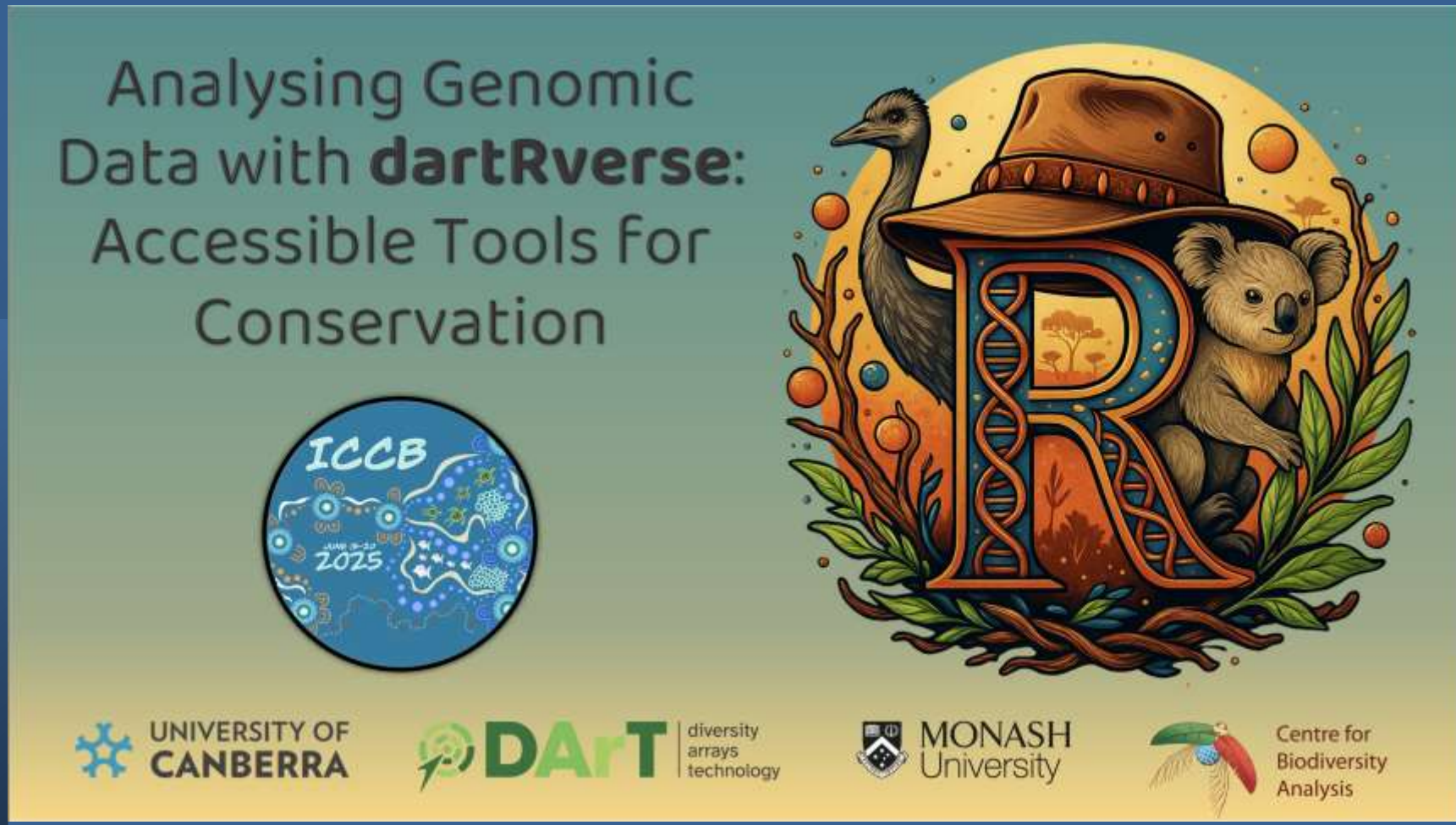
Genetic Diversity in Conservation Key Indicators and Their Applications



Why Genetic Diversity Matters

Genetic diversity is essential for the adaptability and long-term survival of species.

- Enhances resilience to environmental changes

- Reduces risk of inbreeding depression

- Maintains evolutionary potential

- Critical for effective conservation strategies

Genetic composition

Genetic Composition

Species

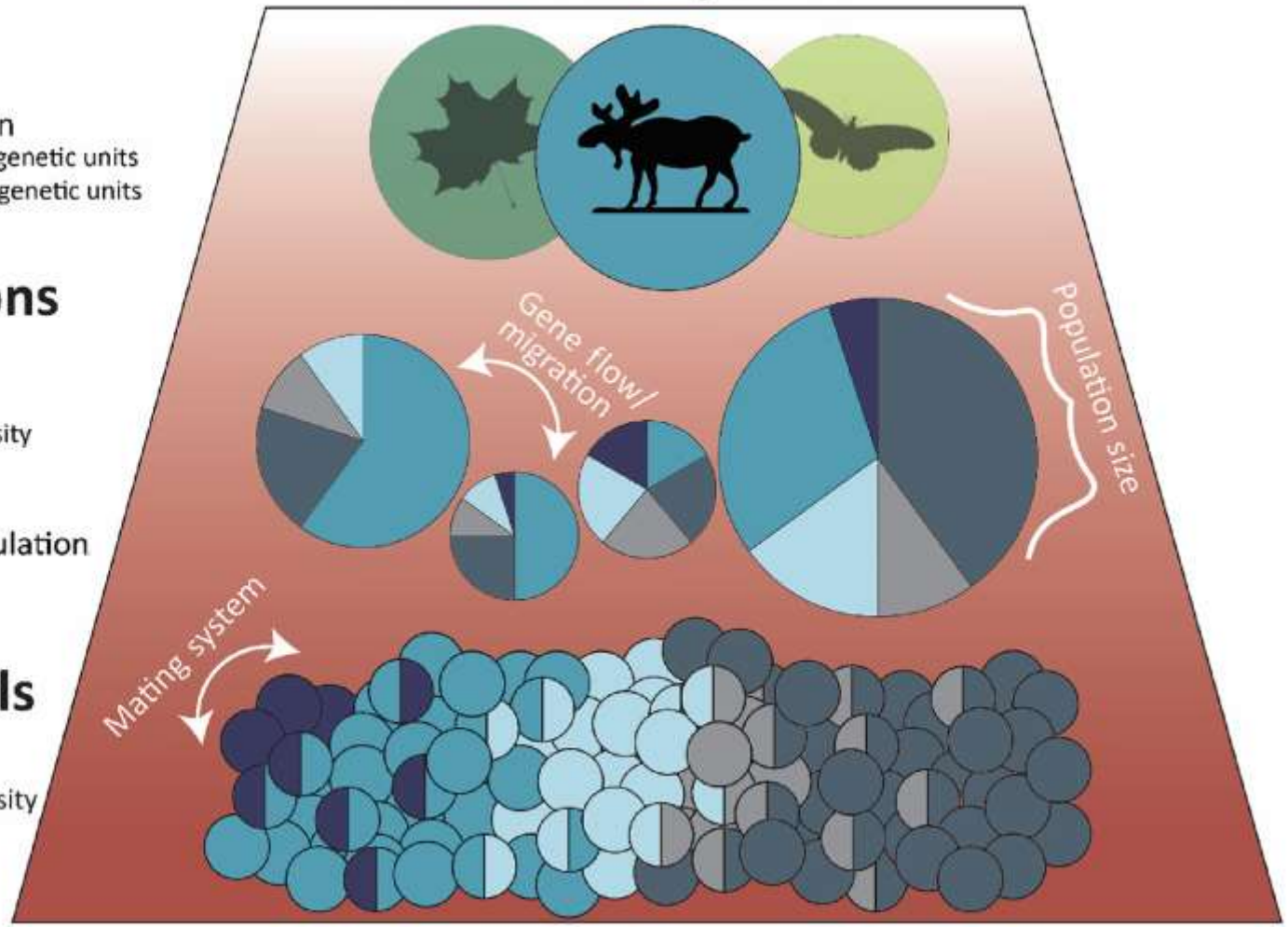
- Differentiation
  - Number of genetic units
  - Distance of genetic units

Populations

- Diversity
  - Richness
  - Heterozygosity
- Inbreeding
- Effective population size

Individuals

- Diversity
  - Heterozygosity
- Inbreeding



BIOLOGICAL REVIEWS
Cambridge Philosophical Society

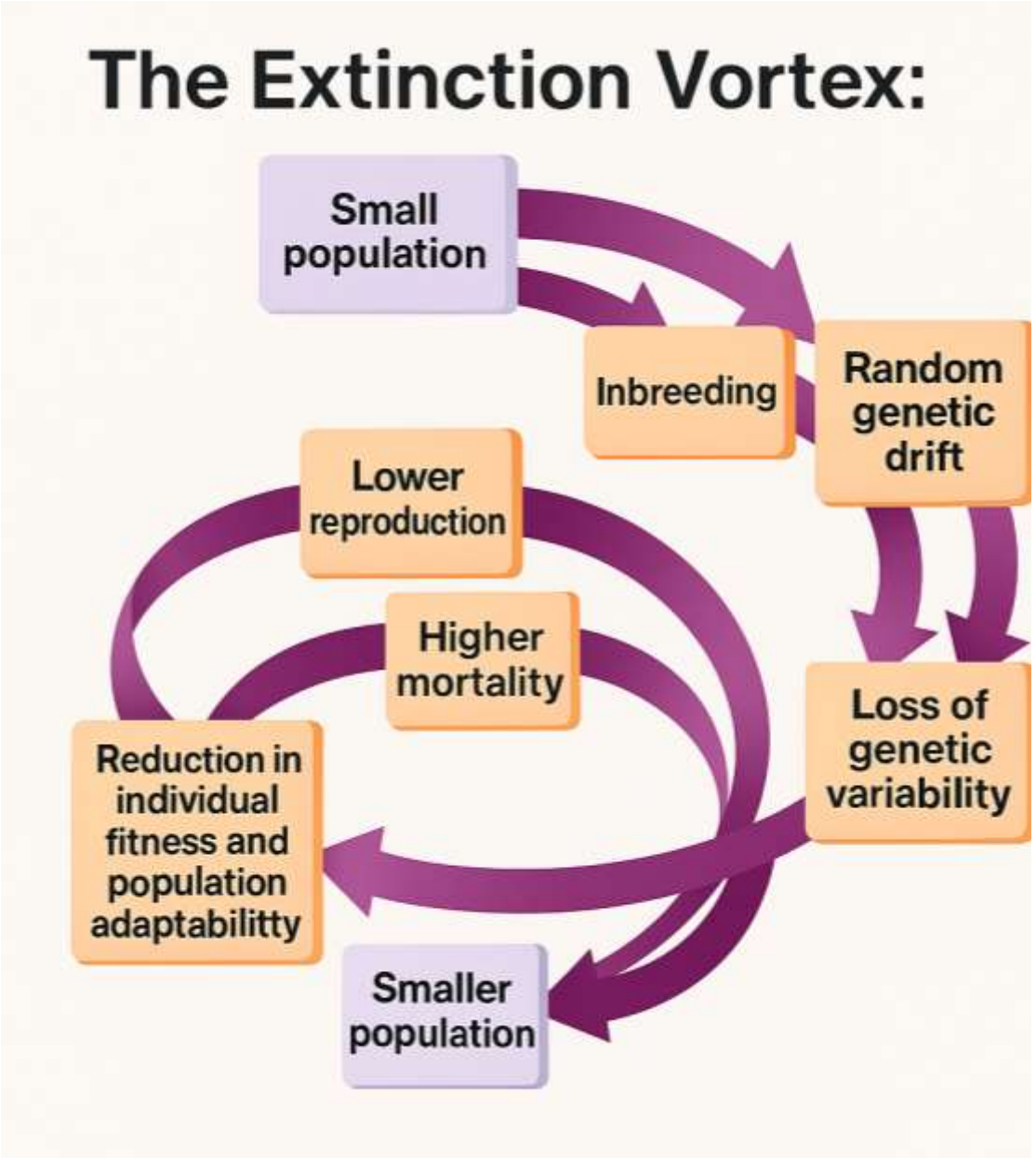
Original Article | Open Access

Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition

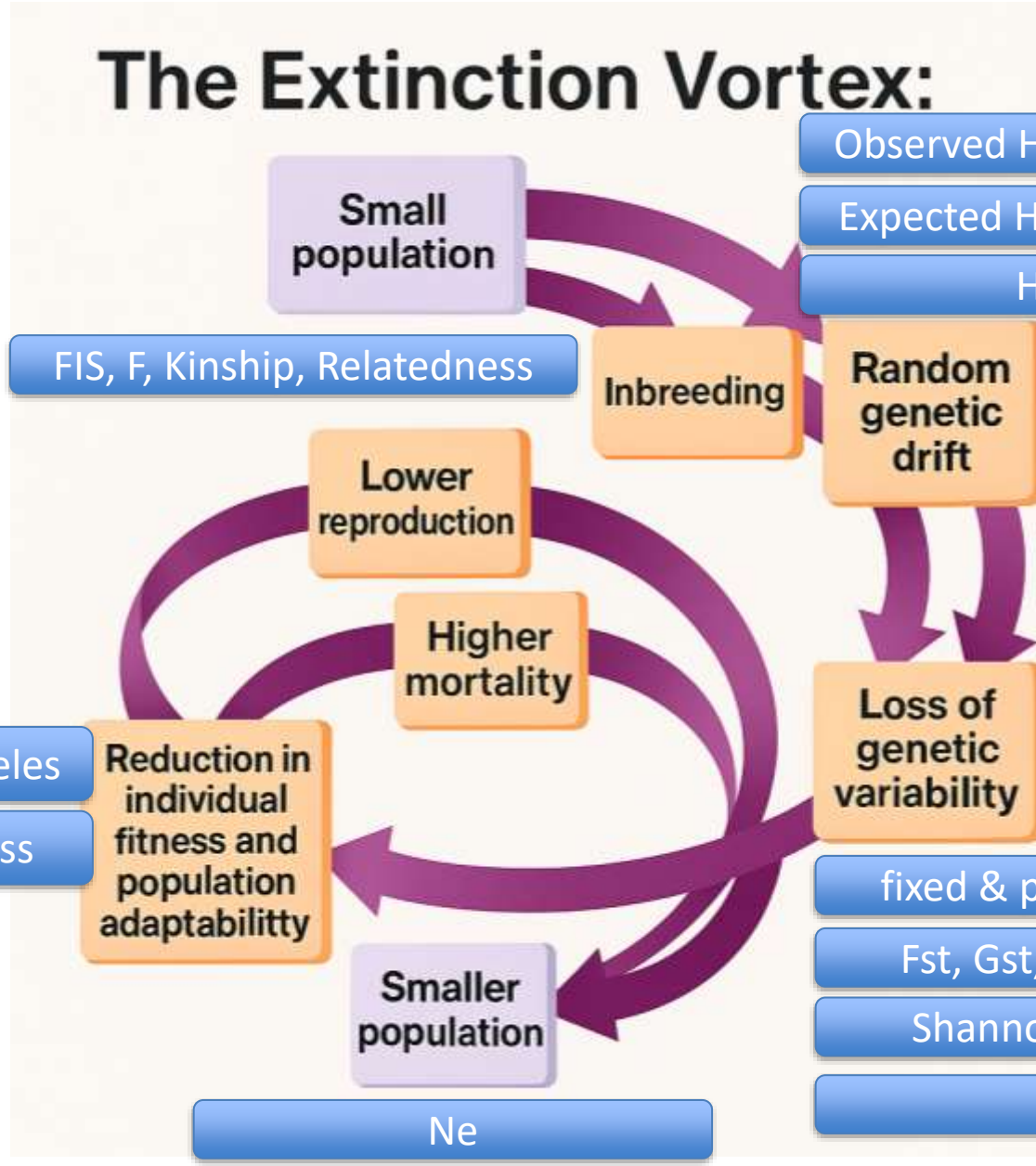
Sean Hoban, Frederick I. Archer, Laura D. Bertola, Jason G. Bragg, Martin F. Breed, Michael W. Bruford, Melinda A. Coleman, Robert Ekblom, W. Chris Funk, Catherine E. Grueber ... See all authors

