# 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...
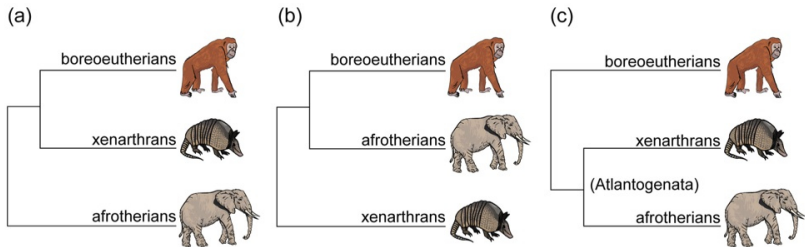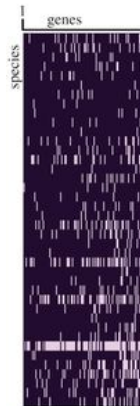
Gonzalez et al. 2015

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

genes
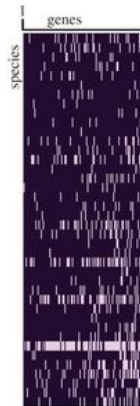
species

Gonzalez et al. 2015

Gonzalez et al. 2015

genes

species

Gonzalez et al. 2015

Gonzalez et al. 2015

Gonzalez et al. 2015

Who says missing data is a problem?

# Morphology

# Morphology

# Morphology

# Morphology
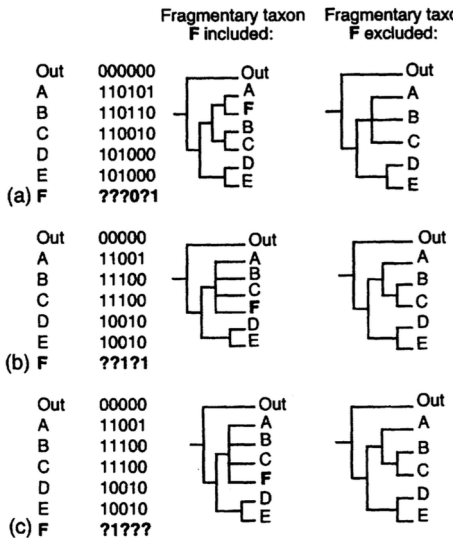
FIGURE 1. Effect of inclusion or exclusion of a wildcard taxon (redrawn from Nixon and Wheeler, 1992). (a) Single tree resulting from analysis of the data set excluding the wildcard taxon G. (b) Strict consensus of eight trees resulting from analysis of the data set including the wildcard taxon G. (c) 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.

Kearney 2002

Fragmentary taxon **F** included:  Fragmentary taxon **F** excluded:

(a)
| Out | 000000 |
| A | 110101 |
| B | 110110 |
| C | 110010 |
| D | 101000 |
| E | 101000 |
| **F** | **???0?1** |

(b)
| Out | 00000 |
| A | 11001 |
| B | 11100 |
| C | 11100 |
| D | 10010 |
| E | 10010 |
| **F** | **??1?1** |

(c)
| Out | 00000 |
| A | 11001 |
| B | 11100 |
| C | 11100 |
| D | 10010 |
| E | 10010 |
| **F** | **?1???** |

(a)
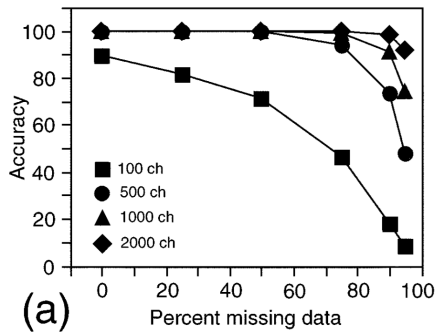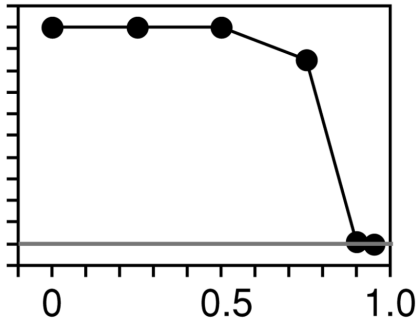
Gonzalez et al. 2015

# Divide the long branches!



Incompleteness
2000 char
brlens = 0.2

# Small genetic datasets

Perhaps it was a problem
eg Novacek, 1992; Wiens  Reeder, 1995;
Wilkinson, 1995;
Huelsenbeck, 1991; Hartmann  Vision, 2008).?
(Weins and Moen 2008)

# empirical studies show no effect



© Michael Quinton

# method doesn't matter

# Simulation studies

# Simulation studies showing problems

can be refuted