# The myth? about phylogenetic trees



## The myth? about phylogenetic trees





### **Complex Evolutionary Relationships**

"Evolutionary relationships between taxa are most often represented as phylogenetic trees and many different algorithms for the tree construction have been developed.

This is of course justified by the assumption that evolution is a branching or tree-like process. However, a set of real data often contains a number of different and sometimes conflicting signals and thus does not always clearly support a unique tree."

Huson 1998

•4 Species

•3 Genes

•3 Trees

•4 Species

•3 Genes

•3 Trees

Gene 1 Species A: ATGCGTGACT...TGA Gene 1 Species B: ATGCGTGCCT...TGA Gene 1 Species C: ATGCGTGCCT...TGA Gene 1 Species D: ATGCGTGACT...TGA



•4 Species

•3 Genes

•3 Trees

Gene 1 Species A: ATGCGTGACT...TGA Gene 1 Species B: ATGCGTGCCT...TGA Gene 1 Species C: ATGCGTGCCT...TGA Gene 1 Species D: ATGCGTGACT...TGA

Gene 2 Species A: ATGCGTGCCT...TGA Gene 2 Species B: ATGCGTGCCT...TGA Gene 2 Species C: ATGCGTGACT...TGA Gene 2 Species D: ATGCGTGACT...TGA





•4 Species

•3 Genes

•3 Trees

Gene 1 Species A: ATGCGTGACT...TGA Gene 1 Species B: ATGCGTGCCT...TGA Gene 1 Species C: ATGCGTGCCT...TGA Gene 1 Species D: ATGCGTGACT...TGA

Gene 2 Species A: ATGCGTGCCT...TGA Gene 2 Species B: ATGCGTGCCT...TGA Gene 2 Species C: ATGCGTGACT...TGA Gene 2 Species D: ATGCGTGACT...TGA

Gene 3 Species A: ATGCGTGCCT...TGA Gene 3 Species B: ATGCGTGACT...TGA Gene 3 Species C: ATGCGTGCCT...TGA Gene 3 Species D: ATGCGTGACT...TGA













Split 1: {ABC | DEF}



**Split 1:** {ABC | DEF} **Split 2:** {FAB | EDC}



Split 1: {ABC | DEF}
Split 2: {FAB | EDC}
Split 3: {EFA | DCB}

## Splits decomposition

•Splits decomposition method, in contrast to MP and ML (parameter optimization), is a transformation based approach.

•Evolutionary data are transformed into a sum of weakly compatible splits and then represented by a split graph.

•For ideal data, this is a tree

•For less ideal data this gives a tree-like network as evidence for different and conflicting phylogenies.

•Split decomposition (Bandelt 1992) does not attempt to force data onto a tree.

# A Split is ...

A "split" is a partition of the taxa into two nonempty subsets, such as the partition obtained when we remove a branch from a phylogenetic tree. For example, removing the branch indicated by the arrow, splits the taxa into two groups {B, C, D, E} and {o, A, F, G, H, I, J, K}.





	Α	В	С	D	$\mathbf{E}$	F	G	Η	weights
1	٠	0	0	0	0	0	0	0	7.92
2	0	٠	0	0	0	0	0	0	3.31
3	0	0	٠	0	0	0	0	0	1.74
4	0	0	0	•	0	0	0	0	3.72
5	0	0	0	0	٠	0	0	0	8.94
6	0	0	0	0	0	٠	0	0	3.88
7	0	0	0	0	0	0	•	0	5.63
8	0	0	0	0	0	0	0	•	6.21
9	٠	٠	0	0	0	0	0	0	1.12
10	•	٠	•	0	0	0	0	0	1.28
11	0	٠	٠	•	•	0	0	0	2.83
12	0		٠	٠	0	0	0	0	3.63
13	0	0	0	0	•	•	•	0	1.28
14	0	0	0	0	0	•	•	•	1.95



