Why cite?

- Citations reflect *the careful and thorough work you have put in* to locating and exploring your sources.
- Citations *help readers understand the context* of your argument, and locate your work within other conversations on your topic.
- Citations allow you to *acknowledge those authors who made possible particular aspects of your work*. Failure to provide adequate citations constitutes plagiarism.
- Citations, by delineating your intellectual debts, also *draw attention to the originality and legitimacy of your own ideas.*

http://www.dartmouth.edu/~sources/about/what.html

When to cite?

- Cite sources for all verbatim quotations of two or more consecutive words.
- Cite sources from which you paraphrase or summarize facts or ideas.
- Cite sources for ideas or information that could be regarded as common knowledge but which you think your reader might still find unfamiliar.

http://www.dartmouth.edu/~sources/about/what.html

What to cite?

Primarily: *refereed, archival* materials. (Archival materials are materials that are available in libraries or bookstores, have an ISBN number, etc.)

- Books
- Journal articles
- Refereed conference proceedings

Avoid:

- Websites, news stories, photocopied workshop handouts, personal communications.

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

Students paper:

The majority of the animal images in the play are Iago's, and all of these are contemptuous or repellent. He refers to a plague of flies, a quarrelsome dog, bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, goats and monkeys. Through these images the general sense of pain and unpleasantness is increased and kept constantly before us.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

Not OK:

Students paper:

The majority of the animal images in the play are Iago's, and all of these are contemptuous or repellent. He refers to a plague of flies, a quarrelsome dog, bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, goats and monkeys. Through these images the general sense of pain and unpleasantness is increased and kept constantly before us.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

Not OK:

Students paper:

I believe that the main image in Shakespeare's tragedy, *Othello,* is that of animals. These creatures are constantly in action, preying upon one another, and they are depicted as mischievous, wanton, cruel or suffering. By Shakespeare's ingenious use of these animal images, the general sense of pain and unpleasantness that pervades the entire story is much increased and kept constantly before the reader.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

Not OK:

Students paper:

In *Othello,* Shakespeare makes frequent use of animal imagery. The specific images he uses are generally distasteful and convey to the reader a constant impression of conflict and misery.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

OK:

Students paper:

In the play, *Othello*, the character of Iago is associated with unpleasant animal imagery[1]....

1. Caroline F. E. Spurgeon, Shakespeare's Imagery (Cambridge: Cambridge UP, 1935) 335.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

Not OK:

Students paper:

The majority of "the animal images in the play are Iago's, and all of these are contemptuous or repellent". He refers to "a plague of flies, a quarrelsome dog," "bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, goats and monkeys." "Through these" images "the general sense of pain and unpleasantness is increased and kept constantly before us."[1]

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

http://www.dartmouth.edu/~sources/about/w

The original text:

The main image in *Othello* is that of animals in action, preying upon one another, mischievous, lascivious, cruel or suffering, and through these, the general sense of pain and unpleasantness is much increased and kept constantly before us.

More than half the animal images in the play are Iago's, and all these are contemptuous or repellent: a plague of flies, a quarrelsome dog, the recurrent image of bird-snaring, leading asses by the nose, a spider catching a fly, beating an offenceless dog, wild cats, wolves, goats and monkeys[1].

1. Caroline F. E. Spurgeon, *Shakespeare's Imagery* (Cambridge: Cambridge UP, 1935) 335.

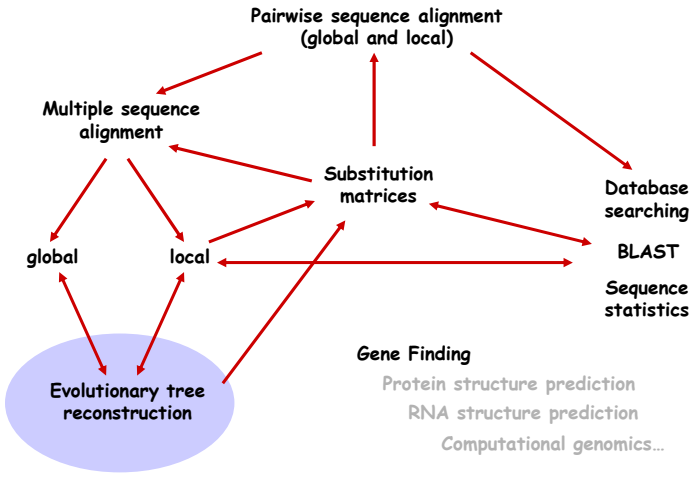
OK:

Students paper:

Caroline Spurgeon uses the words "contemptuous" and "repellent" in describing the animal imagery associated with Iago in Othello [1]. In my opinion, her choice of words indicates that...

