Stable inversion clines in a grasshopper species group despite complex geographical history

Guzmán, Noelia V.

2022-02

Guzmán , N V , Kemppainen , P , Monti , D , Castillo , E R D , Rodriguero , M S , Sánchez-Restrepo , A F , Cigliano , M M & Confalonieri , V A 2022 , ' Stable inversion clines in a grasshopper species group despite complex geographical history ' , Molecular Ecology , vol. 31 , no. 4 , pp. 1196-1215 . https://doi.org/10.1111/mec.16305

http://hdl.handle.net/10138/351358
https://doi.org/10.1111/mec.16305

unspecified
acceptedVersion

DR. VIVIANA CONFALONIERI (Orcid ID : 0000-0003-4499-9816)

Article type : Original Article

Stable inversion clines in a grasshopper species group despite complex geographic history

Running title: Adaptive inversion clines in grasshoppers

Noelia V. Guzmán [†], Petri Kemppainen [±], Daniela Monti [†], Elio R. D. Castillo[‡], Marcela S. Rodriguero [†], Andrés F.Sánchez-Restrepo [†,§], Maria Marta Cigliano[¶], and Viviana A. Confalonieri [†,1]

[†] Departamento de Ecología, Genética y Evolución, FCEyN, Universidad de Buenos Aires (UBA), IEGEBA (Consejo Nacional de Investigaciones Científicas y Tecnológicas (CONICET)/UBA), Intendente Güiraldes 2160, Ciudad Universitaria, C1428EGA, Buenos Aires, Argentina.

[±] Ecological Genetics Research Unit, Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland

[‡]Laboratorio de Genética Evolutiva "*Dr. Claudio J. Bidau*", FCEQyN, Universidad Nacional de Misiones (UNaM), Instituto de Biología Subtropical (IBS) (CONICET/UNaM). LQH, Felix de Azara 1552, N3300, Posadas, Misiones, Argentina.

[§]Fundación para el Estudio de Especies Invasivas (FuEDEI). Gral. Simón Bolívar 1559, Hurlingham, Buenos Aires, Argentina.

¶Museo de La Plata, Universidad Nacional de la Plata, Centro de Estudios Parasitológicos y de Vectores (CEPAVE- CONICET/UNLP). Paseo del Bosque s/n, B1900 La Plata, Buenos Aires, Argentina.

[1]Corresponding Author: E-mail: bibilu@ege.fcen.uba.ar; vaconfalonieri@gmail.com. ORCID ID: 0000-0003-4499-9816

Abstract

Chromosomal inversions are known to play roles in adaptation and differentiation in many species. They involve clusters of correlated genes (i.e loci in linkage disequilibrium, LD) possibly associated with environmental variables. The grasshopper "species complex" *Trimerotropis pallidipennis* comprises several genetic lineages distributed from North to South America in arid and semi-arid high-altitude environments. The southernmost lineage, *Trimerotropis* sp., segregates for 4-7 putative inversions that display clinal variation, possibly through adaptation to temperate environments. We analyzed chromosomal, mitochondrial and genome-wide single nucleotide polymorphism (SNP) data in 19 *Trimerotropis* sp. populations mainly distributed along two altitudinal gradients (MS and Ju). Populations across Argentina comprise two main chromosomally and genetically differentiated lineages: one distributed across the southernmost border of the "Andes Centrales", adding evidence for a differentiation hotspot in this area; and the other widely distributed in Argentina. Within the latter, network analytical approaches to LD found three clusters of correlated loci (LD-clusters), with inversion karyotypes explaining >79% of the genetic variation. Outlier loci associated with environmental variables mapped to two of these LD-clusters. Furthermore, despite the complex geographic history indicated by population genetic analyses, the clines in inversion karyotypes have remained stable for more than 20 generations, implicating their role in adaptation and differentiation

within this lineage. We hypothesize that these clines could be the consequence of a coupling between extrinsic postzygotic barriers and spatially varying selection along environmental gradients resulting in a hybrid zone. These results provide a framework for future investigations about candidate genes implicated in rapid adaptation to new environments.

Keywords

Inversion clines, SNPs, Linkage disequilibrium, adaptation, secondary contact, grasshoppers.

Introduction

A central aim in evolutionary genetics is to quantify the influence of adaptive and selectively neutral mutations on genomic variation (Loewe & Hill, 2010, Monroe et al., 2020). The first genomic data from animal model species revealed that single-nucleotide polymorphisms (SNPs) are an important source of neutral as well as adaptive genetic variation (1000 Genomes Project Consortium, 2010; Langley et al., 2012; Wade et al., 2002). However, the increasing availability of whole genome sequencing analyses have led to the conclusion that structural mutations or structural genomic variants (SVs) are not only ubiquitous, but also much more common than previously thought (Dobigny et al., 2017; Feuk, et al., 2005; Kirkpatrick, 2010). Moreover, several studies have emphasized the role of SVs (mainly inversions) in adaptation and diversification of different animal and plant species (Barth et al., 2019; Hoffmann & Rieseberg, 2008; Wellenreuther & Bernatchez, 2018). The underlying evolutionary mechanism of these phenomena is most often unknown and may be due to random association, disruption of gene function at the breakpoints, or preservation of co-adapted alleles through the suppression of recombination (Kirkpatrick & Barton, 2006; Villoutreix et al., 2021), giving rise to so called "supergenes" (Dobzhansky, 1970; Schwander et al., 2014; Thompson & Jiggins, 2014).

These supergenes can comprise large sets of correlated genes that can exert control on extraordinary complex polymorphisms, as in the case of mimetic butterfly forms (Joron et al., 2011), mating strategies in birds (Küpper et al., 2016; Tuttle et al., 2016), social organization of some ants (Purcell et al., 2014; Wang et al., 2013), local adaptation to altitude in bees (Christmas et al., 2019), or fish migratory behavior (Arostegui et al., 2019). In addition, they may play an important role in restricting gene flow between populations differentiated by the SV resulting in the accumulation of mutational changes that may eventually give rise to reproductive isolation and speciation (Coughlan & Willis, 2019; Fuller et al., 2018, 2019). Finally, the frequency of SVs can be associated with environmental variables resulting in clinal distributions of different karyotypic variants (Barth et al., 2019; Confalonieri, 1994; Dobzhansky, 1948; Manoukis et al., 2008; Umina et al., 2005; Wallberg et al., 2017) thus implicating their possible roles in adaptive evolution.

Gradients of genetic variation —or clines— have been classically interpreted as the outcomes of antagonistic interactions between selection and gene flow. The most commonly invoked processes

are as follows: 1) secondary-contact zone between species or populations that evolved in allopatry (i.e. secondary hybrid zones; Barton & Hewitt 1985), where selection against hybrids, for instance due to the emergence of incompatibilities (Bierne et al., 2011) prevents gene flow and generates clines of alleles frequencies; 2) secondary hybrid zones maintained by both endogenous (i.e. genome incompatibilities) and exogenous (i.e local adaptation or fitness effects associated with the environment) factors, or only exogenous factors (i.e. selection against immigrant hybrids in the parental environment coupled with local adaptation of all populations across the cline; Bierne et al., 2011); 3) mixing of populations adapted to different ecological conditions where the ecological transition occurs over short distances (i.e. primary hybrid zones; James et al., 1997; Mullen & Hoekstra 2008) or over wide geographic areas, such as a continent (Balanyà et al., 2009; Hoffmann et al., 2004; Kennington & Hoffmann, 2013). All invoked processes can be favored by the presence of SVs such as inversions, promoting divergence through the suppression of recombination between fitness associated loci (Kirkpatrick & Barton, 2006), giving them central roles in the emergence of hybrid zones in sympatric/parapatric populations (i.e. primary hybrid zones; Feder et al., 2012). However, clines can also appear through two neutral processes: isolation by distance (IBD) and secondary contact without local adaptation of previously isolated populations or any selective disadvantage of hybrids (Novembre & Di Rienzo 2009). In the latter scenario, the clines are nevertheless expected to disappear resulting in the fusion of the populations. Testing these competing hypotheses is challenging, especially for inversion clines of non-model organisms for which the gene contents of the inversions are unknown. The analysis of thousands of SNPs across the genome offers a unique opportunity to search for clusters of loci in high LD associated with those polymorphic chromosomal inversions (Barth et al., 2019; Fang et al., 2020; Huang et al., 2020; Kemppainen, et al., 2015). Furthermore, if these inversion-associated LD-clusters include non-neutral loci associated with endogenous (e.g. incompatibilities) and/ or exogenous factors (e.g. the same environmental variables explaining the clines), distinguishing among different competing adaptive hypotheses will become easier.

In this study, we focus on the analysis of chromosomal inversion clines in the grasshopper species complex *Trimerotropis pallidipennis* (Orthoptera: Acrididae). Grasshoppers include numerous species and genera that experienced rearrangements in polymorphic or fixed states (Hewitt, 1979;

White, 1973). This is particularly the case for *T. pallidipennis*, which is distributed in arid and semiarid regions from North to South America. It has been considered as a "species complex", with morphologically indistinguishable genetic lineages. Biogeographical analyses suggested a first dispersal event of *T. pallidipennis* populations from North to South America. Subsequent dispersion and vicariance probably led to the differentiation of the endemics now found in South America (Guzmán et al., 2017; Husemann et al., 2013). Indeed, one of the southernmost lineages of the complex is distributed across Argentina (hereafter *Trimerotropis* sp.). It is chromosomally and phylogenetically well differentiated from other populations in South America (Guzmán et al., 2017).

Among the members of the complex so far studied, only *Trimerotropis* sp. shows four chromosomes (4, 6, 7 and 8) carrying seven polymorphic rearrangements (considered as inversions based on previous cytogenetic analyses), which probably changed the standard (ancestral) acrocentric morphology to metacentric (M), submetacentric (SM) and inverted acrocentric (AI) ones (Confalonieri & Colombo, 1989). The frequency of most of these inversions tends to be higher at lower altitudes (Figure 1 c-f), showing similar patterns of variation across independent altitudinal gradients (Confalonieri, 1994, 1995). Indeed, three inversion frequencies (4AI, 7SM2 and 8SM4; Figure 1c,e,f, respectively) have been shown to be correlated with altitude, latitude and minimum temperature, and two inversion frequencies (6M and 8SM3; Figure 1d,f, respectively) are associated with longitude and humidity (Colombo & Confalonieri, 1996). Furthermore, one allozyme marker (Est-5) was in LD with chromosome 8, displaying variation like the inversions in this chromosome (Matrajt et al., 1996). *Trimerotropis* sp. is the only lineage of the species complex distributed in both arid and semi-arid high-altitude environments and in more humid ones at lower altitudes. This has prompted Confalonieri (1994) and Colombo and Confalonieri (1996) to suggest that the emergence of these polymorphic inversions could have facilitated the colonization of new habitats at lower altitudes.

