Module 4 : Building a species tree

A tree of *Coronaviridae* viruses

Concepts:

- Topology
- Branches lengths
- Internal nodes
- Leaves
- Root
- Clades / lineages



Dylus et al. Nat Biotech 2023

Other tree representations



Concepts:

- Rooted vs. unrooted
- Scaled vs. unscaled



HOG:C0620879.1a with 61 members (insulin)

Vertebrata / Gnathostomata / Euteleostomi / Sarcopterygii / Tetrapoda / Amniota / Mammalia / Lower Level >



Graphical viewer

	1 - 5 - 1	2.4.6.8.	10 . 12 . 14 .	18 . 18 . 20 . 22 . 2	4 . 26 . 28 .	30 . 32 . 34 . 3	38 . 38 . 40 . 42	. 44 . 46	. 48 . 50 . 52	2 . 54 . 5f	8 . 58 . 60
Members	CHIGH1/53/		MALWM	LLPLLALLA	LWEPN	PAVAFV	NUHLCUS	HLVE	ALYLVC	GERU	FFYI
	MOUSE26691		MALLV	F L P L L A L L A	LWEPK	PTQAFV	KQHLCGP	HLVE	ALYLVC	GERG	FFYT
	MOUSE57227		MALWM	F L P L L A L L F	LWESH	PTQAFV	K Q H L C G S	HLVE	ALYLVC	GERG	FFYT
Alignment	RATNO17730	MPSCGHCSN	MALWI	FL PLLALLI	LWEPR	PAQAFV	KQHLCGS	HLVE	ALYLVC	GERG	FFYT
	RATNO18008		MALWM	FL PLLALL V	LWEPK	PAQAFV	KQHLCGP	HLVE	ALYLVC	GERG	FFYT
	NANGA19616		MALWM	LL PLLALLA	FWGPN	PGQAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
Ancestral synteny	ICTTR07348		MALWTF	LL PLLALLA	LLGPD	PAQAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	CERAT20372		MALWM	LL PLLALLA	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	CHLSB00063		MALWM	LLPLLALLA	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
Similar HOGs > <	MACFA08228		MALWM	L L P L L A L L A	LWGPD	PAPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	MACMU09924		MALWM	L L P L L A L L A	LWGPD	PAPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	MACNE29303		MALWM	LL PLLALLA	LWGPD	PAPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	MANLE26401		MALWM	LL PLLALLA	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	PAPAN06025		MALWM	L L P L L A L L A	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	COLAP22086		MALWM	LL PLLALLA	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	RHIBE04930		MALWM	L L P L L A L L A	LWGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	RHIRO07577		MALWM	LLPLLALLA	LCGPD	PVPAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	GORGO03436		MALWM	LLPLLALLA	LWGPD	PAAAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	HUMAN03911		MALWM	LLPLLALLA	LWGPD	PAAAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	PANPA05915		MALWM	L L P L L A L L A	LWGPD	PASAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	PANTR02017		MALWM	LL PLL VLL A	LWGPD	PASAFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT
	PONAB02480		MALWM	LLPLLALLA	LWGPD	PA-AFV	NQHLCGS	HLVE	ALYLVC	GERG	FFYT

General Scheme to Build Phylogenetic Trees



Measuring confidence with the Bootstrap

How would our tree inference change if we repeated the experiment with other data that came about through the same process?

Use "pseudo-replicates": resample alignment columns (with replacement), same length.



Slide courtesy Prof. R. Goldstein

How many branches are there in an unrooted bifurcating tree of taxa?



How many topologies?

Number of
"taxa"Number of binary trees
Unrooted31