

Founder-event speciation dramatically improves likelihoods and alters parameter inference in Dispersal-Extinction-Cladogenesis (DEC) analyses

Nicholas J. Matzke¹

¹ CTEG: Center for Theoretical and Evolutionary Genomics
Department of Integrative Biology, University of California, Berkeley



matzke@berkeley.edu

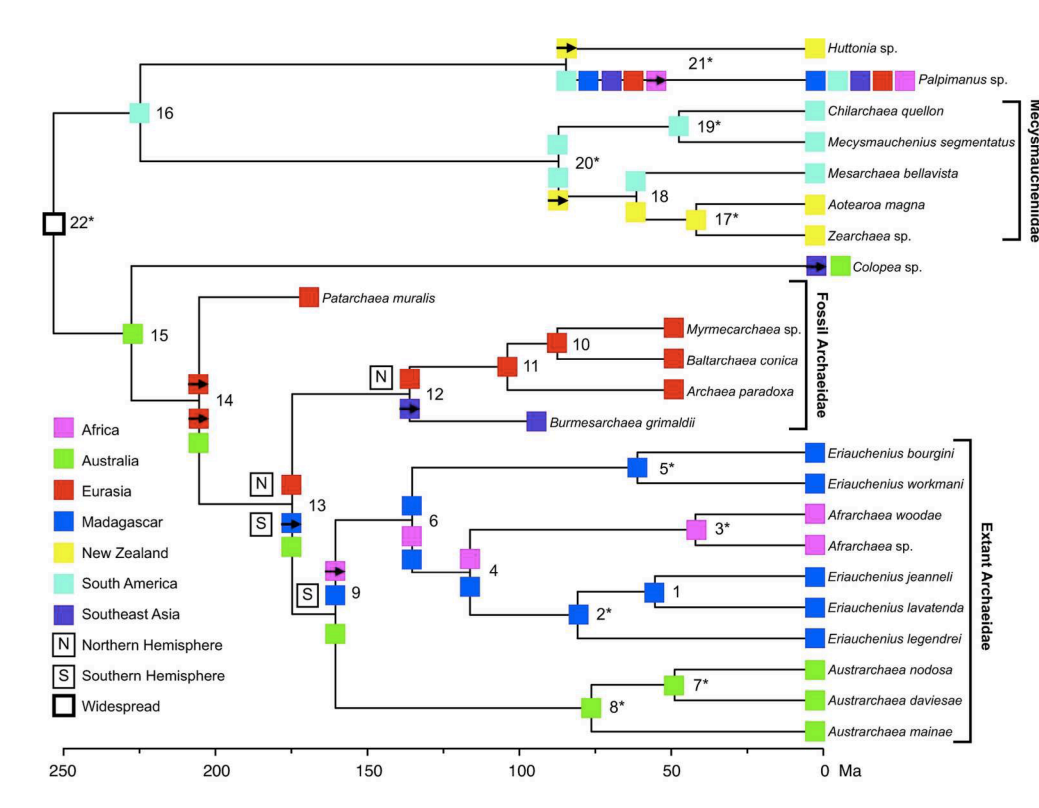


My QR code to get contact
info, CV, website, R code!
Read with e.g. Bakodo

Introduction

Probabilistic modeling of geographic range evolution was a major advance in historical biogeography, making biogeographical problems accessible to model-based maximum likelihood (ML) and Bayesian methodologies. The most popular model is Dispersal-Extinction-Cladogenesis (DEC), implemented in the software LAGRANGE (Ree & Smith 2008).

Example LAGRANGE analysis of assassin spiders (Figure 4 from Wood, Matzke et al. 20121). Phylogeny is tip-dated with amber-preserved fossils. We interpreted the results as support for vicariance with the breakup of Gondwanaland, but LAGRANGE cannot formally test this hypothesis.



