

congenomics ■

2016

Conference on
Conservation Genomics

PROGRAMME

3 – 6 May 2016 | Campus de Vairão, Portugal | CIBIO-InBIO

EUROPEAN
SCIENCE
FOUNDATION

congen. ■



Posters

- P01** An Vanden Broeck | Dispersal, gene flow and sibship analysis of *Phengaris (Maculinea) alcon*: implications for conservation
- P02** Ana Lino | Genetic consequences of habitat loss and fragmentation on vertebrates
- P03** Anne Andersen | Phylogeography of the large blue butterfly (*Maculinea arion*)
- P04** Bárbara Santos | Metabarcoding reveals trophic relations between the giant wall gecko, *Tarentola gigas* and other endemic species in an Integral Reserve in the Atlantic Ocean (Raso islet, Cape Verde)
- P05** Barbora Zemanová | National Animal Genetic Bank in the Czech Republic – infrastructure for conservation genetics and genomics research
- P06** Begoña Martínez-Cruz | Population genomics and the demographic history of Imperial Eagles
- P07** Catarina Silva | Insights into the evolution of phenotypic traits in house sparrows
- P08** Daniel Kleinman-Ruiz | Genomic tools for the conservation of Iberian lynx: A novel and curated set of genome-wide SNPs
- P09** David Stanković | Detection of olm (*Proteus anguinus*) environmental DNA in karst groundwater
- P10** Dora Henriques | Searching for signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) using allele-environment association approaches
- P11** Marketa Zimova | Interspecific competition, hybridization, replacement and adaption: a case study using whole genome sequencing in hares from Scotland
- P12** Federica Costantini | RAD genotyping to evaluate inter and intra specific levels of genetic structuring in Mediterranean biogenic reefs species
- P13** Filipa Martins | Towards Next-generation Biodiversity Monitoring: Improving Freshwater Quality Assessment using DNA Metabarcoding
- P14** Guy Colling | Changes at different altitudinal levels in the metabolic fingerprint of *Arnica montana* L. leaf extracts by 1H NMR-based metabolomics
- P15** Joana Paupério | InBIO Barcoding Initiative: moth reference database using Next Generation Sequencing
- P16** João Pedro Marques | A guide to reconstruct “true” de novo transcriptomes
- P17** João Queirós | Genome-wide associations identify novel candidate loci associated with genetic susceptibility to tuberculosis in wild boar
- P18** Johan Michaux | Using Next Generation Sequencing to characterize species diets: a study case with faeces of wild boar in the Spanish Pyrenees
- P19** Laura Daco | Climate change, plasticity and genetic variation in *Anthyllis vulneraria*
- P20** Lua Lopez Perez | The attack of the clones
- P21** Mafalda Costa | Phylogeography and population genetic structure of the European polecat (*Mustela putorius*)
- P22** Manuel Curto | A new approach of whole mitogenome sequencing and its application for phylogenetic analysis of East African white-eyes (Aves, Zosteropidae)
- P23** Manuel Curto | Mitogenomics using amplicon sequencing with Illumina MiSeq
- P24** Margarida Gonçalves | Population genomics across the distribution range of the sable antelope

SEARCHING FOR SIGNATURES OF SELECTION IN THE IBERIAN HONEY BEE (*APIS MELLIFERA IBERIENSIS*) USING ALLELE-ENVIRONMENT ASSOCIATION APPROACHES

Dora Henriques^{1,2}, Andreas Wallberg³, Julio Chávez-Galarza^{1,2}, Cátia Neves⁴, José Rufino^{4,5}, Filipe O. Costa², Matthew T. Webster³, M. Alice Pinto^{1,4}

¹Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, ²CBMA, Centro de Biologia Molecular e Ambiental, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

³Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

⁴Instituto Politécnico de Bragança, 5301-857 Bragança, Portugal.

⁵Laboratório de Instrumentação e Física Experimental de Partículas, Campus de Gualtar, 4710-057 Braga, Portugal.

INTRODUCTION

In the current context of a global human-mediated environmental crisis, understanding which genes and mechanisms are responsible for adaptation to different climates will enable predictions on how organisms will respond to a rapidly changing world. This is particularly important for the honey bee, a key-stone species for ecosystem functioning and economy, which is facing increasing pressures from the effects of intensified land use, climate change, and the spread of pests and pathogens.