In the present study we analyze the genetic structure of samples belonging to *Trimerotropis* sp., which are distributed along two altitudinal gradients in Argentina, as well as in other localities of this country. One altitudinal gradient extends over Mendoza and San Luis provinces (hereafter MS, Figure 1a) and was discovered 30 years ago, while the other (studied here) lies along Jujuy province (hereafter Ju, Figure 1a). These samples were assayed for chromosomal inversions, mitochondrial and

ddRADseq (Double Digest Restriction-site Associated DNA sequencing) variation (Baird et al., 2008; Peterson et al., 2012) to test to what extent the inversion frequency clines could be maintained by non-neutral processes, and to further discuss the previously suggested hypotheses explaining the origin of these clines. Using network analytical approaches to LD of population genomic data together with karyotype information allowed us to study the genomic architecture of the inversions. In addition, we scanned for non-neutral SNPs within and outside these LD-clusters to evaluate their possible association with the same abiotic variables that had explained these gradients. The joint analysis of karyotypic-, mitochondrial- and SNPs-based genetic markers revealed a complex scenario of population substructure, hybrid zones and adaptation to novel environments along ecological gradients where SVs likely have played a crucial role.

Materials and Methods

Sample collection and COI sequence amplification

We collected 178 specimens of *Trimerotropis* sp. from 19 locations of Argentina (Tables S1 and S2; Figure 1) during the summers of 2007–2010. Seven populations [Susques (N=3), Tres Cruces (N=14), Corral Blanco (N=12), La Quiaca (N=19), Uquía (N=9), Posta de Hornillos (N=2) and León (N=13); Figure 1 and Tables S1 and S2] are located along Ju, and five populations [Puente del Inca, hereafter PI (N=14), Uspallata (N=24), Cacheuta (N=12), Mendoza (N=12) and Chosmes (N=16); Figure 1 and Tables S1 and S2] along MS. Both gradients were analyzed for cytochrome *c* oxidase subunit I (*COI;* N=101) and ddRADseq variation (100 individuals from both gradients except for Susques and Posta de Hornillos), and for chromosomal polymorphisms (N=91; Table S1). The remaining samples (N=29) were analyzed only for *COI* (Table S1). Hind legs were preserved in 100% ethanol. Genomic DNA was extracted using the REDExtract-N-Amp TM Tissue PCR kit (Sigma) or DNeasy Blood and Tissue kit (Qiagen) following manufacturer's protocol. *COI* was amplified as in Husemann et al. (2013). Sequences were inspected, trimmed and aligned using Geneious v 7.0.6 (Kearse et al., 2012) and deposited at GenBank (accession numbers: MT037873 - MT038005).

Cytogenetic analysis

Ninety-one males coming from both ecological gradients (Table S1, Figure 1) were dissected and their testes fixed in 3:1 ratio of ethanol and acetic acid solution. Male meiotic preparations were performed by squashing testes follicles in lactopropionic hematoxylin to examine at least 10 metaphase I cells per individual, which is the only stage allowing the morphological identification of homologous chromosomes (i.e. the bivalent). *T. pallidipennis* has 2n=23/24 chromosomes in males/females, with the three largest chromosome pairs being submetacentric, and the remaining ones acrocentric, which is considered the standard (i.e. ancestral) form (Guzmán & Confalonieri, 2010). We determined if chromosomes 4, 6, 7 and 8 are polymorphic for the seven pericentric inversions as described by Confalonieri and Colombo (1989) and Confalonieri (1994). The morphology of these rearrangements supposedly changed from standard acrocentric (A) to five submetacentric (SM1, SM2, SM3, SM4, and SM5), one metacentric (M), and one inverted acrocentric (AI) forms.

ddRADseq analyses

One hundred individuals from MS and Ju (Table S1, and Figure 1) were screened for nuclear genome wide SNPs by ddRADSeq. This is one of the most cost-effective methods for organisms without a reference genome (Peterson et al., 2012). DNA was quantified using ND-1000 spectrophotometer (Nanodrop Technologies) and a Qubit dsDNA HS Assay Kit (Life Technologies), with assays read on a Qubit v2.0 (Life Technologies). We obtained at least 1 μg of high-quality genomic DNA per sample.

Library preparation and sequencing were carried out by IGAtech (Udine, Italy). The libraries were produced using a custom protocol (IGAtech), with minor modifications (Peterson et al., 2012). Briefly, 100 ng of genomic DNA from each individual was digested with enzymes SphI and HindIII. Fragments were then ligated to barcoded adapters, pooled on multiplexing batches and bead-purified. Target fragment distribution (300-500 base pair fragments) was collected on a BluePippin instrument (Sage Instruments Inc., Freedom, CA, US). The eluted fraction was amplified with specific primers and subsequently bead-purified. Libraries were checked with both Qubit v2.0 (Life Technologies) and the Bioanalyzer DNA assay (Agilent Technologies, US) and processed with Ill Wallberg cBot for cluster generation on the flow cell, following the manufacturer's instruction. Paired-end (2*125 bp)

sequencing was performed on a single lane of Illumina Hiseq 2500 (Illumina, San Diego, US) using the Illumina SBS v4 chemistry.

Raw Illumina sequences were processed with STACKS v2.0 following the protocols of Catchen et al. (2011, 2013). Briefly, they were quality-filtered and demultiplexed using *process_radtags* from the STACKS program. Stacks of similar sequences (reads) were then assembled in each individual using *ustacks* and grouped into RAD-loci. A mean of 1.4 million raw reads were obtained per sample. SNP calling was performed through the *pstacks* program, with a minimum allele depth of three sequences (on individual basis) and was applied only to loci with a minimum coverage of six. We obtained a catalogue of loci from all individuals (*cstacks*), and determined the allelic state at each locus in each individual (*sstacks*). RAD-loci were filtered using the *population* program retaining only loci that were represented in at least 75% of the population. Correction of calls (*genotype* program) was applied using information across individuals in the dataset.

SNPs obtained from the STACK analyses of 100 individuals (hereafter referred to as *Rad-loci_0* data set) were subsequently filtered using *vcftools* v0.1.13 (Danecek et al., 2011) by Minor Allele Frequency (MAF>0.03) and Maximum SNP Missingness Rate (0.7). Only one SNP per locus was retained, using R packages reshape2 (Wickham, 2007) and stringr (Wickham, 2019). This matrix (hereafter referred to as *Rad-loci_1* data set with 24494 SNPs) was subsequently filtered with the PLINK program (Purcell et al., 2007; --indep 300 5 2) to select only variants in linkage equilibrium (hereafter *Rad-loci_2* with 5475 SNPs). These two data sets were used for population structure analyses. For Linkage Disequilibrium network analyses (LDna; Kemppainen et al. 2015), a data set containing only the 56 individuals from MS was extracted from *Rad-loci_0* and subsequently filtered with *vcftools* v0.1.13 (Danecek et al., 2011) by MAF>0.10 and Maximum SNP Missingness Rate (0.7). For LDna, loci with low MAF are typically not informative since these can be caused by a few closely related individuals in the data. A higher MAF also focuses the downstream analyses on "common" rather than "rare" variants. While all SNPs from each RAD locus were used, only one SNP from any LD-cluster from a given RAD locus with a minimum r^2>0.7 was kept. This "pre-clustering" resulted in 7707 SNPs (hereafter *Rad-loci_3*; see further details below). LDna was applied only for the MS gradient (*Rad-loci_3* dataset) because i) samples from Ju (except from León) were

found to be highly differentiated from the remaining populations, based on ddRADseq, mitochondrial variability and chromosomal polymorphisms; and ii) they show polymorphisms for new chromosomal rearrangements (see Results section). Finally, a data set containing the same 7707 SNPs as *Rad-loci_3*, but including only individuals from Ju (excepting León, see below) was also obtained (hereafter *Rad-loci_4)*.

Chromosomal analysis along the MS gradient

To test for differences in the proportion of inversions 4AI, 6M, 7SM2, 8SM3 and 8SM4 observed for *Trimerotropis* sp. among altitudes and to compare the present data with those collected during 1986-1991 (Colombo & Confalonieri, 1996), Generalized Linear Models (GLM) were used. The response variables (i.e. presence of the inversion in the total chromosome samples) were adjusted to a *binomial* distribution using the *glm* function (library: stats; R Core Team, 2021). Overdispersion factors of the response variables were calculated by Pearson residuals. To account for the overdispersion observed in the data, we also proved a *beta-binomial* distribution with the *betabin* function (library: aod; Lesnoff and Lancelot, 2012) and compared both models with Akaike Information Criterion for small samples (AICc). We modelled the inversion frequencies (response variable) using Altitude and Time as explanatory variables (InvF ~ Altitude*time). The "time" was considered a factor variable with two levels (current-data and data from 30 years ago). For *betabinomial* distribution an aleatory factor was included (~1). Best fitting models for each inversion were selected by dropping explanatory variables and their interactions from the models and comparing the resulting change in deviance with a F-test (6M, 7SM2, 8SM3 and 8SM4) and λ^2 test (4AI). AICc for inversions 6M, 7SM2, 8SM3 and 8SM4 show better adjustments of the data to a binomial distribution, so a *quasibinomial* function was chosen to account for the overdispersion on these data (scale parameter ≥ 3). The inversion 4AI shows a higher value of the overdispersion factor (scale parameter = 10.5) and the comparison between binomial and betabinomial indicates that the second one works slightly better explaining the data (AICc = 60.5 vs 59.58).

Phylogeographic analysis

Haplotype networks were constructed from *COI* sequences using PopArt v1.7 (Leigh & Briant, 2015) through TCS v1.21 algorithm (Clement et al., 2000) with 95% confidence intervals. Haplotype diversity along with other summary statistics [D_T (Tajima, 1989), Fu's *F*s neutrality test (Fu, 1997) and R_2 (Ramos-Onsins & Rozas, 2002)] were obtained with DNAsp v5.10 (Librado & Rozas, 2009). These statistics were performed to provide information on past population expansions, a result that can be taken as evidence of successful performance of some genetic lineages in relation to their inversion systems. Then, samples were grouped according to their inversion system. Spearman´s correlation coefficient was used to correlate the frequency of the most frequent haplotypes and altitude.

Population structure analyses

Several population statistics such as observed proportion of heterozygous sites per individual (H_o), deviation from Hardy-Weinberg equilibrium per site within populations, and F_{ST} (Weir & Cockerman, 1984) distances among populations were estimated with *Rad-loci_2* and other filtered datasets (see below) with *vcftools* v0.1.13 (Danecek et al., 2011). Nei´s (1972) genetic distances based on this same matrix were calculated using the "StAMPPNeisD" function (library: StAMPP) (Pembleton et al., 2013). Principal component analysis (PCA) was performed with *Rad-loci_1* and *Rad-loci_2*, using "glPca" function (library:adegenet; Jombart, 2008). The first two principal components were plotted with R packages GGPLOT2 (Wickham, 2016) and factoextra (Kassambara & Mundt, 2020).