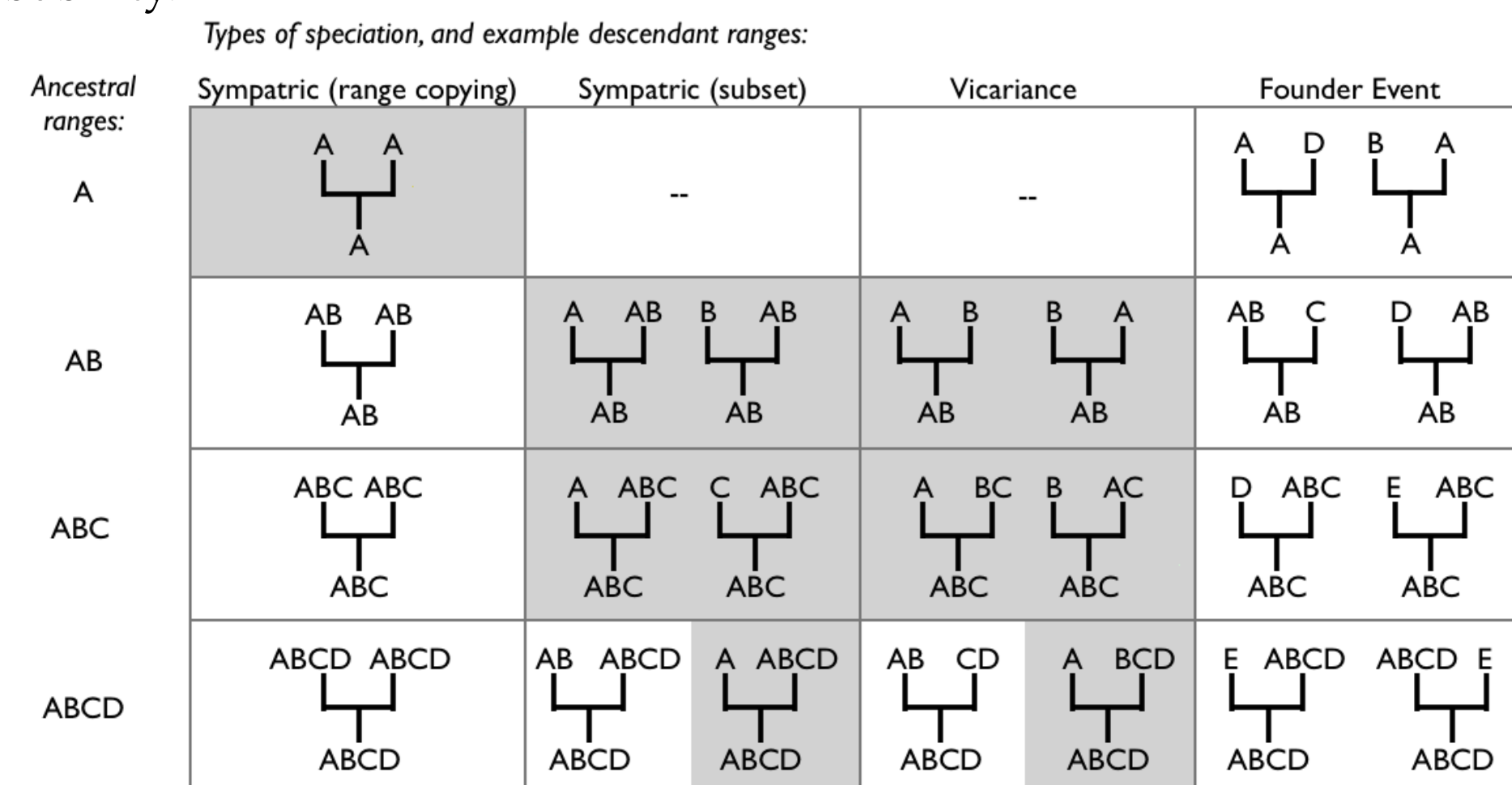
Dispersal and Extinction

Standard LAGRANGE DEC is a model with two free parameters specifying the rate of “dispersal” (range expansion) and “extinction” (range contraction).

Cladogenesis

However, while dispersal and extinction rates are free parameters, the cladogenesis model is fixed in LAGRANGE, such that the geographic range of the ancestral lineage is inherited by the two daughter lineages through a variety of scenarios fixed to have equal probability.

The types of cladogenesis events allowed by LAGRANGE are shown in gray (right).



The fixed nature of the cladogenesis model means that it has been indiscriminately applied in all DEC analyses, and has not been subjected to any inference or formal model testing.

BioGeoBEARS avoids these limitations.

Methods

I re-implement DEC in my R package BioGeoBEARS, which exactly reproduces LAGRANGE 2-parameter inferences and likelihoods.

I also implement additional parameters controlling the probability of new cladogenesis models, such as “founder-event speciation” (Templeton 2008), in which one daughter jumps to an area completely outside the ancestral range. Other parameters control P(vicariance), descendent range sizes, etc.