First published: 12 April 2022 | https://doi.org/10.1111/brv.12852 | Citations: 64

Extinction Vortex



Extinction Vortex



Number of alleles, Allelic richness (A)

- Number of alleles per locus (often corrected for sample size)

- Indicates the potential for adaptive variation

```
gl.report.allelrich()    # allelic richness, number of alleles
gl.report.diversity( , table="D")

gl.report.nall()         # bootstrapped number of allels
                           per sample size in a population
```

Number of alleles

RESEARCH ARTICLE

Journal of Applied Ecology

Assessing the benefits and risks of translocations in depauperate species: A theoretical framework with an empirical validation

Elise M. Furlan[1] | Bernd Gruber[1] | Catherine R. M. Attard[2] | Robert N. E. Wager[3,4] | Adam Kerezsy[3,5] | Leanne K. Faulks[6] | Luciano B. Beheregaray[2] | Peter J. Unmack[1]

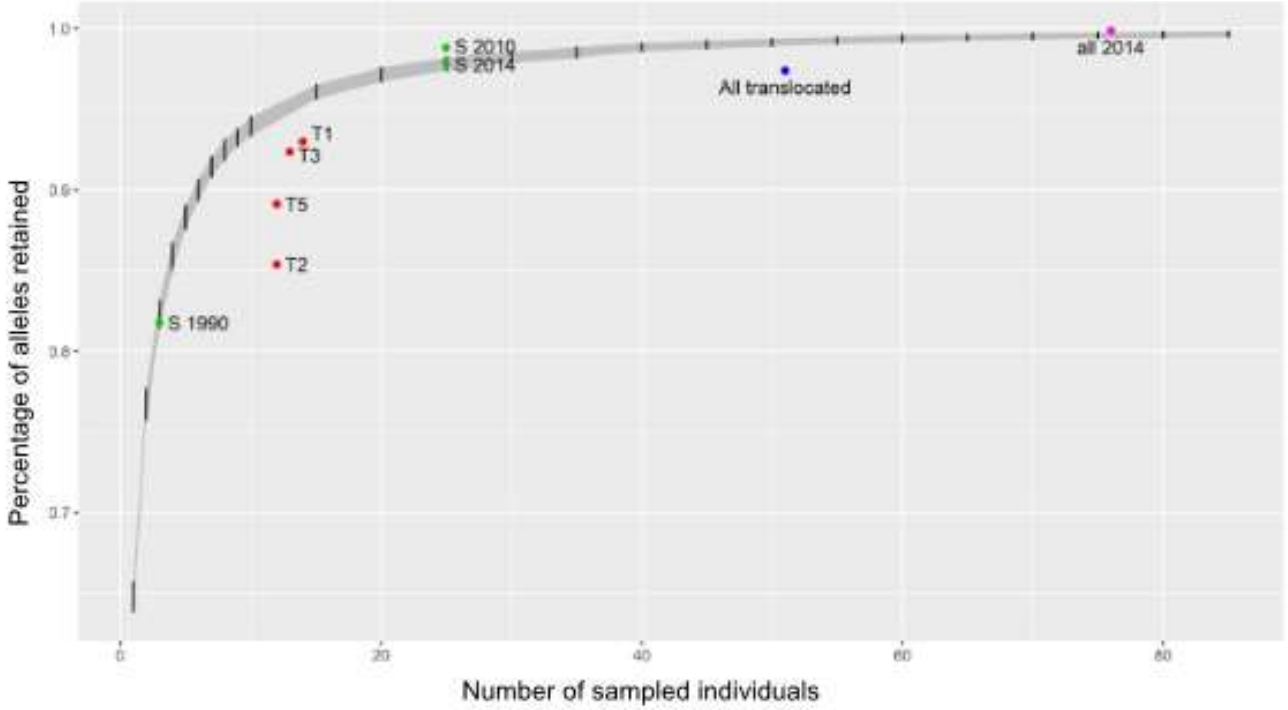


FIGURE 4 The percentage of redfin blue eye *Scaturiginichthys vermeilipinnis* alleles retained for a given number of individuals sampled from a population. The vertical black lines and grey confidence envelope show the expected percentage of alleles retained (95% CI) based on simulations of allele frequencies from the source population (*S*). The green and red dots indicate the actual percentage of alleles retained within the source population (sampled in 1990, 2010 and 2014) and four translocated populations (T1–T3 and T5) respectively. The blue dot indicates the total percentage of alleles actually retained in all translocated populations combined while the pink dot indicates the total percentage of alleles retained in all populations persisting in 2014

Observed heterozygosity (Ho)

- Proportion of individuals heterozygous at a locus
- Reflects current levels of genetic variability
- Count numbers of 1s in the data set
- gl.report.heterozygosity(), gl.filter.heterozygosity
- gl.Ho

Observed heterozygosity (Ho)

- Using SNPs
- Schmidt et al.
- Sopniewski & Catullo

Expected heterozygosity (He)

- Probability that two alleles randomly chosen from the population are different

🔷 **Equation for Expected Heterozygosity at a Single Locus**

For a **biallelic locus** with allele frequencies p and $q = 1 - p$:

$$H_E = 2pq$$

- Key measure of within-population diversity
- gl.report.heterozygosity(), gl.filter.heterozygosity
- gl.He

Unbiased He (uHe)

- He corrected for small sample size
- Important in small or fragmented populations
- gl.report.heterozygosity(), [gl.filter.heterozygosity()]

$$uH_e = H_e \times \left(\frac{2n}{2n-1} \right)$$

Hardy-Weinberg-Equilibrium

❑ What is HWE?

Hardy–Weinberg Equilibrium provides the expected genotype frequencies in a **randomly mating population** under **ideal conditions**. It serves as a **null model** for detecting evolutionary forces.

For a single locus with **two alleles**, A and a:

- **Allele frequencies:**

$$p = \text{frequency of } A, \quad q = \text{frequency of } a$$

  where $p + q = 1$

- **Expected genotype frequencies:**

$$\text{AA} : p^2, \quad \text{Aa} : 2pq, \quad \text{aa} : q^2$$

Hardy-Weinberg-Equilibrium

Assumptions of HWE:

1. **Random mating**
2. **No mutation**
3. **No gene flow (migration)**
4. **Infinitely large population (no drift)**
5. **No selection**
6. **Diploid, sexually reproducing organism**
7. **Non-overlapping generations**

Why HWE Matters in Conservation:

- **Baseline expectation** for genetic variation.

- **Deviations from HWE** can indicate:
  - **Inbreeding** → excess homozygosity
  - **Wahlund effect** → hidden structure across subpopulations
  - **Genotyping errors** (e.g., null alleles)
  - **Selection** or **migration** influencing alleles
- **Used to:**
  - Monitor genetic health
  - Detect structure or fragmentation
  - Validate data quality before downstream analyses
    (Genotyping errors (e.g., null alleles, allelic dropout) Sample contamination or duplicates, Population structure (Wahlund effect), Inbreeding or selection)

Inbreeding FIS

- Probability of identity by descent within individuals
- High values indicate inbreeding risk

- FIS: gl.report.heterozygosity(): FIS
- F: gl.grm, gl.run.emibd9): F= (2*diag(xx)-1)

Inbreeding FIS

FIS (inbreeding coefficient within subpopulations) measures the **reduction in individual heterozygosity** compared to what is expected under random mating within that subpopulation. It quantifies **deviation from Hardy-Weinberg equilibrium** due to inbreeding or assortative mating.

- **FIS > 0** → **deficit** of heterozygotes (inbreeding or population substructure)
- **FIS < 0** → **excess** of heterozygotes (outbreeding or selection)
- **FIS ≈ 0** → random mating
- In conservation, FIS is used to detect signs of inbreeding and assess genetic health within populations.

- FIS: gl.report.heterozygosity(): FIS
- [F: gl.grm, gl.run.emibd9): F= (2*diag(xx)-1)]

Fixed differences/ private alleles

	Private Alleles	Fixed Alleles
Definition	Alleles found only in one population	Alleles at 100% frequency in a population
Indicates	Unique genetic variation or local adaptation	Loss of genetic variation or strong selection
Used for	Assessing inbreeding, drift, or selection pressure	Identifying distinct populations or ESUs

gl.report.pa()
gl.fixed.diff()

Private alleles

Alleles unique to a single population

- Alleles that are **found only in one population** and **absent in others**.
  - Indicate unique genetic variation.
  - Help identify evolutionarily significant units (ESUs).
  - Suggest limited gene flow or long-term isolation.
  - Important for prioritizing populations for conservation or breeding.

Private alleles

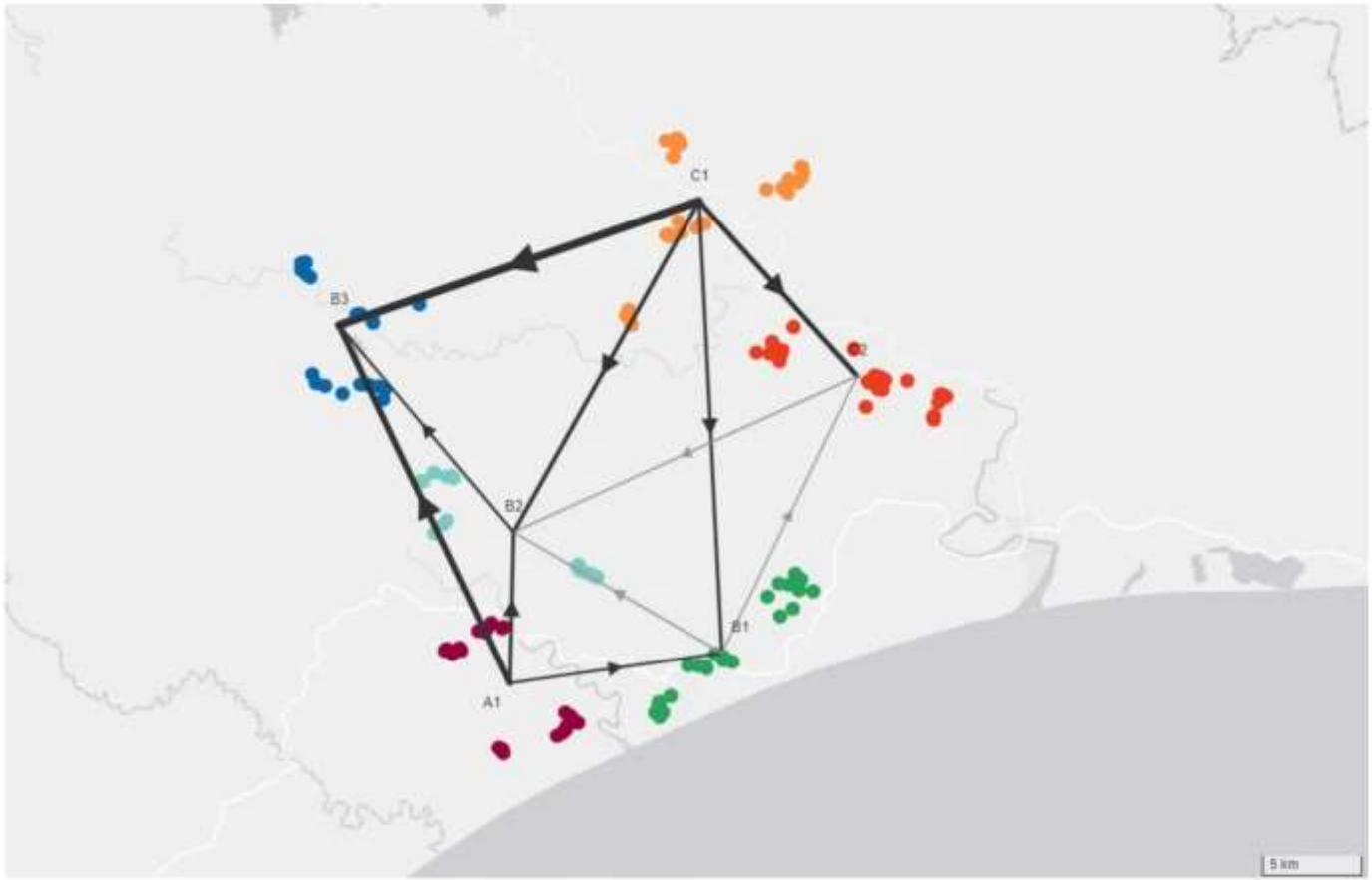
Biol Invasions (2021) 23:3831–3845
https://doi.org/10.1007/s10530-021-02609-1

ORIGINAL PAPER

Has the introduction of two subspecies generated dispersal barriers among invasive possums in New Zealand?

