Landscape genomics in a cosmopolitan earthworm along a climatic gradient

Marchán DF[1], Tilikj N[1], Taboada S[1], Jiménez S[1], Fernández R[2], Novo M[1]

1. Depto. de Biodiversidad, Ecología y Evolución, Facultad de Ciencias Biologicas, Universidad Complutense de Madrid, José Antonio Novais, 2, 28040, Madrid, Spain
2. Institute of Evolutionary Biology (IBE, CSIC-UPF), Barcelona, Spain

Introduction

Some **earthworm species** have shown their capacity to **acclimate or adapt to diverse climate conditions**. Within a context of **climate change**, those species that present **variable** loci in **genes involved in functional responses to changing climate** variables may have **more options to rapidly adapt** to locally changing environments. **Understanding these alleles** putatively associated with acclimation could aid to **predict the responses of species** to the **future conditions** and to mitigate biodiversity loss.

Aporrectodea trapezoides Dugès, 1828 is one of the most **widely distributed cosmopolitan species** in temperate habitats: their geographical distribution (which includes a range of different climatic conditions) make it an **ideal candidate** to explore the relationship between their **genomic variability and climate (landscape genomics).**

Materials and methods

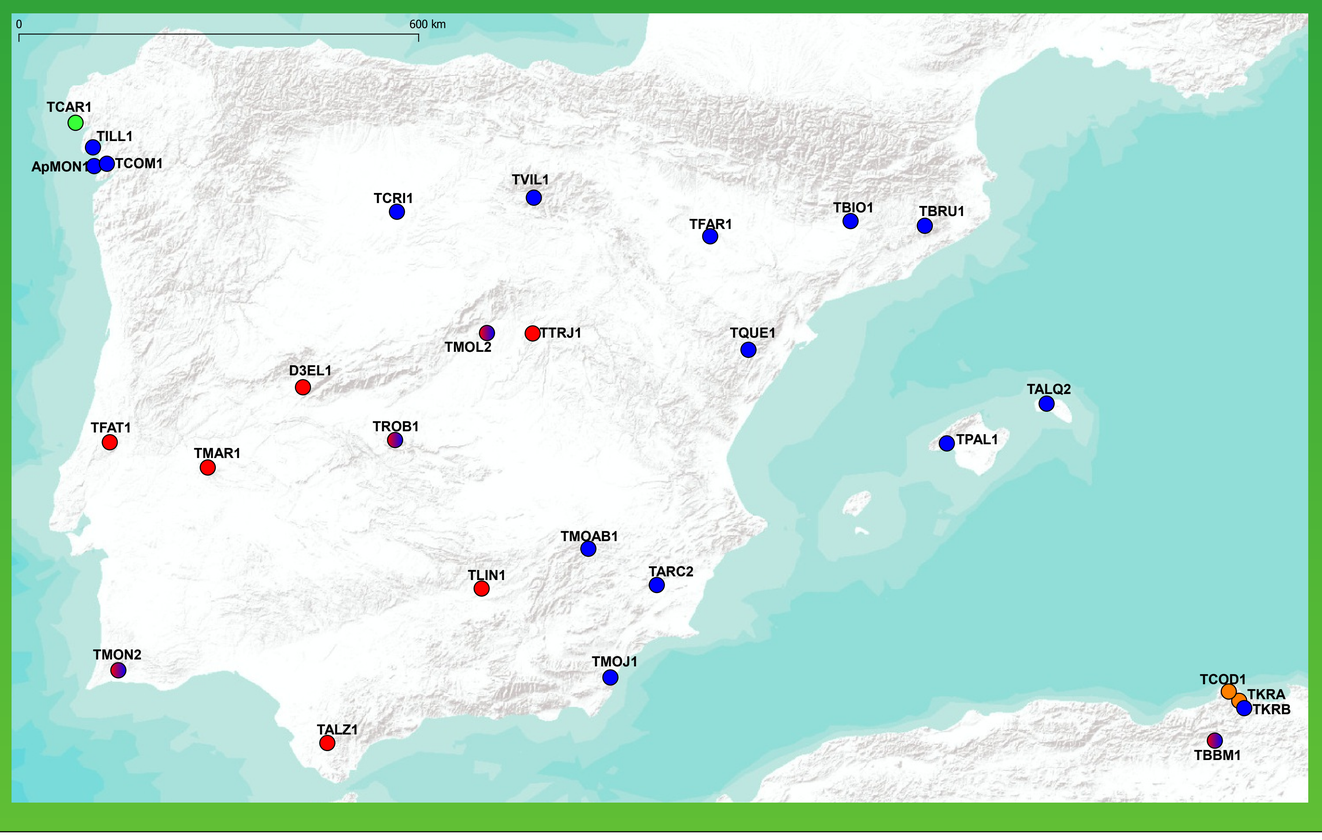
Twenty-eight populations from the **Iberian Peninsula and Algeria** were selected to encompass the **climatic range of the Mediterranean lineage II** (following Fernández et al. 2012) based on a preliminary PCA analysis including the main environmental variables (such as temperature and precipitation). In total, **134 individuals** were sequenced using dGBS, using the restriction enzyme MsII, and an average of 2 million reads per individual were obtained.

