Genetic variation of loci potentially under selection confounds species-genetic diversity correlations in a fragmented habitat

Angeline Bertin, Nicolas Gouin, Alex Baumel, Ernesto Gianoli, Juan Serratosa, Rodomiro Osorio, Stephanie Manel

To cite this version:


HAL Id: hal-01605033
https://hal.archives-ouvertes.fr/hal-01605033
Submitted on 25 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Genetic variation of loci potentially under selection confounds species-genetic diversity correlations in a fragmented habitat

ANGELINE BERTIN\textsuperscript{1}, NICOLAS GOUIN\textsuperscript{1,2,3}, ALEX BAUMEL\textsuperscript{4}, ERNESTO GIANOLI\textsuperscript{1,5}, JUAN SERRATOSA\textsuperscript{6}, RODOMIRO OSORIO\textsuperscript{1} and STEPHANIE MANEL\textsuperscript{7}

\textsuperscript{1}Departamento de Biología, Universidad de La Serena, La Serena, Chile  
\textsuperscript{2}Centro de Estudios Avanzados en Zonas Áridas, La Serena, Chile  
\textsuperscript{3}Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, La Serena, Chile  
\textsuperscript{4}Institut Méditerranéen de Biodiversité et d’Ecologie marine et continentale (IMBE)  
Aix Marseille Univ, Univ Avignon, CNRS, IRD, Marseille, France.  
\textsuperscript{5}Departamento de Botánica, Universidad de Concepción, Concepción, Chile  
\textsuperscript{6}Facultad de Ciencias del Mar, Doctorado en Biología y Ecología Aplicada, Universidad Católica del Norte  
\textsuperscript{7}EPHE, PSL Research University, CNRS, UM, SupAgro, IRD, INRA, UMR 5175 CEFE, F-34293 Montpellier, FRANCE

Keywords: community genetics, species-genetic diversity correlations, outlier loci, habitat connectivity, high Andean wetlands

Corresponding Author:  
Angéline Bertin  
Departamento de Biología, Facultad de Ciencias, Universidad de La Serena, Av. Raúl Bitrán 1305, Casilla 554, La Serena, Chile.  
Fax: (56) (51) 2 204383, electronic mail address: abertin@userena.cl;

Running Title: Influence of outlier loci on SGDCs
Abstract

Positive species-genetic diversity correlations (SGDCs) are often thought to result from the parallel influence of neutral processes on genetic and species diversity. Yet, confounding effects of non-neutral mechanisms have not been explored. Here, we investigate the impact of non-neutral genetic diversity on SGDCs in high Andean wetlands. We compare correlations between plant species diversity (SD) and genetic diversity (GD) calculated with and without loci potentially under selection (outlier loci). The study system includes 2188 specimens from five species (three common aquatic macroinvertebrate and two dominant plant species) that were genotyped for 396 amplified fragment length polymorphism loci. We also appraise the importance of neutral processes on SGDCs by investigating the influence of habitat fragmentation features. Significant positive SGDCs were detected for all five species (mean SGDC = 0.52 ± 0.05). While only a few outlier loci were detected in each species, they resulted in significant decreases in GD and in SGDCs. This supports the hypothesis that neutral processes drive species-genetic diversity relationships in high Andean wetlands. Unexpectedly, the effects on genetic diversity of the habitat fragmentation characteristics in this study increased with the presence of outlier loci in two species. Overall, our results reveal pitfalls in using habitat features to infer processes driving SGDCs and show that a few loci potentially under selection are enough to cause a significant downward bias in SGDC. Investigating confounding effects of outlier loci thus represents a useful approach to evidence the contribution of neutral processes on species-genetic diversity relationships.
Introduction

Investigating spatial or temporal patterns of covariation between species diversity within a community (SD) and genetic diversity within a species (GD) is one of the main objectives of community genetics (Antonovics 1992). It is a key step towards a conceptual synthesis at the interface between ecology and evolutionary biology (Vellend & Geber 2005). It is also a core issue in conservation biology as it provides valuable information to understand and predict different facets of biodiversity (Hu et al. 2009; Lavergne et al. 2010; Mouquet et al. 2012).