I compare the results of these methods on (1) the standard published datasets used by the LAGRANGE authors, and additional datasets; (2) simulated geographical histories.

Range: AB $\xrightarrow[\text{event}]{\text{dispersal}}$ Range: ABC

Range: ABC $\xrightarrow[\text{event}]{\text{extinction}}$ Range: AB

Above: Events on branches

Below: Events at nodes

ABCD —  **sympatric subset event**

ABC

D

Results – inference

Results – inference

Phylogeny	Model	# free parameters	# areas	dispersal (d)	extinction (e)	founder-event weight (f)	variance weight (v)	sympatric weight (remainder)	Desc. ranges small (R=1) or large (R=0)	Variant small (R=0) or large (R=1)	Log-likelihood	AIC	AIC weight compared to LAGRANGE model
Hawaiian <i>Psychotria</i> (Ree & Smith 2008); original LAGRANGE dataset	sympatric (range-copying) only	2	4	0.0287	0.1894	0	0.5	0.5	1	1	-39.72	83.04	5.63E-03
	standard LAGRANGE	2	4	0.0305	0.0289	0	0.5	0.5	0.0001	0.0001	-34.54	78.48	1
	add j (founder-event speciation)	3	4	1E-15	1E-15	0.05608	0.472	0.472	0.0001	0.0001	-20.95	47.47	294893.5
	add v (variance weight)	4	4	1E-15	1E-15	0.2775	0.741	0.2313	0.0001	0.0001	-20.85	49.68	120073.9
	add R (desc. range size)	5	4	1E-15	1E-15	0.02655	0.7523	0.2211	0.121	0.0001	-20.84	51.68	44415.1
	add Rv (desc. range size)	5	4	1E-15	1E-15	0.02656	0.7523	0.2211	0.0001	0.121	-20.84	51.68	44415.1
add both R and Rv	6	4	1E-15	1E-15	0.02725	0.7455	0.2272	0.197	0.4243	-20.84	53.68	16337.9	
Northern Hemisphere <i>Loniceria</i> (honeysuckle); Smith & Donoghue 2010)	sympatric (range-copying) only	2	4	0.0058	0.01716	0	0.5	0.5	1	1	-146.89	297.7	3.05E-10
	standard LAGRANGE	2	4	0.0058	0.00105	0	0.5	0.5	0.0001	0.0001	-124.98	253.95	1
	add j (founder-event speciation)	3	4	0.0026	1E-15	0.01739	0.4913	0.4913	0.0001	0.0001	-106.36	218.71	44892469.8
	add v (variance weight)	4	4	0.0022	1E-15	0.03344	0.099	0.8676	0.0001	0.0001	-106.27	220.53	18098711.6
	add R (desc. range size)	5	4	0.0022	1E-15	0.03244	0.1043	0.8624	0.0016	0.0001	-106.27	222.52	656986.0
	add Rv (desc. range size)	5	4	0.0022	1E-15	0.03324	0.1043	0.8624	0.0001	0.0016	-106.27	222.53	6569975.3
add both R and Rv	6	4	0.0022	1E-15	0.03277	0.1192	0.848	1E-05	0.0466	-106.27	224.53	2448281.4	
Pacific <i>Cyrtandra</i> (Clark et al. 2008)	sympatric (range-copying) only	2	7	0.0014	0.01384	0	0.5	0.5	1	1	-62.57	129.14	9.74E-08
	standard LAGRANGE	2	7	0.0013	1E-15	0	0.5	0.5	0.0001	0.0001	-46.43	96.85	1
	add j (founder-event speciation)	3	7	1E-15	1E-15	0.00471	0.4976	0.4976	0.0001	0.0001	-42.07	90.13	28.8
	add v (variance weight)	4	7	1E-15	1E-15	0.0054	0.4107	0.5839	0.0602	0.0001	-42.10	94.20	3.8
	add R (desc. range size)	5	7	1E-15	1E-15	0.00526	0.4254	0.5693	0.0001	0.3634	-42.09	94.7	3.8
	add Rv (desc. range size)	5	7	1E-15	1E-15	0.00625	0.3402	0.6535	0.0006	0.064	-42.27	96.53	1.2
Southwestern <i>Cypripodontidae</i> (desert pupfish) (with Matt Van Damm)	sympatric (range-copying) only	2	7	0.0103	0.09377	0	0.5	0.5	1	1	-74.06	152.12	8.80E-08
	standard LAGRANGE	2	7	0.0102	0.00942	0	0.5	0.5	0.0001	0.0001	-57.81	119.63	1
	add j (founder-event speciation)	3	7	1E-15	1E-15	0.00436	0.4978	0.4978	0.0001	0.0001	-48.81	103.63	2974.3
	add v (variance weight)	4	7	1E-15	1E-15	0.00281	0.6828	0.3144	0.0001	0.0001	-48.51	105.01	140.2
	add R (desc. range size)	5	7	1E-15	1E-15	0.00635	0.2419	0.7318	0.0001	0.2064	-49.75	109.50	157.9
	add Rv (desc. range size)	5	7	1E-15	1E-15	0.00622	0.2716	0.7222	0.0189	0.1225	-49.69	111.39	61.5
Fossil Miocene horses (with Kaitlin Maguire)	sympatric (range-copying) only	2	5	0.0122	0.04015	0	0.5	0.5	1	1	-452.82	919.68	1.19E-20
	standard LAGRANGE	2	5	0.0099	0.00002	0	0.5	0.5	0.0001	0.0001	-406.94	870.83	1
	add j (founder-event speciation)	3	5	0.0079	1E-15	0.01405	0.463	0.463	0.0001	0.0001	-390.50	787.00	5077290.7
	add v (variance weight)	4	5	0.007	1E-15	0.02646	0.2317	0.7418	0.0001	0.0001	-387.73	783.25	3207937.1
	add R (desc. range size)	5	5	0.007	1E-15	0.02624	0.2361	0.7377	0.1162	0.0001	-387.78	785.25	12203196.5
	add Rv (desc. range size)	5	5	0.0067	1E-15	0.03181	0.1382	0.8299	0.0001	0.0832	-388.15	786.30	720770.0
add both R and Rv	6	5	0.0067	1E-15	0.04067	0.1147	0.8446	1E-05	0.0463	-388.78	789.50	141763.7	