We ran STRUCTURE v2.3.4 (Pritchard et al., 2000) on the *Rad-loci_2* and other filtered datasets (see below), using an admixture model and correlated allele frequencies. The number of genetic clusters (K) was assessed between one and 10 and four runs were made for each K, with 300000 generations each and a burn-in of 30000. The most probable value of K was inferred using the delta K method (Evanno et al, 2005).

Patterns of isolation-by-distance (IBD) were assessed through Mantel tests using the "mantel. randtest" function (library:ade4; Dray & Dufour, 2007) based on 1000 Monte Carlo simulations to test for significance. We calculated pairwise genetic distances between samples based on *Rad-loci_3* and *Rad-loci_4* dataset (see below) using the "dist.genepop" function (method=2) and geographic

distances between collection localities using the "dist" function (library: adegenet; Jombart, 2008). These analyses were also performed based on *Rad-loci_3,* but excluding non-neutral markers from outlier analyses.

Linkage Disequilibrium Network Analysis (LDna)

To search for sets of highly correlated loci possibly located within chromosomal inversions, we applied the methodology of Kemppainen et al. (2015), Li et al. (2018) and Fang et al. (2021) using the R-package LDna (v2.0; https://github.com/petrikemppainen/LDna). LDna uses a pairwise matrix of LD values, estimated by r^2 with ngsLD (Fox et al., 2019), to produce a single linkage clustering tree where branches represent clusters and the joining of branches represents clusters and/or individual loci merging (i.e. become connected by at least one edge) at a particular LD threshold. Prior to this, we performed the following "pre-cluster" procedure first described in Li et al. (2018): in the absence of a reference genome, we assumed that all SNPs came from the same chromosome and within a window size of 100 SNPs we recursively (in a single-linkage tree) looked for clusters with r^2>0.7 between one locus and any other loci in the cluster using function "LDnClusteringEL". Since all SNPs from a given RAD locus are adjacent to each other in the data set, all resulting LD-clusters were formed within RAD loci (LD between loci from different RAD loci never exceeded r^2>0.7). This procedure only keeps the most connected SNP with respect to LD (the SNP with the highest median LD with all other SNPs in its cluster) when many SNPs are correlated, but may nevertheless keep several SNPs from a RAD locus, as long as they are not strongly correlated, thus compressing the data with minimal loss of information. Based on these SNPs, we: 1) performed standard LDna when only analyzing LD-clusters (branches) with a minimum number of edges, $|E|_{min}$=[20, 50, 100]; 2), extracted the two first principal components (PC1 and PC2; PCAngsd; Meisner & Albrechtsen 2018) based on loci from all different LD-clusters in the tree (also those that are nested within others) and 3) calculated the proportion of variance explained (PVE) by the karyotypes (based on Euclidean distances) for each of the four chromosomes with segregating inversions. Thus, the LD-cluster with the highest PVE (within a given tree) represents the nested set of correlated loci that best explain the genetic variation between karyotypes for a given chromosome. The scripts used for these analyses

were modified from Fang et al. (2021) and work also for genotype likelihood data (see Data Accessibility Statement).

There are well-characterized evolutionary phenomena associated with elevated LD among multiple loci: inversions, geographic structure, and local adaptation, as well as specific predictions about the population genetic signal and the distribution of loci in the genome that arise from these evolutionary phenomena (Fang et al., 2020; Kemppainen et al., 2015). First, PCA of clusters of loci with LD signals caused by polymorphic inversions is expected to separate individuals based on karyotypes. In general, the heterokaryotypes are expected to be intermediate between the homokaryotypes, and show higher H_o. Alternatively, a PCA based on loci whose frequencies are shaped by the interplay of genetic drift and gene flow is expected to separate individuals based on geographic location. Finally, an LD signal caused by selection is expected to cluster individuals based on the selected trait, regardless of geographic location.

Outlier Loci and correlation with environmental variables

Genome scans to detect SNPs association with environmental variables was performed with BayPass (Gautier, 2015). The MS and Ju gradients were analyzed separately, based on *Rad-loci_3* and *Rad-loci_4,* respectively. We run BayPass under the STD (Standard) covariate model by estimating for each SNP the Bayes Factor, the empirical Bayesian P-value and the underlying regression coefficient using an Importance Sampling algorithm (Gautier, 2015). This model explicitly corrects for the scaled covariance matrix across population allele frequencies to completely remove isolation by distance or substructure effects (Coop et al. 2010). The decision criterion used to quantify the strength of evidence (i.e. in favor of association between the SNP and the covariable) was based on the Jeffreys' rule (Jeffreys, 1961), with BF being expressed in deciban (dB) units, via the transformation 10 log10 (BF) as follows: "strong evidence" when 10 <BF< 15, "very strong evidence" when 15 <BF <20, and "decisive evidence" when BF > 20. Outlier loci were explored by blasting their sequences in the NCBI (https://blast.ncbi.nlm.nih.gov/) nucleotide (nr/nt) (BLASTn) and non-redundant protein sequence (BLASTx) database. BLASTn was optimized for highly and somewhat similar sequences, with the default algorithm options. BLASTx was optimized with invertebrate mitochondrial genetic code and default algorithm options.

Environmental data were obtained from climate layers available in the WorldClim database (Fick & Hijmans 2017; www.worldclim.org) and ENVIREM databases (Title & Bemmels 2018; http://envirem.github.io; Table S2). Climatic profiles were calculated using raster R package (Hijmans, 2020). Spearman´s correlation coefficients were calculated to avoid multi-collinearity of variables. We then selected eight geographic and climatic variables (Table S2) as follows: Altitude, Minimum Temperature of coldest month (Bio6); Mean temperature of driest quarter (Bio9); Annual Precipitation (Bio12); Annual Potential EvapoTranspiration (measure of the ability of the atmosphere to remove water, hereafter AnnualPET); Thornthwaite aridity index (index of the degree of water deficit below water need, hereafter Aridity); Climatic Moisture index (metric of relative wetness and aridity, hereafter Moisture); and Continentality (difference between the average temperature of warmest and average temperature of coldest month).

Results

Chromosomal analysis along the MS and Ju altitudinal gradients

The karyotype analysis of males from MS revealed the presence of inversion polymorphisms for chromosomes 4, 6, 7 and 8 as those previously identified by Confalonieri & Colombo (1989; Figure 1c-f and Table S1). As previously reported (Confalonieri, 1994, 1995; Colombo & Confalonieri, 1996), populations at lower altitudes [e.g., Chosmes at 565 m, Table S2] show higher inversion frequencies than those at higher altitudes. Accordingly, GLM best fitting models include only altitude as explanatory variable for all the inversions, except for 6M (Tables S3a, b, Figure 2) which includes altitude and time. The inclusion of the variable "time" in the models do not allow to explain the changes observed in the proportion of the inversions 4A, 7SM2, 8SM3 and 8SM4 (Tables S3a, b), meaning that the cline for these inversion frequencies was stable for the last 30 years (Figure 2). On the contrary, 6M shows differences between the two sampling time points (Table S3b); past data shows a stronger relationship with altitude compared with the current data (coefficient estimate: current data= -2.59e-3 vs 30 years ago= -1.988e-3; Figure 2). Inversion 6SM5 was not tested because it occurs only in two samples, at low frequencies (Table S1).

Corral Blanco and La Quiaca samples from Ju show a standard karyotype, just like PI (Table S2). Instead, Tres Cruces and Uquia samples from Ju show polymorphisms for inversion karyotypes

differing in morphology from those described for the remaining Argentine samples (hereafter called 4SMa, 6SMb, 7SMc and 8SMd) and from those found in any other population studied so far (Figures 1c-f and S1; Table S1). For instance, Tres Cruces shows polymorphisms for an inversion on chromosomes 4 (SMa), 6 (SMb) and 8 (SMd), while Uquia shows the same polymorphism for chromosome 4 as Tres Cruces, and for a new inversion involving chromosome 7 (7SMc; Figures 1c-f and S1; Table S1). Finally, the sample from León, located at the lowest altitude in Ju, shows polymorphism for the chromosome 6 with the same inverted morphology as in the remaining Argentine populations herein studied, and in previous reports (Confalonieri, 1994, 1995; Colombo & Confalonieri, 1996; Figures 1c-f and S1a; Table S1). Frequencies of inversion karyotypes in Ju show no correlation with altitude (4SMa: difference of deviance with null model=-10.81, p=0.345; 6SMb: difference of deviance with null model=-2.37, p=0.468; 8SMD: difference of deviance with null model=-0.8144, p=0.832) except for 7SMc whose frequency decreases significantly at higher altitudes (difference of deviance with null model=-20.08, p=2.2e-16; coefficient estimate=-0.056). However, these correlations should be taken with caution, given the small size of some of the karyotyped samples from Ju.

Phylogeographic analysis

The *COI* fragment analysis of 130 individuals yielded a 629-bp matrix. Gaps or stop-codons were not detected, suggesting no pseudo-gene amplifications. Thirteen different haplotypes were found (T1-T13; Figure 1b, Table S1). Haplotype diversity per sample (*Hd*) ranges from 0 to 0.818 (Table S4). The network shows three most frequent haplotypes: T1, T2 and T3 (Figure 1b; Table S1). T1 appears all along MS except in PI, but in the Ju gradient it only appears in León and Posta de Hornillos, being the most frequent and widely distributed haplotype (almost 55% of the samples; Figure 1a; Table S1). T2 is separated from T1 by one mutational step and is mainly represented by individuals from PI and Uspallata along MS, and from Cabra Corral in Salta province. T3 occurs in four of the seven samples along Ju and is differentiated from the remaining haplotypes by several mutational steps. Uspallata has the highest level of haplotype variability and *Hd*, and the only sample containing the most frequent haplotypes T1 and T2, as well as four other endemic haplotypes (T8, T9, T10 and T11; Figure 1a,b; Tables S1 and S4). The frequencies of both T1 and T3 are significantly

correlated with altitude, but with opposite patterns (R=-0.7659, p=0.000131; and R=0.7327, p=0.0046, respectively). Conversely, T2 is not correlated with this variable.

Estimates of D_T (Tajima, 1989), Fu's *Fs* neutrality test (Fu, 1997) and R_2 (Ramos-Onsins & Rozas, 2002; Table 1) indicated significant departures from the neutrality hypothesis when the analyses included: i) populations from MS and ii) all samples from Argentina excluding Ju. Small *Fs* statistic points to a recent population expansion as the most likely explanation. On the contrary, samples from Ju show no departure from neutrality.

Population structure analyses

Despite using very stringent filtering criteria, 1003094 SNPs scattered across 24494 genotyped loci were retrieved from the STACK analysis. Most variants from *Rad-loci_2* are in Hardy-Weinberg equilibrium within each location (Table S4). Mean H_o varies from 6.1 to 13.5%; with the percentage of heterozygosity being higher in MS samples (Table S4).

