.R. 1992. Variation in sex allocation and floral morphology in *Ipomopsis aggregata* (Polemoniaceae). *Am. J. Bot.* **79**: 516–521.
- Cheverud, J.M. 1988. A comparison of genetic and phenotypic correlations. *Evolution* **42**: 958–968.
- Cheverud, J.M. & Marroig, G. 2007. Comparing covariance matrices: random skewers method compared to the common principal components model. *Genet. Molec. Biol.* **30**: 461–469.
- Cheverud, J.M., Wagner, G.P. & Malcolm, M.D. 1989. Methods for the comparative analysis of variation patterns. *Syst. Zool.* **38**: 201–213.
- Cosacov, A. 2010. Patrones de variación de caracteres fenotípicos y frecuencias génicas en el rango de distribución de la especie endémica de Patagonia *Calceolaria polyrhiza*: su relación con variables ambientales, polinizadores y factores históricos. PhD Thesis, Universidad Nacional de Córdoba, Argentina.
- Cosacov, A., Sérsic, A.N., Sosa, V., Johnson, L.A. & Cocucci, A.A. 2010. Multiple periglacial refugia in the Patagonian steppe and post-glacial colonization of the Andes: the phylogeography of *Calceolaria polyrhiza*. *J. Biogeogr.* **37**: 1463–1477.
- Cosacov, A., Johnson, L.A., Paiaro, V., Córdoba, F.E. & Sérsic, A.N. 2013. Precipitation rather than temperature influenced the phylogeography of the endemic shrub *Anarthrophyllum desideratum* in the Patagonian steppe. *J. Biogeogr.* **40**: 168–182.
- Cosacov, A., Cocucci, A.A. & Sérsic, A.N. 2014. Geographical differentiation in floral traits across the distribution range of the Patagonian oil-secreting *Calceolaria polyrhiza*: do pollinators matter? *Ann. Bot.* **113**: 251–266.
- Delph, L.F., Knapczyk, F.N. & Taylor, D.R. 2002. Among-population variation and correlations in sexually dimorphic traits of *Silene latifolia*. *J. Evol. Biol.* **15**: 1011–1020.
- Dennenmoser, S., Nolte, A.W., Vamosi, S.M. & Rogers, S.M. 2015. Phylogeography of the prickly sculpin (*Cottus asper*) in north-western North America reveals parallel phenotypic evolution across multiple coastal–inland colonizations. *J. Biogeogr.* **42**: 1626–1638.
- Diniz-Filho, J.A.F., Fuchs, S. & Arias, M.C. 1999. Phylogeographical autocorrelation of phenotypic evolution in honey bees (*Apis mellifera* L.). *Heredity* **83**: 671–680.
- Drake, A.G. & Klingenberg, C.P. 2010. Large-scale diversification of skull shape in domestic dogs: disparity and modularity. *Am. Nat.* **175**: 289–301.
- Eckert, C.G., Samis, K.E. & Loughheed, S.C. 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molec. Ecol.* **17**: 1170–1188.

