Ecological Genetics – HS18

v0.2

Gleb Ebert

October 13, 2018

Vorwort

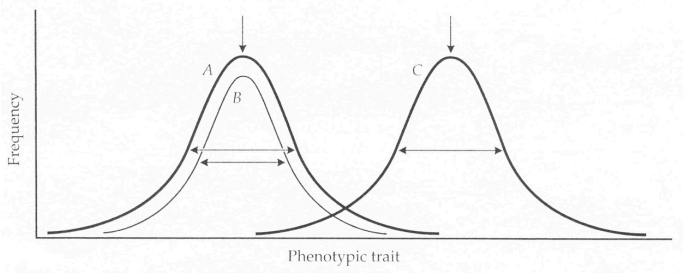
This document aims to summarize the lecture Ecological Genetics as it was taught in the autumn semester of 2018. Unfortunately I can't guarantee that it is complete and free of errors. You can contact me under `glebert@student.ethz.ch` if you have any suggestions for improvement. The newest version of this summary can always be found here: `https://n.ethz.ch/~glebert/`

Contents

1 Introduction

"Nothing in biology makes sense except in the light of evolution"
— Theodosius Dobzhansky

In 1971 Ford wrote that ecological genetics deal with the adjustments and adaptations of wild populations to their environment. According to Conner and Hartl (2004) ecological genetics is the study of the process of phenotypic evolution occurring in present-day natural populations and is concerned with the genetics of ecologically important traits, that is, those traits related to fitness.
Phenotypic evolution is the change in the mean or variance of a trait across generations due to changes in allele frequencies.



Ecologically important traits are closely tied to fitness and are important in determining an organisms adaptation to its natural environment. Adaptation is a heritable phenotypic trait that has evolved in a population in response to a specific environmental factor and improves the survival or reproduction of its carriers. It can also be seen as a process whereby the members of a population become better suited to some feature of their environment through change in a characteristic that affects their survival or reproduction. Of the four key evolutionary processes, only natural selection consistently leads to adaptation (mutations, genetic drift and gene flow don't).

Uses of ecological genetics include

- agriculture (crop improvement)
- medicine (e.g. antibiotics)
- conservation measures (assisted migration)
- geographical differences between populations
- changes in species composition
- habitat adaptation & speciation

Fields related to ecological genetics include

- population genetics
- ecology
- evolutionary biology
- phylogenetics
- quantitative genetics
- statistics
- molecular biology
- epigenetics
- genomics

2 Species

Species are the fundamental unit in ecology, evolution and conservation legislation. Depending on the research question, adequate species identification, assignment of samples to populations or discrimination of individuals may be of relevance.

2.1 Species concepts

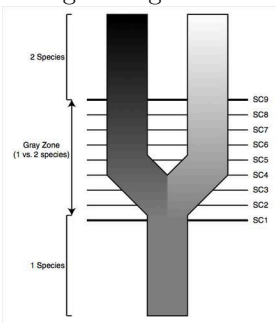
- Morphological / typological species concept: focus on similar morphology
- Biological species concept: group of potentially or actually interbreeding populations which are **reproductively isolated** from other such groups
- Phylogenetic species concept: focuses on monophyletic lineages

2.2 Operational Taxonomic Unit

An OTU is a group of organisms that is treated as a distinct evolutionary unit for the purposes of research underway. They are often applied when one or several species concepts fail. Once identified and research has been completed, OTUs should receive full taxonomic treatment and be given a scientific name if possible. OTUs are sometimes called molecular operational taxonomic unit (MOTU) when molecular methods are used.

2.3 Unified Species Concept

The only necessary property of species is that they form a separately evolving matepopulation (involves dynamics of gene flow and separation) lineage. The concept separates species conceptualization and separation. All criteria can be used for species delimination and any one of the properties is accepted as evidence for the existence of a species. More properties provide a higher degree of corroboration.



2.4 Identification of Species

Difficulties may include

- species-specific traits are not (always) visible
- differences are cryptic
- direct observation may be difficult and traces may be confused
- undescribed species may occur

Parataxonomy may be used when identification is difficult. It sorts the material to species on the basis of external morphology without considering taxonomy.

