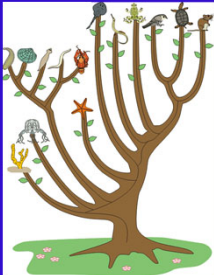


SCIENCE

Phylogenetic tree-building



Prof Peter O'Donoghue

1

Habits of mind:

Human nature to compartmentalize!



spend much of life comparing (similarities) and contrasting (differences)

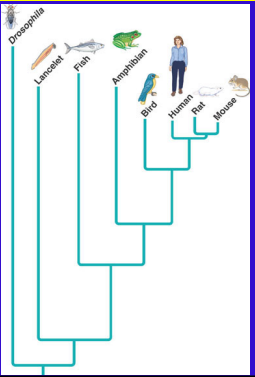
2

Exemplar: **Biological Classification**

Visualization of relationships through tree-building

Use algorithms to quantify similarities and differences

- entities related (taxonomy)
- related by descent (phylogeny)



3

Phylogenetics

Two approaches

Traditional

morphological data

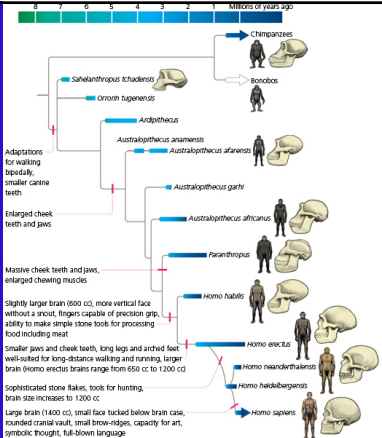
- appearance
- size

Contemporary

molecular data

- nucleotides
- amino acids

4



Adaptations for walking bipedally, smaller canine teeth

Enlarged cheek teeth and jaws

Massive cheek teeth and jaws, enlarged chewing muscles

Slightly larger brain (600 cc), more vertical face without a snout, fingers capable of precision grip, ability to make simple stone tools for processing food including meat

Smaller jaws and cheek teeth, long legs and arched feet: well-suited for long-distance walking and running, larger brain (Homo erectus brains range from 650 cc to 1200 cc)





Sophisticated stone flakes, tools for hunting, brain size increases to 1200 cc

Large brain (1400 cc), small face tucked below brain case, rounded cranial vault, small brow-ridge, capacity for art, symbolic thought, full-blown language

5

PHYLOGENY

Basic tree-building:

A	B	C	D
			

Select entities (A, B,)

Select characters (size,)

Score states (large/small,)

6

AVERAGE LINKAGE CLUSTER METHOD

	A	B	C	D
Gender	male	male	female	male
Complexion	dark	dark	light	dark
Height	short	tall	short	tall
Hirsuteness	hairy	hairy	hairy	bald

Matrix of character states

	A	B	C	D
A	0	1	2	2
B		0	3	1
C			0	4
D				0

4 x 4 matrix of fixed differences

Tree building

7

	A	B	C	D
A	0	1	2	2
B		0	3	1
C			0	4
D				0

Collapse to 3 x 3 matrix (average AB)

	AB	C	D
AB	0	$(2+3)/2=2.5$	$(2+1)/2=1.5$
C		0	4
D			0

Tree building

8

	AB	C	D
AB	0	2.5	1.5
C		0	4
D			0

Collapse to 2 x 2 matrix (average ABD)

	ABD	C
ABD	0	$(2.5+4)/2=3.25$
C		0

Tree building

9

Interpretation

A	B	C	D

Who are most similar?
 Who is most different?
 Who is most evolved?

Tree building

10

Do trees show evolutionary relationships?

Entity selection?	in-groups, out-groups, root
Character selection?	number, informative?
Character states?	ranking, scoring
Tree building?	multiple trees, consensus
Topography?	divergence, convergence

11

12

Group exercise: Pair-up

Let's build some trees.

Handouts available for:

- bacteria
- protozoa
- fungi
- beetles
- butterflies
- flowers
- birds
- fish

13

Identify characters - score states

	A	B	C	D
1.				
2.				
3.				
4.				
5.				

14

Complete distance matrices

	A	B	C	D
A	-			
B		-		
C			-	
D				-

	/		
/	-		
		-	
			-

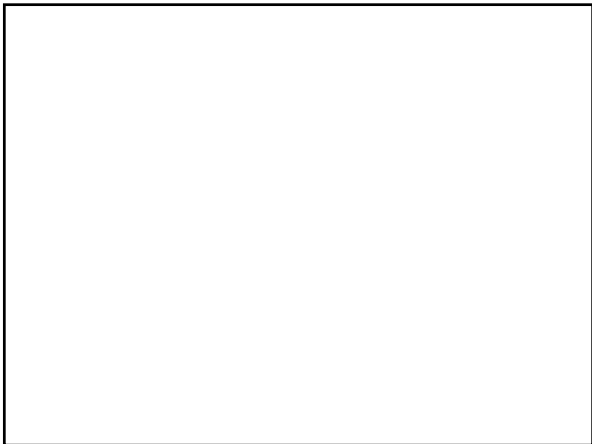
15

Build tree (phenogram)