Studies addressing species genetic diversity correlations (SGDCs) in the wild have shown that the relationship between GD and SD can take different forms depending on the mechanisms at work (Kahilainen et al. 2014), with reports of positive (He et al. 2008; Papadopoulou et al. 2011; Lamy et al. 2013), negative (Silvertown et al. 2009; Xu et al. 2016) and no correlation (Taberlet et al. 2012) between the two components. Concordant spatial patterns between species and genetic diversity can emerge from various processes. They may result from direct causal relationships between the two biodiversity components (Vellend and Geber 2005). For instance, GD can influence SD when genetic variation of habitat-forming species, such as forest trees, structures dependent communities (i.e. micro-organisms, fungi, arthropods and vertebrates, Whitham et al. 2006). Alternatively, SD in the community can impose variable selective pressures on populations, and therefore alter the levels of population GD (Strauss et al. 2005; Vellend and Geber 2005; Johnson and Stinchcombe 2007; Lankau and Strauss 2007). SD and GD can become indirectly associated as well, by responding similarly to common mechanisms (Vellend and Geber 2005). Indeed, both ecological and genetic models posit similar effects of drift and
migration on species and genetic diversity (e.g. depletion effects of drift and diversity
increases from immigration, Hu et al. 2009). Thus, concordant spatial patterns between
species and genetic diversity can be common under conditions of demographic stochasticity
and dispersal limitation (Rosenzweig 1995; Frankham 1996; Frankham 1997). To date,
little research has been conducted into the drivers of negative SGDCs (Xu et al. 2016).
According to a modeling study, negative SGDCs can occur under particular conditions of
high genetic mutation rates, depending on the relative importance of mutation, immigration
and local competition (Laroche et al. 2015). They may also occur in situations where
increases in SD, promoted by environmental heterogeneity, are associated with a reduction
in niche breadth and/or population size (Taberlet et al. 2012; Vellend et al. 2014; Xu et al.
2016).

While both neutral and selective mechanisms can generate positive covariation
between genetic and species diversity (Vellend and Geber 2005), evidence suggests that
neutral processes play a dominant role of in this respect (Odat et al. 2004; Vellend 2004;
Positive SGDCs are common in discrete habitat units like oceanic islands, lakes, or forest
fragments, and less frequent in arbitrarily defined spatial units such as grid cells or
vegetation plots (Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014). In
fragmented landscapes, migration and drift are likely major contributors to within-patch
diversity. Thus, positive SGDCs among islands and habitat patches have been postulated to
arise from the parallel influence of neutral processes on both species and genetic diversity
(Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014). This is supported by a
number of empirical studies, which have demonstrated that local habitat characteristics
influencing rates of stochastic immigration and extinction contribute markedly to species-
genetic diversity relationships in fragmented habitats (Vellend 2005; He et al. 2008; Papadopoulou et al. 2011; Lamy et al. 2013). These habitat features include habitat size, connectivity and perturbation. Indeed, both habitat carrying capacity and heterogeneity are expected to increase with habitat size. Larger patches can thus support larger populations, which are both more genetically diverse and more resilient to genetic drift (Frankham 1997). In addition, by offering more diverse ecological niches, larger habitats are also expected to hold more diverse communities, and are consequently less susceptible to random species loss through ecological drift (MacArthur and Wilson 1967, Hubbell 2001).

Because habitat connectivity influences rates of migration, it is expected to result in increases in both genetic and species diversity, by facilitating the introduction of new alleles in a population or new species in the community (Hu et al. 2009). Regarding habitat perturbation, it can cause parallel effects on species and genetic diversity by provoking extinction of species and alleles irrespectively of their identities (Vellend 2003).

Neutral explanations for positive SGDCs have also come from the wide use of presumed neutral molecular markers to estimate genetic diversity in SGDC investigations (Vellend and Geber 2005; Vellend et al. 2014). Genetic variation of molecular markers with supposedly no effects on fitness, such as microsatellite or amplified fragment length polymorphism (AFLP) loci, is considered to mirror neutral processes, including mutation, genetic drift and gene flow (Holderegger et al. 2006; Kirk and Freeland 2011). SGDCs estimated with these markers have therefore been assumed to emerge from concordant changes in species and genetic diversity induced by neutral processes (Vellend and Geber 2005; Lamy et al. 2013; Vellend et al. 2014). However, no attention has been paid to confounding effects of non-neutral mechanisms, which may lead to erroneous conclusions regarding the mechanisms causing SD-GD covariation (Fig. 1). Indeed, selection generates
changes in allele frequencies at target genes, but also in adjacent, linked regions (genetic hitchhiking; Barton 2000). Such linked DNA regions, even non-coding DNA, can thus bear signatures of selection (confounding effect 1 in Fig. 1; Barton 2000; Oleksyk et al. 2010).

When used to infer neutral processes, such loci may bias population parameter estimates (Luikart et al. 2003). Their expected effects on genetic diversity depend on the type of selection operating, as each selection mode leads to specific patterns of allele frequency changes (Oleksyk et al. 2010). For instance, divergent directional selection decreases genetic variation within populations, while balancing selection maintains variation within populations. A few loci potentially under divergent selection are enough to cause a significant downward bias in the assessment of GD estimated from markers assumed to be neutral (García-Verdugo et al. 2015). The presence of such markers can thus be misleading for SGDC studies. When SGDCs are driven by neutral mechanisms, loci under selection are likely to decrease the correlation between species and genetic diversity, since only neutral GD is expected to covary with SD in such circumstances. In contrast, no such effects are expected when SGDCs arise from non-neutral processes.