		A	D	C	D	Ľ	г	G	п	weights
/	1	٠	0	0	0	0	0	0	0	7.92
1	2	0	٠	0	0	0	0	0	0	3.31
;	3	0	0	٠	0	0	0	0	0	1.74
,	4	0	0	0	•	0	0	0	0	3.72
ļ	5	0	0	0	0	•	0	0	0	8.94
(	6	0	0	0	0	0	•	0	0	3.88
,	7	0	0	0	0	0	0	•	0	5.63
i	8	0	0	0	0	0	0	0	•	6.21
	9	•	•	0	0	0	0	0	0	1.12
1	0	•	•	•	0	0	0	0	0	1.28
1	.1	0	•	•	•	•	0	0	0	2.83
1	2	0		٠	٠	0	0	0	0	3.63
1	3	0	0	0	0	•	•	•	0	1.28
1	4	0	0	0	0	0	•	•	•	1.95

### Split Network

Formally, for a given taxon set X and set of splits S; we define a split network N to be a connected graph in which some of the nodes are labelled by taxa and all edges are labelled by splits, such that

(N1) Removing all edges associated with a given split S in S divides N into two connected components, one part containing all taxa on one side of S and the other part containing all taxa on the other side.

(N2) The edges along any shortest path in  $\overline{N}$  are all associated with different splits.

# Split System

#### •Compatible System of Split Σ:

For any 2 splits S1{A1,A1} and S2={A2,A2} in S one of the 4 intersections:

A1 $\cap$ A2, A1 $\cap$ A2, A1 $\cap$ A2 or A1 $\cap$ A2

is empty.

(Thus any phylogenetic tree gives rise to a compatible  $\Sigma$  system)

#### •Weakly Compatible System of Split Σ:

For any 3 splits S1, S2 and S3 and all Ai in Si at least one of the 4 intersections:

A1∩A2∩A3, A1∩A2∩A3, A1∩A2∩A3 or A1∩A2∩A3

is empty.

(any 2 splits are allowed to be incompatible)

#### ABCDEF

**Split 1:** {AB | CDEF} **Split 2:** {ED | ABCF} Split 1: {ABC | DEF}
Split 2: {FAB | EDC}
Split 3: {EFA | DCB}

SplitsTree: Application of Phylogenetic Networks in Evolutionary Studies, Daniel H. Huson

