Article

Phylogeographic and bioclimatic determinants of the dorsal pattern polymorphism in the Italian wall lizard, *Podarcis siculus*

Francesco Gallozzi[1,2], Paolo Colangelo[2], Gabriele Senczuk[3] and Riccardo Castiglia[1]*

1. Università degli studi di Roma "La Sapienza", Dipartimento di Biologia e Biotecnologie "Charles Darwin", via Borelli 50, 00188, Roma
2. National Research Council, Research Institute on Terrestrial Ecosystems, Via Salaria km 29.300, 00015, Montelibretti (Rome), Italy
3. Department of Agricultural, Environmental and Food Sciences, University of Molise, 86100, Campobasso, Italy

Abstract: The geographic variability of the dorsal pattern (DP) of the Italian wall lizard, *Podarcis siculus*, across its native range was studied with the aim to understand whether the distributions of this phenotypic trait were more shaped by allopatric differentiation rather than adaptive processes. A total of 1298 georeferenced observations scattered across the Italian peninsula and the main islands (Sicily, Corsica and Sardinia) were obtained from citizen science databases and five DPs were characterized by different shapes of the dark pattern ("*reticulated*", "*campestris*", "*reticulated/campestris*" and "*striped*") or by absence of it ("*concolor*"). Frequencies of different DP phenotypes differ between the two main mtDNA lineages settled in central-northern and in southern Italy respectively. This pattern may be indicative of a role of long-term allopatric historical processes in determining the observed pattern. The analysis also identified a putative wide area of secondary contact, in central southern Italy, characterized by high diversity of the DP. Generalized Linear Models (GLMs), used to estimate a possible association between bioclimatic variables and the observed phenotypic variation, showed that each of the five DPs is correlated to different environmental factors and show different distribution of areas with high probability of occurrence. However, for all but one of the DPs, the area with the greatest probability does not correspond exactly to the real distribution of the DP. Conversely, the "*concolor*" phenotype does not seem related to any particular mtDNA lineage and it shows a preference for areas with high temperature and low rainfall. This is in agreement with the expectation of low amount of melanin of the dorsal pattern that, in the study areas, is characterized by a light uniform coloration which could confer a better thermoregulation ability in high temperatures environments avoiding overheating.

Keywords: Citizen science; Colour pattern; geographic diversity; phylogeography

1. Introduction

Geographic phenotypic variation within a species can be due to natural selection driven by different factors as, for example, adaptive divergence to different environments or variation in sexual selection (Camargo et al. 2010; Zamudio et al. 2016). Alternatively, it may reflect a non-adaptive divergence of phenotypes in isolated lineages due to random drift or a combination of all these factors (Lande 1976; Barton 1996). In general, when genetic variation overlaps with the phenotypic variation in similar environmental conditions and mating systems, a dominant role of allopatry and isolation might be evoked. On the other hand, if the phenotypic variability agrees with the environmental variables, natural selection might be most involved (Zamudio et al. 2016). Within reptiles, the dorsal surface of lizard is frequently a polymorphic character, in color and pattern, and it is believed to evolve especially for adaptative purposes, probably to serve functions of crypsis and thermoregulation (e.g., Marshall et al. 2015; Smith et al. 2016; Moreno-Rueda et al.

2019). Even so, a series of studies linked lizard dorsal morphs with different aspects of escape behavior and reproductive involvement (Ortega et al. 2014; Ortega et al. 2015). Finally, DP polymorphism may also be related to sexual dimorphism thus it is largely driven by sexual selection as an outcome of male–male competition (Pérez i de Lanuza et al. 2013).

In natural sciences, the Citizen Science (Eitzel et al. 2017) is playing an ever-increasing role in scientific research due to the possibility to obtain large datasets of species occurrence records across large geographic scales, at a significantly lower cost when compared to traditional scientific surveys (Crall et al. 2010; Tulloch et al. 2013; Simoniello et al. 2019). These large datasets have been used successfully to model species distributions, abundances, (Devictor et al. 2010), phenology (Hurlbert & Liang 2012) and to quantify ecological interactions that affect animal fitness as predation and parasitism (Putman et al. 2021). Studies on the geographic phenotypic variation of populations are less common. In fact, these studies, despite their great potential, are applicable only to abundant species whose phenotypic characteristics under study are easily decipherable from photos (e.g., Silvertown et al. 2011; Kerstes et al. 2019; Drury et al. 2019; Chen et al. 2021). Only a small number of studies have dealt with vertebrates, but also in this case, it is possible to have results on the frequencies of the different phenotypes with geographical unprecedent details (Lehtinen et al. 2020).

The Italian wall lizard, *Podarcis siculus*, represents a good model for studying phenotypic diversity with a citizen science approach since it is not elusive, and very abundant and present in anthropized areas. This species is widespread in peninsular Italy, also occurring in Sicily, Sardinia and Corsica, and along the northern part of the eastern Adriatic coast. The species exhibits a remarkable variability of the dorsal pigmentation pattern (dorsal pattern, DP) (Arnold et al. 1978). The two most widespread phenotypes are the *"reticulated"* one, with a brown or black dorsal reticulation overlying a lighter base color which can range from light brown to green (Fig. 1e), and a phenotype (Fig. 1b) in which the dark reticulation leaves room for two lighter parietal lines that run throughout the body. The situation is made more complex by the presence of numerous intermediate phenotypes between the two and the absence of a specific pattern observable in the case of the so-called *"concolor"* (Fig. 1c) (Corti & Lo Cascio 1999). These DP phenotypes may reflect differences in the quantity or distribution of melanin on the back (Lepetz et al. 2009) and may have an impact on thermoregulation, for example with slower cooling rates, which may increase efficiency of thermoregulation in darker phenotypes (Sherbrooke et al. 1994; Gabirot et al. 2013).

The evolutionary history of *P. siculus* appears to have been strongly influenced by a series of historical fragmentation events due to severe cycles of glacial/interglacial periods during the Early-Middle Pleistocene with the onset of geographical barriers (Podnar et al. 2005; Senczuk et al. 2017). These events resulted in a complex phylogeographic structure characterized by strong mitochondrial discontinuities only partially in accordance with the nuclear signature (Senczuk et al. 2017; Senczuk et al. 2018). In particular, there are two main parapatric lineages, corresponding to the "Siculo-Calabrian" (SC) lineage and the "Central-Northern" (CN) lineage (Fig. 2). The boundary between these two lineages is settled in central Calabria, although there might be a sympatry area as shown in in Figure 2 (see also Senczuk et al., 2017). Within the Siculo-Calabrian lineage, three clades, two in Calabria (S1 and S2) and one in Sicily (S3), were identified, while the central-northern lineage splits into two main clades, the 'Tyrrhenian' (T) and the 'Adriatic' (A) parapatric clades. The latter can be divided into two subclades (A2 and A3). The T-clade is mainly distributed across the northern-central Tyrrhenian coast, while the A-clade is found along the Adriatic coast and in Croatia and extending in the Tyrrhenian coast in southern Italy (the distribution of mtDNA lineages and clades in Italy is indicated with their phylogeny in figure 2).

In this study, photographs obtained from citizen science databases were used to analyze the overall geographic variability of the DPs in *P. siculus.* In order to better understand whether these phenotypic traits were more shaped by allopatric differentiation rather than adaptive processes, phenotypic frequencies were tested in relation to mtDNA clades. A correspondence between DP and mtDNA lineages may be indicative of an origin of the phenotypic variant following the allopatric events. Subsequently, Generalized Linear Models (GLM) were used to estimate the possible link between bioclimatic variables and the observed phenotypic variation for each DP pattern.

Thus, the GLMs were then used to predict the occurrence probability in the range of *P. siculus*: if GLM-based predictions result as reliable, this may indicate that each phenotype is adapted to different climatic conditions.

