# Detecting signatures of positive selection

# Literature for further reading (& sources of figures used here)

**Oleksyk et al., 2010 (doi:10.1098/rstb.2009.0219)**

**Vitti et al., 2013 (doi:10.1146/annurev-genet-111212-133526)**

**Hoban et al., 2016 (10.1086/688018)**

**Weigand et al., 2018 (doi.org/10.1093/zoolinnean/zly007)**

**Koropoulis et al., 2020 (doi.org/10.1007/978-1-0716-0199-0_5)**

# Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions

Sean Hoban, Joanna L. Kelley, Katie E. Lotterhos, Michael F. Antolin, Gideon Bradburd, David B. Lowry, Mary L. Poss, Laura K. Reed, Andrew Storfer, and Michael C. Whitlock

Findin

Solutic

Sean Hoban
Andrew Sto

**Table 1:** Approaches to identifying loci involved in local adaptation

| Approach | Data collected/resources required | What analysis reveals | Review articles/programs |
|---|---|---|---|
| Genetic differentiation outlier tests | Genome-wide SNPs from multiple populations | Allele frequencies for a SNP or SNPs that are differentiated across populations above what is expected from neutrality | See table A1 |
| Genetic-environment association | Genome-wide SNPs from multiple populations and environmental data for each population | Alleles at a SNP or SNPs that are associated with environmental variables over space | See table A1 |
| QTL mapping in a reciprocal transplant field experiment | Hybrids ($F_2$s, BCs, RILs, etc.) between locally adapted populations grown and phenotyped for fitness traits in reciprocal transplant common garden experiment | Use of hybrids allows identifying QTLs involved in local adaptation and the effect size of those QTLs on fitness; can resolve whether trade-offs at individual loci underlie local adaptation | Reviewed in Anderson et al. 2011; Savolainen et al. 2013 |
| GWAS | Genome-wide SNPs from hundreds of individuals grown in one or multiple common gardens; phenotypes and/or fitness for each individual | Identifies SNPs that are associated with traits associated with fitness measured under field conditions | Key example study: Fournier-Level et al. 2011; commonly used: TASSEL (Bradbury et al. 2007); EMMA (Kang et al. 2008); GCTA (Yang et al. 2011) |
| Population-specific selective sweeps | Genome-wide SNPs from at least two populations and a recombination map | DNA sequences with longer-than-expected regions of extended haplotype homozygosity, which is consistent with a recent selective sweep in one of the populations | XP-EHH (Sabeti et al. 2007); hapFLK (Fariello et al. 2013) |

Findin

Soluti

Sean Hoban
Andrew Sto

**Table 1:** Approaches to identifying loci involved in local adaptation

| Approach | Data collected/resources required | What analysis reveals | Review articles/programs |
|---|---|---|---|
| Genetic differentiation outlier tests | Genome-wide SNPs from multiple populations | Allele frequencies for a SNP or SNPs that are differentiated across populations above what is expected from neutrality | See table A1 |
| Genetic-environment association | Genome-wide SNPs from multiple populations and environmental data for each population | Alleles at a SNP or SNPs that are associated with environmental variables over space | See table A1 |
| QTL mapping in a reciprocal transplant field experiment | Hybrids ($F_2$s, BCs, RILs, etc.) between locally adapted populations grown and phenotyped for fitness traits in reciprocal transplant common garden experiment | Use of hybrids allows identifying QTLs involved in local adaptation and the effect size of those QTLs on fitness; can resolve whether trade-offs at individual loci underlie local adaptation | Reviewed in Anderson et al. 2011; Savolainen et al. 2013 |
| GWAS | Genome-wide SNPs from hundreds of individuals grown in one or multiple common gardens; phenotypes and/or fitness for each individual | Identifies SNPs that are associated with traits associated with fitness measured under field conditions | Key example study: Fournier-Level et al. 2011; commonly used: TASSEL (Bradbury et al. 2007); EMMA (Kang et al. 2008); GCTA (Yang et al. 2011) |
| Population-specific selective sweeps | Genome-wide SNPs from at least two populations and a recombination map | DNA sequences with longer-than-expected regions of extended haplotype homozygosity, which is consistent with a recent selective sweep in one of the populations | XP-EHH (Sabeti et al. 2007); hapFLK (Fariello et al. 2013) |