PCAs based on *Rad-loci_2* (Figure 3a) and *Rad-loci_1* yielded the same five groups, composed of samples from: 1) PI; 2) Uspallata; 3) Mendoza, Cacheuta and Chosmes (hereafter group 3); 4) León; 5) Uquía, Tres Cruces and La Quiaca (hereafter group 5). MS is widely spread along PC2 and PC1, with Uspallata placed in between the other two groups from this gradient. Along PC1, León is almost equidistant from the remaining samples of Ju and group 3. The heatmap of pairwise mean F_{ST} and Nei´s genetic distances (Figure 3b) exhibited almost similar patterns of population differentiation as PCA. The samples belonging to groups 3 and 5 show low levels of within group differentiation; genetic distances between PI and the remaining populations are high except for Uspallata. All comparisons among samples from different gradients reveal high levels of differentiation. Finally, León is genetically more similar to group 3 than to group 5, despite its closer geographic proximity with the latter.

Lack of correlation between pairwise geographic and genetic distances for MS gradient indicates that there is no IBD (mantel test, simulated p-value=0.338). The opposite pattern was observed for the Ju gradient (simulated p-value=0.0006).

STRUCTURE analyses based on *Rad-loci*_2 suggests K=4 as the most probable number of panmictic units (Table S5), with Ju, MS, León and PI as separate units and Uspallata as a hybrid

population (Figure 3c; see Figure S2 for more details on K=2 and K=3 clustering). Cluster analysis solely based on MS (*Rad-loci*_3), shows a similar structure, but with K=2 being the most probable number of panmictic units (Figure S3 a, Table S5).

Linkage Disequilibrium Network Analysis (LDna)

To investigate the different evolutionary phenomena causing LD and their relationships with chromosomal inversions in MS, we analyzed the association between the genetic information in all nested sets of LD-cluster in LD-based single linkage clustering trees and the inversion karyotypes for each of the chromosomes 4, 6, 7 and 8. Regardless of the minimum size of clusters analyzed ($|E|_{min}$=[20,50,100]), LDna identified the same four branches (LD-clusters) where the karyotypes explained the highest amount of the genetic variance (based on PC1 and PC2) for each of the four chromosomes: 517 (Chr. 4), 782 (Chr. 6), 845 (Chr. 7) and 632 (Chr. 8) (Figure S4 a-c). For these LD-clusters, karyotype explained 85%, 65%, 79% and 85% of the total variance in PC1 and PC2, respectively and contained 59 (123), 31 (45), 38 (53) and 23 (36) loci, respectively, with values in brackets also including loci that were not used in LDna but were nevertheless linked to the LD-cluster loci with r^2>0.7 (i.e. removed in the pre-clustering). Each karyotype in the PCAs is enclosed by a convex hull in Figure 4a, with a small hull indicating that individuals with the same karyotype are also genetically highly similar (for the loci in each respective LD-cluster). In general, individuals with high heterozygosity tend to be intermediate between individuals that have low heterozygosity, which is consistent with them being heterokaryotypes. When the PCA is performed excluding SNPs from these LD-clusters, individuals are clearly grouped according to their geographical location, with the same ordering pattern as in the former PCA with *Rad-loci_2* and with small variance in heterozygosity (Figures 3a and 5a). When the same analyses were done using only the SNPs from the four LD-clusters and all samples from the MS cline, individuals separated into three groups, but according to their level of H_O, instead of their geographical location (Figure 5b, c, d, e). Cluster 845 shows the most striking pattern (Figure 5d) as it includes two groups with low levels of H_O, and a third one in a middle position consisting of 10 individuals (six males and 4 females) from Uspallata, Cacheuta, Mendoza and Chosmes with higher levels of H_O. Five out of six males of this group have been scored as SM2/A heterokaryotypes.

Note, however, that not all highly heterozygous individuals have been scored as heterokaryotypes and some individuals with high heterozygosity have been scored as homokaryotypes (Figure 4a). This is particularly clear for Chromosome 6, where the karyotypes only explain 65% of the genetic variation. These mismatches indicate either karyotyping errors or the fact that the loci do not correctly reflect the genetic variation among the different karyotypes. The fact that the karyotypes typically also explain a large proportion of the genetic variance for the entire *Rad-loci_3* data set (the LD-cluster representing the root of each LDna-tree; 85%, 62%, 68% and 74%, respectively) indicates that the karyotype frequencies also co-vary strongly with the overall genetic structure in the data (Figure S4), particularly the chromosome 4 inversion, where the full data set explained almost as much information as the 517 LD-cluster (84.5% vs 85.4%). Our results are nevertheless consistent with the notion that chromosomal rearrangements are the most likely cause of the LD-clusters for chromosomes 4, 7 and 8.

Genetic differentiation (F_{ST}) among populations based on SNPs within the putative inversion clusters are in all cases higher than those based on the remaining *Rad-loci_3* loci (Figure 4 b), consistent with lower levels of gene flow between the heterokaryotypes due to suppression of recombination. The STRUCTURE analyses based on *Rad-loci_3* including and excluding SNPs within the four LD-clusters (Figure S3 a and b, respectively, Table S5) revealed K=2 as the most probable number of panmictic units, and similar patterns of structure for the MS samples as with *Rad-loci_2*.

Outlier Loci and correlation with environmental variables

Allele frequencies of up to 51 out of 7707 SNPs in *Rad-loci_3* from the MS cline correlated with different environmental variables with a BF>10 (Figure 6; Table S6). Altitude, AnnualPET and Bio6 explain the variation of most molecular markers, including (55370_7; MAF=0.33) which varies with a BF>30 (i.e. decisive evidence) in relation to these variables and Bio9 (Figure 6; Table S6). Although Aridity, Bio9 and Moisture explain the variability of a small number of markers, some of them show very high BF-values (e.g. 15250_36; MAF=0.12). One outlier SNP belongs to cluster 845 (100240_75; MAF=0.23; correlated with Altitude and Minimum Temperature), and two belong to cluster 632 (1469_198; MAF=0.18; correlated with Altitude, Annual-Pet, Minimum Temperature and

Continentality and 59541_18; MAF=0.16; correlated with Altitude). SNPs 1469_198 and 59541_18 are also correlated with one (1469_212) and six (59541_130, 59541_152, 59541_165, 59541_168, 59541_169, 59541_181) additional SNPs within their respective RAD loci (r^2>0.7). To determine if non neutral SNPs are responsible for the pattern of structure within MS, we performed STRUCTURE and IBD analyses based on *Rad-loci_3* but excluding the 51 SNPs correlated with the different environmental variables (Figure S3 c). We obtained K=2 as the most probable number of panmictic units (Table S5), and a similar pattern of structure for MS as with *Rad-loci_2*; IBD was not significant, as in previous analyses (Mantel test, simulated p-value=0.321).

The analysis of *Rad-loci_4* from Ju (excepting León) revealed 46 loci significantly associated with environmental variables, none of which are shared with the other gradient (Figure 6; Table S6). In this case, Altitude, Bio9, Bio6, Aridity and Bio12 are the most important explanatory variables in terms of associated number of SNPs and strength of evidence. León was excluded from Ju because of its inversion system, type of *COI* haplotype and the results of genetic structure analyses, which indicate that it has more genetic affinity with MS than with Ju populations, despite the geographical proximity with the latter.

BLASTn analysis of outlier loci found within clusters 845 and 632 revealed that the haplotype of 230 bp containing SNP 100240_75 (cluster 845/ chromosome 7) exhibit high similarity with a repeat region from *Locusta migratoria* that correspond to a RTE-2 retrotransposon (Ruiz-Ruano et al., 2018) (Table S7). Less stringent BLASTn searches and BLASTx analysis retrieved homology with myosin and troponin muscle genes (Li et al., 2016; Liang et al., 2020) and with a sulfotransferase enzyme (Chen et al., 2021), all from *L. migratoria* (Table S7). Outlier loci from cluster 632 did not match with any known sequence.

Discussion

The population genomic analysis of *Trimerotropis* sp. lineage revealed new evidence for the simultaneous occurrence of both extrinsic postzygotic barriers and local adaptation maintaining inversion clines and the role these SVs may play in the adaptation to new environments. We found support for the presence of associations between three out of the four chromosomes (4, 7 and 8) in the MS cline (PVE>79%) that also contain outlier loci associated with up to four of the environmental

variables studied here. Our analyses also reveal that, for the MS gradient, different lineages possibly evolved in allopatry under the influence of natural selection. These observations suggest that there is a secondary/primary hybrid zone in addition to selection across the altitudinal gradient. Conversely, most samples from the other altitudinal cline (Ju) belong to a different genetic lineage, bearing an apparently different inversion system. These results uncover an entirely new scenario involving a complex biogeographical structure, as previously seen for other Peruvian populations of the *Trimerotropis pallidipennis* species complex (Husemann et al., 2013; Guzmán et al., 2017), in which Pleistocene glaciations have undoubtedly played a crucial roles.

Population genomic approaches reveal clusters of loci in high LD associated to inversions

The LDna approach (Kemppainen et al., 2015) applied in the MS gradient, revealed the existence of four clusters that most probably correspond to inverted rearrangements. Accordingly, F_{ST} genetic distances among populations from both extremes of the altitudinal cline (e.g. PI vs Chosmes) are higher for SNPs within than outside the clusters, reinforcing the hypothesis that they correspond to inversions (Figure 4 b). Faria et al. (2019) and Westram et al., (2021) used a similar approach in several replicative hybrid zones of the coastal marine snail *Littorina saxatia* and found many large putative inversions "de novo" that could be contributing to divergence between ecotypes in all locations. Similarly, McKinney et al., (2020) identified an overlapping chromosome inversion and sex-determining region in chum salmon, and Huang et al. (2020) discovered inverted regions containing outlier SNPs associated with different ecological parameters, contributing to adaptive divergence of a dune sunflower ecotype. Therefore, the present study is, as far as we know, a first example of application of this network analytical approach to LD to investigate inversions formerly described as "putative" by classical tools.

While there was a strong association between karyotype and LD-clusters for three out of the four chromosomes, the association between genetic cluster, heterozygosity and inversion karyotype was not 1:1 for any of them. There could be several non-mutually exclusive explanations for this. First, chromosomes 6 and 8 bear more than one rearrangement. Consequently, a larger sample size is required to cover all karyotypes and, most importantly, the pattern of PCA segregations expected for

individuals carrying the different rearrangements may be obscured by inversion overlap. Second, gene conversion and double crossovers may result in some genetic exchange ("gene flux") between inverted and standard karyotypes, leading to incomplete suppression of recombination in heterokaryotypes (Guerrero et al., 2012; Pegueroles et al., 2013; Stevison et al., 2011). This may alter the expected pattern for heterokaryotypes with higher levels of H_o. Third, the possibility of cytological misidentification of some rearrangements (e.g. 6M and 7SM2 can be confounded due to subtle differences in chromosome size and centromere position). Finally, the identified LD-clusters may not correspond to any of the chromosomal rearrangements previously identified cytologically for this lineage. Future analysis including more individuals from the same or independent gradients will be necessary to disentangle these explanations. Furthermore, even a simple genetic linkage map would allow us to determine the approximate positions of the LD-cluster loci in the genome (Kemppainen et al., 2015).

