UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea magistrale in Biologia Evoluzionistica

TESI DI LAUREA

Evolution and plasticity of gene expression in *Tribolium castaneum* exposed to a stressful environment

Relatore: Prof. Mauro Agostino Zordan
Dipartimento di Biologia

Correlatore: Prof. Alessandro Grapputo
Dipartimento di Biologia

Laureando: Giacomo Friso

ANNO ACCADEMICO 2024/2025

Table of contents

1) INTRODUCTION

1.1 Adaptation and Evolution

Natural selection is a mechanism that can alter the composition of a set of entities with respect to a given trait, consisting in the differential reproductive success of those entities, inherently linked to that trait (Fusco, 2019). Such demographic phenomenon, grounded in the principles of variation, hereditability and differential fitness (Lewontin, 1985), is the only one that can generate evolutionary change potentially resulting in an adaptation. More generally, the term adaptation refers both to the evolutionary process from which it originates and to the resulting adaptive trait. In Darwin's non-historical sense, such a trait represents a fitness advantage to individuals who possess it. In modern evolutionary biology, adaptation is defined more rigorously as a derived character (apomorphy) that has evolved through the direct action of natural selection (Coddinton, 1988). Under the Darwinian definition of adaptation fall the concept of exaptation. It consists on the cooptation of traits, previously modified and not by selection, for a new use (Gould & Vrba, 1982). Adding this new term helps to understand better the origins of adaptation: adaptive traits in non-evolutionary sense are the key step for adaptation in evolutionary sense. Jointly, the demographical nature of the process highlights the importance of the temporal variable. Following the recent worldwide ecological changes, the time becomes even more critical on the studying of evolutionary dynamics.

The environment provides the framework within which populations evolve. Acting as a selective filter, it shapes genetic variability with respect to survival and reproduction. Adaptation is therefore always relative to the environment in which organisms operate. For an adaptation to evolve, the environment must remain relatively constant - at least in the direction of selection. Genetically it means that advantageous variants arise and become (nearly) fixed, when similar selective pressures last for a sufficiently long period. However, in the Anthropocene, abrupt anthropogenic environmental changes pose a critical fate to population survival. Rapid warming and relate stressors, first and foremost, can exceed the evolutionary capacity of populations to respond adaptively (Bellard et al., 2012; Urban, 2015). In a context where the pace of ecological change could resemble that of past mass-extinction events, the study of adaptation becomes crucial. Investigate thoroughly the evolutionary dynamics of adaptation can help to define effective conservation strategies and so, the future of many species (Hoffmann & Sgrò, 2011). Moreover, adaptation is one of the foundational concepts of evolutionary biology, as the direct product of natural selection. From the birth of the modern genetic theory of evolution, it has long been the object of empirical and theoretical studies aimed at clarifying its genetic bases, temporal dynamics and structural constraints (Futuyma & Kirkpatrick, 2017).

Adaptation requires the presence of heritable genetic variation. Proceeding through natural selection the spread of advantageous alleles is favored. The speed and effectiveness of this process depend on the amount and nature of the genetic variance available for traits under selection, particularly the additive

component (Fisher, 1930; Lynch & Walsh, 1998). If sufficient additive genetic variance exists for a fitness-related trait, adaptation can occur in relatively short timeframes as demonstrated by evolutionary experiments on *Tribolium castaneum* (Koch et al., 2020b).

Understanding the genetic basis of traits under selection is therefore essential. For complex traits such as fitness, adaptation often involves intermediate molecular characters like gene expression, which links genotype to the measurable phenotype (Wray, 2007). Gene expression is highly sensitive to the environment and constitutes a key mechanism for the physiological response to external stresses. Numerous studies show that stable differences in transcriptional profiles are associated with adaptive divergence, both among natural populations and in laboratory-evolved lines (Fraser, 2011; Romero et al., 2012). To contribute effectively to evolutionary adaptation, such transcriptional modifications must be heritable and translate into significant effects on fitness. Experimental evidence demonstrates that the regulation of gene expression can evolve rapidly. In *T. castaneum*, changes in transcript levels emerged within 20 generations in response to extreme environmental conditions suggesting that molecular adaptation can accompany and support phenotypic adaptation (Koch et al., 2020; Koch & Guillaume 2020a). This makes transcriptomics a powerful tool for identifying biomarkers of adaptation and for tracking ongoing evolutionary dynamics.

1.2 GWAS and Adaptation

Genome-wide association studies (GWAS) are a fundamental tool for identifying genetic variants associated with complex phenotypic traits, through the large-scale analysis of millions of single-nucleotide polymorphisms (SNPs). These studies have pinpointed numerous loci linked to diseases, quantitative traits and physiological responses, proving essential also in the study of evolutionary adaptation (Visscher et al., 2017; Tam et al., 2019).

In the context of adaptation, GWAS have highlighted loci subject to positive selection in both human and animal populations. A classic example is adult lactase persistence, associated with regulatory variants of the *LCT* gene, which shows clear signs of selection in populations with a history of livestock farming (Enattah et al., 2002; Bersaglieri et al., 2004). Another emblematic case concerns adaptation to hypoxia in Tibetan populations, mediated by variants of the *EPAS1* gene, whose high frequency is compatible with an origin under positive selection (Yi et al., 2010).

However, traditional GWAS have important limitations in identifying the complete genetic basis of adaptation, especially for quantitative or highly polygenic traits. Most traits under selection are not controlled by a few large-effect variants but by many modest-effect variants whose detection often falls below conventional statistical-significance thresholds (Boyle et al., 2017).

Faced with this complexity, the infinitesimal model originally proposed by Fisher offers an alternative theoretical framework for understanding the inheritance of quantitative traits (Fisher, 1918). This model postulates that such traits are determined by a virtually infinite number of loci, each with an infinitesimally

small additive effect on the phenotype. Although simplified, the infinitesimal model has proved robust in explaining selection responses in many polygenic traits. Particularly it underscores how adaptation can proceed through small, coordinated changes in allele frequencies across numerous genes rather than through major selective sweeps at a few loci (Barton et al., 2017; Walsh & Lynch, 2018). This perspective clarifies why GWAS, designed to detect strong, localised signals, may struggle to capture the entire genetic basis of complex adaptations that conform more closely to a diffuse polygenic architecture.
Moreover, GWAS tend to identify associations rather than direct causality and are influenced by population structure, linkage disequilibrium and uncontrolled environmental factors (Tam et al., 2019).
To grasp the complexity of the genetic basis of adaptation, the concept of adaptive architecture has been developed. It expands the classic notion of genetic architecture by including aspects such as the distribution of allelic effects, pleiotropy, genetic redundancy and variability among replicated populations. This approach is particularly useful when traditional GWAS are insufficient to explain the observed adaptive response (Barghi et al., 2020; Boyle et al., 2017).
Evidence from experimental-evolution studies, such as those on *Drosophila simulans*, has shown that adaptation can occur through different allelic combinations. Similar phenotypic responses rest on distinct genetic foundations. This reveals notable genetic redundancy and low parallelism among replicated populations, highlighting the highly polygenic and contingent nature of many evolutionary trajectories (Barghi et al., 2019). In this context, integrating genomic data with functional, phenotypic and transcriptomic analyses becomes essential for a more comprehensive understanding of the mechanisms underlying adaptation, as also demonstrated by selection experiments on *Tribolium castaneum*, which underpin my work.

1.3 Differential Expression (DE)

Gene expression constitutes a “molecular phenotype”: the amount of transcript produced by each gene under given conditions. Easily to define thanks to the well-established RNA-seq, it arise from underlying complex regulatory mechanisms (Wray, 2007). When measurements are taken at whole-organism level, as for *Tribolium castaneum* in this project, the transcriptional profile represents the weighted sum of tissue-specific activities. Such approach provides a systemic picture of the response to selective pressures without pinpointing its cellular origin (Romero et al., 2012).
Heritable differences in average expression levels can become fixed within a few dozen generations when they confer a fitness advantage, making transcriptional regulation a rapid route to polygenic adaptation (Fraser, 2011). It follows that regulation can contribute to species divergence and differences among locally adapted population. Laboratory-evolution experiments have documented these processes in insects. *T. castaneum* exposed to hot-dry stress (Koch & Guillaume, 2020b) or *Drosophila* lines selected for high-temperature tolerance (Zhou et al., 2012) reveals functional convergence on stress-response genes, energy metabolism and proteostasis, and differentiation between groups as well.

Identifying genes potentially under selection relies on differential-expression (DE) analysis between adapted populations and reference lines. This method captures phenotypic divergence at the transcriptomic level, and under specific experimental design it helps to distinguish between plasticity and evolved expression shifts. Moreover, when combined with fitness data, DE patterns can be used to infer whether plastic responses are adaptive or maladaptive (Hodgins-Davis & Townsend, 2009). Lastly, integrating time-point analysis over generations allows to understand how short-term selection pressures are linked to long-term optimum expression levels. The difference between short and long-term gene expression changes will depend on trade-offs between the benefit of immediate stress responses and their long-term costs (Koch & Guillaume, 2020b).

Because genes operate within networks, co-expression analysis offers a structured view of regulatory re-organization. Approaches based on Weighted Gene Co-expression Network Analysis (WGCNA) identify modules of strongly correlated genes whose changes in connectivity or centrality reflect selection acting on shared functional processes (Langfelder & Horvath, 2008). In adapted *T. castaneum* lines, modules centred on heat-shock chaperones and water-stress responses show increased connectivity, suggesting that selection strengthens key regulatory hubs (Koch et al. 2024). Comparative analyses in other species reveal that the topologically most central genes in networks evolve under tighter constraints or experience stronger directional selection, depending on the ecological context (Jordan et al., 2019; Alvarez et al., 2015). These studies integrate the molecular and organismal levels, linking expression-network restructuring to measured fitness gains.

1.4 Phenotypic Plasticity, the Baldwin Effect and Counter-Gradient Variation

The mechanisms (time and modes) that enable populations to adapt cannot be assessed without taking into account plasticity. Plasticity refers to the ability of a genotype to produce different phenotypes in response to environmental variation without altering the DNA sequence (Pigliucci, 2001; West-Eberhard, 2003). While plasticity allows a rapid adjustment during the initial exposure to a novel or stressful environment, it does not constitute an evolutionary adaptation. However, it possess a crucial importance in the adaptive process: by changing the distribution of phenotypes on which selection can act, plasticity interferes with the process of evolution in a population (Koch & Guillaume, 2020b).

When the plastic response increases fitness of an individual, the profile response is adaptive, with two possible distinct outcomes. If evolution proceeds in the same direction of the induced plastic response, without reaching the phenotypic optimum of the plastic trait, it results in the Baldwin effect (Crispo, 2017). Described by Baldwin and initially referred to plasticity in behavior and learning (Baldwin, 1896), nowadays it refers broadly to plasticity. More generally referred to as cogradient variation (Conover et al., 2009), it potentially favours the most plastic individuals, causing evolved populations to exhibit a higher plasticity than their ancestors(Crispo, 2017; Lande, 2009). The second possible outcome is genetic assimilation: “a phenotypic character, which initially is produced only in

response to some environmental influence, becomes, through a process of selection, taken over by the genotype, so that it is formed even in the absence of the environmental influence which had at first been necessary" (Waddington, 1961). This is translate in a loss of plasticity together with a phenotypic change fixed and continuously expressed even in the ancestral environment. Such phenomenon has been independently described by Scmalhausen with the term "stabilizing selection"(Shcmalhausen, 1949), where selection acts both against mutation and changes induced by environments. The distinct theoretical framework of the two descriptions lays on the value of the initial environmental perturbations. While these would be always adaptive for Waddington, on the contrary they are mainly maladaptive for Shcmalhausen (Crispo, 2017). Interesting, both the response profiles of plasticity are found in the results of this thesis. It is important to note that the two outcomes can be seen as two sequential phases of the same process.

By contrast, if plasticity decrease the individual fitness, the plastic response is maladaptive. Therefore evolutionary trait response and plasticity are in opposite directions. While initial plasticity shifts the phenotype away from the optimum, selection acts in a compensatory sense, producing counter-gradient variation (Conover et al., 2009), or genetic compensation (Grether, 2005). In this case plasticity may provide only a short-term advantage in the first generations but becomes progressively costly as the genetic background evolves in the opposite direction (Ghalambor et al., 2007). Increasing the strength of selection on the phenotypes, this mechanism configures to maladaptive plasticity the potential role of promoting evolution. Empirical examples of this dynamic come from fish populations along latitudinal thermal gradients and from *Drosophila* lines selected for heat tolerance. Here, genetic adaptation restores expression patterns or physiological performance close to the ancestral state despite the persistence of the original plastic response (Ho & Zhang, 2018). It clearly demonstrates that plastic responses are not beneficial for long-term adaptation.

From an evolutionary perspective, phenotypic plasticity can therefore be a decisive advantage in the initial generations, preserving the demographic size required for selection to act. In the medium to long term, however, the energetic costs and trade-offs associated with continuous environmental regulation might make inheritable genetic control preferable: selection then tends to reduce (genetic assimilation) or reverse (counter-gradient variation) the plastic response (Koch & Guillaume, 2019). Distinguishing between these two outcomes requires analyses that combine the trajectory of ancestral plasticity, genetic change across multiple generations and matching fitness measurements.

1.5 The *Tribolium castaneum* Model

Tribolium castaneum (Herbst, 1797), commonly known as the red flour beetle, is a cosmopolitan tenebrionid beetle that lives in close association with humans and is particularly abundant in grain warehouses and flour products. Since the 1960s it has been recognised as an emerging model organism for ecological genetics thanks to its short life cycle (≈ 25–30 days at 33 °C), high fecundity, ease of laboratory culture under continuous darkness on a flour substrate, and the

possibility of maintaining large populations with minimal logistical effort (Sokoloff, 1972).
Scientific interest in *T. castaneum* grew further after its genome was sequenced (Tribolium Genome Sequencing Consortium, 2008). The compact genome (~ 200 Mb) showed extensive functional homology to the main developmental and stress-response pathways of other hemi- and holometabolous insects. Consolidated molecular tools, as systemic RNA interference or high-density genetic maps, make the species particularly well suited to functional-genetics and association studies (Brown et al., 2009). Moreover, its tolerance of a wide range of temperatures and humidities allows reproducible, controlled simulation of climate-change scenarios in the laboratory (Milutinović et al., 2013).
These characteristics, combined with the possibility of running multi-generational experimental-evolution studies on replicated populations, make *T. castaneum* an ideal system for investigating the genetic and transcriptional mechanisms of adaptation to complex environmental stresses.

1.6 Aims and Structure of the Present Work

This thesis builds on the experimental project of Koch & Guillaume (2020), extends its analysis with new quantitative approaches applied to two key generations of the *Tribolium castaneum* Cro1 population. The starting data comprised: VCF files with 358 k high-quality SNPs derived from RNA-seq on 822 individuals, gene-count matrices (featureCounts) for 17 078 genes, individual fitness measurements (number of offsprings produced) and functional metadata (hub genes and Parallelism values) taken from a dataset of the original study.
The work was guided by four sets of questions:

1. Which genomic variants contribute to fitness and/or environmental divergence? How do they change over generation?
   → eight linear-mixed GWAS (EMMAX - Kang et al., 2010) across two generations, with kinship correction (GRM) and false-discovery-rate control (Benjamini–Hochberg).
2. What is the transcriptional imprint of plasticity and evolution after 20 generations?
   → differential-expression analysis (edgeR + limma/voom) between conditions and generations, with $\log_2$FC thresholds calibrated to highlight biologically relevant variation.
3. Do topologically central genes in the networks (hub genes) display distinctive signals of association or selection?
   → enrichment tests (Fisher exact and permutation) on genes of SNP and transcript sets, integrating the WGCNA-derived hub-gene list from the original study.
4. How does the intensity of selection on expression vary across generations and gene categories? Is maladaptive or adaptive the evolutionary and plastic response over generations?
   → estimation of trait–fitness covariance on $\log_2$-CPM-normalised expression, compared with permuted distributions to detect positive or negative directional selection (Robertson, 1966; Price, 1970).