# Selection scan design

## Important aspects:

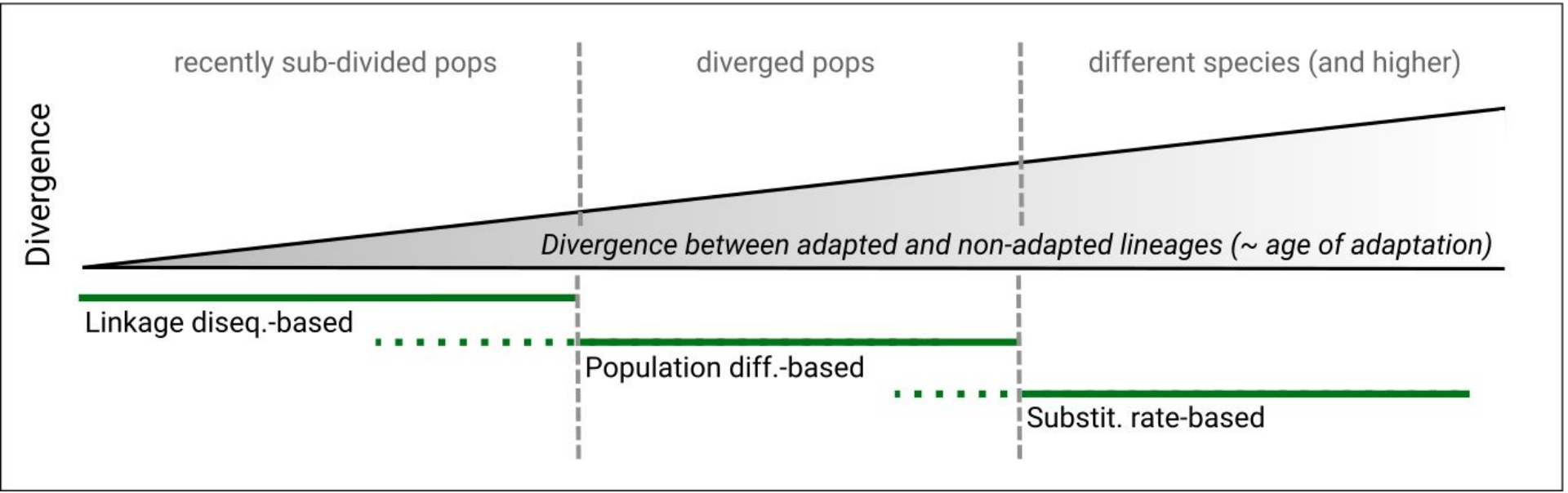
What is the **time frame** (macroevolution / **microevolution**)

What is the **scope of the comparison** (within / **among populations**)

How are populations **distributed across the environmental condition** of interest (**clinally** / **only at extremes**)

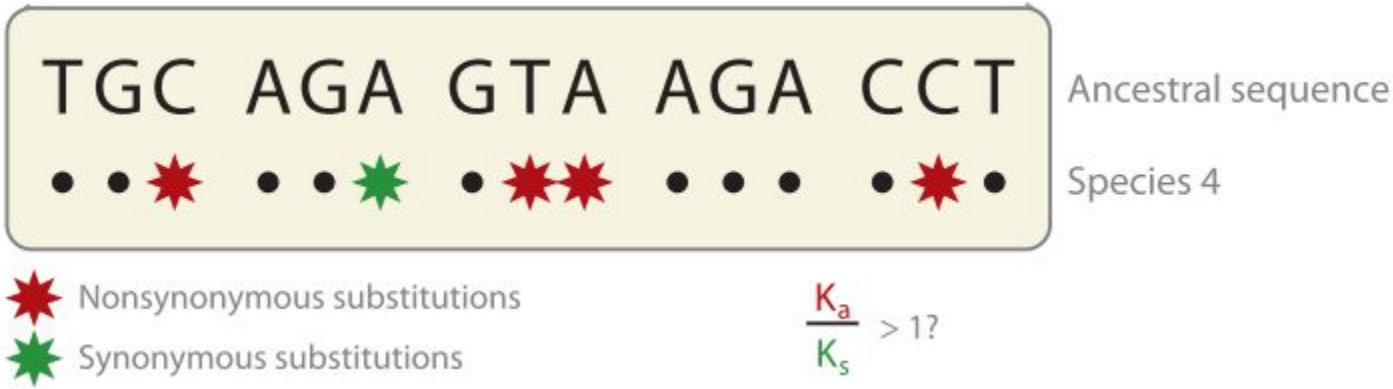
What are your **data** (**individual-based resequencing** / phased / poolseq)

