

LANDSCAPE GENETICS USING SNPS

Bernd Gruber
Institute for Applied Ecology
University of Canberra
Australia



Overview

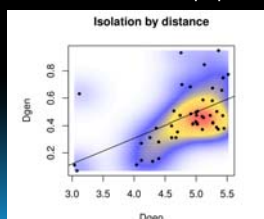
- Isolation by distance
- What is a landscape genetic analysis based on resistance layers and how does it work?
 - Cost distances
 - Partial mantel tests
- The simulator
- A "good" differentiation measure
- Tutorial

Isolation by distance

- Pairwise genetic distances against pairwise geographic distances
 - Based on individuals or subpopulations

Isolation by distance

- Pairwise genetic distances against pairwise geographic distances
 - Based on individuals or subpopulations

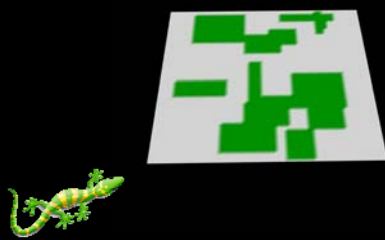


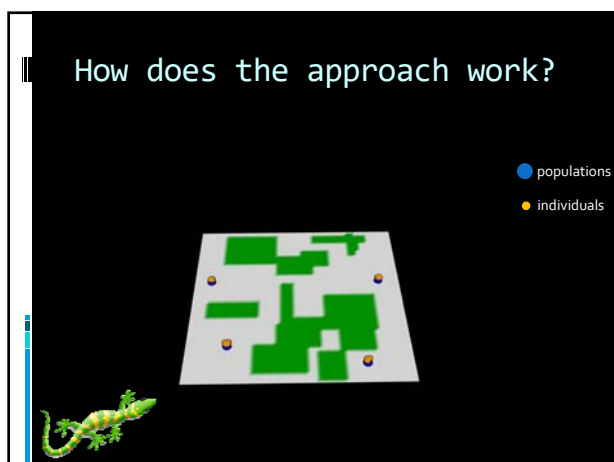
Landscape genetic approach based on resistance values

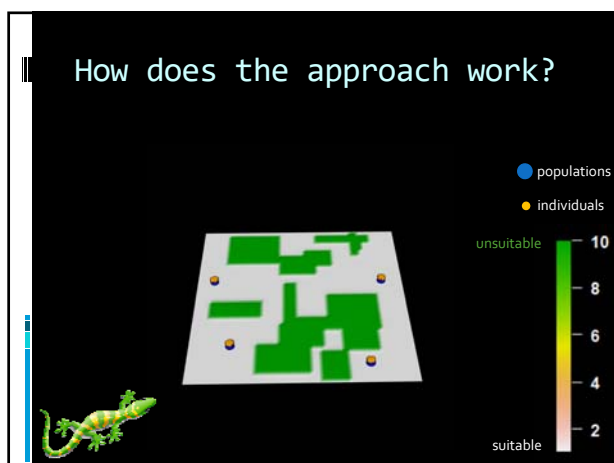
- Pairwise genetic distances against pairwise geographic distances
 - Based on individuals or subpopulations

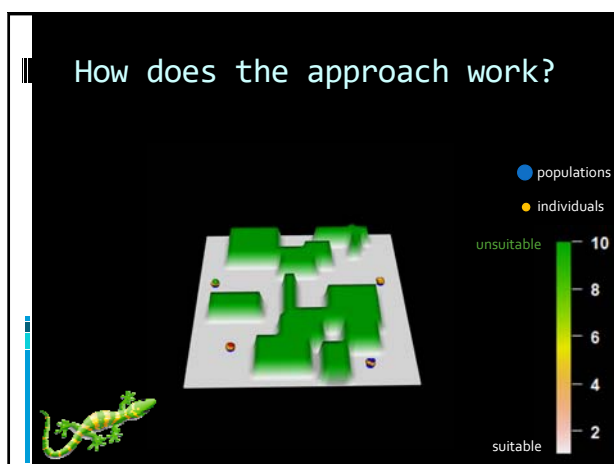
How does the approach work?