1. Caroline F. E. Spurgeon, Shakespeare's Imagery (Cambridge: Cambridge UP, 1935) 335.

http://www.dartmouth.edu/~sources/about/w



Selecting data for tree reconstruction

- For reconstructing recent events, use DNA sequences
- For reconstructing distant events, use amino acid sequences
- Select sequences that
  - Are present in all taxa
  - Contain a conserved region
  - Exhibit variation within that region
  - e.g., Ribosomal (16sRNA) genes were used to reconstruct the tree of life. These genes encode products use in all organisms from bacteria to mammals.
- Pitfalls: duplicated genes, horizontal gene transfer, mosaic genes.

Finding the optimal tree

Given *k* taxa,

- Consider all trees with *k* leaves
- ➢Score each tree with respect to chosen evolutionary model.
- Select highest scoring tree(s)

Criteria for evaluating which tree best fits the data:

- ➢Maximum parsimony (character data)
- Minimum evolution (distance data)
- Maximum Likelihood (character data)

Finding the most parsimonious tree

Given *k* taxa and *n* characters (columns),

For each tree with *k* leaves

score(t) = 0

For each character, c /* 1≤c≤n */

Fitch's algorithm

- Find the optimal labeling of internal nodes
- *score(t)* = *score(t)* + *count_mutations(c)*

Return the tree(s) with minimum score.

Some problems with parsimony

- Not all characters are informative
- Data may not be parsimonious
- There may be more than one parsimonious tree

Number of unrooted trees for *k* taxa

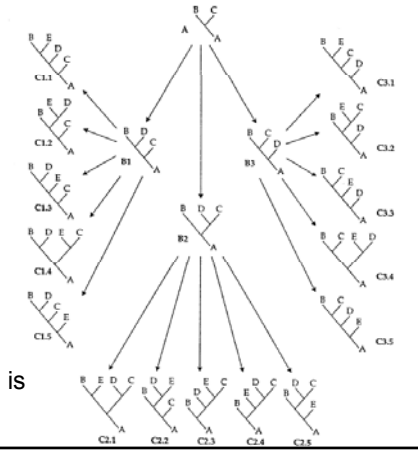
$$E(k) = E(k-1) + 2 = 2k - 3$$

The number of trees gets big fast

Number of leaves	Number of unrooted binary trees
3	**1**
4	**3**
5	**15**
6	**105**
10	**2,027,025**
20	**2.2×10^{20}**
50	**2.8×10^{74}**
500	**1×10^{1074}**

How do you find the optimal tree?

1. Exhaustive search (<12 taxa)



(Phylogeny reconstruction is NP-complete.)

How do you find the optimal tree?

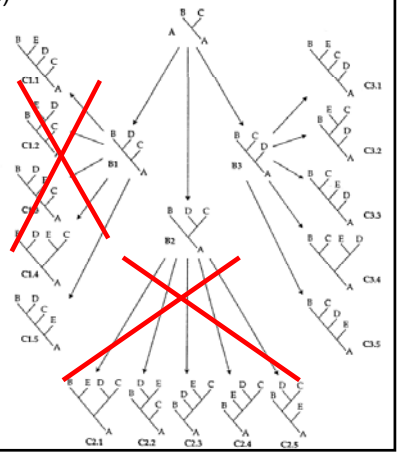
Method	Result	Time	Typical k
Exhaustive search	Optimal solution	*T(k)*	12

How do you find the optimal tree?