The effects are dramatic in several test datasets. For example, for Hawaiian *Psychotria*, the original test clade for LAGRANGE, addition of founder-event speciation to a geographically unconstrained analysis improves the log-likelihood from -34.5 to -20.9, and the ML inferences for dispersal and extinction rates are moved to zero. The 3-parameter model is ~300,000 times more probable in terms of relative AIC weight.

New features in BioGeoBEARS

Parameter / feature	LAGRANGE	BioGeoBEARS
Probability of vicariance (v)	<i>Fixed</i> to have equal weight	Vary v from 0-1 and estimate
Rangesize smaller descendant (R)	Smaller range <i>always</i> = 1 area	$r = 0.001$: always 1 $r = 0.5$: equal prob for all sizes $r = 1$: always largest possible
Distance dependence for dispersal, founder-events ($dmat$)	Users can hard-core a dispersal matrix; but this is usually subjective	Make <i>dmat</i> a function of distance and estimate
Probability of detection (p)	Assumes ranges at phylogeny tips are known perfectly	Occupancy model from ecology used to <i>estimate</i> P(presence) in each area for each tip
Inclusion of fossils	Fossils can only be included if each is a separate species (phylogeny tip) and the observed fossil range is assumed to be complete.	Detection model, and control of the speciation model, allows inclusion of partial range data from fossil species at multiple timepoints over their lifespan

Conclusions

Careful consideration of cladogenesis models in historical biogeography is recommended. (1) Models for the evolution of biogeography should not be *assumed*, they should be *tested*. (2) Founder-event speciation is likely to be a dominant process structuring phylogenetic patterns between regions. (3) Fossils and detection models can be added.

Results - simulations

Dispersal rates are usually accurately estimated, but extinction rates are not unless rates are low and the model is correct and simple. Estimating j accurately is difficult but any nonzero j has a large impact. Regardless of the model, areas are accurate at low rates (but exact states are not very accurate).

