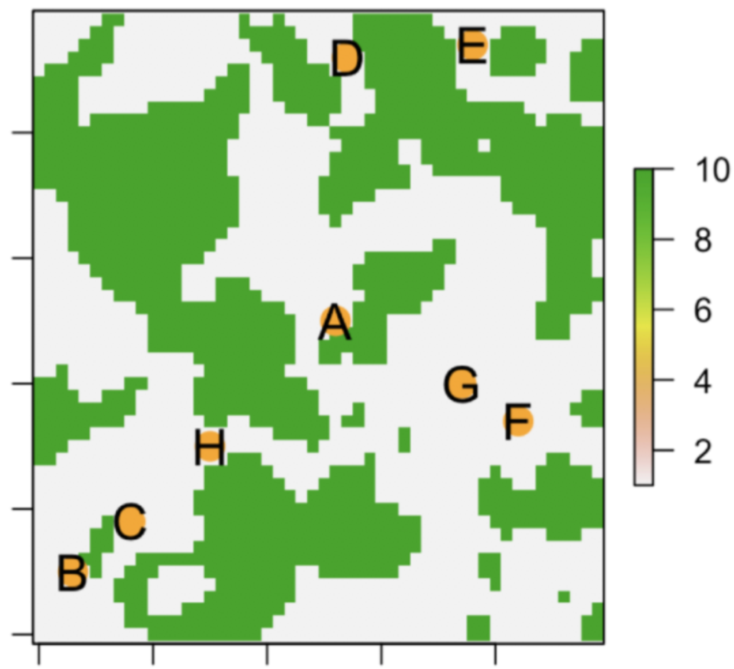
Simulation Experiments in R

Helene Wagner, University of Toronto

Goals:

- Simulate a metapopulation on a resistance landscape
- Compare performance of partial Mantel test and Sunder



Methodological Challenges:

Video 1:

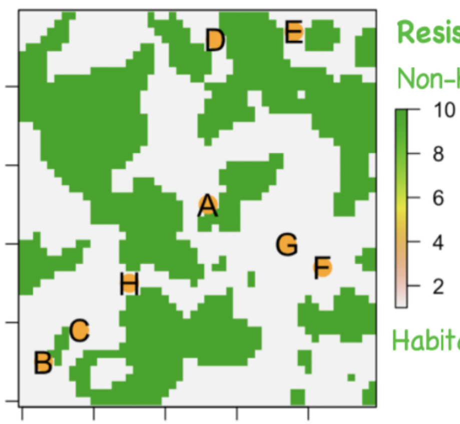
- Workflow of a simulation experiment
- Testing statistical methods with simulations
- Partial Mantel test vs. Sunder

Video 2: Efficient R

Simulation Workflow

1. Initialize

- Landscape map — constant
- Populations (A – H) — constant
- Individuals (genotypes) — variable



Create random maps with 'secr':
- Habitat amount (A)
- Habitat aggregation (p)

2. Time step

- Demographic model
- Mating and reproduction
- Dispersal and recruitment

```
> str(para)
List of 15
 $ n.pops     : num 8
 $ n.ind      : num 100
 $ sex.ratio  : num 0.5
 $ n.cov      : num 3
 $ n.offspring: num 2
 $ mig.rate   : num 0.1
 $ disp.max   : num 50
 $ disp.rate  : num 0.05
 $ n.allels   : num 10
 $ n.loci     : num 20
 $ mut.rate   : num 0.001
```

3. Run a single simulation

- Initialize genotypes
- Run for many time steps
- Collect genotype data
- Summarize results

Fst: degree of differentiation
Decide: IBD or IBR?

4. Batch run simulations

- Replicate runs with same parameters
- Run scenarios across parameter space
- Store results and settings

Parameter space:
- time: # generations
- rep: # replicate sims

```
> para.space
  rep time
1   1    5
2   2    5
3   3    5
4   1   25
5   2   25
6   3   25
7   1   45
8   2   45
9   3   45
```

5. Synthesize results

- Extract summary data
- Visualize in parameter space
- Sensitivity analysis

Robust vs. sensitive

Partial Mantel Tests

IBD IBR

```
PopGenReport::wassermann(eucl.mat = eucl.mat, cost.mats = list(cost=cost.mat),
                         gen.mat = gen.mat, plot=F)$mantel.tab
```
```

| | model <chr> | | r <chr> | p <chr> |
|---|---|---|---|---|
| 1 | Gen ~cost \| Euclidean | IBR \| IBD | 0.5366 | 0.041 |
| 2 | Gen ~Euclidean \| cost | IBD \| IBR | -0.4753 | 0.983 |

Some issues with (partial) Mantel tests:

- Low statistical power?
- Inflated type I error rates if spatial autocorrelation?

Use simulations to test and compare methods!
```

Alternative with 'Sunder'

'Bedassle' (Bradburd et al. 2013), alternative implementation in 'Sunder' (Botta et al. 2014)

Run the analysis (parameter settings: http://www.nbi.dk/~botta/Sunder.html#overview)

IBD
IBR
Iterations

```
D.G <- as.matrix(dist(para$locs))
D.E <- cost.mat
nit <- 10^3   ## just for the example, should be much larger, e.g. 50000
output <- Sunder::MCMCCV(Array,D.G,D.E,
                         nit=nit,thinning=max(nit/10^3,1),
                         theta.max=c(10,10*max(D.G),10*max(D.E),1,0.9),
                         theta.init=c(1,2,1,1,0.01),
                         run=c(1,1,1), ud=c(0,1,1,0,0),
                         n.validation.set=dim(Array)[1]*dim(Array)[2]/10,
                         print.pct=FALSE)
```

```
print(output$mod.lik)
```

```
                IBD       IBR
      G+E         G         E
-9050.244 -9058.499 -8974.353
```

Likelihood

```
> names(which.max(output$mod.lik))
[1] "E"
```

Testing Method Performance

Assessing error rates requires MANY replicate samples!

Resistance values

1	1	1	1	1
1	B	1	1	1
1	1	1	1	1
1	1	1	1	1
A	1	1	1	C

Type I error rate

- Simulate under null hypothesis
- Expect alpha % false positives

Statistical power to detect effect

- Simulate under alternative hypothesis
- Assess True Positive Rate (TPR)
- Larger effect size -> higher power
- Larger sample size -> higher power

3	3	7	7	7
3	B	2	7	3
1	1	2	2	3
1	5	5	2	1
A	5	5	5	C

Compare power between methods!

Statistical Power