Separating the neutral and non-neutral components of GD not only opens new perspectives for the study of SGDCs, but can also be critical to interpreting apparent effects of local habitat features in a neutral framework. Several studies have investigated how disturbance regime, habitat size and/or connectivity relate to SGDCs in order to infer underlying mechanisms (Vellend and Geber 2005; Papadopoulou et al. 2011; Lamy et al. 2013; Vellend et al. 2014). These studies postulate that these local features become associated with neutral components of SD/GD due to their modulating role on rates of migration and stochastic loss of species/alleles. However, these relationships can be mistakenly attributed to neutral processes (confounding effect 2 in Fig. 1) if disturbance
regime, habitat size and/or connectivity and selective environmental factors are spatially
correlated, as they would lead to indirect associations between non-neutral components of
GD and the local habitat features. These confounding effects have been documented for
species richness-island area associations (Ricklefs and Lovette 1999), and are expected at
the genetic level since populations at the margins of a species distribution range are not
only more fragmented and isolated than central populations, but also subject to different
ecological constraints (central-marginal hypothesis, Eckert et al. 2008). In such situations,
factoring out the effects of disturbance regime, habitat size and/or connectivity bears the
risk of overestimating the contribution of neutral processes on SD and GD, as it is likely to
control for some influence of selective factors as well.

In this study, we aimed to disentangle the confounding effects of non-neutral
genetic diversity on species-genetic diversity relationships in a highly fragmented
ecosystem, the high Andean wetlands of Chile’s Norte Chico. These wetlands, formed by
groundwater upwelling, resemble oases in an arid matrix (Fig. 2, Squeo et al. 2006). They
occupy a latitudinal gradient of both aridity (increasing northwards) and wetland density
(Fig. 2). This suggests that wetlands differing in connectivity also experience different
selective pressures. Given the high degree of fragmentation of these ecosystems, we
hypothesized that neutral processes should play a critical role in species-genetic diversity
relationships. Thus, we expected loci influenced by selection to decrease SGDCs. We also
postulated that such loci can complicate the evaluation of the importance of neutral
processes on SGDC by factoring out the effects of habitat connectivity on SD and GD.
Indeed, due to the spatial covariation between habitat fragmentation and the latitudinal
environment gradient, GD of loci under environmental gradient pressure is also expected to
correlate with fragmentation characteristics. To test these conjectures, we investigated
SGDCs between species diversity of high Andean wetland vegetation and GDs of five species (two plants and three macroinvertebrates) with contrasting life-histories. We used a genome scan approach to screen for AFLP markers potentially under selection (i.e. outlier loci; Storz et al. 2004) and analyzed SGDCs and the effects of habitat characteristics (size, connectivity and stability) on SGDCs with and without these loci. Our results showed that, as expected, confounding effects of outlier loci on GD decrease SGDCs. This trend was particularly evident with loci potentially under divergent selection. We also found that, in some species, GD at outlier loci correlate with habitat features supposedly linked to neutral processes, revealing pitfalls in using such characteristics to infer processes driving SGDCs.

**Material and methods**

**Study system**

We analyzed species and genetic diversity for 21 high Andean wetlands located along a ca. 600 km latitudinal range (26°S–32°S) in north-central Chile (Fig. 2, and see Bertin et al. 2015). The region, known as Norte Chico, is a biodiversity hotspot, and is characterized by remarkably high levels of endemism (Arroyo et al. 2004). It includes five hydrologically unconnected river basins. The climate of the region varies from hyperarid in the north to mediterranean in the south, with mean annual precipitation ranging between 35 mm and 200 mm for the northernmost and southernmost valleys, respectively.

**Data collection**

*Community and population sampling*
Sampling of plant and benthic macroinvertebrate specimens was carried out between March and April 2011. For plant diversity assessment, the length of each wetland was divided into five sectors and a 30 x 30 cm quadrat was randomly placed within each sector. Plant species were separated and identified in the laboratory.

For genetic diversity evaluation, we selected five common species: two dominant plants, *Carex gayana* (Cyperaceae) and *Patosia clandestina* (Juncaceae), and three abundant aquatic macroinvertebrates, *Andesiops peruvianus* (Insecta, Ephemeroptera), *Austrelmis sp.* (Insecta, Coleoptera) and *Hyallela fossamancini* (Crustacea, Amphipoda).

Samples were collected by hand or using a small fishing net in the case of aquatic organisms. The entire wetland surface area was sampled in each case, with individual wetlands ranging in size from 3.7 to 38 ha. Leaf samples were conserved individually in silica gel until DNA extraction, whereas macroinvertebrates were stored in 95% alcohol. The plant *C. gayana* was the only species found to occur within all 21 wetlands sampled. Details regarding sampling sites and size for each species are given in Table S1.