● populations



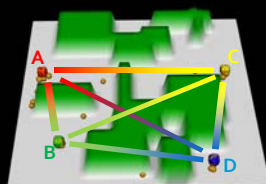






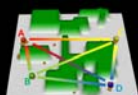
How does the approach work?

Pairwise Genetic Distances


 $G =$

	A	B	C	D
A	-	-	-	-
B	10	-	-	-
C	20	12	-	-
D	14	16	13	-

How does the approach work?


 $G =$

Pairwise Euclidean Distances


 $E =$

	A	B	C	D
A	-	-	-	-
B	10	-	-	-
C	20	24	-	-
D	24	16	13	-

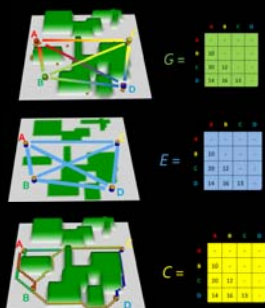
How does the approach work?

Pairwise Cost Distances


 $C =$

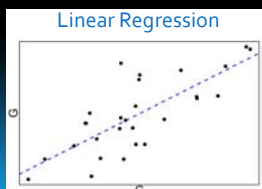
	A	B	C	D
A	-	-	-	-
B	15	-	-	-
C	12	6	-	-
D	8	9	13	-

How does the approach work?



How does the approach work?

- Three pairwise distance matrices
 - G: Genetic Distance matrix
 - E: Euclidean Distance matrix
 - C: Least-Cost Path Distance Matrix
- Is there a correlation between $G \sim C$ (but not $G \sim E$) ??



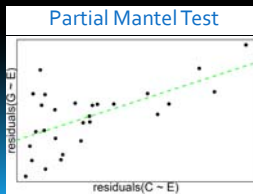
How does the approach work?

- Three pairwise distance matrices
 - G: Genetic Distance matrix
 - E: Euclidean Distance matrix
 - C: Least-Cost Path Distance Matrix
 - Is there a correlation between $G \sim C$ (but not $G \sim E$) ??
- Linear Regression**
- Not Good for two reasons:**
- we ignore non independence
 - we need to correct for E



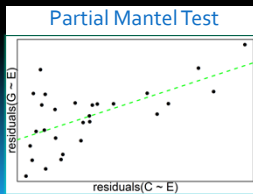
How does a landscape genetic approach work?

- Three pairwise distance matrices
 - G: Genetic Distance matrix
 - E: Eukclidean Distance matrix
 - C: Least-Cost Path Distance Matrix
- Is there a correlation between $\text{resid}(G \sim E) \sim \text{resid}(C \sim E)$?



How does a landscape genetic approach work?

- Three pairwise distance matrices
 - G: Genetic Distance matrix
 - E: Eukclidean Distance matrix
 - C: Least-Cost Path Distance Matrix
- Is there a correlation between $\text{resid}(G \sim E) \sim \text{resid}(C \sim E)$?



Good for two reasons:

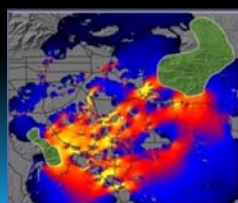
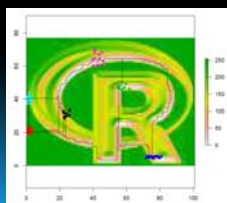
- we treat non independence using a bootstrap approach
- we correct for E

Cost distance algorithms

- Algorithm
 - Least-cost
 - Circuitscape [<http://maps.tnc.org/migrations-in-motion/#3/19.00/-78.00>]
 - rSPDistance ???

Cost distance algorithms

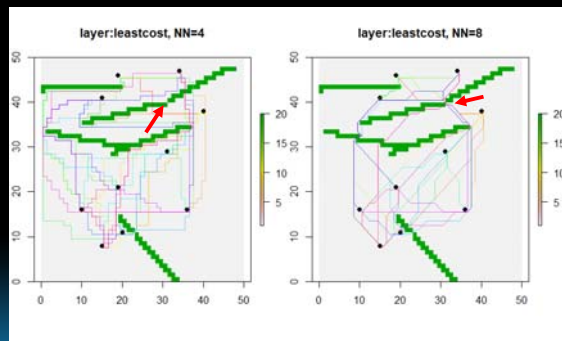
- Algorithm
 - Least-cost [a single path]
 - Circuitscape [summarises over all possible= commute] [<http://maps.tnc.org/migrations-in-motion/#3/19.00/-78.00>]