# Time frame matters



# Let’s start the activity!

# Macroevolution - substitution-based

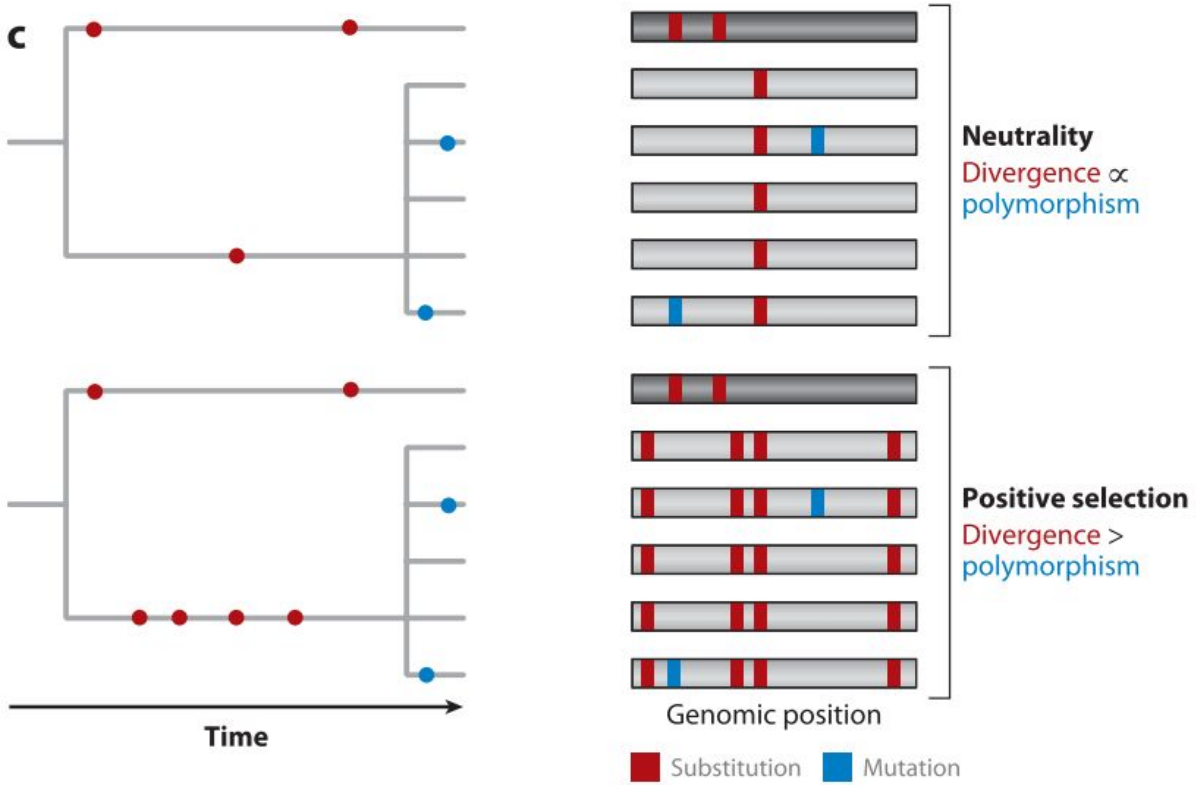



Ka/Ks (= dN/dS or ω)