Catriona D. Campbell · Phil Cowan · Bernd Gruber · Anna J. MacDonald · Clare E. Holleley · Stephen D. Sarre

Fig. 7



Asymmetry in geneflow estimated using the number of private alleles. Black lines indicate significant asymmetries in geneflow based on 95% confidence intervals obtained through 1000 replicates of resamples of 20 individuals; the level of significance is indicated by the thickness of the lines; arrows indicate the predominant direction of gene flow. Grey arrows indicate non-significant asymmetries between pairs of sub-populations

Full size image ›

Relatedness vs Kinship

- gl.grm() or gl.run.emibd9()

	gl.grm (Relatedness Matrix)	EMIBD9 (Kinship Matrix)
Definition	Measures realized genomic relatedness between individuals	Measures expected IBD-based kinship between individuals
Typical Range	-0.5 to 1 (centered around 0 for unrelated pairs)	0 to 0.5 (0.25 for siblings, 0.125 for half-sibs, etc.)
Diagonal Values	Vary around 1 (self-relatedness), can exceed 1	≈ 0.5 for diploid individuals with no inbreeding
Interpretation	Based on allele sharing, reflects realized genomic similarity	Classical pedigree-based or IBD-based probability of sharing alleles
Use Case	Estimating genomic relationships	Inferring pairwise inbreeding or pedigree relationships
Statistical Basis	Built from SNP sharing and allele frequencies	Built from IBD segment counts or probabilistic IBD sharing estimates
Inbreeding Estimation	Inbreeding Estimation $F = GRM_{ii} - 1$	Inbreeding Estimation $F = 2 \times kinship_{ii} - 1$

Fst luis

- **Description:** Genetic differentiation among populations
- **Importance:** Detects structure; informs translocation decisions
- **dartR function or note:** gl.fst.pop()

Gst, G’st, Jost’s D luis

- **Description:** Alternatives or complements to Fst
- **Importance:** Address biases with highly variable markers
- **dartR function or note:** gl.fst.pop() or manual calculations

AMOVA (Φst) luis

- **Description:** Hierarchical analysis of variance in molecular data

- **Importance:** Quantifies structure at multiple levels

- **dartR function or note:** gl.amova()

Shannon or Simpson’s diversity luis

- **Description:** Entropy-based diversity measures
- **Importance:** Occasionally used to compare marker distributions
- **dartR function or note:** gl.report.diversity()

Effective population size (Ne)

Effective population size (Ne) is the size of an idealized population that would experience **genetic drift** or **inbreeding** at the same rate as the observed population.

- It is almost always **smaller than the actual census size (N)** due to factors like unequal sex ratios, variation in reproductive success, or population size fluctuations (Ne/Nc ~0.1-0.2).
- Target in conservation: **Ne ≥ 50 (short term)**, **Ne ≥ 500 (long term)** *(Franklin–Lynch rule)*

- gl.LDNe() [Neestimator from Do & Waples 2016]

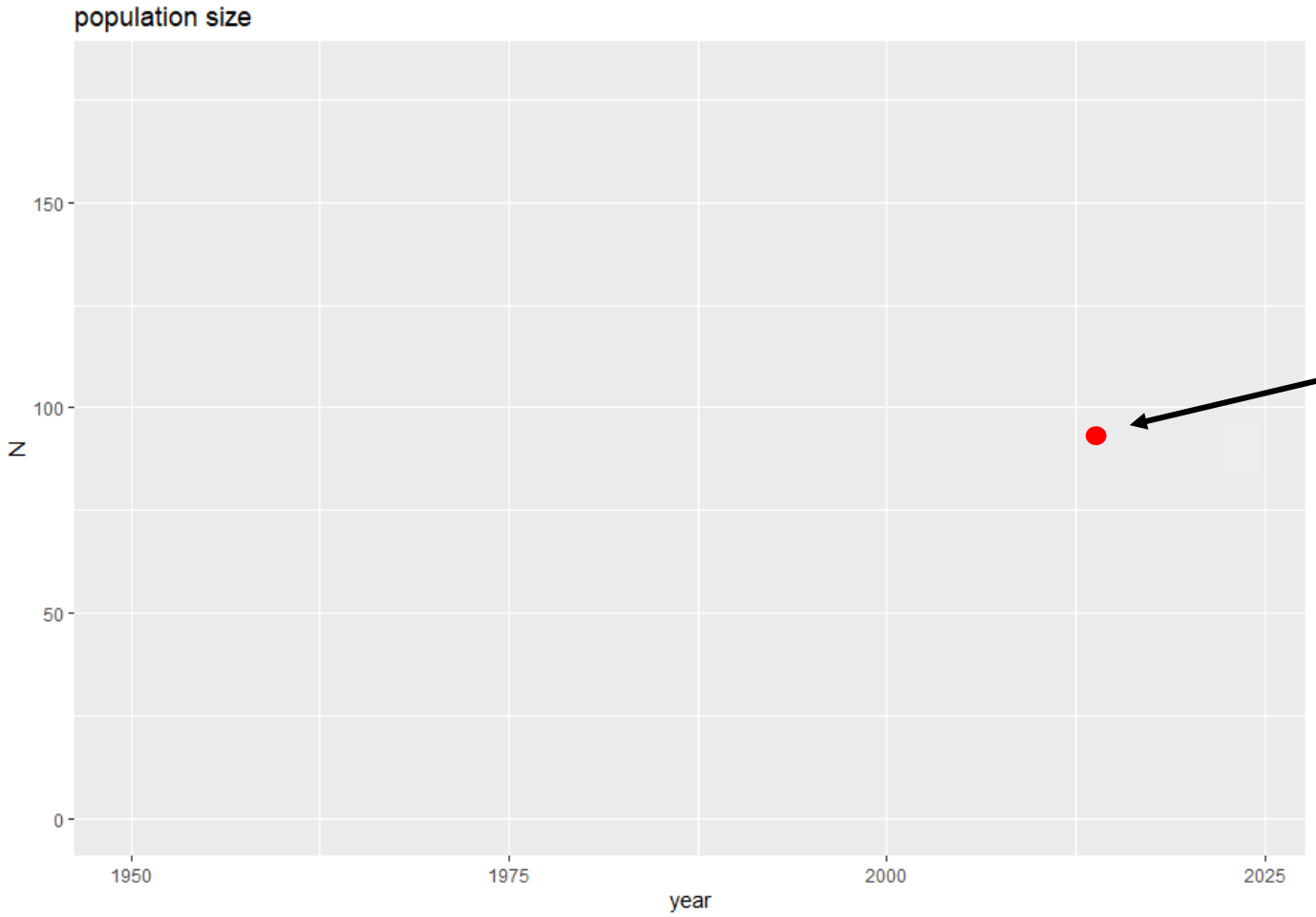
Why? Example Mahony's Toa

Recently discovered species (Clulow et al. 2016)

- Is it endangered?
- Was there a recent decline?
- If yes, what are the reasons for that decline?

Usual approach:

- Estimate current population size



A single population estimate in time.

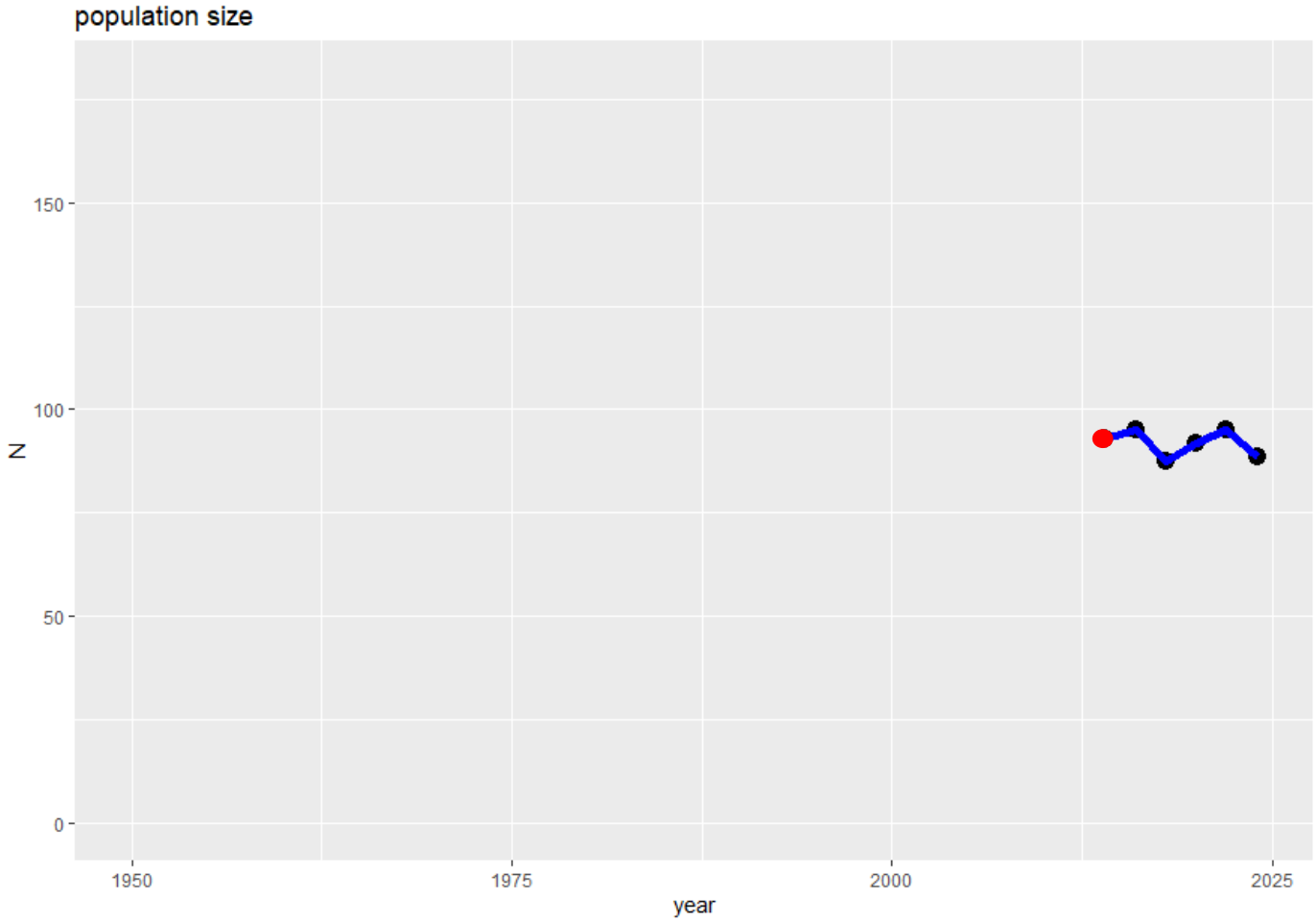
Example Mahony's Toadle

Recently discovered species (Clulow et al. 2016)

- Is it endangered?
- Was there a recent decline?
- If yes, what are the reasons for that decline?

Usual approach:

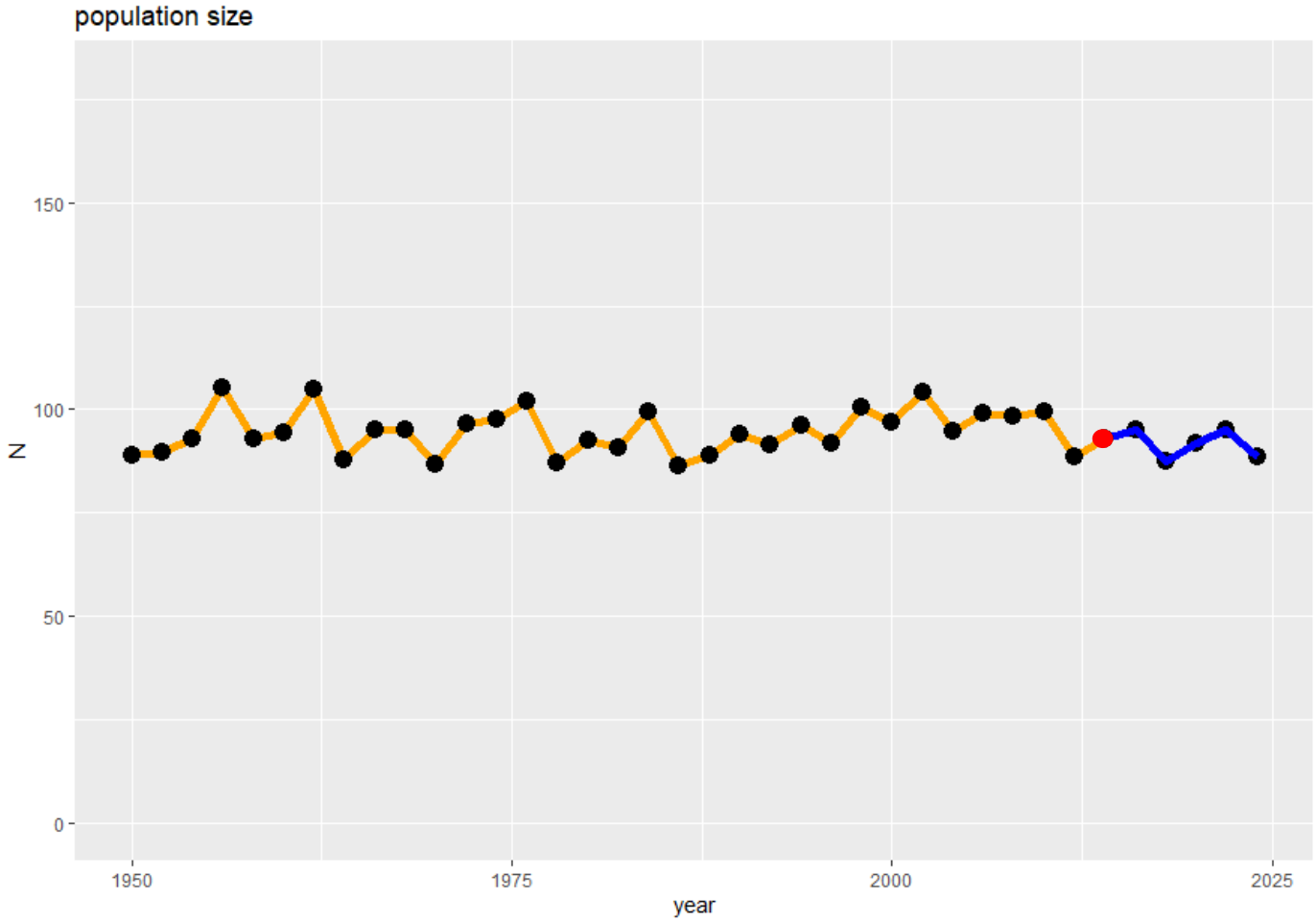
- Estimate current population size
- Monitor for some years



Example Mahony's Toadlet

Recently discovered species (Clulow et al. 2016)

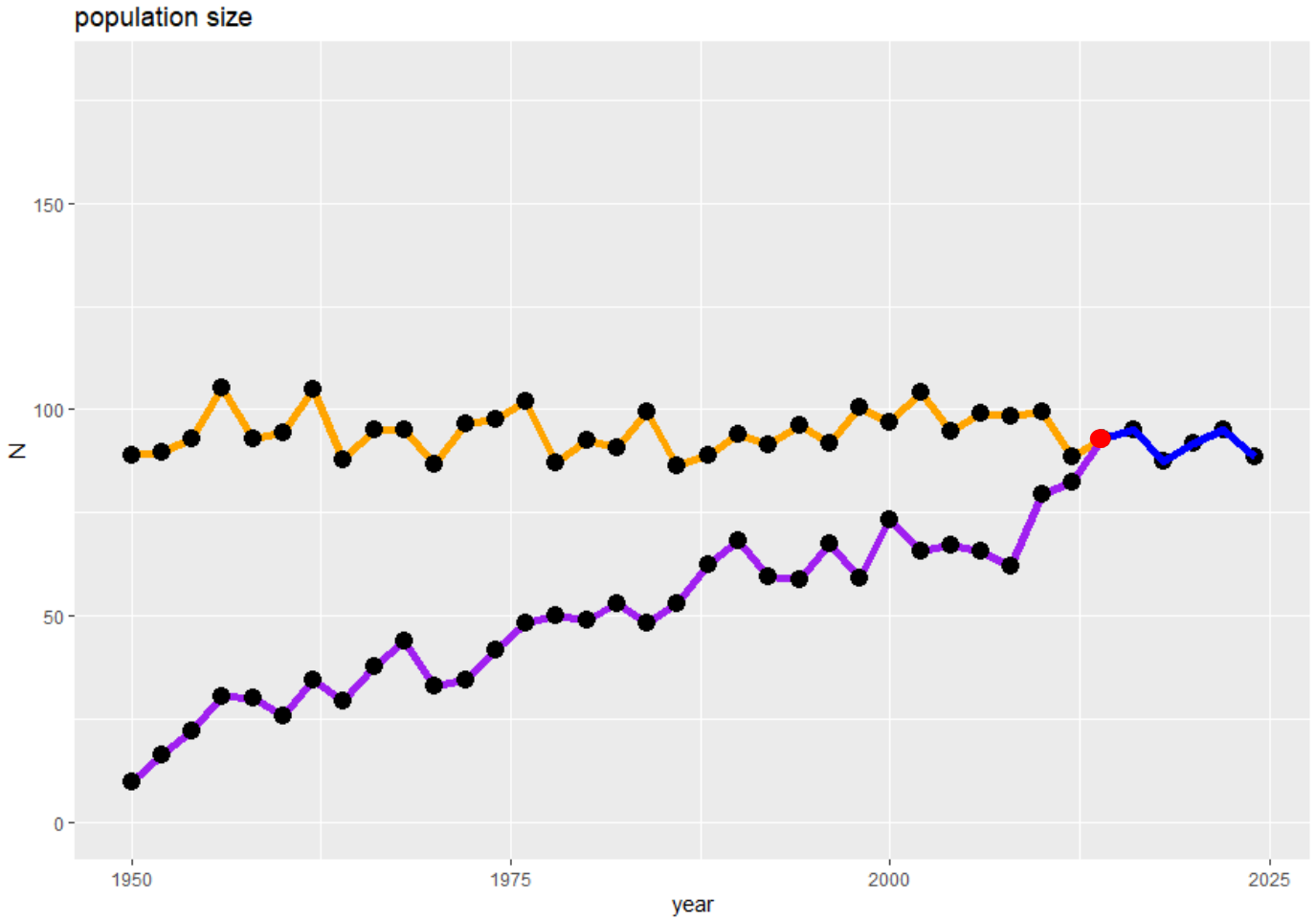
- Is it endangered?
- Was there a recent decline?
- If yes, what are the reasons for that decline?



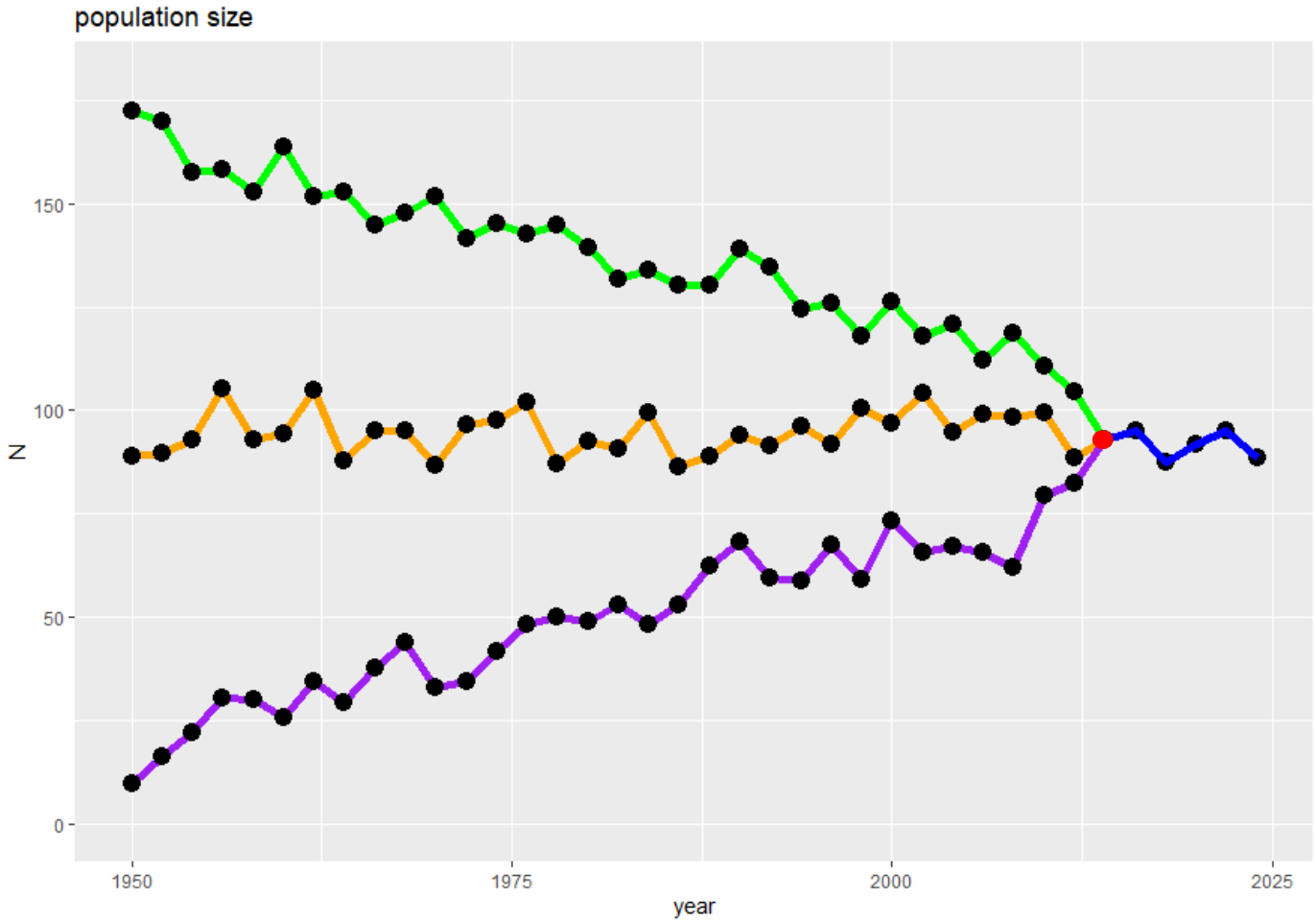
Usual approach:

- Estimate current population size
- Monitor for some years
- Historic population sizes

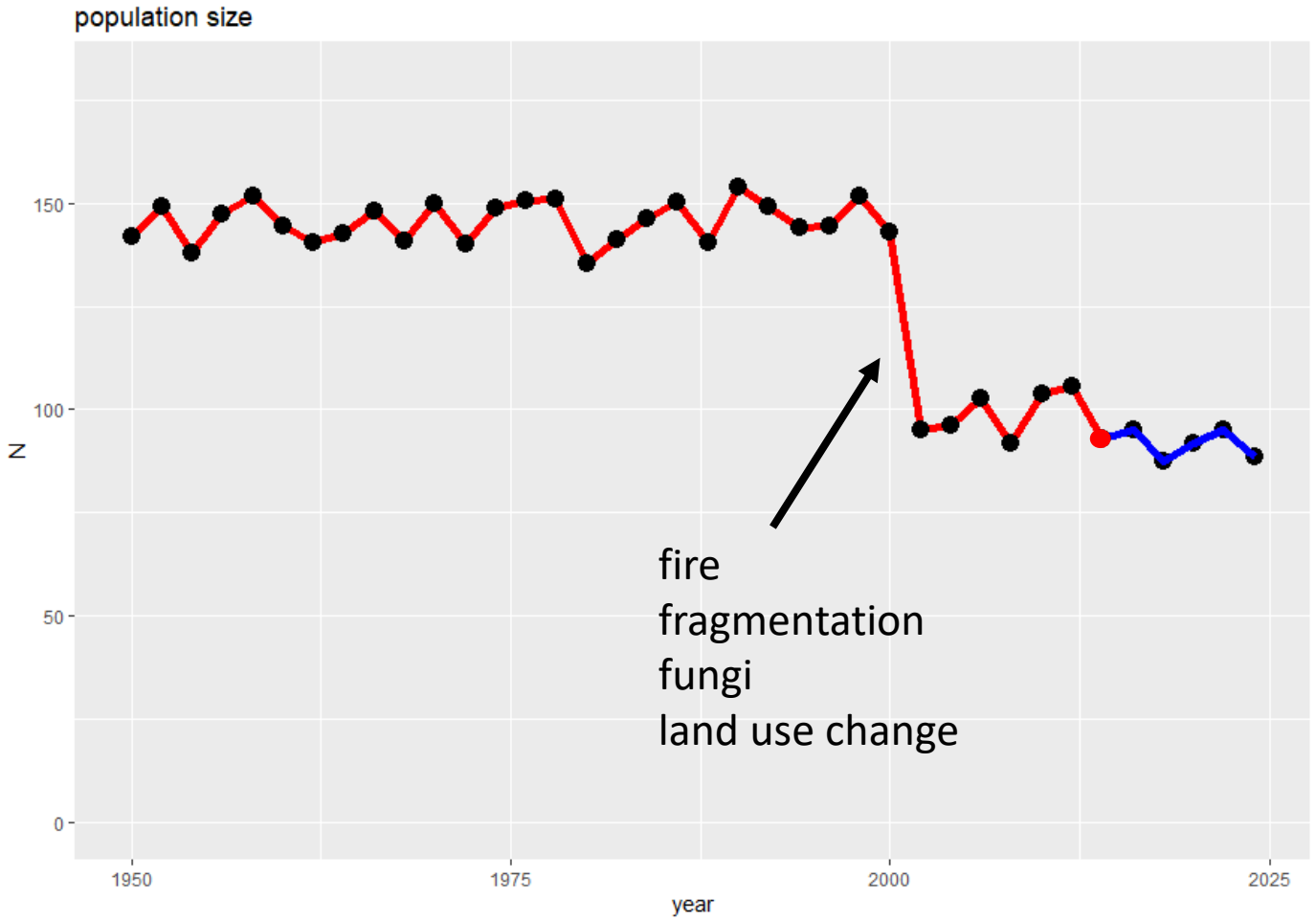
Example Mahony's Toadlet



Example Mahony's Toadlet



Example Mahony's Toadlet



A) Have a hypothesis on the trajectory	• The Canberra Grassland Earless Dragon “a recent decline due to urbanisation” • Historical data
B) Preprare your data	• Population structure • Missing data, How to filter
C) Create and check the SFS	
D) Run epos and stairways	• Guesstimate L and mu • Estimate current Ne
E) Validate your trajectory Does it make sense?	• Run several populations if you have them • Run simulations to check if your data were good enough
F) How to run	• Scripts on posit cloud

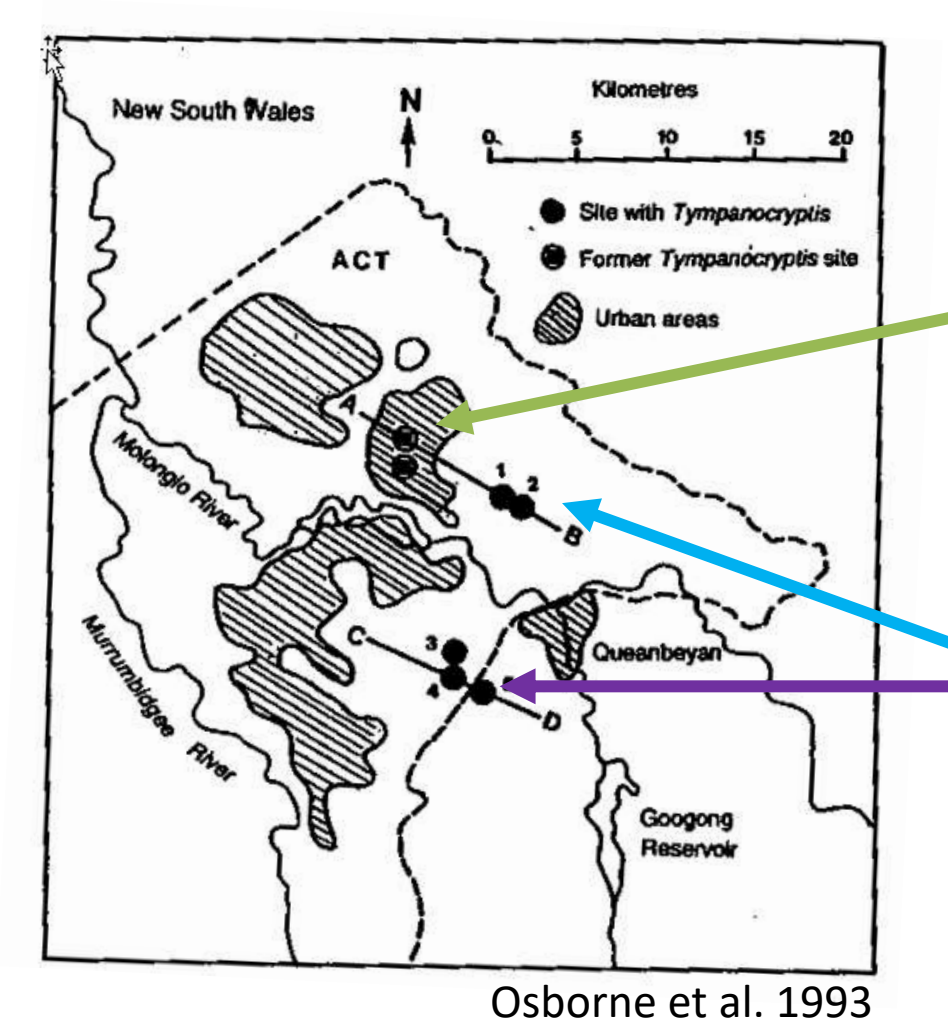
An example – The Canberra Grassland Earless Dragon

- A) Have your hypothesis on the trajectory
  - The Canberra Grassland Earless Dragon
  - “a recent decline due to urbanisation” [habitat loss, fragmentation]
  - Historical data
- B) Preprare your data
  - Population structure
  - Missing data
  - How to filter
- C) Create and check the SFS
- D) Run epos and stairways
  - Guesstimate L and mu
  - Estimate current Ne
- E) Validate your trajectory Does it make sense?
  - Run several populations if you have them
  - Run simulations to check if your data were good enough
- F) How to run