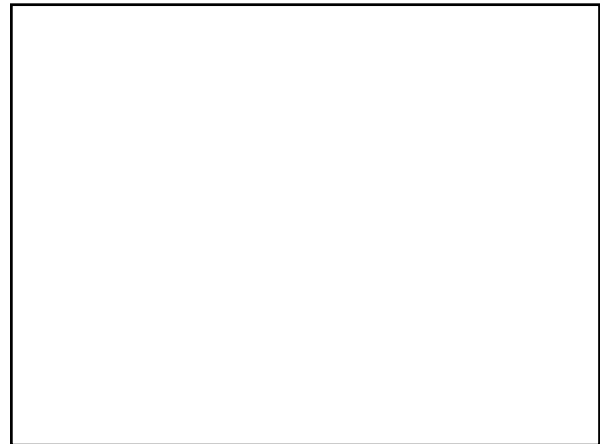
5 4 3 2 1 0

Number of fixed differences

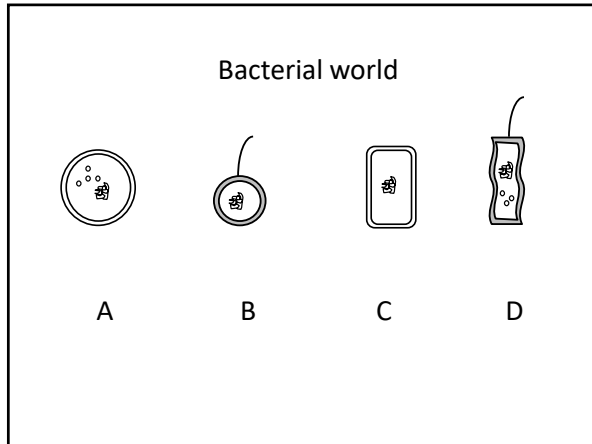
16



17



18



19

Character states

	A	B	C	D
1. Cell size (s/l)	large	small	large	small
2. Cell shape (r/s/w)	round	round	square	wavy
3. Cell wall (g/w)	white	grey	white	grey
4. Flagella (+/-)	absent	present	absent	present
5. Vacuoles (+/-)	present	absent	absent	present

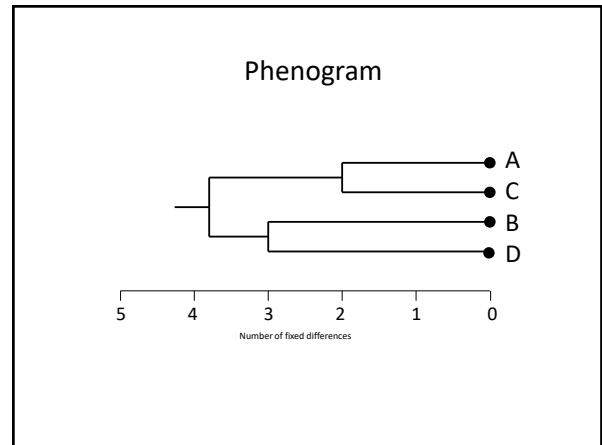
20

Distance matrix

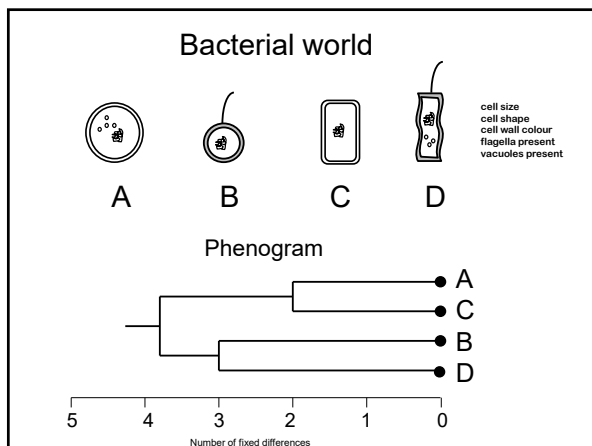
	A	B	C	D
A	-	4	2	3
B		-	4	3
C			-	4
D				-

	A/C	B	D
A/C	-	4	3.5
B		-	3
D			-

21



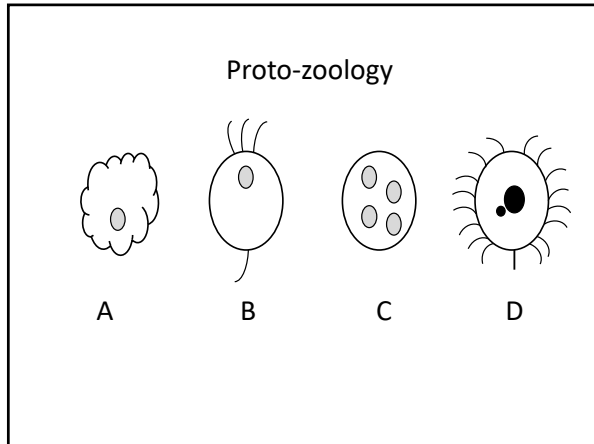
22



23



24



25

Character states

	A	B	C	D
1. Cell shape (p/r)	polymorphic	round	round	round
2. Flagella (+/-)	absent	present	absent	absent
3. Cilia (+/-)	absent	absent	absent	present
4. Nuclei (no.)	1	1	4	2
5. Nuclei (m/b)	mono	mono	mono	bi

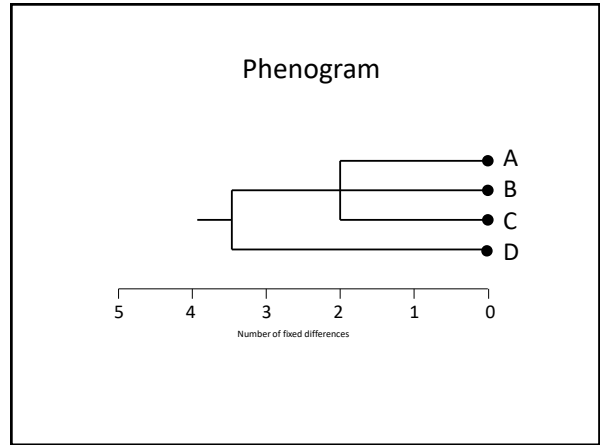
26

Distance matrix

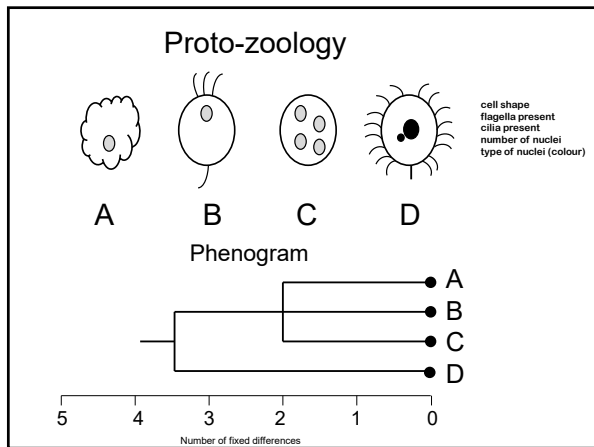
	A	B	C	D
A	-	2	2	4
B		-	2	4
C			-	3
D				-