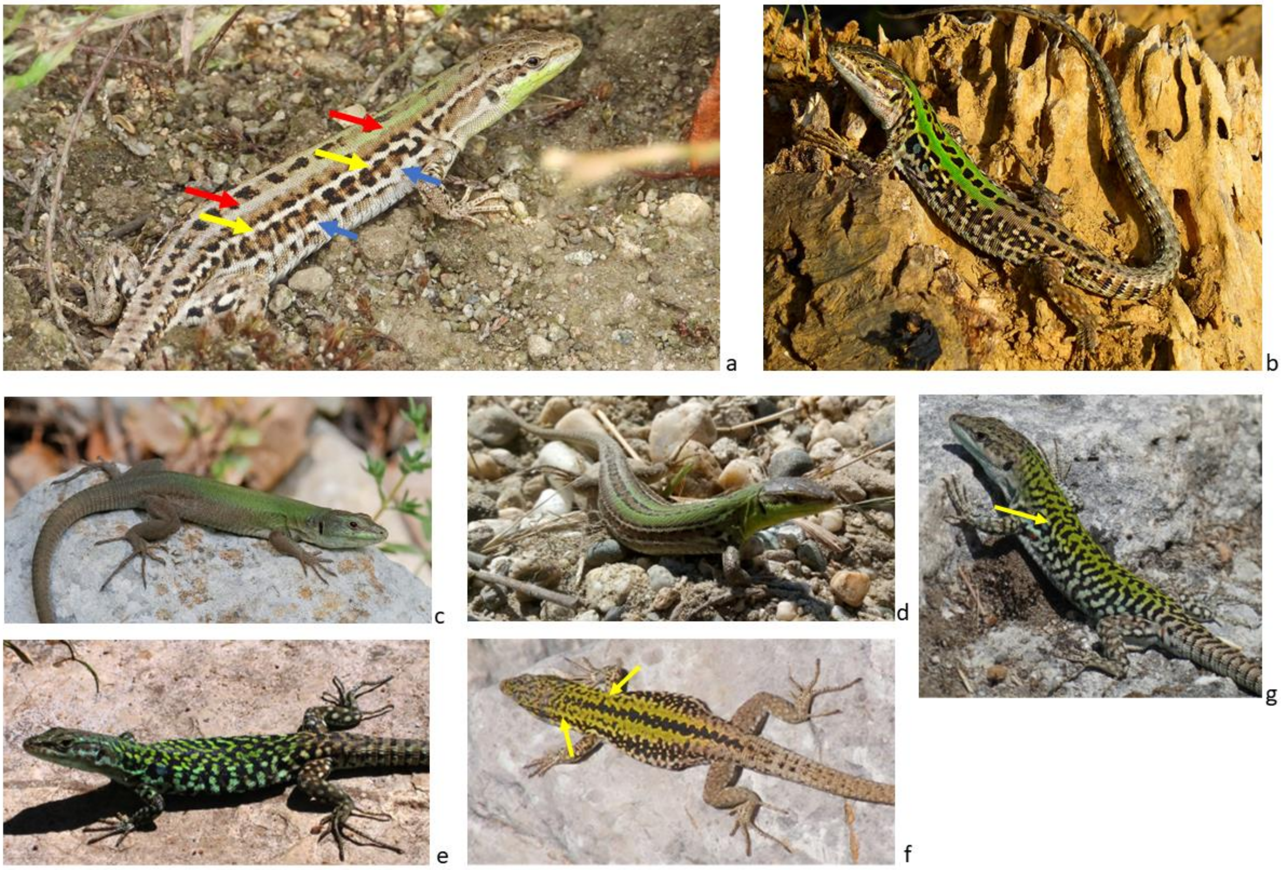


Figure 1 (a) Position of the characters used to define the dorsal pattern phenotypes; parietal, supraciliar, and lateral stripes are indicated by red, blue, and yellow arrows respectively (observation by Mirko Tomasi https://www.inaturalist.org/observations/51132755). (b) “campestris” phenotype (observation by Renato Franzi, https://www.inaturalist.org/observations/42845479). (c) “Concolor” phenotype (observation by Peter Kennerley, https://www.inaturalist.org/observations/76986247). (d) “Striped” phenotype (observation by Daniele Seglie, https://www.inaturalist.org/observations/12779595), (e) “Reticulated” phenotype (observation by Riccardo Castiglia, https://www.inaturalist.org/observations/42465364). The “reticulated/campestris” phenotypes, (f) and (g), (observation by Giorgio De Simone, https://www.inaturalist.org/observations/48902707 and "martind" https://www.inaturalist.org/observations/28726011); the interrupted parietal stripe can be noted in both specimens (yellow arrows, see text for explanations).

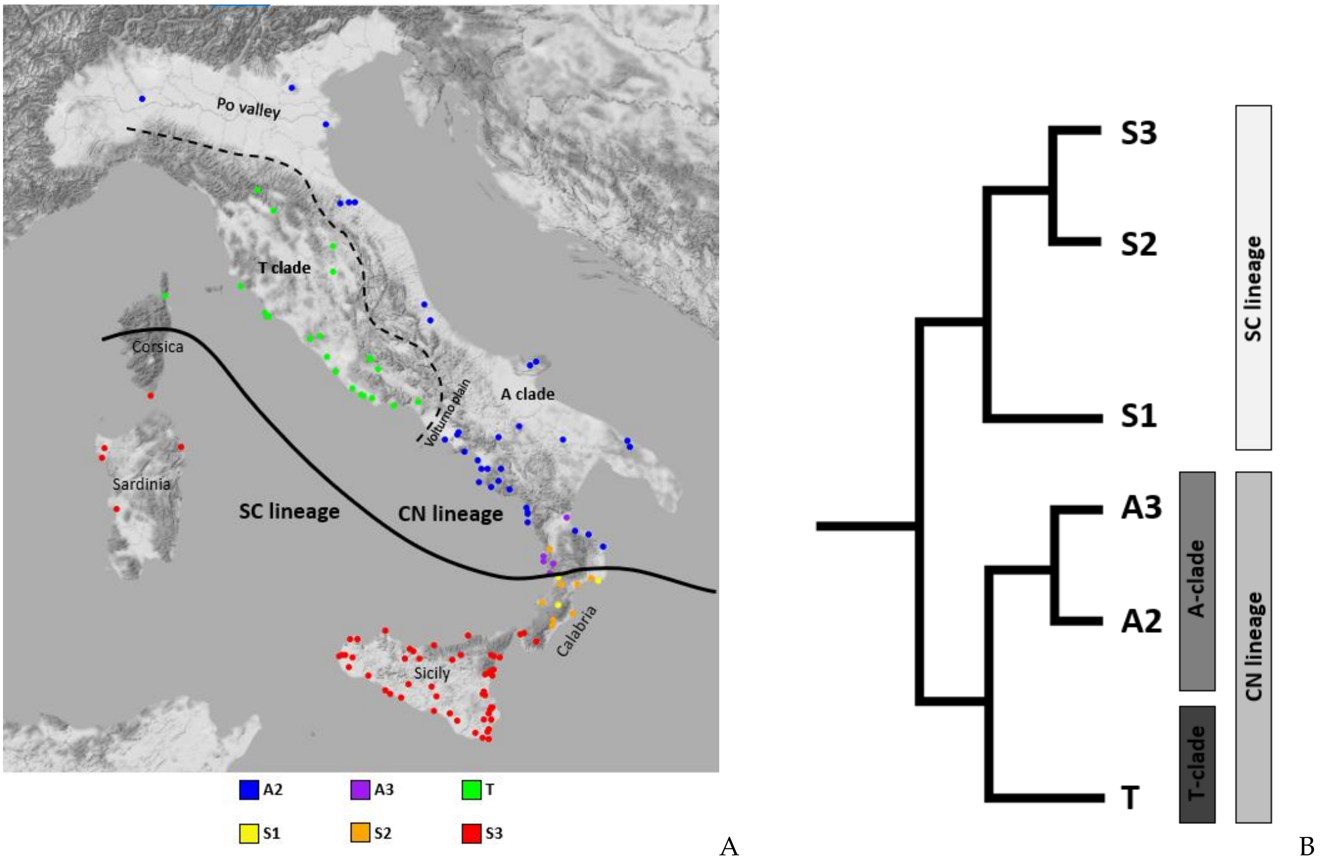


Figure 2. The geographic distribution of the main clades (A) and their phylogeny (B). Each dot represents a genotyped sample. The map and the phylogenetic tree are redrawn from Senczuk et al., 2017. In the map, the geographical region cited in the text are shown. The boundaries of the distribution of the different lines and clades are also indicated.

