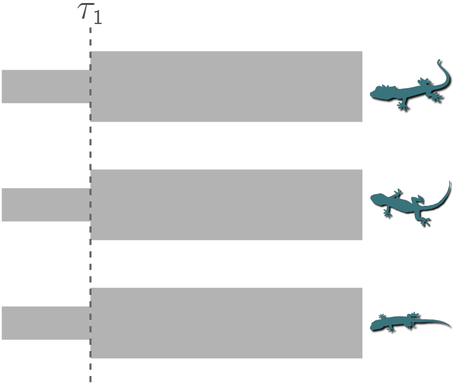
Inferring shared demographic changes from genomic data

Jamie R. Oaks

Department of Biological Sciences &
Museum of Natural History,
Auburn University

June 3, 2018



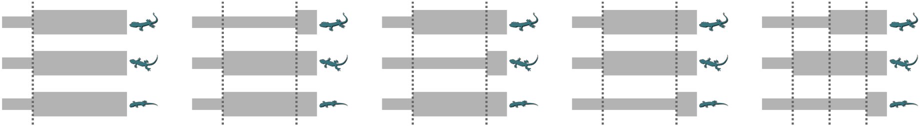
- Recent interest in testing shared demographic changes
- Several nice ABC approaches[1,2,3]
- It's a tricky inference problem

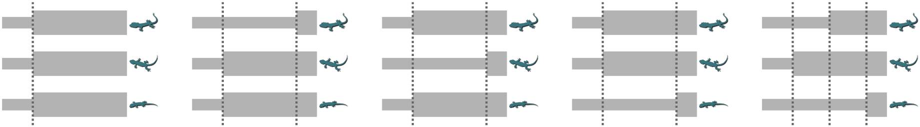
---

[1] Y. L. Chan et al. (2014). *Molecular Biology and Evolution* 31: 2501–2515

[2] A. T. Xue et al. (2015). *Molecular Ecology* 24: 6223–6240

[3] X. A. T. et al. (2017). *Molecular Ecology Resources* 17: e212–e224





Given genomic data, can we infer the correct model and the timing of the demographic events?

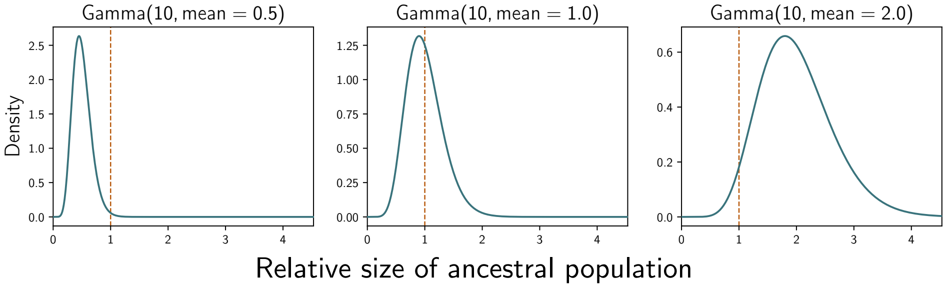
Ecoevolity: Estimating evolutionary coevality

- Simulate datasets with 500k characters
- Use Bayesian model averaging with full likelihood[1,2]
- No model misspecification
- Time of change $\sim$ Exponential(mean $= 0.01$)

---

[1] J. R. Oaks (2018). *bioRxiv*

[2] D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



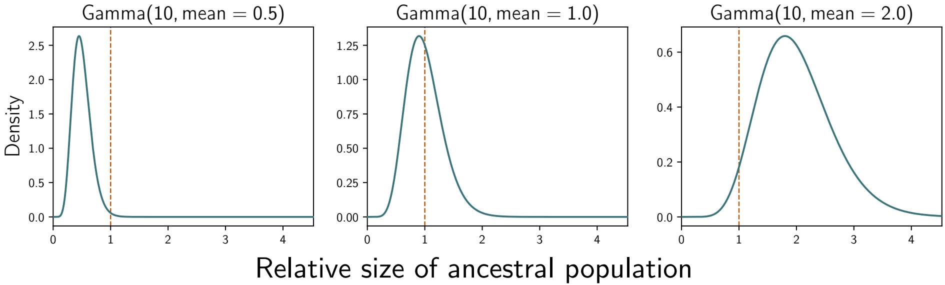
Ecoevolity: Estimating evolutionary coevality