**Genetic data**

For all five species, DNA extraction was performed using a CTAB protocol. AFLP amplifications and genotyping were carried out using standard procedures and following the protocol of Meudt and Clarke (2007), available at http://clarkeresearch.org/aflp_2012-01-26/aflp.html, with very few modifications. In each species, four combinations of specific primers were selected based on their reliability and number of fragments amplified. AFLP fragments were separated by capillary electrophoresis and genotyping was performed with the software GeneMarker v2.4 (Softgenetics). Following Bonin *et al.* (2007), we retained only AFLP loci with genotyping error rate < 10%, with the exception
of the crustacean *H. fossamancinii*, for which we included loci with up to 12% genotyping error in order to maintain a comparable number of loci for all species. Loci with redundant information were filtered using the software AFLPOP (Duchesne and Bernatchez 2002). Detailed procedures and AFLP primer pairs are described in Supplementary Information. Only polymorphic markers (those with band frequencies between 5% and 95%) were kept in the analysis (Bonin *et al*. 2007).

**Habitat parameters**

Habitat attributes associated with connectivity, size and stability were assessed for each of the 21 sampled wetlands. Attributes were derived from maps of wetland distribution generated based on spectral analysis of Landsat 8 OLI satellite imagery (http://glovis.usgs.gov/, see Fig. 2) and elevation data. Normalized Difference Vegetation Index, a measure of actively photosynthesizing vegetation (NDVI, Rouse *et al*. 1973), was first calculated using Landsat data (January 2011) for a training dataset of 87 high Andean wetlands mapped on-site in the Norte Chico region by Chile’s Ministry of Environment and Agricultural and Livestock Service. Based on the mean NDVI value of all pixels comprising the 87 field-mapped wetlands, and applying the minimum observed elevation value as a cut-off, we defined wetlands here as all vegetated pixels with NDVI values > 0.2, located above 1800 m a.s.l. To exclude sparse vegetation from the mapping, only vegetated areas ≥ 0.36 ha were considered.

We calculated various local structural connectivity metrics for each wetland using Graphab 1.0 (Foltete *et al*. 2012), including node degree (Dg), clustering coefficient (CC), closeness centrality (CCe), eccentricity (Ec), betweenness centrality (BC) and wetland density. These metrics were quantified by considering all the wetlands mapped within a 20
km radius from the focal site. In addition, we estimated the surface of each focal wetland from the GIS-based map (Fig. 2). As a proxy for wetland stability, we used the temporal range of the mean NDVI of each site calculated over 20 years (from 1991 to 2002 and from 2004 to 2011). Before evaluating the impact of the habitat metrics on diversity, we removed strongly correlated predictors by excluding one variable when a Pearson’s correlation between a pair of habitat metrics was > 0.7. The final set of predictors included wetland size, Dg, CCe, Ec, BC and the temporal range of NDVI wetland values.

**Statistical analysis**

**Identification of outlier loci**

We used the Bayesian likelihood $F_{ST}$-based method implemented in the software BayeScan (Foll and Gaggiotti 2008) to identify outlier loci. This approach estimates for each locus the posterior probabilities of two alternative models: a neutral one and a model including selection. We performed the analyses using the default parameters and considered a conservative value of 10 for the prior odds (10:1 odds in favor of the neutral model). Loci with $F_{ST}$ values substantially higher than the observed norm are considered potentially under diversifying selection, while those with substantially lower $F_{ST}$ values are considered potentially under balancing or purifying selection (Foll and Gaggiotti 2008). These outlier loci were identified with the R function plot_bayescan as described in the user’s manual, using a false discovery rate (FDR, the q-value threshold) of 5%. Accordingly, we examined the effects of outlier loci by creating three marker datasets and considering their $F_{ST}$ values: a dataset free of outliers (DS1), one with all the genotyped loci, thus including both the
non-outlier and all the outlier loci (i.e. with high and low $F_{ST}$, DS2) and one dataset including the non-outlier loci and the outlier loci with high $F_{ST}$ only (DS3).

**Species richness, genetic diversity and the influence of outlier loci on genetic diversity estimates**

We used the statistical software R (R Core Team 2015) to calculate diversity indices and perform SGDC analyses. We estimated species diversity within each wetland as species richness of all the collected quadrats, and within-population genetic diversity as Nei’s gene diversity (Nei 1987). We calculated the bootstrapped 95% confidence interval of species richness and Nei’s indices by considering 10,000 bootstrap replications using the R package rich (Rossi 2011) and R functions of AFLPdat (Ehrich 2006), respectively. Nei’s gene indices were computed for the three marker datasets separately (DS1, DS2 and DS3).

We tested the effects of outlier loci on GD by carrying out linear mixed models (LMM) of Nei’s gene diversity (dependent variable) in relation to the presence of outlier loci (fixed effect), and including species and site as crossed random factors. We performed two such analyses, testing the effects of all the outlier loci (with datasets DS1 and DS2), and of those with high $F_{ST}$ only (with datasets DS1 and DS3). These analyses were also performed for each species separately.