	A/B	C	D
A/B	-	2	4
C		-	3
D			-

27



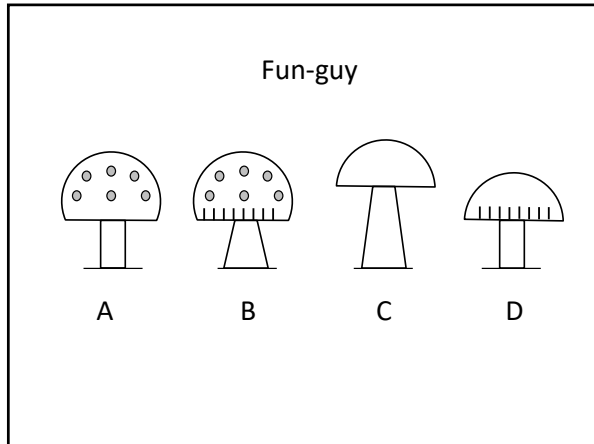
28



29



30



31

Character states

	A	B	C	D
1. Cap size (s/l)	large	large	Small	small
2. Gills (+/-)	absent	present	absent	present
3. Stalk shape (r/t)	rectangular	trapezoid	rectangular	trapezoid
4. Stalk length (l/s)	short	short	long	short
5. Spots (+/-)	present	present	absent	absent

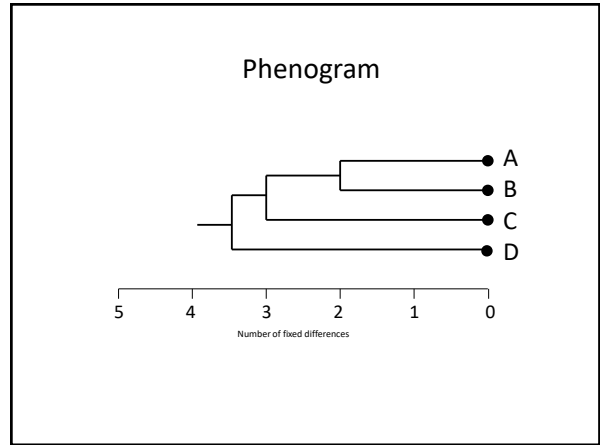
32

Distance matrix

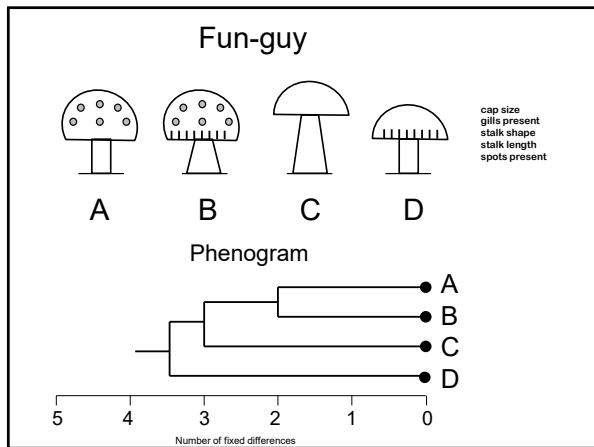
	A	B	C	D
A	-	2	4	3
B		-	4	3
C			-	3
D				-

	A/B	C	D
A/B	-	4	3
C		-	3
D			-

33



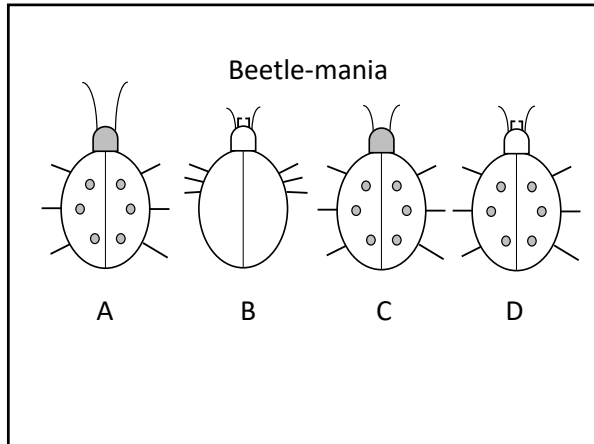
34



35



36



37

Character states

	A	B	C	D
1. Antennae (s/l)	long	short	short	short
2. Jaws (a/i)	inapparent	apparent	inapparent	apparent
3. Head (g/w)	grey	white	grey	white
4. Spots (+/-)	present	absent	present	present
5. Legs (a/l)	lateral	anterior	lateral	lateral

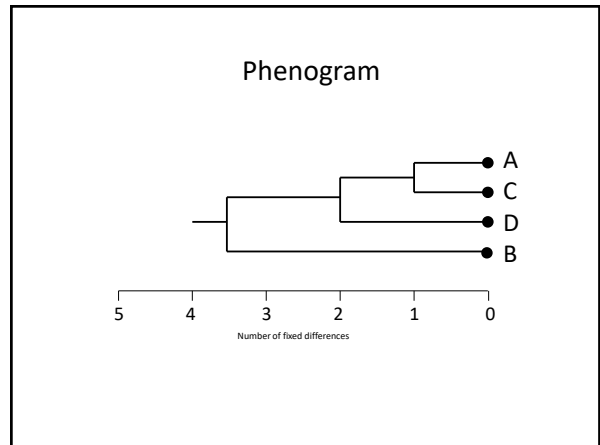
38

Distance matrix

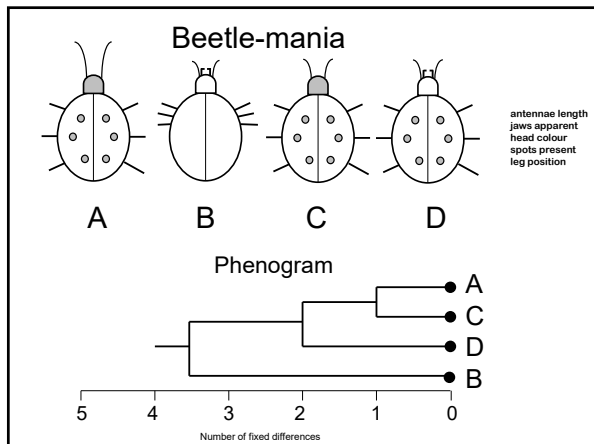
	A	B	C	D
A	-	5	1	3
B		-	4	2
C			-	2
D				-