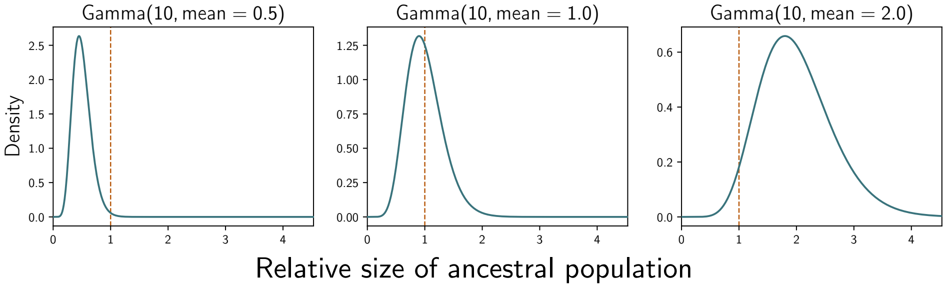
- Simulate datasets with 500k characters
- Use Bayesian model averaging with full likelihood[1,2]
- No model misspecification
- Time of change $\sim$ Exponential(mean $= 0.01$)

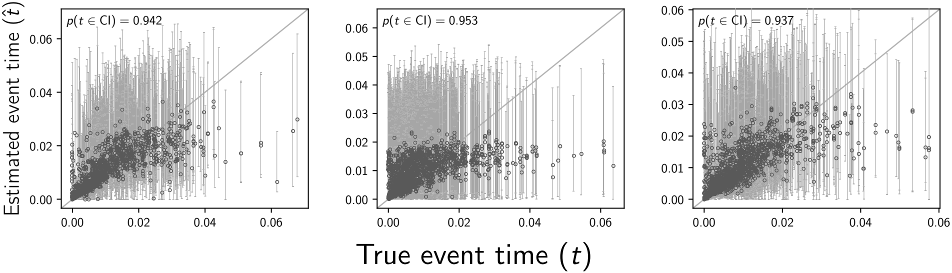
---

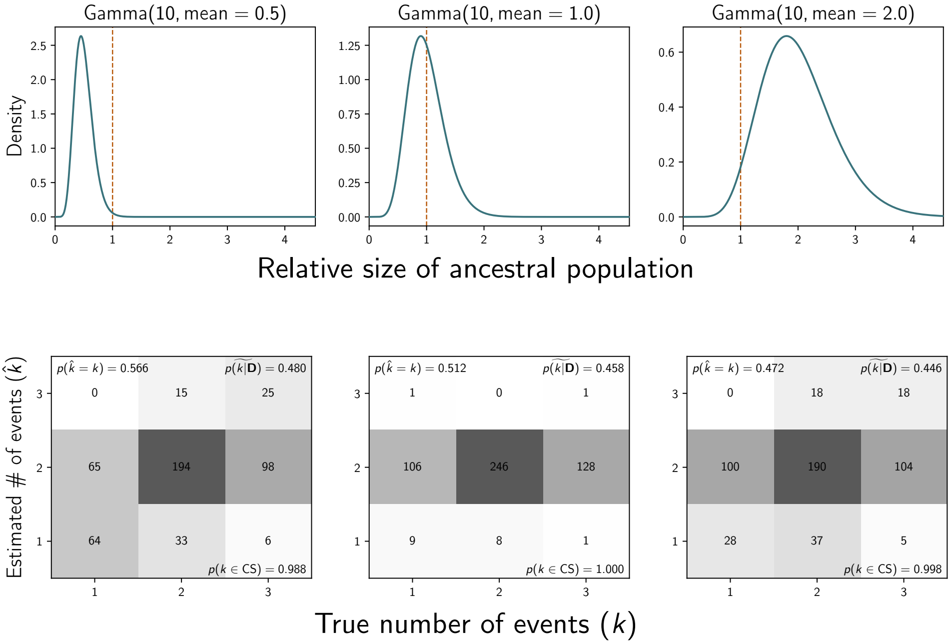
[1] J. R. Oaks (2018). *bioRxiv*

[2] D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932


Gamma(10, mean = 0.5)
Gamma(10, mean = 1.0)
Gamma(10, mean = 2.0)
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.00
0.25
0.50
0.75
1.00
1.25
0.0
0.2
0.4
0.6
0
1
2
3
4
Relative size of ancestral population