Stability of MS altitudinal cline, divergent genetic lineages and hybrid zones

The cytological analysis of MS revealed that polymorphisms due to inversion rearrangements remained stable for at least over 20 generations. Confalonieri and Colombo (1989) reported similar variation patterns for an altitudinal gradient in La Rioja Province that is not far from the Ju gradient (Figure 1 c-f). However, populations from Ju exhibit a different inversion system (Figure 1 c-f, and S1) and are genetically differentiated. This genetic lineage is geographically distributed across the southernmost border of the "Andes Centrales", adding evidence that this mountain landscape is a differentiation hotspot (Guzmán et al., 2017). According to the phylogenetic reconstruction performed by Guzmán et al., (2017), it separated from the other *Trimerotropis* sp. populations about 300,000 years ago, most probably because of Pleistocene glaciations.

Within MS, Uspallata emerges in all analysis as an admixed population, showing the highest COI diversity with many endemic haplotypes. The pattern of admixture is maintained even when outlier SNPs and those from the four clusters are excluded (Figure S3), suggesting the existence of a barrier to gene flow. We hypothesize that Uspallata most likely is a "hybrid" population originated by the admixture of the populations at both extremes of the altitudinal cline. Puente del Inca might have been isolated for some time from the rest of the populations of the *Trimerotropis* sp. lineage or could

have maintained a continuous level of gene flow with the remaining samples. Therefore, we propose that the MS cline represents a typical secondary or primary hybrid zone (Barton & Hewitt, 1985; Harrison, 1993; Harrison & Larson, 2016) maintained by selection. The use of model-based approaches will be necessary to disentangle these two scenarios. Non-neutrality is supported by the absence of an IBD pattern of variability distribution (Novembre & Di Rienzo, 2009) and by the fact that the cline has remained stable for at least 20 generations. In the absence of selection against hybrids, a secondary contact between populations that diverged in isolation would have resulted in the gradual fusion of populations and therefore in an unstable cline in the MS population.

It is important to note that such a hybrid zone could give rise to genetic patterns like inversions when applying LDna. Furthermore, if inversions correlate with this zone, it is difficult to distinguish these two patterns from each other. However, the comparison between loci within and outside LD-clusters showed different heterozygosity patterns and levels of genetic differentiation, making the case that these indeed are directly influenced by segregating inversions. This suggests that LDna can be sensitive enough to correctly detect inversion polymorphism in population genomic data despite being strongly confounded by population demographic history (the admixture zone).

Hybrid zones are formed through the emergence of intrinsic or extrinsic postzygotic barriers to gene flow (Bierne et al., 2011; Barth et al., 2019). Theory predicts that genes controlling hybrid fitness depression distributed across large portions of the genome is the most efficient selection process to prevent the flow of neutral genes (Barton & Bengtsson, 1986). As a consequence of their recombination-suppression effect, chromosomal inversions can facilitate the accumulation of many reproductive barrier loci (Noor et al., 2001; Lowry & Willis, 2010; Ayala, et al., 2012); they can even protect a particular array of locally adapted alleles from disruption through recombination (e.g., Fuller et al., 2017; Kapun et al., 2016; Kirkpatrick & Barton, 2006; Rane et al., 2015). Inversions 517, 782, 845 and 632 are prime candidates for a barrier mechanism in *Trimerotropis* sp. because higher levels of genomic differentiation were found within than outside these genomic regions in high LD. The accumulation of alleles involved in incompatibilities does not require environmental changes (Bierne et al., 2011). However, intrinsic incompatibilities or even extrinsic postzygotic barriers where hybrids are unfit in the local environment (i.e. Uspallata) seem unlikely because 1) problems during meiotic chromosome pairing in heterozygotes can be ruled out (Confalonieri, 1986), and 2) most SNPs (Table

S4) and chromosomal inversions (Table S8) conform HW expectations. Furthermore, the observed stability of most inversion clines indicates that the tension zone has not moved at least during this period of time, suggesting that only intrinsic postzygotic barriers (i.e. incompatibilities) cannot explain the *Trimerotropis sp.* gradient (Bierne et al., 2011). The absence of chromosome rearrangements in PI and the gradual decrease in their frequencies at lower altitudes, may suggest the existence of extrinsic post-zygotic barriers with reduced fitness of hybrids in the parental environments. Selection may have favored inversions at the lowest end of the hybrid zone in the MS cline, keeping together combinations of locally adapted alleles and preventing gene flow with migrant haplotypes that would produce less fit individuals. In this way, *Trimerotropis* sp. might have been able to reach "temperate grasslands savannas and shrublands" at its southernmost distribution during the Late Pleistocene (Guzmán et al., 2017). Successful colonization of these new habitats would have been principally achieved by T1 haplotype carriers, which would have experienced a significant increase in population numbers (Table 1). This may also explain the significant and negative correlation of T1 carriers with altitude. Similar adaptive barrier to gene flow and immigrant inviability was detected for genetically differentiated Atlantic cod types (*Gadus morhua*) coexisting in a confined fjord environment (Barth et al., 2019). Behavioral data indicated that North Sea cod and individuals homozygous for one of the three inversions found had lower fitness in the fjord environment.

Does spatially varying (clinal) selection along environmental gradients also explain chromosomal variation?

Our genome-scan analysis of adaptive loci from MS revealed that altitude, annual potential evapotranspiration and minimum temperature of the coldest month are the most important geographic and environmental variables that significantly explain the variation of the majority of outlier markers, after correction for population structuring. One outlier (100240_75) belongs to inversion 7SM2 (LD-cluster 845), associated with altitude and Minimum Temperature, the same variables that explain the variation of this inversion across 28 populations from Argentina, along latitudinal and longitudinal gradients between 26º and 40º S Lat. (1500 km) and between 64º and 71º E. Long. (900 km), respectively, in a range of altitudes from 79 to 2900 m. (Colombo & Confalonieri, 1996). Other two

outliers (1469_198 and 59541_18) belong to chromosome 8 (LD-cluster 638), associated with variables that also explain the variation of 8SM4 (altitude and Minimum Temperature) and 8SM3 (Annual Potential EvapoTranspiration as a measure of Humidity; Colombo & Confalonieri, 1996). Such results suggest that the frequencies of these markers depend not only on the interplay of migration and local adaptation across a hybrid zone, but also on the environmental conditions that determine spatially varying selection coefficients. These selective advantages most probably paved the way to conquest new environments, thereby leading to niche diversification. Similar patterns of adaptive clinal variation across wide geographic areas were found in other insects, like *Drosophila* (e.g. Anderson et al., 2005; Kapun & Flatt, 2019) and *Anopheles* (e.g. Cheng et al., 2012; Ayala et al., 2012, 2019). This comparison adds further support for association between markers (100240_75, 1469_198 and 59541_18) and chromosome inversions (7SM2 and 8SM3/8SM4) and suggests that their clines are maintained by these non-neutral processes. Adaptive genes residing in inversions have been described in many organisms (see Villoutreix et al., 2021 and reference stherein), including the threespine sticklebacks (Jones et al., 2012), *Heliconius* butterflies (Huber et al., 2015), the willow warbler (Lundberg et al., 2017), and more recently the chum salmon (*Oncorhynchus keta*) (McKinney et al., 2020), in which genes associated with life-history variation were found within an inversion also exhibiting spatial variation in frequency throughout western Alaska.

The nucleotide query containing the marker 100240_75 residing in the inversion 7SM2 has homology with the RTE-2 retrotransposon (Ruiz-Ruano et al, 2018) and with a sulfotransferase (ST) enzyme (Chen et al., 2021), both from *Locusta migratoria***.** Villoutreix et al. (2021) recently highlighted the importance of adaptative mutations at breakpoints during the evolution of inversions in natural populations. Transposable Elements (TEs) can be involved in these breakpoints (Mathiopoulos et al. 1998; Tevy et al., 2007; Miller et al., 2020) because their insertions may serve as substrates for recombination events (Sharakhov et al., 2002). They can even control the expression of genes in adjacent genomic regions (Elbarbary, et al. 2016; Lerman et al., 2003; Sundaram & Wysocka, 2020; Whitelaw & Martin, 2001), leading to adaptive changes. Sulfotransferasa enzyme is involved in *Locusta* in the behavioral shift from gregarious to solitarious state, thus regulating its behavioral plasticity through molecular sulfatation in the brain (Chen et al., 2021). Long-read technologies such as Nanopore (Jain et al., 2016) and PacBio (Eid et al., 2009) will contribute to the

knowledge of the genomic structure and location of RTE-2 and ST gene, and their possible role as adaptive breakpoints and in the behavioral plasticity in *Trimerotropis* sp lineage.

Finally, it is important to note that most markers behaving as non-neutral were not found within the inversion associated LD-clusters (Figure 6). All these SNPs may not be physically linked, but they could exhibit some level of LD because they are associated with the same environmental gradients (i.e. non-gametic-phase disequilibrium). This result contradicts the theory predicting that increased linkage is favored for alleles with similar pleiotropic effects, while increased recombination is favored among alleles with contrasting pleiotropic effects, when local adaptation is driven by complex and non-covarying stresses (Lotterhos et al., 2018). To conclude, the way genomic architecture evolves can be very complex across environmental gradients. The challenge of future studies will be to identify these loci. Whole genome sequencing and a highly continuous reference genome will be crucial to achieve this goal.

Acknowledgement

We gratefully acknowledge the suggestions made by the three anonymous reviewers. This contribution was supported by grants from Agencia Nacional de Promoción Científica y Tecnológica (PICT-2016-2396) and UBA (20020170100409BA) to VAC, and from the Finnish Cultural Foundation (00190489) to PK. NVG, DM, ERDC, MSR, MMC and VAC are members of the Research Career of CONICET, Argentina.

