ContextPhylogenies...Distance treesParsimonLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...000000000000000000000000000000000000

A (short) introduction to phylogenetics

Thibaut Jombart, Marie-Pauline Beugin

MRC Centre for Outbreak Analysis and Modelling Imperial College London

Genetic data analysis with PR~Statistics, Millport Field Station 16 Aug 2016
 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 0000
 000
 0000
 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 0

Outline

Context

- Phylogenies...
- **Distance trees**
- Parsimony
- Likelihood/Bayesian
- Uncertainty
- Pitfalls & best practices
- And more...



Phylogenetics: from the origins...



'From the first growth of the tree, many a limb and branch has decayed and dropped off; and these fallen branches of various sizes may represent those whole orders, families, and genera which have now no living representatives, and which are known to us only in a fossil state.'

C. Darwin, Notebook, 1837.



Phylogenetics: ...to the present



Bininda-Emonds *et al.*, 2007, Nature.

- phylogenetic trees are part of the standard toolbox of genetic data analysis
- represent the evolutionary history of a set of (sampled) taxa

Context 0000

Phylogenies... Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

And the main difference is...





Current trees look better!

(and some other minor differences)

Context 0000

Phylogenies... Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

And the main difference is...





Current trees look better!

(and some other minor differences)

Context 0000

Phylogenies... Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

And the main difference is...





Current trees look better!

(and some other minor differences)



About the minor differences...

- DNA sequencing revolution
- huge data banks freely available (e.g. GenBank)
- easier, cheaper, faster to obtain DNA sequences
- increasing number of full genomes available



Different ways to exploit this information.



About the minor differences...

- DNA sequencing revolution
- huge data banks freely available (e.g. GenBank)
- easier, cheaper, faster to obtain DNA sequences
- increasing number of full genomes available



Different ways to exploit this information.

 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

Substitution: replacement of a nucleotide (e.g. a \rightarrow t)

Phylogenies...

0000



Pitfalls & best practices And more ...

Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Substitution: replacement of a nucleotide (e.g. $a \rightarrow t$)

Phylogenies...



Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Substitution: replacement of a nucleotide (e.g. $a \rightarrow t$)

Phylogenies...



Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Substitution: replacement of a nucleotide (e.g. $a \rightarrow t$)

Phylogenies...



Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Substitution: replacement of a nucleotide (e.g. $a \rightarrow t$)

Phylogenies...



Substitution: replacement of a nucleotide (e.g. a \rightarrow t)

Phylogenies...

•••••



Substitution patterns reflect the evolutionary history

Pitfalls & best practices And more...

Using substitution patterns to reconstruct the evolutionary history

Pitfalls & best practices



Phylogenies...

Using substitution patterns to reconstruct the evolutionary history

Pitfalls & best practices

Phylogenies...



Using substitution patterns to reconstruct the evolutionary history

Pitfalls & best practices

Phylogenies...

0000



Phylogenetics aim to reconstruct evolutionary trees (*phylogenies*) from genetic sequence data.

Using trees to represent the evolutionary history

Distance trees Parsimony Likelihood/Bayesian Uncertainty



Phylogenies...

Pitfalls & best practices And more ...

Phylogenies... 0000

Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Using trees to represent the evolutionary history



analysed sequences

- "leaves"
- "Operational Taxonomic Units (OTUs)"

Using trees to represent the evolutionary history

Phylogenies...

0000

Most Recent Common Ancestors (MRCA)



Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Using trees to represent the evolutionary history



- "edge"

Phylogenies...

- length = amount of evolution (**not time**, as a rule)
- length is optional

Pitfalls & best practices And more...

 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0
 0000
 000
 000
 0000
 0000
 0000
 0000

Using trees to represent the evolutionary history



distances between tips

- "patristic" distance: sum of branch lengths
- other measures of distance/dissimilarity
- vertical axis meaningless

Using trees to represent the evolutionary history

Phylogenies...

0000



Pitfalls & best practices And more...



How to we build them?



Prepare data

• align sequences: alignment software + manual refinement

Build the tree

- distance-based methods
- maximum parsimony
- likelihood-based methods (ML, Bayesian)

- assess uncertainty
- test phylogenetic signal
- model trait evolution



Prepare data

• align sequences: alignment software + manual refinement

Build the tree

- distance-based methods
- maximum parsimony
- likelihood-based methods (ML, Bayesian)

- assess uncertainty
- test phylogenetic signal
- model trait evolution



Prepare data

• align sequences: alignment software + manual refinement

Build the tree

- distance-based methods
- maximum parsimony
- likelihood-based methods (ML, Bayesian)

- assess uncertainty
- test phylogenetic signal
- model trait evolution



Prepare data

• align sequences: alignment software + manual refinement

Build the tree

- distance-based methods
- maximum parsimony
- likelihood-based methods (ML, Bayesian)

- assess uncertainty
- test phylogenetic signal
- model trait evolution

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000</

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...