The results show that fitness, although a highly polygenic trait, exhibits a few associated SNPs in the first generation, detectable only with high statistical power (*n* = 485). By contrast, environment association already reveals hundreds of variants in G1, rising to more than 1 600 SNPs (902 genes) in G22, indicating a strong hot–dry selective gradient. DE analysis confirms the largely polygenic nature of the response: about 5 000 plastic genes in G1 and 2 300 in G22, with consistent enrichments for detoxification, carbohydrate metabolism and stress-response functions. Integrating GWAS and DE reveals limited overlap, suggesting distinct regulatory pathways for expression and structural polymorphisms. Nonetheless, certain hub genes emerge as both GWAS loci and transcriptional drivers, supporting the idea that network-central nodes are privileged targets of selection (Jordan et al., 2019).

By comparing DE results of the two generations, the general evolutionary dynamics are framed. Some genes, initially plastic, undergo evolution suggesting progressive genetic assimilation of part of the initial plastic response, in line with the Baldwin-effect model (Crispo, 2007). Nevertheless, the decrease over time of the amount of plastic genes suggests a counter-gradient variation. Plasticity might be expensive to maintain in the long-term. Overall, the study demonstrates how integrating GWAS, differential transcriptomics and network metrics allows the contributions of point variants, gene regulation and functional topology to be disentangled in the polygenic adaptation of *T. castaneum* to a hot–dry regime.

2) METHODS

The work carried out for this thesis stems from a research project by a PhD student at the University of Zurich started in 2018, which has produce the following scientific papers:

- Additive and mostly adaptive plastic responses of gene expression to multiplestress in *Tribolium castaneum* (Koch & Guillaume, 2020a).
- Restoring ancestral phenotypes is a general pattern in gene expression evolution during adaptation to new environments in *Tribolium castaneum* (Koch & Guillaume, 2020b).
- Genetic variance in fitness and its cross-sex covariance predict adaptation during experimental evolution (Koch et al., 2020)
- Gene expression evolution is predicted by stronger indirect selection at more pleiotropic genes (Koch et al., 2024).

My project comprises two main types of analyses, both performed entirely in RStudio: (1) genome-wide association studies (GWAS) and (2) differential-expression (DE) analyses of gene transcripts. No association studies had ever been run on these data sets, whereas part of the PhD work addressed DE analysis - albeit with a stricter statistical pipeline and a smaller sample size in generation 22.

2.1 Original work

The original scientific study is an experimental-evolution experiment in which an ancestral population of *Tribolium castaneum* was split among different laboratory environments and allowed to grow and reproduce for 22 generations. RNA-seq and fitness assays were carried out on the first and last generations to evaluate various aspects of adaptation and evolutionary change over time.

2.1.1 General design and methods

The experimental-evolution study was performed on the Cro1 strain of *T. castaneum*, originally isolated from a wild population in 2010. Before starting the experiment, all the individuals were adapted to laboratory condition for more than 20 generations under 33 °C, 70 % RH (relative humidity), continuous darkness, and a sterilized wheat-yeast flour medium. From this ancestral stock, six replicate lines were established for each of four environmental conditions: Control (CT, 33 °C / 70 % RH), Dry (D, 33 °C / 30 % RH), Hot (H, 37 °C / 70 % RH) and Hot–Dry (HD, 37 °C / 30 % RH); each founded with 120 pupae (60 males, 60 females). At every generation, 120 pupae were randomly chosen to start the next generation, thereby mimicking natural selection while keeping groups' size constant: in nature, individuals, depending on their fitness, do not contribute equally to the next generation. At generation 15 the six replicate lines within each treatment were intercrossed four times in equal proportions (20 individuals from each replicate line), resulting in four mixed lines with 120 individuals each. The mixing was to prevent loss of genetic diversity by gene drift and inbreeding, which might impede adaptation. At the final generation (22th) the total number of lines was 39 (one line in D became extinct).

To assess adaptation in each environment, two fitness assays were run: one on the initial generation (G1) and one on the final generation (G22). In both, virgin females were paired with unrelated males from the same environment for 24 h in vials containing 1 g of substrate; after seven days of oviposition, fresh medium was supplied and, after four weeks (CT/H) or five weeks (D/HD), the number of adult offspring per female was counted. In G1 a paternal half-sib design was also applied: each male mated with three females in three successive replicates, generating half-sib groups and three full-sib families, allowing estimation of additive-genetic variances.

At generation 20 all selection lines were returned for two generations (20 and 21) to common-garden conditions (33 °C, 70 % RH) to eliminate potential maternal or epigenetic effects arising from their divergent histories. Thus the descendants of every line began from the same environmental "starting point". From these common-garden beetles (G21) thirteen full-sib families per line were created (G22); the eggs from each family were divided among four replicated vials and assigned to the four experimental conditions (CT, D, H, HD). This split-brood phase ensured that, for every selection line, the impact of each stressor was evaluated on genotypes sharing the same maternal background. On reaching the pupal stage, four males and four females per family and treatment (G22) were isolated and kept under the target conditions until the fitness assay, which followed the same protocol as in G1. (See Figure 1)

Parallel to the fitness assays, 208 female samples per generation and condition were collected, frozen at –80 °C and later processed for RNA-seq. Each individual was homogenised in TRI Reagent and total RNA extracted with a commercial kit; only samples with RIN > 9 were retained. mRNA-seq libraries (500 ng RNA) were prepared with LEXOGEN kits, checked on a TapeStation, quantified by qPCR, normalised and pooled (33–36 samples per pool), then sequenced single-end (75 bp) on an Illumina NextSeq 500. Reads were quality-filtered and length-trimmed, mapped to the Tcas 3.30 reference genome with STAR v2.5, and counted at gene level with featureCounts within the SUSHI framework.
For variant discovery, the 822 RNA-seq samples from G1 and G22 were processed under GATK 4.2.3 best practices: read-group assignment, duplicate marking, splitting of CIGAR strings, HaplotypeCaller in GVCF mode, and joint calling via GenomicsDBImport / GenotypeGVCFs. High-quality biallelic SNPs were annotated with SnpEff and filtered for call rate ≥ 50 %, yielding a final VCF of ~ 358 k SNPs. Missing genotypes (~ 38 %) were imputed with Beagle 5.4, achieving > 98 % accuracy on a control data set.
Finally, allele frequency changes (AFC) between G1 and G22 were calculated separately for CT and HD (and analogously for D and H), polarising on the minor allele in G1. To quantify Parallelism, in each HD line at G22 the significance of AFC increases was tested against the reference distribution from CT via non-parametric bootstrap (10 000 replicates, $\alpha = 2.5$ %), and parallelism was defined as the number of HD lines in which the same SNP showed a significant change.
For a more detailed description of the methods, see Koch & Guillaume (2020a, 2020b).

2.1.2 Data sets and analyses

The experimental methodology described above generated several data sets and files that I used for my analyses. For both generations, my starting data consisted of fitness tables, VCF files (variant call format), featureCounts tables with gene-expression counts. Additionally, SNP and gene tables with diverse information such as the Parallelism value or gene type (hub genes) were used. Missing genotypes (~ 38 %) were imputed with Beagle 5.4, achieving > 98 % accuracy on a control data set.
Finally, allele frequency changes (AFC) between G1 and G22 were calculated separately for CT and HD (and analogously for D and H), polarising on the minor allele in G1. To quantify Parallelism, in each HD line at G22 the significance of AFC increases was tested against the reference distribution from CT via non-parametric bootstrap (10 000 replicates, $\alpha = 2.5$ %), and parallelism was defined as the number of HD lines in which the same SNP showed a significant change.
For a more detailed description of the methods, see Koch & Guillaume (2020a, 2020b).

2.2 GWAS

To study the association of genetic variants with fitness and with environment (selection regime), four GWAS were run for each generation (eight in total). Fitness was examined in three GWAS: first considering individuals from each

treatment separately, then pooling all individuals regardless of selection regime. The fourth GWAS addressed the association with environment, coding CT individuals as 0 and HD individuals as 1 to enable a direct contrast (Table 1).
Because the R-Studio script followed a similar structure in every case and used the same statistical approach, I wrote dedicated helper functions whenever possible, making the workflow smoother:

 # gc() - build a genotype matrix for emmax() from a .vcf file
 # grmx() - build a genetic relationship matrix for emmax()
 # adjpvgn() - build a vector with p-values from emmax() adjusted with 'fdr' method, and a dataframe with the SNP ID and the respective GENE of the significant adjusted p-values (-log10 pval > 1.3)
 # ftest() - perform a Fisher´s exact test on the gene category specified
 # permtest() - perform a permutation test on the gene category specified

Datasets	*N.SNP*	*N.Genes*	*N.samples*	*N.GWAS*	*GO*
FTCT.G1	0	0	244	1	-
FTHD.G1	0	0	242	2	-
FT.G1	19	13	485	3	no significant
ENV.G1	256	100	485	4	1 MF, 2 BP, 1CC
FTCT.G22	0	0	87	5	-
FTHD.G22	0	0	67	6	-
ENV.G22	1699	902	154	7	3 MF, 2 BP, 2CC
FT.G22	0	0	184	8	-

Table 1. List of the 8 GWAS performed with respectively number of significant SNP detected, genes associates to the SNP, sample size of the dataset, and number of GO categories.

The first part of the script handles filtering, checking and formatting the input so that each individual is paired with its correct genotype and fitness value. All VCF files had to be decoded with *gc()*, converting the classic VCF genotype codes into a numeric format suitable for analysis: 0|0 → 0, 0|1 or 1|0 → 1, 1|1 → 2.
To account for population structure and correct possible distortions due to relatedness or unobserved stratification, the GWAS were run with *emmax()*. This function implements a linear mixed model (LMM) in which the marker effect is treated as a fixed effect, whereas genetic relatedness among individuals is modelled as a random effect via a genomic-relationship matrix (GRM). The GRM, representing genetic similarity among individuals, is built from genotype data and captures the variance component not explained by the markers themselves. Unlike simpler approaches such as genomic control or principal-component covariates, EMMAX provides a more precise correction for population structure and, thanks to its pre-estimated variance approach, retains good computational performance on large datasets (Kang et al., 2010).

The kinship matrix used in the model was generated from the genotype matrix with the SNPRelate package in R (Zheng et al., 2012). Genotypes were converted to the binary GDS (Genomic Data Structure) format, which is optimised for handling genomic data efficiently. The function *snpgdsGRM()* with the "GCTA" method (Yang et al., 2011) calculated genetic similarity among individuals based on shared alleles at each locus. This approach is widely used in mixed-model frameworks to estimate the genetic component of phenotypic variance and is particularly suitable when relatedness exists among samples.
P-values returned by the GWAS models were then corrected for multiple testing with *adjpvgn()*, controlling the false discovery rate (FDR) by the Benjamini–Hochberg method, which is standard in high-dimensional genomic studies to limit the proportion of false positives among significant results (Benjamini & Hochberg, 1995).
Finally, for downstream comparisons and functional analyses, custom functions *ftest()* and *permtest()* apply classical statistics: Fisher's exact test, ideal for assessing enrichment in 2 × 2 contingency tables with low expected counts (Fisher 1922), and the permutation test, a non-parametric approach that estimates empirical significance via random re-sampling of the dataset. Such approach is especially useful when parametric assumptions are violated (Good 2005).
This thesis does not report the statistical analyses performed on the GWAS results with respect to the various gene categories (Parallelism, hub-gene, eQTL, ecc.) defined in the original project. Only plots are shown that relate GWAS outcomes to Parallelism values and logFC (Figure 2). Nevertheless, the general helper functions described above are reused in the subsequent analyses.

2.2.1 Visualisation and in-depth plotting

For every GWAS that yielded significant results, both static (Figure 3; Figure 4) and interactive (Figure 5; Figure 6) Manhattan plots were generated to show the genomic distribution of association signals. In parallel, a series of diagnostic and comparative plots was produced to assess the statistical reliability of the findings and to explore the impact of the models used.
A central role was played by quantile–quantile plots (QQ plots – Figure 3; Figure 4). The first QQ plot, based on the raw *p*-values returned by the EMMAX model, was used to evaluate the uncorrected performance of the GWAS model. In the absence of true associations, the *p*-values should follow a uniform distribution and therefore align with the diagonal. Early departures from the line can signal inflation due to insufficient correction for population structure, whereas tail deviations point to real associations.
A second QQ plot compares the distribution of FDR-adjusted $-\log_{10}$ *p*-values (Benjamini–Hochberg correction) with the theoretical uniform distribution of expected quantiles under the null hypothesis. Good alignment of the points along the diagonal indicates effective control of false positives, whereas deviations restricted to the upper tail reflect genuinely significant signals that remain after adjustment.

A third plot directly contrasts the $-\log_{10}$ *p*-values obtained with EMMAX against those from a general linear model (LM) with no structure correction. This comparison makes the effect of the adjustment clear: the LM typically shows stronger *p*-value inflation because relatedness among individuals is not accounted for, whereas EMMAX corrects for this confounding.
Finally, a scatter plot of the β-coefficients estimated by EMMAX versus those from the LM allowed a direct comparison of allelic-effect estimates. Marked differences for specific SNP indicate effect sizes distorted in the LM by population structure. Examining the direction and magnitude of effects helps to identify SNPs that are especially sensitive to kinship correction.
Taken together, these plots provide a deeper and more robust view of the GWAS results, going beyond the mere listing of significant SNPs. They aid interpretation of statistical significance, support validation of signals and strengthen confidence in the conclusions drawn from the association analyses.

2.3 Gene Ontology

To support the biological interpretation of the GWAS findings, the genes of interest were functionally characterised. Descriptive information for each gene was retrieved from the Ensembl Metazoa portal (Howe et al., 2020) using the annotation available for *Tribolium castaneum*. An enrichment analysis of Gene Ontology (GO) terms was then carried out with the g:Profiler tool (Raudvere et al., 2019), focusing solely on the three main categories: Molecular Function (MF), Biological Process (BP) and Cellular Component (CC). Significantly enriched GO terms were identified with the hypergeometric test and multiple-testing correction by the g:SCS method, which is tailored to hierarchical ontologies.
The same GO-enrichment procedure was applied to the results of the DE analysis described in the following section.

2.4 Differential-expression analysis (DE)

The differential-expression (DE) analysis was carried out entirely in R, combining the edgeR and limma packages. The starting point was a DGEList object built from the raw RNA-seq counts, to which an initial filter was applied to remove poorly expressed genes: only genes with an expression value greater than 1 count-per-million (CPM) in at least two samples were retained.
Library normalisation was performed with the TMM (Trimmed Mean of M-values) method implemented in edgeR, in order to correct for differences in sequencing depth among samples (Robinson et al. 2010).
To adapt the data to limma's linear-model framework, the *voom()* function (Law et al. 2014) was applied; this transforms the counts to $\log_2$-CPM, estimates a mean-variance trend for each gene and assigns a precision weight to every observation. This step is essential to ensure the appropriateness of the linear model for RNA-seq data.

Figure 2. Parallelism significance plot for Generation 22 (GWAS 7) on the left and Volcano plot of absolute value of logFC for Generation 22 on the right. Vertical axis of each plot defines the log10 adjusted P-values starting from the significance threshold of negative log10 1,3 (equal to 0,05). Each SNP is associated with the color of the respective chromosome visualized in the lateral label. Information with SNP and gene ID are added to the SNP with higher values. Both Parallelism and logFC values derived from the original work on Tribolium castaneum.

