

x
x
x
x
x

TUTORIAL 2: Tree Inference

Building Phylogenetic Trees using RAxML and FastTree algorithms



Assignment

- Build trees for the 16S rRNA sequence using RaxML and FastTree algorithms
- Explain how underlying algorithms work
- Comment on differences in trees based off the different algorithms



x x x x

What are Trees?

- Trees show how groups of organisms, nucleic acids, or protein sequences are related evolutionarily
- Show how these groups came about through evolution
- Allows scientists to follow evolutionary changes that have occurred over time



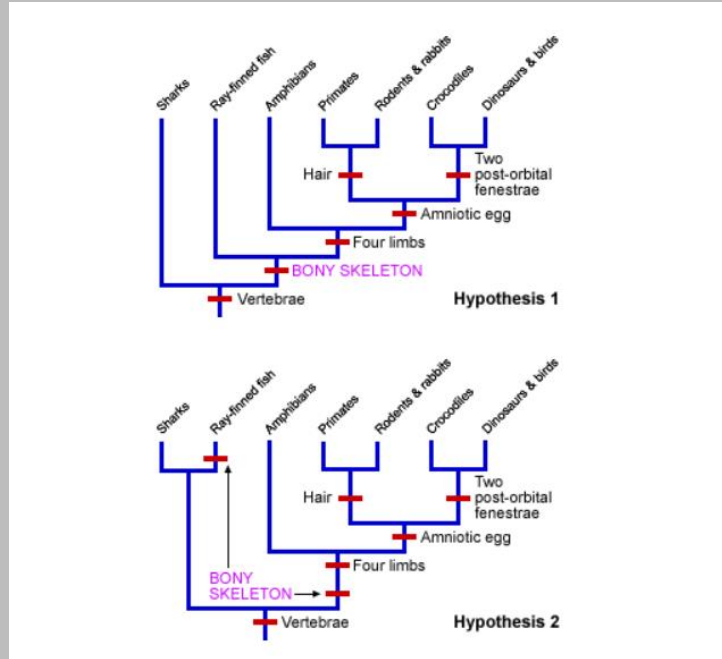
Types of Tree Algorithms

- Maximum Likelihood (ML) uses an evolutionary model and tree topology finds the tree which has the highest likelihood of giving the observed data
- ML scores represent the probability of the observed sequence data given a specific phylogenetic tree and evolutionary model
- Distance Methods calculate evolutionary distances and build trees based off of these distances. Are also much less computationally intensive
- Maximum Parsimony (MP) makes tree that has minimum number of changes to explain observed data



x x x x

Parsimony



<https://evolution.berkeley.edu/phylogenetic-systematics/reconstructing-trees-cladistics/reconstructing-trees-parsimony/>



x x x x

Bootstrapping

- Bootstrapping is like taking many different snapshots of your data by randomly selecting observations from it.
- Each bootstrap sample is the same size as the original dataset, but it's created by randomly selecting sites from the original dataset
- Some observations may be selected more than once, while others may not be selected at all
- Reconstruct a phylogenetic tree for each replicate, and then evaluate the frequency that branches appear
- The support value associated with each branch indicates the proportion of bootstrap replicates that support that specific branch. Higher support values indicate greater confidence in the presence of that branch in the true evolutionary tree.