OBJECTIVES

The aim of this work is searching for signatures of selection along the genome of 87 individuals using two different allele-environment association approaches.

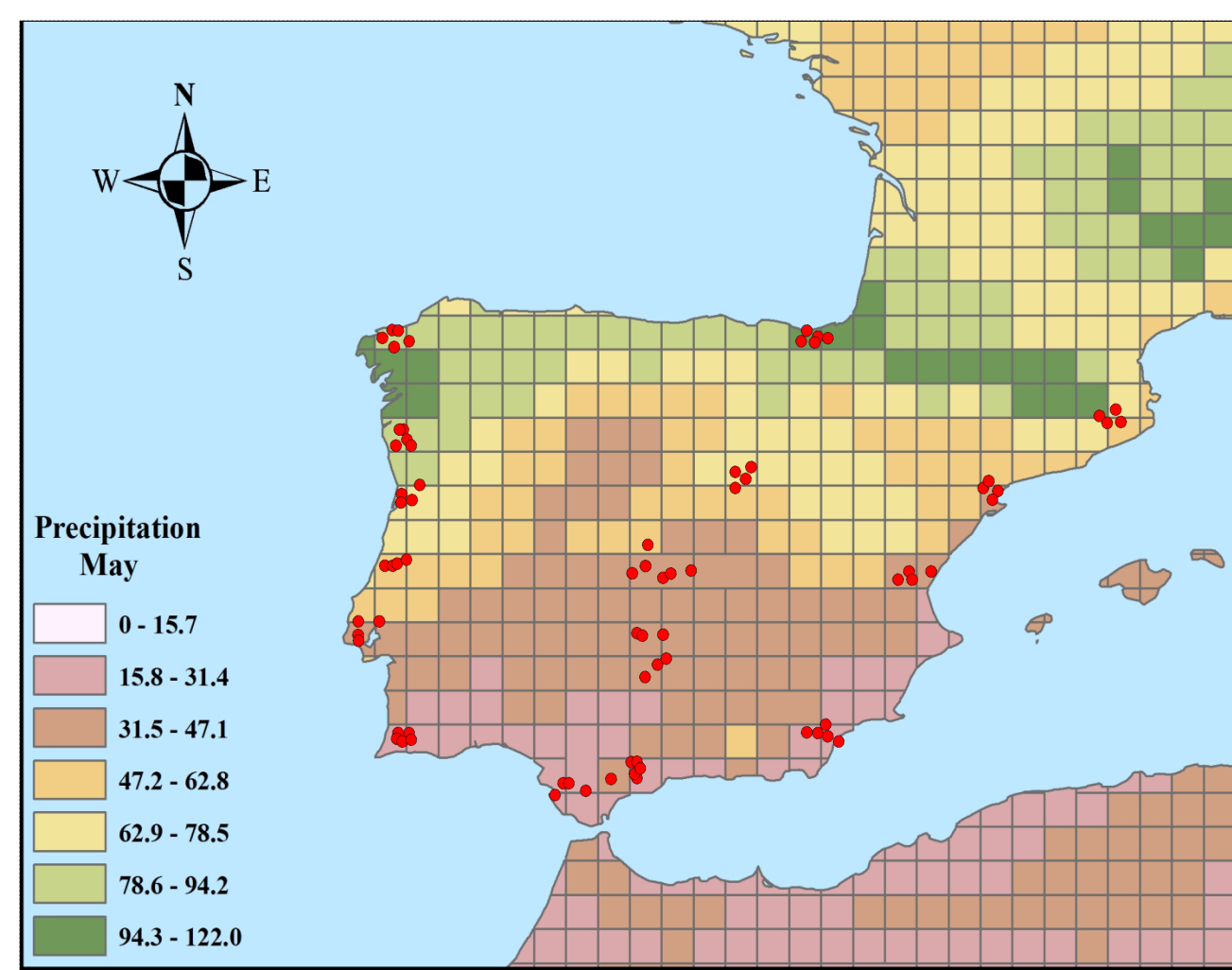


Figure 1: Map representing the distribution of precipitation in May. The red spots indicates the position of the individuals used in this study.