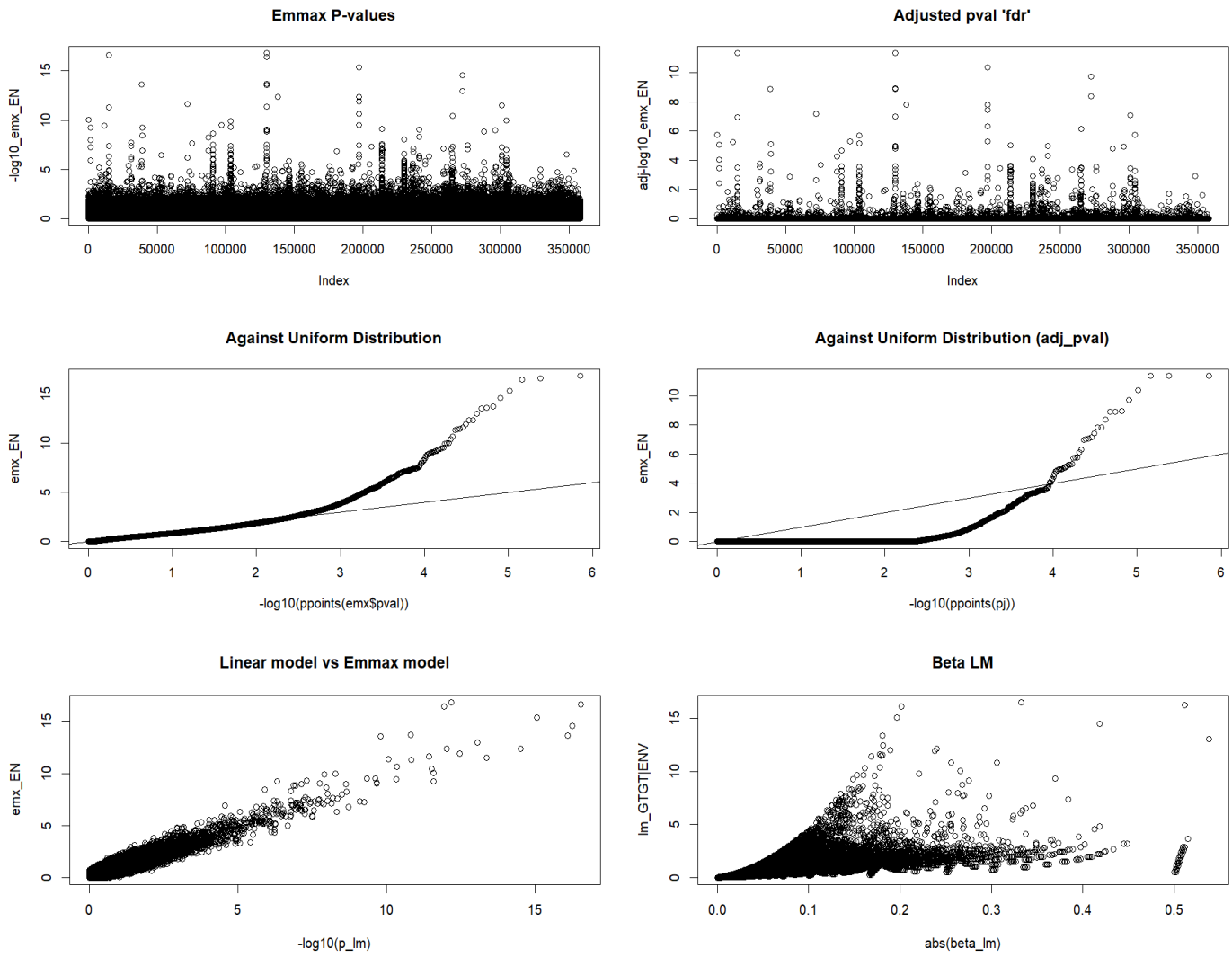


Figure 3. Static Manhattan plots and diagnostic plot of Generation 1 (GWAS 4). Starting from the top left and proceeding to the right:

- Manhattan plot without correction of SNP
- Manhattan plot with adjusted SNP for false discovering rate (BH)
- QQ plot of uncorrected SNP
- QQ plot of adjusted SNP
- Comparative scatter plot of –log10 P-values between EMMAX model and LM model
- Scatter plot of the β-coefficients estimated by EMMAX versus those from the LM

Only a handful of SNPs pass the significance threshold, and both QQ-plots hug the diagonal apart from a short upper-tail, indicating limited inflation and few true associations. The LM-vs-EMMAX scatter clearly shows how structure correction lowers the –log$_{10}$ p values, preventing false positives.

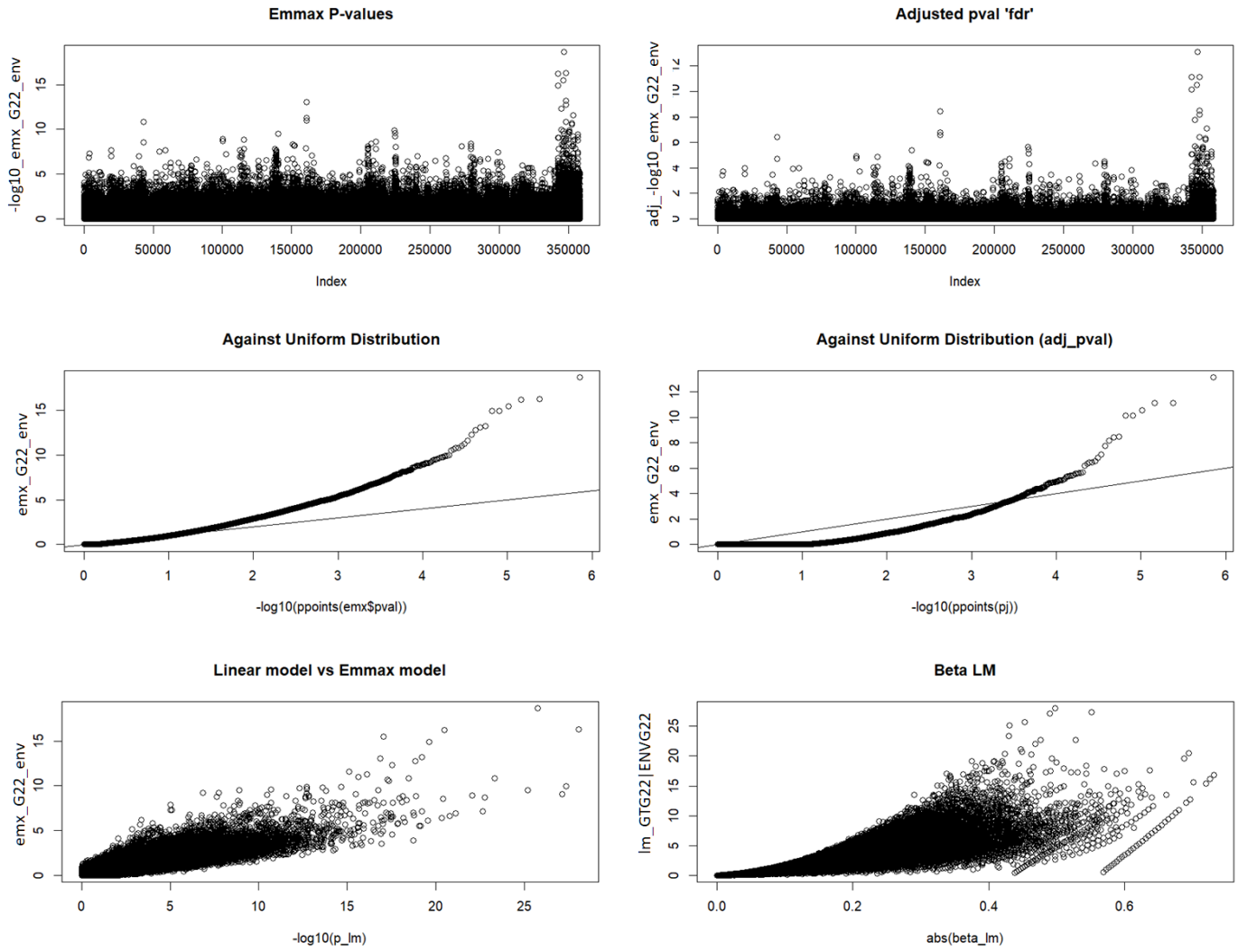


Figure 4. Static Manhattan plots and diagnostic plot of Generation 22 (GWAS 7). Starting from the top left and proceeding to the right:

- Manhattan plot without correction of SNP
- Manhattan plot with adjusted SNP for false discovering rate (BH)
- QQ plot of uncorrected SNP
- QQ plot of adjusted SNP
- Comparative scatter plot of –log10 P-values between EMMAX model and LM model
- Scatter plot of the β-coefficients estimated by EMMAX versus those from the LM

Numerous genome-wide hits emerge, with QQ-plots displaying a long right-hand tail that reflects extensive divergence after 20 generations. Yet the LM-vs-EMMAX panel reveals pronounced inflation in the uncorrected model, underscoring the importance of the kinship correction even when many genuine signals are present.

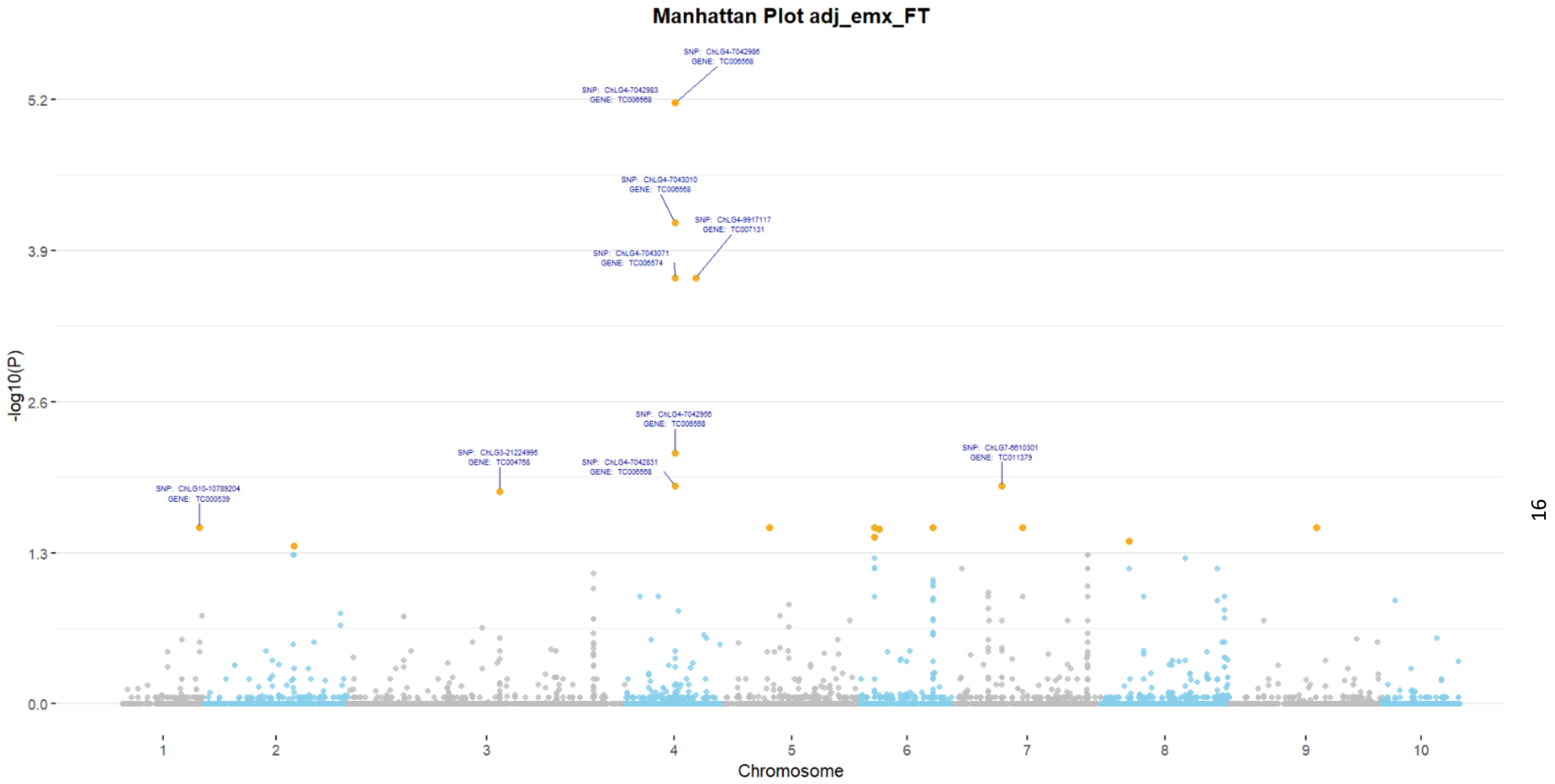


Figure 5. Frame of the interactive Manhattan plot of Generation 1 (GWAS 3). 19 adjusted significant SNP are founded from 485 samples, visualized with the orange color. In the horizontal axis chromosomes are represented with alternated colors of gray and blue. P-value on the vertical axis is shown as negative log10. Information with SNP and gene ID are added to the SNP with higher significance.

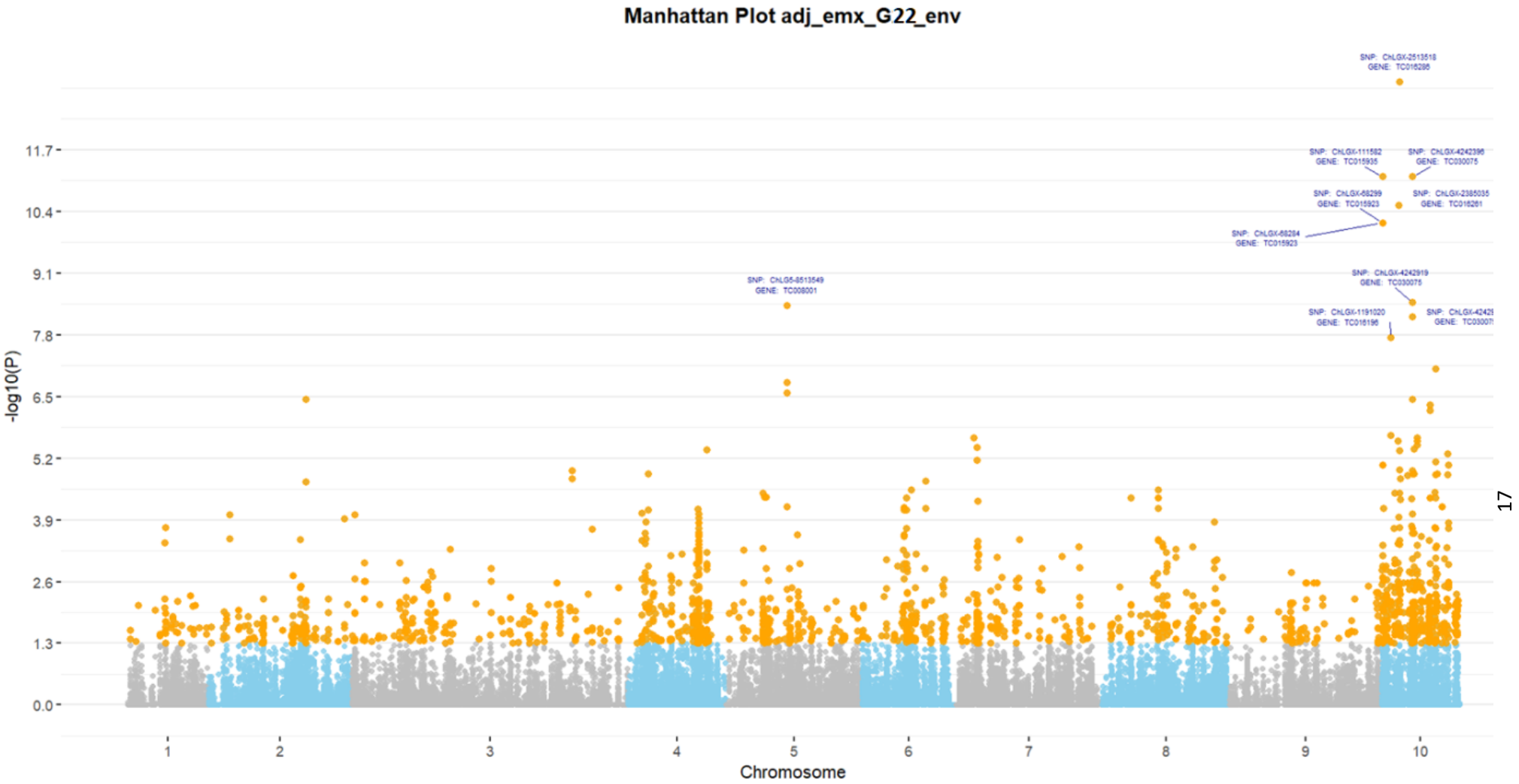


Figure 6. Frame of the interactive Manhattan plot of Generation 22 (GWAS 7). 1699 adjusted significant SNP are founded from 154 samples, visualized with the orange color. In the horizontal axis chromosomes are represented with alternated colors of gray and blue. P-value on the vertical axis is shown as negative log10. Information with SNP and gene ID are added to the SNP with higher significance.