	A/C	B	D
A/C	-	4.5	2.5
B		-	2
D			-

39



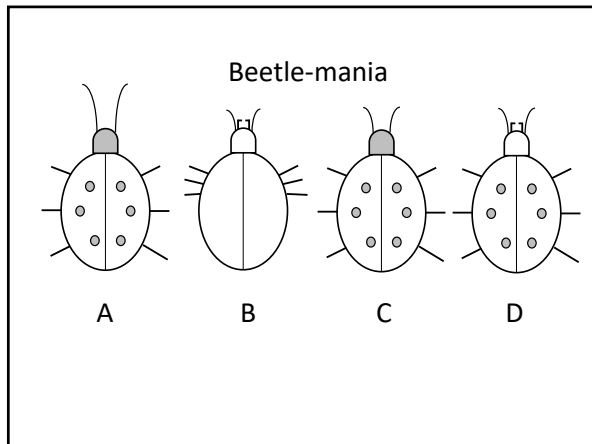
40



41



42



43

Character states

	A	B	C	D
1. Antennae (s/l)	long	short	short	short
2. Jaws (a/i)	inapparent	apparent	inapparent	apparent
3. Head (g/w)	grey	white	grey	white
4. Spots (+/-)	present	absent	present	present
5. Legs (a/l)	lateral	anterior	lateral	lateral

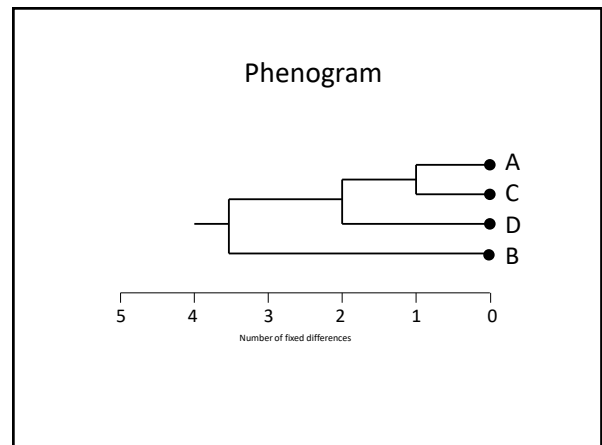
44

Distance matrix

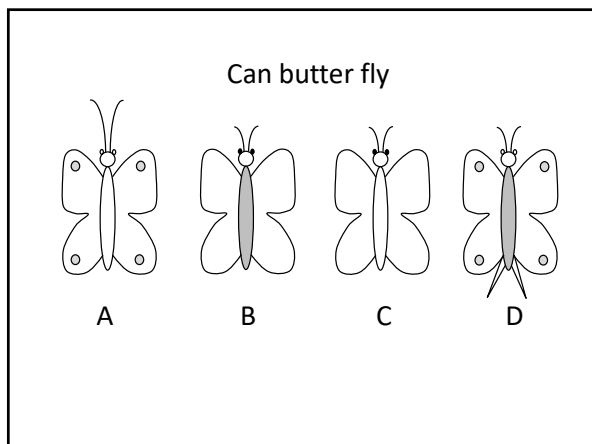
	A	B	C	D
A	-	5	1	3
B		-	4	2
C			-	2
D				-

	A/C	B	D
A/C	-	4.5	2.5
B		-	2
D			-

45



46



47

Character states

	A	B	C	D
1. Antennae (s/l)	long	short	short	short
2. Eye color (b/w)	white	black	black	white
3. Body color (g/w)	white	grey	white	grey
4. Spots (+/-)	present	absent	absent	present
5. Tail (+/-)	absent	absent	absent	present

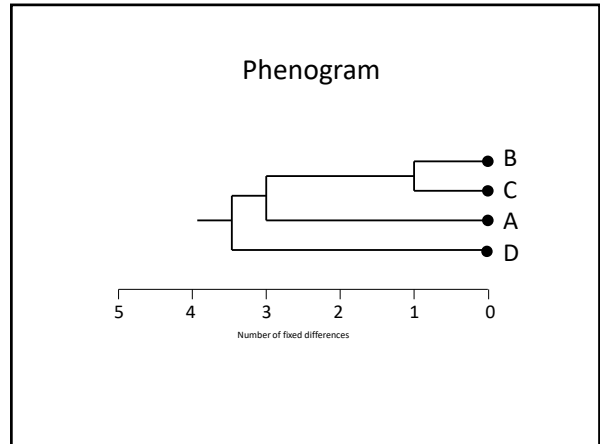
48

Distance matrix

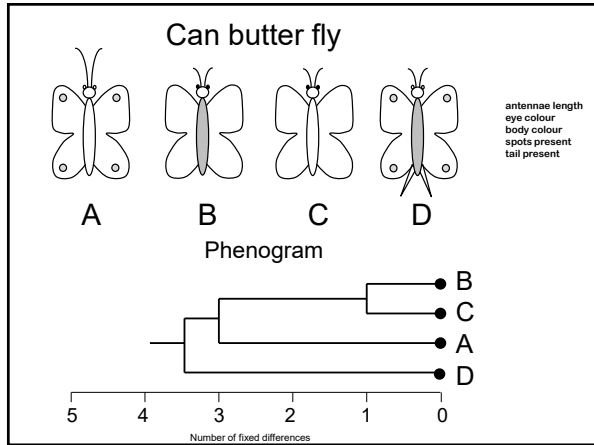
	A	B	C	D
A	-	4	3	3
B		-	1	3
C			-	4
D				-