**Effects of habitat characteristics on species and genetic diversity**

We used partial least-square (PLS) regressions to analyze the influence of habitat characteristics on species and genetic diversity. PLS regressions are recommended in analyses involving a large number of potentially correlated predictor variables and few observations (Mevik and Wehrens 2007), as in the case of the present study. Such analyses
first produce a set of uncorrelated linear combinations of the original predictors, the PLS components, calculated so as to maximize the covariance between the predictor and response variables (Mevik and Wehrens 2007). Then, the PLS components are used to derive the linear regression of the dependent variable (i.e. species or genetic diversity in this case). To establish the optimal number of PLS components, we used a leave-one-out cross-validation strategy (Mevik and Cederkvist 2004). Significance of the PLS model was assessed by testing the significance of the Pearson correlation between the observed and predicted values. To determine which habitat characteristics contributed most to the model, we examined the loading weights and performed approximate t-tests based on jackknife variance estimates of regression coefficients, calculated based on the optimal number of PLS components.

Species and genetic diversity correlations

For each marker dataset, we computed Pearson correlations between plant species richness and Nei’s gene diversity index. One-tail tests with Bonferroni-Holm corrections were applied to correct for multiple comparisons. In order to assess the overall trend in SGDCs, we calculated Student’s t-tests to determine whether the average correlation calculated from the five species was greater than zero. To investigate whether the relationships between species and genetic diversity were mediated by habitat isolation, stability, and size, we performed partial Pearson correlations after controlling for significant effects of habitat characteristics on species and genetic diversity. Significance of the residual correlations was tested as described above for the raw SGDC coefficients. LMMs were performed to test for differences in SGDCs between DS1 and DS2 (dataset as fixed effect), using species as a random factor.
Results

Identification of the outlier loci

Depending on the species, between 66 and 90 polymorphic AFLP loci were retained in our study (Table S2). The BayeScan analysis identified between one and ten outlier loci per species, representing between 1.2% and 15.1% of the AFLP loci, respectively, with an average of 6.8% over all species (Table S2). Most of these outlier loci had high $F_{ST}$ values, but in *P. clandestina* and *Austrelmis sp.* two loci showed low $F_{ST}$ values (Table S2).

Species richness and genetic diversity

Plant species richness ranged from 6 to 21 species per wetland (Table S3) and decreased with latitude (Fig. S1). The two plant species showed low genetic diversity overall (range of average Nei’s gene diversity indices: 0.11–0.13; Table S4), varying moderately to substantially between wetlands (more than six-fold in *C. gayana*; Table S4). Genetic diversity estimates of macroinvertebrates were higher (range of average Nei’s gene diversity indices: 0.11–0.28; Table S4) but less variable (Table S4). The presence of outlier loci significantly influenced Nei’s gene diversity. These effects were detected in four of the five species (Fig. 3), and as an overall trend across species (LMM, dataset effect: $LR = 12.52, df = 1, P < 0.001$ for DS1 vs DS2; $LR = 24.77, df = 1, P < 0.001$ for DS1 vs DS3).

The only species for which we failed to detect outlier loci effects on genetic diversity was *C. gayana*, but only one outlier was identified for this species. Overall, GD estimates decreased with the presence of outlier loci (Fig. 3, Fig. S1), and this trend was more pronounced when only outliers with high $F_{ST}$ were included in the genetic datasets than when both outliers with high and low $F_{ST}$ were included.
Species-genetic diversity correlations and the influence of outlier loci

Our analyses revealed positive correlations between species richness and genetic diversity (Table S5). The average SGDCs of the five species was high (ranging from 0.51 ± 0.07 for DS3 to 0.54 ± 0.05 for DS1, Table S5) and significantly positive overall ($t$-tests: $t = 17.4–25.5$, $df = 4$, $P < 0.001$ in all cases). All SGDCs were positive and significant with respect to each species ($P < 0.05$ in all cases after correcting for multiple comparisons). They ranged from 0.47 to 0.60 for DS1 (Table S5). The presence of outlier loci significantly influenced the SGDCs (Fig. 4A, LMM, dataset effect: $LR = 6.05$, $df = 2$, $P = 0.05$), causing a decrease in SGDCs overall (Fig 4A, Table S5). This trend was particularly marked in *P. clandestina* (Fig. 4A), for which the SGDC estimate dropped by 13.5% when outlier loci with high $F_{ST}$ were included (Fig. 4A).

Influence of habitat fragmentation on species richness, genetic diversity and SGDCs

The PLS regressions explained 49% of the variation in wetland plant richness (Table 1). Significant influence of wetland size and connectivity on genetic diversity was found for the plants and the mayfly species (*A. peruvianus*), but no such effects were detected for the long-lived aquatic species (*H. fossamancinii* and *Austrelmis sp*). Wetland isolation was an important determinant of plant species richness and genetic diversity. We found a negative influence of the average distance to neighboring patches (i.e. closeness centrality, Table 1) and positive effects of the number of neighboring patches (i.e. node degree, Table 1) on genetic and species diversity. A positive effect of the distance to the farthest wetland (i.e. eccentricity) on species richness and genetic diversity was observed for the plants and the mayfly species (Table 1).
The effects of habitat fragmentation on GD were not lowered by the presence of outlier loci (i.e. DS2 and DS3, Table 1). In fact, the opposite trend was observed in *P. clandestina* and *A. peruvianus*, with habitat connectivity explaining a greater proportion of the variance in GD for DS2/DS3 than for DS1 (Table 1).