Cost distance algorithms

- Algorithm:
 - least-cost vs circuitscape vs SPDistance
- Next Neighbours,
 - NN=4 vs NN=8
- Resistance values

Cost distance algorithms



Cost distance algorithms

- Algorithm:
least-cost vs circuitscape vs SPDistance
- Next Neighbours,
NN=4 vs NN=8
- Resistance values
 - Additional data [telemetry]
 - Advocate strongly initial, simple hypothesis of strongly different non-correlating layers
 - though often "fishing for correlation" is done

Tutorial

- Example [possums.gl]

```
> possums.gl
/// GENLIGHT OBJECT //////////
// 300 genotypes, 200 binary SNPs, size: 499 kb
0 (0 %) missing data

// Basic content
@gen: list of 300 SNPbin
@ploidy: ploidy of each individual (range: 2-2)

// Optional content
@ind.names: 300 individual labels
@loc.names: 200 locus labels
@pop: population of each individual (group size range: 30-30)
@other: a list containing: xy latlong
```

Tutorial

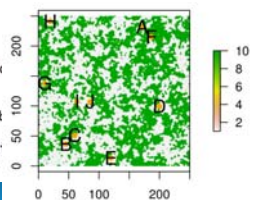
▪ Example [possums.gl]

```
> possums.gl
/// GENLIGHT OBJECT //////////

// 300 genotypes, 200 binary SNPs. size: 499 kb
0 (0 %) missing data

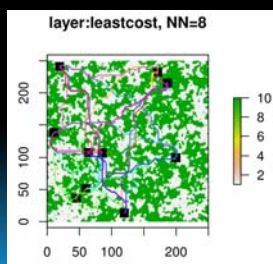
// Basic content
@gen: list of 300 SNPbin
@ploidy: ploidy of each indivi

// Optional content
@ind.names: 300 individual labels
@loc.names: 200 locus labels
@pop: population of each indivi
@other: a list containing: xy
```



Tutorial

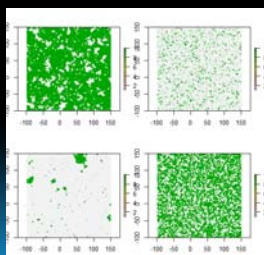
▪ Example [possums.gl]



```
## $mantel.tab
##          model      r      p
## 1 Gen -cost | Euclidean 0.4945 0.028
## 2 Gen -Euclidean | cost -0.2502 0.816
```

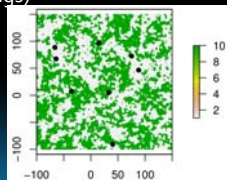
Tutorial - Simulator

▪ Create a landscape



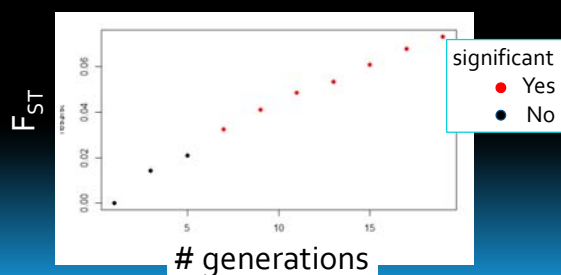
Tutorial - Simulator

- Create a metapopulation on the landscape
 - (npop populations with nind individuals)
 - Define dynamics (n offsprings)
 - Define genetics (n SNPs)
 - Define the dispersal



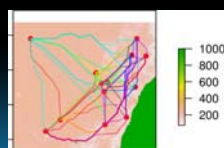
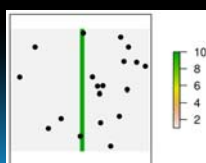
Tutorial - Simulator

- Run and analyse



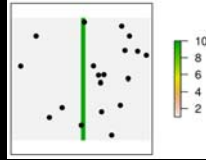
Tutorial - Use your own data

- Create a resistance layer
 - Download a map



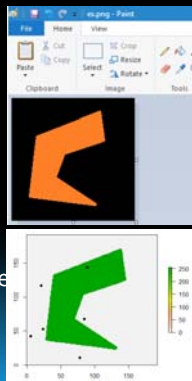
Tutorial – Use your own data

- Create a resistance layer
 - Download a map
 - "Draw" with R



Tutorial – Use your own data

- Create a resistance layer
 - Download a map
 - "Draw" with R
 - Use image editor
- Run a landscape genetics analysis based on resistance layers.