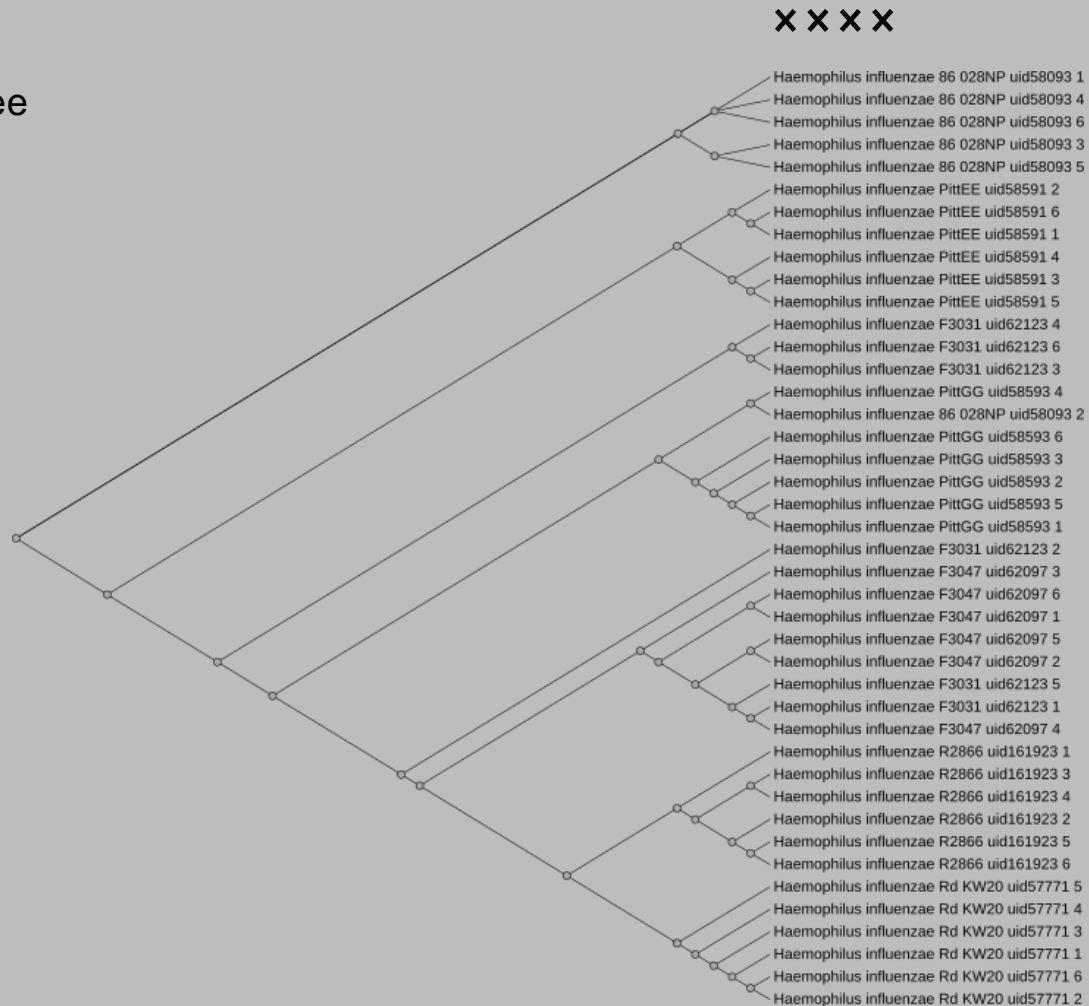


RAxML

- RAxML (Randomized Axelerated Maximum Likelihood) is an ML algorithm
- Uses randomly generated trees and parsimony based starting trees
- Uses variety of heuristic search algorithms to explore possible tree topologies and parameter values - Subtree prune and regraft (SPR) and Nearest-Neighbor Interchange (NNI)
- Iteratively adjusts the tree topology and branch lengths to improve the likelihood score
- RAxML can do bootstrapping to assess the confidence levels of each branch
- Can calculate Robinson-Foulds (RF) distances between tree iterations
- The current leading method for large-scale ML estimation
- Can require weeks for datasets with thousands of molecular sequences