	B/C	A	D
B/C	-	3.5	3.5
A		-	3
D			-

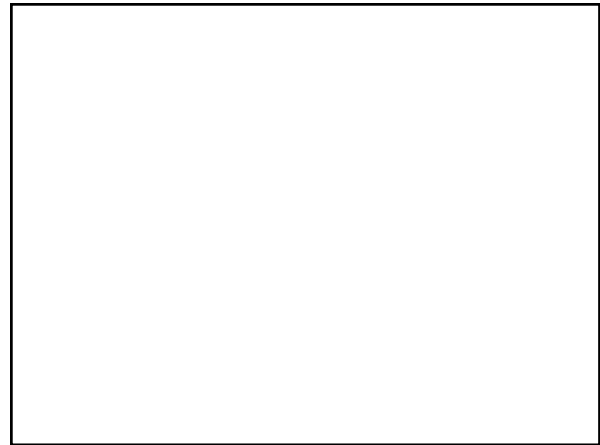
49



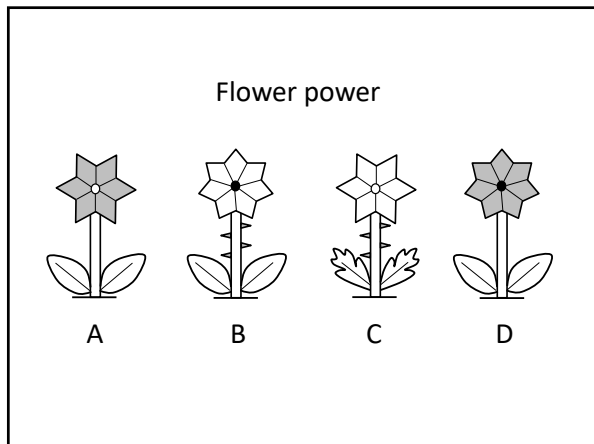
50



51



52



53

Character states

	A	B	C	D
1. Flower color (g/w)	grey	white	white	grey
2. Petals (no.)	6	7	6	7
3. Carpel color (b/w)	white	black	white	black
4. Leak shape (r/v)	round	round	variegated	round
5. Thorns (+/-)	absent	present	present	absent

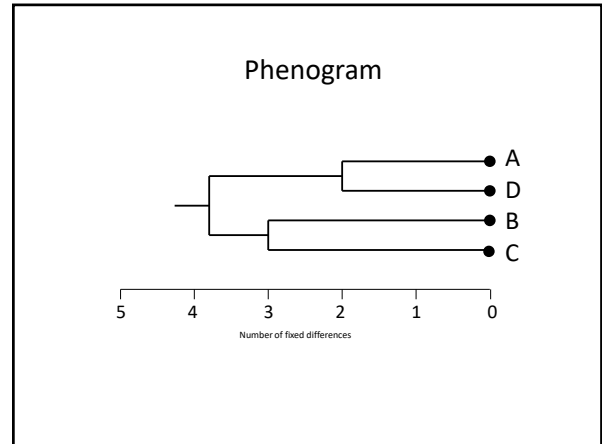
54

Distance matrix

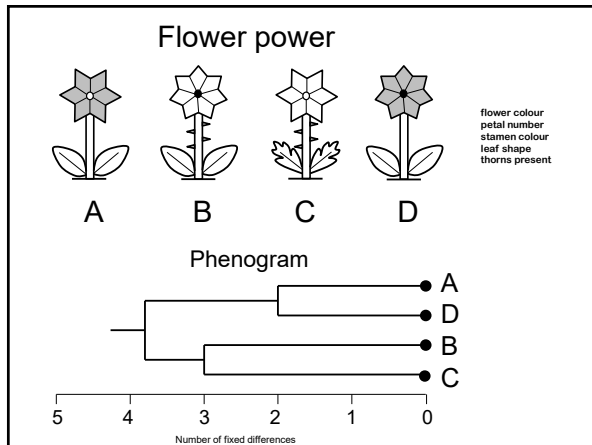
	A	B	C	D
A	-	4	3	2
B		-	3	2
C			-	5
D				-

	A/D	B	C
A/D	-	3	4
B		-	3
C			-

55



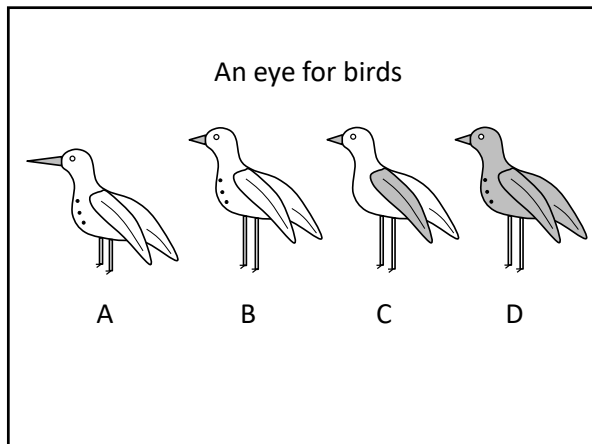
56



57



58

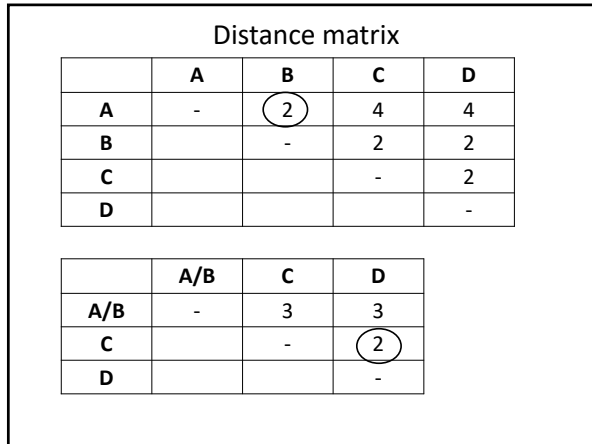


59

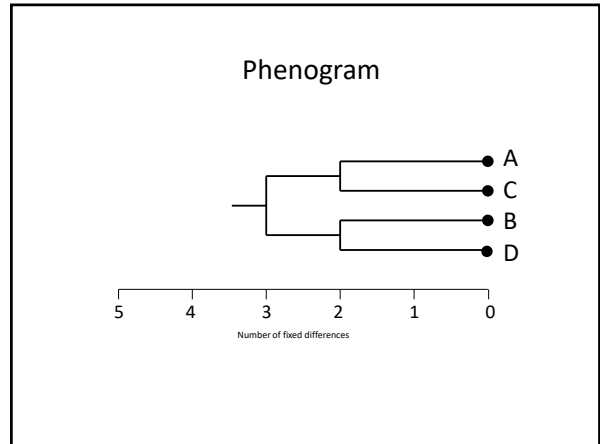
Character states

	A	B	C	D
1. Leg length (s/l)	short	long	long	long
2. Beak length (s/l)	long	short	short	short
3. Body color (g/w)	white	white	white	grey
4. Wing color (g/w)	white	white	grey	grey
5. Spots (+/-)	present	present	absent	present