SplitsTree Demo



# Case Study: Salmonella lineage

	K	L	M	N	0	Р	Q	R
1	S.bongori 12419	E.coli EDL933	E.coli CFT073	S.flezneri 2a 301	E.coli MG1655	Tree Num	Diff-ConsNetw	tree
2	SBG0351	Z0495	c0505	SF0334	b0397	165	2	(ECO,((EDL,SF),(SBO,((((MCC,SSP),(STY,TY2)),((SEN,SG),(SL,(STM,SDT)))),SAR))),CFT);
3	SBG0494	Z0725	c0671	SF0497	Ь0585	242	2	((SB0,(SAR,((SL,(STM,SDT)),(((MCC,SSP),(TY2,STY)),(SG,SEN))))),((EDL,SF),ECO),CFT);
4	SBG1146	Z2802	c2172	SF1455	b1768	710	2	(EDL,(ECO,(CFT,(SBO,((((SG,SEN),(SL,(STM,SDT))),(((SSP,MCC),TY2),STY)),SAR)))),SF);
5	5 SBG1948	Z3245	c2602	SF2141	b2076	869	2	((SBO,(SAR,(((TY2,STY),(MCC,SSP)),((SEN,SG),((SDT,STM),SL))))),((ECO,SF),EDL),CFT);
E	SBG2960	Z4576	c3973	SF3252	b3212	1392	2	((ECO,(CFT,(SBO,(SAR,((((SSP,MCC),STY),TY2),((SEN,SG),(SL,(STM,SDT))))))),SF,EDL);
7	SBG3398	Z5198	c4630	SF3758	b3706	1600	2	((SB0,(SAR,(((MCC,SSP),(TY2,STY)),(SEN,(SG,((STM,SDT),SL))))),((SF,EDL),ECO),CFT);
8	SBG3293	Z5047	c4446	SF3660	b3620	1641	2	((ECO.(CFT.(SBO.(SAR.((SG.SEN).(((TY2,STY).(SSP,MCC)).(SL.(STM,SDT))))))),SF,EDL);
9	SBG3219	Z4956	c4354	SF3574	b3539	1678	2	(((SAR,(((SEN,SG),(SL,(STM,SDT))),((STY,TY2),(SSP,MCC)))),SBO),(EDL,(ECO,SF)),CFT);
1	D SBG3118	Z4771	c4193	SF3439	b3416	1732	2	((CFT,(SBO,(SAR,(((SEN,SG),((STM,SDT),SL)),((MCC,SSP),(TY2,STY)))))),(ECO,SF),EDL);
1	1 SBG3673	Z5631	c5003	SF4172	b4033	1812	2	(((SF,EDL),ECO),((SAR,(((TY2,STY),(MCC,SSP)),((SG,SEN),(SDT,(SL,STM))))),SBO),CFT);
1	2 SBG3693	Z5652	c5031	SF4151	b4054	1828	2	((SB0,(((STY,TY2),(((SEN,SG),((STM,SDT),SL)),(SSP,MCC))),SAR)),((EDL,SF),ECO),CFT);
1	3 SBG0008	Z0009	c0013	SF0010	Ь0009	6	4	((((MCC,SSP),(TY2,STY)),(SAR,((SEN,SG),((STM,SDT),SL)))),(((SF,ECO),EDL),CFT),SBO);
1	4 SBG0092	Z0075	c0082	SF0061	Ь0066	45	4	((SAR,(SBO,(((STY,TY2),(SSP,MCC)),((SG,SEN),((SDT,STM),SL))))),(SF,(EDL,ECO)),CFT);
1:	5 SBG0093	Z0076	c0083	SF0062	Ь0067	46	4	((SBO,(SAR,(((SL,(STM,SDT)),((MCC,SSP),(TY2,STY))),(SG,SEN)))),((ECO,SF),EDL),CFT);
1	6 SBG0342	Z0482	c0493	SF0322	b0386	159	4	((SBO,(SAR,((SSP,MCC),((TY2,STY),(((STM,SL),SDT),(SG,SEN)))))),((SF,EDL),ECO),CFT);
1	7 SBG0357	Z0501	c0513	SF0340	b0403	171	4	((SBO,(SAR,(((SEN,SG),(SDT,(STM,SL))),((MCC,SSP),(STY,TY2))))),(EDL,(ECO,SF)),CFT);
1	8 SBG0422	Z0581	c0584	SF0410	b0465	208	4	((SBO,(SAR,((SSP,MCC),((TY2,STY),((SEN,SG),((SDT,SL),STM))))),((EDL,SF),ECO),CFT);
1	9 SBG0443	Z0604	c0604	SF0429	b0484	222	4	(((EDL,CFT),SF),(SBO,(SAR,(((SL,(STM,SDT)),(SG,SEN)),((TY2,STY),(SSP,MCC))))),ECO);
2	D SBG0496	Z0727	c0673	SF0498	b0586	243	4	(((((STY,TY2),((MCC,SSP),((SG,SEN),(STM,(SDT,SL))))),SAR),SBO),((EDL,SF),ECO),CFT);
2	1 SBG0862	Z1297	c1083	SF0948	b0947	442	4	((SBO,(SAR,((SEN,SG),(((TY2,STY),(MCC,SSP)),((STM,SL),SDT))))),((EDL,SF),ECO),CFT);
2	2 SBG1025	Z1722	c1353	SF1088	b1084	492	4	(((ECO,CFT),((SAR,(((SDT,STM),SL),((SG,SEN),((STY,TY2),(MCC,SSP))))),SBO)),EDL,SF);
2	3 SBG1575	Z2532	c1745	SF1280	b1276	563	4	(((ECO,EDL),(((((SG,SEN),(SL,(STM,SDT))),((MCC,SSP),(TY2,STY))),SAR),SBO)),SF,CFT);
2	4 SBG1531	Z2450	c1797	SF1331	b1325	584	4	((SBO,(SAR,((TY2,STY),((SSP,MCC),((SEN,SG),(SDT,(STM,SL)))))),((SF,EDL),ECO),CFT);
2	5 SBG1322	Z2581	c1982	SF1612	b1591	623	4	((ECO,(SF,EDL)),(SBO,(((SG,SEN),(((MCC,SSP),(STY,TY2)),(STM,(SDT,SL)))),SAR)),CFT);
2	6 SBG1298	Z2610	c2001	SF1632	b1609	633	4	(((CFT,(SBO,(SAR,((STY,TY2),((SG,SEN),((MCC,SSP),(SL,(STM,SDT))))))),ECO),EDL,SF);
2	7 SBG1264	Z2661	c2038	SF1673	b1646	655	4	((ECO,EDL),(CFT,(SBO,(SAR,((((SG,SEN),((SDT,STM),SL)),(MCC,SSP)),(TY2,STY))))),SF);
2	8 SBG1210	Z2711	c2078	SF1713	b1683	662	4	(EDL,(ECO,((SBO,(SAR,((SL,(STM,(SDT,(SG,SEN)))),((MCC,SSP),(STY,TY2))))),CFT)),SF);
2	9 SBG1153	Z2792	c2161	SF1465	b1760	703	4	((SBO,(SAR,(((SDT,STM),SL),((SEN,SG),(((TY2,STY),MCC),SSP))))),((EDL,SF),ECO),CFT);
3	D SBG1704	Z2882	c2244	SF1390	b1835	765	4	(CFT.(ECO.(EDL,((SAR,(((SEN,SG),(SL,(STM,SDT))),((MCC,SSP),(STY,TY2)))),SBO))),SF);
3	1 SBG1717	Z2897	c2258	SF1857	b1846	772	4	((SBO,(SAR,((SG,SEN),((SL,(STM,SDT)),((STY,TY2),(MCC,SSP)))))),((SF,EDL),CFT),ECO);
3	2 SBG1813	Z3047	c2374	SF2000	b1956	833	4	(((SAR.((STY,TY2),((((STM,SDT),SL),(SEN,SG)),(SSP,MCC)))),SBO),(EDL,(ECO,SF)),CFT);
3	3 SBG1977	Z3379	c2660	SF2215	b2130	884	4	(((SAR,(((TY2,STY),(MCC,SSP)),((SG,SEN),((STM,SDT),SL)))),SBO),((ECO,CFT),SF),EDL);
3	4 SBG2063	Z3471	c2754	SF2297	b2213	929	4	(((SF,EDL),ECO),(SBO,(SAR,((STY,TY2),((((SL,STM),SDT),(SEN,SG)),(MCC,SSP))))),CFT);
3	5 SBG2080	Z3501	c2784	SF2325	b2243	943	4	((ECO,(CFT,(SBO,(SAR,(((SSP,MCC),(TY2,STY)),(SDT,(SL,(STM,(SEN,SG))))))),EDL,SF);
3	6 SBG2090	Z3515	c2799	SF2336	b2257	947	4	((ECO,(EDL,SF)),(SBO,(SAR,(((SG,SEN),((STY,TY2),(SSP,MCC))),((SL,STM),SDT)))),CFT);
1 01	T ODOOMOD	70004	0000	050440	1.0044	40.05		