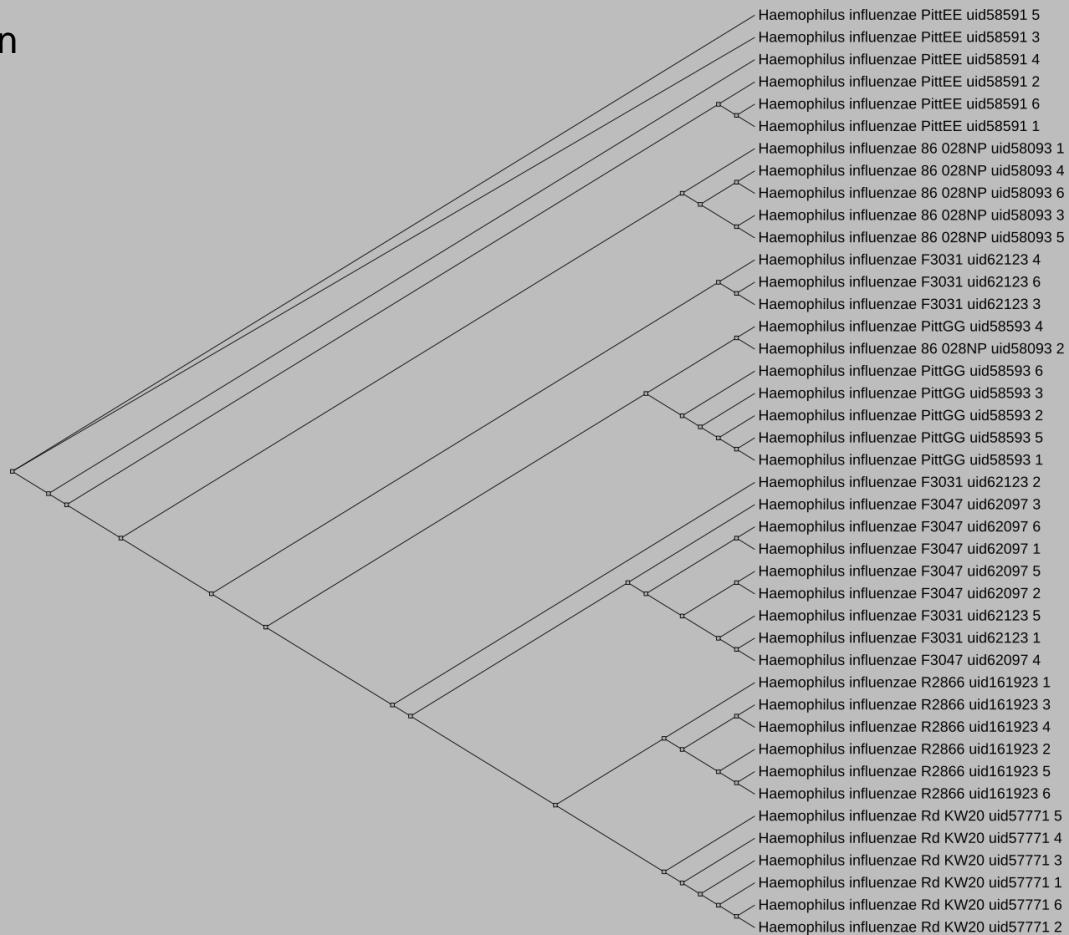


RxML – Best Tree



RAxML – Bipartition

XXXX



Phylogenetic tree showing the relationships between various *Haemophilus influenzae* strains. The tree is rooted on the left and branches out to the right. Strains are listed on the right, grouped by color-coded labels: 028NP (grey), PittEE (green), PittGG (blue), F3031 (red), F3047 (orange), R2866 (purple), and Rd KW20 (pink). Each strain name is followed by a unique identifier (uid). The tree shows the evolutionary relationships between these strains, with some internal nodes marked by small circles.

Strains listed (from top to bottom):

- Haemophilus influenzae 86 028NP uid58093 1
- Haemophilus influenzae 86 028NP uid58093 6
- Haemophilus influenzae 86 028NP uid58093 3
- Haemophilus influenzae 86 028NP uid58093 4
- Haemophilus influenzae 86 028NP uid58093 5
- Haemophilus influenzae F3031 uid62123 4
- Haemophilus influenzae F3031 uid62123 6
- Haemophilus influenzae F3031 uid62123 3
- Haemophilus influenzae PittEE uid58591 2
- Haemophilus influenzae PittEE uid58591 6
- Haemophilus influenzae PittEE uid58591 1
- Haemophilus influenzae PittEE uid58591 3
- Haemophilus influenzae PittEE uid58591 5
- Haemophilus influenzae PittEE uid58591 4
- Haemophilus influenzae 86 028NP uid58093 2
- Haemophilus influenzae PittGG uid58593 3
- Haemophilus influenzae PittGG uid58593 4
- Haemophilus influenzae PittGG uid58593 5
- Haemophilus influenzae PittGG uid58593 6
- Haemophilus influenzae PittGG uid58593 1
- Haemophilus influenzae PittGG uid58593 2
- Haemophilus influenzae F3047 uid62097 3
- Haemophilus influenzae F3047 uid62097 6
- Haemophilus influenzae F3047 uid62097 1
- Haemophilus influenzae F3031 uid62123 1
- Haemophilus influenzae F3031 uid62123 5
- Haemophilus influenzae F3047 uid62097 4
- Haemophilus influenzae F3047 uid62097 5
- Haemophilus influenzae F3047 uid62097 2
- Haemophilus influenzae F3031 uid62123 2
- Haemophilus influenzae R2866 uid161923 5
- Haemophilus influenzae R2866 uid161923 3
- Haemophilus influenzae R2866 uid161923 2
- Haemophilus influenzae R2866 uid161923 1
- Haemophilus influenzae R2866 uid161923 6
- Haemophilus influenzae R2866 uid161923 4
- Haemophilus influenzae Rd KW20 uid57771 2
- Haemophilus influenzae Rd KW20 uid57771 3
- Haemophilus influenzae Rd KW20 uid57771 4
- Haemophilus influenzae Rd KW20 uid57771 6
- Haemophilus influenzae Rd KW20 uid57771 5
- Haemophilus influenzae Rd KW20 uid57771 1