METHODS



Genome sequencing

Illumina technology was used to sequence the whole genomes of 87 Iberian honey bees collected across three north-south transects in the Iberian Peninsula (Fig.1).

The reads were mapped against the reference genome *Amel_4.5* using BWA.

Several steps of quality control were performed:

Identify and mark PCR duplicates: Picard

Identify regions of poorly and inconsistently mapped reads: GATK45

SNP calling: FreeBayes

Remove SNPs with: quality below 50; depth below 123 or higher than 3000 present in less than 37 alleles

In total the dataset is composed by 1,289,449 SNPs.

Environmental data

The environmental data were obtained from the Climatic Research Unit (www.cru.uea.ac.uk) and ArcGIS was used to manipulate the data

Searching for signals of selection

The allele-environmental association methods LFMM and Samβada were used to search signals of selection along the genome.

The database PANTHER v.8.0 (protein annotation through evolutionary relationships) classification system (<http://www.pantherdb.org/>) was used to examine if adaptively evolving loci were enriched for specific functional annotation.

RESULTS AND DISCUSSION

➤ A total of 1,416 (FDR<0.05), 360 (FDR<0.02), and 220 SNPs (FDR<0.01) were identified by the LFMM approach (Fig. 2).

➤ Samβada detected a total of 2,145 SNPs with a P-value < 0.0001).

➤ Of the 1,416 outlier's SNPs detected by LFMM 814 were simultaneously detected by Samβada (Table 1) and these were located in or near 302 genes. The environmental variables that showed a greater number of associations with the 814 SNPs were long (113), lat (350), prec1 (143), prec5 (308), and cld4 (108) (Table1).

➤ The climate-driven candidate gene with more SNPs associated is GB49881 (37), most of them (28) associated with prec1. GB49881 is only 1,864 bp away from another candidate GB49882, which contained 15 SNPs.

➤ The homolog gene GB49882 in *Drosophila melanogaster* encodes the protein SHAW, that is a member of shaker family of voltage-gated K⁺ channels. This protein is widely expressed in the nervous system, more specifically in clock neurons and helps to regulate resting membrane potential in central neuron. Parisky *et al.* 2008 have shown that this gene plays an important rule in the regulation of the sleep circuit or circadian clock in *Drosophila melanogaster*.

➤ Of a total of 14 GO categories were significantly over-represented (threshold of P-value<0.05) (Fig. 3)

➤ We have an over-representation of the terms *receptor binding* and *receptor activity* for the molecular function that are crucial for the organism recognize the extracellular messages and translate these signals into cellular response.

Table 1. Number of overlapping SNPs detected by the allele-environment approaches and number of genes detected just by Samβada (P-value < 0.0001) and LFMM (FDR<0,05) for each environmental variable.

	Overlapping	LFMM unique	Samβada unique
long	113	0	564
lat	350	35	64
prec1	143	21	105
prec5	308	288	10
prec8	79	272	58
tmin1	57	19	11
tmin6	5	2	4
cld4	108	89	23
cld7	8	26	3
rh1	16	17	3
rh3	19	92	14
rh6	6	17	14
ins4	20	48	28
land	11	21	0

Note: prec (precipitation), tmin (minimum temperature), cld (cloud cover)
rh (relative humidity), ins (insolation) and land (Land Cover)
numerals 1 to 12 on front of each variable specify the month