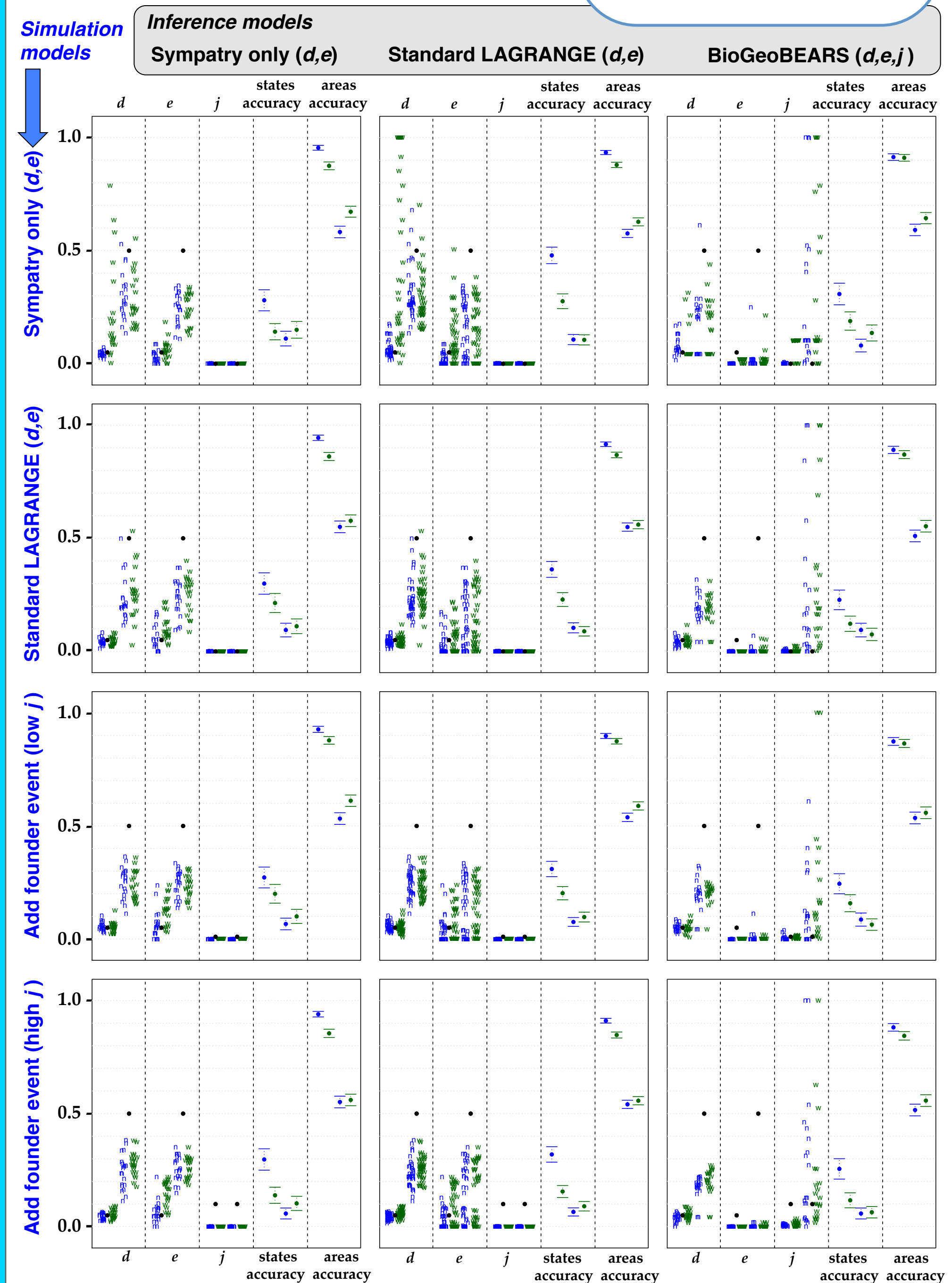
Simulation conditions:

Row 1. Sympatry (range-copying) only
Row 2. Standard LAGRANGE
Row 3. LAGRANGE + low weight(founder)
Row 4. LAGRANGE + high weight(founder)

Each with:
low rates ($d, e = 0.05 = \bullet$) or high ($d, e = 0.5 = \bullet$)
Ancestral range narrow (n) or wide (w)

Inference conditions:

Column 1. Sympatry (range-copying) only
Column 2. Standard LAGRANGE
Column 3. LAGRANGE + Founder-events



References & more info

Use QR code or go to: <http://phylo.wikidot.com/nicholas-j-matzke>

Acknowledgements: John Huelsenbeck, Tony Barnosky, Roger Byrne, David Jablonski, Scott Steppan, Rick Ree, Stephen Smith. Kaitlin Maguire could not attend the meeting, but is a coauthor on the fossils-in-lagrange section. This work was funded by U.C. Berkeley and IBS travel grants, a Google Summer of Code grant, a Wang Fellowship, a Tien Fellowship, and the NSF grant "Bivalves in Time and Space." Thanks also to Mat Van Damm and Kaitlin Maguire.