Assignment Testing Applications (and Misapplications)

With special emphasis on using Geneclass2

Assignment Methods and short timescales

Traditional popgen models assume equilibrium (stable balance)

Events on shorter timescales (ecological) are “noise” in those models

AMs allow study of contemporary events

Classification vs. **Clustering**

How much prior information is available?
Family? Population? Species?

Clustering- categories are not predefined and must be constructed from the data (STRUCTURE)

Classification- individuals are assigned to pre-defined categories (GENECLASS2)

Classification problems

- Compute discriminant function from samples from potential source populations
- Classify unknowns to the groups with highest discriminant score
- Discriminant function: expected genotypic frequency distribution
  - Assumes H-W and linkage equil. in each source population

Biological Questions

- Origin of Specific Individuals
- Dispersal
- Genetic Mixture Analysis
- Hybridization

Biological Questions

- Origin of Specific Individuals
- Dispersal
- Genetic Mixture Analysis
- Hybridization

Origin of Specific Individuals

- Detection of illegal harvests/trade routes
- Captive Breeding (exclusion of non-target individuals)
- Control Mechanisms for Bioinfestation

Detection of Poaching- Manel et al. 2002

- Comparison of Bayesian (STRUCTURE) and partially Bayesian (Geneclass) methods applied to wildlife forensics
- Fully Bayesian method performed better than the exclusion test **BUT**:

It assumes ALL populations were sampled (Geneclass exclusion method doesn't)

Best thing to do= Use Both Methods!

Manel et al. 2002

Table 2. Performance of the exclusion simulation test (Geneclass) and Bayesian assignment test (Structure) used to assign individuals to a population of origin.

			Geneclass		*Structure*	
Species	F_{ST}[a]	H_{mean}[a]	*individuals correctly assigned (%)*[b]	*cases where the correct population is excluded (%)* ($p < 0.001$)	*individuals correctly assigned (%)* ($T > 0.999$)	*cases where the correct population is excluded (%)* ($T < 0.001$)
Bighorn sheep	0.262 (0.056–0.020)	0.574 (0.540–0.640)	64.39	0.719	78.41	0.360
Taurine cattle	0.116 (0.076–0.149)	0.500 (0.473–0.557)	22.13	0.791	62.45	0.000
Fish	0.377 (0.097–0.569)	0.503 (0.358–0.684)	37.37	1.476	39.48	0.000
Ibex	0.164 (.067–0.278)	0.399 (0.309–0.464)	1.77	0.588	27.65	0.000
Goats	0.140 (0.049–0.195)	0.596 (0.328–0.703)	17.23	2.153	60.77	0.239
Drosophila	0.295 (0.124–0.541)	0.430 (0.334–0.550)	51.74	0	65.97	0.000
Bumble bees	0.401 (0.108–0.660)	0.401 (0.122–0.628)	35.37	1.22	60.97	0.000
Brown trout	0.285 (0.114–0.499)	0.289 (0.197–0.426)	17.30	0	45.67	0.000
Apis mellifera	0.341 (0.102–0.564)	0.504 (0.262–0.653)	43.62	1.064	51.06	0.000
Bears	0.324 (0.076–0.149)	0.552 (0.360–0.694)	75.43	0.683	84.982	0.341
Median (SD)			36.37 (21.750)	0.76 (0.620)	60.87 (16.483)	0.00 (0.149)

[a] *Mean F_{ST} and mean gene diversity (H_{mean}) are given for each data set, with the range for all population pairs in parentheses.*
[b] *All populations that are not the true population of origin were excluded ($p < 0.001$) but not the true population of origin ($p > 0.001$).*

The One That Did Not Get Away…

Primmer et al. 2000:

- Fishing competition in Finland
- Competitor claimed to have caught a 5.5kg salmon
- 7 Msats excluded the competition location as a source

The One That Did Not Get Away…

Primmer et al. 2000:

- Fishing competition in Finland
- Competitor claimed to have caught a 5.5kg salmon
- 7 Msats excluded the competition location as a source

The offender confessed to purchasing the fish at a local market!

Biological Questions

- Origin of Specific Individuals
- Dispersal
- Genetic Mixture Analysis
- Hybridization

Dispersal- Berry et al. 2004

- Comparison of skink dispersal via mark-recapture vs. genetic assignment testing
  - AT require less time, less fieldwork than m-r
- Previous studies addressed cases where dispersal was unlikely
- Fully vs. Partially Bayesian- both performed well, partial slightly better.
- Higher Stringency=Increased Accuracy but more exclusions, especially with partial Bayesian (geneclass)

Take home message: AT are a satisfacotory replacement for m-r, but use Full AND Partial Bayesian methods

Biological Questions

- Origin of Specific Individuals
- Dispersal
- Genetic Mixture Analysis
- Hybridization