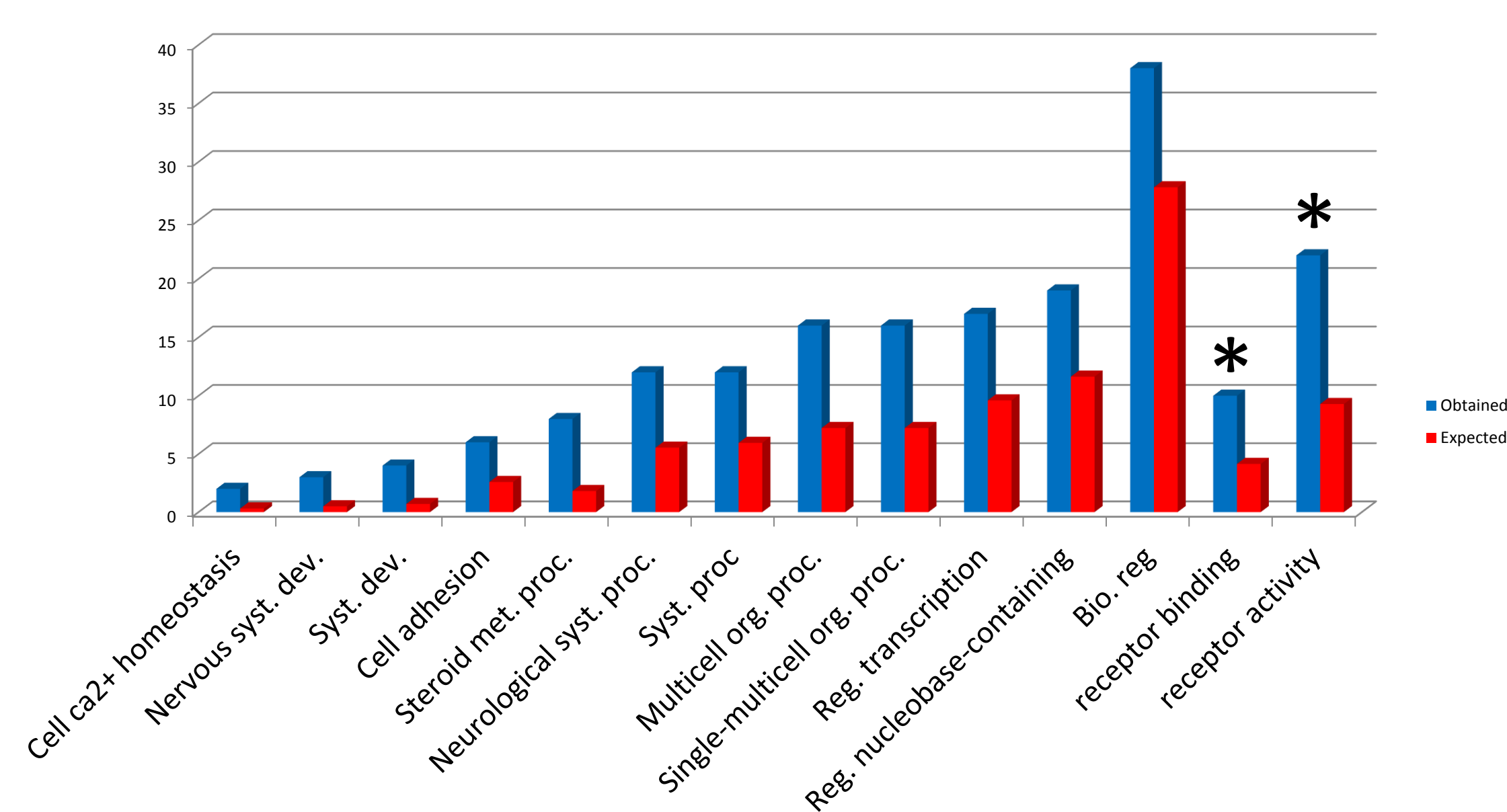


Figure 3: Enrichment of GO-Terms related with biological and molecular functions (the molecular functions are marked by *). Abbreviations : syst (system), dev (development), met (metabolic), proc (process), org (organism), reg (regulation), bio (biological).

➤ Some of the biological functions over-represented are important for the cellular response such as *steroid metabolism process*, *cell adhesion*, *neurological system process*, *nervous system development*.