2. Material and Method

2.1. *Selection of observations*

Photo-vouchered observations of the Italian Wall Lizard (*P. siculus*) were sourced from iNaturalist (www.inaturalist.org). Over 4000 observations of this species had been uploaded online as of 1st January 2021 and they have been all checked. We only included in the analysis georeferenced observations of adult specimens with accuracy less than 5 km and in which the DPs (see below) of photographed animals were clearly identifiable. For observations less than 100 m apart, we visually compared the size, colour, and dorsal patterns of the lizards in the images to eliminate potential duplicates of the same individual. After pruning not usable observations, a total of 1298 observations were included in this study. Because for most individuals (60%) we were not able to determine the sex, we decided to not consider it in the analysis (see Discussion).

2.2. *Character and phenotype*

Raw data used in this study are available in supplementary materials (Table S1), including the iNaturalist ID number of the observations and geographical coordinates. Six characters of the dorsal pattern were included in our study (Fig. 1). For each observation, presence/absence of each character was evaluated and a presence/absence matrix was created (presence = 1, absence = 0). The six characters considered in this study include: 1) presence of anterior parietal bands (a double green or brownish dorsal stripe stretching dorsally from the parietal scales to the midbody); 2) presence of a posterior parietal band (a double green or brownish dorsal stripe stretching dorsally from the midbody to the tail); 3) presence of anterior supraciliar stripe (a usually light stripe stretching dorsolaterally from the supraciliary scales to the midbody); 4) presence of posterior supraciliar stripe (a usually light stripe stretching dorsolaterally from the midbody to the tail); 5) presence of a lateral stripe (a usually light lateral stripe from the forelimb to the hindlimb); 6) presence of a dorsal reticulation (a dorsal dark reticulum that interrupts a stripe or a band in at least one point). Bands and stripes were considered as present only when they were continuous and showed no interruption due to the darker dorsal pattern (Fig. 1b, d).

Following the identification of the 6 characters, each individual was assigned to the following five phenotypic categories according to presence/absence combinations of the characters: *"campestris"* (presence of anterior and posterior parietal bands only, absence of other characters), *"reticulated"* (presence of complete dorsal reticulation, absence of other characters); *"striped"* (complete parietal bands with at least one between supraciliar or lateral stripe, absence of other characters); *"concolor"* (0 for all characters, i.e., absence of any dark pattern or stripe, uniform coloration). Although some individuals (n = 47; Fig. 1f, g) showed parietal bands, they were very different from classical *"campestris"*, since their bands presented jagged edges or were interrupted in at least one point in the anterior or posterior portion of the body. These individuals were classified as "*reticulated/campestris*" (Fig. 1f, g).

2.3. Distribution of dorsal pattern phenotypes and mtDNA lineages

Dorsal pattern phenotype distributions were plotted on a map using *ggmap* package in R (Kahle and Wickham, 2013). The relative frequencies of the different DP phenotypes were estimated for twenty-two groups of nearby observation. Groups of populations were as follows: three groups for Sicily, two groups for Calabria, Puglia, Sardinia and Corsica, one group for each of the remaining Italian regions. The distribution of mtDNA clades and lineages has been extrapolated by data points extracted by Senczuk et al. (2017, 2018) with a minimum convex polygon approach and according to the orographic characteristics of the study area (Fig. 2). Therefore, according to their geographic origin, each observed individual was assigned to a lineage/clade. This method leads to a certain approximation in the correspondence between the observation on iNaturalist and their assignment to a specific lineage/clade. However, the lineages/clades have a very wide distribution and are parapatric/allopatric with no case of co-occurrence of lineages in the same population except for an isolated individual in Calabria (Fig. 2 and Senczuk et al. 2017). Furthermore, we avoided comparing the frequencies of the DPs between the clades in Calabria as the available data cannot precisely identify their boundaries. For this reason, we were confident that, at the geographic scale considered in this work, the approximation in the assignment of the observations to the various lineages/clades does not significantly affect the results. To test differences in phenotypic frequencies in relation to mtDNA lineages/clades, the χ^2 test was conducted to compare frequencies of different phenotypes between the two main lineages (CN and SC) and between the A and T clades. Moreover, we compared DP frequencies of pooled specimens of the S1 and S2 clades with specimens from the Sicilian S3-clade. Since we found two haplotypes belonging to the S3-clade in the very southern portion of Calabria, DP data from this area were pooled with other Calabrian samples.

2.4. GLM modelling

To model the probability of encountering a phenotype across the range of the species, a set of GLMs (one for each phenotype) was estimated considering phenotypes' presence/absence as function of environmental heterogeneity. To describe environmental heterogeneity, we selected nine bioclimatic variables across the whole range of the species with a 0.86 km^2 resolution (WorldClim database: https://www.worldclim.org) among those considered as most affecting for the presence of the *P. siculus* and other related species and that are not highly autocorrelated (Pearson correlation coefficient r < 0.75) according to previous studies (Senczuk et al. 2017; Senczuk, Harris et al. 2019). Selected variables were as follows: mean diurnal range (BIO2), isothermality (BIO3), temperature seasonality (BIO4), minimum temperature of the coldest month (BIO6), temperature annual range (BIO7), mean temperature of the driest quarter (BIO9), annual precipitation (BIO12), precipitation seasonality (BIO15), precipitation of the warmest quarter (BIO18). We performed a raster PCA using the *rasterPCA* function implemented in the *raster* package in R (Hijmans and van Etten, 2012) on the selected bioclimatic variables. This function performs a PCA directly on raster data and we used principal components (PCs) as proxies for environmental heterogeneity. We then considered the highest loadings values to interpret PCs. This kind of approach was already successfully used to reduce the dimensionality of bioclimatic data in several studies (Avramo et al., 2021; Yoon & Lee, 2021; Xie et al., 2021; Gusmão et al., 2021). Based on Kaiser-Guttmann criterion, PCs with eigenvalues greater than 1, PC1 and PC2 in our case, should be retained. A widely used methods is to interpret only those components that contribute more than 5% of the total variance (Janžekovič and Novak, 2012) and PC1 to PC4 meet this criterion. Since we wanted to select PCs that could explain more than 90% of the variations in the bioclimatic variables, we retained all the four of them to get 94.79% of the variation. Even if only PC1 and PC2 have eigenvalues greater than 1, they explain 80.94% of total variation. PCA loadings, eigenvalues and importance of PCs are shown in supplementary materials (Table S3 and S4). PCs scores were extracted for each observation using the *extract* function of *raster* package in R (Hijmans and van Etten, 2012). Full GLMs were estimated for each phenotype using a binomial approach and considering the phenotype presence/absence as response variables and the first four PCs as predictors. All the models were compared to null models in which a constant (x~1) was fitted to evaluate random effect. For model estimation we used the *glm* function in the *stats* package in R (R Core Team, 2013). Then, we compared full and null models through AIC to evaluate the actual effect of environmental heterogeneity on phenotypic frequencies. The probability of encountering a phenotype was then modelled by using the *predict* function of *stats* package in R (R Core Team, 2013) and we tested the actual prediction performances of our models by an area under the ROC curve (AUC) (Merow et al., 2013) with *pROC* package in R (Robin et al., 2011). Our predictions were plotted with *ggmap* package in R and we obtained occurrence probability maps for each phenotype.