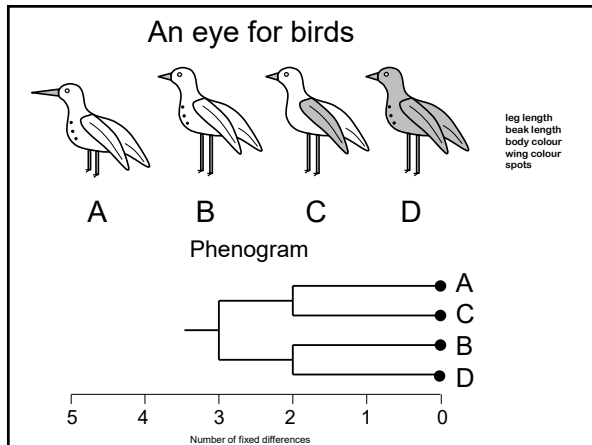
60



61



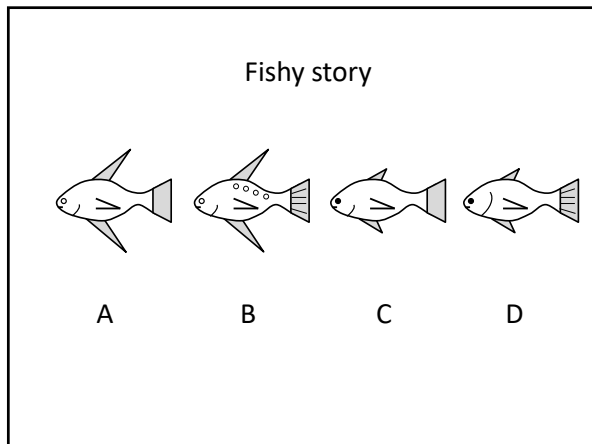
62



63



64



65

Character states

	A	B	C	D
1. Eye color (b/w)	white	white	black	black
2. Tail (s/st)	smooth	striped	smooth	striped
3. Fin length (l/s)	long	long	short	short
4. Spots (+/-)	absent	present	absent	absent
5. Gill cover (l/s)	short	short	short	long

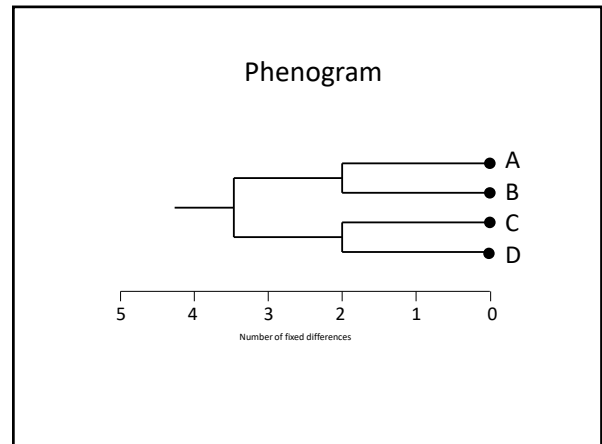
66

Distance matrix

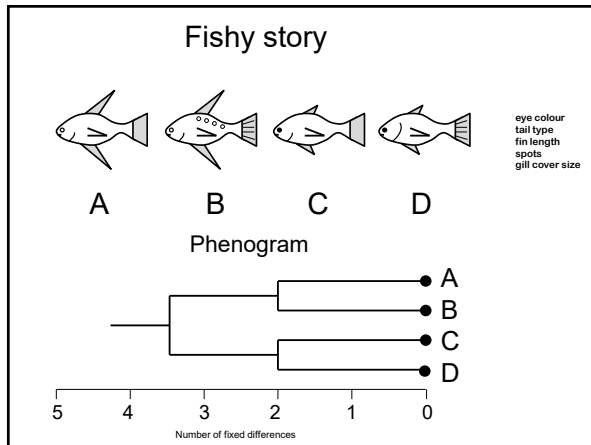
	A	B	C	D
A	-	2	2	4
B		-	4	4
C			-	2
D				-

	A/B	C	D
A/B	-	3	4
C		-	2
D			-

67



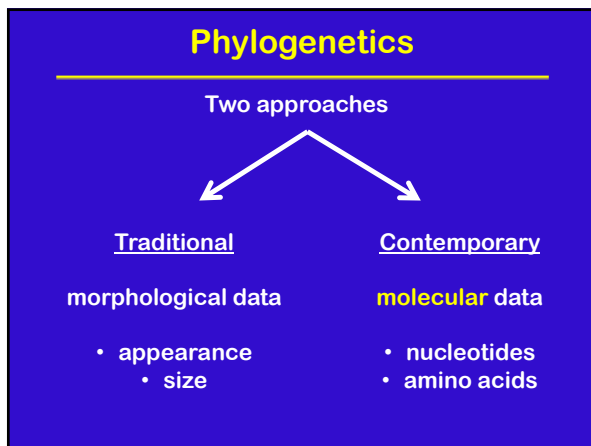
68



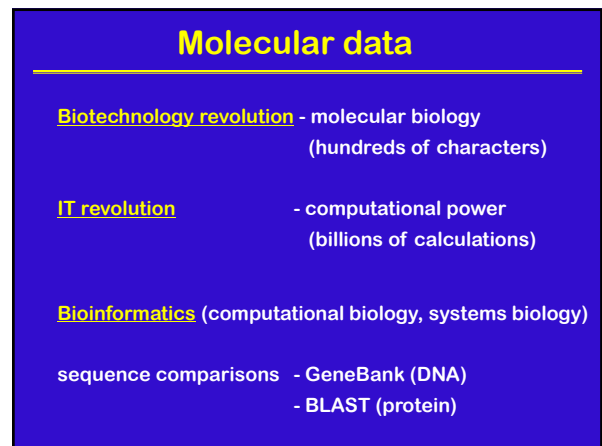
69



70



71



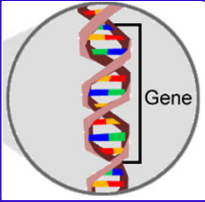

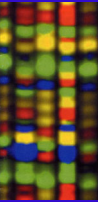
72

Obtaining DNA sequences

Select region to amplify ('gene' fragment)

Use flanking primers for PCR amplification

Read nucleotide sequences (A, T, G, C)