Software PAML

Input: (consensus) fasta

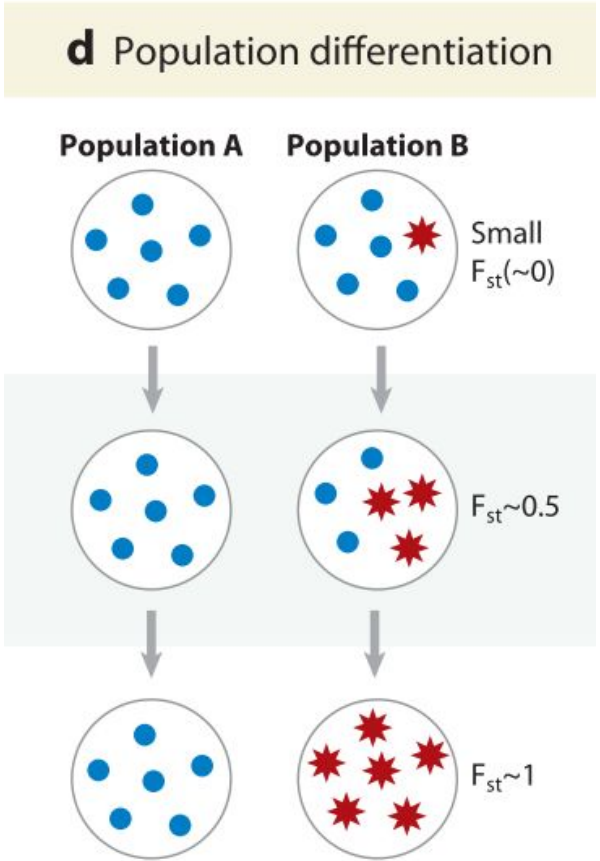
# Macroevolution - substitution-based




McDonald-Kreitman test

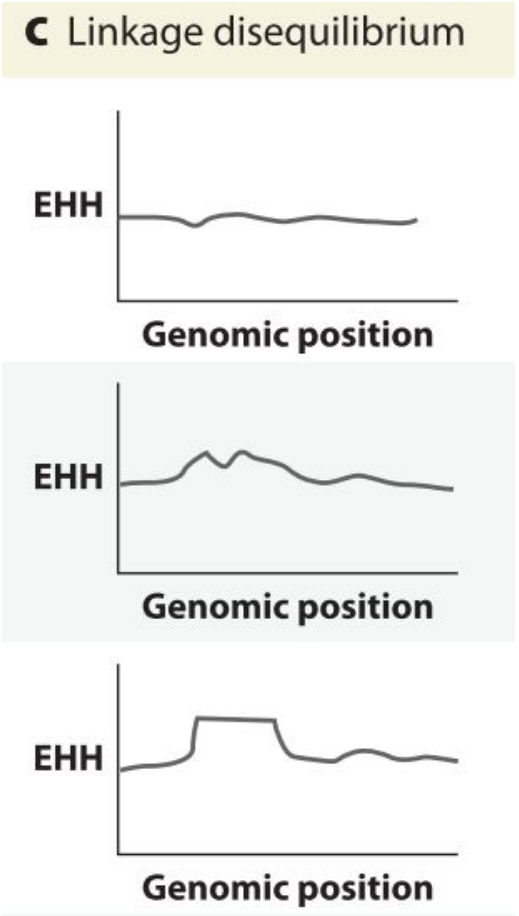
# Microevolution - differentiation-based

- use of the focal adapted and “background” non-adapted populations
- **Measures of differentiation:**
  - **Fst**, Rho, Dxy, XtX,...
- Model-based test for excess differentiation
- **Software** BayPass, BayeScan, FLK,...
- Input: allele counts/frequencies, population-level sampling