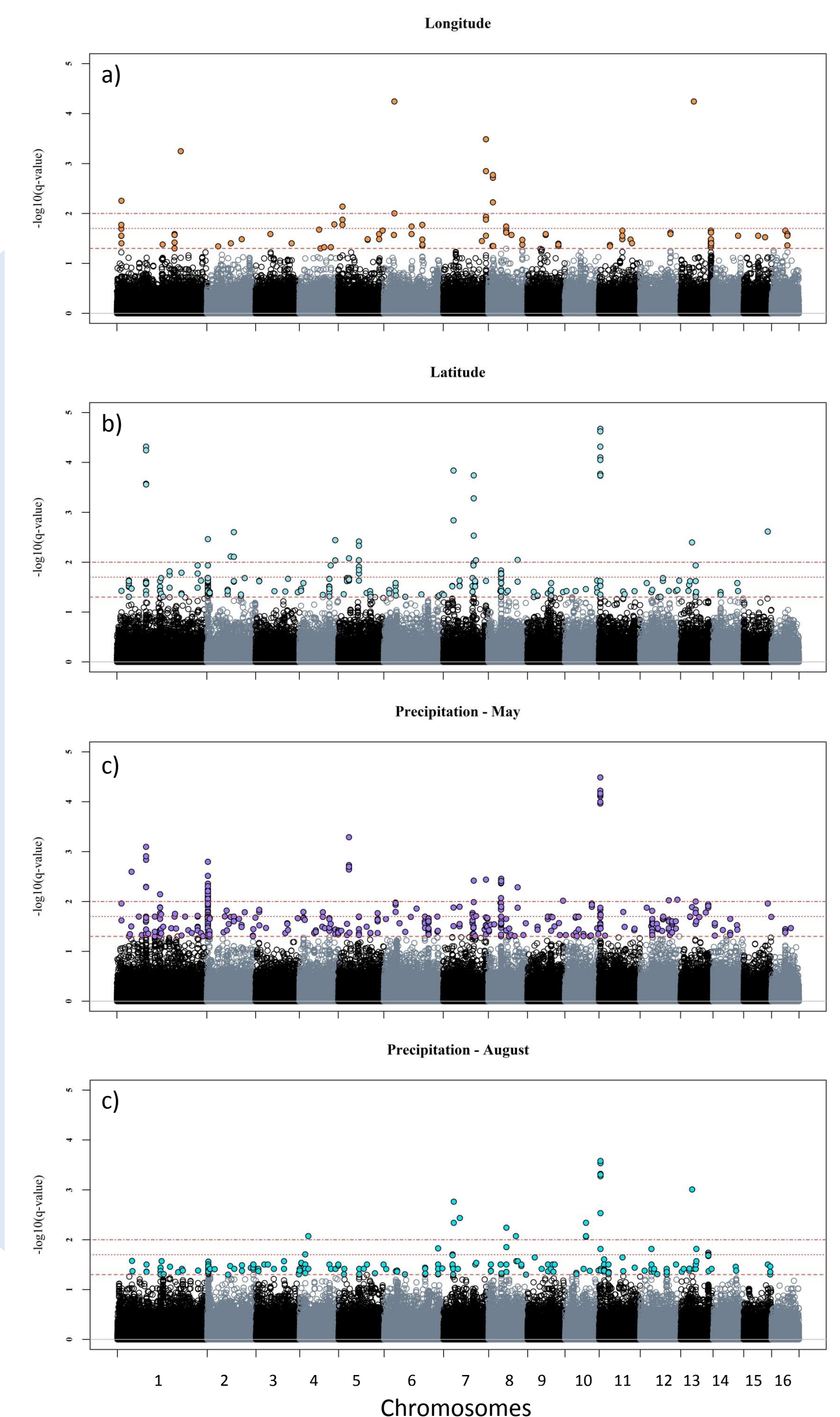


Figure 2: Genome-wide distribution of significance values $-\log_{10}(q\text{-value})$ obtained by the allele-environment approach LFMM for the variables with higher number of outliers a) longitude b) latitude c) precipitation in May c) precipitation in August. The red lines indicate FDR values of 0.05, 0.02 and 0.01

➤ We found nine genes related with insecticide response, GB54460 was previously described by Chávez-Galarza *et al.* (2013) and Harpur *et al.* (2014). Five out nine genes are associated with longitude. These could be due differences in agriculture practice between the countries.

References

- Chavez-Galarza J, Henriques D, Johnston JS, *et al.* (2013) Signatures of selection in the Iberian honey bee (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. *Mol Ecol* 22, 5890-5907.
Harpur BA, Kent CF, Molodtsova D, *et al.* (2014) Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A* 111, 2614-2619
Parisky KM, Agosto J, Pulver SR, *et al.* (2008) PDF cells are a GABA-responsive wake-promoting component of the *Drosophila* sleep circuit. *Neuron* 60, 672-682.

Acknowledgements

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. Bioinformatic analyses were performed using resources at the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX)