

# Procurando sinais de selecção no genoma nuclear da abelha ibérica (*Apis mellifera iberiensis*)

DORA HENRIQUES, ANDREAS WALLBERG, JULIO CHÁVEZ-GALARZA, FILIPE O. COSTA, JOSÉ RUFINO, MATTHEW WEBSTER, M. ALICE PINTO



## Introdução

Estudo da adaptação local

Quais os factores/genes cruciais que permitiram a sobrevivência dos indivíduos num certo habitat

Prever a resposta da espécie a um mundo em acelerada transformação.



## Objectivo

Procurar sinais de selecção no genoma nuclear da abelha



Correlacionar frequências alélicas com variáveis ambientais

## Vantagem:

Perceber quais os factores ambientais que actuam como pressão selectiva

## Material e Métodos

### 1) Amostragem

- Representar a distribuição natural da abelha ibérica;
- 87 indivíduos;
- 3 transectos longitudinais



### 2) Variáveis Ambientais

Retiradas das bases de dados

- [www.worldclim.org](http://www.worldclim.org)
- [www.cru.uea.ac.uk](http://www.cru.uea.ac.uk)

Apenas a variáveis que tinham  $|r| < 0.80$  permaneceram



### 3) Sequenciação

Illumina HiSeq 2500 Paired-end (2X150pb)

### 4) Filtros e obtenção de SNPs

- Remoção dos duplicados de PCR (Picard)
- Realinhamento (GATK)
- Obtenção de SNPs (Freebayes)

- Retirar SNPs :**
  - > 2 alelos
  - 123 > depth < 3000
  - < 37 amostras

**1 289 449 SNPs**

### 5) Análises

**SAMβADA (autocorrelação espacial)**

3000 Permutações  
25 "nearest neighbours"

**LFMM (estrutura populacional)**

2 factores latentes  
50000 iterations  
burnin-in 25000

Península Ibérica

Laboratório natural

Engloba diversos habitats como o Mediterrânico, Deserto, Alpino e Atlântico

Como é que um organismo se adapta aos diferentes ambientes

## Resultados e Discussão

Ordenar os modelos tendo em a estatística Wald (1.73 e 26.14)