A design matrix (model.matrix) was then defined that included sequencing batch (batch) and experimental group (group) effects, without an intercept. For generation 22, where multiple individuals belonged to the same selection lines, the intra-line correlation was estimated with *duplicateCorrelation()* in limma, using the factor line as the blocking variable. This procedure models the non-independence among individuals from the same line as a random effect, analogous to a mixed model (Law et al. 2014; Ritchie et al. 2015).
The final model (lmFit) was fitted to the voom-transformed, precision-weighted data (V2), incorporating the estimated correlation among samples from the same line as a parameter in the linear model.
Differentially expressed genes were identified with *eBayes()*, which applies Bayesian moderation to the gene-wise variance estimates. Genes were deemed significant if they showed an FDR ≤ 5 %, calculated with the Benjamini–Hochberg procedure (Benjamini & Hochberg 1995).
Plastic genes in generation 1 were obtained from the contrast between CT and HD individuals in G1.
For generation 22, plastic genes were defined as those differing between HD.CT individuals (beetles adapted for 20 generations to HD but placed in CT in the last generation) and HD.HD individuals (kept in HD throughout).
Genes showing evolutionary change were extracted from the contrast between CT.HD and HD.HD groups.
For these three main gene sets an arbitrary filter was applied to retain only biologically relevant changes: genes were kept only if they showed an expression difference greater than 10 % [abs(logFC) > 0.136].
From these, sub-categories later used in the analyses were defined (Table 2).

Code	Category name	Operational definition
MPC1.50	Most 50 % plastic genes	Genes plastic in G1 with $\|\log_2 FC\| > 0.585$ (≥ 50 % higher / lower expression)
MPC1	Most plastic genes	Genes plastic in G1 with $\|\log_2 FC\| > 1$ (≥ 100 % higher / lower expression)
NWPC	New plastic genes	Genes plastic in G22 that were not plastic in G1
LSPC	Lost-plasticity genes	Genes plastic in G1 but not plastic in G22
PRPC	Preserved plasticity genes	Genes plastic in both G1 and G22 (intersection)
PCEC	Evolved plastic genes	Genes that evolved in G22 and were also plastic in G1 (intersection)

Table 2. Sub-categories definition list.

For both generations (G1 and G22), the intensity of selection on gene expression was then quantified. Starting from the TMM-normalised counts in edgeR, the data were transformed to $\log_2$-CPM and cleaned of batch effects with limma's *removeBatchEffect()*. For each experimental condition (CT and HD) the

corresponding expression matrix was extracted. For every gene, the covariance between expression level and relative fitness - defined as the number of adults produced divided by the group mean - was computed, following the classic framework that links selection to the covariance between trait and fitness (Robertson 1966; Price 1970). In parallel, Pearson's correlation between expression and relative fitness was calculated, providing a standardised index of selection intensity that can be compared across generations and conditions.
To detect signals of adaptive or maladaptive selection, a permutation test was implemented to compare the distribution of gene-wise correlations for up-regulated or down-regulated genes against the rest of the genome. The custom function *sel.resp()* evaluated, for each group, whether selection intensity differed significantly: the difference between the mean correlations of the two groups was contrasted with the distribution obtained from 10 000 random permutations of the labels. Up-regulated genes were tested for significantly more positive selection, down-regulated genes for more negative selection.

2.5 Hub genes

The hub genes used in the present analyses are taken directly from the original study, which applied a weighted gene co-expression network analysis (WGCNA) to the G1 data under control conditions (CT). In that approach, genes were clustered into modules on the basis of the correlation among their expression levels, thereby identifying groups of highly co-expressed, functionally connected genes. Within each module, intramodular connectivity was calculated with WGCNA's *intramodularConnectivity()* function, and the ten genes with the highest connectivity were designated as hub genes. These genes represent the central nodes of the transcriptional modules and are considered especially important for regulating biological processes. The genes identified in this way were subsequently used as a functional category in the analyses of this thesis.

2.6 Statistical tests and DE plots

To explore significant relationships among the various categories described above, statistical tests based on Fisher's exact test and permutation testing were performed. These analyses, implemented with the generic helper functions created for the GWAS step - *ftest()* and *permtest()* - also allowed us to examine how the intensity of selection changed across generations.
Finally, as for the GWAS, a series of plots were generated to visualise and better understand the results.

3) RESULTS

3.1 GWAS

3.1.1 Fitness is a complex trait

The genetic-association analyses for fitness yielded just one significant result, and, as expected, this came from the group with the largest sample size.

No significant variants emerged in the control group of generation 1 (FTCT.G1 - GWAS 1, 244 individuals) or in its Hot-Dry counterpart (FTHD.G1 - GWAS 2, 242 individuals).

Likewise, neither the control group (FTCT.G22 - GWAS 3, 87 individuals) nor the Hot-Dry group (FTHD.G22 – GWAS 4, 67 individuals) of generation 22 produced any significant variants. Combining the two G22 subsets (FT.G22 – GWAS 8, 184 individuals) still yielded no hits.

By contrast, the full data set for generation 1 (FT.G1 – GWAS 6, 485 individuals) revealed 19 significant SNPs mapping to 13 distinct genes.

These findings suggest that significant associations between variants and fitness can be detected only when statistical power is sufficient, that is, when sample size is large enough (Table 1).

3.1.2 Gene annotation for fitness-associated variants

Using the genome-centric Ensembl Metazoa portal, eight of the thirteen genes were functionally annotated (Table 3).

Gene ID	Protein name	Description
TC000516	Pathogenesis-related protein 5-like	Involved in immune response; expressed in larvae, pupae and adults, with highest levels in adults.
TC002663	Leucine-rich repeat-containing protein 15-like	Probably participates in cell–cell interactions and extracellular signalling; in vertebrates, LRR proteins are linked to cell adhesion and signalling.
TC007136	Muscle LIM protein Mlp84B-like	Associated with muscle structure; in Drosophila, Mlp84B is essential for cardiac function and Z-disc organisation.
TC009706	Heat-shock protein 68a	Heat-shock protein involved in thermal-stress response; expressed at several developmental stages.
TC010678	Dolichyl-diphosphooligosaccharide–protein glycosyl-transferase 48 kDa subunit	Subunit of the oligosaccharyl-transferase complex; catalyses N-linked glycosylation of proteins in the endoplasmic reticulum.

TC011000	Cathepsin L	Lysosomal protease involved in protein degradation; in *T. castaneum* it has been examined for potential enzyme-replacement therapy applications.
TC015197	Centromere protein J-like	Probably involved in centromere assembly and centriole elongation during cell division; in other organisms CENP-J is essential for centriole formation.
TC015480	Equilibrative nucleoside transporter 3-like	Nucleoside transporter that mediates equilibrative transport across cell membranes; in mammals ENT3 participates in nucleoside transport and nucleotide homeostasis.

Table 3. Informative table of eight genes functionally annotated derived from 19 SNPs of GWAS 6. Five genes of the thirteen were not characterized.

3.1.3 GO enrichment for fitness-associated variants

A Gene Ontology enrichment analysis with g:Profiler was performed on the set of thirteen genes to identify over-represented biological functions and processes. No GO terms were significantly enriched: the genes do not share biological roles or pathways beyond what would be expected by chance.

3.1.4 Variants associated with environment

In contrast to the fitness-based GWAS, the association studies contrasting the two environments produced many variants significantly linked to environment in both the first and the last generation (Table 1).

- *Generation 1 (ENV.G1 - GWAS 5):*
  From 485 individuals - 243 in CT and 242 in HD - 256 significant SNPs were detected, mapping to 100 genes. This indicates that selective or evolutionary processes were already acting from the very first generation.
- *Generation 22 (ENV.G22 - GWAS 7):*
  Despite the smaller sample size (184 individuals: 102 CT.XX and 82 HD.XX), 1 699 significant variants were identified, belonging to 902 genes. Evidently, evolutionary change over the 20 generations produced a marked divergence between the two groups, reflected in the large number of environment-associated variants.

3.1.5 GO enrichment for environment-associated variants

Gene-ontology enrichment yielded significant results for both gene sets.

- *Generation 1:*
  76 of the 100 genes were annotated and used in the analysis. Five

enriched GO profiles were identified. One molecular-function profile, three biological-process profiles and one cellular-component profile (Table 4).

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:0004553	Hydrolase activity, hydrolyzingO-glycosyl compounds	Molecular Function	7.341×10^{-6}	Enzymatic activity involved in the degradation of complex carbohydrates such as chitin.
2	GO:0042026	Protein refolding	Biological Process	1.412×10^{-4}	Proper protein folding by chaperones in response to cellular stress.
3	GO:0006032	Chitin catabolic process	Biological Process	3.123×10^{-3}	Degradation of chitin, essential for remodeling during molting.
4	GO:0005975	Carbohydrate metabolic process	Biological Process	2.981×10^{-2}	General carbohydrate metabolic processes that provide energy and structural components.
5	GO:0005576	Extracellular region	Cellular Component	3.932×10^{-3}	Extracellular localisation of proteins, often linked to defence.

Table 4. GO profiles from GWAS 4.

- *Generation 22:*
  Of the 902 genes, 787 were annotated and included in the analysis. Three molecular-function profiles, two biological-process profiles and two cellular-component profiles were enriched (Table 5).

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:0016740	Transferase activity	Molecular Function	1.373×10^{-2}	Enzymes that transfer functional groups between

					molecules (e.g. phosphates, methyl groups).
2	GO:0036094	Small molecule binding	Molecular Function	2.186×10^{-2}	Proteins' ability to bind small molecules (e.g. substrates, cofactors).
3	GO:0000166	Nucleotide binding	Molecular Function	2.341×10^{-2}	Interaction with nucleotides (e.g. ATP, GTP), often linked to enzymatic activity.
4	GO:0006091	Generation of precursor metabolites and energy	Biological Process	1.560×10^{-2}	Production of precursor metabolites and energy (e.g. glycolysis, cellular respiration).
5	GO:0043412	Macromolecule modification	Biological Process	4.605×10^{-2}	Post-translational modifications of proteins or other macromolecules.
6	GO:0005622	Intracellular anatomical structure	Cellular Component	6.881×10^{-17}	Structures inside the cell (e.g. organelles, cytoskeleton, compartments).
7	GO:0032991	Protein-containing complex	Cellular Component	1.697×10^{-3}	Molecular complexes composed of multiple proteins (e.g. ribosomes, proteasomes).

Table 5. GO profiles from GWAS 7.

Among these results, one profile is strikingly significant (ID 6, adj. $p = 6.881 \times 10^{-17}$) and concerns variants in genes involved in intracellular anatomical structure. This suggests that long-term adaptation to a stressful environment may proceed, directly or indirectly, through modifications of intracellular architecture.

3.1.6 Parallelism and logFC values of genes linked to significant SNPs

Additional insight into the significant variants and their genes was obtained by using results from the original study. By extracting the Parallelism value for each

SNP and the logFC for its gene, one can appreciate the difficulty of drawing firm conclusions - reflecting the inherent complexity of biological systems.

- *GWAS 3 (generation 1, 19 SNPs):*
  Most SNPs occur in only one of the six replicate lines; just three variants appear in two lines. The logFC values of the corresponding thirteen genes lie mostly between 0 and 0.2 (0 - 15 % change in expression between CT and HD); only two genes show larger shifts of ~ 40 %.

- *GWAS 4 (generation 1, 256 SNPs):*
  Nine variants are shared by three lines, three by four lines, and one by all four lines; the rest occur in one or two lines. Of the 100 genes, most have logFC < 0.3; only sixteen genes show logFC between 0.4 and 0.55 (≈ 32 – 45 % change).

- *GWAS 7 (generation 22, 1 699 SNPs):*
  As before, higher Parallelism corresponds to fewer variants: 17 SNPs are found in four of the six lines and three in five lines. LogFC values for the 902 genes extend to higher levels than in earlier data sets: twelve genes change by ~ 1.3 logFC (≈ 146 %), and three genes by ~ 2.6 logFC (≈ 506 %); the remaining genes fall between 0 and 1 logFC.

3.2 DE

3.2.1 Generation 1 - Plasticity and GO

The DE analysis contrasting CT and HD in generation 1 (485 individuals in total) identified 5 037 genes as significantly differentially expressed, and thus plastic, out of the 11 216 annotated *Tribolium* genes.

A GO-enrichment test with g:Profiler, based on 3 905 annotated genes out of the 5 037, yielded 22 significantly over-represented profiles (Table 6). The most significant were:

- oxidoreductase activity(MF, adj.p = 3.890×10^{-13})
- serine-type peptidase activity(MF, adj.p = 5.806×10^{-9})
- organic acid metabolic process(BP, adj.p = 9.901×10^{-13})
- extracellular region(CC, adj.p = 2.555×10^{-14})

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:0016491	oxidoreductase activity	GO:MF	3.890×10^{-13}	Enzymes that catalyse redox reactions, important in metabolism and energy production
2	GO:0008236	serine-type peptidase activity	GO:MF	5.806×10^{-9}	Protein degradation, tissue remodelling
3	GO:0005506	iron ion binding	GO:MF	1.675×10^{-5}	Involved in iron transport and

					metabolism
4	GO:0022857	transmembrane transporter activity	GO:MF	3.105×10^{-5}	Active or passive transport of molecules across membranes
5	GO:0008061	chitin binding	GO:MF	3.706×10^{-4}	Involved in chitin degradation or interaction
6	GO:0046906	tetrapyrrole binding	GO:MF	7.073×10^{-4}	Interaction with cofactors such as haem or chlorophyll
7	GO:0008484	sulfuric ester hydrolase activity	GO:MF	9.123×10^{-4}	Hydrolysis of sulphate esters; detoxification
8	GO:0004568	chitinase activity	GO:MF	2.087×10^{-3}	Chitin degradation, defence or digestion
9	GO:0004180	carboxypeptidase activity	GO:MF	3.633×10^{-3}	Removal of terminal amino acids from proteins
10	GO:0004315	3-oxoacyl-[ACP] synthase activity	GO:MF	1.124×10^{-2}	Fatty-acid biosynthesis
11	GO:0004312	fatty acid synthase activity	GO:MF	1.124×10^{-2}	Long-chain fatty-acid production
12	GO:0016418	S-acetyltransferase activity	GO:MF	3.944×10^{-2}	Transfer of acetyl groups; enzymatic regulation
13	GO:0006082	organic acid metabolic process	GO:BP	9.901×10^{-13}	Metabolism of organic acids such as lactate and pyruvate
14	GO:0005975	carbohydrate metabolic process	GO:BP	6.187×10^{-6}	Digestion, storage and utilisation of carbohydrates
15	GO:0006022	aminoglycan metabolic process	GO:BP	2.249×10^{-5}	Metabolism of glycosaminoglycans (structural)
16	GO:0055085	transmembrane transport	GO:BP	5.795×10^{-5}	Movement of molecules into or out of the cell
17	GO:0043603	amide metabolic process	GO:BP	4.869×10^{-4}	Metabolism of amides (e.g. proteins, peptides)
1	GO:0009060	aerobic	GO:BP	3.348×10^{-2}	ATP production

8		respiration			using oxygen
19	GO:0044248	cellular catabolic process	GO:BP	3.391×10^{-2}	Controlled degradation of cellular components
20	GO:0006508	proteolysis	GO:BP	3.669×10^{-2}	Protein degradation, proteome renewal
21	GO:0005576	extracellular region	GO:CC	2.555×10^{-14}	Secreted proteins or proteins located outside the cell
22	GO:0005688	U6 snRNP	GO:CC	1.948×10^{-2}	Spliceosome component involved in RNA splicing

Table 6. GO profiles of plastic genes of generation 1 (PCG1).

3.2.2 Filtering plastic genes (generation 1)

Filtering the plastic genes with two arbitrary logFC thresholds allowed a more realistic definition of plastic genes in generation 1:

- *MPC1.50* comprises the plastic genes whose absolute logFC is greater than 0.585 (≈ 50 % change), for a total of 766 genes.
- *MPC1* comprises the plastic genes whose absolute logFC is greater than 1 (≈ 100 % change), for a total of 325 genes.