73

Input DNA sequences

Amylavorax dehorityi (novel endosymbiotic ciliate from stomach of kangaroo)

```

0      AACCTGGTTA ATCCTGCCAG TAGTCATATG CTTGTCTCAA AGACTAAGCC
50     ATGCATGCTT AAGTATAAAT AACTACACAG TAAAACCTCG AATGGCTCAT
100    TAAAACAGTT ATAGTTTATT TGATACATTA AATGGATAAC TGTAGAAAAA
150    CTAGAGCTAA TACATGCTGA GGCCGCAAGG TCGTATTAT TAGATATTCC
200    AATTAAGGTG AATCATAATA ACTTCGCAAA TCACGATTTT GTCGTGATAA
250    ATCATCCAAG TTCTGCCCTT ATCATGCTTT CGATGGTAGT GTATTGGATT
300    ACCATGGCTT TTACGGGTAA CGGGGAATTA GGCTTGATT CCGGAGAAGG
350    AGCCTGAGAA ACGGCTACTA CATCTACGSA AGGCAGCAGG CCGCTAATTT
400    ACCCAATCCT GACTCAGGGA GGGGTGACAA AGATATAACG ACGTGATTAA
450    AATCGCGATT GGTAGTGAGG GTTTCOTACA CCGAACCACT AGTACGATTA
500    GAGGGCAAGT CTGGTGCCAG CAGCCGCGGT AATTCAGCTT CTAATAGCCT
550    ATATTAAGT  TGCTGCAGT  AAAAGCTCG  TAGTTGGATT  TCAAGGATTA
600    TAATCACCTT  CTGGTGACTA  TACCCTACTA  CCCTTTTAGG  TGTACTGTG
650    AGAAAATTAG  AGTGTTTAAA  GCAGGCTATT  GCAAGAATAC  ATTAGCATGG
700    AATAACGAAT  GTGTTAGAA  TCTTGTTAA  TTCTAGACGC  GGTTAATAGG
750    GACATTTGGG  GGCATTAGTA  TTTAATAGTC  AGAGGTGAAA  TTCTTGGAAT
800    TTCTTAAAGA  CCTAAGTAT  GCAARAGAT  TTGCCAAGGA  TCTTTTCATT
850    AATCAAGAAC  GAAAGATAG  GGATCAAAGA  CAATCAGATA  CTGCTCATG
    
```

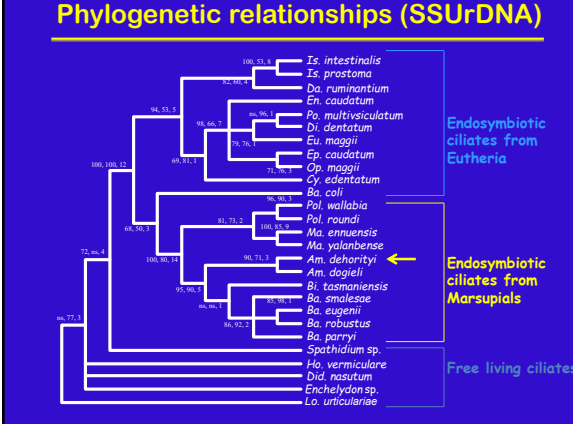
74

Comparing DNA sequences

	10	20	30	40
<i>Sa. muris</i>	AACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCT--TAAAGATTAAG			
<i>Is. intestinalis</i>	---C..
<i>Is. prostoma</i>	---C..
<i>D. ruminantium</i>	---C..
<i>B. coli</i>	---C..
<i>En. caudatum</i>	---C..
<i>P. multivesiculatum</i>	---C..
<i>E. maggii</i>	---C..
<i>D. dentatum</i>	---C..
<i>Ep. caudatum</i>	---C..
<i>O. maggii</i>	---C..
<i>C. edentatum</i>	---C..
<i>P. turniae</i>	---C..
<i>Po. roundi</i>	---C..
<i>Ma. ennuensis</i>	---C..
<i>Ma. yalanbense</i>	---C..
<i>Am. dehorityi</i>	---C..
<i>Am. dogieli</i>	---C..
<i>Bi. tasmaniensis</i>	---C..
<i>Ba. smalesae</i>	---C..

75

Phylogenetic relationships (SSUrDNA)



76

Sequence mutations

point in-del cross-over chromosomal

↓ ↓ ↓ ↓

..ACAGA-CGTAACGC..

..ACTGAACCATAGCG..

- point mutation (substitution)
- in-del mutation (insertion, deletion)
- recombination mutation (cross-over)
- chromosomal mutation (large blocks)

77

Under-estimation

consider point mutations (substitutions)

single multiple coincidental parallel convergent back

↑ ↑ ↑ ↑ ↑ ↑

A A A T T

ACTGAACGTAACGC

↓ ↓ ↓ ↓ ↓

C G A C

↓ ↓

T T

↑ ↑ ↑ ↑

AATGAAAGAATCGC

↑ ↑ ↑ ↑

ACTGTAGGAATCGC

although 12 mutations have occurred, only 3 are recorded by current sequences

78

Analytical methods

Bifurcating process

Assuming single ancestral gene

- MRCA = most recent common ancestor
- LUCA = last universal common ancestor

79

Group exercise: Pair-up

Let's build a tree.

Handout available for:

- protozoa (unicellular motile protista)

80

Protozoan assemblages

Morphotypic relationships

A Amoebae B Flagellates C Sporozoa D Ciliates

temporary flagellated stages

undulipodia (2+9 cilia & flagella)

81

Let's build a tree

DNA sequence data

A = amoebae (*Entamoeba*) causes amoebic dysentery
 B = flagellate (*Trichomonas*) causes STD
 C = sporozoa (*Plasmodium*) causes malaria
 D = ciliate (*Balantidium*) causes bloody diarrhoea

Protozoan	SSUrDNA sequence (40 nucleotides)
A = <i>Entamoeba</i>	TTGGACTCAGATCTTCTGCCCCGCCGAGCGCTTGCCGATG
B = <i>Trichomonas</i>	TACAACCTCACCTATTCTGCCGAGCCGGGGCGGTGGCAAAG
C = <i>Plasmodium</i>	TTGGTCTCAGCTCTCCTGGCGCCCCGGGCGCGAGCGGTTG
D = <i>Balantidium</i>	TTGGTCTCTGCTCTCCTGGCGCCCTGGGCGCGAGCGGATG

