

Ancestral Reconstruction

Phil Hughes (12346487), Saya Maratova (12344373), Alexander Veith

Background

Example Case

What is Ancestral Reconstruction?

Why do we do it?

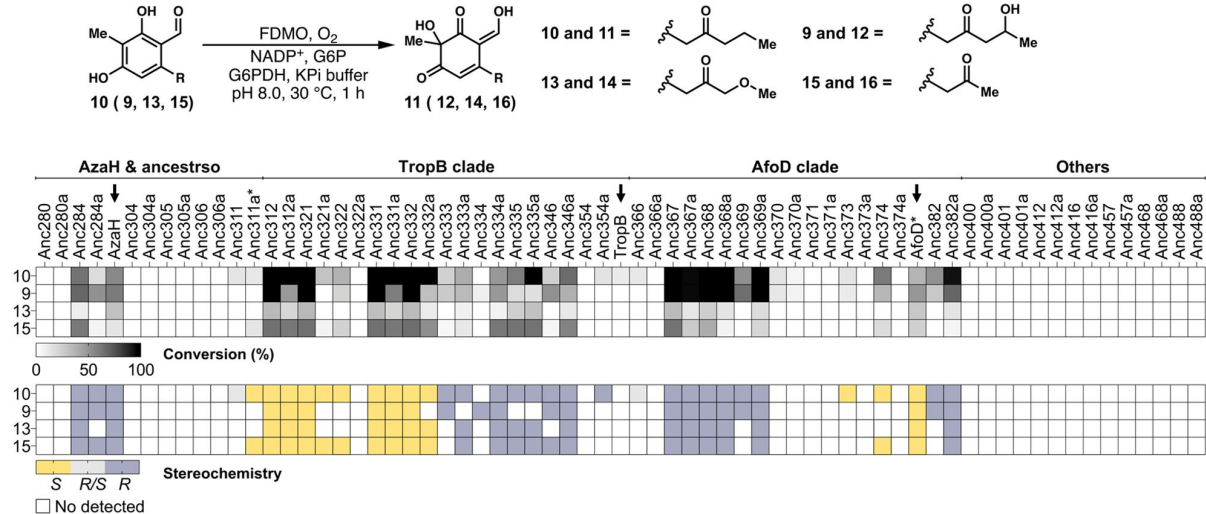
What do you need?

explain/mention packages (phangorn, ape)

Deciphering the Evolution of Flavin-Dependent Monooxygenase Stereoselectivity Using Ancestral Sequence Reconstruction