Genetic Mixture Analysis

- What proportion of individuals come from each source population
  - e. g. Fisheries mgmt, Colonization
- What proportion of genes in an admixture come from each source: Beaumont et. al 2001, Genetic diversity and introgression in Scottish Wildcat

Scottish Wildcat diversity and introgression

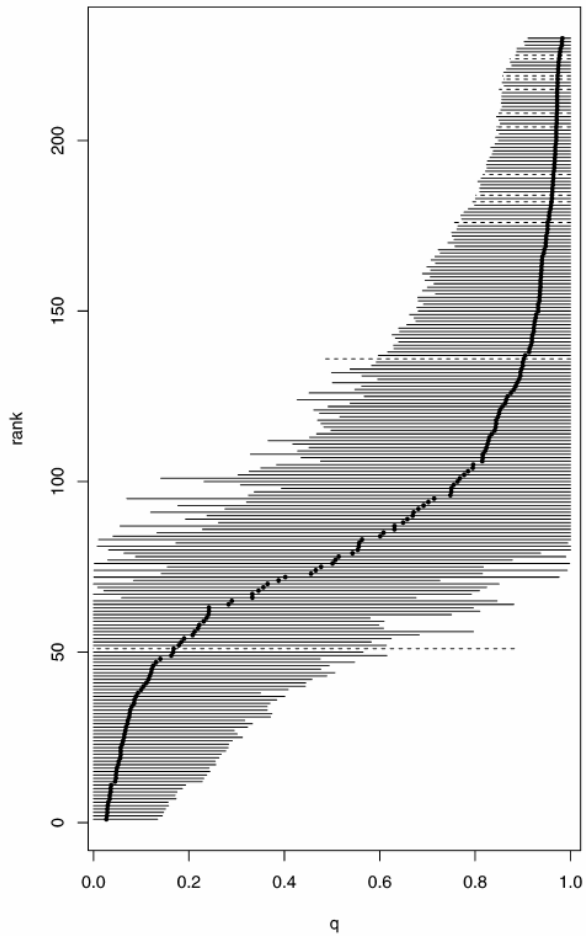

Fig. 5 This figure illustrates the distribution of $\hat{q}$ among individuals. The values of $\hat{q}$ have been ranked and the ranks are plotted against $\hat{q}$. Also shown are lines giving the 95% equal-tail posterior probability intervals for each individual. The dotted lines refer to museum specimens.

- Proportion of an individual's genome that comes from wildcats= q
- Supports the existence of 2 groups of wild-living cats, with genetic intermediaries

Biological Questions

- Origin of Specific Individuals
- Dispersal
- Genetic Mixture Analysis
- Hybridization

Hybridization & management

- Captive Breeding programs
- Translocation
- Identifying/excluding wild populations for protection

Anderson and Thompson 2002- methods for identifying individuals to the F2 generation

Randi and Lucchini 2002

Admixture analysis of Wolves Dogs, Hybrids, and Captive-reared Wolves

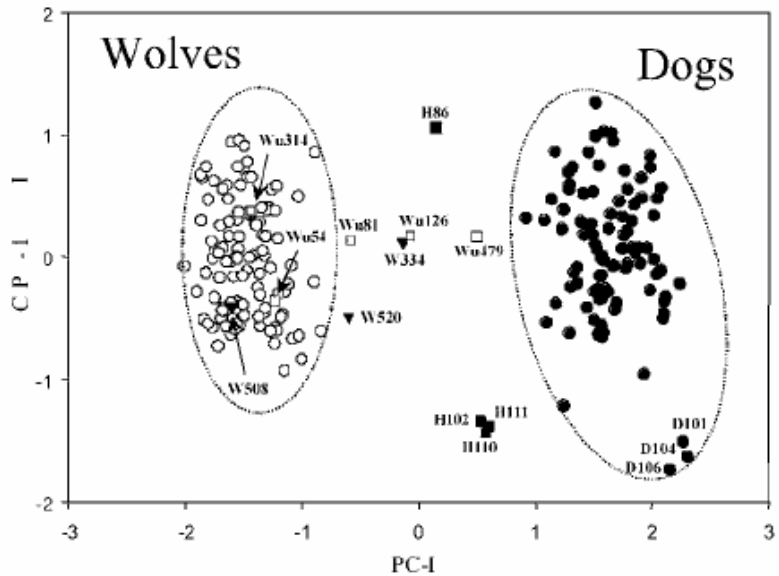

Figure 3. Scores of individual wolf and dog microsatellite genotypes plotted on the first two axes (PC-I, PC-II) of a principal coordinate analysis performed using Pcagen. H = known captive-reared hybrid wolves; Wu = captive-reared wolves of unknown origin; W334 and W508 = "black wolves"; W520 = "fifth finger wolf".

Further Questions…

- What is a population???
- Clustering performance with low Fst values?
- Realisitic migration detection
- Effects of unsampled sources
- Detection of hybrids/backcrosses beyond F1
- Effects of selection, linked loci?