Gradients or clusters? The genomic architecture of adaptive radiations

MSC PROJECTS · DATA ANALYSIS · SIMULATIONS · MODELLING

Supervisors: Stephan Peischl, Interfaculty Bioinformatics Unit, Univ. of Bern

Pooja Singh, Institute of Ecology and Evolution, Univ. of Bern and EAWAG

Contact: stephan.peischl@unibe.ch · pooja.singh@unibe.ch

Background

Hundreds of cichlid fish species have evolved in the East African Great Lakes Victoria and Malawi within a short time. Each species is a discrete *combination* of a limited set of ecological and mating traits associated with adaptations to diet, water depth, habitat as well as behaviours such as attracting mates through body colour variation. The genetic basis of these traits is polygenic and often **redundant**: many different genotypes give the same phenotype, and the same trait can have a different genetic basis in different species. Under gene flow, recombination keeps breaking up favourable allele combinations, and theory predicts a **smooth gradient** of intermediate phenotypes rather than **discrete clusters**. There are two candidate mechanisms can hold trait combinations together: **chromosomal inversions**, which suppress recombination and can lock co-adapted alleles into a single block, and **assortative mating**, which reduces gene flow between trait combinations directly. Large inversions segregate with habitat in Lake Malawi, while in Lake Victoria the genomic architecture is highly modular and largely unlinked, and long-range linkage disequilibrium is hypothesised to be maintained by non-random mating. Furthermore, genetic redundancy is reported in Lake Victoria modular trait architectures. However, the landscape of inversions has not been investigated in Lake Victoria cichlids. Critically, no theory exists to understand the interaction of redundant genomic architectures and inversions in speciation with gene flow.

The idea

We want to understand when selection in a heterogeneous environment produces discrete ecotypes, and when it produces a continuum. Our working hypothesis is that the answer depends on how the **genetic architecture** (number of loci, effect sizes, genomic redundancy) interacts with the **environmental architecture** (a continuous gradient vs discrete habitats, and how many traits differ between niches), and on mechanisms that maintain linkage, such as inversions and assortative mating (Figure 1). Preliminary simulations show that inversions, even without any fitness effect of their own, can turn a smeared-out phenotype distribution into discrete ecotypes under gene flow.

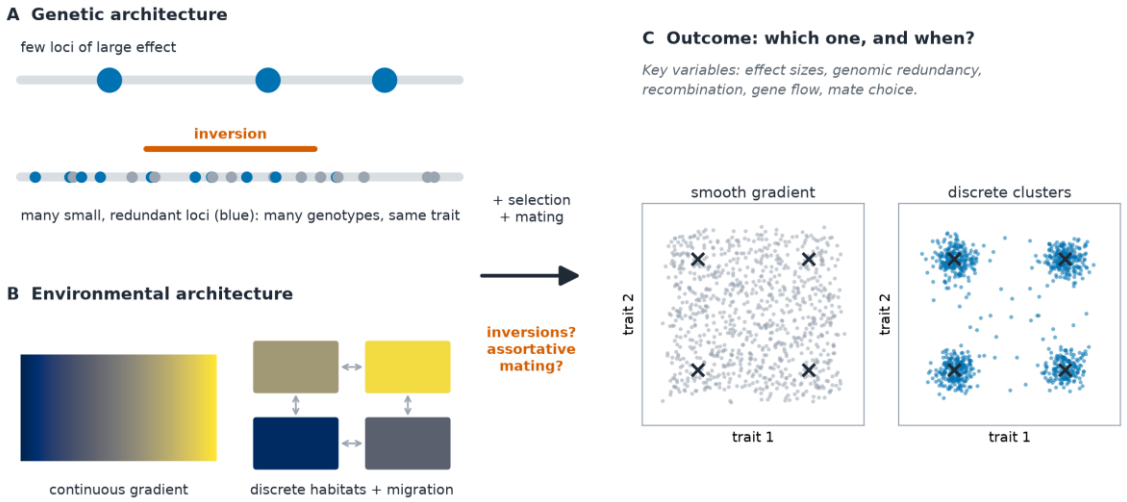


Figure 1. The idea. (A) The same trait can be built by a few large-effect loci or by many small, redundant loci; an inversion suppresses recombination among the loci it captures. (B) Selection can vary along a continuous gradient or between discrete habitats. (C) The outcome can be a smooth gradient of intermediates or discrete clusters on the niche optima (×). Which outcome do we get, and when?