2. Branch-and-bound (<18 taxa)

Score is non-decreasing as you add edges

L = {T_3}, *C = infinity*
For i = 3 to k {
  For each tree, *t,* in L{
    If *Score(t) > C, skip*
    If *Score(t) < C, C = Score(t).*
    For every edge, *e*, in *t* {
      t'= t plus a new edge at *e*
      NewL = NewL U {t'}
    }
  }
  L = NewL
}



How do you find the optimal tree?

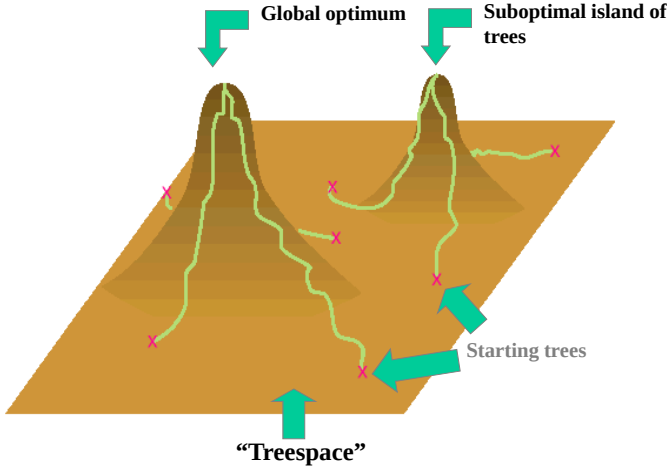
Method	Result	Time	Typical k
Exhaustive search	Optimal solution	*T(k)*	12
Branch and bound	Optimal solution	≤*T(k)*	18

How do you find a pretty good tree?

3. Heuristic search

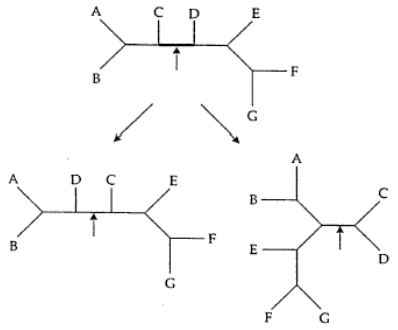
   Search for optimal trees by finding good trees and then rearranging them in the hopes of finding an even better tree

Heuristic search



Branch swapping

Nearest-neighbor interchange (NNI)



How do you find the optimal tree?

Method	Result	Time	Typical k
Exhaustive search	Optimal solution	*T(k)*	12
Branch and bound	Optimal solution	*≤T(k)*	18
Heuristic search	Suboptimal solution	*You choose*	*You choose*

Finding the optimal tree

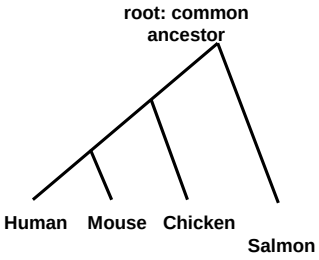
Given *k* taxa,

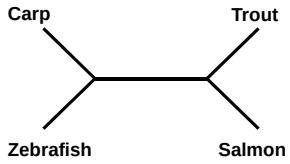
- Consider all trees with *k* leaves
  - ➢Score each tree with respect to chosen evolutionary model.
- Select highest scoring tree(s)

Criteria for evaluating which tree best fits the data:

- Maximum parsimony (character data)
- ➢Minimum evolution (distance data)
- Maximum Likelihood (character data)

Rooted vs unrooted trees:





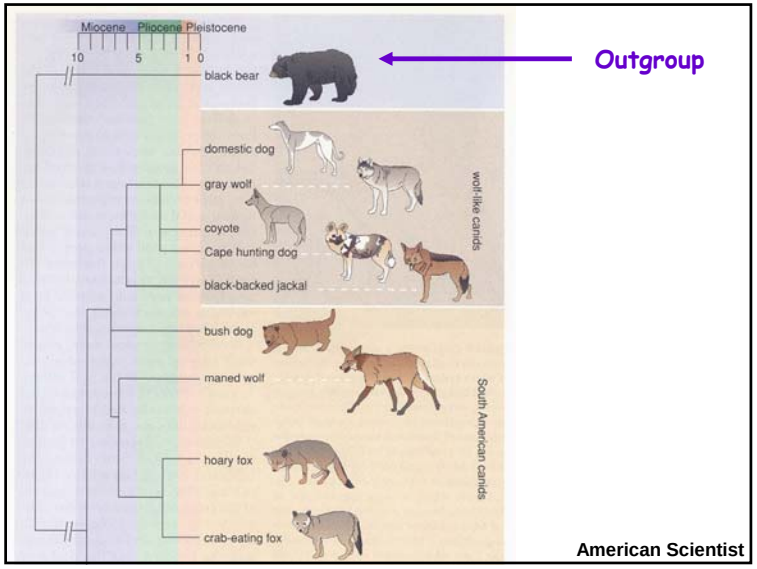
Unroooted trees give no information about the order of speciation events