FastTree

- Algorithm used to construct phylogenetic trees from alignments of nucleotide or protein sequences
- Specifically designed to efficiently build phylogenetic trees from large datasets
- Can handle alignments with up to a million sequences
- Jukes-Cantor or generalized time-reversible (GTR) models of nucleotide evolution
- Jones-Taylor-Thornton (JJT), Whelan & Goldman (WAG), or Le and Gascuel (LG) models of amino acid evolutions
- Nearest Neighbor Interchange (NNI)
- A limitation of FastTree's neighbor joining phase is that it does not correct the distance for multiple for multiple substitutions, but it gets corrected in later steps

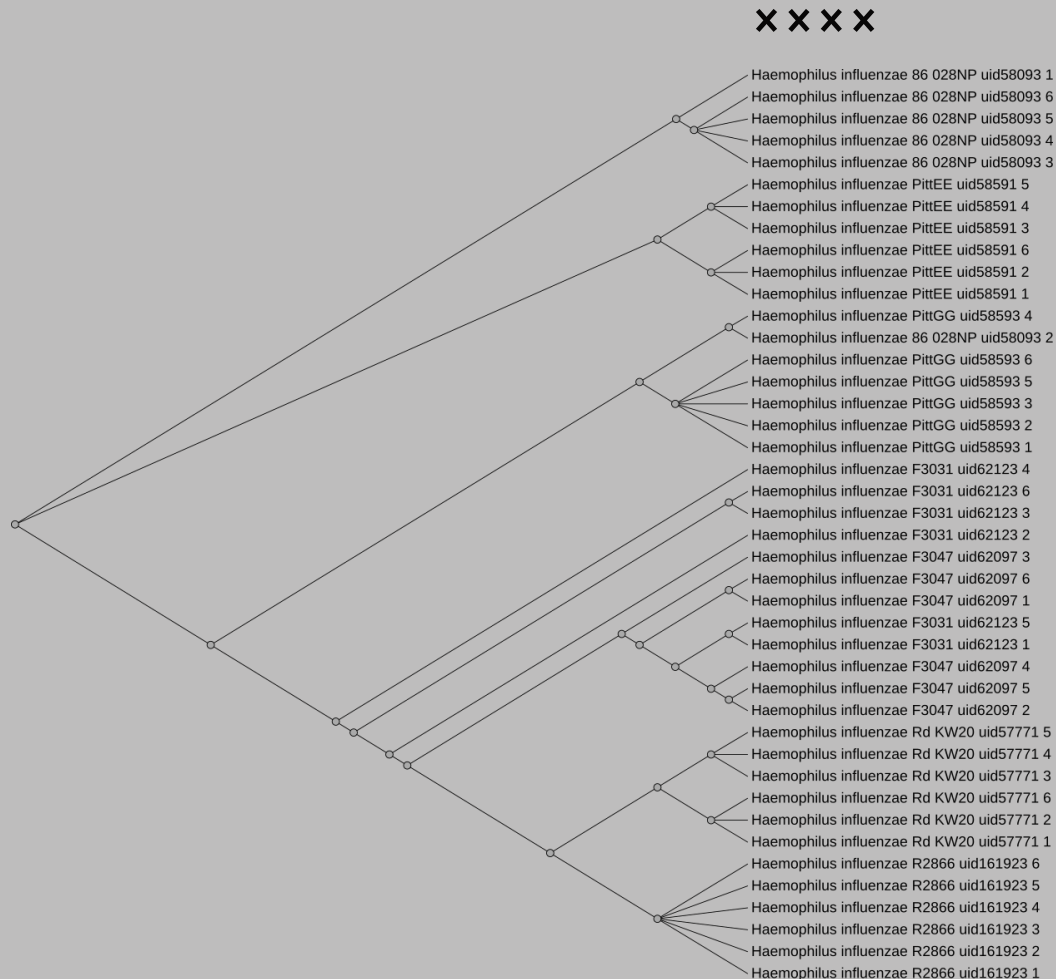


FastTree - Xsede

Rooted Tree

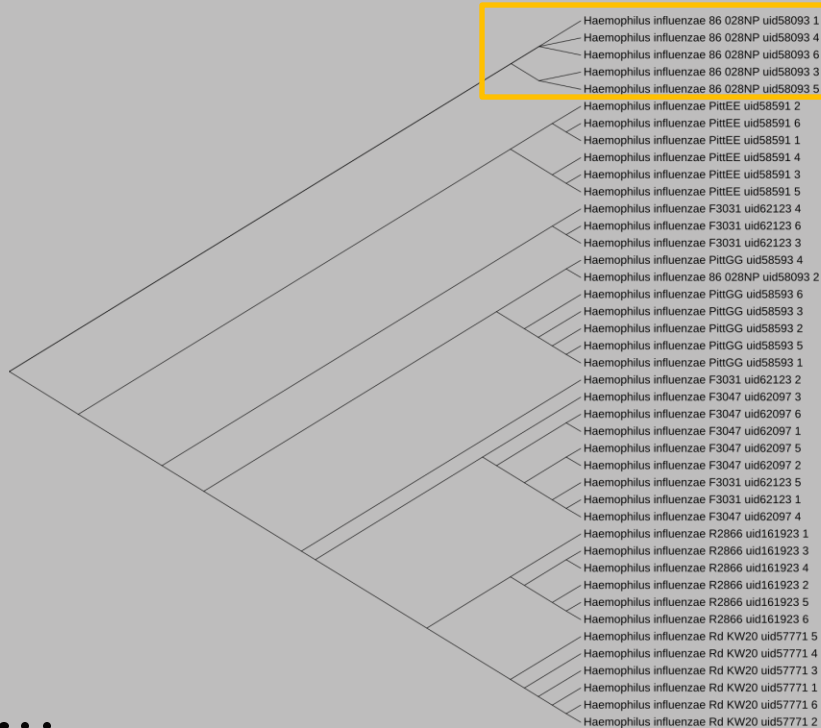