# Case Study: Salmonella lineage

	K	L	M	N	0	Р	Q	R	
1	S.bongori 12419	E.coli EDL933	E.coli CFT073	S.flezneri 2a 301	E.coli MG1655	Tree Num	Diff-ConsNetw	tree	
2	SBG0351	Z0495	c0505	SF0334	Ь0397	165	2	(ECO.((EDL,SF).(SBO.((((MCC,SSP).(STY,TY2)).((SEN,SG).(SL.(STM,SDT)))).SAR))).CFT);	
3	SBG0494	Z0725	c0671	SF0497	b0585	242	2	((SB0,(SAR,((SL,(STM,SDT)),(((MCC,SSP),(TY2,STY)),(SG,SEN))))),((EDL,SF),ECO),CFT);	
4	SBG1146	Z2802	c2172	SF1455	b1768	710	2	(EDL,(ECO,(CFT,(SBO,((((SG,SEN),(SL,(STM,SDT))),(((SSP,MCC),TY2),STY)),SAR)))),SF);	
5	SBG1948	Z3245	c2602	SF2141	b2076	869	2	((SBO,(SAR,(((TY2,STY),(MCC,SSP)),((SEN,SG),((SDT,STM),SL))))),((ECO,SF),EDL),CFT);	
6	SBG2960	Z4576	c3973	SF3252	b3212	1392	2	((ECO,(CFT,(SBO,(SAR,((((SSP,MCC),STY),TY2),((SEN,SG),(SL,(STM,SDT))))))),SF,EDL);	
7	SBG3398	Z5198	c4630	SF3758	b3706	1600	2	((SBO,(SAR,(((MCC,SSP),(TY2,STY)),(SEN,(SG,((STM,SDT),SL))))),((SF,EDL),ECO),CFT);	
8	SBG3293	Z5047	c4446	SF3660	b3620	1641	2	((ECO,(CFT,(SBO,(SAR,((SG,SEN),(((TY2,STY),(SSP,MCC)),(SL,(STM,SDT))))))),SF,EDL);	
9	SBG3219	Z4956	c4354	SF3574	b3539	1678	2	(((SAR.(((SEN,SG),(SL,(STM,SDT))),((STY,TY2),(SSP,MCC)))),SBO),(EDL,(ECO,SF)),CFT);	
10	SBG3118	Z4771	c4193	SF3439	b3416	1732	2	((CFT,(SBO,(SAR,(((SEN,SG),((STM,SDT),SL)),((MCC,SSP),(TY2,STY))))),(ECO,SF),EDL);	
11	SBG3673	Z5631	c5003	SF4172	b4033	1812	2	<pre>(((SE_EDL)ECD).((SAR.(((TY2,STY).(MCC,SSP)).((SG,SEN).(SDT.(SL,STM))))).SBO).CFT);</pre>	
12	SBG3693	Z5652	c5031					Y2).(((SEN,SG).((STM,SDT),SL)).(SSP,MCC))).SAR)).((EDL,SF).ECO).CFT);	
13	SBG0008	Z0009	c0013					(TY2,STY)),(SAR,((SEN,SG),((STM,SDT),SL)))),(((SF,ECO),EDL),CFT),SBO);	
14	SBG0092	Z0075	c0082			TP	<b>100</b>	STY,TY2),(SSP,MCC)),((SG,SEN),((SDT,STM),SL))))),(SF,(EDL,ECO)),CFT);	
15	SBG0093	Z0076	c0083		) J L	L		SL.(STM,SDT)).((MCC,SSP).(TY2,STY))).(SG,SEN)))).((ECO,SF).EDL).CFT);	
16	SBG0342	Z0482	c0493					SSP,MCC).((TY2,STY).(((STM,SL),SDT).(SG,SEN)))))).((SF,EDL).ECO).CFT);	
17	SBG0357	Z0501	c0513	SF0340	60403	171	4	<pre>((SB0,(SAF,(((SEN,SG),(SDT,(STM,SL))),((MCC,SSP),(STY,TY2))))),(EDL,(EC0,SF)),CFT);</pre>	
18	SBG0422	Z0581	c0584	SF0410	b0465	208	4	((SBO,(SAR,((SSP,MCC),((TY2,STY),((SEN,SG),((SDT,SL),STM))))),((EDL,SF),ECO),CFT);	
19	SBG0443	Z0604	c0604	SF0429	b0484	222	4	(((EDL,CFT),SF),(SBO,(SAR,(((SL,(STM,SDT)),(SG,SEN)),((TY2,STY),(SSP,MCC))))),ECO);	
20	SBG0496	Z0727	c0673	SF0498	b0586	243	4	(((((STY,TY2).((MCC,SSP).((SG,SEN).(STM.(SDT,SL))))),SAR),SBO).((EDL,SF),ECO),CFT);	