Fazer intervalos

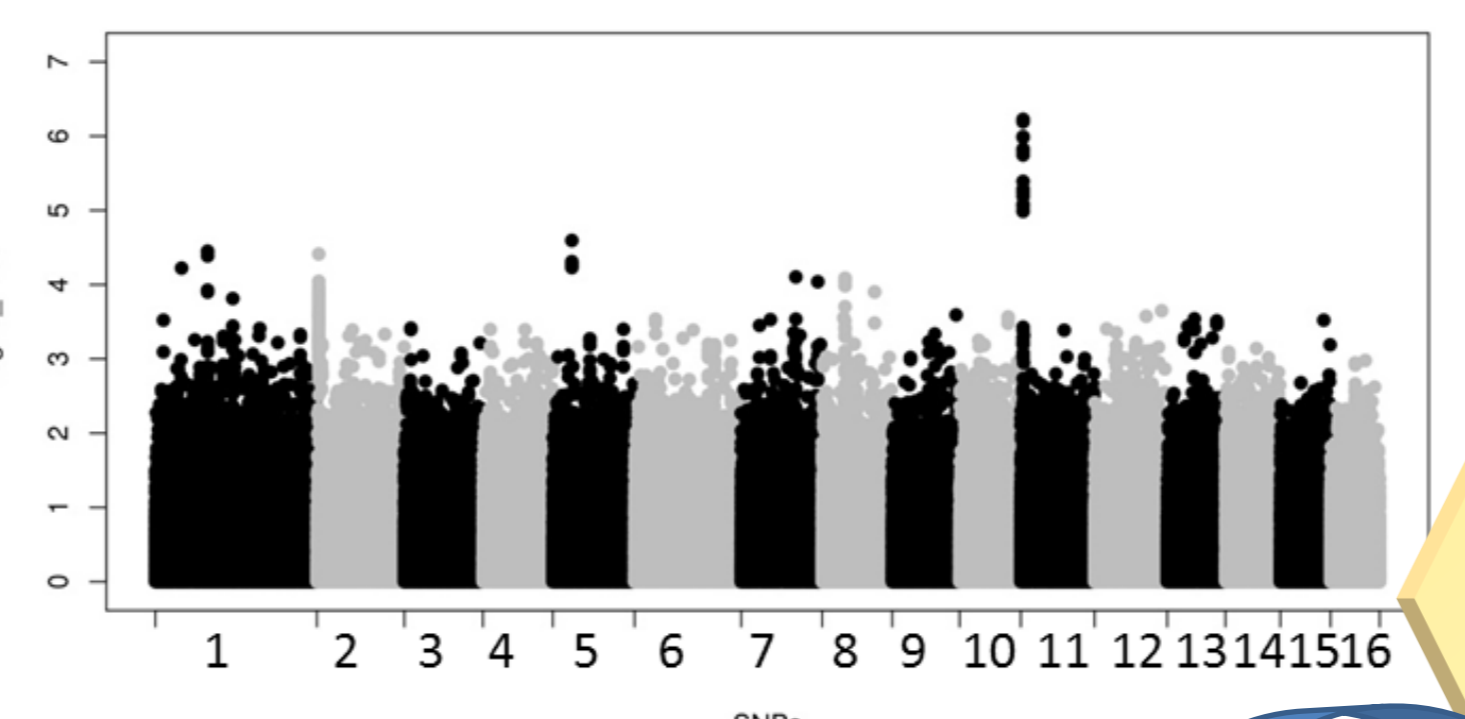
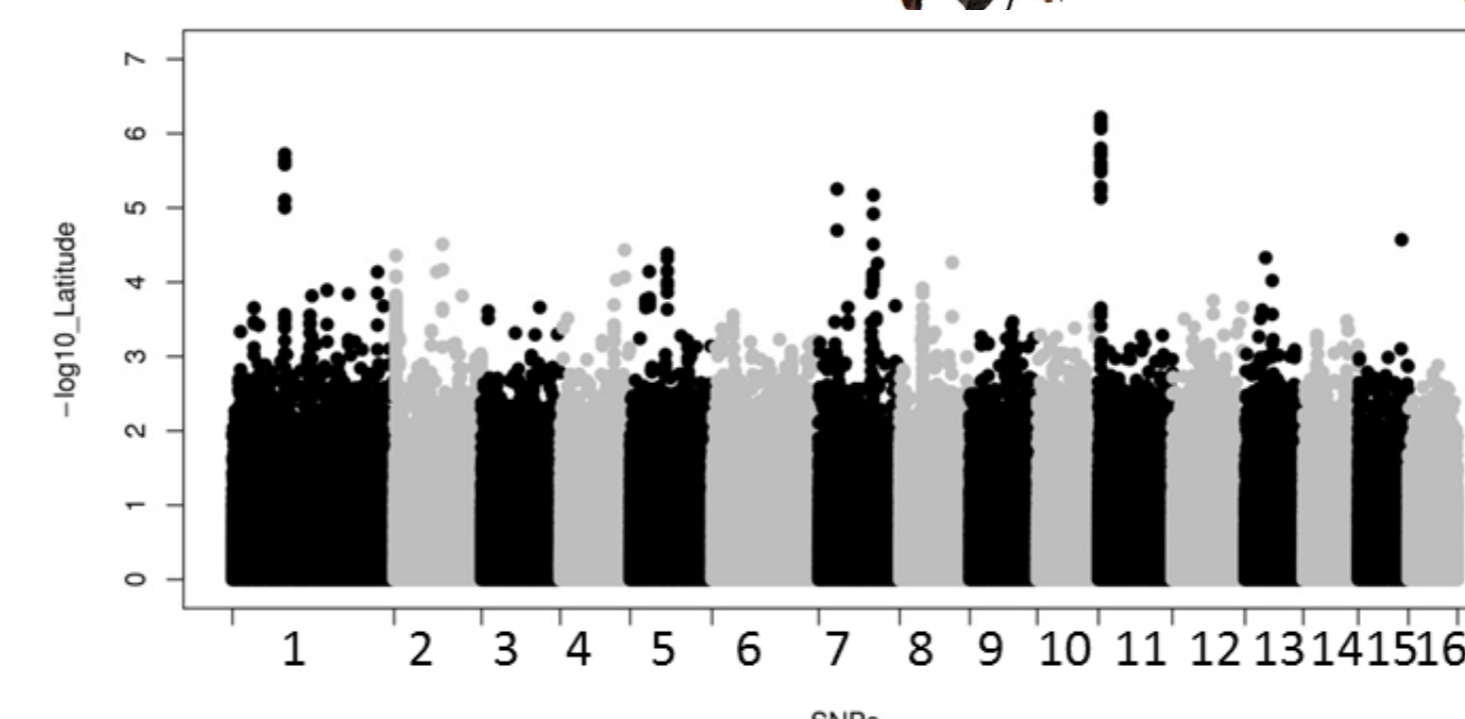