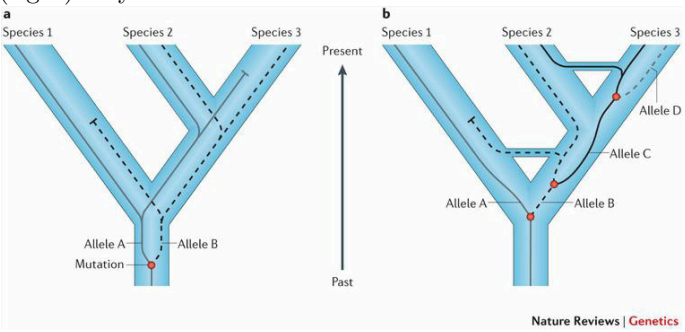
2.5 Species delimination

Species delimitation is the act of identifying species-level biological diversity (independent evolutionary lineages). Most methods fit models to collected data to make (often different) simplifying assumptions. Incongruence across methods may occur due to differences in the power to differentiate lineages or due to violations of one or several assumptions made by a given method. Fundamentally there are two approaches. Some models can assign samples to groups without being given information first (STRUCTURE, Structurama, Geneland). Others need the user to assign samples

to putative lineages (BPP, iBPP, spedeSTEM, DISSECT, tr2).

2.5.1 Problems of Species Trees

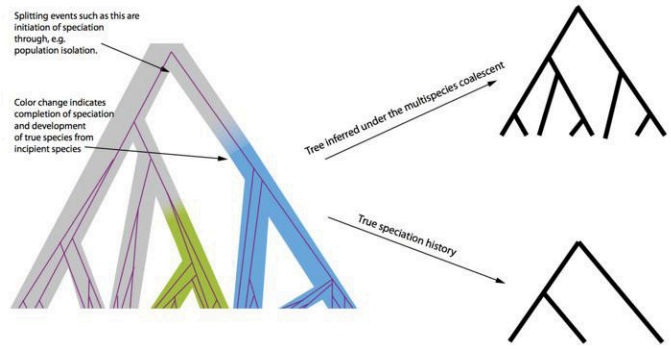
Problems like imcomplete lineage sorting (left) or gene flow (right) may occur.



2.5.2 Bayesian Species Identification under the Multispecies Coalescent

This method is currently the most used approach for species delimination. The multispecies coalescent describes the genealogical relationships between DNA samples from several species. Simplifying assumptions include:

- species phylogeny unknown
- complete isolation after divergence
- no recombination



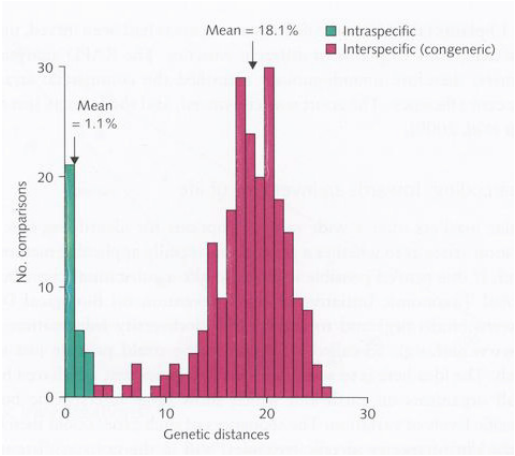
The above graph shows that the MSC approach identified populations as separate species from a simulated data set. MSC delimits structure, not species.

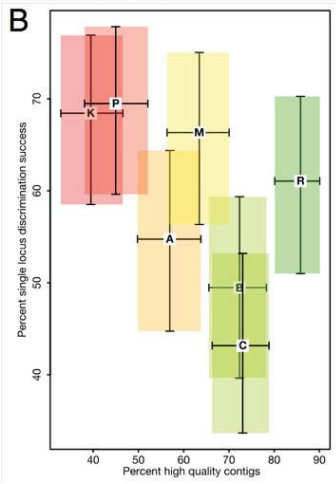
2.5.3 Recommendations for Species Delimination

- use at least 10 samples from all lineages
- simulate before analysis
- use several complementary methods
- combine genetic with nongenetic data
- define the used species concept
- be cautious

2.5.4 DNA Barcoding

DNA barcoding is a method for rapid identification of species through the analysis of short, standardized gene regions. These have to be universal for all animals or plants, have to be amenable to the production of bidirectional sequences with little ambiguity and allow for discrimination of most species. The barcoding gap defines the distance between species. Genetic distances are based on the extent of nucleotide sequence devergence. Some groups like orchirds lack the barcoding gap.