The GO-enrichment analysis for MPC1.50 annotated 573 of the 766 genes and identified 14 over-represented profiles (Table 7).

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:0016787	hydrolase activity	GO:MF	2.056×10^{-16}	Enzymes that catalyse the hydrolysis of chemical bonds
2	GO:0008061	chitin binding	GO:MF	2.495×10^{-8}	Interaction with chitin, relevant to cuticular structure
3	GO:0045735	nutrient reservoir activity	GO:MF	1.940×10^{-3}	Proteins that store nutrients, e.g. ferritin
4	GO:0004312	fatty acid synthase activity	GO:MF	2.942×10^{-2}	Synthesis of long-chain fatty acids
5	GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthase activity	GO:MF	2.942×10^{-2}	Catalyses steps in fatty-acid biosynthesis
6	GO:0003993	acid	GO:MF	3.841×10^{-2}	Removal of

		phosphatase activity			phosphate groups in an acidic environment
7	GO:0005975	carbohydrate metabolic process	GO:BP	2.453×10^{-13}	Sugar metabolism
8	GO:0006022	aminoglycan metabolic process	GO:BP	8.460×10^{-9}	Breakdown or synthesis of nitrogen-containing glycans
9	GO:0006508	proteolysis	GO:BP	3.564×10^{-5}	Protein degradation for control or recycling
10	GO:0042026	protein refolding	GO:BP	1.059×10^{-4}	Restoration of the native structure of denatured proteins
11	GO:0016042	lipid catabolic process	GO:BP	1.393×10^{-4}	Lipid breakdown for energy or renewal
12	GO:0009408	response to heat	GO:BP	1.188×10^{-2}	Activation of genes/proteins in response to heat stress
13	GO:0006820	monoatomic anion transport	GO:BP	2.708×10^{-2}	Movement of inorganic anions across membranes
14	GO:0005576	extracellular region	GO:CC	5.971×10^{-2}	Extracellular zone where secreted enzymes or proteins act

Table 7. GO profiles of most plastic genes (50% change) of generation 1.

The GO-enrichment analysis for MPC1 annotated 186 of the 325 genes and identified 9 over-represented profiles (Table 8).

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:0016798	hydrolase activity, acting on glycosyl bonds	GO:MF	1.444×10^{-16}	Hydrolytic breakdown of sugars linked by glycosidic bonds
2	GO:0008061	chitin binding	GO:MF	1.221×10^{-11}	Interaction with chitin, important for cuticular structure
3	GO:0008236	serine-type	GO:MF	8.391×10^{-5}	Enzymes that

		peptidase activity			degrade proteins using a serine residue in the active site
4	GO:0006022	aminoglycan metabolic process	GO:BP	7.587×10^{-13}	Metabolism of nitrogen-containing carbohydrates (e.g. chitin)
5	GO:0005975	carbohydrate metabolic process	GO:BP	1.438×10^{-10}	Transformation and utilisation of sugars
6	GO:0042026	protein refolding	GO:BP	1.255×10^{-7}	Restoration of proper protein folding
7	GO:0009408	response to heat	GO:BP	4.235×10^{-4}	Cellular response to heat stress
8	GO:0006508	proteolysis	GO:BP	3.303×10^{-2}	Protein degradation involved in regulation and renewal
9	GO:0005576	extracellular region	GO:CC	3.081×10^{-19}	Extracellular localisation, often the site of secreted enzymes

Table 8. GO profiles of most plastic genes (1000% change) of generation 1.

3.2.3 Generation 22 – Plasticity, evolutionary change and GO

In generation 22 the DE analysis detected 2 282 plastic genes in a sample of 56 individuals and 731 genes that underwent evolutionary change in a sample of 54 individuals.

For the plastic genes, 1 882 of 2 282 were annotated and 18 over-represented GO profiles were found (Table 9).

ID	ID GO	Term ID	Source	Pval-adj	Biological meaning
1	GO:1901505	carbohydrate derivative transmembrane transporter activity	GO:MF	1.659×10^{-5}	Transport of carbohydrate derivatives across membranes
2	GO:0015932	nucleobase-containing compound transmembrane	GO:MF	2.571×10^{-4}	Transport of nitrogenous bases, nucleosides and nucleotides

		transporter activity			
3	GO:0016787	hydrolase activity	GO:MF	8.917×10^{-4}	Enzymes that catalyse hydrolytic reactions
4	GO:0015605	organophosphate ester transmembrane transporter activity	GO:MF	2.157×10^{-3}	Transport of organophosphate esters across membranes
5	GO:0140662	ATP-dependent protein folding chaperone	GO:MF	3.043×10^{-3}	Proteins that assist folding by using ATP
6	GO:0051082	unfolded protein binding	GO:MF	6.178×10^{-3}	Binding to unfolded proteins to stabilise or refold them
7	GO:0004568	chitinase activity	GO:MF	1.554×10^{-2}	Chitin degradation; defence or digestion
8	GO:1901264	carbohydrate derivative transport	GO:BP	2.443×10^{-5}	Intracellular or extracellular movement of carbohydrate derivatives
9	GO:0042026	protein refolding	GO:BP	3.405×10^{-4}	Reactivation of denatured proteins
10	GO:0045333	cellular respiration	GO:BP	1.273×10^{-3}	Production of cellular energy through respiration
11	GO:0009408	response to heat	GO:BP	2.550×10^{-3}	Cellular or systemic responses to heat stress
12	GO:0043603	amide metabolic process	GO:BP	9.240×10^{-3}	Metabolism of molecules containing amide bonds
13	GO:1901072	glucosamine-containing compound catabolic process	GO:BP	9.486×10^{-3}	Degradation of glucosamine-containing compounds
14	GO:0009056	catabolic process	GO:BP	3.302×10^{-2}	Breakdown processes for complex

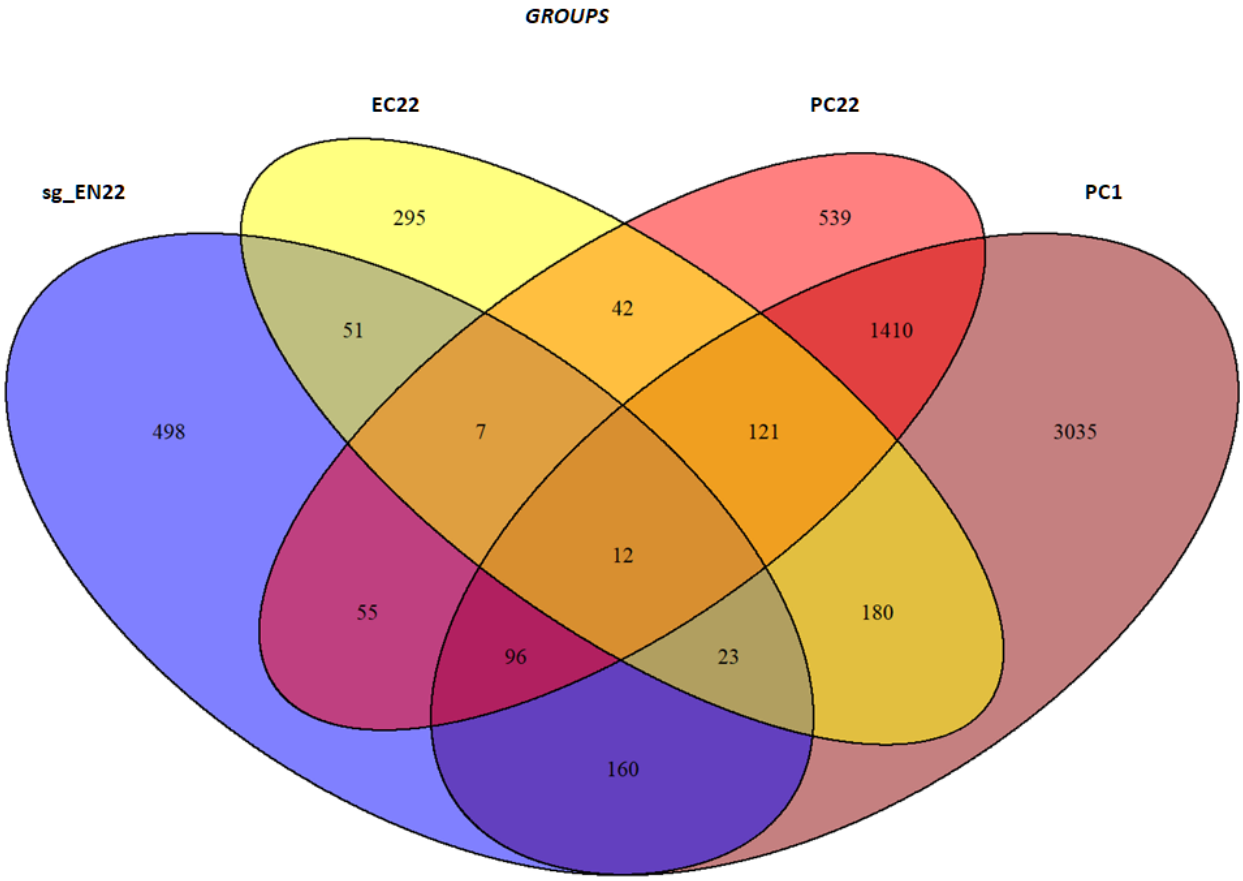
					substances
15	GO:0019673	GDP-mannose metabolic process	GO:BP	4.955×10^{-2}	GDP-mannose metabolism, involved in glycosylation
16	GO:1905369	endopeptidase complex	GO:CC	7.535×10^{-4}	Enzyme complexes for intracellular protein degradation
17	GO:0005737	cytoplasm	GO:CC	1.580×10^{-3}	Cellular compartment where most metabolism occurs
18	GO:0140534	endoplasmic reticulum protein-containing complex	GO:CC	5.469×10^{-3}	Protein complexes resident in the endoplasmic reticulum

Table 9. GO profiles of plastic genes of generation 22.

For the genes with evolutionary change, 581 of 731 were annotated and 5 over-represented GO profiles were found (Table 10).

ID	ID GO	Term ID	Source	Pval -adj	Biological meaning
1	GO:0003824	catalytic activity	GO:MF	6.449×10^{-5}	Enzymatic activity that accelerates biochemical reactions
2	GO:0043167	ion binding	GO:MF	2.844×10^{-3}	Ability of proteins to bind ions (e.g. Ca^{2+}, Fe^{2+})
3	GO:0005524	ATP binding	GO:MF	2.175×10^{-2}	Interaction with ATP, essential for energy supply and signalling
4	GO:1903047	mitotic cell cycle process	GO:BP	7.717×10^{-3}	Processes involved in mitotic cell division
5	GO:0006270	DNA replication initiation	GO:BP	4.127×10^{-2}	Initiation of DNA replication prior to division

Table 10. GO profiles of evolved genes of generation 22.



	PC1	PC22	EC22	MPC1.50	MPC1	NWPC	LSPC	PRPC	PCEC	GWAS 7	HUB.GENES
Genes	**5037**	**2282**	**731**	**766**	**325**	**643**	**3398**	**1639**	**336**	**894**	**259**
N. samples	485	56	54	485	485	–	–	–	–	184	–

Figure 7. Up: Venn diagram showing the overlap among four gene sets: sg_EN22 (SNP-associated genes from GWAS 7), EC22 (genes that evolved in generation 22), PC22 (genes plastic in generation 22), and PC1 (genes plastic in generation 1).
Down: Summary table of gene sets. Columns list the eleven gene categories, while the two rows report the number of genes in each set and the corresponding sample size (missing where not applicable).

3.2.4 Categories derived from the two generations

Combining the DE results from the two generations allows additional categories to be defined and clarifies how plasticity changed from generation 1 to generation 22 during the evolutionary process (Figure 7).

- *NWPC* - genes that gained plasticity over the generations: 643 genes, of which 550 were annotated and show 8 over-represented profiles (Table 11).

ID	ID GO	Term ID	Source	Pval -adj	Biological meaning
1	GO:0070628	proteasome binding	GO:MF	4.646×10^{-2}	Interaction with the proteasome, involved in protein degradation
2	GO:0042692	muscle cell differentiation	GO:BP	1.126×10^{-2}	Process of muscle-cell differentiation
3	GO:0005622	intracellular anatomical structure	GO:CC	8.833×10^{-5}	Any structure inside the cell
4	GO:0098588	bounding membrane of organelle	GO:CC	1.593×10^{-2}	Membrane delimiting an intracellular organelle
5	GO:0008250	oligosaccharyltransferase complex	GO:CC	1.982×10^{-2}	Complex involved in protein glycosylation
6	GO:0030126	COPI vesicle coat	GO:CC	1.982×10^{-2}	Coat of vesicles implicated in intracellular trafficking
7	GO:0005838	proteasome regulatory particle	GO:CC	2.104×10^{-2}	Regulatory subunit of the proteasome that controls substrate entry
8	GO:1905368	peptidase complex	GO:CC	2.246×10^{-2}	Enzyme complex that catalyses the cleavage of peptide bonds

Table 11. GO profiles of new plastic genes.

- *LSPC* - genes that lost plasticity: 3 398 genes; 2 573 annotated, with 10 over-represented profiles (Table 12).

ID	ID GO	Term ID	Source	Pval -adj	Biological meaning
1	GO:0016491	oxidoreductase activity	GO:MF	9.654×10^{-10}	Catalysis of redox reactions, fundamental in metabolism
2	GO:0005506	iron ion binding	GO:MF	2.245×10^{-6}	Interaction with iron ions, essential for many enzymatic activities
3	GO:0008236	serine-type peptidase activity	GO:MF	8.766×10^{-5}	Protein-cleaving activity that uses a serine residue in the active site
4	GO:0020037	heme binding	GO:MF	1.624×10^{-3}	Binding to haem groups, important in transport and redox reactions
5	GO:0022857	transmembrane transporter activity	GO:MF	8.567×10^{-3}	Transport of molecules across cellular membranes
6	GO:0015116	sulfate transmembrane transporter activity	GO:MF	4.586×10^{-2}	Specific transport of sulphate ions across the membrane
7	GO:0019752	carboxylic acid metabolic process	GO:BP	2.219×10^{-8}	Carboxylic-acid metabolism, involved in many biochemical pathways
8	GO:0055085	transmembrane transport	GO:BP	2.586×10^{-3}	Movement of substances across cellular membranes
9	GO:0006820	monoatomic anion transport	GO:BP	6.372×10^{-3}	Transport of simple inorganic anions such as Cl^- or NO_3^-
10	GO:0005576	extracellular region	GO:CC	1.712×10^{-7}	Extracellular area where many processes take place

Table 12. GO profiles of lost plasticity genes.

- *PRPC* - genes that preserved plasticity: 1 639 genes; 1 332 annotated, with 14 over-represented profiles (Table 13).

ID	ID GO	Term ID	Source	Pval -adj	Biological meaning
1	GO:0004568	chitinase activity	GO:MF	4.735×10^{-4}	Chitin degradation, important in digestion or defence
2	GO:1901505	carbohydrate derivative transmembrane transporter activity	GO:MF	7.083×10^{-4}	Transport of carbohydrate derivatives across the membrane
3	GO:0015932	nucleobase-containing compound transmembrane transporter activity	GO:MF	1.442×10^{-3}	Transport of nitrogenous bases, nucleotides or their derivatives
4	GO:0051082	unfolded protein binding	GO:MF	1.818×10^{-3}	Interaction with unfolded proteins, linked to the stress response
5	GO:0140662	ATP-dependent protein folding chaperone	GO:MF	2.290×10^{-3}	Assistance in protein folding through ATP consumption
6	GO:0016787	hydrolase activity	GO:MF	3.733×10^{-3}	Catalysis of hydrolysis of various substrates
7	GO:0008061	chitin binding	GO:MF	1.566×10^{-2}	Ability to bind chitin, involved in defence or digestion
8	GO:0042026	protein refolding	GO:BP	1.905×10^{-5}	Functional reactivation of denatured proteins
9	GO:1901072	glucosamine-containing compound catabolic process	GO:BP	1.033×10^{-4}	Degradation of glucosamine-containing compounds
10	GO:1901264	carbohydrate derivative	GO:BP	4.503×10^{-4}	Transport of carbohydrate

		transport			derivatives
11	GO:0009408	response to heat	GO:BP	6.336×10^{-4}	Cellular or systemic response to heat stress
12	GO:0009060	aerobic respiration	GO:BP	1.130×10^{-3}	Energy production via oxidation with O_2
13	GO:0051049	regulation of transport	GO:BP	2.026×10^{-2}	Control of substance movement between compartments
14	GO:0005576	extracellular region	GO:CC	1.265×10^{-3}	Extracellular environment where signalling and defence functions occur

Table 13. GO profiles of preserved plasticity genes.