82

Proto-zoology

A *Entamoeba* B *Trichomonas* C *Plasmodium* D *Balantidium*

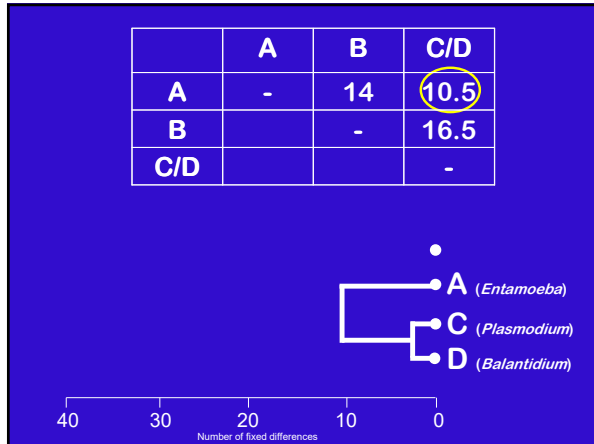
	Sequence (small subunit ribosomal DNA)																																							
A	T	T	G	G	A	C	T	C	T	T	C	T	G	C	C	C	G	C	G	A	G	C	G	C	T	T	G	C	C	G	A	T	G							
B	T	A	C	A	A	C	T	C	A	C	T	A	T	T	C	T	G	C	C	G	A	G	C	C	G	G	G	C	G	C	T	T	G	G	C	A	A	A	A	G
C	T	T	G	G	T	C	T	C	A	G	C	T	C	T	C	C	T	G	G	C	G	C	C	C	C	G	G	G	C	G	C	G	A	G	C	G	G	T	T	G
D	T	T	G	G	T	C	T	C	T	G	C	T	C	T	C	C	T	G	G	C	G	C	C	C	T	G	G	G	C	G	C	G	A	G	C	G	G	A	T	G

83

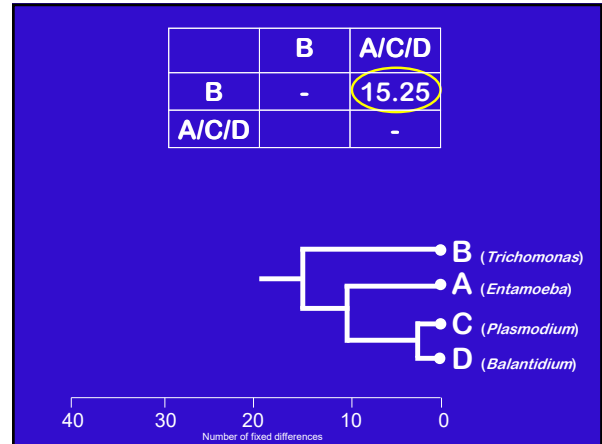
	A	B	C	D
A	-	14	10	11
B		-	16	17
C			-	3
D				-

Number of fixed differences

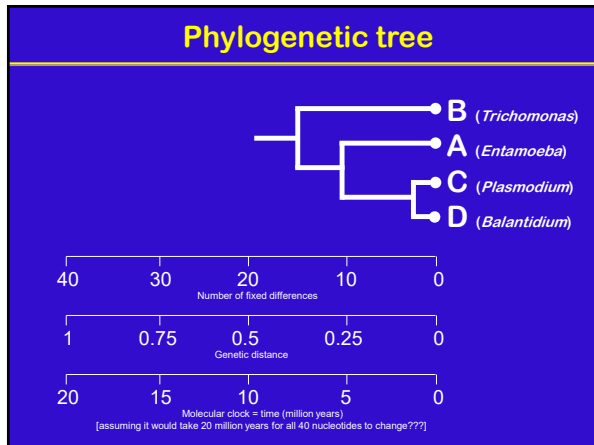
84



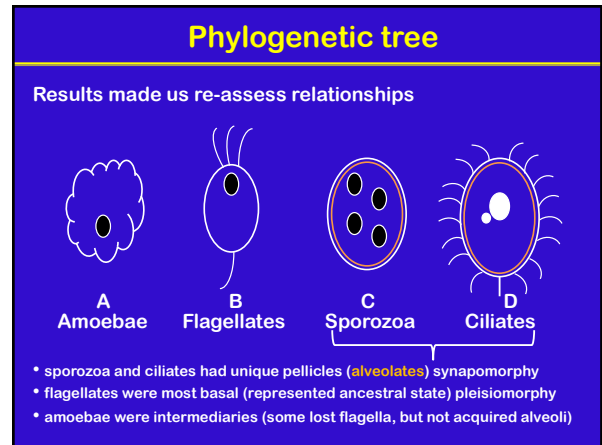
85



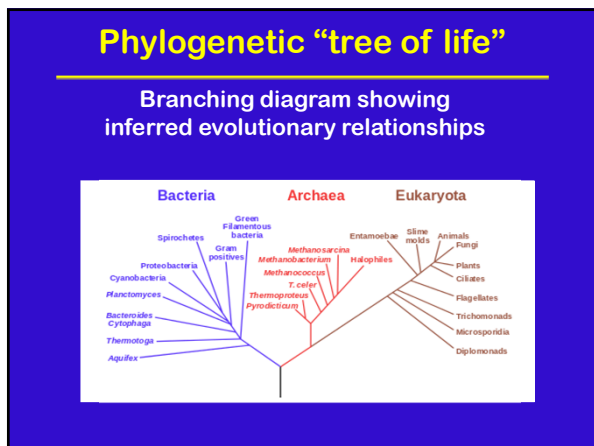
86



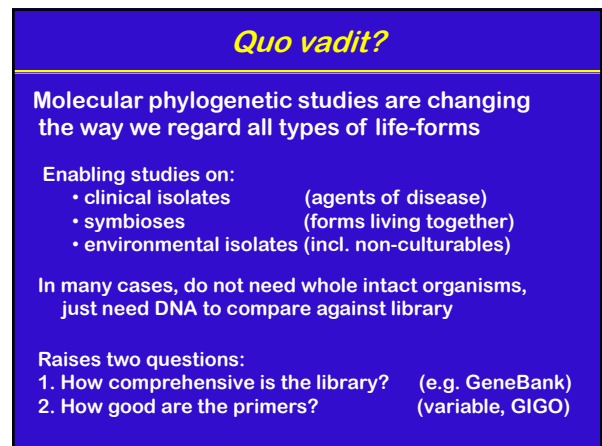
87



88



89



90