External Nodes: 42

Internal Nodes: 26



XXXX

Comparing RAxML and FastTree



RAxML – Best Tree

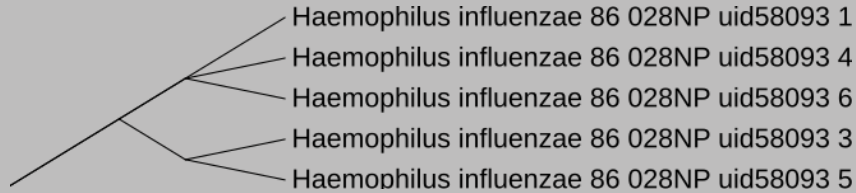


FastTree - Xsede

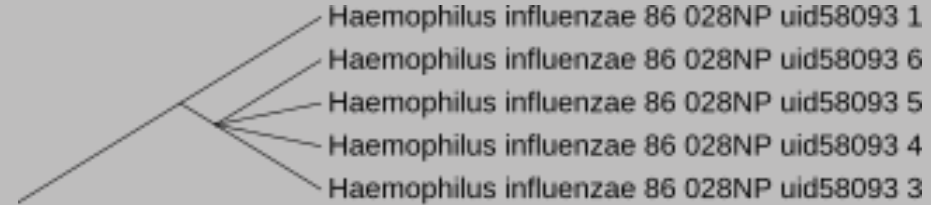
x x x x

Comparing RAxML and FastTree

RAxML – Best Tree



FastTree - Xsede



Comparing RAxML and FastTree

- RAxML generally considered more accurate and FastTree considered faster
- "We find that when RAxML and FastTree are constrained to the same running time, FastTree produces topologically much more accurate trees in almost all cases. We also find that when RAxML is allowed to run to completion, it provides an advantage over FastTree in terms of the ML score, but does not produce substantially more accurate tree topologies."