References

Anderson, A. R., Hoffmann, A. A., Mckechnie, S. W., Umina, P. A., & Weeks, A. R. (2005). The latitudinal cline in the In(3R)Payne inversion polymorphism has shifted in the last 20 years in Australian *Drosophila melanogaster* populations. *Molecular Ecology, 14(3),* 851–858. https://doi:10.1111/j.1365-294x.2005.02445.x

Arostegui, M.C., Quinn, T.P., Seeb, L.W., Seeb, J.E. & Lerman, G.J. (2019). Retention of a chromosomal inversion from an anadromous ancestor provides the genetic basis for alternative

freshwater ecotypes in rainbow trout. *Molecular Ecolology*, **28** (6) 1412–1427. https://doi.org/10.1111/mec.15037

Ayala, D., Guerrero, R. F., & Kirkpatrick, M. (2012). Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution*, **67**(4), 946–958. https://doi:10.1111/j.1558-5646.2012.01836.x

Ayala, D., Zhang, S., Chateau, M., Fouet, C., Morlais, I., Costantini, C., Matthew W. H., Besansky, N. (2019). Association mapping desiccation resistance within chromosomal inversions in the African malaria vector *Anopheles gambiae. Molecular Ecology,* **28**(6), 1333-1342. https://doi.org/10.1111/mec.14880

Balanyà, J., Huey, R., Gilchrist, G. W. & Serra, L. (2009). The chromosomal polymorphism of *Drosophila subobscura*: a microevolutionary weapon to monitor global change. *Heredity,* **103,** 364–367. https://doi.org/10.1038/hdy.2009.86

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE,* **3**(10), *e3376.* https://doi.org/10.1371/journal.pone.0003376

Barth, J. M. I., Villegas-Ríos, D., Freitas, C., Moland, E., Star, B., André, C.,… Jentoft, S.(2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. *Molecular Ecol*ogy, **28**(6), 1394– 1411. https://doi.org/10.1111/mec.15010

Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics,* **16**(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553.

Barton, N., Bengtsson, B. (1986) The barrier to genetic exchange between hybridising populations. *Heredity,* **57,** 357–376 (1986). http://sci-hub.tw/10.1038/hdy.1986.135

Bierne, N., Welch, J., Loire, E., Bonhomme, F., David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**(10), 2044-2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics,* **1**(3), 171–182. https://doi.org/10.1534/g3.111.000240

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology,* **22**(11), 3124–3140. https://doi.org/10.1111/mec.12354

Chen, B., Tong, X., Zhang, X., Gui, W., Ai, G., Huang, L., Ding, D., Zhang, J.,Kang, L. (2021) Sulfation modification of dopamine in brain regulates aggregative behavior of animals, National Science Review, nwab163, https://doi.org/10.1093/nsr/nwab163

Cheng, C., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., & Besansky, N. J. (2012). Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing approach. *Genetics, 190(4),* 1417–1432. https://doi.org/10.1534/genetics.111.137794

Christmas, M.J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O. & Webster, MT. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecol*ogy, **28** (6) 1358– 1374. https://doi.org/10.1111/mec.14944

Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology,* **9**(10), 1657–1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

Colombo, P. C., & Confalonieri, V. A. (1996). An adaptive pattern of inversion polymorphisms in *Trimerotropis Pallidipennis* (Orthoptera). Correlation with environmental variables: an overall view. *Hereditas,* **125**(2-3), 289–296. https://doi.org/10.1111/j.1601-5223.1996.00289.x

Confalonieri, V.A. & Colombo, P.C. (1989) Inversion polymorphisms in *Trimerotropis pallidipennis* (Orthoptera): clinal variation along an altitudinal gradient. *Heredity,* **62**(1), 107-112. https://doi.org/10.1038/hdy.1989.14

Confalonieri, V. A. (1994). Inversion Polymorphisms and Natural Selection in *Trimerotropis Pallidipennis* (Orthoptera): Correlations with Geographical Variables. *Hereditas,* **121**(1*),* 79–86. https://doi.org/10.1111/j.1601-5223.1994.00079.x

Confalonieri, V.A. (1995). Macrogeographic patterns in B-chromosome and inversion polymorphisms of the grasshopper *Trimerotropis pallidipennis*. *Genetics Selection Evolution,* **27**(4), 305-311. https://doi.org/10.1186/1297-9686-27-4-305

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics,* **185**(4), 1411–1423. https://doi.org/10.1534/genetics.110.114819

Coughlan, JM, Willis, JH. (2019) Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. *Molecular Ecology* , **28**, 1343–1357. https://doi.org/10.1111/mec.14804

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., …Durbin, R. 1000 Genomes project analysis group. (2011). The variant call format and VCFtools. *Bioinformatics*, **27**(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dobigny, G., Britton-Davidian, J. & Robinson, T.J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, **92**(1), 1-21. https://doi.org/10.1111/brv.12213

Dobzhansky, T. (1948). Genetics of natural populations. XVI. Altitudinal and seasonal changes produced by natural selection in certain populations of *Drosophila pseudoobscura* and *Drosophila persimilis*. *Genetics,* **33**, 158-176. https://doi.org/10.1093/genetics/33.2.158

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. Columbia University Press, New York.

Dray, S. & Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software,* **22**(4), 1-20. https://doi.org/10.18637/jss.v022.i04.

Eid,J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., …Turner S. (2009) Real-time DNA sequencing from single polymerase molecules. *Science*, **323**(5910),133-138. https://doi.org/10.1126/science.1162986

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. *Science*, **351**(6274), aac7247. https://doi.org/10.1126/science.aac7247

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology,* **14**(8), 2611–2620. https://doi.org/j.1365-294x.2005.02553.x

Fang, B., Kemppainen, P., Momigliano, P., Feng X. & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology and Evolution,* **4,** 1105–1115. https://doi.org/10.1038/s41559-020-1222-6

Fang, B., Kemppainen, P., Momigliano, P., Feng, X. & Merilä, J. (2021)., Population structure limits parallel evolution in sticklebacks. *Molecular Biology and Evolution*, **38**(10), 4205–4221, https://doi.org/10.1093/molbev/msab144

Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular ecology*, **28**(6), 1375–1393. https://doi.org/10.1111/mec.14972

Feder, J.L., Egan, S.P., Nosil, P. (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**(7), 342-350, https://doi.org/10.1016/j.tig.2012.03.009.

Feuk, L., MacDonald, J.R., Tang, T., Carson, A.R., Li, M., Rao, G.,…,Scherer, S.W. (2005). Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLoS Genet.*, **1**(4),e56. https://doi.org/10.1371/journal.pgen.0010056

Fick, S.E., & Hijmans, R.J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology,* **37** (12), 4302-4315. https://doi.org/10.1002/joc.5086

Fox, E.A., Wright, A.E., Fumagalli, M., Vieira, F.G. (2019) ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, **35**(19), 3855-3856. https://doi.org/10.1093/bioinformatics/btz200

Fu Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**(2), 915–925. https://doi.org/10.1093/genetics/147.2.915

Fuller, Z. L., Haynes, G. D., Richards, S., & Schaeffer, S. W. (2017). Genomics of natural populations: Evolutionary forces that establish and maintain gene arrangements in *Drosophila pseudoobscura* . *Molecular Ecology,* **26**(23), 6539–6562. https://doi.org/10.1111/mec.14381

Fuller, Z. L., Leonard, C. J., Young, R. E., Schaeffer, S. W., & Phadnis, N. (2018). Ancestral polymorphisms explain the role of chromosomal inversions in speciation. *PLOS Genetics,* **14**(7), *e1007526.* https://doi.org/10.1371/journal.pgen.1007526

Fuller, Z. L., Koury, S. A., Phadnis, N., & Schaeffer, S. W. (2019). How chromosomal rearrangements shape adaptation and speciation: Case studies in *Drosophila pseudoobscura* and its sibling species *D. persimilis. Molecular Ecology,* **28**(6), 1333-1342. https://doi.org/10.1111/mec.14923

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, **201**(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Goudet, J & Jombart, T (2020). hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.5-7. https://CRAN.R-project.org/package=hierfstat

Guerrero, R.F., Kirkpatrick, M. & Perrin, N. (2012). Cryptic recombination in the ever-young sex chromosomes of *Hylid* frogs. *Journal of Evolutionary Biology,* **25**(10), 1947-1954. https://doi.org/10.1111/j.1420-9101.2012.02591.x

Guzmán, N. V., & Confalonieri V. A. (2010). The evolution of South American populations of *Trimerotropis pallidipennis* (Oedipodinae: Acrididae) Revisited: dispersion routes and origin of chromosomal inversion clines. *Journal of Orthoptera Research*, **19**(2), 253-260. https://doi.org/10.1665/034.019.0211

Guzmán, N. V., Pietrokovsky, S. M., Cigliano, M. M., & Confalonieri, V. A. (2017). Unraveling the diversification history of grasshoppers belonging to the "*Trimerotropis pallidipennis*" (Oedipodinae: Acrididae) species group: a hotspot of biodiversity in the Central Andes. *PeerJ,* **5***, e3835.* https://doi.org/10.7717/peerj.3835

Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspectives. In R. G. Harrison [ed.] *Hybrid zones and the evolutionary process*, (pp. 3-12). Oxford University Press, Oxford.

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology,* **25**(11), 2454–2466. https://doi.org/10.1111/mec.13582

Hewitt, G. H. (1979). Orthoptera. Grasshoppers and crickets. In John B. (ed.): *Animal Cytogenetics*, Vol. 3. Insecta 1. (pp. 1–170). Gebrüder Borntraeger, Berlin,

Hijmans, R. J. (2020). raster: Geographic data analysis and modeling. R package version 3.4-5. https://CRAN.R-project.org/package=raster

Hoffmann, A. A. & Rieseberg L.H. (2008).Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics,* **39**(1), 21-42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532

Hoffmann, A. A, Sgro, C. M., Weeks, A. R. (2004). Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology & Evolution*, **19**(9), 482–88. https://doi.org/10.1016/j.tree.2004.06.013.

Huang, K., Andrew, R.L., Owens, G.L., Ostevik, K.L. & Rieseberg, L.H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*. **29**(14), 2535– 2549. https://doi.org/10.1111/mec.15428

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., … Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius butterflies*. *Heredity*, **114**(5), 515–524. https://doi.org/10.1038/hdy.2015.22

Husemann, M., Guzmán, N.V., Danley, P.D., Cigliano, M.M. & Confalonieri, V.A. (2013). Biogeography of *Trimerotropis pallidipennis* (Acrididae: Oedipodinae): deep divergence across the Americas. *Journal of Biogeography,* **40**(2),261–273. https://doi.org/10.1111/jbi.12007

Jain, M., Olsen, H. E., Paten, B., & Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology*, **17**(1), 239. https://doi.org/10.1186/s13059-016-1103-0

James, A. C., Azevedo, R. B., & Partridge, L. (1997). Genetic and environmental responses to temperature of *Drosophila melanogaster* from a latitudinal cline. *Genetics*, 146(3), 881–890. https://doi.org/10.1093/genetics/146.3.881

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**(7392), 55–61. https://doi.org/10.1038/nature10944

Jombart, T. (2008). Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* **24**(11), 1403-1405. https://doi.org/10.1093/bioinformatics/btn129

Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee. S.F., Haag, C.R., … ffrench-Constant, R.H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature,* **477**, 203–206. http://sci-hub.tw/10.1038/nature10341

Jeffreys, H. (1961). *Theory of probability* (3rd ed.), Oxford University Press, New York

Kapun, M., Fabian, D. K., Goudet, J., & Flatt, T. (2016). Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. *Molecular Biology and Evolution,* **33**(5)*,* 1317–1336. https://doi.org/10.1093/molbev/msw016

Kapun, M., & Flatt, T. (2019). The adaptive significance of chromosomal inversion polymorphisms in *Drosophila melanogaster*. *Molecular Ecology,* **28**(6), 1263–1282. https://doi.org/10.1111/mec.14871

Kassambara, A. & Mundt, F. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics,* **28**(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Kemppainen, P., Knight, C. G., Sarma, D. K., Hlaing, T., Prakash, A., Maung Maung, Y. N., …& Walton, C. (2015). Linkage disequilibrium network analysis (LDna) gives a global view of chromosomal inversions, local adaptation and geographic structure. *Molecular ecology resources*, **15**(5), 1031–1045. https://doi.org/10.1111/1755-0998.12369

Kennington, W.J. & Hoffmann, A.A. (2013). Patterns of genetic variation across inversions: geographic variation in the In(2L)t inversion in populations of *Drosophila melanogaster* from eastern Australia. *BMC Evolutionary Biology*,**13**, 100. https://doi.org/10.1186/1471-2148-13-100.