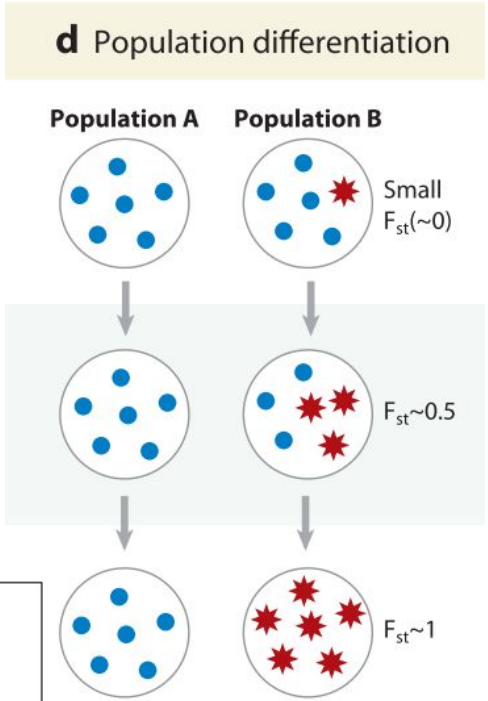
# Microevolution - linkage disequilibrium-based

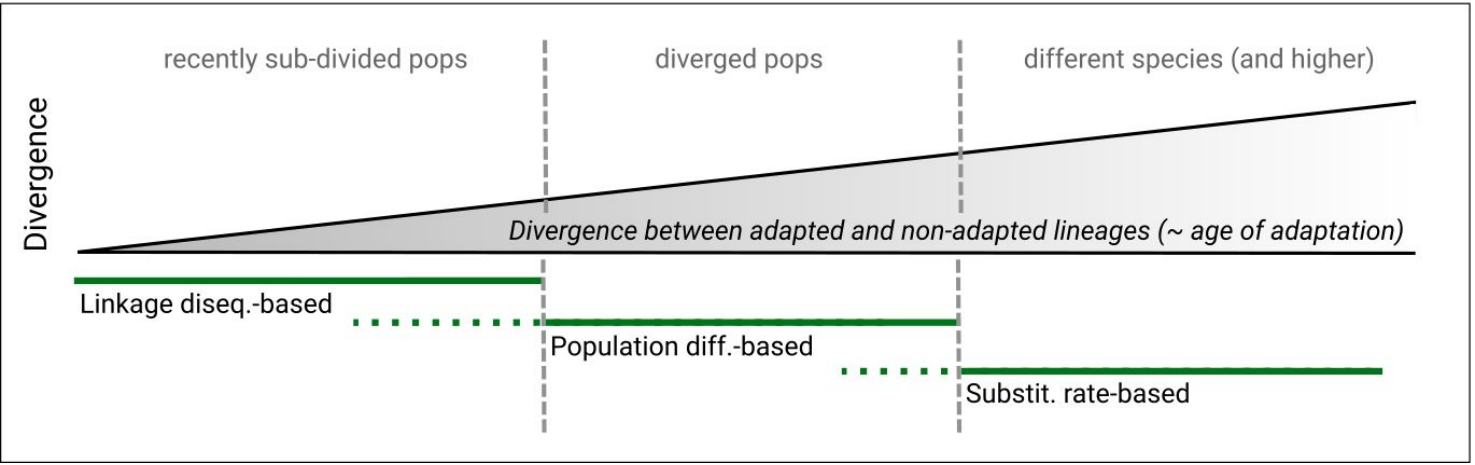
- High prevalence of longer (selected) haplotype
- Suitable for soft sweeps
- Integrated haplotype score (iHS), Cross-population extended haplotype homozygosity (XP-EHH),...
- Software: selScan
- Input: haplotypes, even single population is ok