3. Results

3.1. Phenotype geographical distribution and χ^2 test

Among 1298 individuals included in this study 817 were categorized as "*campestris*", 39 as "*concolor*", 130 as "*striped*", 265 as "*reticulated*" and 47 as "*reticulated/campestris*" (see table S1 and Fig. S1 for details). In Fig. 3, the frequencies of the different phenotypes on selected groups of populations are indicated on map. The "*campestris*" phenotype is widely distributed all along continental Italy and Northern Corsica. This phenotype is absent in Sicily. The "*reticulated*" phenotype is mainly distributed in Calabria and in other Southern Italian areas (Fig. 3); it is also common in Sicily, Sardinia and in Southern Corsica. The distribution of the "*striped*" phenotype matches with that of "*campestris*". However, its frequency is considerably lower in all the southern part of the peninsular range corresponding to Calabria region. The intermediate "*reticulated/campestris*" phenotype is at low frequency respect to the other phenotypes and its distribution overlaps the one of the "*reticulated*" phenotype. The "*concolor*" phenotype is distributed at a low frequency in different parts of the species range, especially in the southern region. It seems more frequent in one central Italian region (Campania) and very abundant in South-Eastern Sicily, as it is the most common phenotype and the 20% of observations of this phenotype are found in this area.

The χ^2 test confirmed the presence of statistically significant differences in the frequencies of various DPs between CN and SC lineages. "*Campestris*" and "*striped*" are more frequent in the CN lineage, whereas "*reticulated*" and "*reticulated/campestris*" are more frequent within the SC lineage and "*concolor*" showed no association with any lineage (figure 4). Statistically significant differences were also detected between A and T clades, due to the higher frequencies of reticulated individuals in the latter, and between Calabria (S1+S2) and Sicily (S3).

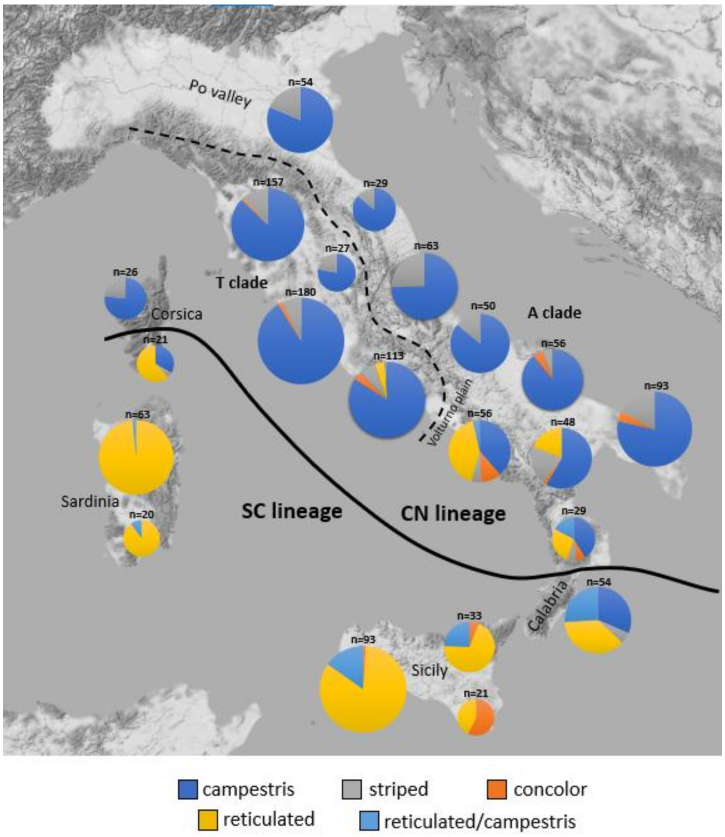
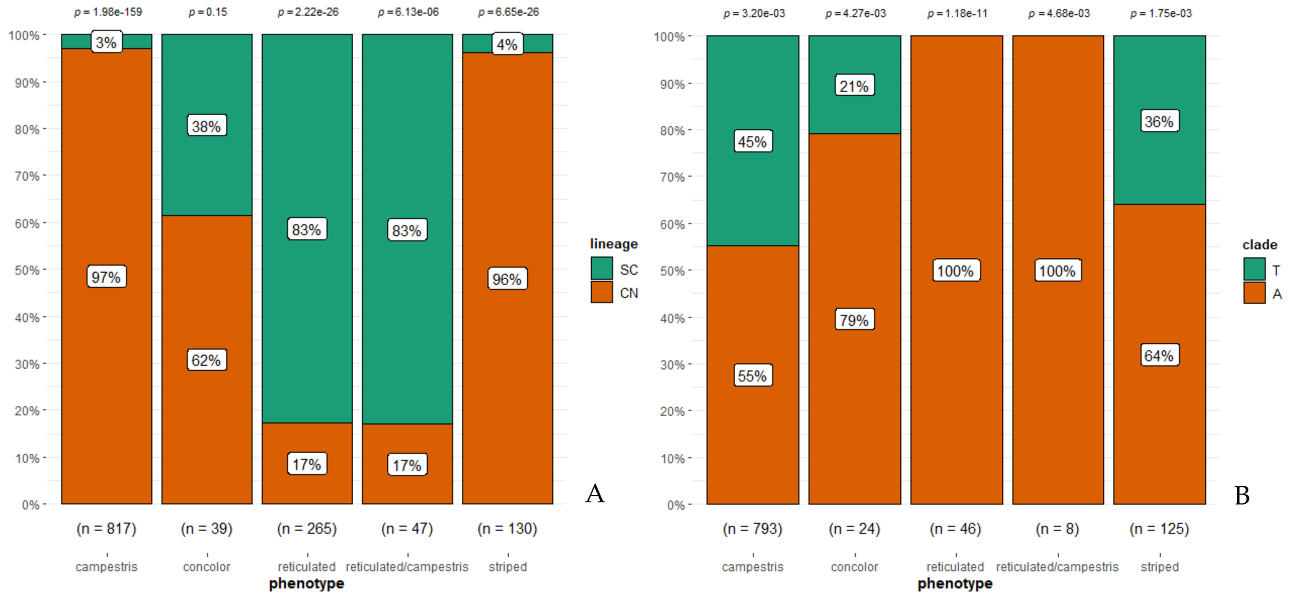


Figure 3. Geographical distribution of phenotypic frequencies and mtDNA lineages/clades of *Podarcis siculus* in Italy.



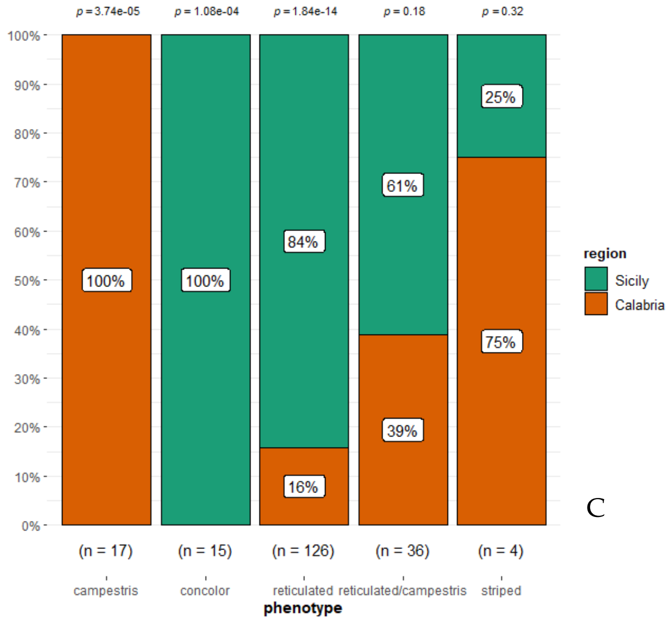


Figure 4. χ^2 test results. Bar charts indicate the percentage of association of each phenotype with lineages (A), clades (B) or regions (C).