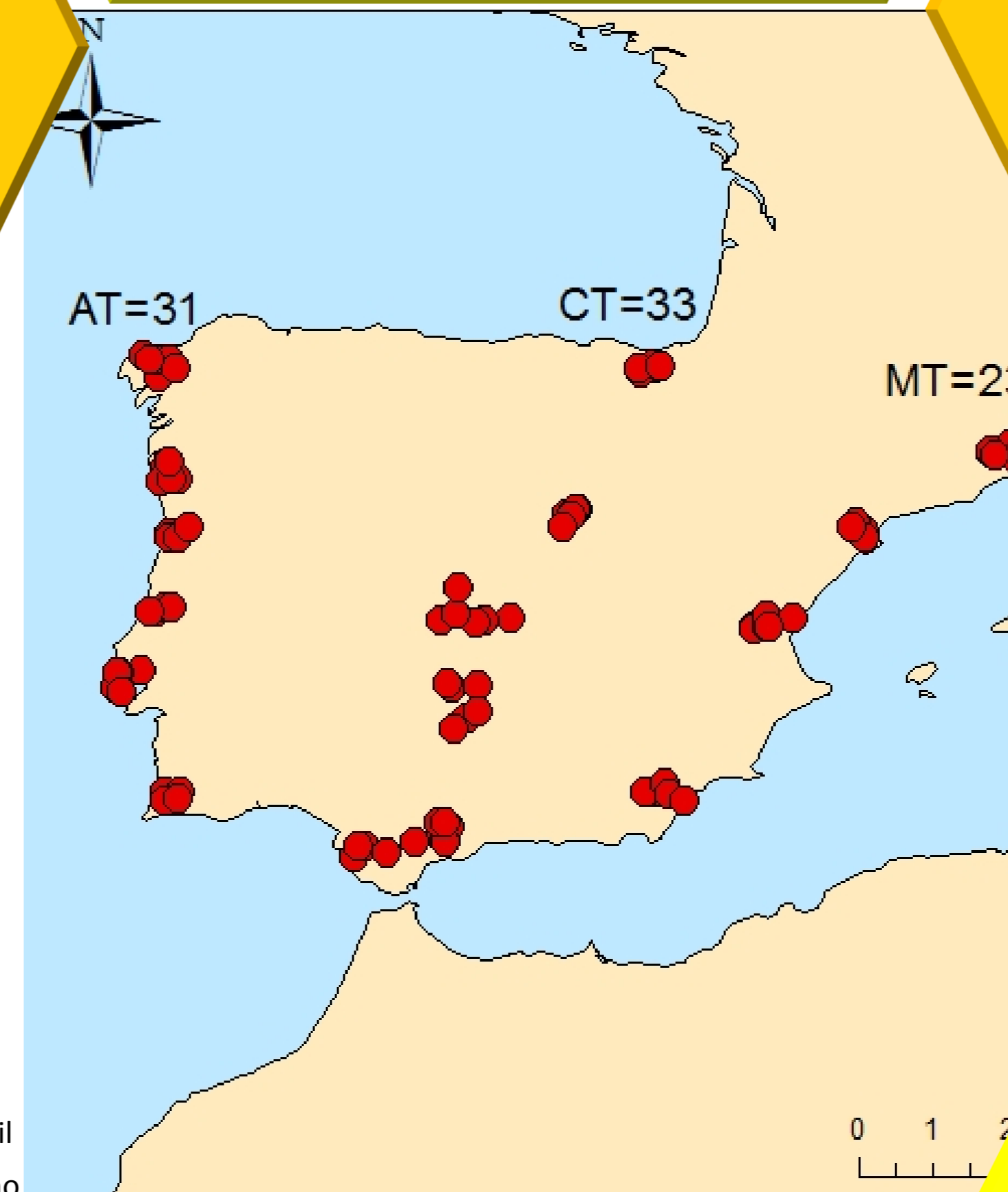