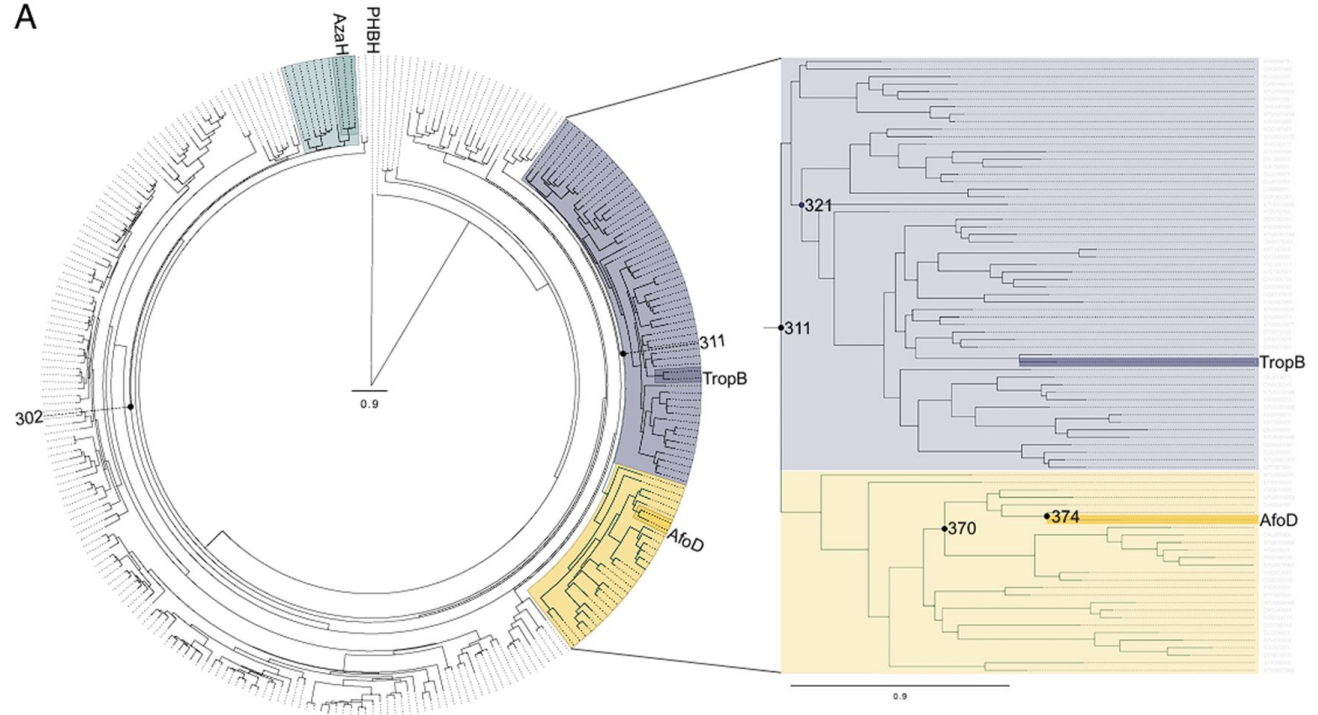
Example Case

- ASR Completed
- Enzyme activity screened, with predictions
- Stereocontrol mechanism of AzaH revealed
- Residues changed in enzyme to switch enantiomeric preference



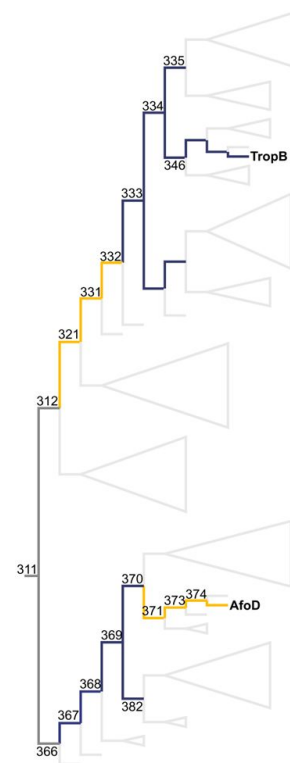
What is AR?

- Extant sequences compared
- Phylogenetic Tree constructed
- Predictions made for ancestors



Why do we do it?

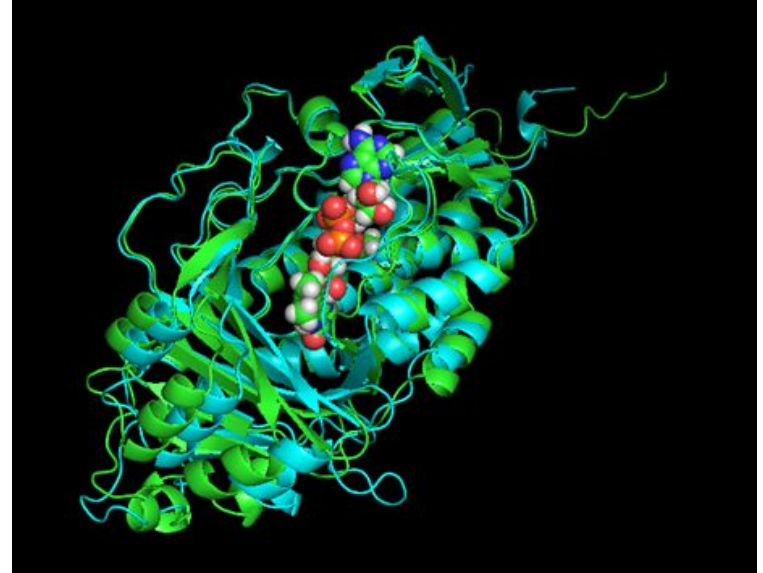
- Produce likely functional proteins
- Different properties
- Base for further protein engineering
- Identify mutations leading to different properties
- Chiral Products