3.2. Generalized Linear Models

The AIC values for GLMs of each phenotype compared to the respective null model AIC are shown in Table 1. Every model shows always a better (lower) AIC value respect to the null models. Therefore, b heterogeneity is a better explainer of the observed phenotypic distributions rather than a random effect. GLM-based predictions have an AUC > 0.7 for every phenotype except for *"striped"* (AUC=0.660). In particular, prediction performances are excellent for *"reticulated"* (AUC=0.934), good for *"campestris"* and *"reticulated/campestris"* (AUC > 0.8) and moderately good for *"concolor"* (AUC=0.713).

	AIC		
Phenotype	**Null**	**Env. Het.**	**AUC**
campestris	1713.4	1251.5	0.843
concolor	352.21	317.83	0.713
striped	846.79	817.84	0.660
reticulated	1315.9	694.46	0.934
reticulated/campestris	406.21	340.37	0.841

Table 1. AIC values for null and environmental heterogeneity-based models and AUC values for each phenotype.

The "*campestris*" phenotype shows a positive correlation with PC1 (temperature seasonality, minimum temperature of the coldest month - negative loading), precipitation of the warmest quarter) and PC2 (mean diurnal range, isothermality) and a negative relation with PC3 (annual precipitation - negative loading) and PC4 (mean temperature of the driest quarter - negative loading). All these relations are statistically supported (p-values < 0.05, Table 2). Therefore, this phenotype appears to be associated to lower temperatures, higher precipitations and high seasonality. According to our predictive models based on environmental heterogeneity, the highest probability to find "*campestris*" individuals is in mountain areas of continental Italy with the exception of the highest altitude mountains, and in the Po valley at the northern part of the species range. Conversely, coastal areas, Sardinia, Corsica and southern region show lower probability values (Figure 5).

The "*concolor*" phenotype shows statistically supported relations (positive correlation) with PC4 only (mean temperature of the driest quarter - negative loading) (Table 2). The GLM predicts "*concolor*" individuals to be very uncommon in all *P. siculus* distribution range, except for south-eastern Sicily showing the highest probability, west Sardinia and a few more coastal areas (Figure 5).

The "*striped*" phenotype shows statistically supported relations with PC1 and PC4 (Table 2). There is a positive correlation with PC1 (temperature seasonality, minimum temperature of the coldest month - negative loading; precipitation of the warmest quarter) and a negative correlation with PC4 (mean temperature of the driest quarter - negative loading). As for *"campestris"*, this may represent an association to lower temperatures, higher precipitations and high seasonality. This phenotype is predicted to be mainly distributed in continental Italy but with lower probabilities values than *"campestris"* (Fig. 5).

The "*reticulated*" phenotype shows a positive correlation with PC1 (temperature seasonality, minimum temperature of the coldest month - negative loading; precipitation of the warmest quarter) and PC3 (annual precipitation - negative loading) and a negative relation with PC2 (mean diurnal range, isothermality) and PC4 (mean temperature of the driest quarter - negative loading) (Table 2). All these relations seem to be statistically supported, except for PC3. Therefore, the *"reticulated"* phenotype appears to be associated to higher temperatures, low precipitations and seasonality and small temperature ranges. According to GLM predictions, the highest probability to find "*reticulated*" individuals is located in Sicily, Sardinia, Corsica and in some scattered areas in southern Italy and in the Apennines. Other areas with moderate probabilities are found along other coastal areas and in all Calabria (Fig. 5).

The "*reticulated/campestris*" appears to be positively correlated with PC4 (mean temperature of the driest quarter - negative loading) and negatively correlated with PC1 (temperature seasonality, minimum temperature of the coldest month - negative loading; precipitation of the warmest quarter), PC2 (mean diurnal range, isothermality) and PC3 (annual precipitation - negative loading) but only the relationships with PC1, PC2 and PC4 are statistically supported (Table 2). As for "reticulated", it seems to be associated with narrow temperature ranges and low precipitations. According to GLM the probability to find this phenotype is higher in Southern Italy, specifically in some areas of Sicily, Sardinia, Corsica, Calabria and central Apennines (Fig. 5).

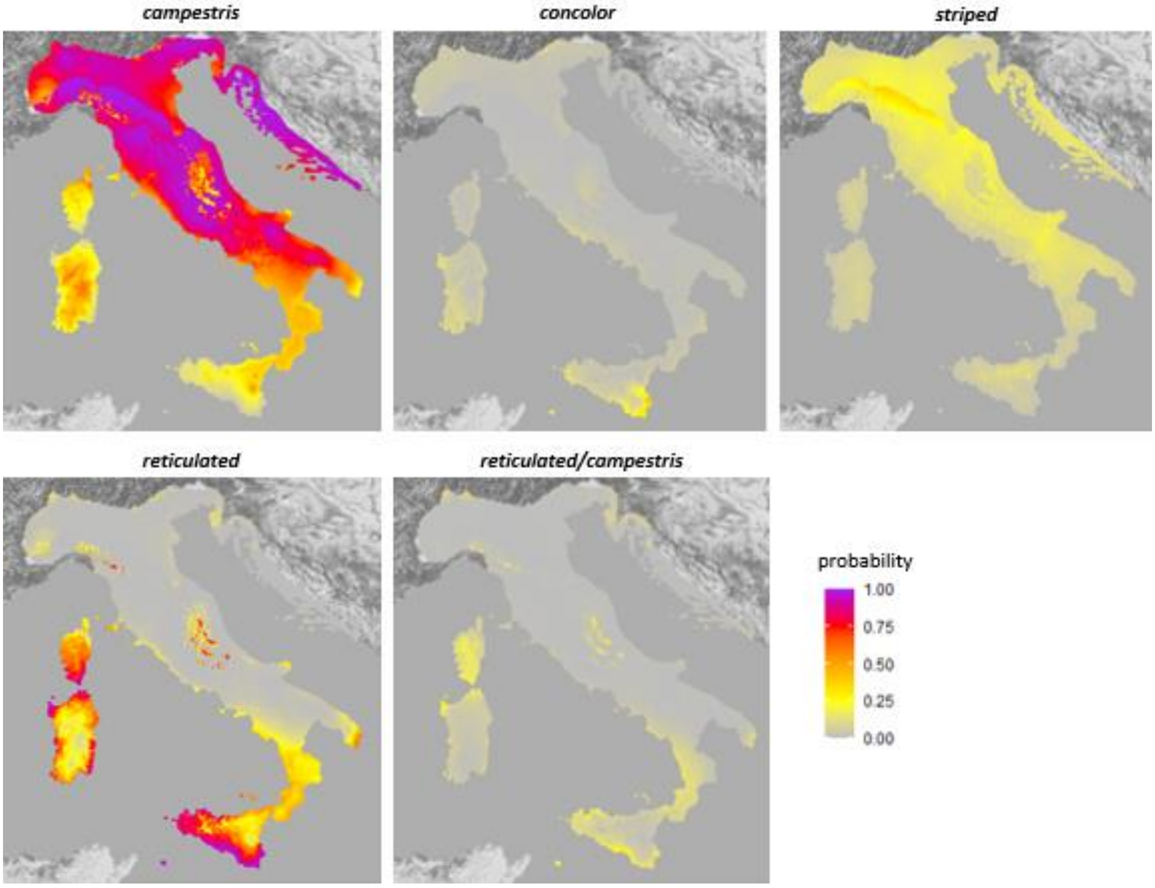


Figure 5. Maps of predicted probabilities to find each phenotype according to GLM estimation.