Canberra Grassland Earless Dragon
Tympanocryptis lineata

Ecology and conservation status

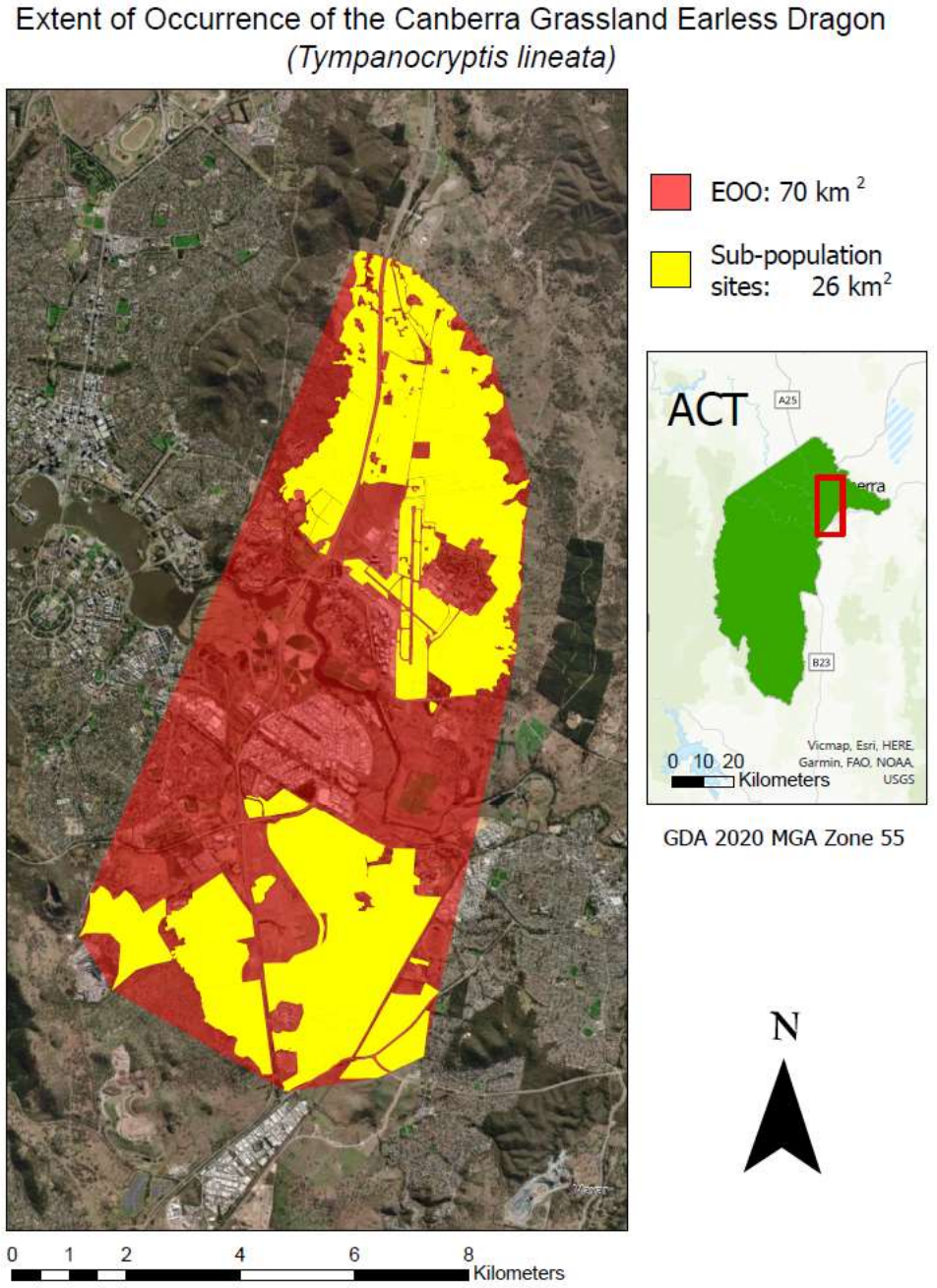
Canberra Grassland Earless Dragon



Osborne et al. 1993

- before 1970 commonly found in **urban areas** of Canberra
- rediscovered in 1991 by Will Osborne at sites **north** and **south** of the airport

Today

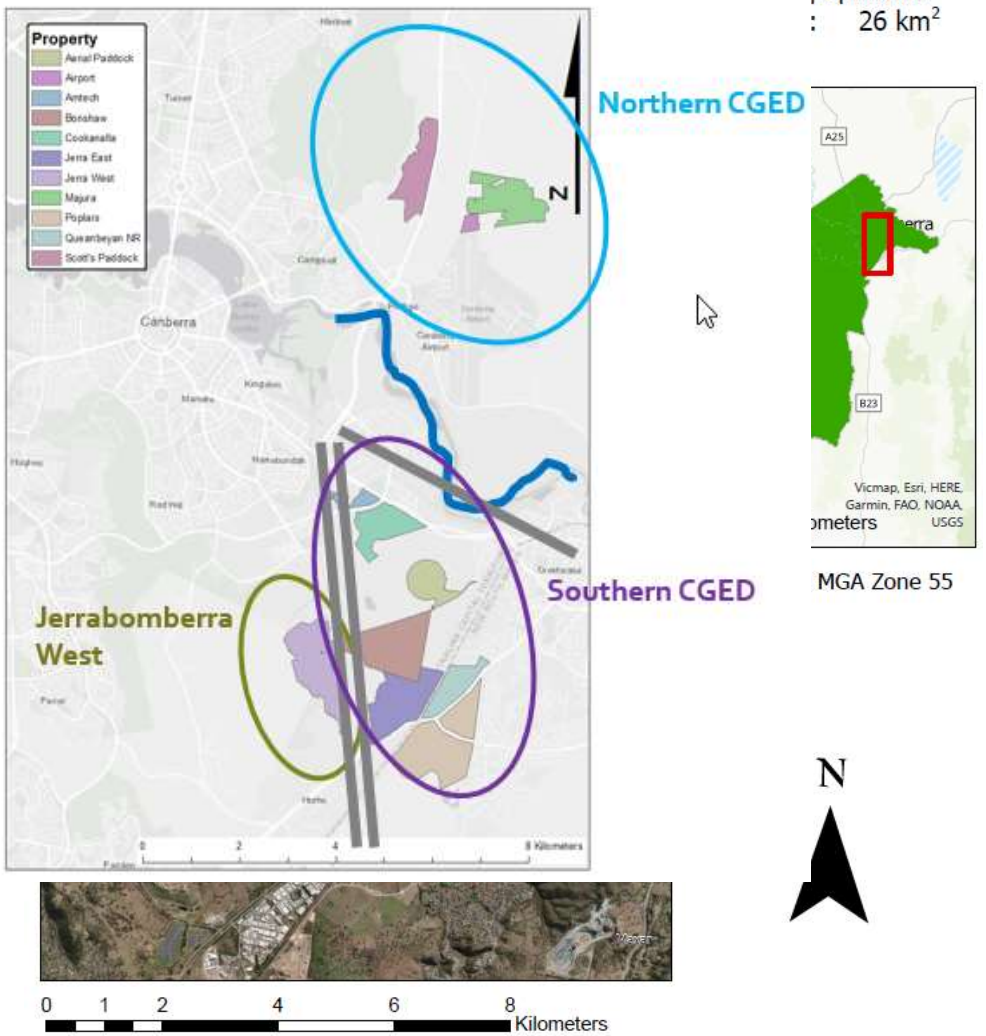

Extent of Occurrence of the Canberra Grassland Earless Dragon
(Tympanocryptis lineata)
EOO: 70 km 2
Sub-population sites: 26 km2
ACT
A25
B23
0 10 20
Kilometers
Vicmap, Esri, HERE, Garmin, FAO, NOAA, USGS
GDA 2020 MGA Zone 55
N
0 1 2 4 6 8
Kilometers

Today

Extent of Occurrence of the Canberra Grassland Earless Dragon
(Tympanocryptis lineata)

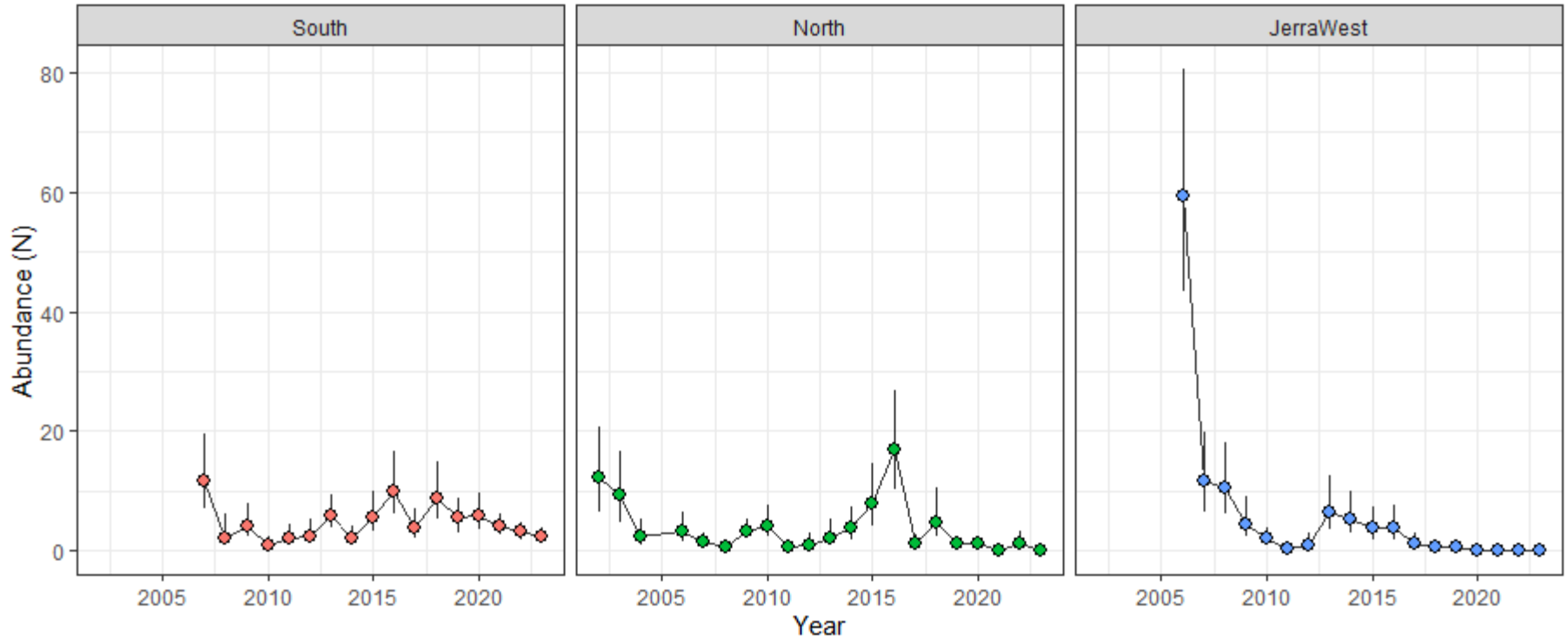
EOO: 70 km ^{2}

Sub-population
: 26 km^2

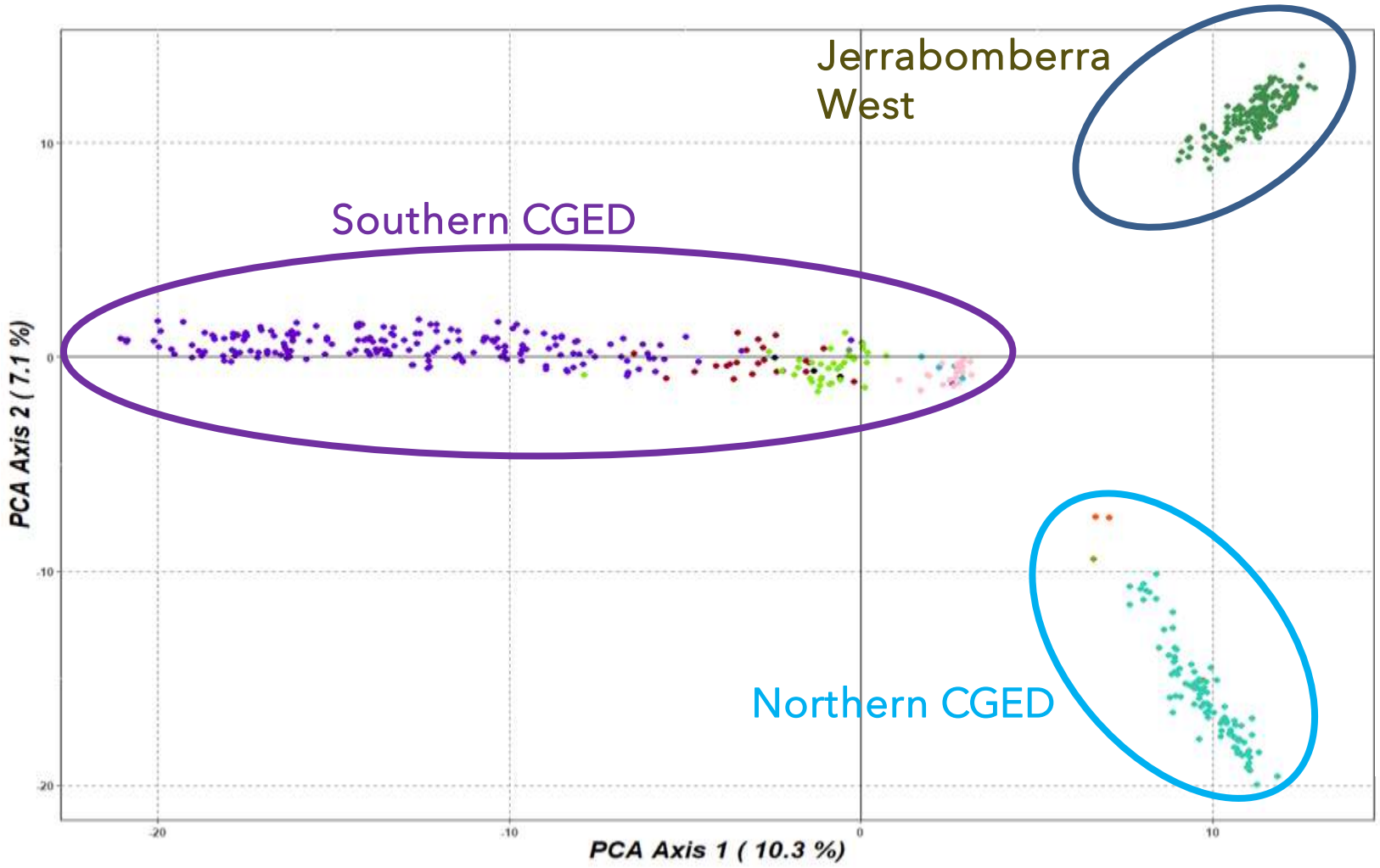


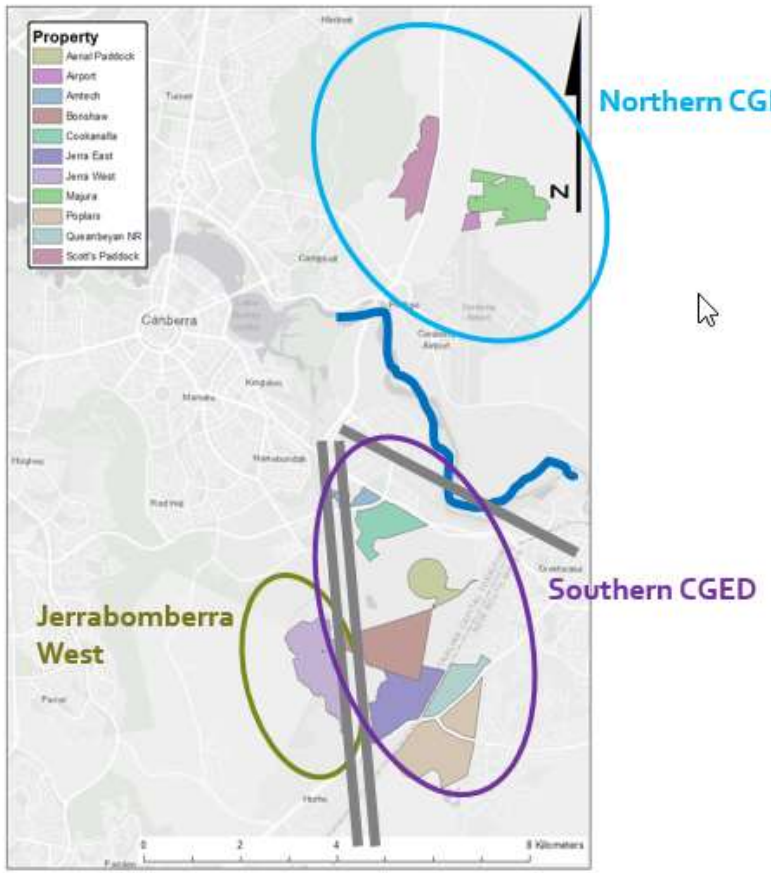
Canberra Grassland Earless Dragon

- Long term genetic data set (>15 years)
- Suspicion: decline in the recent years (collapsed ~ 2007 in JW)



