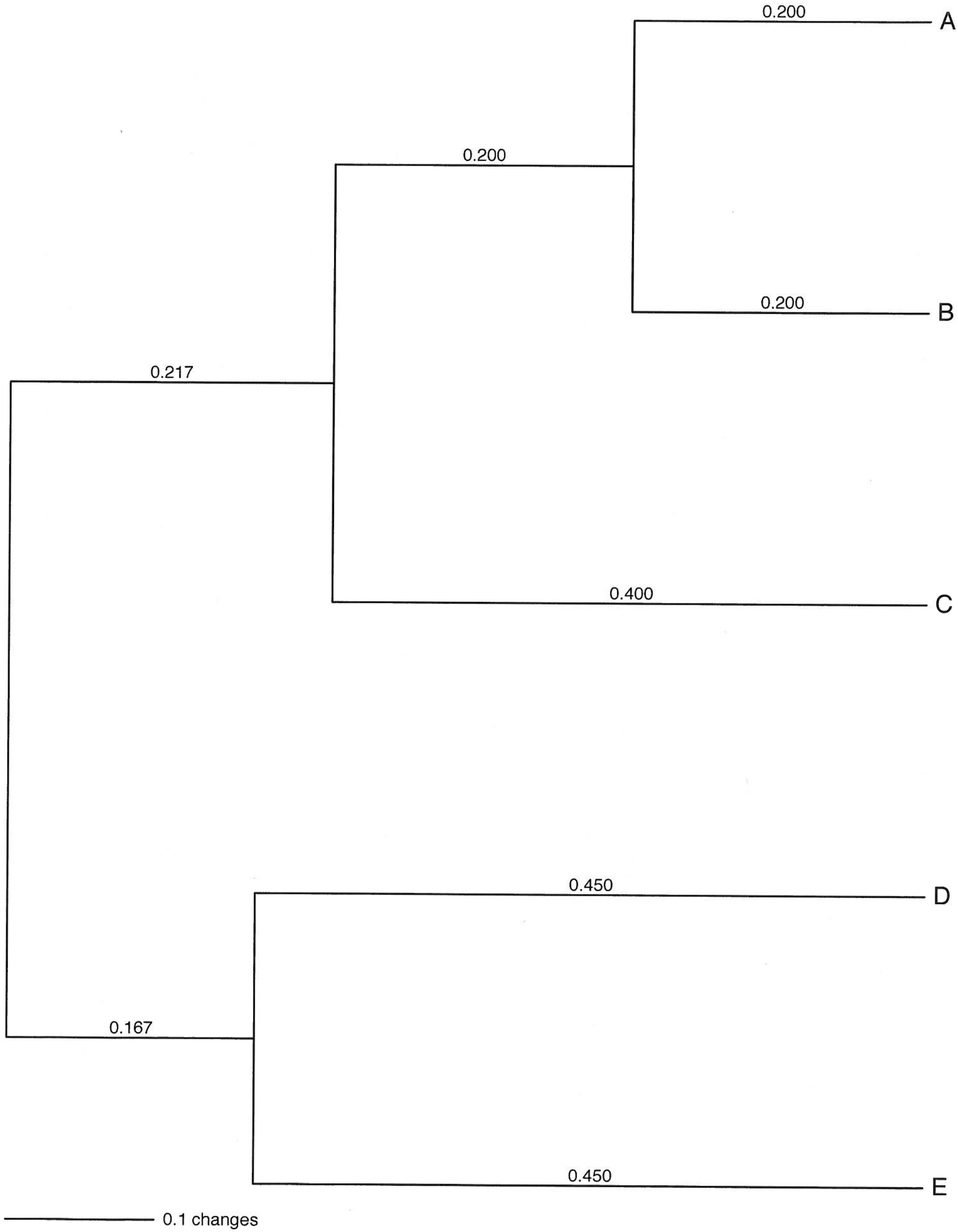
UPGMA tree



Conclusions:

UPGMA:

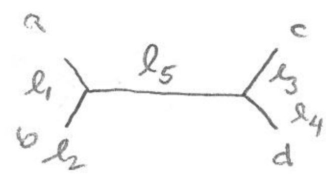
- constructs a rooted, ultrametric tree (binary)
- is <u>fast</u>! Instead of searching over $(2n-3)!!$ trees like parsimony, UPGMA quickly constructs a tree from dissimilarity data
- could be reasonable if one assumes/believes a molecular clock is at work.

Before continuing, review solving m equations in n unknowns...

Suppose the taxa under study are X, with $|X|=n$. Then any dissimilarity table has $\binom{n}{2} = \frac{n(n-1)}{2}$ dissimilarities, but an (unrooted) binary tree has only $2n-3$ edges with lengths l_i, $i=1,\dots,2n-3$

n	# pairwise diss.	# edge length l_i
2	1	1
3	3	3
4	6	5
	⋮	⋮
10	45	17

$\binom{n}{2} \gg 2n-3$ as $n \to \infty$



Viewing the l_i as unknowns we have

$n=10$: 45 equations in 17 unknowns

$n=4$: 5 equations in 6 unknowns

$\Rightarrow$ overdetermined system (more equations than unknowns)

$\rightsquigarrow$ likely inconsistent (i.e. <u>no</u> solution)

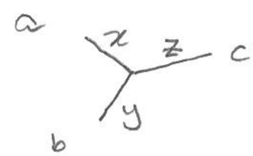
$\rightsquigarrow$ distance tables are usually $\underline{\underline{\text{not}}}$ from tree metrics.

However, when $n=3$ there are 3 equations in 3 unknowns and the system has a solution.

Eg.

	a	b	c
a		4	3
b			5

Single unrooted tree:
with $l_i = a, b, c$.



E1: $d(a,b) = 4 = x + y$

E2: $d(a,c) = 3 = x + z$

E3: $d(b,c) = 5 = y + z$

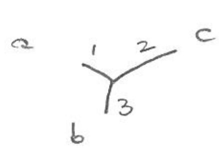
3 linear equations in 3 unknowns
$\rightsquigarrow$ consistent

Solve using linear algebra OR common sense.

$$x = \frac{d(a,b) + d(a,c) - d(b,c)}{2} = \frac{4+3-5}{2} = 1 \qquad \boxed{x=1}$$

$$y = d(a,b) - x = 4 - 1 = 3 \qquad \boxed{y=3}$$

$$z = d(a,c) - x = 3 - 1 = 2 \qquad \boxed{z=2}$$



We use this idea → 3 pairwise distances exactly fit a tree ←
to get a new distance method that does not produce ultrametric trees.