Gamma(10, mean = 0.5)
Gamma(10, mean = 1.0)
Gamma(10, mean = 2.0)
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.00
0.25
0.50
0.75
1.00
1.25
0.0
0.2
0.4
0.6
0
1
2
3
4
Relative size of ancestral population


$p(t \in \mathrm{CI}) = 0.942$
$p(t \in \mathrm{CI}) = 0.953$
$p(t \in \mathrm{CI}) = 0.937$
Estimated event time ($\hat{t}$)
0.00
0.01
0.02
0.03
0.04
0.05
0.06
True event time (t)



$p(\hat{k} = k) = 0.566$, $p(\widetilde{k|\mathbf{D}}) = 0.480$, $p(k \in \text{CS}) = 0.988$

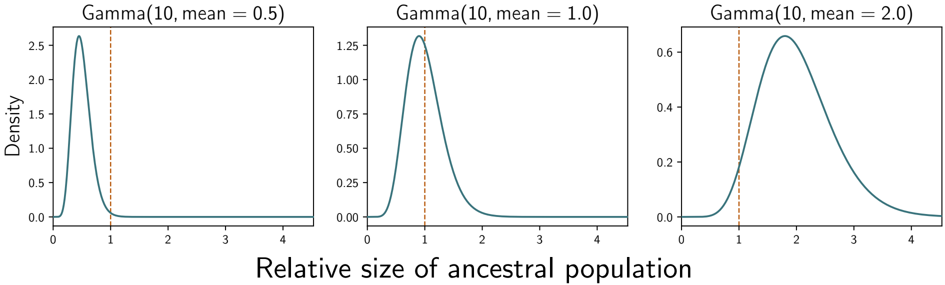
$\hat{k}$ \ k	1	2	3
3	0	15	25
2	65	194	98
1	64	33	6

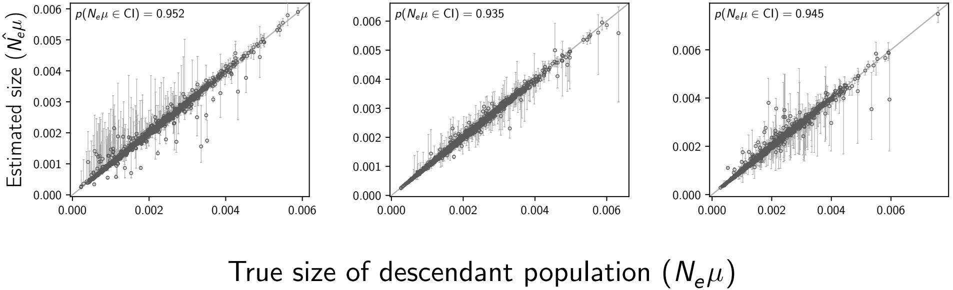
$p(\hat{k} = k) = 0.512$, $p(\widetilde{k|\mathbf{D}}) = 0.458$, $p(k \in \text{CS}) = 1.000$

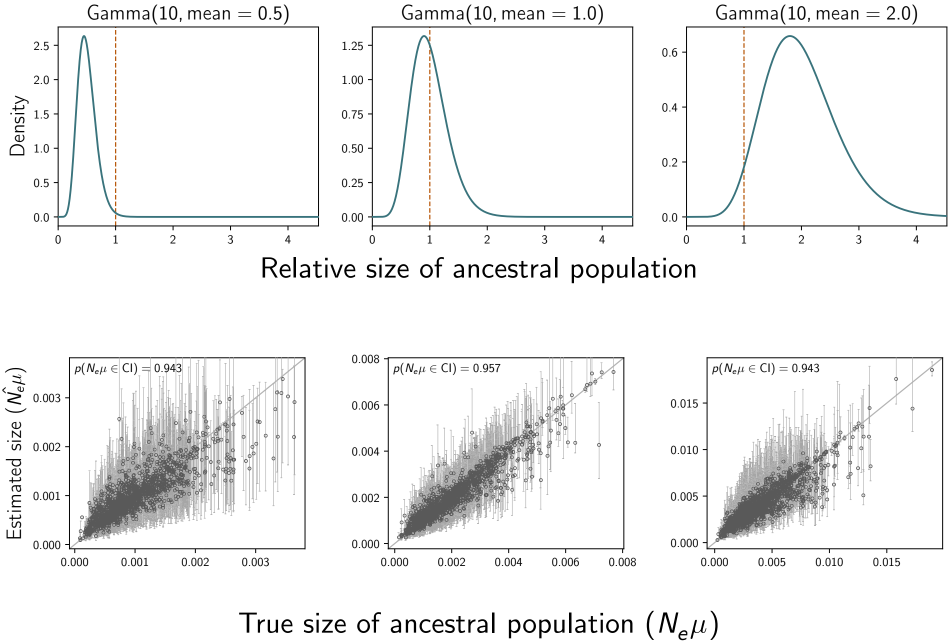
$\hat{k}$ \ k	1	2	3
3	1	0	1
2	106	246	128
1	9	8	1