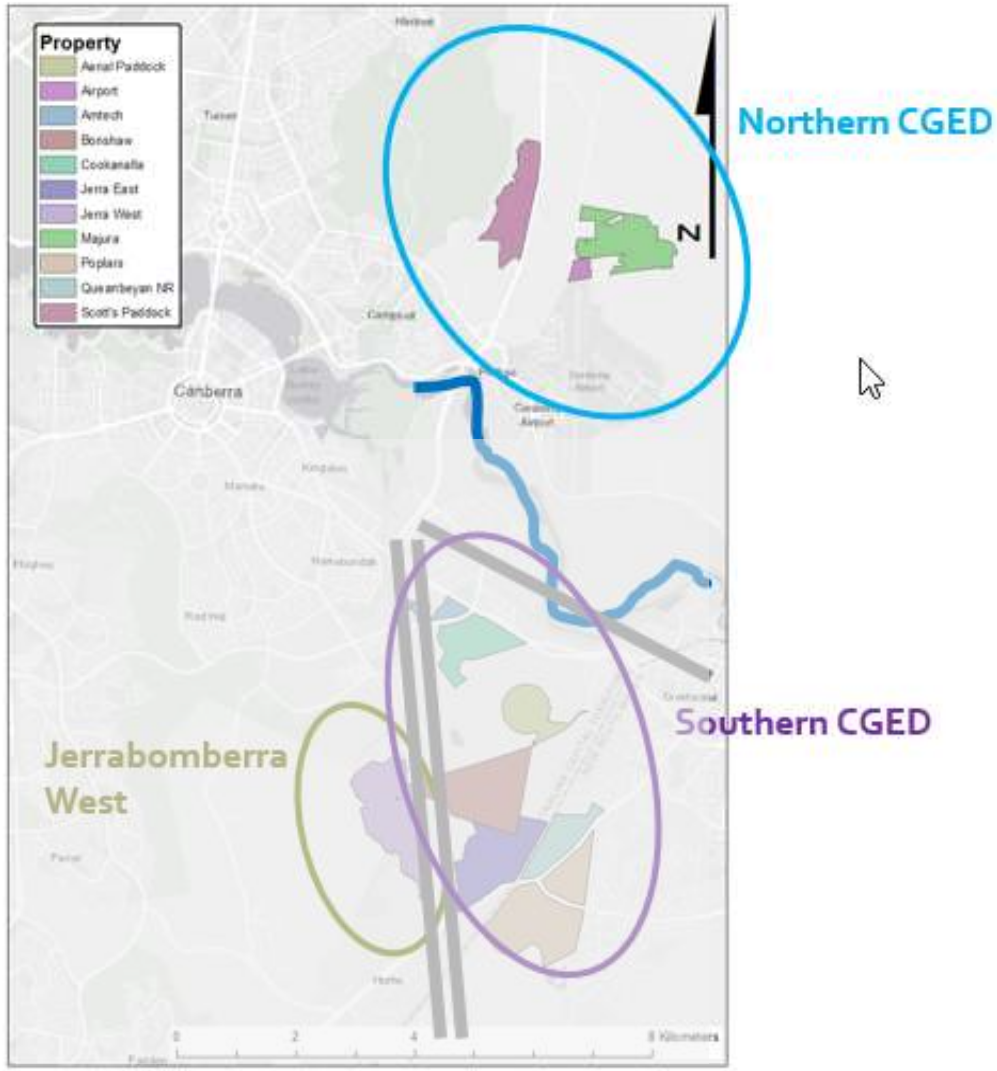
Canberra Grassland Earless Dragon fragmentation





Canberra North

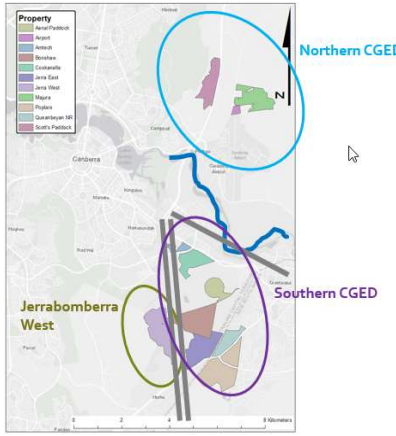
- Small for some time
- Was there are crash?



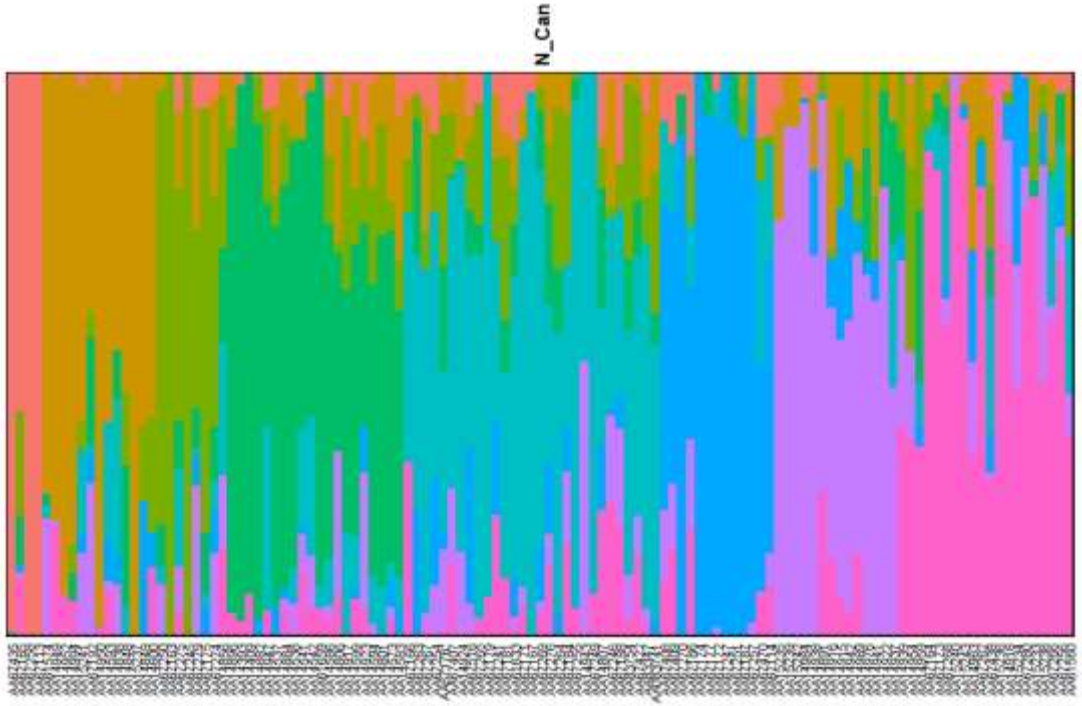
An example – The Canberra Grassland Earless Dragon

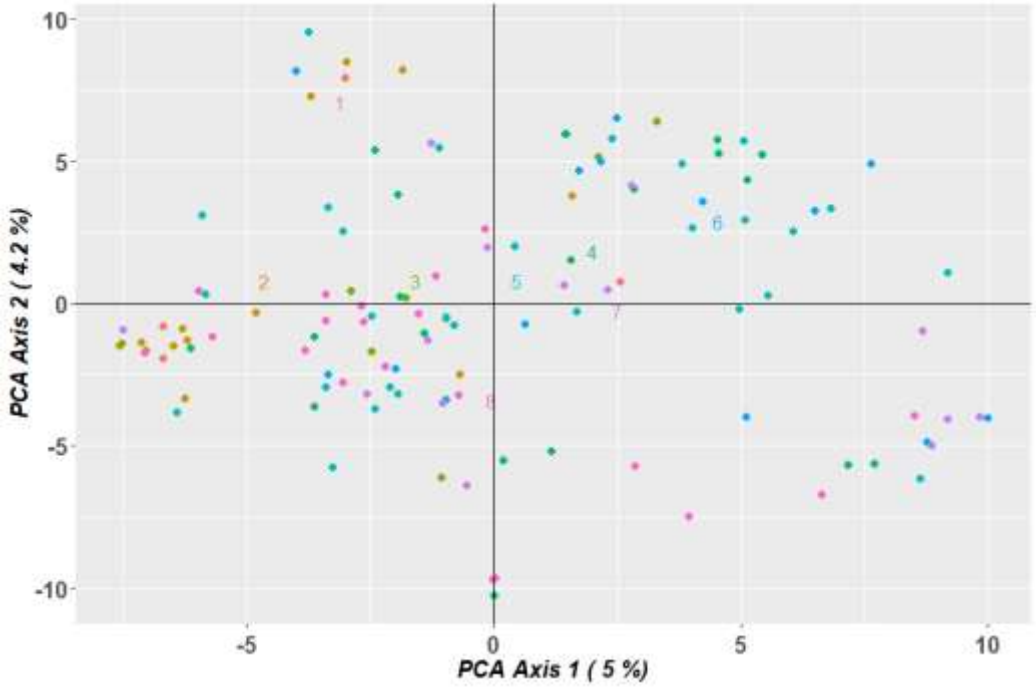
- A) Have your hypothesis on the trajectory
  - The Canberra Grassland Earless Dragon
  - “a recent decline due to urbanisation [habitat loss, fragmentation]
  - Historical data
- B) Preprare your data
  - Population structure
  - Missing data
  - How to filter
- C) Create and check the SFS
- D) Run epos and stairways
  - Guesstimate L and mu
  - Estimate current Ne
- E) Validate your trajectory Does it make sense?
  - Run several populations if you have them
  - Run simulations to check if your data were good enough
- F) How to run

Population structure



- Necessary to identify a single populations
- snmf to identify structure within the samples of North Canberra
- a individuals [over the last 10 y





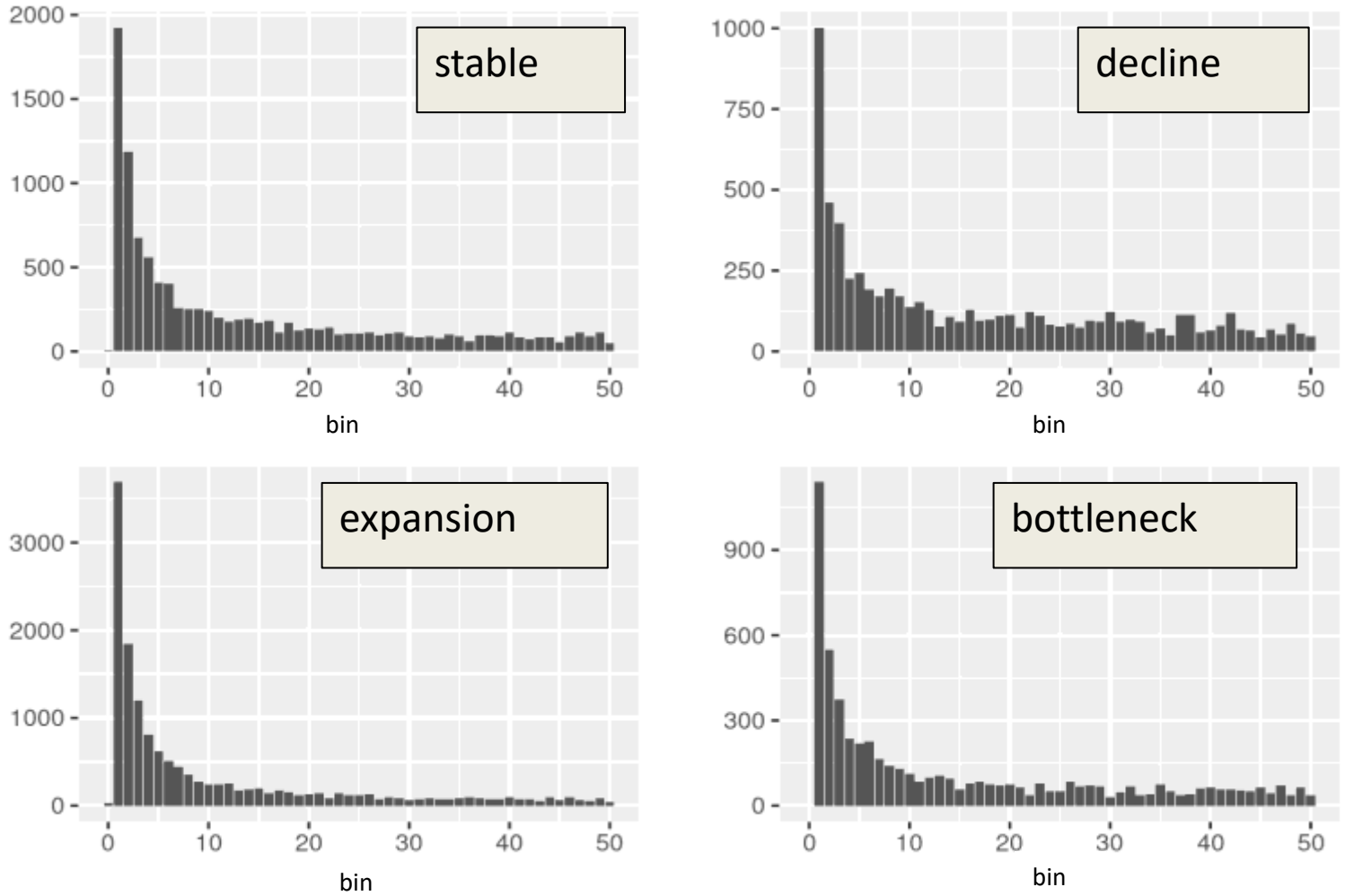
How to filter

- Aim: Site frequency spectrum based on high quality SNPs [low frequency bins are important]
  - Filter for missing data (call.rate)
  - [filter for reproducibility (dart)]
  - Filter for read depth
  - Filter callrate for individuals [you want to reduce missing data]
- Missing data should be less than 5%, better 1%
- [Imputation for missing data ? Needs further exploration]