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology,* **8**(9), e1000501. https://doi.org/10.1371/journal.pbio.1000501

Kirkpatrick, M. & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics,* **173**(1), 419–434. https://doi.org/10.1534/genetics.105.047985

Küpper, C., Stocks, M., Risse, J., dos Remedios, N., Farrell, L. L., McRae, S.B., … Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics* **48,** 79–83. https://doi.org/10.1038/ng.3443

Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C. G.,Schrider, D. R.,Pool, J. E.,...Begun, D.J.(2012). Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics,* **192** (2), 533–598, https://doi.org/10.1534/genetics.112.142018

Leigh, J. W., & Bryant, D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology &. Evolution*, **6**(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Lerman, D. N., Michalak, P., Helin, A. B., Bettencourt, B. R., & Feder, M. E. (2003) Modification of heat-shock gene expression in *Drosophila melanogaster* populations via transposable elements. *Molecular Biology and Evolution*, **20** (1), 135–144, https://doi.org/10.1093/molbev/msg015

Lesnoff, M., Lancelot, R. (2012). aod: Analysis of overdispersed data. R package version 1.3.1. http://cran.r-project.org/package=aod

Li, J., Z. Lu, Z., He, J. Chen, Q. ,Wang, X., Kang, L., & Li ,X.-d. (2016) Alternative exon-encoding regions of *Locusta migratoria* muscle myosin modulate the pH dependence of ATPase activity. *Insect Molecular Biology,* **25**(6), 689–700. https://doi.org/10.1111/imb.12254

Li, Z., Kemppainen, P., Rastas, P., Merilä; J. (2018). Linkage disequilibrium clustering-based approach for association mapping with tightly linked genome-wide data. *Molecular Ecology Resources*, **18**(4), 809-824. https://doi.org/10.1111/1755-0998.12893

Liang, H. F., Li, J., & Li, X. D. (2020). Identification and characterization of troponin genes in *Locusta migratoria*. *Insect molecular biology*, **29**(4), 391–403. https://doi.org/10.1111/imb.12644

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**(11), 1451–1452. https://doi.org/10.1093/bioinformatics/btp187.

Loewe, L., & Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **365**(1544), 1153–1167. https://doi.org/10.1098/rstb.2009.0317

Lotterhos, K. E., Yeaman, S., Degner, J., Aitken, S., & Hodgins, K. A. (2018). Modularity of genes involved in local adaptation to climate despite physical linkage. *Genome biology*, **19**(1), 157. https://doi.org/10.1186/s13059-018-1545-7

Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, **8**(9), e1000500. https://doi.org/10.1371/journal.pbio.1000500

Lucek, K., Gompert, Z., & Nosil, P. (2019). The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. *Molecular Ecology*, **28**(6),1224-1237. https://doi.org/10.1111/mec.15016

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., … Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution letters*, **1**(3), 155–168. https://doi.org/10.1002/evl3.15

Manoukis, N. C., Powell J. R., Touré M. B., Sacko A., Edillo F. E., Mamadou B, ... Besansky, N.J.(2008). A test of the chromosomal theory of ecotypic speciation in Anopheles gambiae. *Proceedings of the National Academy of Sciences*, **105** (8), 2940-2945. https://doi.org/10.1073/pnas.0709806105

Mathiopoulos, K. D., della Torre, A., Predazzi, V., Petrarca, V., & Coluzzi, M. (1998). Cloning of inversion breakpoints in the *Anopheles gambiae* complex traces a transposable element at the inversion junction. *Proceedings of the National Academy of Sciences of the United States of America*, **95**(21), 12444–12449. https://doi.org/10.1073/pnas.95.21.12444

Matrajt, M., Confalonieri, V. A., & Vilardi, J. C., (1996). Parallel adaptive patterns of allozyme and inversion polymorphisms on an ecological gradient. *Heredity,* **76**(4), 346–354. https://doi.org/10.1038/hdy.1996.52

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, **210**(2), 719–731. https://doi.org/10.1534/genetics.118.301336

McKinney, G., McPhee, M. V., Pascal, C., Seeb, J. E., & Seeb, L. W. (2020). Network analysis of linkage disequilibrium reveals genome architecture in Chum Salmon. *G3 Genes|Genomes|Genetics,* **10**(5), 1553–1561. https://doi.org/10.1534/g3.119.400972

Miller, D. E., Kahsai, L., Buddika, K., Dixon, M. J., Kim, B. Y. Calvi, B. R. … Cook, K. R. (2020) Identification and characterization of breakpoints and mutations on *Drosophila melanogaster* balancer chromosomes. *G3 Genes|Genomes|Genetics*, **10** (11), 4271–4285, https://doi.org/10.1534/g3.120.401559

Monroe,G., McKay, J. Weigel, D. & Flood P.J. (2021) The population genomics of adaptive loss of function. *Heredity*, **126**, 383–395. https://doi.org/10.1038/

Mullen, L.M. & Hoekstra, H.E. (2008). Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution,* **62**(7), 1555-1570. https://doi.org/10.1111/j.1558-5646.2008.00425.x

Nei, M. (1972). Genetic distance between populations. *The American Naturalist,* **106**(949), 283-292. http://www.jstor.org/stable/2459777

Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences,* **98***(21),* 12084–12088. doi:10.1073/pnas.221274498

Novembre, J., & Di Rienzo, A. (2009). Spatial patterns of variation due to natural selection in humans. *Nature reviews. Genetics*, **10**(11), 745–755. https://doi.org/10.1038/nrg2632

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources,* **13**(5), 946–952. https://doi.org/10.1111/1755-0998.12129

Pegueroles, C., Aquadro, C., Mestres, F. & Pascual, M.(2013). Gene flow and gene flux shape evolutionary patterns of variation in *Drosophila subobscura*. *Heredity,* **110,** 520–529. https://doi.org/10.1038/hdy.2012.118

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE,* **7**(5*), e37135*. https://doi.org/10.1371/journal.pone.0037135

Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics,* **155**(2), 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, … Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics,* **81**(3), 559–575. https://doi.org/10.1086/519795

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. *Current Biololgy*, ,**24**(22), 2728-2732. https://doi.org/10.1016/j.cub.2014.09.071

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Ramos-Onsins, S. E., & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution,* **19**(12)*,* 2092–2100. https://doi.org/10.1093/oxfordjournals.molbev.a004034

Rane, R.V., Rako, L., Kapun, M., Lee, S.F., Hoffmann, A.A. (2015). Genomic evidence for role of inversion 3RP of *Drosophila melanogaster* in facilitating climate change adaptation. *Molecular Ecology,* **24**(10)*,* 2423-2432. https://doi.org/10.1111/mec.13161

Ruiz-Ruano, F.J., Cabrero, J., López-León, M.D., Sánchez, A. & Camacho, J.P.M. (2018). Quantitative sequence characterization for repetitive DNA content in the supernumerary chromosome of the migratory locust. *Chromosoma,* **127**(1)**,** 45–57. https://doi.org/10.1007/s00412-017-0644-7

Schwander, T., Libbrecht, R. & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology.,* **24**(7), PR288-R294. https://doi.org/10.1016/j.cub.2014.01.056

Sharakhov, I.V., Serazin, A.C., Grushko, O.G., Dana, A., Lobo, N., Hillenmeyer, M.E., … Besansky, N.J.(2002). Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science,* **298**(5591),182-185. https://doi.org/10.1126/science.1076803.

Stevison, L. S., Hoehn, K. B., & Noor, M. A. (2011). Effects of inversions on within- and between-species recombination and divergence. *Genome biology and evolution*, **3**, 830–841. https://doi.org/10.1093/gbe/evr081

Sundaram, V. & Wysocka, J. (2020). Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society of London B Biological Sciences,* **375**(1795):20190347. https://doi.org/10.1098/rstb.2019.0347

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics,* **123**(3), 585-595. https://doi.org/10.1093/genetics/123.3.585

Tevy, F., Guzmán, N., Gonzalez, G., Lia, V., Poggio, L. & Confalonieri V. A. (2007). Mobile elements and inverted rearrangements in *Trimerotropis pallidipennis* (Orthoptera: Acrididae), *Caryologia*, **60** (3), 212-221. https://doi.org/10.1080/00087114.2007.10797939

Title, P. O., & Bemmels, J. B. (2018). ENVIREM: an expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. *Ecography,* **41**, 291–307. https://doi.org/10.1111/ecog.02880.

Thompson, M.J. & Jiggins, C.D. (2014). Supergenes and their role in evolution. *Heredity,* **113**(1),1-8. doi: 10.1038/hdy.2014.20.

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P.,… Balakrishnan, C. N. (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Current biology*, **26**(3), 344–350. https://doi.org/10.1016/j.cub.2015.11.069

Umina, P.A., Weeks, A.R., Kearney, M.R., McKechnie, S.W. & Hoffmann, A.A. (2005). A rapid shift in a classic clinal pattern in *Drosophila* reflecting climate change. *Science,* **308**(5722), 691-3. https://doi.org/10.1126/science.1109523.

Villoutreix, R., Ayala, D., Joron, M., Gompert, Z., Feder, J.L. & Nosil, P. (2021). Inversion breakpoints and the evolution of supergenes. *Molecular Ecology,* **30** (12), 2738-2755. https://doi.org/10.1111/mec.15907

Wade, C. M., Kulbokas, E. J., 3rd, Kirby, A. W., Zody, M. C., Mullikin, J. C., Lander, E. S., Lindblad-Toh, K., & Daly, M. J. (2002). The mosaic structure of variation in the laboratory mouse genome. *Nature*, 420(6915), 574–578. https://doi.org/10.1038/nature01252.