Node	5	5	9	9	1	2	2	2	2	2	2	2	2	3	3	3	3	4	(R)-11: (S)-11
	4	5	6	8	1	0	2	2	3	3	3	5	2	3	3	3	9	2	
					9	6	6	8	5	7	9	0	2	9	0	1	2	7	
311	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
311a	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
312	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
312a	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
321	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
321a	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
331	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
331a	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
332	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
332a	I	A	L	W	F	R	Q	M	H	L	F	V	F	P	H	H	G	Y	W
333	I	A	L	W	F	R	Q	M	H	L	Y	T	F	P	H	H	G	Y	W
333a	I	A	L	W	F	R	Q	M	H	I	Y	T	F	P	H	H	G	Y	W
334a	I	A	L	W	F	R	Q	M	H	I	Y	T	F	P	H	H	G	Y	W
346	I	A	L	W	F	R	Q	M	H	I	Y	T	F	P	H	H	G	Y	W
346a	I	A	L	W	F	R	Q	I	H	I	Y	T	F	P	H	H	G	Y	W
TropB	M	A	L	W	F	R	L	M	H	I	Y	A	F	P	H	H	G	Y	W
335	I	A	L	W	F	R	Q	M	H	I	Y	T	F	P	H	H	G	Y	W
335a	I	A	L	W	F	R	Q	M	H	I	Y	T	F	P	H	H	G	Y	W
354a	I	A	L	W	F	R	F	N	N	L	Y	T	F	P	H	H	G	H	W
366	I	A	Y	Y	F	R	Q	M	H	L	F	V	F	P	F	H	G	Y	W
367	I	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
367a	I	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
368	F	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
368a	F	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
369	F	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
369a	F	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
382	F	A	N	Y	Y	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
382a	F	A	N	Y	Y	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
370	F	A	N	Y	F	R	Q	M	H	L	Y	V	F	P	F	H	G	Y	W
371	F	A	V	Y	F	R	Q	M	H	V	F	F	F	P	F	H	G	Y	W
373	F	A	V	Y	F	R	Q	M	H	V	F	F	F	P	F	H	G	Y	W
374	F	A	V	Y	Y	R	Q	A	H	V	F	F	F	P	F	H	G	Y	W
AfoD	F	A	V	Y	Y	R	Q	A	H	V	F	F	F	P	F	H	G	Y	W

Requirements for AR

- Suitable library of existing sequences (nucleotide or Amino acid)
- A matrix; assessing the likelihood of different changes occurring
- Predicting Deletions, Insertions, Substitutions etc
- Software!



Alignment of AfoD (green) with FAD (crystal structure against an AlphaFold generated protein structure of Anc311(blue), aligned in PYMOL. (AA sequence from paper supplementary)

Parsimony

- Tries to generate a tree that is the most likely; with the fewest nodes and least extreme or unlikely changes; minimal steps
- Trees constructed with nodes; 'deeper' in the tree are 'internal nodes'
- Maximum parsimony; formation of a tree with the fewest evolutionary changes (most economical evolutionary path)

Gap occurrences (deletions, insertions); major evolutionary events

Gap may be treated as a character or 'additional state'; how they are treated will affect tree topology

Appropriate 'Matrix' must be selected that gives appropriate significance to gap events

Marginal Reconstruction vs Joint Reconstruction

Joint Reconstruction- 'Maximise the overall likelihood' over the whole tree; works it out simultaneously; selects the joint most likely

- Focus is the 'whole tree'; individual nodes are considered less important
- Can lead to inaccuracy for individual nodes
- Best for looking at full genomes or gene sequences

Marginal Reconstruction- Most likely on a node-by-node, or site-by-site basis individually

- Reconstructions for nodes may not jointly all fit together
- 'Up the tree and back again'
- Best for looking at particular positions or sites (Cysteine residues later!)

Both aspects of Maximum Likelihood