Distance-based phylogenetic reconstruction

Approaches relying on **agglomerative clustering** algorithms (e.g. Single linkage, UPGMA, Neighbor-Joining)

Rationale

- 1. compute pairwise genetic distances \mathbf{D}
- 2. group closest sequences
- 3. update \mathbf{D}
- 4. go back to 2) until all sequences are grouped

ContextPhylogenies...Distance treesParsimonyLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...000000000000000000000000000

Distance-based phylogenetic reconstruction



Context
0000Phylogenies...Distance trees
000Parsimony
000Likelihood/Bayesian
000Uncertainty
0000Pitfalls & best practices
0000And more...
000

Distance-based phylogenetic reconstruction



Context
0000Phylogenies...Distance trees
000Parsimony
000Likelihood/Bayesian
000Uncertainty
0000Pitfalls & best practices
0000And more...
000

Distance-based phylogenetic reconstruction




ContextPhylogenies...Distance treesParsimonLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...000000000000000000000000000000000000

Distance-based phylogenetic reconstruction





ContextPhylogenies...Distance treesParsimonLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...000000000000000000000000000000000000

Distance-based phylogenetic reconstruction





Context
0000Phylogenies...Distance trees
000Parsimon
000Likelihood/Bayesian
000Uncertainty
0000Pitfalls & best practices
0000And more...
000

Distance-based phylogenetic reconstruction



Hierarchical clustering:

0000

Distance trees



- single linkage: $D_{k,g} = \min(D_{k,i}, D_{k,j})$
- complete linkage: $D_{k,g} = \max(D_{k,i}, D_{k,j})$

Pitfalls & best practices And more...

• UPGMA:
$$D_{k,g} = \frac{D_{k,i} + D_{k,j}}{2}$$

Neighbor joining:

Hierarchical clustering:

0000

Distance trees



- single linkage: $D_{k,g} = \min(D_{k,i}, D_{k,j})$
- complete linkage: $D_{k,g} = \max(D_{k,i}, D_{k,j})$

Pitfalls & best practices And more...

• UPGMA:
$$D_{k,g} = \frac{D_{k,i} + D_{k,j}}{2}$$

Neighbor joining:

Hierarchical clustering:

0000

Distance trees



- single linkage: $D_{k,g} = \min(D_{k,i}, D_{k,j})$
- complete linkage: $D_{k,g} = \max(D_{k,i}, D_{k,j})$

Pitfalls & best practices And more...

• UPGMA:
$$D_{k,g} = \frac{D_{k,i} + D_{k,j}}{2}$$

Neighbor joining:

1

Hierarchical clustering:

0000

Distance trees



• single linkage:
$$D_{k,g} = \min(D_{k,i}, D_{k,j})$$

• complete linkage: $D_{k,g} = \max(D_{k,i}, D_{k,j})$

Pitfalls & best practices And more...

• UPGMA:
$$D_{k,g} = \frac{D_{k,i} + D_{k,j}}{2}$$

Neighbor joining:

1

Hierarchical clustering:

0000

Distance trees



• single linkage:
$$D_{k,g} = \min(D_{k,i}, D_{k,j})$$

• complete linkage: $D_{k,g} = \max(D_{k,i}, D_{k,j})$

Pitfalls & best practices

• UPGMA:
$$D_{k,g} = \frac{D_{k,i} + D_{k,j}}{2}$$

Neighbor joining:

Distance-based phylogenetic reconstruction

Advantages

- simple
- flexible (many distances and clustering algorithms)
- fast and scalable (applicable to large datasets)

- sensitive to distance/clustering chosen
- evolutionary rates are not estimated
- no measure of uncertainty for the tree obtained

Distance-based phylogenetic reconstruction

Advantages

- simple
- flexible (many distances and clustering algorithms)
- fast and scalable (applicable to large datasets)

- sensitive to distance/clustering chosen
- evolutionary rates are not estimated
- no measure of uncertainty for the tree obtained

ContextPhylogenies...Distance treesParsimonyLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...0000000000000000000000000000000000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

Approaches relying on finding the tree with the smallest number of character changes (substitutions)

Rationale

1. start from a pre-defined tree

Distance trees Parsimony

- 2. compute initial parsimony score
- 3. permute branches and compute parsimony score
- 4. accept new tree if the parsimony score is improved
- 5. go back to 3) until convergence

Pitfalls & best practices And more ...