Examples of application are the identification of amphibian species diversity and abundance after epidemic diseases in Panama or restriction of shaving brushes to only use hair from the Hog badger instead of the Eurasian badger.

3 Molecular Markers

Molecular markers are polymorphic proteins or DNA sequences and reveal different alleles within individuals, populations or species. Ideally they can be used as indicators of genome-wide variation.

- **Chromosome based markers**
  - Numbers and staining patterns
- **Enzyme based**
  - Allozymes
- **DNA based**
  - Restriction fragment length polymorphisms (RFLPs)
- **DNA & PCR based**
  - Random amplified polymorphic DNA (RAPD)
  - Amplified fragment length polymorphism (AFLP)
  - Microsatellites (SSRs)
  - DNA sequencing and SNPs

3.1 Genome

The size of genomes can differ between species by large factors. Size and complexity are not coupled, as non-coding regions are the main reason for big genomes. These regions contain repeated sequences like tandem repeats or interspersed repeats (transposable elements). The number of

chromosomes is limited, because at some point there would be problems with the spindle apparatus. Introns are spliced after transcription while intergenic regions aren't.
Mitochondria and chloroplasts have their own genome (mtDNA and cpDNA respectively). Depending on the species, different organelle genomes should be used for comparisons. There is much more organelle DNA in a cell than there is nuclear nDNA and it is easier to amplify because of its high conservation.

3.2 Widely used genetic markers

Marker	Advantages	Disadvantages
Allozymes	• Cheap • Universal protocols	• Requirement for fresh or frozen material • Potentially direct target of selection • Limited number of available markers • 'No longer used' (<1998)
Microsatellites	• Informative (large number of alleles, high heterozygosity) • Easy to isolate	• High mutation rate • Complex mutation behaviour • Difficult to automate • Cross-study comparisons are difficult
AFLPs	• Cheap • Produces a large number of markers • Easy to establish in the lab	• Mainly dominant • Difficult to analyse • Difficult to automate • Cross-study comparisons are difficult
DNA sequencing	• Highest level of resolution possible • Not biased • Cross-study comparisons are easy • Data repositories already exist (e.g. NCBI)	• Sanger sequencing: significantly more expensive than the other techniques • NGS: cost per base (bp) very low • NGS: computational intense analyses
SNPs arrays	• Low mutation rate • High abundance • Easy to type • Cross-study comparisons are easy; data repositories already exist	• Expensive to isolate • Ascertainment bias • Low information content of a single SNP

3.2.1 Microsatellites

- SSR (simple sequence repeat) and STR (short tandem repeat)
- highly polymorphic: mutation rates between 10^{-6} and 10^{-2} per locus per generation
- widely used to assess genetic variation in animals, plants and fungi as they are highly variable between individuals
- codominant
- mostly evolutionary neutral, as they are in intragenic regions
- PCR-based (primers, agarose-gel electrophoresis)
- capillary sequencers use fluorescence labelled primers
- 10-20 SSRs per study
- the mutation mechanism is called slipped-strand mispairing: polymerase slips of and when rejoining doesn't know which repeat it already copied; insertions and excisions happen, but the latter seem to be corrected in nature

3.2.2 Structural variation (SV)

- Microsatellite repeats
- 1bp indels
- More complex insertions and deletions
- Copy number variants (CNVs) (> 1kb)
- ...

3.2.3 AFLPs

- 100-4'000 random markers
- dominant (homo- and heterozygotes indistinguishable)

1) DNA extraction
2) Digestion by restriction enzyme MseI and EcoRI
3) Ligation of the adaptors MseI and EcoRI
4) Selection amplification (+3/+3 bp)
5) Sequencing
6) Binary data matrix (peak present/absent)

3.2.4 SNP microarrays

DNA is hybridised to predefined probes. Only common and known SNPs are called. This is the ascertainment bias.