$p(\hat{k} = k) = 0.472$, $p(\widetilde{k|\mathbf{D}}) = 0.446$, $p(k \in \text{CS}) = 0.998$

$\hat{k}$ \ k	1	2	3
3	0	18	18
2	100	190	104
1	28	37	5

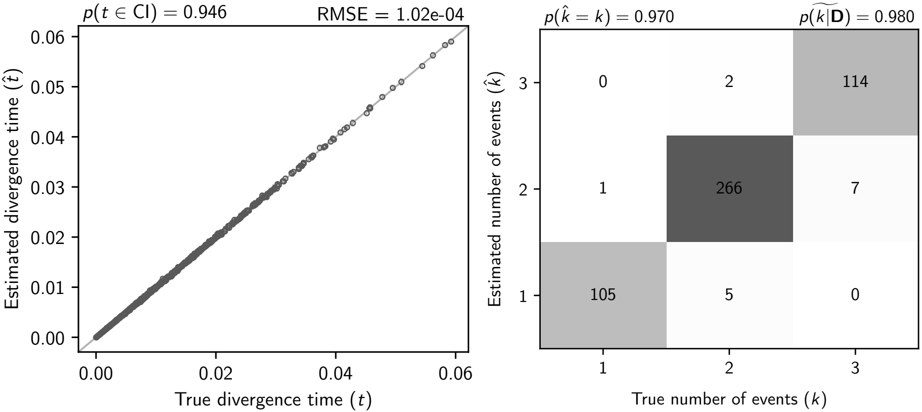

Gamma(10, mean = 0.5)
Gamma(10, mean = 1.0)
Gamma(10, mean = 2.0)
Density
0.0
0.5
1.0
1.5
2.0
2.5
0.00
0.25
0.50
0.75
1.00
1.25
0.0
0.2
0.4
0.6
0
1
2
3
4
Relative size of ancestral population


$p(N_e\mu \in \text{CI}) = 0.952$
$p(N_e\mu \in \text{CI}) = 0.935$
$p(N_e\mu \in \text{CI}) = 0.945$
Estimated size ($\hat{N_e\mu}$)
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.000
0.002
0.004
0.006
True size of descendant population ($N_e\mu$)


Gamma(10, mean = 0.5)
Gamma(10, mean = 1.0)
Gamma(10, mean = 2.0)
Density
2.5
2.0
1.5
1.0
0.5
0.0
1.25
1.00
0.75
0.50
0.25
0.00
0.6
0.4
0.2
0.0
0
1
2
3
4
Relative size of ancestral population
$p(N_e\mu \in \text{CI}) = 0.943$
$p(N_e\mu \in \text{CI}) = 0.957$
$p(N_e\mu \in \text{CI}) = 0.943$
Estimated size $(\hat{N_e\mu})$
0.000
0.001
0.002
0.003
0.004
0.006
0.008
0.005
0.010
0.015
True size of ancestral population $(N_e\mu)$

Thoughts

- Yes, this is a difficult inference problem
- Next step:
  - Are there regions of parameter space where this works?



	1	2	3
3	0	2	114
2	1	266	7
1	105	5	0

Good news: Inferring shared *divergence* times does seem to work[1]

---

[1] J. R. Oaks (2018). *bioRxiv*

Everything is on GitHub…

Software:

- Ecoevolity: `http://phyletica.org/ecoevolity`

Open-Science Notebooks:

`https://github.com/phyletica/ecoevolity-demog-experiments`
`https://github.com/phyletica/ecoevolity-experiments`

Acknowledgments

Ideas and feedback:

- Phyletica Lab
- Leaché Lab
- Minin Lab
- Mark Holder

Computation:

AUBURN UNIVERSITY

Funding:

NSF

Photo credits:

- PhyloPic!

Questions?

joaks@auburn.edu