XXXX

Build a Tree using iTOL

<input type="checkbox"/>	PROCESS_OUTPUT	STDOUT	104.59 KB	View	Download
<input type="checkbox"/>		STDERR	189.00 bytes	View	Download
<input type="checkbox"/>	all_outpuffiles	stderr.txt	189.00 bytes	View	Download
<input type="checkbox"/>		RAxML_bestTree.result	3.59 KB	View	Download
<input type="checkbox"/>		scheduler.conf	114.00 bytes	View	Download
<input type="checkbox"/>		stdout.txt	104.59 KB	View	Download
<input type="checkbox"/>		start.txt	40.00 bytes	View	Download
<input type="checkbox"/>		RAxML_bootstrap.result	1.55 MB	View	Download
<input type="checkbox"/>		done.txt	49.00 bytes	View	Download
<input type="checkbox"/>		RAxML_info.result	100.94 KB	View	Download
<input type="checkbox"/>		_scheduler_stderr.txt	870.00 bytes	View	Download
<input type="checkbox"/>		RAxML_bipartitions.result	3.66 KB	View	Download
<input type="checkbox"/>		infile.reduced	37.17 KB	View	Download
<input type="checkbox"/>		term.txt	314.00 bytes	View	Download
<input type="checkbox"/>		RAxML_bipartitionsBranchLabels.result	3.73 KB	View	Download
<input type="checkbox"/>		infile	75.56 KB	View	Download
<input type="checkbox"/>		_JOBINFO.TXT	402.00 bytes	View	Download

<input type="checkbox"/>	PROCESS_OUTPUT	STDOUT	166.00 bytes	View	Download
<input type="checkbox"/>		STDERR	1.28 KB	View	Download
<input type="checkbox"/>	all_results	logfile.txt	38.10 KB	View	Download
<input type="checkbox"/>		stderr.txt	1.28 KB	View	Download
<input type="checkbox"/>		scheduler.conf	76.00 bytes	View	Download
<input type="checkbox"/>		stdout.txt	166.00 bytes	View	Download
<input type="checkbox"/>		start.txt	40.00 bytes	View	Download
<input type="checkbox"/>		fasttree_result.tree	2.40 KB	View	Download
<input type="checkbox"/>		done.txt	49.00 bytes	View	Download
<input type="checkbox"/>		_scheduler_stderr.txt	870.00 bytes	View	Download
<input type="checkbox"/>		term.txt	314.00 bytes	View	Download
<input type="checkbox"/>		infile	75.56 KB	View	Download
<input type="checkbox"/>		_JOBINFO.TXT	347.00 bytes	View	Download

Download output data from CIPRES



x x x x

Build a Tree using iTOL

Upload a new tree

Tree name:

Paste your tree into the box below, or select a file using the **Tree file** selector. You can also simply drag and drop the tree file onto the page (only a regular plain text file, not QIIME QZA files).

Tree text:

Tree file:




 No file chosen

Interactive Tree of Life: <https://itol.embl.de/upload.cgi>



XXXXX

Build a Tree using iTOL

Control panel   

Basic **Advanced** Datasets Export



Mode Circular Rectangular Unrooted




Mode options Rotation: 0° Invert tree: Yes No Slanted: Yes No

Labels Display Hide

Label options Font: Arial Font style: 20 px Alignment: Left Right Rotation: On Off Shift: 0 px

Branch options Line style: 1 px Color gradient: On Off Dashed lines: 0.3 px

Tree views  Undo  Reset tree

Control panel   

Basic Advanced **Datasets** Export

Scaling factors: 1 x horiz. 1 x vert.



Leaf sorting: Default None

Invert sort order: Yes No

Branch metadata display Node IDs: Display Hide

Node options Leaf node symbols: Display Hide Internal node symbols: Display Hide no class available Collapse

Other functions Label functions: Multi-style Bulk edit Auto assign taxonomy: NCBI GTDB Root the tree midpoint: Midpoint root




Tree views  Undo  Reset tree

Edit the tree





xxxxx

Build a Tree using iTOL

Control panel   



Basic **Advanced** **Datasets** **Export**






Format: PNG: Portable Network Graphics 

Resolution: 120 dpi 

Export area: **Screen** **Full image**

File name: optional

 Share  **Export**

 Tree views   Undo   **Reset tree**

Export the Tree



x
x
x
x
x

Q&A