Wallberg, A., Schöning, C., Webster, M.T. & Hasselmann, M. (2017). Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genetics,* **13**(5), e1006792. https://doi.org/10.1371/journal.pgen.1006792

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.C., Shoemaker, D. & Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature,* **493,** 664–668. http://sci-hub.tw/10.1038/nature11832

Weir, B., & Cockerham, C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution,* **38**(6), 1358-1370. https://doi.org/10.2307/2408641

Westram, A.M., Faria, R., Johannesson, K. & Butlin, R. (2021) Using replicate hybrid zones to understand the genomic basis of adaptive divergence. *Molecular Ecology*, **30**(15), 3797-3814. https://doi.org/10.1111/mec.15861.

Wellenreuther, M. & Bernatchez. L. (2018). Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends in Ecology & Evolution,* **33**(6),427-440. https://doi.org/10.1016/j.tree.2018.04.002.

White, M. J. D. (1973). *Animal Cytology and Evolution*, 3rd Edition. Cambridge University Press, London

Whitelaw, E. & Martin, D.I. (2001) Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nature Genetics,* **27**(4),361-365. https://doi.org/10.1038/86850

Wickham, H. . (2007). Reshaping Data with the reshape Package. *Journal of Statistical Software*, **21**(12), 1–20. http://sci-hub.tw/10.18637/jss.v021.i12.

Wickham H. (2016) Programming with ggplot2. In: ggplot2. Use R!. Springer, Cham. http://sci-hub.tw/10.1007/978-3-319-24277-4_12

Wickham, H. (2019). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. https://CRAN.R-project.org/package=stringr

Data Accessibility Statement

Mitochondrial DNA sequences produced for this study are deposited in GenBank (Accession numbers MT037873 - MT038005), and Illumina reads in the NCBI Sequence Read Archive (BioProject accession number PRJNA733099).

SNP vcf files are available from Figshare Data Repository as:

Rad-loci_1: https://doi.org/10.6084/m9.figshare.17034323

Rad-loci_2 : https://doi.org/10.6084/m9.figshare.17035328

Rad loci_3: https://doi.org/10.6084/m9.figshare.16899985

Rad-loci_4 : https://doi.org/10.6084/m9.figshare.17035331

Cluster_517, Cluster_782, Cluster_ 845 and Cluster_ 632:

https://figshare.com/articles/dataset/VCF_Clusters_517-782-845-632/17035370/1.

Scripts and other data are available from Figshare Data Repository as:

https://figshare.com/articles/software/Scripts_Guzm_n_etal/17069483/1

https://doi.org/10.6084/m9.figshare.17069483

Author contributions

N.V.G. and V.A.C designed the project. E.R.D.C. and M.M.C. collected and identified sample materials. N.V.G. performed DNA extractions. N.V.G., E.R.D. C. and V.A.C performed cytological preparations and identified the karyotype of male individuals. N.V.G., P.K., V.A.C., D.M. and A.S.R. analyzed data. M.M.C and M.S.R. contributed substantially to the intellectual development of these projects and experimental design. V.A.C., N.V.G. and P.K. wrote the paper. All authors contributed to the editing and approved the final manuscript.

Table 1 Summary statistics for demographic changes of *Trimerotropis* sp. populations from Argentina based on a mitochondrial COI sequence. D_T (Tajima, 1989), Fu's *F*s test of neutrality (Fu, 1997) and R_2 (Ramos, Onsins and Rozas, 2002) are shown. Mendoza and San Luis: the analyses included populations from these provinces; Jujuy: the analysis included populations from Jujuy Province; All populations: refers to all populations sampled and described in Table S1. **p*<0.05; ****p*<0.01

	D_T Tajima	Fu's *Fs*	R_2
Mendoza and San Luis	-1,90916 *	-8,083***	0,0446 *
Jujuy	0,18829 NS	0,493 NS	0,1192 NS
All populations (except Jujuy)	-1,87829 *	-7,638 ***	0,0388 *
All populations	-0,99295 NS	-4,557 NS	0,0556 NS

Figure legends:

Figure 1. (a) Geographic sampling locations for populationsof *Trimerotropis* sp. in Argentina analyzed for mitochondrial COI and ddRADseq variation. Red bars indicate the altitudinal gradients from Mendoza –San Luis (MS) and Jujuy (Ju). Pie charts indicate the relative frequencies of each haplotype found in each population. Red circles indicate samples also analyzed for ddRADseq variation. (b) Haplotype network based on the COI gene, showing the relationships among haplotypes under the 95% criterion for a parsimonious connection among them, with colors matching those of Figure a. Each dash represents one mutational step. Circle size represents the proportion of the haplotypes. (c-f) Geographic sampling locations (1-30) for populations of *Trimerotropis* sp. in Argentina analyzed for chromosomal variation in the present paper (red circles) and in Confalonieri & Colombo (1989), Colombo & Confalonieri (1996) and Confalonieri (1994, 1995) (pink circles). Pie charts indicate the relative frequencies of each medium-sized (4, 6, 7 and 8) chromosome morphologies. A: standard (ancestral) acrocentric morphology, AI: Inverted acrocentric morphology, SM1-SM5 and SMa-SMc: different submetacentric morphologies and M: Metacentric morphology (for a description of these morphologies see Confalonieri & Colombo, 1989). 1: La Quiaca, 2: Corral Blanco, 3: Tres Cruces, 4: Uquía, 5: León, 6: Ampimpa, 7: Las Gredas, 8: Famatina, 9: Chilecito, 10: Nonogasta, 11: Vichigasta, 12: Catinzaco, 13: Puente del Inca, 14: Uspallata, 15: Cacheuta, 16: Mendoza, 17: El Chacay, 18: Chacras de Coria, 19: Maipú, 20: San Martín, 21: Chosmes, 22: Capilla del Monte, 23: Pescadores, 24: Balde, 25: San Carlos, 26: Cañon del Atuel, 27: Bariloche, 29: Laguna Blanca, 30: Chocón. Localities 6-12 correspond to La Rioja gradient.

Figure 2. GLMs predicted proportion of inversions of *Trimerotropis* sp. populations from Mendoza-San Luis altitudinal gradient. A. Inverted submetacentric morphology for medium chromosome 4 (4AI), predicted response variable (black line) with the model InvF ~ 7.24 - 4.37e-3*A, ~1 (φ = 7.84e-2); B. Inverted submetacentric morphology for medium chromosome 7 (7SM2), predicted response variable (black line) with the model InvF ~ 3.33 - 2.61e-3*A, explained deviance = 82.78%; C. Inverted submetacentric morphology for medium chromosome 8 (8SM3), predicted response variable (black line) with the model InvF ~ 2.28 - 1.52e-3*A, explained deviance = 80.47%; D. Inverted submetacentric morphology for medium chromosome 8 (8SM4), predicted response variable

(black line) with the model InvF ~ 0.31 - 1.29e-3*A, explained deviance = 57.49%; F. Inverted metacentric morphology for medium chromosome 6 (6M), predicted response variable for present data (black line) and data from 30 years ago (dashed black line) with the model InvF ~ 3.8 - 2.21e-3*A -1.07*t, explained deviance = 94.87%. Black circles: data from this study; grey triangles: data from Colombo & Confalonieri (1996). Altitudes are indicated in meters above sea level.

Figure 3. (a) Principal component analysis (PCA) based on *Rad-loci_2* dataset (Rad-loci_1 gave similar results). The first two principal components are shown, and colored dots refer to the Jujuy and Mendoza –San Luis gradients. Ellipses include individuals from the same population. (b) Heatmap of pairwise mean Weir and Cockerman (1984) F_{ST} (below the diagonal) ranging from low (light blue) to high (blue) and Nei´s (1972) genetic distances (upper the diagonal) ranging from low (white) to high (red), calculated for population from both gradients. (c) STRUCTURE (Pritchard et al., 2000) cluster analysis on *Trimerotropis* sp. populations based on *Rad-loci_2* dataset for K=4. The y-axis reports the probability of each individual (Q-value) assigned to one of the genetic groups identified, which are represented by different colors. Each bar corresponds to an individual. Individuals with 100% assignment to one group are identified by a single color, and those with mixed ancestry by different colors. Population and gradient limits are indicated with horizontal black lines below the plot.

Figure 4. (a) PCA analysis based on the genetic information from the four branches detected in LD-based single linkage clustering trees (LD-clusters 517, 782, 845 and 632) where the karyotypes explained the highest amount of the genetic variance for each of the four chromosomes (4, 6, 7 and 8, respectively). For these LD-clusters, karyotype explained 85%, 65%, 79% and 85% of the total variance in PC1 and PC2, respectively. PVE: Proportion of Variance Explained. The first two principal components are shown, and the colored symbols refer to the individual observed heterozygosity (H_o) values, ranging from high H_o (red colors) to low H_o (light-blue colors). The different symbols represent the karyotype of each individual (according to Table S1). Each karyotype in the PCAs is enclosed by a convex hull (hidden lines). (b-e) Heatmap of pairwise mean Weir and Cockerman (1984) F_{ST} values among populations from MS gradient using *Rad-loci*_3 matrix without outlier loci (upper the diagonal) and using SNPs from **(b)** Cluster 517, **(c)** Cluster 782, **(d)** Cluster 845, (e) Cluster 632 loci (below the diagonal) ranging from low (light-blue) to high (blue).

Figure 5. PCA based on **(a)** *Rad-loci_3* minus 151 SNPs belonging to clusters 517, 782, 845 and 632, **(b)** SNPs from cluster 517, (c) SNPs from cluster 782, (d) SNPs from cluster 845, (e) SNPs from cluster 632. The first two principal components are shown and the colored dots refer to the individual observed heterozygosity (H_o) values, ranging from high H_o (red colors) to low H_o (light-blue colors). Acronyms on each dot correspond to each sample ID according to Table S1.

Figure 6: Markers (SNPs) simultaneously associated with altitude and bioclimatic variables, for *Rad-loci_3* and *Rad-loci_4*. The line (marker) is the same in the different columns. Variables from WorldClim (Fick& Hijmans, 2017) - Bio6: Min Temperature of Coldest Month, Bio9: Mean Temperature of Driest Quarter, Bio12: Annual Precipitation. Variables from ENVIREM (Title & Bemmels, 2018)-Annual PET: Annual Potential EvapoTranspiration, Aridity: Thornthwaite aridity index, Moisture: Climatic Moisture Index, Continentality. Analyses were performed including populations of *Trimerotropis* sp from Jujuy and Mendoza-San Luis altitudinal gradients. This Figure was constructed based on Table S6 where the identification (ID) of markers is indicated in detail. One SNP was found in cluster 845 (green, solid arrow) and two SNPs in cluster 632 (red, dot line arrow).