A suitable genetic distance measure

- F_{st} or other (D , G_{st} , G'_{st} , G_{st} Hedrick and lots of others)
- SNPs are special markers
 - biallelic
- How do we measure "being" different
 - Basically we have number of alleles and allele frequency per loci

A suitable genetic distance measure

- Diversity using q values
 - S (number of alleles)
 - p (allele frequency)
 - q (0, 1, 2, ...)

$${}^qH = \left(1 - \sum_{i=1}^S p_i^q\right) / (q - 1).$$

A suitable genetic distance measure

- Diversity using q values

q	H alpha (within)	H beta (between)
0	# alleles	Soerensen, Jaccard
1	# of possible arrangements (=Shannon)	I (=mutual information)
2	heterozygosity (chance of choosing two different alleles)	Jost-D (Fst, Gst)

A suitable genetic distance measure

$${}^qH = \left(1 - \sum_{i=1}^S p_i^q\right) / (q - 1).$$

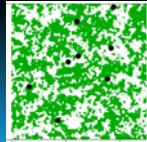
- Diversity using q values (the higher the more influenced by more common alleles!!)

q	H alpha (within)	H beta (between)
0	# alleles	Soerensen, Jaccard
1	# of possible arrangements (=Shannon)	I (=mutual information)
2	heterozygosity (chance of choosing two different alleles)	Jost-D (Fst, Gst)

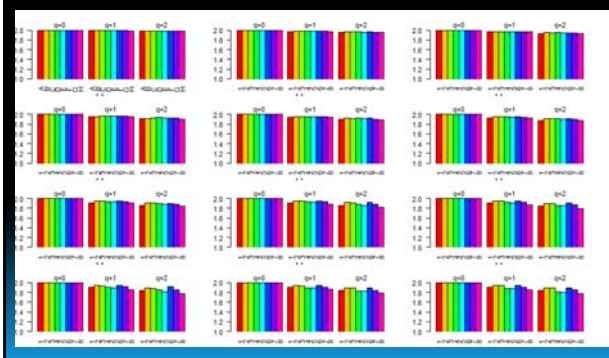
Common alleles

A suitable genetic distance measure

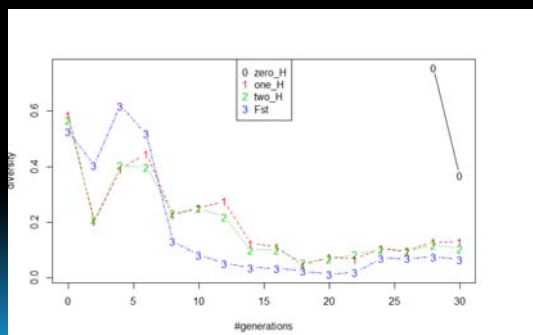
- Change of diversity
- As before run a simulation over several generations on a resi
- Check efficiency of measures of diversity (partial mantel test) o



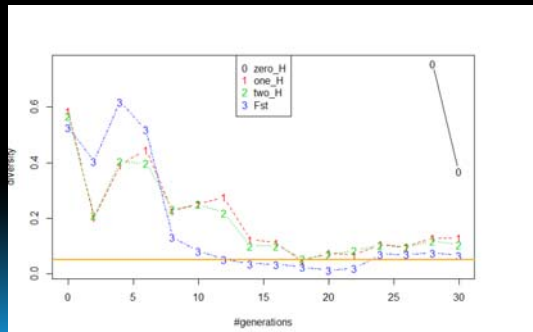
A suitable genetic distance measure



A suitable genetic distance measure



A suitable genetic distance measure



Tutorial:

<https://github.com/green-striped-gecko/dartRworkshop>

Dart_CBA_Tutorial_LandscapeGenetics.pdf