- *PCEC* - plastic genes in generation 1 that also experienced evolutionary change: 336 genes; GO enrichment on 265 of these revealed 4 over-represented profiles (Table 14).

ID	ID GO	Term ID	Source	Pval -adj	Biological meaning
1	GO:0003824	catalytic activity	GO:MF	1.731×10^{-3}	General enzymatic activity that facilitates biochemical reactions
2	GO:0016209	antioxidant activity	GO:MF	2.253×10^{-2}	Neutralisation of free radicals to protect cells
3	GO:0072593	reactive oxygen species metabolic process	GO:BP	1.008×10^{-3}	Metabolism of reactive oxygen species (ROS) associated with oxidative stress
4	GO:0098869	cellular oxidant detoxification	GO:BP	4.565×10^{-2}	Removal or neutralisation of harmful oxidants inside the cell

Table 14. GO profiles of evolved plastic genes.

In summary, of the 5 037 plastic genes in generation 1, 3 398 appear to lose plasticity, whereas 1 639 remain plastic in generation 22, representing about 70% of the total that retain some plastic response.
The 731 genes that show evolutionary change consist mostly of plastic genes (Figure 7):

- 203 genes were plastic in generation 1 but later lost plasticity;
- 133 genes remained plastic across generations;
- 49 genes (NWPCEC) become plastic only in generation 22.

Thus only 346 evolved genes were never plastic, accounting for roughly 47% of the total evolved set.
Finally, of the 902 genes highlighted by GWAS 7 (genes presumably shaped by evolutionary processes) 93 belong to the EC21 category, meaning they also changed in expression level (Figure 7).

3.3 Hub genes

A total of 259 hub genes were identified in the original study and were used in all subsequent analyses.

3.3.1 Fisher tests - enrichment analysis

Detecting categories that are significantly over-represented relative to random expectation provides insight into the genetic structure of the response and into the relationship between plasticity and evolutionary change. Fisher's exact tests showed that the following categories are significantly enriched within the 259 hub genes (Table 15):

- MPC1 and MPC1.50 (most plastic genes)
- PCEC (plastic genes in G1 that also evolved)
- EC22 (all genes with evolutionary change)
- GWAS 7 (genes tagged by significant SNPs in the generation-22 environment GWAS)

Conversely, the full set of 731 evolved genes is significantly enriched for PC1, PC22, PRPC, PCEC, hub genes and GWAS.7 genes (Table 15).
Finally, the 902 genes highlighted by GWAS.7 are themselves over-represented in the NWPC and PCEC categories. In short, every category that reflects an evolutionary shift (EC22, NWPC and PCEC) is enriched for genes whose SNPs display signatures of evolution (Table 15).

3.4 Selection

3.4.1 Selection intensity across generations

Permutation tests were used to assess how strongly each category is selected and how that intensity changes over generations (Table 15).
Categories that experience, and retain, a significantly stronger selection pressure are PC1, MPC1, MPC1.50, hub genes, LSPC, PRPC, PCEC and PC22.
Categories showing no evidence of intensified selection are
NWPC, NWPCEC and GWAS.7 genes.
Only the evolved genes (EC22) exhibit a significantly greater selection intensity in generation 22 than in generation 1 (Figure 8).

3.4.2 Direction of selection (positive or negative)

Additional permutation tests determined whether each category is under significantly positive or negative selection, thereby clarifying whether the plastic and evolutionary responses are adaptive or maladaptive (Table 16).

Generation 1 plastic genes, both PC1 and MPC1, show positive selection for genes that are up-regulated and for those that are down-regulated.

LSPC genes display positive selection only on down-regulated genes.

NWPC genes show no significant selection in generation 1; in generation 22 they are negatively selected when up-regulated and positively selected when down-regulated.

The same pattern holds for PC22 genes.

For evolved genes (EC22), considering the logFC that captures evolutionary change, up-regulated genes are under positive selection, whereas down-regulated genes are under negative selection (Figure 9).

The adaptive or maladaptive value of each plastic or evolutionary response is summarised in the table below (Table 16) and will be discussed in detail in the Discussion section.

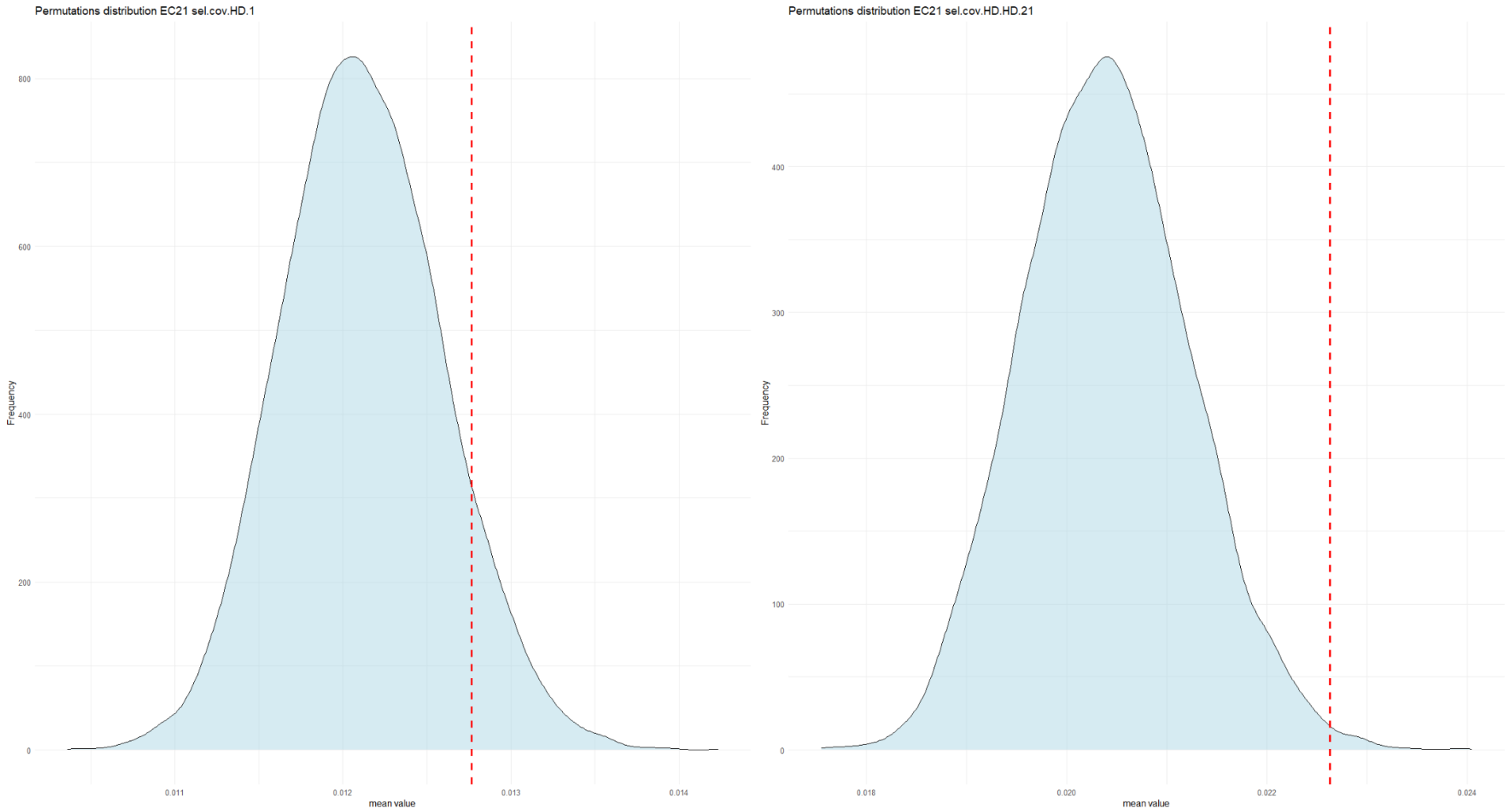


Figure 8. Graphic representation of the permutation test aimed to detect a significant stronger selection on evolved genes at generation 22 (EC22) . Horizontal axis is the mean selection value of the permutation; vertical axis is the frequency of the mean value. The light-blue distribution represent the distribution of the mean values of 10000 permutations executed with the EC22 gene set sample size on the all-genes selection values. The red dashed line points the mean selection value for the EC22 group. The left plot is on generation 1 selection values, while the right one on generation 22 values. Selection on EC22 is significantly stronger only in generation 22.

PERMUT.TEST (BH - FILTERED 10%)			
MAINTAINED STRONGER SELECTION OVER GENERATIONS	N	STRONGER S_G1	STRONGER S_G22
PC1	5037	0,00009	0,00009
MPC1	235	0,00009	0,00009
MPC1.50	766	0,00009	0,00009
HUB GENES	259	0,001	0,00009
LSPC	3398	0,00009	0,00009
PRPC	1639	0,00009	0,00009
PCEC	336	0,0006	0,00009
PC22	2282	0,0008	0,00009

NO STRONGER SELECTION OVER GENERATIONS	N	STRONGER S_G1	STRONGER S_G22
NWPC	643	1	1
NWPCEC	49	0,88	0,69
GWAS.7	893	1	1

STRONGER SELECTION ONLY IN GENERATIONS 22	N	STRONGER S_G1	STRONGER S_G22
EC22	731	0,08	0,004

FISHER TESTS - ENRICHMENTS									
HUB-GENES ENRICHED OF:									
PC1	**MPC1**	**MPC1.50**	PC22	NWPC	LSPC	PRPC	PCEC	EC22	**GWAS 7**
EVOLVED GENES (EC22) ENRICHED OF:									
PC1	MPC1	MPC1.50	PC22	NWPC	LSPC	PRPC	PCEC	HUB GENES	**GWAS 7**
NWPC & PCEC ENRICHED OF:									
-	-	-	-	-	-	-	-	-	**GWAS 7**

Table 15. Summarized results derived from Permutation test and Fisher test. Permutation test table shows the P-values associated with the statistical analysis on each group to infer the intensity of selection both at the first and last generation.
Fisher test table shows the qualitative enrichment of specific genes sets on hub genes, EC22, NWPC and PCEC categories.
Significant results are green coloured, while absence of significant P-values is red coloured.

PERMUT.TEST - SIGNIFICANT POSITIVE OR NEGATIVE SELECTION						
CATEGORY	**logFC PC or EC**	**UP or DOWN REGULATED**	**COV.SEL G1 or G22**	**POSITIVE or NEGATIVE**	**PVAL**	**MEANING**
PC1	PC	UP	G1	POSITIVE	0,008	ADAPTIVE PLASTIC RESPONSE OF UP-REGULATED GENES
PC1	PC	DOWN	G1	NEGATIVE	1	
PC1	PC	DOWN	G1	POSITIVE	0	MALADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES
MPC1	PC	UP	G1	POSITIVE	0	ADAPTIVE PLASTIC RESPONSE OF UP-REGULATED GENES
MPC1	PC	DOWN	G1	NEGATIVE	1	
MPC1	PC	DOWN	G1	POSITIVE	0	MALADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES
LSPC	PC	UP	G1	POSITIVE	0,052	
LSPC	PC	DOWN	G1	NEGATIVE	1	
LSPC	PC	DOWN	G1	POSITIVE	0	MALADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES
NWPC	PC	UP	G1	POSITIVE	0,8	
NWPC	PC	DOWN	G1	NEGATIVE	0,8	
EC22	PC	UP	G1	POSITIVE	0,7	
EC22	PC	DOWN	G1	NEGATIVE	0,008	ADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES
PC22	PC	UP	G22	NEGATIVE	0	MALADAPTIVE PLASTIC RESPONSE OF UP-REGULATED GENES
PC22	PC	DOWN	G22	POSITIVE	0	MALADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES
EC22	EC	UP	G22	POSITIVE	0	ADAPTIVE EVOLUTIONARY RESPONSE OF UP-REGULATED GENES
EC22	EC	DOWN	G22	NEGATIVE	0,0036	ADAPTIVE EVOLUTIONARY RESPONSE OF DOWN-REGULATED GENES
NWPC	PC	UP	G22	NEGATIVE	0	MALADAPTIVE PLASTIC RESPONSE OF UP-REGULATED GENES
NWPC	PC	DOWN	G22	POSITIVE	0,0008	MALADAPTIVE PLASTIC RESPONSE OF DOWN-REGULATED GENES

Table 16. Summarized results of Permutation tests used to infer the adaptive and maladaptive profile of both evolutionary and plastic response. Significant P-values are green colored; conversely are red.

4) DISCUSSION

4.1 GWAS on fitness

Fitness, here measured as fecundity (number of offspring produced), represents the ultimate outcome of multiple physiological and regulatory processes operating at different genetic and molecular levels in response to a complex, multivariate environment. The offspring count reflects many intertwined phenotypic traits, themselves influenced by a frequently redundant genetic architecture, where multiple genotypes can lead to similar phenotypes due to pleiotropy, epistasis, or compensatory interactions (Wagner & Zhang, 2011). Under such polygenic complexity, it is rare for single polymorphic variants to show significant association with traits like fitness, especially given the small effect sizes and the pervasive contribution of background noise and genetic interactions (Boyle, Li & Pritchard, 2017).

Moreover, before an individual can reproduce it must first survive: the fecundity observed is therefore conditioned by viability, which is genetically complex as well (Orr, 2009). The selective context is also decisive. In a system where direct selection acts on a trait such as speed - e.g. in a resource-competition scenario - one might plausibly detect a direct genotype-fitness association from the first generation, particularly if the selected trait is highly heritable (Barrett & Hoekstra, 2011). In the present experiment, however, the selective context is defined by a combined, stressful heat-and-drought regime that acts on many traits simultaneously and in a non-linear fashion. In such a scenario it is highly improbable to detect significant associations as early as generation 1. Indeed, even within the same line, mean offspring number was lower in HD than in the control, indicating a generalised effect of environmental stress.

For a genetic variant to emerge as fitness-associated after viability selection, it would need to be sufficiently common among survivors and to exert a measurable effect on the offspring they produce. A negative association would be paradoxical, as selection should already have reduced its frequency; a positive association would require an effect strong enough to stand out despite limited sample size, which is unrealistic in the early generations. Given a large enough sample, significant associations in the final HD generation would be expected only if adaptation to stress - driven mainly by viability selection over generations - had produced a tight correlation between survival selection and fecundity selection.

In line with these expectations, the sole significant result arose from the analysis that combined all generation-1 individuals irrespective of environment. This confirms a key methodological principle in GWAS: increasing sample size boosts statistical power, and hence the likelihood of detecting significant signals (Visscher et al., 2017).

Although the fitness associations did not return many significant variants, functional annotation of genes near the detected variants provides some intriguing leads. Eight of the thirteen mapped genes have biological characterisations that, while not enriched for specific GO terms, point to roles in cellular processes potentially relevant to stress adaptation. These include genes

linked to heat-shock response, metabolite transport, protein glycosylation and muscle structure. The functional heterogeneity observed, together with the absence of significant enrichment, fits the notion of a diffuse polygenic architecture in which adaptation involves the concerted action of many genes with modest, pathway-scattered effects (Boyle, Li & Pritchard, 2017). Such a broad signal is consistent with what is typically seen in highly complex quantitative traits like fitness (Koch & Guillaume, 2020a).

4.2 GWAS on environment

Compared with the previous studies on fitness, the associations with environment yielded more consistent results; in fact, by contrasting individuals from the two environments (CT - HD), the direct effect of viability selection was observed. Putative variants that enable the survival of individuals under stress were identified.