Unrooted vs. Rooted Trees

- An ***unrooted*** tree gives information about the relationships between taxa.
- A ***rooted*** gene tree gives information about the order of events.
- Most tree reconstruction programs output unrooted trees.

How to root an unrooted tree

- If the mutation rate is constant in all lineages, the distance from root to leaf is the same for all leaves.
  - → We can obtain a rooted tree algorithmically.
  - → This is ***midpoint rooting***.
- Otherwise, use an ***outgroup***, a taxon that is distantly related to all other leaf taxa.



Finding the optimal tree

Given *k* taxa,

– Consider all trees with *k* leaves

➢Score each tree with respect to chosen evolutionary model.

– Select highest scoring tree(s)

Criteria for evaluating which tree best fits the data:

- Maximum parsimony (character data)
- Minimum evolution (distance data)
- Maximum Likelihood (character data)

Multiple Sequence Alignment as Character Data

Each column (or site) is one character.

```
~~~~ALTEKQEALLKQSWEVLKQNIPAHSRLFALIIEAA…
~~~MALTEKQEALLKQSWEVLKQNIPAHSRLFALILEAA…
~~~MALTERQEALLKQSWEVLKQNIPGHSRLFALIIEAA…
~~~~~~~~~~EALLKQSWEVLKQNIPGHSCLFALIIEAA…
```

	C1	C2	C3	C4
Bees	A	H	S	R
Moths	A	H	S	R
Ants	G	H	S	R
Centipedes	G	H	S	C

Multiple Sequence Alignment as Distance Data

```
Human   ~~~~ALTEKQEALLKQSWEVLKQNIPAHSLRLFALIIEAA…
Rabbit  ~~~MALTEKQEALLKQSWEVLKQNIPAHSLRLFALILEAA…
Pig     ~~~MALTERQEALLKQSWEVLKQNIPGHSLRLFALIIEAA…
Chicken ~~~~~~~~~~EALLKQSWEVLKQNIPGHSLCLFALIIEAA…
```

	Human	Rabbit	Pig	Chicken
Human	0	3	7	9
Rabbit		0	6	8
Pig			0	6
Chicken				0

Assumptions:
- Selection dominates
- Mutations are rare
- No multiple substitutions

Parsimony:

Character data

Find the tree that requires the fewest changes to explain the data

Assumptions:
- Neutral mutation dominates
- Multiple substitutions occur

Minimum Evolution:

Distance data

Find the tree that best fits the pairwise distances between taxa

Maximum Likelihood:

Character data

Find the most likely tree

Assumptions:
- Selection dominates
- Mutations are rare
- No multiple substitutions

Parsimony:

Character data

Find the tree that requires the fewest changes to explain the data

Assumptions:
- Neutral mutation dominates
- Multiple substitutions occur

Minimum Evolution:

Distance data

Find the tree that best fits the pairwise distances between taxa

Maximum Likelihood:

Character data

Find the most likely tree

Distance-based methods

- ➢ How to obtain a distance matrix
- Correcting for multiple substitutions
- Fitting distances to a tree
  - Conditions for obtaining an exact fit
    - Ultrametric distances
    - Additive distances
  - Minimum Evolution: finding the tree with the best fit
  - Greedy algorithms
    - UPGMA
    - NeighborJoining

How distance matrices are obtained

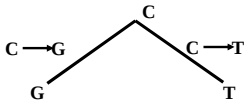
Given sequences from *k* taxa

- Construct a multiple sequence alignment
- Determine pairwise distance from each pair of taxa *using the MSA*
- Correct for multiple substitutions

