NETGEN
A Simulator for Creating Phylogenetic Networks

Monique M. Morin

UNM Computer Science
www.compbio.unm.edu
February 2006
Outline

- Motivation for network generators
- **NETGEN** simulator
  - specific attributes of code
  - treatment of diploid hybrids
- Features for network customization
- **NETGEN** simulator characterization
- Network visualization options
- Conclusion
Phylogenetic research aims to reconstruct the past and predict the future … accurately.

- Traditional phylogenies are trees.
  
  While applicable in some cases, non-tree events are known to occur:
  
  - hybridization
  - lateral gene transfer

- These events transform trees into networks.
  
  Only through understanding networks can the phylogenetic picture be complete.

Tools are needed to generate, compare, and reconstruct phylogenetic networks.
By comparing generated and reconstructed networks, confidence in reconstruction algorithms is achieved.

**Simulator Generated Phylogenetic Network**

**Proposed Phylogenetic Reconstruction**

**Network Comparison Techniques** (e.g. Extended Robinson-Foulds¹)

NETGEN allows the generation of both phylogenetic trees and networks in a controlled manner.

NETGEN is a variable rate, event-driven phylogenetic network simulator with simultaneous sequences.

Adaptation of traditional birth/death model

- birth-death-diploid hybrid model (inputted rate for each event type)

Sequences are evolved simultaneously via Seq-Gen$^2$

- enables sequence-dependent hybridization

Dual tracking of evolutionary and clock branch lengths

- evolutionary – required for evolving sequences
- clock – used for event queue

Event queue driven

- permits easy expansion of model to other events

Input/Output format provides flexibility

- Newick compatible output for trees, modified Newick for networks

To implement diploid hybridizations, sequences are evolved simultaneously with the topology.

Hybrid Event Process:
- Update all active lineages
  - branch lengths
  - sequences (via Seq-Gen)
- Find minimum hamming distance
- Compare with user specified threshold
  - (YES) 2nd parent found
    - cancel prior 2nd parent event
  - (No) no hybrid made
- Define appropriate subsequent events

NETGEN tracks intermediate sequences to determine existence of second hybrid parent based on hamming criteria.
NETGEN options add flexibility which facilitates the generation of customized phylogenies.

Optional features:

- variable birth, death, and hybrid rates
differing rates for each lineage

- ultrametric or non-ultrametric evolutionary times

- pseudo-random or specified root sequence(s)

- arbitrary chromosome and chromatid numbers
  up to memory/processing limitations

- optional maximum hybrid number limit

- regulated outgroup generation
Phylogenetic trees can be rooted with an outgroup, while phylogenetic networks are rooted by definition.

Outgroups, often created with expert knowledge, are used to root trees.

In NETGEN, outgroup sequences are generated last and according to user specified bounds.

User specifies:
- similarity range
  - hamming distance
  - or percentage
- maximum attempts

Outgroups add topological insight useful in phylogenetic reconstruction.
Simulator tests with various birth/death rates confirm correct population growth.

Population $m$ grows as: $m(t) = n_0 e^{(B-D)t}$
NETGEN simulations show hybridizations have a similar (but different) effect on growth as do births.

Without hybrids, we have exponential growth: \( n_0 e^{(B-D)t} \)

However, cancellations alter the \( B+H-D \) expected slope.
NETGEN offers Newick and SplitsTree³ compatible output for phylogeny visualization.

Original Newick Format (Trees Only)  

Modified Newick Format (Hybrid Inclusive)


A visualization tool for modified Newick is still needed.  

In the absence of hybrids, the modified Newick format reverts to original Newick.

**NETGEN** is a variable rate, event-driven phylogenetic network simulator with simultaneous sequences.

Key features:

- sequence-dependent diploid hybridizations
- optional variable birth, death, and hybrid rates
- event-driven structure allows easy expansion: mass extinction, lateral gene transfer, etc.
- outgroup option for root perspective
- Newick compatible output format

Software and documentation available for download at:

```
http://www.phylo.unm.edu/~morin/
```
Acknowledgements

For discussions, guidance, and advice, my advisor:

- Bernard Moret

For brainstorming sessions:

- Randy Linder
- Luay Nakhleh

This work has been supported in part by National Science Foundation grants IIS 01-21377, DEB 01-20709, and EF 03-31654.

Software and documentation available for download at:

http://www.phylo.unm.edu/~morin/
Auxiliary Slides
Modified Newick Format works for networks with multiple hybrid nodes.
Outgroup sequences are generated last and according to user specified bounds.

Outgroups add topological insight useful in phylogenetic reconstruction.

Outgroup Generation Process:
- establish “sub-root”
- execute simulation
- generate outgroup sequence
- check sequence against user specified criteria
- repeat sequence generation until:
  a) suitable sequence found
  b) maximum attempts made

User specifies:
- similarity range
  hamming distance
  or percentage
- maximum attempts