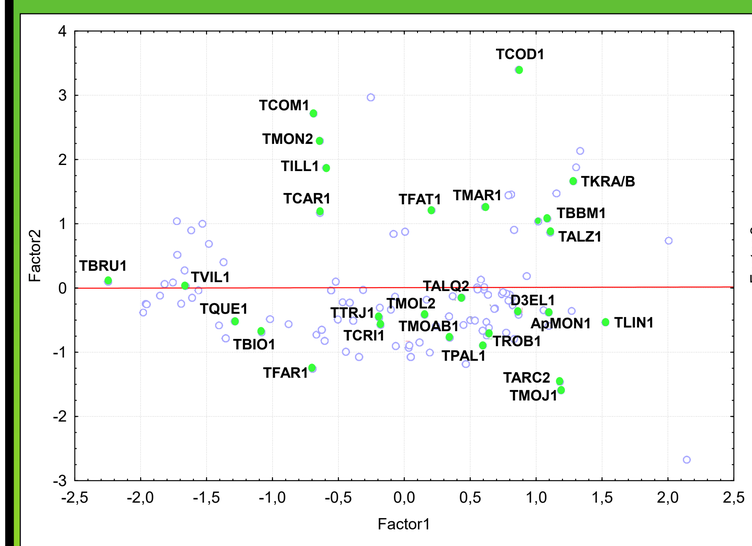
-**Quality filtering and locus assembly** was conducted with the **Stacks** pipeline version 2.57. RAD-tags were processed using process_radtags with the rescue (-r) and trimming (-t) features. The Stacks populations module was next used to retain SNPs present in at least 80% of individuals (r = 0.8) and the first SNP from each RAD-tag using –write_single_SNP, in order to reduce the linkage disequilibrium among loci.

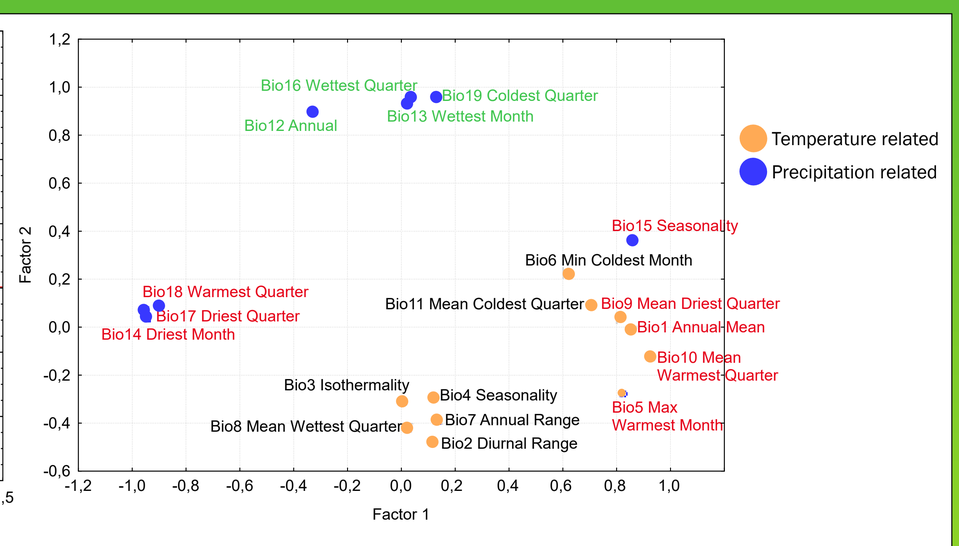
-**Population structure** was assessed using two different methods:

STRUCTURE was run using the admixture model, putative K from 1 to 9 and 15 replicates for each run. We used STRUCTURE HARVESTER and CLUMPP version 1.1.2 to determine the most likely number of clusters and to average each individual's membership coefficient across the K value replicates, respectively.

