

Missing Data in Phylogenomics

Rachel Schwartz

University of Rhode Island

January 10, 2017

Talk to me about research opportunities
working on methods for phylogenomic data...

@rachelss

<http://rachelss.github.io>

<http://github.com/rachelss>

Be Afraid

Be Very

Afraid



**KEEP
CALM
AND
CARRY
ON**



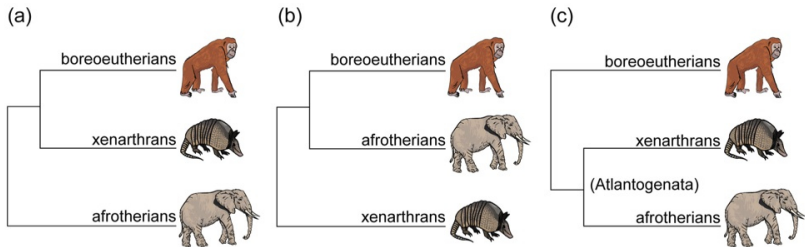
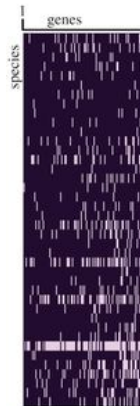


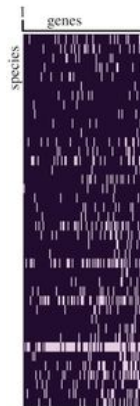
Figure 1. The root of the evolutionary tree of living placental mammals. (a) Afrotherian root. (b) Xenarthran root. (c) Atlantogenatan root.



Gonzalez et al. 2015



Gonzalez et al. 2015



Gonzalez et al. 2015







**KEEP
CALM
AND
CARRY
ON**

Who says missing
data is a problem?



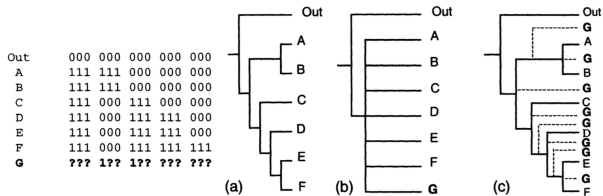
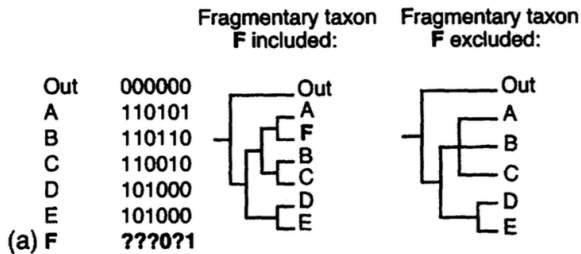
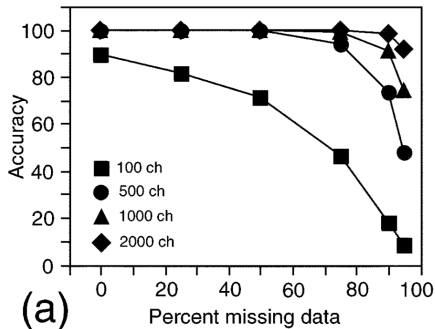


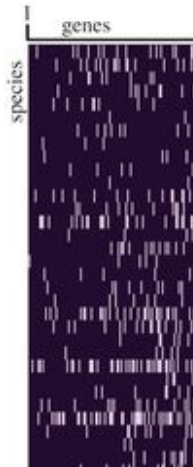
FIGURE 1. Effect of inclusion or exclusion of a wildcard taxon (redrawn from Nixon and Wheeler, 1992). (a) Single tree resulting from analysis of the data set excluding the wildcard taxon G. (b) Strict consensus of eight trees resulting from analysis of the data set including the wildcard taxon G. (c) The eight possible positions of G are indicated with dashed lines, a result of analysis of the data set with the program HENNIG86 (Farris, 1988) or PAUP (Swofford, 1993). If analyzed in NONA (Goloboff, 1993), four of those eight trees are found due to the algorithm's different approach to ambiguous character optimizations. The strict consensus of those four is still completely unresolved, however, and the basic wildcard problem remains.



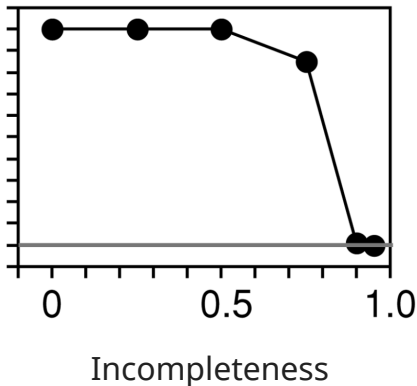
Just need enough good data



More taxa?



Divide the long branches!



2000 char
brlens = 0.2

Empirical studies show no effect



Empirical studies show no effect

Driskell et al 2004

Phillippe et al 2004

Weins 2005

Fulton and Strobeck 2006

Cho et al 2011

Jiang et al 2014

Streicher et al 2016

Fernandez et al 2016

Dufort 2016

Irisarri and Meyer 2016

Sharma et al 2015

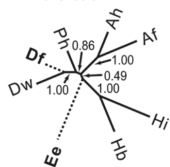
Syst. Biol. 58(1):130–145, 2009
Copyright © Society of Systematic Biologists
DOI:10.1093/sysbio/syp017
Advance Access publication on May 21, 2009

The Effect of Ambiguous Data on Phylogenetic Estimates Obtained by Maximum Likelihood and Bayesian Inference

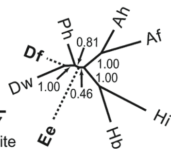
ALAN R. LEMMON^{1,2,3,*}, JEREMY M. BROWN¹, KATHRIN STANGER-HALL⁴, AND EMILY MORIARTY LEMMON^{1,3}

Distant Invariable

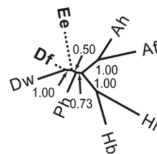
0 sites



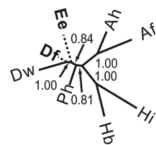
100 sites

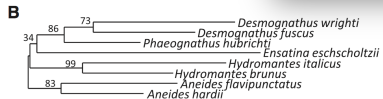
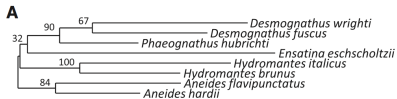


550 sites

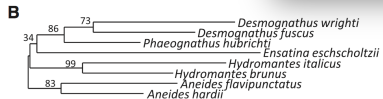
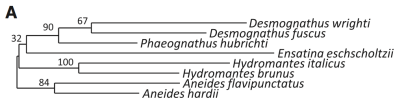


1,000 sites





Roure et al. 2013



Roure et al. 2013

Weins and Morrill 2011



**KEEP
CALM
AND
CARRY
ON**