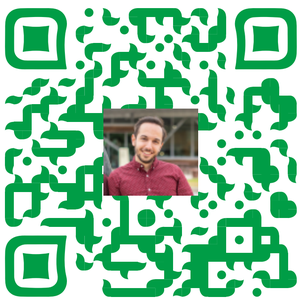
EMBL's European Bioinformatics Institute (EMBL-EBI)

Extensive Gene x Environment interactions in the outbred vertebrate genetics of a standard physiological phenotype

Saul Pierotti[1#*], Bettina Welz[2#], Jakob Gierten[2], Beate Wittbrodt[2], Philip Watson[2], Sebastian Stricker[2], Marcio Ferreira[1], Fanny Defranoux[1], Risa Suzuki[2], Jana Fuß[2], Tomas Fitzgerald[1], Thomas Thumberger[2], Kiyoshi Naruse[4], Felix Loosli[3], Joachim Wittbrodt[2], Ewan Birney[1]

[1] European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[2] Centre for Organismal Studies, Heidelberg University, Campus Im Neuenheimer Feld 230, 69120, Heidelberg, Germany
[3] Institute of Biological and Chemical Systems, Biological Information Processing (IBCS-BIP), Karlsruhe Institute of Technology, 76131, Karlsruhe, Germany
[4] Laboratory of Bioresources, National Institute for Basic Biology, Okazaki, Japan
[#] these authors contributed equally
[*] presenting author



Summary

We demonstrate that the Medaka Inbred Kiyosu Karlsruhe (MIKK) panel can be used for discovering novel genetic associations in a phenotype of relevance for human physiology (heart rate), and use this resource to quantify the importance of interaction terms (gene-by-gene, GxG, and gene-by-environment, GxE) in shaping individual's phenotypes. We use our results to challenge the linear approximation typically used in human Genome-Wide Association Studies (GWAS).

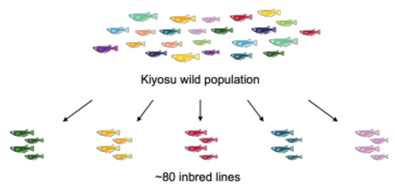
Phenotype = f(Genotype, Environment)

Phenotype =

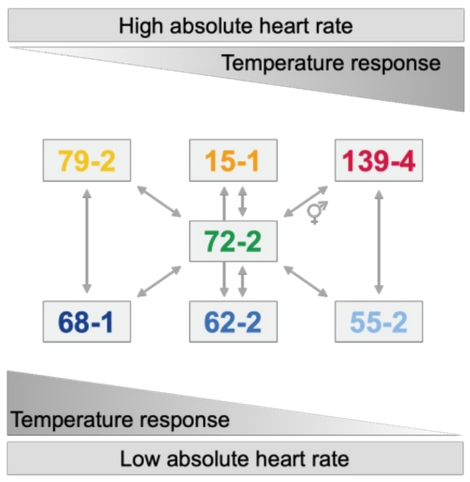
$$\sum f(G_i) + \sum f(E_j) + \sum f(G_i, G_j) + \sum f(G_i, E_j) + \sum f(E_i, E_j)$$

The MIKK panel

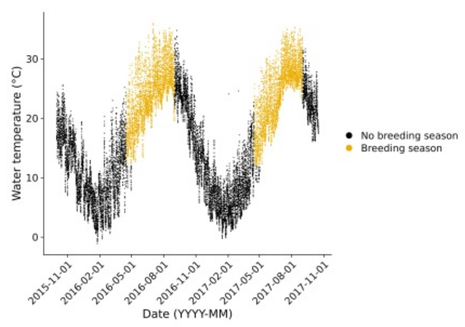
The Medaka Inbred Kiyosu Karlsruhe (MIKK) panel (Fitzgerald et al, 2022) is the first near-isogenic panel of 80 inbred lines in the medaka fish (*Oryzias latipes*) derived from a wild founder population. It is currently maintained in Germany (KIT and University of Heidelberg) and it is the core resource used in this work. The GWAS work presented here is the first to be conducted on the MIKK panel.



F2 cross setup among inbred MIKK panel lines used in this work

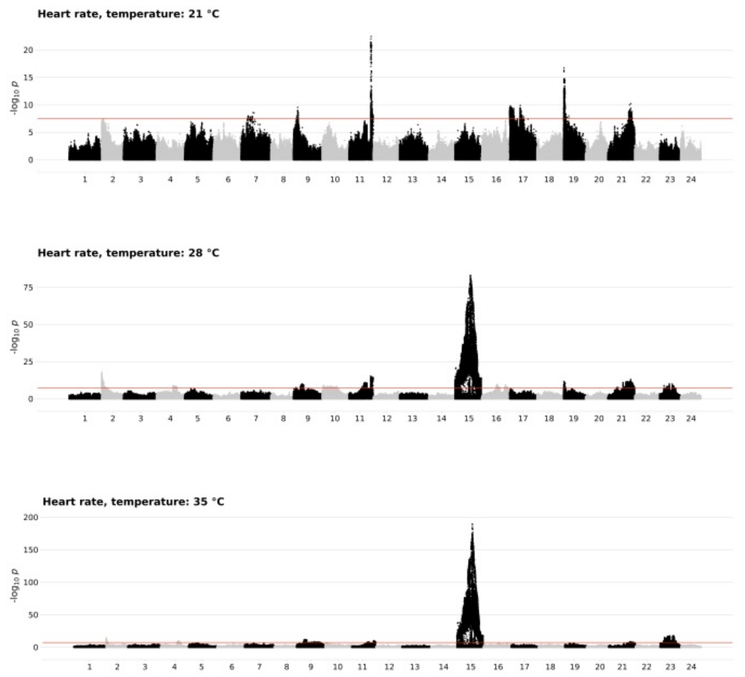


Water temperature in the area of origin of the MIKK panel
data kindly provided by the Yoshimura lab, Nagoya University

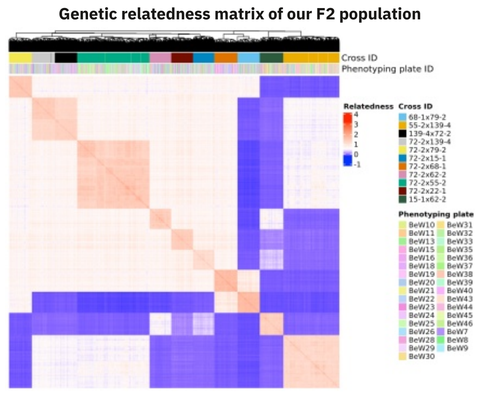


Temperature-dependent genetic associations for heart rate in medaka fish embryos

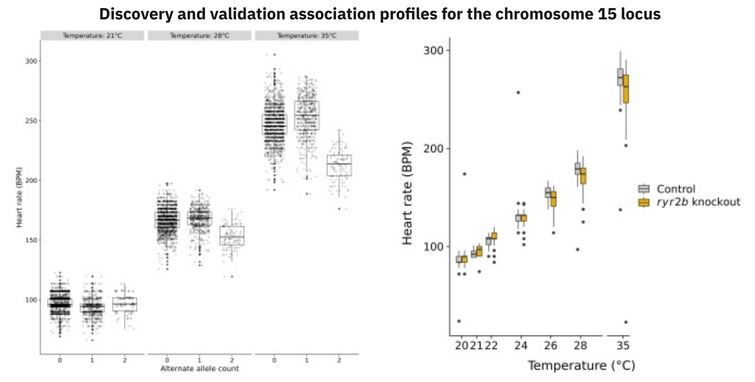
We ran a Genome-Wide Association Study (GWAS) on a multiparental medaka (*Oryzias latipes*) F2 population of 2475 fish obtained by crossing 8 lines from the MIKK panel in 11 different combinations. We detect 16 Quantitative Trait Loci (QTLs), some of which are active only at certain temperatures (suggesting the presence of gene-by-environment interactions, GxE). We were able to identify and validate (using CRISPR genome editing) the genes responsible for the associations for 4 of the 16 QTLs. As an example, we show here the discovery and validation association profiles for the prominent chromosome 15 locus, which we were able to map to a loss of function mutation on the *ryr2b* gene (cardiac isoform of the ryanodine receptor). Notice how the effect is only present for temperatures above 24 °C in both the discovery and validation datasets, a clear example of GxE.



Genetic relatedness matrix of our F2 population



Discovery and validation association profiles for the chromosome 15 locus

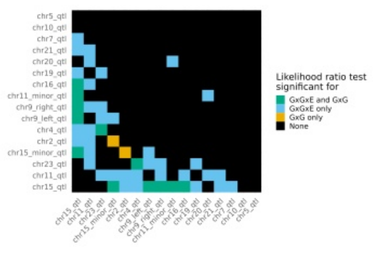


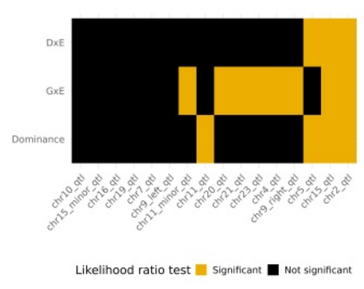
Embryonic heart rate assay

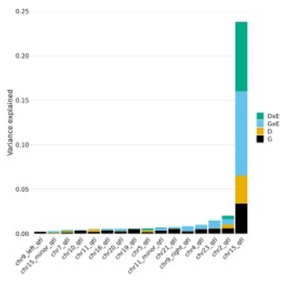
We image medaka embryos at different temperatures (21 °C, 28 °C, 35 °C) in a high-throughput 96-well plate system using the Acquifer imaging machine (www.acquifer.de). We then extract heart rate information using a Fourier transform-based image analysis approach.

Dominance, GxG and GxE effects

We tested for the presence of gene-by-gene (GxG) and gene-by-environment (GxE) interactions among the 16 Quantitative Trait Loci (QTLs) that we discovered using likelihood ratio tests and a Bonferroni correction. We detected that several loci exhibit GxG associations, some of which are dependent on the environment (that is GxGxE effects). Dominance and/or GxE are detectable in 10 of 16 QTLs. The fraction of the overall variance explained by different interaction terms is highly variable among QTLs, highlighting the complex relationship among genotypes, phenotypes, and the environment.

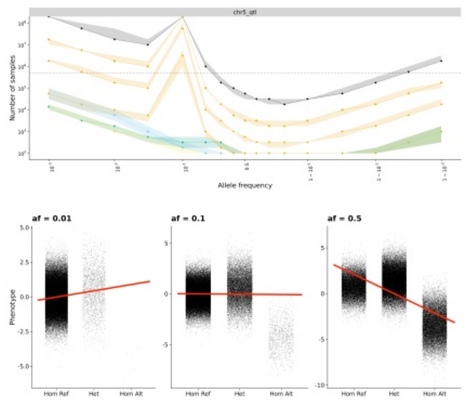






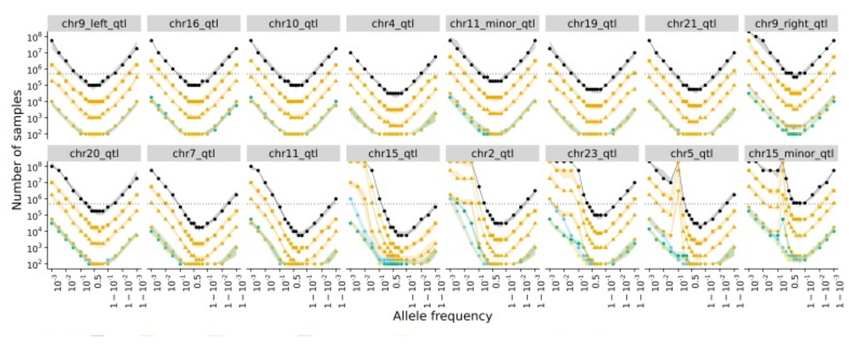
Over-dominance can cause wrong effect direction estimates and missed associations

While doing our simulations we noticed that over-dominance causes an interesting behaviour when the data is analysed under a model that ignores dominance effects: depending on allele frequency, the estimated effect under a linear model can be reversed (when the recessive allele is very rare) or undetectable (when the heterozygous and recessive phenotypes cancel each other out, resulting in a flat linear estimate). This is the reason for the abrupt spike in the required minimum sample size for some of the loci in our simulation.



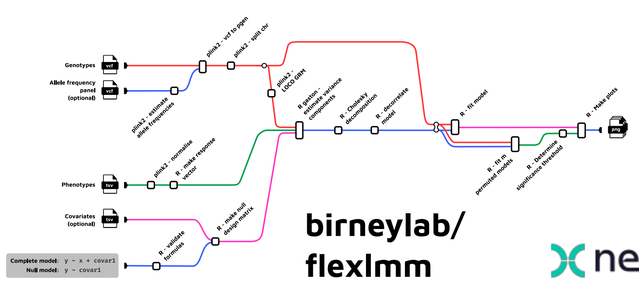
Simulation of the effect of model misspecification

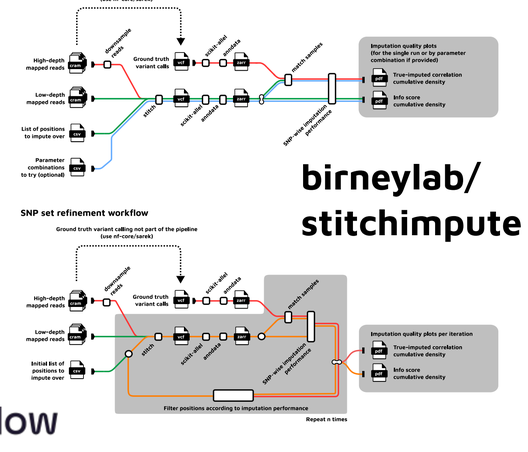
We tested what happens when data generated under a model that includes higher order interactions (GxG, GxE, dominance) is discovered with a simplified model that ignores some of the interaction terms. We asked ourselves what would be the minimum sample size required to discover the associations ($p < 5*10^{-8}$) when using models at different levels of misspecification. We also tested the effect of uncertainty in the measurement of the environmental factor. We observe that the sample size required increases drastically when the environment is completely ignored. The effect sizes used in the simulation are the real effect sizes measured in the 16 QTLs that we discovered. Genotypes are drawn from a binomial distribution at different minor allele frequencies. The environment is sampled from real water temperature measurements (data source: Yoshimura lab).



Computational pipelines

We developed a genomic imputation pipeline (github.com/birneylab/stitchimpute, Pierotti et al, 2023) based on STITCH (Davies et al, 2016), and a linear mixed model GWAS pipeline (github.com/birneylab/flexlmm) for this project, and made them available to the community for facilitating similar studies.





www.ebi.ac.uk

European Research Council
Established by the European Commission

EMBL's European Bioinformatics Institute (EMBL-EBI)
Wellcome Genome Campus, Hinxton,
Cambridgeshire. CB10 1SD. UK.
T +44(0)1223 494 444