Factoring out the effects of habitat fragmentation on species and genetic diversity significantly influenced SGDC estimates (LMM, fragmentation habitat control effect: \( LR = 8.42, df = 1, P < 0.01 \)), decreasing average partial SDGCs relative to average original values (Fig. 4B, Table S5). The decreases in partial SGDCs were more consistent and pronounced with DS1 than with DS2 (Fig. 4B). In *A. peruvianus*, controlling for habitat fragmentation had strikingly contrasting effects depending on whether or not the outlier loci were included in the analysis (Fig. 4B). As in other species, the partial SGDC decreased with DS1, but increased considerably with DS2, reaching a maximum of 0.7 (Table S5). To confirm that this discrepancy was specifically linked to the three outlier loci detected, we applied a bootstrap procedure to simulate the null distribution of the difference in partial SGDCs between DS1 and DS2 when randomly eliminating three non-outlier loci from DS2. None of the 1,000 bootstrap replications resulted in SGDC differences greater than or equal to the one observed between DS1 and DS2.

**Discussion**

**Confounding effects of outlier loci in SGDC studies**

There is increased awareness that neutral molecular markers are not always free of selection pressures; disregarding this fact can bias population genetic inferences (Luikart et al. 2003, Landguth & Balkenhol 2012). To avoid this problem, Luikart et al. (2003)
advocated excluding $F_{ST}$ outlier loci from genetic analyses when the focus is on neutral processes. In this study, we screened for AFLP outlier loci and investigated their impact on SGDC estimates. We also examined how these loci influence the apparent contribution of habitat characteristics linked to neutral processes (migration rates and rates of stochastic loss of alleles/species) on SGDCs. Our results show that outlier loci led to a marked decrease in GD, which in turn downplayed the strength of SGDCs.

Outlier loci have been suggested to be involved in adaptive processes (Hancock et al. 2011; Manel et al. 2012). Yet, alternative mechanisms, which may or may not be indirectly associated with selective forces, have also been proposed to explain outlier loci, such as correlations in co-ancestry between subpopulations in hierarchical structures, gene surfing, background selection, and even endogenous genetic barriers (Bierne et al. 2011, 2013). Regardless of the actual mechanisms involved in the atypical differentiation patterns of the outlier loci, those with markedly high $F_{ST}$ demonstrated the most pronounced effects in terms of reductions in GD. This result concurs with previous empirical observations based on microsatellite data (Garcia-Verdugo et al. 2015), and thus suggests that GD might often be underestimated when such loci are included in population genetic analyses.

We found a significant downward effect of outlier loci on SGDCs, indicating that SD correlates more strongly with GD of non-outlier loci than with GD of outlier loci. This pattern is consistent with the expectation that SD is more closely related to neutral GD than to non-neutral GD, due to the parallel influence of neutral processes on both diversity levels. Investigating the effects of filtering out outlier loci from SGDC studies is therefore a potentially useful approach to evidence the contribution of neutral processes. In addition, it may contribute to revealing SGDCs, and explain some of the variation in species-genetic
diversity relationships observed in the field. Indeed, the effects of outlier loci are likely to depend on the strength of neutral mechanisms on both diversity components. In high Andean wetlands, and highly fragmented ecosystems in general, neutral evolutionary processes are likely key determinants of biodiversity. As a result, they are expected to generate strong species-genetic diversity covariations (Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014), as were found here, with these SGDCs ranking among the highest values reported to date (Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014). In such conditions, non-neutral processes, and thus outlier loci, should only have minor confounding effects. The influence of outlier loci on SGDCs is likely to become increasingly problematic as the relative importance of neutral processes in shaping diversity patterns decreases. Extending the present framework to less fragmented ecosystems may potentially uncover further SGDC patterns, thereby advancing current understanding of species-genetic diversity relationships.