Multiple Sequence Alignment

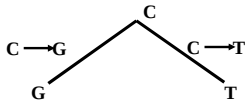
~~~~ALTEKQEALLKQSWEVLKQNIPAHSLRLFALIIEAA…
~~~~

Substitution patterns

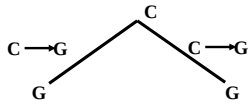


Coincidental substitution:

Substitution patterns



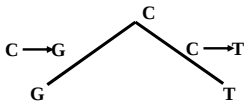
Coincidental substitution:
- 2 changes, 1 difference



Parallel substitution:

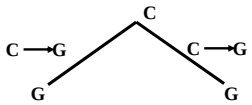
```
...G...
...T...
```

Substitution patterns



Coincidental substitution:
- 2 changes, 1 difference

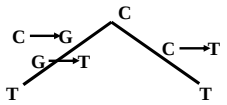
```
...G...
...T...
```



Parallel substitution:
- 2 changes, no difference

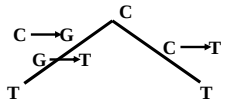
```
...G...
...G...
```

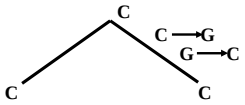
Substitution patterns



Convergent substitution:

Substitution patterns



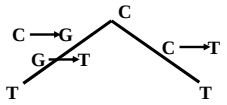


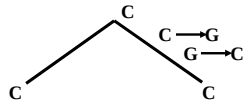
Convergent substitution:
- 3 changes, no difference

Back substitution:

...T...
...T...

Substitution patterns





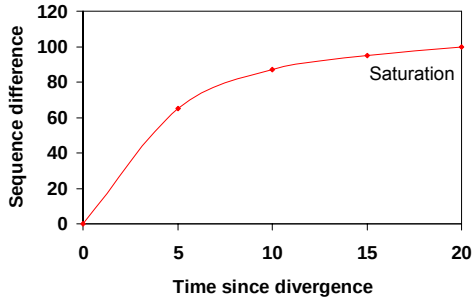
Convergent substitution:
- 3 changes, no difference

...T...
...T...

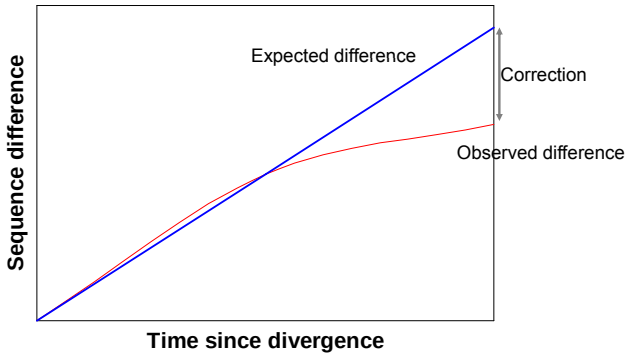
Back substitution:
- 2 changes, no difference

...C...
...C...

Multiple substitutions

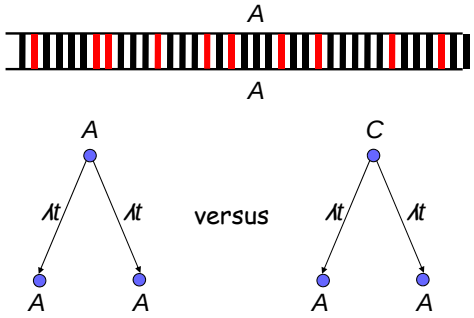


Correcting for multiple substitutions



Correcting for multiple substitutions

Given *m* mismatches in a PW alignment of length *n*, estimate the actual number of substitutions



Correcting for multiple substitutions

Given *m* mismatches in a PW alignment of length *n*, estimate the actual number of substitutions

Note that:

- *p' = m/n* is an estimator for the underlying probability of a mismatch, *p = P(mismatch)*
- The number of substitutions is *2λt*

Strategy

- Propose a Markov model of substitution
- Derive p=f(*λt*)
- $\lambda t = f^{-1}(p)$
- $E[subs/site] = 2 f^{-1}(p)$