An introduction to trees and networks

David Penny

Allan Wilson Centre for Molecular Ecology and Evolution
overview

trees and networks (with cycles)
spanning and phylogenetic (Steiner) trees
exploratory data analysis, visualization
multiscale
splits
compatibility/incompatibility
application - the MinMax squeeze
tree terminology

1c, spanning tree

1f, phylogenetic (Steiner) tree
multiscale, popn growth
lose internal nodes
minimum spanning subtrees
even using complete sequences
lose nodes near the root
mixture of spanning and Steiner trees
minimum spanning subtrees
even using complete sequences
the question of multiscale

effects of time scale

sequences abundant ↔ sequences sparse
ancestral states still present ↔ speciation only at tips
multifurcations and cycles ↔ binary phylogenetic trees
sites in same rate class (gamma) ↔ sites on or off, covarion model
neutral plus deleterious ↔ hypervariable sites ↔ average over all sites

TIME SCALE

generations populations species families orders classes phyla
microevolution ↔ macroevolution
dataset used to argue poor performance of RASA

Data set against RASAdataset used to argue poor performance of RASA

some reasons for cycles?

repeated mutations – random (insufficient information)
  - systematic [e.g. GC bias]
  - positive selection

hybridization

gene conversion/concerted evolution

lateral transfer (incl. endosymbiosis)

alternative (non-Steiner tree) models
  (incorrect assumptions about the mechanism)

exploratory data analysis - EDA
consensus methods lose information

Three data sets from Bandelt et al (1995). The consensus of each would be a polytomy as for (a). However, the median network of each shows that each has a different data structure
are sites always in the same rate class?

observed
NJ
LowG+C cyanobacteria
HighG+C green plants.

expected under i.i.d
Split graph/d_{cov}
LowG+C cyanobacteria
HighG+C green plants.

AtpB

Similar results for RecA, 16SrDNA, Hsp60, FtsZ
(2000) MBE 17, 835-838
Lento plot, ex Kerryn

* * * * * * * * * * * + + + + + + + +

support clashes
overview (contd)

trees and networks (with cycles)
spanning and phylogenetic (Steiner) trees
exploratory data analysis, visualization
multiscale
splits
compatibility/incompatibility
application - the MinMax squeeze
when MP and ML are equivalent

- for a single column of data (this leads to),
- there is no common mechanism shared by the characters,
- maximum evolutionary pathway likelihood (ML_{ep}) is equivalent to maximum parsimony,
- there are effectively infinite character states (the same mutation never occurs twice)

- abundant sequences?
overview

trees and networks (with cycles)
spanning and phylogenetic (Steiner) trees
exploratory data analysis, visualization
multiscale
splits
compatibility/incompatibility
application - the MinMax squeeze
splits

- from trees and networks
- from data - sequences (character states)
- from data - distances
what is a split - from trees
what is a split - from trees

\[ \{a,b\} \quad \{c,d,e\} \]

\[ A \mid B. \]
splits in a network

\{ a, b, c \} \{ d, e \}

\{ a, b, e \} \{ c, d \}
summary of introduction

trees and networks (with cycles)
spanning and phylogenetic (Steiner) trees
exploratory data analysis, visualization
multiscale – generations to macroevolution
splits – from trees, characters or distances
compatibility/incompatibility
application - the MinMax squeeze