		Estimate	**Std. Error**	**Z value**	***p-value***
campestris	**Intercept**	1.20836	0.12066	10.015	< 0.001
	PC1	0.49878	0.06012	8.297	< 0.001
	PC2	0.43560	0.05718	7.617	< 0.001
	PC3	-0.19821	0.09257	-2.141	< 0.05
	PC4	-1.35633	0.20879	-6.496	< 0.001
concolor	**Intercept**	-4.38803	0.42049	-10.436	< 0.001
	PC1	-0.18869	0.18397	-1.026	> 0.05
	PC2	0.20144	0.14434	1.396	> 0.05
	PC3	-0.05147	0.24937	-0.206	> 0.05
	PC4	1.66756	0.55162	3.023	< 0.01
striped	**Intercept**	-2.025614	0.142466	-14.218	< 0.001
	PC1	0.174922	0.075795	2.308	< 0.05

	PC2	-0.001489	0.079173	-0.019	> 0.05
	PC3	0.243598	0.144348	1.688	> 0.05
	PC4	-0.500197	0.237406	-2.107	< 0.05
reticulated	**Intercept**	-3.65142	0.28494	-12.815	< 0.001
	PC1	-0.82899	0.12458	-6.654	< 0.001
	PC2	-0.76792	0.09469	-8.110	< 0.001
	PC3	-0.11727	0.13413	-0.874	> 0.05
	PC4	2.49372	0.42512	5.866	< 0.001
reticulated/campestris	**Intercept**	-4.7651	0.5621	-8.478	< 0.001
	PC1	-0.4345	0.2523	-1.722	> 0.05
	PC2	-0.5734	0.1727	-3.321	< 0.001
	PC3	-0.7313	0.2540	-2.879	< 0.001
	PC4	1.1870	0.7948	1.493	> 0.05

Table 2. GLM Estimated regression coefficients, standard errors, z-values and p-values for each variable for each phenotype.

4. Discussion

Among the genus *Podarcis*, the Italian wall lizard *P. siculus* shows one of the greatest genetic and phenotypic diversity, so much that a species complex has been argued (Speybroeck et al., 2020; Senczuk et al., 2017). In this paper, frequencies and distributions of different DPs have been investigated in order to tentatively assess the relative role of historical (allopatric) and/or adaptive processes in driving their distributions. Other studies have investigated some characteristics of the dorsal surface of *P. siculus* in relation to environmental variables (e.g., Pellitteri-Rosa et al. 2020; Storniolo et al. 2021), but none in relation to the global genetic structure of the species. Here, data from a Citizen Science database (inaturalist.org) were obtained allowing to study the DPs distribution with an unprecedent level of detail.

Our results showed that the DP phenotypes, except the "*concolor*", are significantly associated with the two main mitochondrial lineages (CN and SC lineages, Senczuk et al., 2017). Indeed, while the "*campestris*" and the "*striped*" phenotypes are mainly distributed within the CN lineage, the "*reticulated*" and the "*reticulated /campestris*" DP are mainly distributed within the range of the SC lineage (Fig. 2 and 3).

Despite this correspondence, a certain discrepancy between distribution of DPs and mitochondrial data can be noted. In fact, while the contact zone of the two mtDNA lineages is restricted to an area of about 50 km in central Calabria, where both lineages can be found (Fig. 2), there is a mixture of DP phenotypes characterizing both lineages ("*reticulated/campestris*", "*campestris*", "*striped*" and "*reticulated*") spanning from southern Calabria reaching the Volturno Plain (Campania) that is approximately 350 km northern the border of the two mtDNA lineages. This area was repeatedly flooded by Middle Pleistocene marine transgressions, thus constituting an effective geographic barrier between previously separated populations (Romano et al. 1994, Barra er al. 1996) and represents a contact zone for many taxa (Canestrelli et al. 2008; Bisconti et al. 2018; Chiocchio et al.

2021). The overall pattern is congruent with a wide introgression area with ongoing gene flow between CN and SC lineages spanning from southern Calabria to Volturno plain and this is also supported by microsatellites analyses (Senczuk et al., unpublished data).

A similar pattern with different extent of cline variation between genetic and phenotypic traits has also been found in northern Italy in the common wall lizard (Michaelides et al., 2015; Yang et al., 2018; 2020). In this case, differences in phenotypic traits seem to have triggered the asymmetric introgression via male-male competition, therefore a similar mechanism to explain our phenotype/genetic discordance cannot be excluded and future studies should be addressed in such a direction.

The match between phenotypes and phylogeographic structure is also observable within the SC lineage, where differences in the frequency of DP phenotypes between Sicily (S3-clade) and Calabria (S1 + S2 clades) were found. These differences are due to the absence of "*campestris*" DP phenotypes in Sicily and to the high frequency of "*concolor*" phenotype in the south-eastern part of the island. This distinction in DP frequencies between Sicily and mainland Italy likely corresponds to the ancient origin of the Italian wall lizard in the island. In fact, the so-called Sicilian clade (Senczuk et al., 2017) separated from the mainland clades around 1.54 Mya and it is almost exclusive to Sicily with only two haplotypes found in the most southern part of Calabria. Although in this work a very wide geographical scale has been considered, it cannot be excluded that similar differences may also be found on a regional scale. For example, there are 7 partially geographically separated haplogroups in Sicily (Senczuk et al., 2017) that can be related to other characteristics of the DP.

The dorsal coloration of *P. siculus* is known to be sexually dimorphic with males being more conspicuous than females (Storniolo et al. 2021) and this observation also is extended to the DP pattern here considered (Arnold and Burton 1985, Corti et al. 2010, Corti and Lo Cascio 1999). Specifically, within the CN lineage, the "*striped*" phenotype is known to be characteristic of females while the "*campestris*" is often found in males. However, this is not a rule and it became clear to us from the photos on iNaturalist that many of the "*striped*" phenotypes are males and the "*campestris*" are females (data not shown). As regards the other common phenotypes ("*reticulated*" and "*reticulated/campestris*") it does not seem to us that they are related to sex but we have not analyzed in detail the issue. Not having considered the sex of the individuals, due to the Citizen science approach, can lead to erroneous estimates of morph frequencies if the data are sex-biased. Such biases can arise if there are more pictures of one sex than the other in the iNaturalist database or if the adult sex ratio vary among different populations. In *P. siculus*, the males are probably the most photographed as they have larger dimensions and more vivid colours. However, it seems very unlikely that this phenomenon varies with geographical areas as the greater conspicuousness of males seems to be a constant feature throughout the species range (Di Matteo et al. 2019). Concerning the different sex ratio in different populations, the issue is difficult to assess because there are very scant data, in literature, on sex ratio of mainland populations (Zuffi et al. 2012). However, in other *Podarcis* species the observed geographical variation in adult sex ratio among mainland populations is of negligible magnitude (or absent) (Dietvorst et al. 1980; Chondropoulos and Lykakis 1983).

The results of GLMs show that each of the five DP patterns is correlated to different bioclimatic factors. Thus, while "*campestris*" and "*striped*" DPs show a greater suitability in elevated areas with low temperature and high precipitation, "*reticulated*" and "*reticulated/campestris*" have their greatest suitability areas along coastal areas of southern Italy, characterized by high temperatures and low precipitations. However, for "*striped*" DP the predictions showed lower AUC and ΔAIC between the null and the environmental heterogeneity-based model. This suggests that the distribution of this DP phenotype is mainly shaped by the allopatric history of the genetic lineages rather than by bioclimatic factors.

It may also be noted that some of the areas with the greatest bioclimatic suitability for every phenotype, especially "*campestris*" and "*reticulated*", does not correspond exactly to the actual distribution of these DPs. In fact, for these phenotypes there are highly suitable areas where they are absent/sporadic and non-suitable areas where they are present at high frequency (see supplementary materials Fig. S1). Furthermore, given the correspondence between mtDNA and phenotypes, it is difficult to establish whether the environmental preferences highlighted by the GLMs are linked to have a specific DP pattern or to belong to a clade and then having any other clade-related phenotypic trait that we did not considered in this paper. In the meantime, an adaptive component of the dorsal pattern cannot be excluded since it also possible that a recent adaptation of the DP patterns to the bioclimatic condition occurred in allopatry (Diaz et al. 2017).