The results suggest that selective pressure began acting from the very first generation, revealing 256 variants; however, only a small subset is likely to result from direct selection. For this to occur, alleles must confer large fitness effects and exhibit high selection coefficients. Indeed only four variants were found in three or more replicate lines, making them potential candidates as key adaptive loci. Yet, when integrating these generation-1 results with the 1,699 SNPs detected in the final generation, no shared variants were found. This suggests that even those few variants were either not maintained by selection or were lost due to other evolutionary processes such as drift or recombination. Overall, the majority of the observed SNPs in generation 1 likely reflect stochastic effects or linkage disequilibrium rather than targets of strong selection (Barghi et al., 2020).

A further point of reflection arises from the observation that 16 of the 19 SNPs associated with fitness are found among the 256 variants, which suggests that some genetic variants may contribute simultaneously to survival and reproductive capacity, providing a selective advantage already in the first generation. This convergence could reflect a form of pleiotropic selection, in which a single genetic locus affects multiple components of fitness, or it could stem from an indirect effect linked to the initial genetic structure of the population. However, the absence of associated functional enrichments and the possible influence of the initial genetic structure demand caution in interpretation, leaving open the possibility that these signals reflect indirect effects rather than direct causality.

In contrast, in the final generation, despite a sample size three times smaller, the number of environmentally associated variants increased nearly sixfold. This substantial rise is consistent with expectations from long-term experimental evolution under constant stress. It likely reflects the cumulative effect of direct selection, which incrementally increased the frequency of adaptive alleles, amplifying the genomic signal (Burke et al., 2014). While linkage and drift cannot be entirely ruled out, the functional and transcriptional coherence among many of the associated genes supports a key role for selection in driving the observed genomic divergence. This interpretation is reinforced by parallelism data, which

show that several variants are shared across three or four replicate lines, pointing to repeated, potentially adaptive responses across independent populations.
The GO-category enrichment analysis on the genes associated with generation 22 support the above idea: adaptation to prolonged stress involves an articulated polygenic response distributed over different functional levels. The molecular and biological categories that emerged reflect the activation of general cellular-survival mechanisms, including energy metabolism, cofactor binding and intracellular-restructuring processes. In particular, the high enrichment (pval.adj 6.881×10^{-17}) of genes linked to "intracellular anatomical structures" indicates that adaptation to the hot–dry environment may involve modifications in organelle organisation and structural-maintenance mechanisms. Studies on animal cells under heat shock (for example *Saccharomyces cerevisiae* at 38 °C) have highlighted significant reorganisations of organelles - including vacuoles, mitochondria and vacuolar complexes - with well-documented structural and functional impacts (Keuenhof et al., 2021) thus supporting the hypothesis that analogous transformations may also occur in insects' adaptation. Although some of the enriched categories are broad, their convergence around key physiological processes, such as protein refolding, lipid catabolism, and glycan metabolism, suggests a coordinated, systemic response. This pattern fits well with the genetic and transcriptomic results presented earlier and underscores an adaptive strategy rooted in moderate but functionally widespread cellular reprogramming, rather than reliance on a few high-effect mutations.

4.3 Plasticity as first response

The first line of defence for most insects against rapid environmental change is gene-expression regulation. Form of phenotypic plasticity, it allows organisms to respond flexibly to new conditions (King & Wilson, 1975; West-Eberhard, 2003). By transcriptionally adjusting gene activity, an individual can temporarily exceed the limits of its optimal environmental range, activating cytoprotective, metabolic or structural mechanisms that raise the likelihood of survival (Schlichting & Smith, 2002).
Although this response has a genetic basis, it does not involve heritable change and therefore constitutes a non-evolutionary acclimation. Consistent with this dynamic are the differential-expression results from generation 1: about half of the expressed genes are plastically regulated, although for many of them the magnitude of change is modest. Nevertheless, a consistent subset of highly plastic genes emerges: 766 genes show an expression change ≥ 50 %, and roughly half of these reach values ≥ 100 % relative to the control group. Overall, about 6 % of the total transcriptome displays marked plasticity, a figure in line with previous observations in environmental-stress experiments on *Tribolium* and *Drosophila* (Gibson, 2008; Chen et al., 2015).
Functional-enrichment (GO) analysis of the most plastic genes (≥ 50 % change) reveals strong involvement of molecular categories implicated in the cellular response to environmental stress. Among the most significant molecular-function terms stand out Hydrolase activity (p-adj 2.06×10^{-16}), indicating a

general enzymatic activity to degrade damaged or obsolete molecules; Chitin binding (2.50×10^{-8}), crucial for cuticle remodelling in response to external changes; and Nutrient reservoir activity (1.94×10^{-3}), reflecting an energy- or protein-storage strategy under stress.
Regarding biological processes, terms linked to carbohydrate metabolism (2.45×10^{-13}), proteolysis (3.56×10^{-5}), protein refolding (1.06×10^{-4}) and response to heat (1.19×10^{-2}) are significantly enriched, confirming a multifactorial plastic response involving detoxification, cellular recycling and molecular-chaperone activation (Kültz, 2005; Feder & Hofmann, 1999).
Repeating the analysis for the subset of the most highly plastic genes (≥ 100 % change) reinforces these observations, with similar significant enrichments for Hydrolase activity, acting on glycosyl bonds (1.44×10^{-16}), Chitin binding (1.22×10^{-11}), Protein refolding (1.26×10^{-7}) and Response to heat (4.24×10^{-4}).
The significant presence of the term extracellular region (p-adj 3.08×10^{-19}) among the highly plastic genes also implies an active role for secreted proteins in modulating interaction with the external environment.
This plastic response, though transient, thus appears highly functional and integrated, representing the organism's first attempt to compensate for environmental shock by metabolic, structural and molecular reorganisation.

4.4 Temporal dynamics of plasticity

While gene-expression plasticity provides an immediate buffer against environmental stress, it is not in itself an evolutionary solution. Over time, plastic responses may be preserved, lost, refined, or genetically assimilated depending on their adaptive value and cost-benefit balance (Lande, 2009; Crispo, 2007; Ghalambor et al., 2007). In the long term, selection may favour individuals in which the most beneficial expression levels are achieved not via short immediate response but through stable, heritable regulation.
Consistent with this framework, the data from generation 22 show a clear reshaping of transcriptional plasticity. While 5,037 genes displayed plastic regulation in generation 1, only 2,282 were plastic at the end of the experiment, indicating that approximately two-thirds of the initial plastic response was not maintained. A total of 1,639 genes conserved their plasticity across both generations, suggesting they still respond to the environment in the same way. At the same time, 643 genes became plastic only in generation 22, proposing either evolved plasticity or a response to conditions that was not detectable in the earlier samples. These observations align with the idea that plasticity is itself an evolvable trait, subject to selection depending on context and fitness consequences (Pigliucci, 2005).
Among the 731 genes that underwent evolutionary divergence in expression, 336 were already plastic in generation 1. This partial overlap may reflect cases where an initially plastic response became fixed over time. Such dynamic is compatible with the overall Baldwin effect, whereby plasticity facilitates adaptation by exposing phenotypes to selection before they become genetically canalised (Baldwin, 1896; Crispo, 2007). Of these 336 genes, 203 lost their

plasticity in the final generation, acquiring a strong likelihood of genetic assimilation.
On the other hand, the remaining 133 retained some level of environmental responsiveness despite having diverged in baseline expression. Such preserved plasticity, together with the full set of PRPC, might indicating that plastic response still need to reach the optimum, thereby constituting, by definition, the Baldwin effect. The approaching to the optimum is likely occurring even to the NWPL genes, therefore under the same effect. Additionally, 182 of the evolved genes were also plastic in generation 22, indicating that plasticity and evolutionary change can sometimes emerge in parallel, potentially contributing to more flexible or robust expression dynamics.
It is important to note that the number of samples differs considerably between generations - 485 individuals in generation 1 versus 56-54 in generation 22 - which may influence the power to detect differential expression. As such, some apparent losses of plasticity could reflect reduced sensitivity in the smaller dataset rather than genuine biological change. Nonetheless, the overall shift in the composition and number of plastic genes over time strongly suggests a selective refinement of the initial transcriptional response in favour of more stable regulatory strategies.

4.5 GO analysis at the end of experimental evolution

The functional-enrichment analysis of plastic and evolutionary genes in generation 22 provides an overview of the transcriptional systems most responsive to prolonged environmental stress. It is important to note, however, that the GO analysis performed here does not distinguish between genes whose expression increases and those whose expression decreases. The direction of the transcriptional response is not captured. Since $\log_2$-fold change values tend to follow a near-normal distribution, across each category, approximately half the genes are up-regulated and half down-regulated. While this limits the specificity of biological interpretations it still allows for the identification of functional domains undergoing strong regulatory turnover, which is in itself informative.
Among the plastic genes of generation 22, enriched terms span membrane transport (e.g. carbohydrate derivative transporter activity, nucleobase transport), protein quality control (protein refolding, chaperone binding), metabolic restructuring (cellular respiration, amide metabolism), and structural defence (chitinase activity, response to heat). This constellation of functions suggests that prolonged exposure to the HD environment continues to require a coordinated activation of energy production, proteostasis, and cellular remodelling.
Genes showing evolved expression changes are significantly enriched for more general processes such as catalytic activity, ion and ATP binding, and crucial cell-cycle functions like mitotic progression and DNA replication initiation. This shift hints at a more fundamental rewiring of cellular physiology, possibly reflecting evolved differences in developmental timing or tissue renewal under chronic stress exposure.

Further insights emerge from the categorisation of plastic genes according to their trajectory across generations. Genes that gained plasticity (NWPC) are enriched in terms related to intracellular organisation (oligosaccharyltransferase complex, COPI vesicle coat, peptidase complex), suggesting that evolved plasticity may target …suggesting that evolved plasticity may target the way proteins are moved inside the cell and chemically modified after they are made. Conversely, genes that lost plasticity (LSPC) are enriched for oxidative metabolism (oxidoreductase activity, heme and iron binding, carboxylic acid metabolism), consistent with a steadier, energy-efficient metabolic program that no longer needs to adjust to the environment.
PRPC gene set retain enrichment in categories tied to protein refolding, membrane transport, cuticle dynamics, and heat response, likely indispensable functions for survival under prolonged intense stress. Finally, genes that were both plastic in generation 1 and evolved in generation 22 (PCEC) show enrichment for catalytic activity, antioxidant function, and ROS detoxification, suggesting that oxidative stress management was both an immediate and a long-term target of selection.
Together, these patterns highlight a modular, multi-layered response to environmental stress: some regulatory programs remain plastic across generations, others become fixed, and new ones emerge. This reflects a complex interplay between plasticity and adaptation shaped by selection and constraint. Such evolutionary fine-tuning of plastic responses is well-documented in experimental systems and supports the view that long-term adaptation often acts on pre-existing plastic regulatory frameworks (Levis & Pfennig, 2016).

4.6 Network and evolutionary dynamics

Understanding how plasticity and evolutionary change are structured across the gene-expression network provides key insights into the molecular architecture of adaptation. In this context, network-based approaches - particularly those focused on modules of co-expressed genes - have shown that a gene's position within a network is not neutral but strongly influences its evolutionary potential (Langfelder & Horvath, 2008; Koch et al., 2025). Genes occupying central positions (hub genes) are often subject to stronger indirect selection due to their high connectivity, which increases the likelihood that their expression covaries with fitness-related traits. In *Tribolium castaneum*, Koch et al. (2025) demonstrated that hub genes exhibited stronger genetic selection coefficients and more pronounced evolutionary shifts in expression compared to peripheral genes. This pattern reflects the importance of indirect selection within highly connected expression modules, where adaptation is shaped not only by direct pressure on individual genes but also by their correlated partners.
Consistent with this interpretation, the enrichment analyses reveal that relevant categories are significantly overrepresented among hub genes. Specifically, the 259 hub genes are enriched in both highly plastic categories (MPC1, MPC1.50) and in those associated with evolutionary change (EC22, PCEC, GWAS 7), suggesting a tight connection between network centrality and adaptive potential. Moreover, the enrichment of GWAS 7 genes in categories such as NWPC and

PCEC highlights that many SNPs identified at generation 22 are located in genes that either gained plasticity during the evolutionary process or exhibited plasticity that became genetically fixed, pointing to a convergence between selection, plasticity, and network architecture.
These findings support the view that network centrality is not merely a statistical feature but a biologically meaningful determinant of evolutionary dynamics. In the context of adaptation, hub genes serve as regulatory nodes whose evolution can affect systemic responses. As suggested by theoretical frameworks (Hansen & Houle, 2008; Chevin et al., 2010), the alignment between modular network structure and selective pressures may enhance the speed and directionality of adaptive responses, thereby increasing the overall evolvability of the system.

4.7 Selection Intensity

The analysis of selection intensity across generations reveals important patterns in the dynamics of gene regulation under environmental stress. Permutation-based tests show that several gene categories experienced significantly stronger selection pressures that persisted consistently from generation 1 to generation 22. These include genes that were initially plastic (PC1), highly plastic (MPC1, MPC1.50), maintained plasticity (PRPC), lost plasticity (LSPC), or underwent evolutionary shifts (PCEC, PC22), as well as hub genes identified in the original expression network. The persistence of strong selection in these groups suggests that expression levels were closely aligned with fitness across time.
Some of the sets might represent core components of the adaptive response. In contrast, genes that gained plasticity during evolution (NWPC), or showed both new plasticity and evolutionary change (NWPCEC), did not display elevated selection intensity, nor did the GWAS 7 genes. This may indicate that newly recruited plastic responses, although present, were not under strong or consistent selection, possibly reflecting their more marginal role in the adaptive dynamic.
Notably, a significant increase in selection intensity was detected in generation 22 only among the genes that had undergone evolutionary changes (EC22). In the first generation, the large plastic shift likely masked genetic differences in expression, so individual EC22 genes experienced little detectable selection. After heritable expression changes became established (presumably advantageous) selection acted mainly to stabilise those new levels, keeping them aligned with fitness. Thus, early adaptation relied on plasticity, whereas later adaptation fixes and preserves the specific genetic changes that proved beneficial.