Context
0000Phylogenies...
0000Distance trees
0000Parsimon
000Likelihood/Bayesian
000Uncertainty
0000Pitfalls & best practices
0000And more...
000

Maximum parsimony phylogenies





Initial tree

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000</

Maximum parsimony phylogenies



 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000

Maximum parsimony phylogenies



 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000</

Maximum parsimony phylogenies



score: 5

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000</

Maximum parsimony phylogenies



Maximum parsimony phylogenies

Advantages

- applicable to any discontinuous characters (not just DNA)
- intuitive explanation: 'simplest' evolutionary scenario

- evolutionary rates are not estimated
- no measure of uncertainty for the tree obtained
- computer-intensive
- different types of substitutions ignored
- evolution not necessarily parsimonious
- sensitive to heterogeneous rates of evolution (*long branch attraction*)

Maximum parsimony phylogenies

Advantages

- applicable to any discontinuous characters (not just DNA)
- intuitive explanation: 'simplest' evolutionary scenario

- evolutionary rates are not estimated
- no measure of uncertainty for the tree obtained
- computer-intensive
- different types of substitutions ignored
- evolution not necessarily parsimonious
- sensitive to heterogeneous rates of evolution (*long branch attraction*)

ContextPhylogenies...Distance treesParsimonyLikelihood/BayesianUncertaintyPitfalls & best practicesAnd more...000000000000000000000000000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000

Likelihood-based phylogenies (ML / Bayesian)

Approaches relying on a model of sequence evolution:

- ML: find tree and evolutionary rates with highest likelihood
- **Bayesian**: find tree and evolutionary rates to posterior probability

Rationale

- 1. start from a pre-defined tree
- 2. compute initial likelihood/posterior
- 3. permute branches, sample new parameters and compute likelihood/posterior
- 4. accept new tree and parameters based on likelihood/posterior improvement
- 5. go back to 3) until convergence

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000
 0000

Likelihood-based phylogenies (ML / Bayesian)

Approaches relying on a model of sequence evolution:

- ML: find tree and evolutionary rates with highest likelihood
- **Bayesian**: find tree and evolutionary rates to posterior probability

Rationale

- 1. start from a pre-defined tree
- 2. compute initial likelihood/posterior
- 3. permute branches, sample new parameters and compute likelihood/posterior
- 4. accept new tree and parameters based on likelihood/posterior improvement
- 5. go back to 3) until convergence



Likelihood-based phylogenies (ML / Bayesian)



 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000

Likelihood-based phylogenies (ML / Bayesian)

Advantages

- very flexible
- consistent with a model of evolution
- statistically consistent (model comparison)
- parameter estimation
- (Bayesian) several trees \rightarrow measure of uncertainty

- computer-intensive
- choice of the model of evolution
- (ML) no measure of uncertainty for the tree obtained
- (Bayesian) need to find a consensus tree

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000

Likelihood-based phylogenies (ML / Bayesian)

Advantages

- very flexible
- consistent with a model of evolution
- statistically consistent (model comparison)
- parameter estimation
- (Bayesian) several trees ightarrow measure of uncertainty

- computer-intensive
- choice of the model of evolution
- (ML) no measure of uncertainty for the tree obtained
- (Bayesian) need to find a consensus tree

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 0000
 000
 0000
 0000
 0000
 0000
 0000
 0000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

Pitfalls & best practices And more...

How do we know the tree is robust?

Main issue: assess the uncertainty of the tree topology / individual nodes

- ML: model selection to compare trees (whole tree)
- Bayesian methods: between-samples variability (individual nodes)
- any method: bootstrap (individual nodes)



Pitfalls & best practices And more ...

How do we know the tree is robust?

Main issue: assess the uncertainty of the tree topology / individual nodes

- ML: model selection to compare trees (whole tree)
- Bayesian methods: between-samples variability (individual nodes)
- any method: bootstrap (individual nodes)



Pitfalls & best practices And more ...

How do we know the tree is robust?

Main issue: assess the uncertainty of the tree topology / individual nodes

- ML: model selection to compare trees (whole tree)
- Bayesian methods: between-samples variability (individual nodes)
- any method: • bootstrap (individual nodes)



Pitfalls & best practices And more ...