In this context, an exception is the *"concolor"* phenotype that it is the only phenotype whose distribution better corresponds to what is indicated by the GLM model. The "*concolor*" is an "extreme" phenotype characterized by uniform coloration, without any dorsal pattern and, although it can be very dark on some islands, in the study area the amount of melanin on the back is less than in the other DPs (Fig. 1). The GLM results for "*concolor*" indicate a preference for areas with high temperature and low rainfall. Based on these observations, it can be reasonably assumed that "*concolor*" is a positively selected phenotype in areas with these particular climatic conditions. This could be due to its low amount of melanin, which could confer a better thermoregulation ability in high temperature environments avoiding overheating (Sherbrooke et al. 1994; Gabirot et al. 2013).

It should be pointed out that we have considered only some characters of the dorsal pattern, those that are commonly used to distinguish the different "morphs" of *P. siculus*. It is rather clear now that other characteristics of the dorsal surface are more closely related to environmental variables. In fact, it is already known that *P. siculus* is capable of altering dorsal coloration, to improve crypsis during seasonal change varying from green at the onset of spring, to brownish in the middle of summer and to a greyish colour in October (Pellitteri-Rosa et al. 2020). Furthermore, we noticed in our dataset that the extension of the reticulation pattern in terms of amplitude of the dark area on the dorsal surface is variable and could be related to climate as well as to particular substrates (Miñano et al. 2021).

To conclude, the association between the two main mtDNA lineages and DP phenotypes suggests a role of long-term allopatric historical processes in molding the characteristics of the DPs. An adaptative divergence is particularly evident for the "*concolor*" phenotype that is not related to any mtDNA lineage, but cannot be excluded for the other DPs in an allopatric scenario. A limitation of this study is that we did not consider the sex of the individuals. However, it would be interesting to include the sex in the study of DP *P. siculus* as well as other characteristic of the DP that has not been here considered. Moreover, a possible role of sexual selection at a more local level is possible and further studies should focus on the contact area identified in central-southern Italy. Finally, given the obvious limitations, the potential of using Citizen Science databases in obtaining a large amount of information in a short time on aspects related to the phenotypic variability of vertebrate species was highlighted.

Acknowledgments:

We thank Claudia Corti for her suggestions helping to improve the manuscript. Thanks are extended to Flavia Annesi for English revision.

References

1. Arnold, E. N., Burton, J. A., & Ovenden, D. (1978). Field guide to the reptiles and amphibians of Britain and Europe. Collins.
2. Avramo, V., Senczuk, G., Corti, C., Böhme, W., Capula, M., Castiglia, R., & Colangelo, P. (2021). Evaluating the island effect on phenotypic evolution in the Italian wall lizard, *Podarcis siculus* (Reptilia: Lacertidae). Biological Journal of the Linnean Society, 132(3), 655-665.
3. Barra D, Romano P, Santo A, Campajola L, Roca V, Tuniz C. The versilian transgression in the Volturno River plain (Campania, Italy): palaeoenvironmental evolution and chronological data. Il Quaternario. 1996; 9:445–58.
4. Barton, N. H. (1996). Natural selection and random genetic drift as causes of evolution on islands. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351(1341), 785-795.

5. Bisconti, R., Porretta, D., Arduino, P., Nascetti, G., & Canestrelli, D. (2018). Hybridization and extensive mitochondrial introgression among fire salamanders in peninsular Italy. *Scientific reports*, *8*(1), 1-10.
6. Canestrelli, D., Cimmaruta, R., & Nascetti, G. (2008). Population genetic structure and diversity of the Apennine endemic stream frog, Rana italica–insights on the Pleistocene evolutionary history of the Italian peninsular biota. *Molecular Ecology*, *17*(17), 3856-3872.
7. Chiocchio, A., Arntzen, J., Martinez-Solano, I., de Vries, W., Bisconti, R., Pezzarossa, A., ... & Canestrelli, D. (2021). Reconstructing hotspots of genetic diversity from glacial refugia and subsequent dispersal in Italian common toads (Bufo bufo). *Scientific reports*, *11*(1), 1-14.
8. Chen, T. (2021). A Deep Learning-based Computer Vision Approach for Comparative Monarch Butterfly Phenotype Identification in Citizen Science. *The FASEB Journal*, *35*.
9. Chondropoulos, B. P., & Lykakis, J. J. (1983). Ecology of the Balkan wall lizard, Podarcis taurica ionica (Sauria: Lacertidae) from Greece. *Copeia*, *4*, 991-1001.
10. Corti, C., Capula, M., Luiselli, L., Razzetti, E., & Sindaco, R. (2010). Fauna d'Italia, Reptilia. *Edizioni Calderini de Il Sole*, *24*.
11. Corti, C., Lo Cascio, P. (1999). *Lacertidi italiani*. L'epos.
12. Crall, A. W., Jordan, R., Holfelder, K., Newman, G. J., Graham, J., and Waller, D. M. (2012). The impacts of an invasive species citizen science training program on participant attitudes, behavior, and science literacy. *Publ. Understand. Sci.* 22, 745–764.
13. Devictor, V., Whittaker, R. J. & Beltrame, C. Beyond scarcity: Citizen science programmes as useful tools for conservation biogeography. *Divers. Distrib.* 16, 354–362 (2010).
14. Dietvorst, P. J. M., Strijbosch, H., & Bonnemayer, J. J. A. M. (1980). The northernmost population of Podarcis muralis (Lacertilia, Lacertidae). *Amphibia-Reptilia*, *1*(2), 161-172.
15. Drury, J. P., Barnes, M., Finneran, A. E., Harris, M., & Grether, G. F. (2019). Continent-scale phenotype mapping using citizen scientists' photographs. *Ecography*, *42*(8), 1436-1445.
16. Eitzel, M., Cappadonna, J., Santos-Lang, C., Duerr, R., West, S. E., Virapongse, A., et al. (2017). Citizen science terminology matters: exploring key terms. *Citiz. Sci.* 2, 1–20.
17. Gabirot, M., Balleri, A., López, P., & José, M. (2013, August). Differences in thermal biology between two morphologically distinct populations of Iberian wall lizards inhabiting different environments. In *Annales Zoologici Fennici* (Vol. 50, No. 4, pp. 225-236). Finnish Zoological and Botanical Publishing Board.
18. Gusmão, A. C., Evangelista-Vale, J. C., Pires-Oliveira, J. C., Barnett, A. A., & da Silva, O. D. (2021). New records and modelling the impacts of climate change on the black-tailed marmosets. Plos one, 16(9), e0256270.
19. Hijmans, R. J., & van Etten, J. (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12.
20. Hurlbert, A. H., & Liang, Z. (2012). Spatiotemporal variation in avian migration phenology: citizen science reveals effects of climate change. *PloS one*, *7*(2), e31662.
21. Janžekovič, F., & Novak, T. (2012). PCA–a powerful method for analyze ecological niches. Principal component analysis–multidisciplinary applications, 127-142.
22. Kahle, D., Wickham, H., (2013). ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161
23. Kerstes, N. A., Breeschoten, T., Kalkman, V. J., & Schilthuizen, M. (2019). Snail shell colour evolution in urban heat islands detected via citizen science. *Communications biology*, *2*(1), 1-11.
24. Lafi, S. Q., & Kaneene, J. B. (1992). An explanation of the use of principal-components analysis to detect and correct for multicollinearity. Preventive Veterinary Medicine, 13(4), 261-275.
25. Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 30, 314-334.
26. Lehtinen, R. M., Carlson, B. M., Hamm, A. R., Riley, A. G., Mullin, M. M., & Gray, W. J. (2020). Dispatches from the neighborhood watch: Using citizen science and field survey data to document color morph frequency in space and time. *Ecology and evolution*, *10*(3), 1526-1538.
27. Lepetz, V., Massot, M., Chaine, A. S., & Clobert, J. (2009). Climate warming and the evolution of morphotypes in a reptile. *Global Change Biology*, *15*(2), 454-466.
28. Li, E. *et al.* An urban biodiversity assessment framework that combines an urban habitat classification scheme and citizen science data. *Front. Ecol. Evol.* 7, 277 (2019).
29. Marshall K.L.A., Philpot K.E., Damas-Moreira I., Stevens M. *Intraspecific colour*
30. variation among lizards in distinct island environments enances local camouflage, PLoS ONE ed.,
31. 2015. Vol. 10 (9).
32. Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058-1069.
33. Michaelides, S. N., While, G. M., Zajac, N., & Uller, T. (2015). Widespread primary, but geographically restricted secondary, human introductions of wall lizards, Podarcis muralis. *Molecular Ecology*, *24*(11), 2702-2714.
34. Miñano, M. R., While, G. M., Yang, W., Burridge, C. P., Sacchi, R., Zuffi, M., ... & Uller, T. (2021). Climate shapes the geographic distribution and introgressive spread of color ornamentation in common wall lizards. *The American Naturalist*, *198*(3), 379-393.
35. Moreno-Rueda, G., González-Granda, L. G., Reguera, S., Zamora-Camacho, F. J., & Melero, E. (2019). Crypsis decreases with elevation in a lizard. *Diversity*, *11*(12), 236.
36. Ortega, J., López, P., & Martín, J. (2014). Conspicuous blue tails, dorsal pattern morphs and escape behaviour in hatchling Iberian wall lizards (Podarcis hispanicus). *Biological Journal of the Linnean Society*, *113*(4), 1094-1106.