4 Sampling

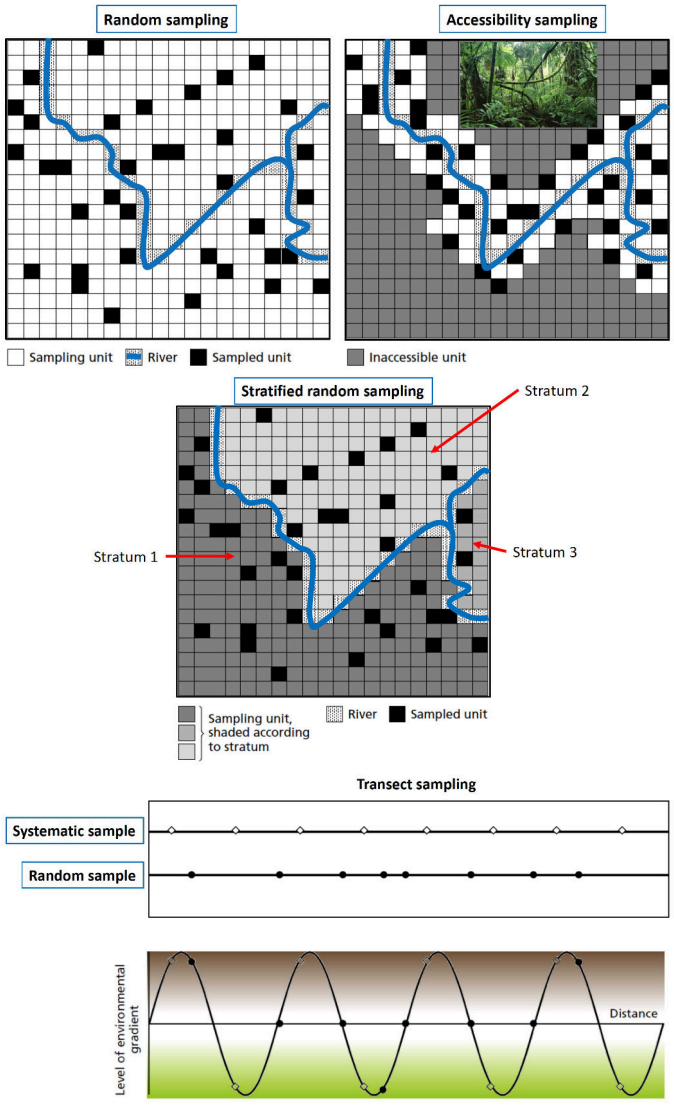
Sampling determines...

- The chances of answering your research questions
- The time you may have to spend to answer your questions
- Research costs
- The likelihood of obtaining research permits
- ...

4.1 Populations

Population type	Definition
Genetic	All individuals which are connected by gene-flow
Ecological	Group of organisms occuring in a particular area at a particular time
Statistical	The universe of times that are under study

4.2 Individuals



Do you sample proportionally or equally across strata? Randomly or systematically? There is no single right solution. One should always consider the circumstances.
Roughly 20-30 individuals represent a population ($> 80\%$ of all alleles).

4.3 Important considerations

- Documentation: archived, reproducable, verifiable, new technologies
- Storage: adequate storage and transport, test in advance!
- Permits: sampling, handling animals, export and import

4.4 Nagoya protocol

The convention on biological diversity from 1993 had the objectives of conservation and sustainable use of biological diversity as well as the sharing to benefits arising from the utilisation of genetic resources. The Nagoya protocol from 2010 is a supplementary agreement that expands on the fair and equitable sharing of benefits arising out of the utilisation of genetic resources.

- **Access obligations**: Provide fair and non-arbitrary application procedures and issue permits when access is granted
- **Benefit-sharing obligations**: Share the value of genetic resources and traditional knowledge with developing countries
- **Compliance obligations**: Ensure that genetic resources and traditional knowledge have been accessed in accordance with prior informed consent