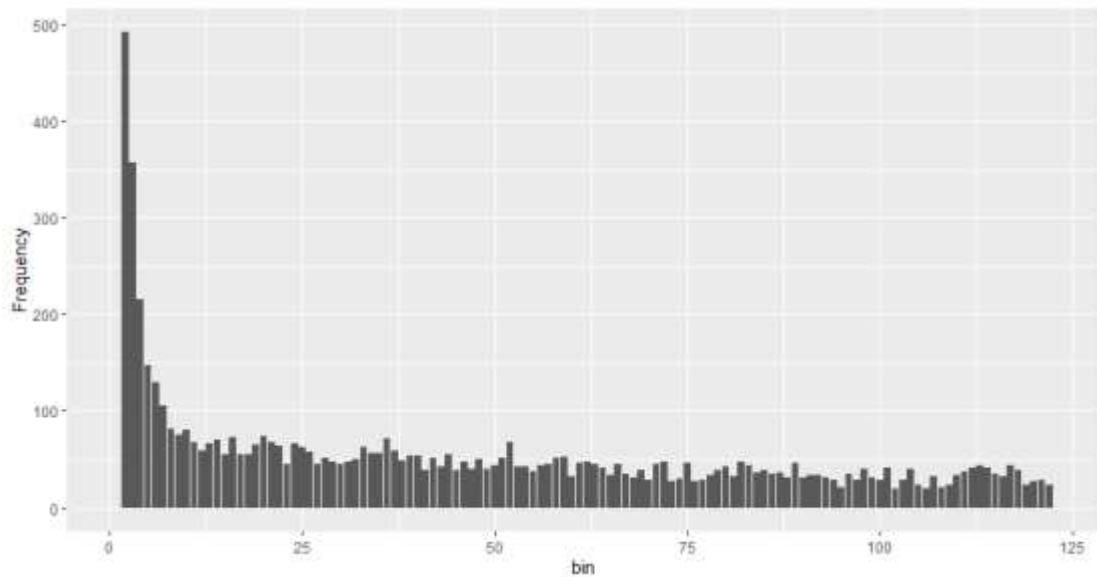
An example – The Canberra Grassland Earless Dragon

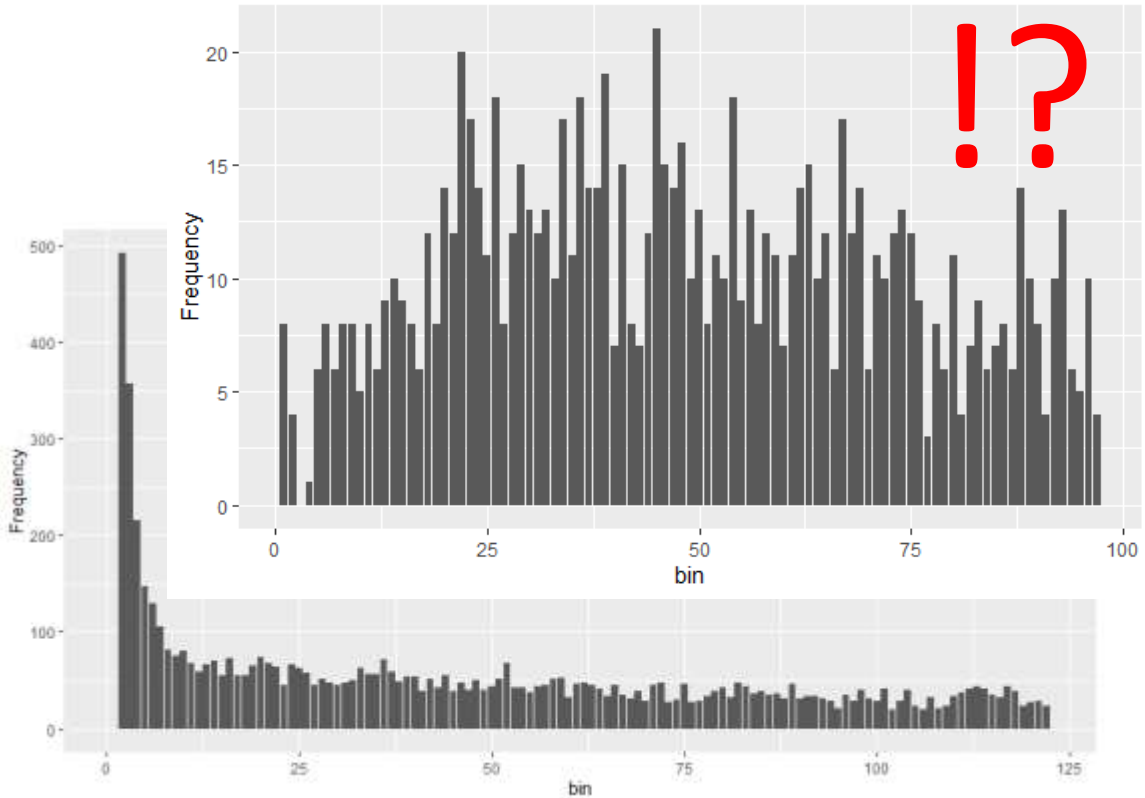
- A) Have your hypothesis on the trajectory
  - The Canberra Grassland Earless Dragon
  - “a recent decline due to urbanisation [habitat loss, fragmentation]
  - Historical data
- B) Preprare your data
  - Population structure
  - Missing data
  - How to filter
- C) Check the SFS
- D) Run epos and stairways
  - Guesstimate L and mu
  - Estimate current Ne
- E) Validate your trajectory Does it make sense?
  - Run several populations if you have them
  - Run simulations to check if your data were good enough
- F) How to run

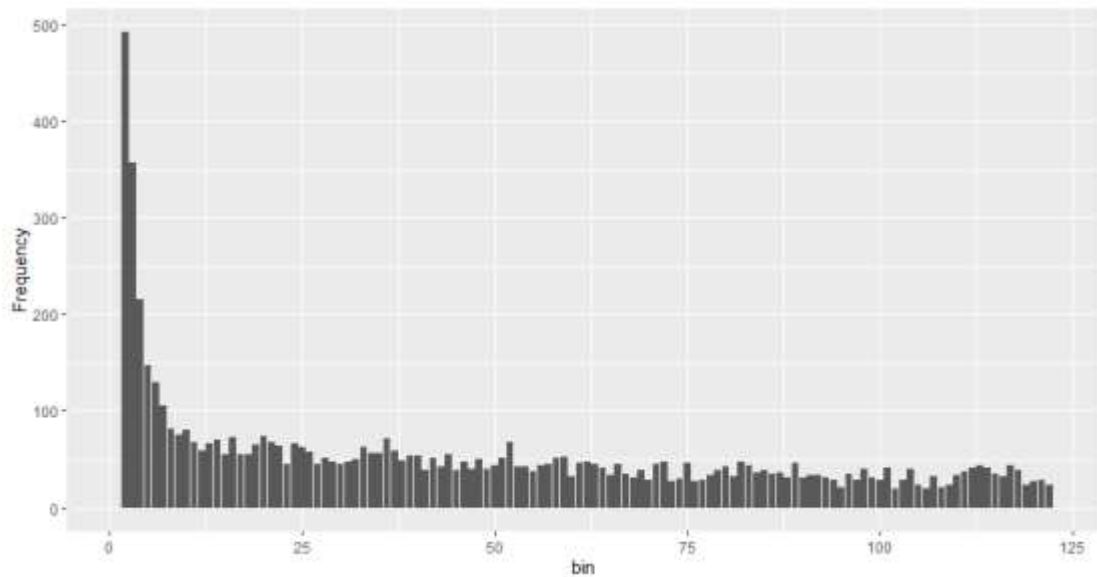
Coalescent –> SFS (Site Frequency Spectrum)



Beichman et al. 2018


500
400
300
200
100
0
Frequency
0
25
50
75
100
125
bin


!?
Frequency
bin
0
5
10
15
20
0
25
50
75
100
Frequency
bin
0
100
200
300
400
500
0
25
50
75
100
125


500
400
300
200
100
0
Frequency
0
25
50
75
100
125
bin

An example – The Canberra Grassland Earless Dragon

- A) Have your hypothesis on the trajectory
  - The Canberra Grassland Earless Dragon
  - “a recent decline due to urbanisation [habitat loss, fragmentation]
  - Historical data
- B) Preprare your data
  - Population structure
  - Missing data
  - How to filter
- C) Check the SFS
- D) Run epos and stairways
  - Guesstimate L and mu
  - Estimate current Ne
- E) Validate your trajectory Does it make sense?
  - Run several populations if you have them
  - Run simulations to check if your data were good enough
- F) How to run

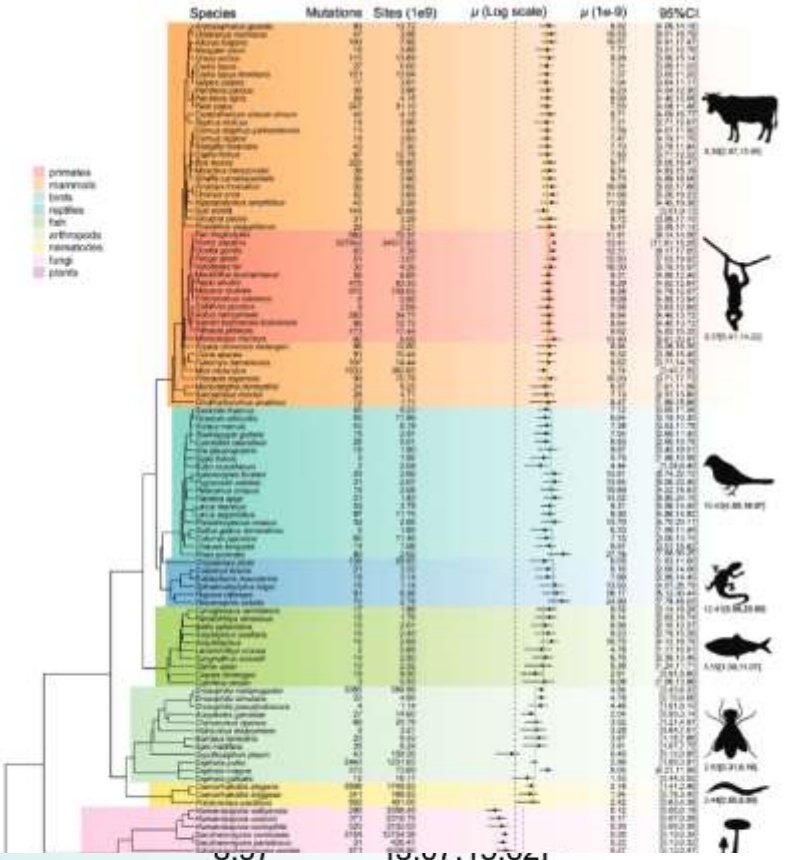
Guesstimate L an

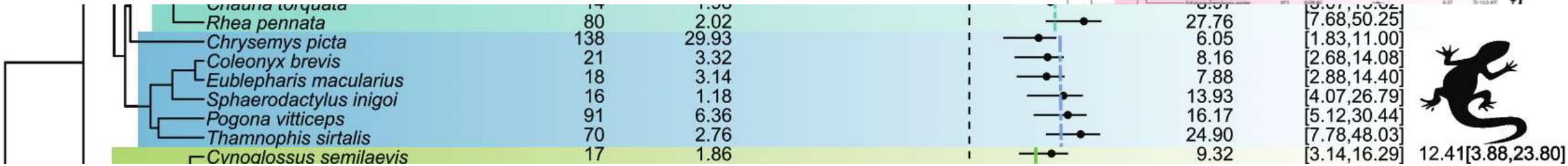
- Mu: mutation rate ~1e-8, [16.17e-9 for GE
- L: length of the sequenced genome

Evolution Letters, 2023, 7(4), 216–226
https://doi.org/10.1093/evlett/qrad027
Letter

Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis

Yiguan Wang and Darren J. Obbard





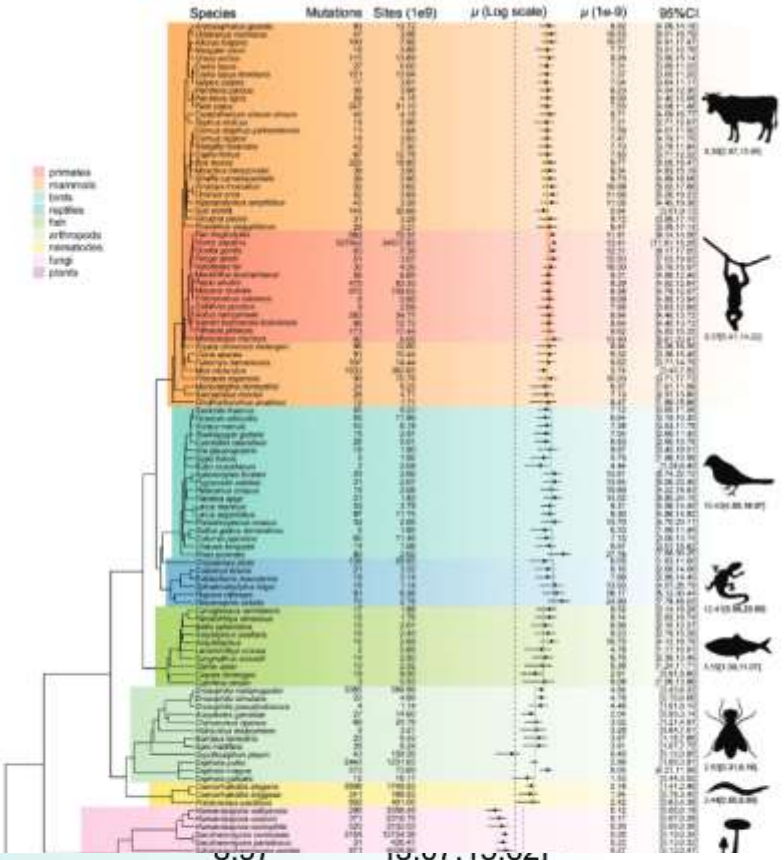
Guesstimate L an

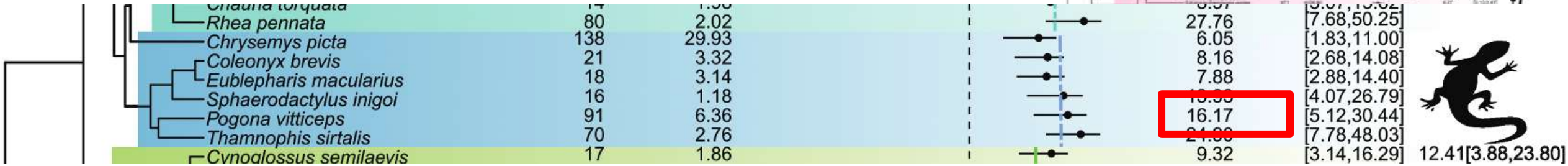
- Mu: mutation rate ~1e-8, [16.17e-9 for GE
- L: length of the sequenced genome