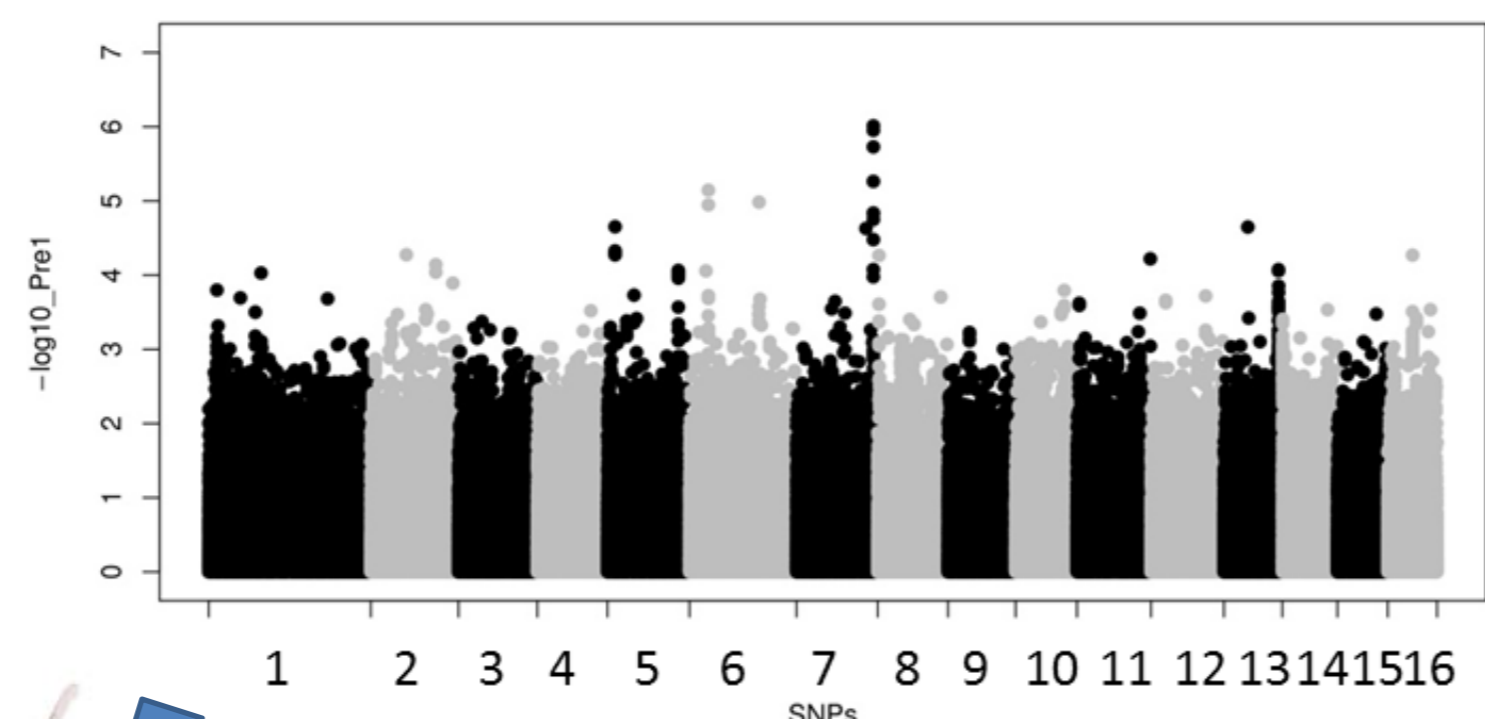
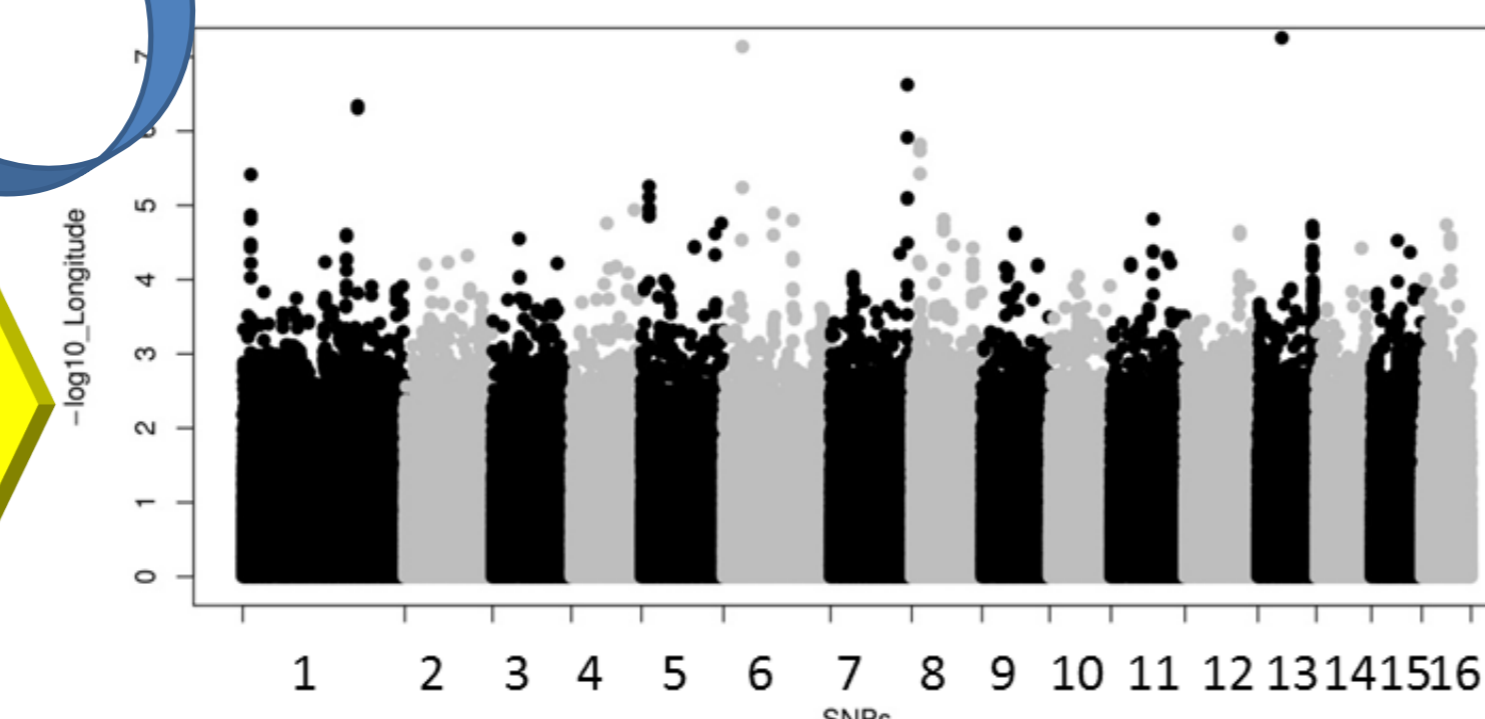