The presence of outlier loci also complicated the evaluation of the contribution to SGDCs of habitat characteristics supposedly linked to neutral processes. Without such loci, the partial SGDCs, calculated after filtering out the effects of habitat fragmentation on SD and GD, decreased compared to the corresponding non-partial estimates. This indicates concordant responses of SD and GD at non-outlier loci to these habitat characteristics. Outlier loci somehow blurred these trends however, as reduction of partial SGDCs was less consistent and less pronounced among species with outlier loci present. In the mayfly A. peruvianus, the SGDC with outlier loci actually increased when habitat effects were factored out (Fig. 4B). This indicates that contrasting effects of habitat characteristics on SD and GD at outlier loci were partly masking the species-genetic diversity relationship in this species. It further shows that GD at outlier loci and neutral GD can correlate differently.
with habitat features, thus confounding the apparent importance of neutral mechanisms on GD. In fact, for both *P. clandestina* and *A. peruvianus*, the habitat features showed stronger correlations with GD at outlier than at non-outlier loci (Table 1). Habitat stability, size, and connectivity, being intrinsic determinants of population demography, are expected to influence all neutral loci equally (Luikart *et al.* 2003). Their stronger association with outlier loci in two species thus suggests that they can also be indirectly associated with selective processes and local adaptation. In our case, high altitude wetland densities decline in parallel with latitude, and both wetland size and connectivity show spatial structure along this gradient (Fig. S2, Supplementary Material). In such circumstances, the standard approach to disentangling the influence of neutral processes on SGDCs – i.e., factoring out the effects of habitat features – may be ineffective, since it can simultaneously omit the effects of selective environmental factors. Our results thus illustrate the difficulties in inferring evolutionary processes from partial SGDCs when habitat density is paired with environmental gradient, a situation likely to be common in ecosystems subject to climatic and/or topographic influence. To overcome pitfalls in interpreting the contribution of disturbance regime, habitat size and/or connectivity to SGDCs, it would be advisable to analyze the spatial distribution of these variables and evaluate their correlation with other environmental factors as potential actors of selection.

**Species-genetic diversity correlations in high altitude wetlands: patterns and processes**

We found strong spatial associations between species richness and AFLP genetic diversity estimates in high altitude wetlands. Consistent with simulation models suggesting stronger species-genetic diversity relationships when GD is measured in more common species (Vellend 2005), highest SGDC was obtained for *C. gayana*, the most abundant plant
species of high altitude wetlands of Chile’s Norte Chico. Overall, our results add to previous meta-analysis evidence that SGDCs are widespread in discrete habitats functioning as islands (Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014). Furthermore, our results show that such correlations are not limited to component species of the community. For instance, we detected strong SGDCs between plant species richness and GD of the benthic macro-invertebrate taxa, despite the fact that the latter possess very different life-history traits, particularly regarding their dispersal modes.

SGDCs in patchy ecosystems are thought to result from parallel influences of neutral forces governing variation at the genetic and species diversity levels (Vellend and Geber 2005; Whitlock 2014; Vellend et al. 2014). Here, the contribution of such mechanisms is substantiated both by the lowering effect of outlier loci on SGDCs, as well as the apparent common responses of SD and GD at non-outlier loci to habitat features (size, connectivity and stability). The persistence of positive, and overall significant, partial SGDCs after removing all confounding effects (outlier loci and habitat features) suggests, however, that other causal mechanisms might also be involved. These can include both parallel responses to processes unrelated to habitat features, and direct relationships between SD and GD. While our results indicate that potentially selected loci tend to decrease SGDCs overall, they do not rule out the possibility that some selective mechanisms may contribute to species-genetic diversity relationships. For instance, a strong selective agent, e.g. a voracious herbivore or an aggressive exotic plant, acting with increasing intensity upon plant communities located along a spatial gradient might cause a gradual reduction in SD (due to differences in species’ abilities to cope with increased predation/competition) together with a reduction of GD in the “surviving” species, where only some tolerant genotypes could remain; this would generate a positive correlation.
between SD and GD. In our study system, plant SD may affect soil and water characteristics and, as such, may influence local GD by determining the abundance of plant and macroinvertebrate populations. Alternatively, genetic variation within foundation species such as *P. clandestina*, which hosts other plant species, can affect the composition and diversity of associated communities (Whitham *et al.* 2006). Nevertheless, this conjecture would imply a significant correlation between GD at non-outlier loci and GD of specific genes encoding ecologically relevant traits influencing plant assemblages.

Further investigation is required to fully unravel the mechanisms driving SGDCs in high Andean wetlands. Future studies will benefit from next-generation sequencing applied to genome scans, which will very soon allow for more reliable estimates of genes under selection (Manel *et al.* 2016), essential to distinguishing neutral from adaptive markers. Ideally, field studies could complement these efforts by identifying strong interactions between species in the community, putative selective pressures, and patterns of selection and adaptive evolution.

**Acknowledgements**

We thank F. Squeo, G. Arancio, D. Rodriguez, R. Inzulza Ayala, L. Cifuentes, E. Alvarez, C. Urqueta, R. Hereme and A. Troncoso for their help with field work and/or for their technical assistance, and Craig Weideman for revising the English of the manuscript. Our thanks go to P. Jarne and F. Laroche for insightful discussion during the preparation of this manuscript. We also thank four anonymous reviewers for their comments on previous versions of the manuscript. The authors are grateful to La Junta de Vigilancia del Elqui, la Comunidad Agrícola los Huascoaltinos, La Sociedad de Parceleros de Coiron, La Sociedad
de Parceleros de San Agustín, La Sociedad de Parceleros Hacienda Illapel, la Hacienda Tulahuén Oriente, la Hacienda El Maiten de Pedregal and la Hacienda El Bosque for giving us access to private wetlands. This research was funded by a grant from the Chilean National Research Agency FONDECYT #1110514 and the exchange research grant ECOS-CONICYT C12B02.