Evolution Letters, 2023, 7(4), 216–226
https://doi.org/10.1093/evlett/qrad027
Letter

Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis

Yiguan Wang and Darren J. Obbard



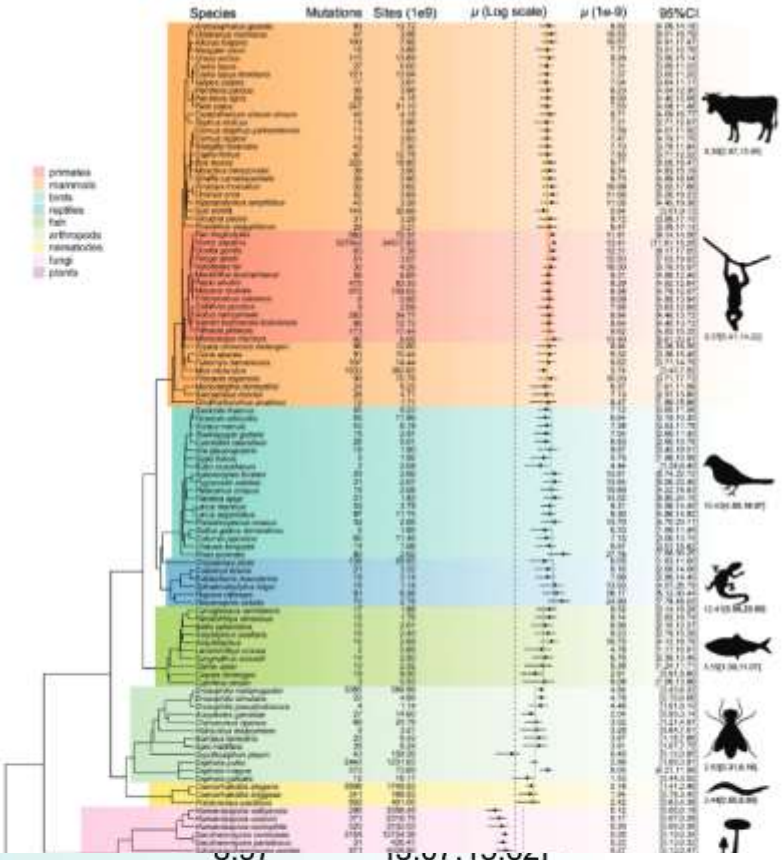


Guesstimate L an

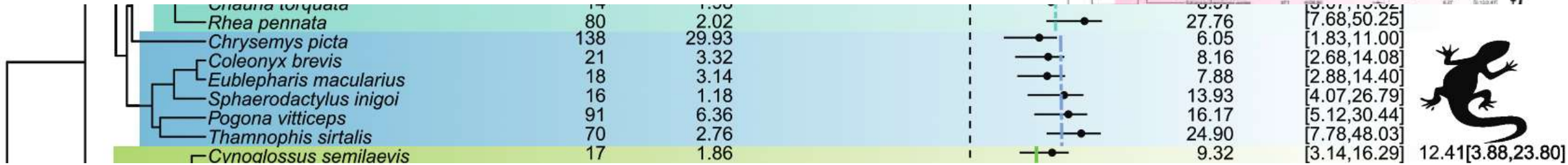
Evolution Letters, 2023, 7(4), 216–226
https://doi.org/10.1093/evlett/qrad027
Letter

Experimental estimates of germline mutation rate in eukaryotes: a phylogenetic meta-analysis

Yiguan Wang and Darren J. Obbard



- Mu: mutation rate ~1e-8, [16.71e-9 for GE
- L: length of the sequenced genome
  - Tricky because of filters and restriction enzymes
  - mu * L = number of mutations
  - Often used for dart L = 69 * nLoc [complete underestimate]

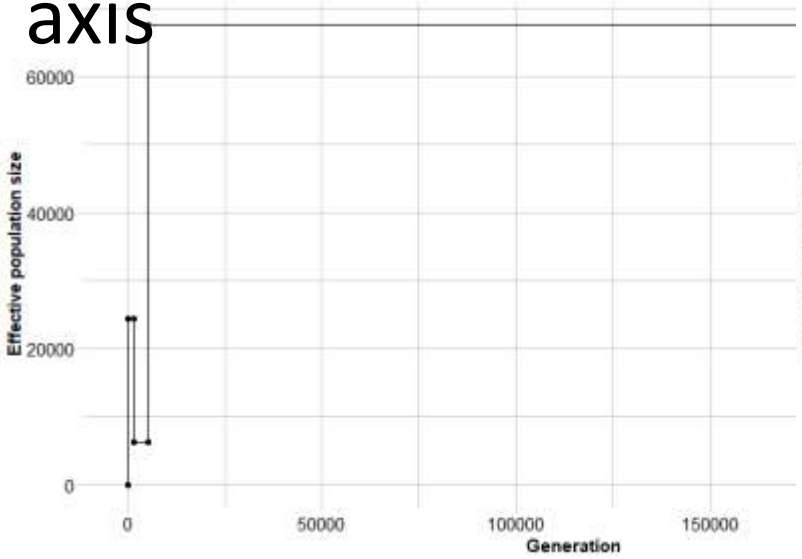


L * mu = number of mutations

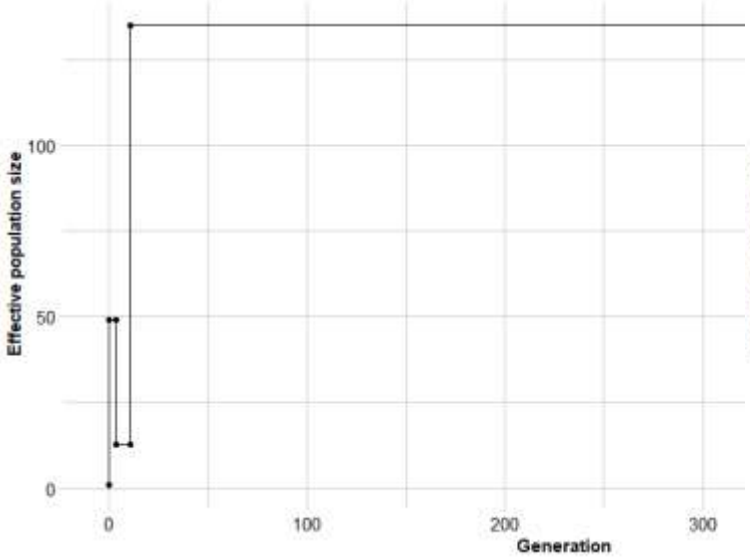
- The good news: Being wrong here does not influence the trajectory

- The bad news: being wrong here does influence that x and y axis

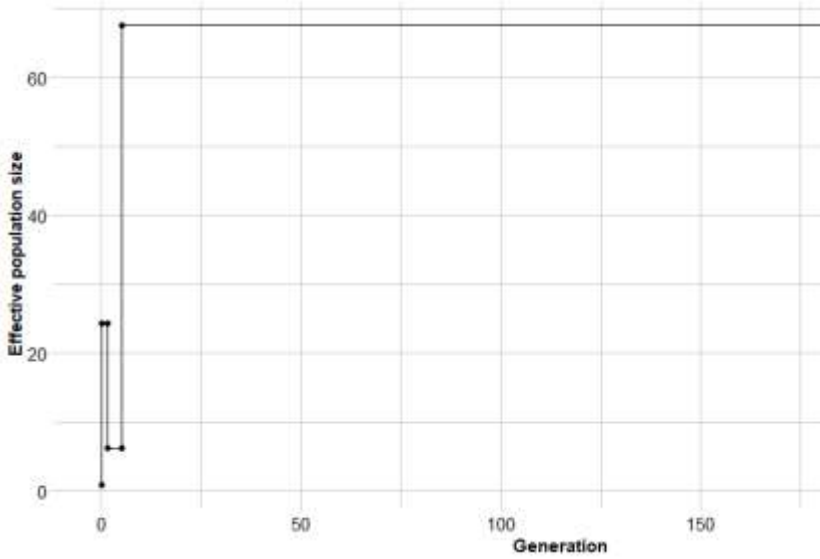
L=nLoc(gl)*75 * 1



L=nLoc(gl)*75 * 500



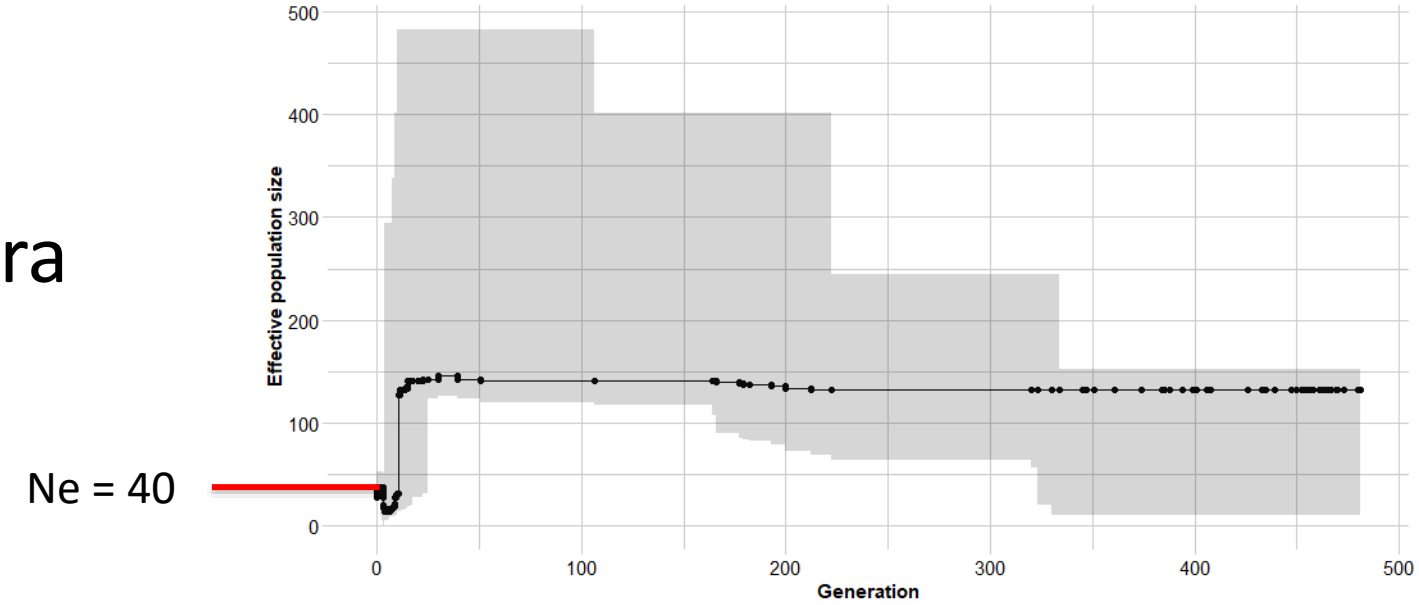
L=nLoc(gl)*75 * 1000



Estimate current Ne

- Idea: use linkage equilibrium to estimate current Ne and "adjust" L in such a way that the current Ne fits the trajectory
- L should be constant for different data sets of the species
- L <- nLoc(gl)*75*500 [500 = fudge factor]

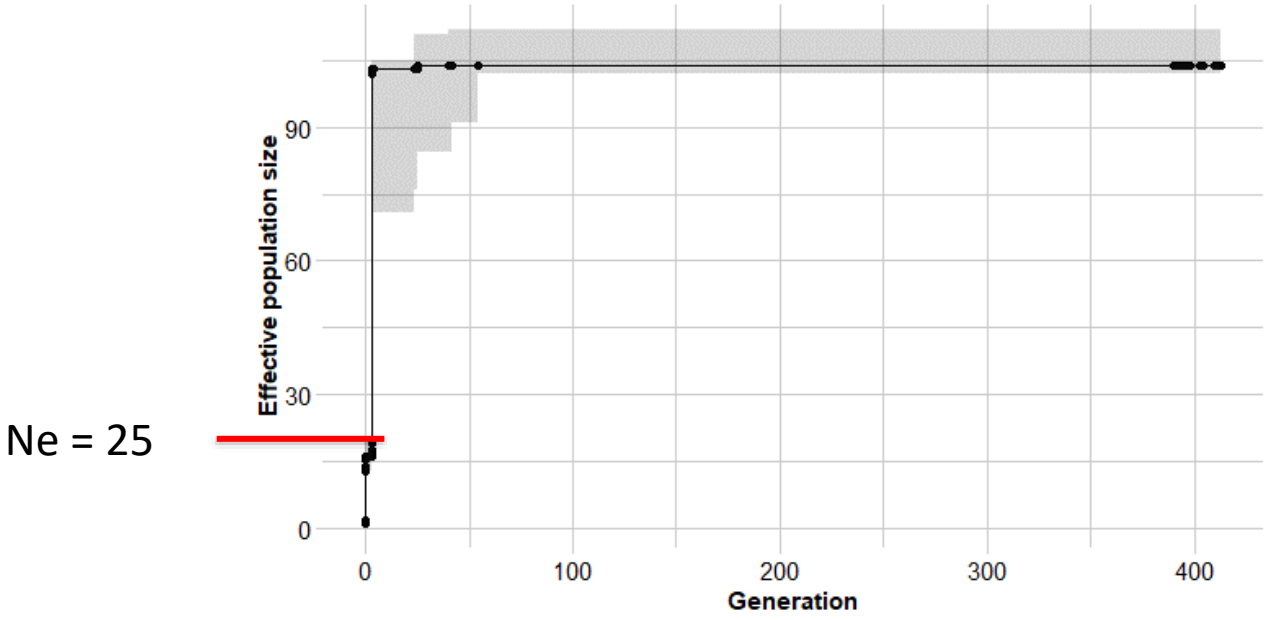
- NeEstimator North Canberra
- Ne = 40



Estimate current Ne

- Idea: use linkage equilibrium to estimate current Ne and “adjust” L in such a way that the current Ne fits the trajectory
- L should be constant for different data sets of the species (FF=500)

- South Canberra
- Ne ~ 25



Run epos and stairways [takes a minute]

- Canberra North