21	SBG0862	Z1297	c1083	SF0948	b0947	442	4	((SBO,(SAR,((SEN,SG),(((TY2,STY),(MCC,SSP)),((STM,SL),SDT))))),((EDL,SF),ECO),CFT);	
22	SBG1025	Z1722	c1353	SF1088	b1084	492	4	(((ECO,CFT),((SAR,(((SDT,STM),SL),((SG,SEN),((STY,TY2),(MCC,SSP))))),SBO)),EDL,SF);	
23	SBG1575	Z2532	c1745	SF1280	b1276	563	4	(((ECO,EDL),(((((SG,SEN),(SL,(STM,SDT))),((MCC,SSP),(TY2,STY))),SAR),SBO)),SF,CFT);	
24	SBG1531	Z2450	c1797	SF1331	b1325	584	4	((SB0,(SAR,((TY2,STY),((SSP,MCC),((SEN,SG),(SDT,(STM,SL)))))),((SF,EDL),ECO),CFT);	
25	SBG1322	Z2581	c1982	SF1612	b1591	623	4	((ECO,(SF,EDL)),(SBO,(((SG,SEN),(((MCC,SSP),(STY,TY2)),(STM,(SDT,SL)))),SAR)),CFT);	
26	SBG1298	Z2610	c2001	SF1632	b1609	633	4	(((CFT,(SBO,(SAR,((STY,TY2),((SG,SEN),((MCC,SSP),(SL,(STM,SDT))))))),ECO),EDL,SF);	
27	SBG1264	Z2661	c2038	SF1673	b1646	655	4	((ECO,EDL),(CFT,(SBO,(SAR,((((SG,SEN),((SDT,STM),SL)),(MCC,SSP)),(TY2,STY))))),SF);	
28	SBG1210	Z2711	c2078	SF1713	b1683	662	4	(EDL,(ECO,((SBO,(SAR,((SL,(STM,(SDT,(SG,SEN)))),((MCC,SSP),(STY,TY2))))),CFT)),SF);	
29	SBG1153	Z2792	c2161	SF1465	b1760	703	4	((SB0,(SAR,(((SDT,STM),SL),((SEN,SG),(((TY2,STY),MCC),SSP))))),((EDL,SF),ECO),CFT);	
30	SBG1704	Z2882	c2244	SF1390	b1835	765	4	(CFT,(ECO,(EDL,((SAR,(((SEN,SG),(SL,(STM,SDT))),((MCC,SSP),(STY,TY2)))),SBO))),SF);	
31	SBG1717	Z2897	c2258	SF1857	b1846	772	4	((SB0,(SAR,((SG,SEN),((SL,(STM,SDT)),((STY,TY2),(MCC,SSP)))))),((SF,EDL),CFT),ECO);	
32	SBG1813	Z3047	c2374	SF2000	b1956	833	4	(((SAR,((STY,TY2),((((STM,SDT),SL),(SEN,SG)),(SSP,MCC)))),SBO),(EDL,(ECO,SF)),CFT);	
33	SBG1977	Z3379	c2660	SF2215	b2130	884	4	(((SAR,(((TY2,STY),(MCC,SSP)),((SG,SEN),((STM,SDT),SL)))),SBO),((ECO,CFT),SF),EDL);	
34	SBG2063	Z3471	c2754	SF2297	b2213	929	4	(((SF,EDL),ECO),(SBO,(SAR,((STY,TY2),((((SL,STM),SDT),(SEN,SG)),(MCC,SSP))))),CFT);	
35	SBG2080	Z3501	c2784	SF2325	b2243	943	4	((ECO.(CFT.(SBO.(SAR.(((SSP,MCC).(TY2,STY)).(SDT.(SL.(STM.(SEN,SG)))))))),EDL,SF);	
36	SBG2090	Z3515	c2799	SF2336	b2257	947	4	((ECO,(EDL,SF)),(SBO,(SAR,(((SG,SEN),((STY,TY2),(SSP,MCC))),((SL,STM),SDT)))),CFT);	

#### **One tree – 15 Genomes**

⊢−10.01



#### **One Network – 15 Genomes**



# 15 Genomes and the big picture



Differences against	Number of Trees							
consensus Tree	Staphylococcus	Streptococcus	Neisseria	Salmonella				
0	0	10	11	0				
2	1	34	49	11				
4	7	70	183	43				
6	27	76	467	90				
8	56	95	434	158				
10	74	62	46	207				
12	118	51	0	261				
14	155	16	0	344				
16	161	12	0	335				
18	108	2	0	259				
20	22	1	0	153				
22	4	0	0	76				
24	1	0	0	15				