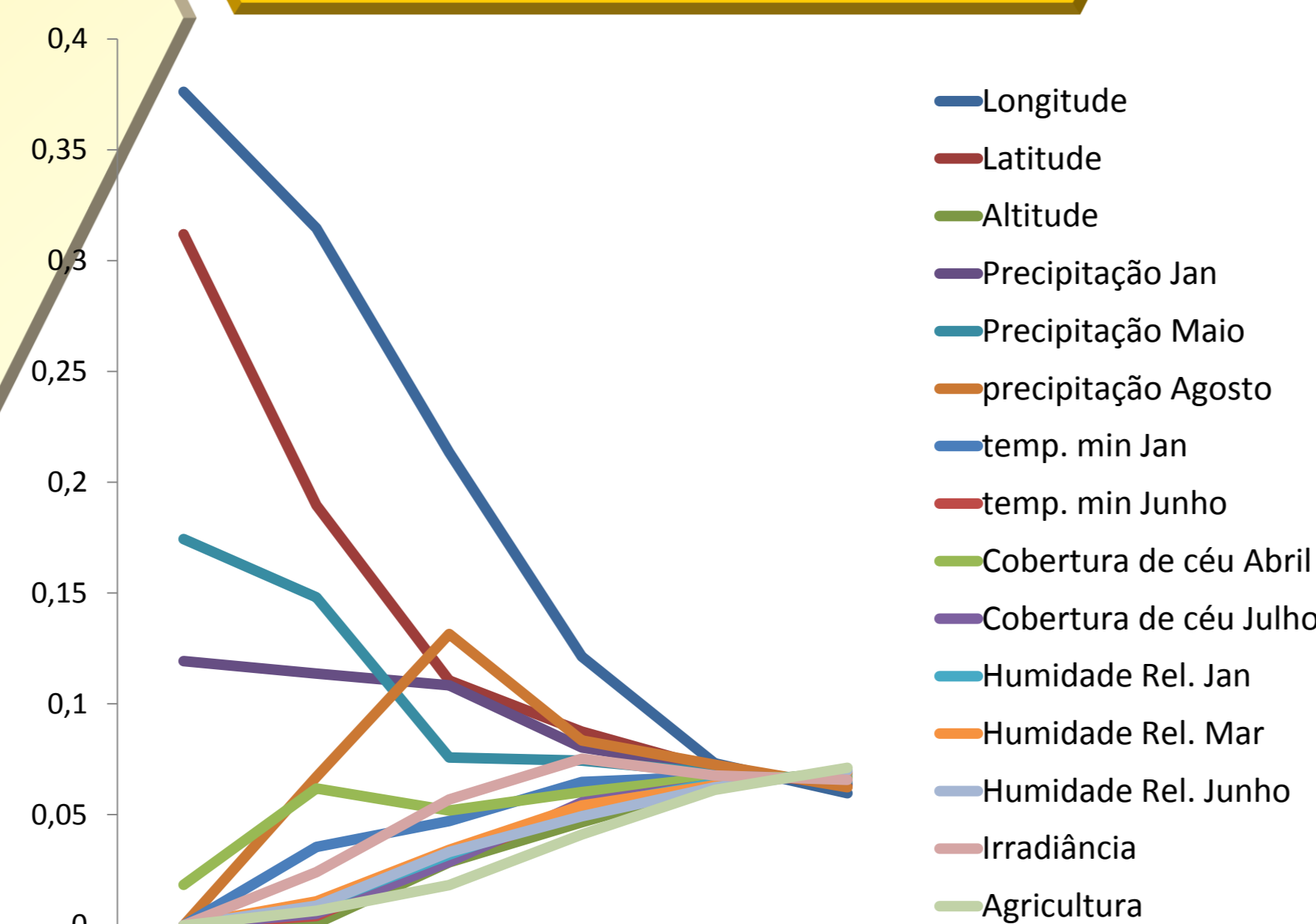
As variáveis ambientais com maiores valores de Wald:

- Longitude;
- Latitude;
- Precipitação em Janeiro;
- Precipitação em Maio.

A distribuição dos valores de  $-\log_{10}$  (p-values) obtidos por mostram a existência de "picos" em diversos cromossomas

Para os SNPs que contem um  $-\log_{10}$  (p-values) maior de 4 verificamos quais os genes onde se encontram.

Os genes encontrados pelos 2 programas coincidem



Gene	Variavel ambiental	Função	Termo
GB40077	long;pre1	metabolismo	metabolismo de lipidos
GB41118	long	metabolismo	processo metabolico de um carbono
GB45499	long	transporte	transportador de sódio
GB47279	long	metabolismo	reposta a insecticida
GB48105	pre1	morfogenesis	neurogeneses
GB48709	lat	regulação	regulação negativa da expressão génica
GB51396	long	metabolismo	desidrogenase hidroxisteroide
GB55263	long	metabolismo	processo de biossíntese de cera

Para os genes que contêm SNPs que levam a uma mutação não sinónima, verificou-se qual a sua função (NCBI, DAVID DATABASE, FLYBASE)

Os termos que mais aparecem estão correlacionados com o metabolismo

We owe special thanks to beekeepers from Spain and Portugal for their valuable help in obtaining samples. Antonio Pajuelo provided the contacts of Spanish beekeepers. Special thanks to Margarida Neto, Andreia Brandão and Irene Muñoz for collaborating in the sampling. JC-G and DH are supported by Fundação para a Ciência e Tecnologia (FCT) through the scholarships SFRH/BD/68682/2010 and SFRH/BD/84195/2012, respectively. This research was funded by FCT and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 and BiodivERsA-FACCE2014-91.