---

37. Ortega, J., Pellitteri-Rosa, D., López, P., & Martín, J. (2015). Dorsal pattern polymorphism in female Iberian wall lizards: differences in morphology, dorsal coloration, immune response, and reproductive investment. *Biological Journal of the Linnean Society*, *116*(2), 352-363.
38. Pellitteri-Rosa, D., Gazzola, A., Todisco, S., Mastropasqua, F., & Liuzzi, C. (2020). Lizard colour plasticity tracks background seasonal changes. *Biology open*, *9*(6), bio052415.
39. Pérez i De Lanuza, G., Font, E. & Carazo, P. (2013). Color-assortative mating in a color-polymorphic lacertid lizard. *Behavioural Ecology*. *24*, 273–279.
40. Pérez i de Lanuza, G., Bellati, A., Pellitteri-Rosa, D., Font, E., & Carretero, M. A. (2019). Colour variation between different lineages of a colour polymorphic lizard. *Journal of Zoology*, *308*(3), 175-187.
41. Podnar, M., Mayer, W., & Tvrtković, N. (2005). Phylogeography of the Italian wall lizard, Podarcis sicula, as revealed by mitochondrial DNA sequences. *Molecular Ecology*, *14*(2), 575-588.
42. Putman, B. J., Williams, R., Li, E., & Pauly, G. B. (2021). The power of community science to quantify ecological interactions in cities. *Scientific reports*, *11*(1), 1-8.
43. R Core Team (2013). R: A language and environment for statistical computing.
44. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*, *12*(1), 1-8.
45. Romano P, Santo A, Voltaggio M. Evoluzione geomorfologica della piana del fiume Volturno (Campania) durante il tardo Quaternario (Pleistocene medio-superiore - Olocene). Il Quaternario. 1994; 7: 41–56.
46. Ruiz Minano, M., G. M. While, W. Yang, C. P. Burridge, R. Sacchi, M. Zuffi, S. Scali, D. Salvi, and T. Uller. "Climate shapes the geographic distribution and introgressive spread of color ornamentation in common wall lizards." *The American Naturalist* (2021).
47. Senczuk, G., Colangelo, P., De Simone, E., Aloise, G., & Castiglia, R. (2017). A combination of long term fragmentation and glacial persistence drove the evolutionary history of the Italian wall lizard Podarcis siculus. *BMC evolutionary biology*, *17*(1), 1-15.
48. Senczuk, G., Havenstein, K., Milana, V., Ripa, C., De Simone, E., Tiedemann, R., & Castiglia, R. (2018). Spotlight on islands: on the origin and diversification of an ancient lineage of the Italian wall lizard Podarcis siculus in the western Pontine Islands. *Scientific reports*, *8*(1), 1-12.
49. Senczuk, G., Harris, D. J., Castiglia, R., Litsi Mizan, V., Colangelo, P., Canestrelli, D., & Salvi, D. (2019). Evolutionary and demographic correlates of Pleistocene coastline changes in the Sicilian wall lizard *Podarcis wagleriana*. Journal of Biogeography, 46(1), 224-237.
50. Sherbrooke, W. C., de L. Castrucci, A. M., & Hadley, M. E. (1994). Temperature effects on in vitro skin darkening in the mountain spiny lizard, *Sceloporus jarrovi*: a thermoregulatory adaptation?. *Physiological Zoology*, *67*(3), 659-672.
51. Silvertown, J., Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., ... & Juan, X. (2011). Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PloS one*, *6*(4), e18927.
52. Simoniello, C., Jencks, J., Lauro, F. M., Loftis, J. D., Weslawski, J. M., Deja, K., et al. (2019). Citizen-Science for the future: advisory case studies from around the globe. *Front. Mar. Sci.* 6, 225.
53. Simpson, J. A., and Weiner, E. S. C. (2014). "Citizen science" in *The oxford English dictionary*, Oxford, UK: Clarendon Press.
54. Smith, K. R., Cadena, V., Endler, J. A., Kearney, M. R., Porter, W. P., & Stuart-Fox, D. (2016). Color change for thermoregulation versus camouflage in free-ranging lizards. The American Naturalist, 188(6), 668-678.
55. Speybroeck, J., Beukema, W., Dufresnes, C., Fritz, U., Jablonski, D., Lymberakis, P., ... & Crochet, P. A. (2020). Species list of the European herpetofauna–2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. *Amphibia-Reptilia*, *41*(2), 139-189.
56. Storniolo, F., Zuffi, M. A., Coladonato, A. J., Di Vozzo, L., Giglio, G., Gini, A. E., ... & Sacchi, R. (2021). Patterns of variations in dorsal colouration of the Italian wall lizard *Podarcis siculus*. *Biology Open*, *10*(10), bio058793.
57. Suárez, N. M., Pestano, J., & Brown, R. P. (2014). Ecological divergence combined with ancient allopatry in lizard populations from a small volcanic island. Molecular ecology, 23(19), 4799-4812.Hurlbert, A. H. & Liang, Z. Spatiotemporal variation in avian migration phenology: Citizen science reveals effects of climate change. *PLoS ONE* 7, e31662 (2012).
58. Tulloch, A. I., Possingham, H. P., Joseph, L. N., Szabo, J., and Martin, T. G. (2013). Realising the full potential of citizen science monitoring programs. *Biol. Conserv.* 165, 128–138.
59. Xie, C., Zhang, G., Jim, C., Liu, X., Zhang, P., Qiu, J., & Liu, D. (2021). Bioclimatic Suitability of Actual and Potential Cultivation Areas for *Jacaranda mimosifolia* in Chinese Cities. Forests, 12(7), 951.
60. Yang, W., While, G. M., Laakkonen, H., Sacchi, R., Zuffi, M. A., Scali, S., ... & Uller, T. (2018). Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard. *Molecular Ecology*, *27*(21), 4213-4224.
61. Yang, W., Feiner, N., Laakkonen, H., Sacchi, R., Zuffi, M. A., Scali, S., ... & Uller, T. (2020). Spatial variation in gene flow across a hybrid zone reveals causes of reproductive isolation and asymmetric introgression in wall lizards. *Evolution*, *74*(7), 1289-1300.
62. Yoon, S., & Lee, W. H. (2021). Methodological analysis of bioclimatic variable selection in species distribution modeling with application to agricultural pests (*Metcalfa pruinosa* and *Spodoptera litura*). Computers and Electronics in Agriculture, 190, 106430.
63. Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, *113*(29), 8041-8048.