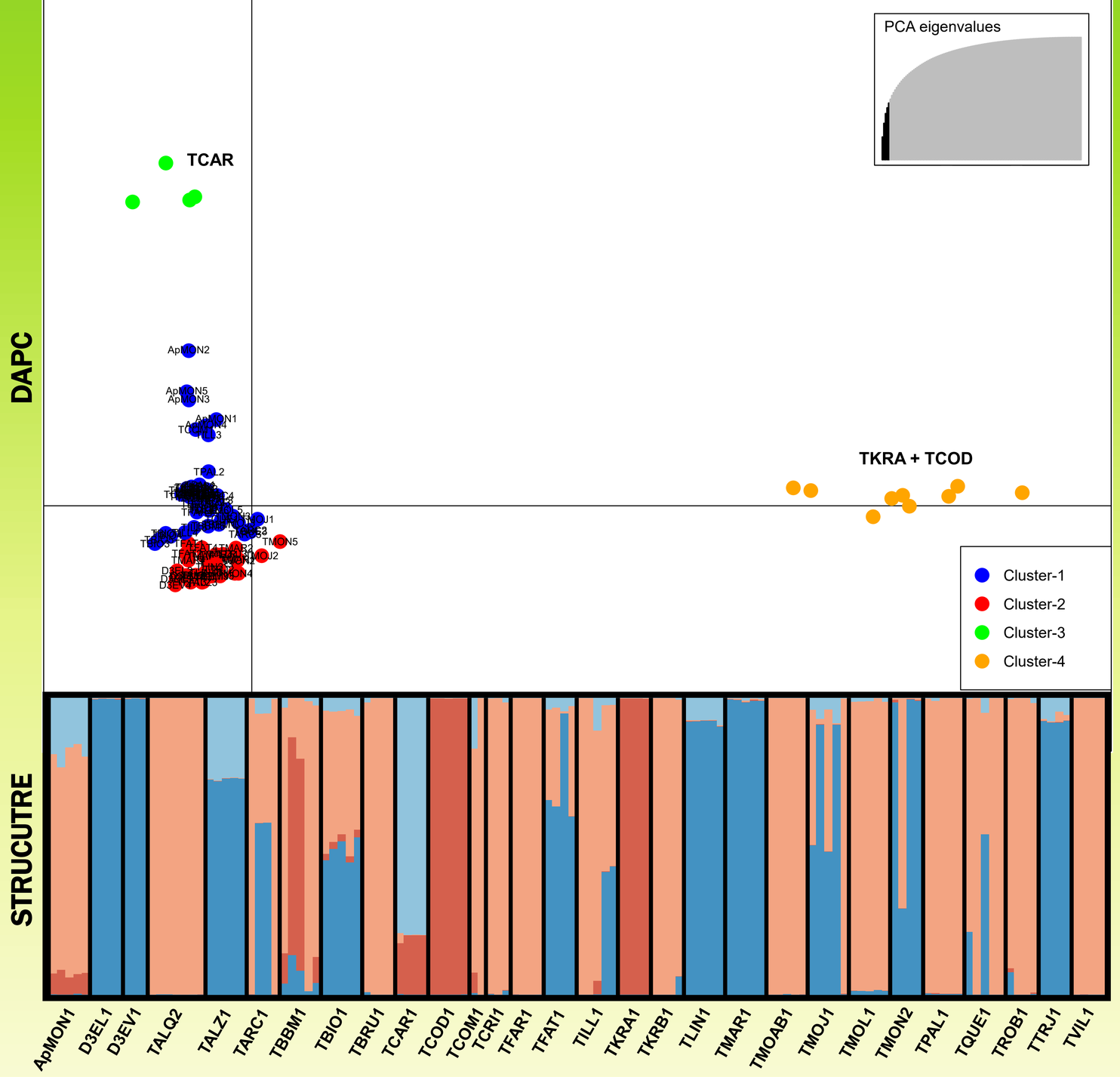
Population structure in **DAPC** was assessed by the function snapclust using the genetic clustering mode snapclust.choose.k and the Akaike Information Criterion (AIC) function. To identify the optimal number of clusters, k-means was run. The number of retained principal components (PCs) axes and eigenvalues were chosen using the cross-validation xvalDapc function from the adegenet R package. The DAPC function assignplot was used to plot the probabilities of assignment of the different individuals to the different clusters, while the function scatter.plot was used to produce scatterplots of PCs with eigenvalues as inset.







PCA analysis of the main bioclimatic variables characterizing the studied localities. Left: Factor scores of each locality. Right: Factor loadings of the bioclimatic variables; in red if FL >70 (factor 1), in green if FL >70 (factor 2)



Preliminary results

After filtering, a total of **664,331,462 reads** from the initial 670,929,490 raw reads (99 % of reads retained) were retained, with an average of **4,957,697 reads per sample** (1,149,183 - 15,798,547).

Four individuals were discarded due to high missing data, resulting in a final dataset of 128 individuals. The number of **SNPs** obtained was **814.**

Four different genetic clusters were identified by **DAPC.** The **most divergent cluster** comprised samples from **TCOD** and **TKRA (Algeria).** The other **three clusters** identified were **separated in a latitudinal gradient** within the **Iberian Peninsula**, with samples of **TCAR (northernmost population, green) separated** from the rest of samples, and then **two groups genetically more similar** including samples from **Northwestern-Central-Mediterranean Iberian Peninsula, Balearic Islands and Algeria(in blue)** and samples from **Southwestern-Central Iberian Peninsula (in red).**

Four different genetic clusters (consistent with the DAPC clusters) were identified by **STRUCTURE** with samples from **TCOD and TKRA (Algeria)** as a distinct group and the rest showing different degrees of admixture.

Future steps

-**Detection of SNPs under selection** will be performed using Bayescan, PCAdapt (among others)

-LFMMs (Latent factor Mixed Models) and the Samßada program will be used to study the **association between SNPs and environmental variables**, taking into account the population structure, which could be a confounding factor.

-**SNPs under selection** and those **associated with environmental variability** will be **mapped against** a de novo assembled **transcriptome** to obtain **candidate genes** potentially involved in **adaptation to the environment in earthworms.**