## Differentiation-based scans:

- intuitive
- frequently used
- applicable to wide spectrum of divergences



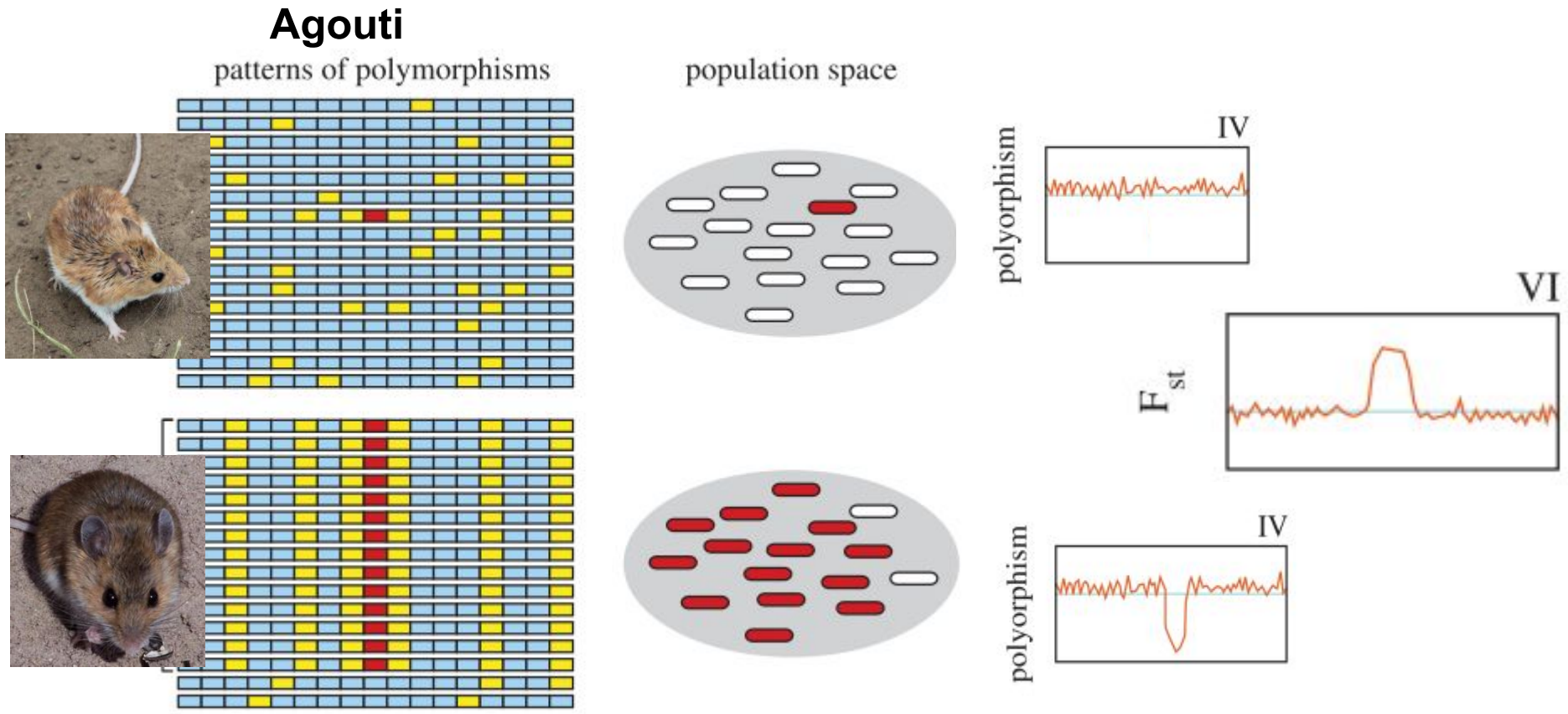



# Fst

*“the most widely used descriptive statistics in population and evolutionary genetics”*

- Combines **genetic differentiation between populations** and **diversity within** them.
- High values - high differentiation between populations, low diversity within
- Between 0 - 1

# Signal of positive selection in a gene



# How to design selection scan for your study - best practices

1. Try to understand your species of interest - Heritable difference? Mating system? Genetic diversity? Divergence between non-adapted and adapted populations?
2. Get familiar with available (up-to-date) methods - they develop quickly!
3. Pick multiple (>1) complementary methods, i.e. Fst + LD-based + BayeScan
4. Available literature involving selection scans and simulations/functional validation might be good inspiration.
5. Re-analyze with multiple thresholds - are the results consistent?
6. Visualize!!!

# How to design selection scan for your study - best practices

7. Try to find as much as possible about your candidate genes - GO enrichment/KEGG pathways/protein-protein interaction/modeling the impact of changes on protein structure/detecting the origin of the selected allele,... Be creative!

8. Combine with QTL mapping/GWAS

9. Functional validation of your candidate genes is optimal - but laborious and often limited to model organisms