- Eroukhmanoff, F. & Svensson, E.I. 2011. Evolution and stability of the G-matrix during the colonization of a novel environment. *J. Evol. Biol.* **24**: 1363–1373.
- Fenster, C.B., Reynolds, R.J., Williams, C.W., Makowsky, R. & Dudash, M.R. 2015. Quantifying hummingbird preference for floral trait combinations: the role of selection on trait interactions in the evolution of pollination syndromes. *Evolution* **69**: 1113–1127.
- Game, E.T. & Caley, M.J. 2006. The stability of P in coral reef fishes. *Evolution* **60**: 814–823.
- González, A.V., Murúa, M.M. & Pérez, F. 2015. Floral integration and pollinator diversity in the generalized plant-pollinator system of *Alstroemeria ligtu* (Alstroemeriaceae). *Evol. Ecol.* **29**: 63–75.
- Harder, L.D. & Johnson, S.D. 2009. Darwin's beautiful contrivances: evolutionary and functional evidence for floral adaptation. *New Phytol.* **183**: 530–545.
- Hewitt, G.M. 2000. The genetic legacy of the quaternary ice ages. *Nature* **405**: 907–913.
- Hosner, P.A., Sánchez-González, L.A., Peterson, A.T. & Moyle, R.G. 2014. Climate-driven diversification and Pleistocene refugia in Philippine birds: evidence from phylogeographic structure and paleo-environmental niche modeling. *Evolution* **68**: 2658–2674.
- Jones, A.G., Arnold, S.J. & Bürger, R. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**: 1747–1760.
- Kolbe, J.J., Revell, L.J., Székely, B., Brodie, E.D. III & Losos, J.B. 2011. Convergent evolution of phenotypic integration and its alignment with morphological diversification in Caribbean *Anolis* ecomorphs. *Evolution* **65**: 3608–3624.
- Kuriyama, T., Brandley, M.C., Katayama, A., Mori, A., Honda, M. & Hasegawa, M. 2011. A time-calibrated phylogenetic approach to assessing the phylogeography, colonization history and phenotypic evolution of snakes in the Japanese Izu Islands. *J. Biogeogr.* **38**: 259–271.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**: 402–416.
- Lande, R. & Arnold, S.J. 1983. The measurement of selection on qualitative characters. *Evolution* **37**: 1210–1226.
- Lynch, M. 1988. Design and analysis of experiments on random drift and inbreeding depression. *Genetics* **120**: 791–807.
- Molau, U. 1988. Scrophulariaceae—Part I. Calceolariae. *Fl. Neotrop. Monogr.* **47**: 1–325.
- Ordano, M., Fornoni, J., Boege, K. & Domínguez, C.A. 2008. The adaptive value of phenotypic floral integration. *New Phytol.* **179**: 1183–1192.
- Phillips, P.C. & Arnold, S.J. 1999. Hierarchical comparison of genetic variance covariance matrices. I. Using the Flury Hierarchy. *Evolution* **53**: 1506–1515.
- Phillips, P.C. & McGuigan, K.L. 2006. Evolution of genetic variance-covariance structure. In: *Evolutionary Genetics, Concepts and Case Studies* (C.W. Fox & J.B. Wolf, eds), pp. 310–325. Oxford University Press, Oxford.
- Phillips, P.C., Whitlock, M.C. & Fowler, K. 2001. Inbreeding changes the shape of the genetic covariance matrix in *Drosophila melanogaster*. *Genetics* **158**: 1137–1145.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/ISBN 3-900051-07-0>.
- Roff, D. 2000. The evolution of the G matrix: selection or drift? *Heredity* **84**: 135–142.
- Roff, D.A. & Fairbairn, D.J. 2012. A test of the hypothesis that correlational selection generates genetic correlations. *Evolution* **66**: 2953–2960.
- Roff, D. & Mousseau, T.A. 2005. The evolution of the phenotypic covariance matrix: evidence for selection and drift in *Melanoplus*. *J. Evol. Biol.* **18**: 1104–1114.
- Sérsic, A.N. 2004. Pollination biology in the genus *Calceolaria*. *Stapfia* **82**: 1–121.
- Sérsic, A.N., Cosacov, A., Cocucci, A.A., Johnson, L.A., Pozner, R., Avila, L.J. et al. 2011. Emerging phylogeographical patterns of plants and terrestrial vertebrates from Patagonia. *Biol. J. Linne. Soc.* **103**: 475–494.
- Steppan, S.J., Phillips, P.C. & Houle, D. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol. Evol.* **17**: 320–327.
- Torices, R. & Méndez, M. 2014. Resource allocation to inflorescence components is highly integrated despite differences between allocation currencies and sites. *Int. J. Pl. Sci.* **175**: 713–723.
- Torices, R. & Muñoz-Pajares, A.J. 2015. PHENIX: An R package to estimate a size-controlled phenotypic integration index. *Appl. Plant Sci.* **3**: 1400404.
- Tribisch, A. & Schönswetter, P. 2003. Patterns of endemism and comparative phylogeography confirm palaeoenvironmental evidence for Pleistocene refugia in the Eastern Alps. *Taxon* **52**: 477–497.
- Wagner, G.P. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a non random organization of quantitative character variation. *J. Math. Biol.* **21**: 77–95.

## Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article: **Table S1** Geographical and genetic information 1 of sampled populations.

**Table S2** Nested ANOVA results.

**Table S3** Detailed information of P-matrix properties.

**Appendix S1** Phenotypic variance–covariance matrices (P) of *Calceolaria polyrhiza* populations.

**Appendix S2** Random Skewers routines for R.

Received 18 September 2015; revised 22 April 2016; accepted 25 April 2016