References


**Data accessibility**

Raw files of the AFLP genotypes (i.e. band presence and absence) for each of the five species and R scripts will be made available from the Dryad Digital Repository upon acceptance of the manuscript.

**Author contributions**

ABertin, NG, ABaumel, EG and SM conceived and designed the study. ABertin, NG, JS and RO performed the experiments. ABertin, NG and ABaumel analyzed the data. ABertin, NG, ABaumel and SM wrote the manuscript. All the authors contributed substantially to revisions.
**Tables**

**Table 1.** Effects of habitat characteristics on genetic diversity and taxa diversity estimated using partial least-square regressions (PLSR). The direction of the effect (positive/negative) of each habitat characteristic on the diversity parameters is reported in parentheses. DS1 refers to the AFLP dataset excluding outlier loci, DS2 to the dataset including both non-outlier and outlier loci (potentially under diversifying or balancing selection), and DS3 to the dataset including non-outlier and outlier loci with high $F_{ST}$ (potentially under diversifying selection).

<table>
<thead>
<tr>
<th>Species diversity</th>
<th>Optimal</th>
<th>Explained variance ($R^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFLP dataset</td>
<td>Number of sites</td>
<td>Number of PLSR axes</td>
</tr>
<tr>
<td>Species diversity</td>
<td>21</td>
<td>3</td>
</tr>
</tbody>
</table>

**Genetic Diversity**

- **C. gayana**
  - DS1: 21 sites, 3 PLSR axes, Node degree (>0), Closeness centrality (<0), Eccentricity (>0)
    - 0.51***
  - DS2 / DS3: 21 sites, 3 PLSR axes, Node degree (>0), Closeness centrality (<0), Eccentricity (>0)
    - 0.50***

- **P. clandestina**
  - DS1: 16 sites, 1 PLSR axes, Node degree (>0), Capacity (>0)
    - 0.19†
|                | DS2 | DS3 | Closeness centrality (<0), Eccentricity (>0), Node degree (>0), Capacity (>0) |  
|----------------|-----|-----|--------------------------------------------------------------------------------|---
|                | 16  | 16  |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
| H. fossamancini |     |     |                                                                                  |  
|                | 14  | 14  | Node degree (>0), Capacity (>0)                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
| A. peruvianus   | DS1 | 14  | 0                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                | DS2 / DS3 | 14 | 0                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
| Austrelmis sp   | DS1 | 15  | 0                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                | DS2 | 15  | 0                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                | DS3 | 15  | 0                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  
|                |     |     |                                                                                  |  

‡: P < 0.1, *: P < 0.05 and ***: P < 0.001
Figure 1. Relationship between species diversity within a community and neutral genetic diversity within a species. Species-genetic diversity correlations (SGDCs) with neutral genetic diversity are assumed to result from parallel responses to neutral processes (black arrows), which relative importance is expected to be determined by habitat features such as disturbance regime, habitat size and/or connectivity. Yet, non-neutral processes can have confounding effects (grey arrows) in the investigation of neutrally-driven SGDCs. They can confound the apparent level of SGDCs driven by neutral processes (confounding effect 1) influencing neutral genetic diversity through linked selection. The contribution of habitat characteristics regulating rates of stochastic immigration and extinction to SGDCs may also be confounded by linked selection if adaptive genetic markers are selected by environmental conditions covarying with those habitat characteristics (confounding effect 2).
Figure 2. Picture of a high altitude wetland and geographical distribution of the sampling sites along Chile’s Norte Chico. The rectangle depicts the remote-sensing area used for identification and mapping of high Andean wetlands (shown in grey).
Figure 3. Average population difference + or – SD between Nei’s gene diversity indices estimated from the AFLP datasets including outlier loci (DS2 and DS3) and the dataset free of outliers (DS1). Asterisks indicate significant differences (P < 0.05) in average population genetic diversity estimated with and without outlier loci for each comparison in each species.
Figure 4. Differences between species genetic diversity correlations (SGDCs) estimated under various scenarios. A) Differences between SGDCs estimated without (SGDC\textsubscript{DS1}) and with outlier loci (SGDC\textsubscript{DS2} and SGDC\textsubscript{DS3}). SGDC\textsubscript{DS2} \textendash SGDC\textsubscript{DS1} refers to the difference between SGDCs calculated with the full AFLP dataset including both non-outlier and outlier loci (DS2) and the dataset excluding all the outlier loci (DS1), and SGDC\textsubscript{DS3} \textendash SGDC\textsubscript{DS1} to the difference between SGDCs calculated with the dataset including non-outlier and outlier loci with high $F_{ST}$ only (DS3) and the dataset excluding all the outlier loci (DS1). B) Difference between raw SGDCs and partial SGDCs, calculated after accounting for habitat characteristics, for each focal species and for datasets without (DS1) and with all the outlier loci (DS2).