
PREDICTION OF MALADAPTATION FROM
ECOLOGICAL AND GENOMIC DATA
USING GENOMIC OFFSETS

FRANCISCO DE BORJA CAMPUZANO JIMÉNEZ, au726682

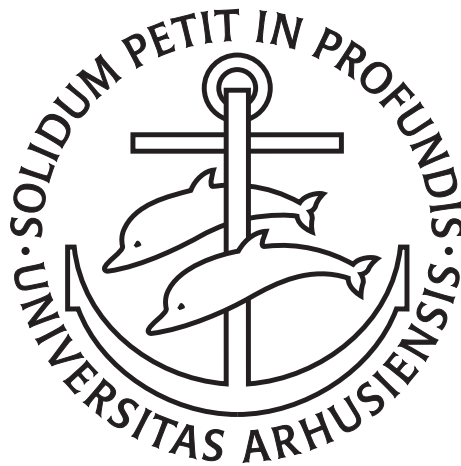
MASTER OF SCIENCE IN BIOINFORMATICS

June 2024

Supervisors: Thomas Bataillon and Genis Garcia Erill

PREDICTION OF MALADAPTATION FROM ECOLOGICAL AND
GENOMIC DATA USING GENOMIC OFFSETS

FRANCISCO DE BORJA CAMPUZANO JIMÉNEZ



Master of Science in Bioinformatics

Bioinformatics Research Centre
Faculty of Natural Sciences
Aarhus University

June 2024

Francisco de Borja Campuzano Jiménez: *Prediction of maladaptation from ecological and genomic data using genomic offsets*, Master of Science in Bioinformatics © June 2024

CONTENTS

1	INTRODUCTION	1
2	STATE OF THE ART	3
2.1	Maladaptation and local adaptation	3
2.2	Why use observational data to predict maladaptation?	4
2.3	Phenotypic offset	5
2.4	Geometric genomic offset	6
2.4.1	Latent factor mixed model (LFMM)	10
2.5	Identifying putatively adaptive loci	11
2.6	Alternative genomic offset statistics	12
2.6.1	Risk of non-adaptedness	13
2.6.2	Gradient forest genomic offset	14
2.6.3	Redundancy analysis	15
3	RESULTS	17
3.1	Conceptual issues on genomic offsets	17
3.1.1	Association of maladaptation and fitness loss	17
3.1.2	Assuming populations are locally adapted or within their adaptive optimum	20
3.2	Methodological research	22
3.2.1	Forward genetics simulations for local and non-local adaptation scenarios	22
3.2.2	Empirical genomic offsets	23
3.2.3	Association between the causal and empirical genomic offset	24
3.2.4	Association between the empirical genomic offset and shifted fitness	26
3.2.5	Genomic offsets distribution for nonlocal adaptation scenarios	26
3.2.6	Identifying putatively adaptive loci	28
3.2.7	Selection of environmental covariates	31
3.2.8	Uncertainty quantification	34
3.2.9	Software implementation in Julia	38
3.2.10	Study case with simulated data from Mediterranean thyme	39
4	METHODS	45
4.1	Forward genetics simulations for local and non-local adaptation scenarios	45
4.2	Genomic offsets for local and non-local adaptation scenarios	47
4.3	Identifying putatively adaptive loci	48
4.3.1	Minimizing false positives or false negatives?	48

4.3.2	Randomly selected loci	49	
4.4	Selection of environmental covariates	50	
4.4.1	Including uncorrelated environmental variables		50
4.4.2	Including confounded variables with population structure	50	
4.5	Uncertainty quantification	51	
4.6	Software implementation in Julia	51	
4.7	Study case with simulated data from Mediterranean thyme	52	
4.7.1	Simulations	52	
4.7.2	Genomic offsets	55	
5	DISCUSSION AND FUTURE WORK	57	
5.1	Conceptual issues on genomic offsets	57	
5.2	Methodological research	58	
5.2.1	Genomic offset under different local and non-local adaptation scenarios	58	
5.2.2	Identifying putatively adaptive loci	59	
5.2.3	Selection of environmental covariates	60	
5.2.4	Uncertainty quantification	62	
5.2.5	Software implementation in Julia	63	
5.2.6	Study case with simulated data from Mediterranean thyme	64	
6	CONCLUSION	67	
	BIBLIOGRAPHY	69	
A	APPENDIX	75	
A.1	Local and non-local adaptation scenarios	75	
A.2	Uncertainty quantification	80	
A.3	Runtimes of the bootstrapped genomic offset	80	
A.4	Numerical issues with the geometric genomic offset	81	

LIST OF FIGURES

Figure 2.1	Conceptual figure of the genomic offset framework. 8
Figure 2.2	Diagram of the alternative genomic offset statistics. 13
Figure 3.1	Diagram of the foreign-local and home-away criteria. 18
Figure 3.2	Conceptual issues of the interpolation of a genotype \times environment association (GEA) model. 21
Figure 3.3	Geometric genomic offsets under different local and non-local adaptation scenarios. 25
Figure 3.4	Prediction of shifted fitness with different genomic offset statistics. 27
Figure 3.5	Prediction of shifted fitness with different sets of loci. 30
Figure 3.6	Prediction of shifted fitness with spurious genomic offsets. 32
Figure 3.7	Prediction of shifted fitness with increasing number of uncorrelated environmental covariates. 33
Figure 3.8	Prediction of shifted fitness with an environmental predictor confounded with population structure. 35
Figure 3.9	95% population rank bootstrapping confidence intervals. 37
Figure 3.10	Prediction of shifted fitness of thyme with historical environmental data. 42
Figure 3.11	Prediction of future fitness of thyme with historical environmental data. 43
Figure 4.1	Schematic diagram of the thyme simulations. 53
Figure A.1	risk of non-adaptedness (RONA) genomic offsets under different local and non-local adaptation scenarios. 76
Figure A.2	redundancy analysis (RDA) genomic offsets under different local and non-local adaptation scenarios. 77
Figure A.3	Gradient forest genomic offsets under different local and non-local adaptation scenarios. 78
Figure A.4	Distribution of geometric genomic offsets under different local and non-local adaptation scenarios. 79

Figure A.5	Spearman's correlation of bootstrapped genomic offsets.	80
Figure A.6	Runtime of bootstrapping genomic offsets.	81
Figure A.7	Numerical issues of the geometric genomic offsets with binary phenotypic traits.	82

ACRONYMS

SNP	single-nucleotide polymorphism
QTL	quantitative trait locus
RONA	risk of non-adaptedness
RDA	redundancy analysis
FDR	false-discovery ratio
PCA	principal component analysis
LFMM	latent factor mixed model
GEA	genotype \times environment association