Possible projects

We combine genomic data from cichlid radiations with theory. Projects can range **from data analysis to large-scale simulations or mathematical modelling, or a combination of these**, depending on the interests and background of the student. Possible directions include:

Data analysis. Large-scale genomic analysis of inversions in hundreds of genomes of >100 Lake Victoria cichlid fishes. Overlap inversions with known Genotype-Phenotype genetic outliers for ecological and mating traits. Integrate inversion analysis with calculations of linkage disequilibrium across chromosomes to infer recombination landscape. Quantify genetic redundancy in Genotype-Phenotype maps. Integrate results into an evolutionary interpretation of inversions, genetic redundancy, and adaptive radiation.

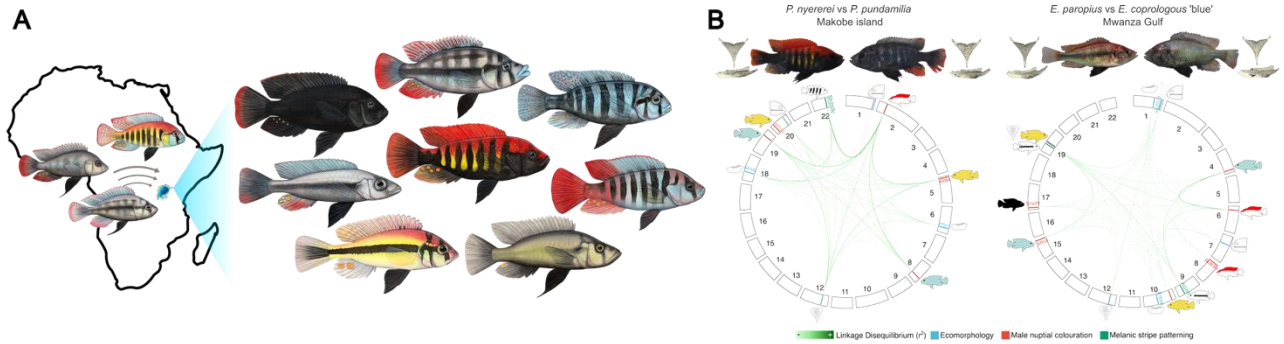


Figure 2. A) Overview of the Lake Victoria cichlid radiation from ancient hybridisation. Hundreds of species evolved with diverse phenotypes in <16,000 years. B) Genetic architecture of speciation from modules dispersed across the genome. Linkage disequilibrium among modules of different traits emerged from non-random mating. Polygenic modules of hybrid ancestry are combined in different ways in different species (i.e. combinatorial speciation)

Large-scale simulations. Individual-based simulations (SLiM or our own simulator) of polygenic adaptation in structured populations. Vary effect sizes, redundancy, migration and habitat structure, and ask when gradients or clusters arise; compare inversions and assortative mating as alternative or complementary routes to discreteness; and derive genomic signatures (F_{ST} profiles, linkage disequilibrium, inversion frequencies) that can be tested in the data.

Mathematical modelling. Simple deterministic models that capture the core of the problem, for example a few loci under migration–selection balance. When can linkage disequilibrium between traits be maintained against recombination and gene flow? Under which conditions can an inversion that captures a favourable allele combination invade? Analytical results give intuition and make the simulations interpretable.

Combinations. For example, use simulations or theory to generate predictions and test them in the cichlid data, or fit simulated scenarios to the observed number of discrete trait combinations.

Aims

(a) Determine when selection in a heterogeneous environment produces a smooth gradient and when it produces discrete clusters of phenotypes.

(b) Test whether, and under which conditions, inversions and/or random mating promote discrete ecotypes.

(c) Characterise the role of inversions in rapid combinatorial speciation and adaptive radiation when genomic architecture is polygenic and redundant.]

Specific aims are set together with the student.

Methods and requirements

Depending on the project: population-genomic analyses (whole-genome re-sequencing data, SNP calling, inversion identification with PCA-based methods), SLiM, R or Python, mathematical modelling (e.g. Mathematica or SymPy), command line, High-Performance Computing (HPC). An interest in evolution and some programming experience are helpful; no prior experience with SLiM or with the data is required. Students with a background in biology, bioinformatics, mathematics or physics are all welcome.

Further reading: Kirkpatrick & Barton (2006) *Genetics* 173: 419–434 · Yeaman (2015) *Am Nat* 186: S74–S89 · Marques, Meier & Seehausen (2019) *TREE* 34: 531–544 · Singh et al. (2025) *bioRxiv* doi:10.1101/2025.07.07.663194 · Yeaman S, Gerstein AC, Hodgins KA, Whitlock MC (2018). PLoS Genet 14(10): e1007717. https://doi.org/ 10.1371/journal.pgen.1007717