How do we know the tree is robust?

Main issue: assess the uncertainty of the tree topology / individual nodes

- ML: model selection to compare trees (whole tree)
- Bayesian methods: between-samples variability (individual nodes)
- any method: bootstrap (individual nodes)



Pitfalls & best practices And more ...

How do we know the tree is robust?

Main issue: assess the uncertainty of the tree topology / individual nodes

- ML: model selection to compare trees (whole tree)
- Bayesian methods: between-samples variability (individual nodes)
- any method: bootstrap (individual nodes)





- assess variability due to sampling the genome and conflicting signals
- relies on analysing resampled datasets

Rationale

- 1. obtain a reference tree
- 2. resample the sites with replacement
- 3. obtain a tree for the resampled dataset
- 4. go back to 2) until the desired number of bootstrapped trees is attained
- 5. compute the frequency of each bifurcation of the reference tree occuring in bootstrapped trees



- assess variability due to sampling the genome and conflicting signals
- relies on analysing resampled datasets

Rationale

- 1. obtain a reference tree
- 2. resample the sites with replacement
- 3. obtain a tree for the resampled dataset
- 4. go back to 2) until the desired number of bootstrapped trees is attained
- 5. compute the frequency of each bifurcation of the reference tree occuring in bootstrapped trees







Advantages

- standard
- simple to implement

- possibly computer-intensive
- assumes that the genome has been sampled randomly (often wrong)



Advantages

- standard
- simple to implement

- possibly computer-intensive
- assumes that the genome has been sampled randomly (often wrong)
Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000
 000</

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more...

Plotting trees as rooted



Never plot an unrooted tree as rooted.

Context
0000Phylogenies...Distance trees
0000Parsimony
000Likelihood/Bayesian
000Uncertainty
0000Pitfalls & best practices
0000And more...
000

Plotting trees as rooted



Never plot an unrooted tree as rooted.



Interpreting distances





Interpreting distances





Interpreting distances





The paradox of divergent clusters



MRCA and genetic distances may give different information.



The paradox of divergent clusters



MRCA and genetic distances may give different information.



The paradox of divergent clusters



MRCA and genetic distances may give different information.

Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more... 00000

Taking uncertainty into account

193 HIV-1 sequences from DRC (Strimmer & Pvbus 2001)



At best, the tree is an estimate of the likely evolutionary history of the taxa studied.

193 HIV-1 sequences from DRC

00000

Likelihood/Bayesian Uncertainty Pitfalls & best practices

Taking uncertainty into account



Collapsed tree (threshold length 0.01)

At best, the tree is an estimate of the likely evolutionary history of the taxa studied.

193 HIV-1 sequences from DRC

00000

Pitfalls & best practices

Taking uncertainty into account



Collapsed tree (threshold length 0.01)

At best, the tree is an estimate of the likely evolutionary history of the taxa studied.

(Over, Mis)Interpreting temporal trends



"Time trees" only make sense under a near-perfect molecular clock.

(Over, Mis)Interpreting temporal trends



"Time trees" only make sense under a near-perfect molecular clock.

(Over, Mis)Interpreting temporal trends



"Time trees" only make sense under a near-perfect molecular clock.

 Context
 Phylogenies...
 Distance trees
 Parsimony
 Likelihood/Bayesian
 Uncertainty
 Pitfalls & best practices
 And more...

 0000
 0000
 000
 000
 000
 0000
 0000
 000
 000

Outline

Context

Phylogenies...

Distance trees

Parsimony

Likelihood/Bayesian

Uncertainty

Pitfalls & best practices

And more ...

This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ...



Phylogenies... Distance trees Parsimon Likelihood/Bayesian Uncertainty Pitfalls & best practices And more... 0000 0000 000 0000

This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ...





This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ...





This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ..



Phylogenies... Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more... 0000 0000 000 000 0000

This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ..



Phylogenies... Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more... 0000 0000 000 00000 0000 0000

This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ..





This is only the beginning

Many things can be done with trees

- estimate divergence time
- model trait evolution (phylogenetic comparative method)
- reconstruct ancestral states
- measure diversity
- infer past demographics/effective population size (coalescence)

• ...



Distance trees Parsimony Likelihood/Bayesian Uncertainty Pitfalls & best practices And more...

Time to get your hands dirty!



The pdf of the practical is online:

```
http://adegenet.r-forge.r-project.org/
```

or

 $Google \rightarrow adegenet \rightarrow documents \rightarrow "GDAR August 2016"$

0.