4.8 Adaptive and maladaptive trajectories of gene expression under environmental stress

The adaptive value of transcriptional responses, whether plastic or evolutionary, depends not merely on the occurrence of gene-expression changes, but on whether these changes occur in the same direction as selection. If selection favours higher expression of a gene (positive covariance), an up-regulation is adaptive, while a down-regulation is maladaptive. Conversely, when selection

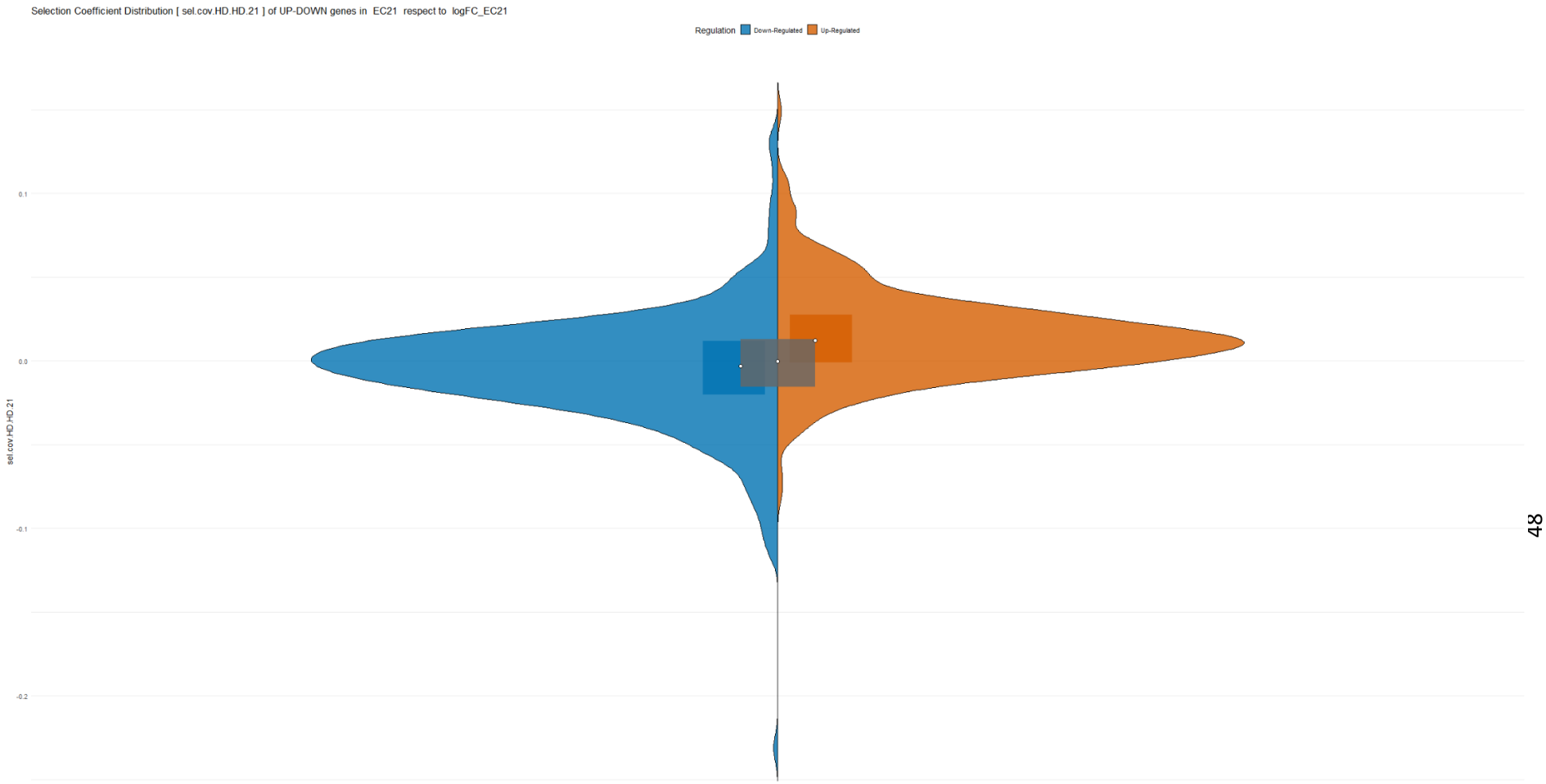


Figure 9. Graphic representation of generation 22 selection values of evolved gene (EC22) with statistical insights. The red distribution defines the values of the up-regulated genes while the red distribution define the down-regulated genes. The respective colored boxes within the distribution represent the range between the 25th and 75th quantile, and the white dot the mean. The central box on the middle line refers to the all-genes values. Up-regulated genes are under significant positive selection while down-regulated ones are under significant negative, indicating a fully adaptive profile of the evolutionary response.

disfavours high expression (negative covariance), a down-regulation is adaptive and an up-regulation maladaptive (Ghalambor et al., 2007; Ho & Zhang, 2018). Permutation tests applied to gene sets provide a refined picture of the selective landscape shaping expression dynamics under environmental stress.

In generation 1, transcriptional plasticity appears to be a mixed strategy, comprising both adaptive and maladaptive components. Among all plastically regulated genes (PC1), up-regulated genes are under significantly positive selection, suggesting an adaptive plastic response, as described by Waddington (Crispo, 2017). However, down-regulated genes in the same category are also under positive selection, indicating that their repression is maladaptive, following the Shcmalhausen's pattern. This dual configuration becomes even more evident when focusing on the highly plastic subset (MPC1), where the same asymmetry is reinforced: up-regulated genes are adaptively regulated, while down-regulated genes are under selection to increase expression. The asymmetry suggests that, when confronted with acute environmental stress, increasing gene expression is generally more beneficial than reducing it. It possibly reflects the need to activate energy-intensive molecular and cellular processes, such as detoxification, protein folding, or cellular repair (Feder & Hofmann, 1999; Kültz, 2005). In harsh conditions, suppressing gene activity may impair stress tolerance mechanisms, explaining why plastic down-regulation appears counter-selected.

A similar maladaptive signal is observed in genes that lost plasticity over time (LSPC). Here, selection in generation 1 is significantly positive for genes that had been down-regulated, implying that their repression was disadvantageous, and that expression recovery over time may have been favoured. Conversely, genes that gained plasticity only in generation 22 (NWPC) show no selective signal in generation 1, but exhibit maladaptive profiles in generation 22 both in up-regulated and down-regulated genes. This suggests a decoupling of plasticity and fitness as the system matures, highlighting that even new plasticity might become costly to maintain in the long-term.

The category of plastic genes in generation 22 (PC22) shows an analogous pattern: both up- and down-regulated genes are under selection in the opposite direction of their expression changes, marking them as maladaptive responses. These observations, collectively, point again to the result that plasticity may provide only a short-term advantage but becomes progressively costly as the genetic background evolves in the opposite direction (Ghalambor et al., 2007).

In contrast, evolutionary changes in expression (EC22) exhibit a fully adaptive profile: up-regulated genes are under positive selection, and down-regulated ones under negative selection. This is in line with the basic expectations: if under a strong selective background gene expression evolves, evolutionary changes on the regulation are presumably adaptive for that environment. Interestingly, these genes were partially adaptive even considering the plastic response of them in generation 1.

4.9 Methodological limitations and interpretive caution

While these patterns are biologically coherent and consistent with evolutionary theory, it is important to underscore that gene-wise selection coefficients were

not individually significant, likely due to limited sample sizes and the polygenic nature of fitness traits. Moreover, plastic and evolutionary responses were categorized based on expression fold changes, not the biological role of individual genes, meaning that gene functions remain inferred rather than confirmed. Nonetheless, when interpreted at the level of gene sets, these results collectively offer a compelling narrative: initial plasticity offers a mix of helpful and harmful responses, depending on the increase or decrease of the expression respectively.

5) CONCLUSION

This thesis shows how *Tribolium castaneum* copes with a harsh hot-dry environment by first "bending" and then "building" its transcriptome. In generation 1 almost half of the expressed genes changed their activity, giving the beetles a quick plastic buffer against stress. Twenty generations later the picture had shifted: two-thirds of those early plastic shifts had disappeared, 731 genes had evolved new baseline levels and 1 699 SNPs now marked the genome regions most closely tied to the environment.

This pattern mirrors the classic ideas on phenotypic plasticity and adaptation. The initial surge of plastic expression fits the Baldwin effect: a rapid, reversible change that moves the phenotype toward the new optimum and buys time for the population. As generations pass, some of those helpful shifts become hard-wired, a process known as genetic assimilation, so the beetles no longer need the environmental cue to show the advantageous expression profile. At the same time, plastic responses that were neutral or harmful are countered by genetic changes in the opposite direction, producing the counter-gradient pattern in which DNA evolution restores the trait to its optimal level while the original plastic shift fades away. Together, the DE data capture the Baldwin phase, and the GWAS highlights the alleles that underlie assimilation or counter-gradient compensation. The result is a clear sequence: flexibility first, fixation (or correction) later, illustrating how plasticity and selection work together to secure long-term adaptation.

[Additional information with scripts, plots, and results can be found here: https://github.com/Bioyako/Tribolium-Traineeship1]

REFERENCES

Alvarez, M., Schrey, A.W. & Richards, C.L., 2015. Ten years of transcriptomics in natural populations: What have we learned about adaptation? *Molecular Ecology*, 24, 710-725.

Baldwin, J.M., 1896. A new factor in evolution. *The American Naturalist*, 30, 441-451.

Barghi, N. *et al.*, 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLoS Biology*, 17, e3000128.

Barghi, N. *et al.*, 2020. Polygenic adaptation: a unifying framework for understanding positive selection. *Nature Reviews Genetics*, 21, 769-781.

Barrett, R.D.H. & Hoekstra, H.E., 2011. Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, 12, 767-780.

Barton, N.H. *et al.*, 2017. The infinitesimal model: definition, derivation and implications. *Theoretical Population Biology*, 118, 50-73.

Bellard, C. *et al.*, 2012. Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15, 365-377.

Benjamini, Y. & Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57, 289-300.

Bersaglieri, T. *et al.*, 2004. Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, 74, 1111-1120.

Boyle, E.A., Li, Y.I. & Pritchard, J.K., 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169, 1177-1186.

Brown, S.J. *et al.*, 2009. The red flour beetle *Tribolium castaneum* as a model for comparative developmental studies. *Cold Spring Harbor Protocols*, 2009(8), pdb-emo126.

Burke, M.K. *et al.*, 2014. Standing genetic variation drives repeatable experimental evolution in *Saccharomyces cerevisiae*. *Molecular Biology and Evolution*, 31, 3228-3239.

Chevin, L-M., Lande, R. & Mace, G.M., 2010. Adaptation, plasticity and extinction in a changing environment: towards a predictive theory. *PLoS Biology*, 8, e1000357.

Chen, H. *et al.*, 2015. After domestication selection acts on maternal-type mitochondrial DNA in Tibetan chickens. *PLoS ONE*, 10, e0120652.

Coddington, J., 1988. Cladistic tests of adaptational hypotheses. *Cladistics*, 4, 3-22.

Conover, D.O., Duffy, T.A. & Hice, L.A., 2009. The covariance between genetic and environmental influences across ecological gradients. *Annals of the New York Academy of Sciences*, 1168, 100-129.

Crispo, E., 2007. The Baldwin effect and genetic assimilation: revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. *Evolution*, 61, 2469-2479.

Darwin, C., 1859. *On the Origin of Species.* London: John Murray.

Enattah, N.S. *et al.*, 2002. Identification of a variant associated with adult-type hypolactasia. *Nature Genetics*, 30, 233-237.

Feder, M.E. & Hofmann, G.E., 1999. Heat-shock proteins, molecular chaperones and the stress response. *Annual Review of Physiology*, 61, 243-282.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399-433.

Fisher, R.A., 1922. On the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42, 321-341.

Fisher, R.A., 1930. *The Genetical Theory of Natural Selection.* Oxford: Clarendon Press.

Fraser, H.B., 2011. Genome-wide approaches to the study of adaptive gene expression evolution. *BioEssays*, 33, 469-477.

Fraser, H.B., 2013. Gene expression drives local adaptation in humans. *Genome Research*, 23, 1089-1096.
Futuyma, D.J. & Kirkpatrick, M., 2017. *Evolution*, 4th ed. Sunderland, MA: Sinauer.
Fusco, G., 2019. Lesson of Theory of evolution, Master class in Evolutionary Biology, University of Padua.
Ghalambor, C.K. *et al.*, 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, 21, 394-407.
Gibson, G., 2008. The environmental contribution to gene expression profiles. *Nature Reviews Genetics*, 9, 575-581.
Gould, S.J. & Vrba, E.S., 1982. Exaptation—a missing term in the science of form. *Paleobiology*, 8, 4-15.
Grether, G.F., 2005. Environmental change, phenotypic plasticity, and genetic compensation. *American Naturalist*, 166, E115-E123.
Hansen, T.F. & Houle, D., 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, 21, 1201-1219.
Herbst, J.F.W., 1797. *Natursystem der ungeflügelten Insekten*. J.G. Müller.
Ho, W-C. & Zhang, J., 2018. Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. *Nature Communications*, 9, 350.
Hodgins-Davis, A. & Townsend, J.P., 2009. Evolving gene expression: from G to E to G×E. *Trends in Ecology & Evolution*, 24, 649-658.
Hoffmann, A.A. & Sgrò, C.M., 2011. Climate change and evolutionary adaptation. *Nature*, 470, 479-485.
Howe, K.L. *et al.*, 2020. Ensembl 2021. *Nucleic Acids Research*, 49, D884-D891.
Jordan, I.K., Mariño-Ramírez, L. & Koonin, E.V., 2019. Evolutionary significance of gene network reorganizations. *Genome Biology*, 20, 24.
Kang, H.M. *et al.*, 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42, 348-354.
Keuenhof, K. *et al.*, 2021. Mapping the genetic architecture of polygenic adaptation. *Molecular Ecology*, 30, 1533-1547.
King, M.C. & Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. *Science*, 188, 107-116.
Koch, E.L. & Guillaume, F., 2020a. Additive and mostly adaptive plastic responses of gene expression to multiple stress in *Tribolium castaneum*. *PLOS Genetics*, 16, e1008768.
Koch, E.L. & Guillaume, F., 2020b. Restoring ancestral phenotypes is a general pattern in gene expression evolution during adaptation to new environments in *Tribolium castaneum*, *Molecular Ecology*, doi: 10.1111/mec.15607.
Koch, E.L. *et al.*, 2020. Genetic variance in fitness and its cross-sex covariance predict adaptation during experimental evolution, *Evolution*, 74(11), 2570–2584, doi: 10.1111/evo.14119.
Koch, E.L., Rocabert, C., Beeravolu Reddy, C. & Guillaume, F. 2024. Gene expression evolution is predicted by selection, genetic covariance and network topology, *bioRxiv*, doi: 10.1101/2024.07.22.604294.
Kültz, D., 2005. Molecular and evolutionary basis of the cellular stress response. *Annual Review of Physiology*, 67, 225-257.
Lande, R., 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology*, 22, 1435-1446.
Langfelder, P. & Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.
Law, C.W. *et al.*, 2014. voom: precision weights unlock linear model analysis of RNA-seq read counts. *Genome Biology*, 15, R29.

Levis, N.A. & Pfennig, D.W., 2016. Evaluating "plasticity-first" evolution in nature: key criteria and empirical approaches. *Trends in Ecology & Evolution*, 31, 563-574.
Lewontin, R.C., 1985. The genetic basis of evolutionary change. *Journal of Heredity*, 76, 1-3.
Lynch, M. & Walsh, B., 1998. *Genetics and Analysis of Quantitative Traits.* Sunderland, MA: Sinauer.
Milutinović, B. *et al.*, 2013. Red flour beetle as a model for bacterial oral infections. *PLoS ONE*, 8, e64638.
Orr, H.A., 2009. Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10, 531-539.
Pigliucci, M., 2001. *Phenotypic Plasticity: Beyond Nature and Nurture.* Baltimore: Johns Hopkins University Press.
Price, G.R., 1970. Selection and covariance. *Nature*, 227, 520-521.
Raudvere, U. *et al.*, 2019. g:Profiler—web server for functional enrichment analysis. *Nucleic Acids Research*, 47, W191-W198.
Ritchie, M.E. *et al.*, 2015. limma powers differential expression analyses for RNA-sequencing and microarray. *Nucleic Acids Research*, 43, e47.
Robinson, M.D., McCarthy, D.J. & Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139-140.
Robertson, A., 1966. A mathematical model of the culling process in dairy cattle. *Animal Production*, 8, 95-108.
Romero, I.G., Ruvinsky, I. & Gilad, Y., 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13, 505-516.
Schlichting, C.D. & Smith, H., 2002. Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. *Evolutionary Ecology*, 16, 189-211.
Shmalhausen, I.I., 1949. *Factors of Evolution: The Theory of Stabilizing Selection.* Philadelphia: Blakiston.
Sokoloff, A., 1972. *The Biology of Tribolium.* Oxford University Press.
Tam, V. *et al.*, 2019. Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, 20, 467-484.
Tribolium Genome Sequencing Consortium, 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452, 949-955.
Urban, M.C., 2015. Accelerating extinction risk from climate change. *Science*, 348, 571-573.
Visscher, P.M. *et al.*, 2017. 10 years of GWAS discovery: biology, function and translation. *Nature Reviews Genetics*, 18, 473-490.
Wagner, A. & Zhang, J., 2011. The pleiotropic structure of the genotype–phenotype map. *Nature Reviews Genetics*, 12, 204-213.
Waddington, C.H., 1961. Genetic assimilation. *Advances in Genetics*, 10, 257-293.
Walsh, B. & Lynch, M., 2018. *Evolution and Selection of Quantitative Traits.* Oxford: Oxford University Press.
West-Eberhard, M.J., 2003. *Developmental Plasticity and Evolution.* Oxford: Oxford University Press.
Wray, G.A., 2007. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8, 206-216.
Yang, J. *et al.*, 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88, 76-82.
Yi, X. *et al.*, 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, 329, 75-78.

Zheng, X. *et al.*, 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28, 3326-3328.
Zhou, X. *et al.*, 2012. Polygenic adaptation to high altitude in Tibetan chickens. *Proceedings of the National Academy of Sciences USA*, 110, 11970-11975.