Consider a hypothetical alignment of a conserved protein, but one taxa may have errors.

	1	2	3	4	5	6	7	8	9	10
	R	E	K	V	L	I	D	T	N	V
	R	E	K	V	M	I	D	S	S	V
	K	E	K	V	L	I	D	G	S	V
	R	S	R	H	I	A	A	S	R	V
	K	E	K	I	L	I	D	G	S	V
	R	E	K	I	L	I	D	S	S	I
	K	E	K	V	F	I	D	G	Q	V
	K	E	K	V	F	I	D	T	S	V

Consensus
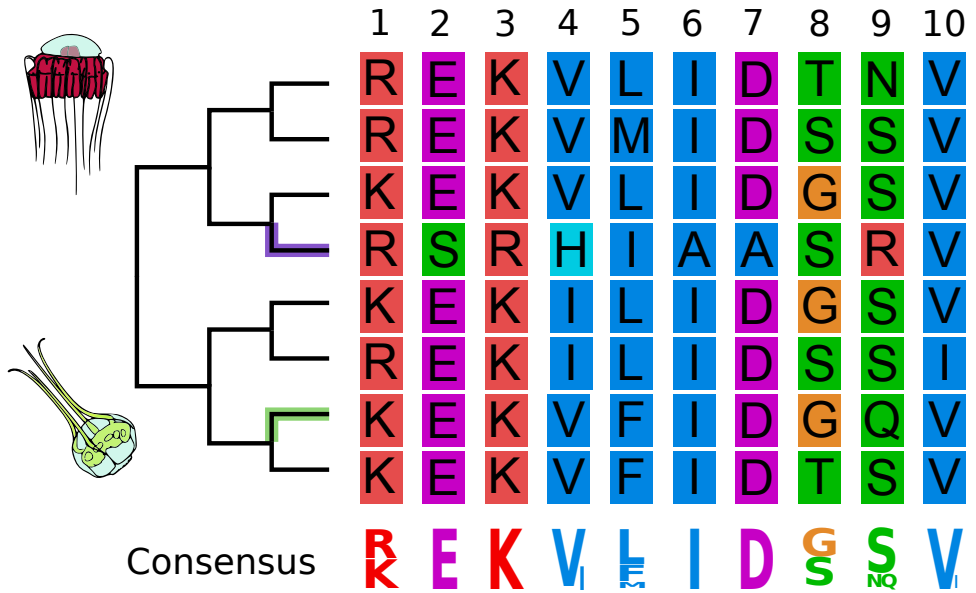
Real alignments alternate between constant and variable sites every few amino acids. This feature is used to set the penalty threshold.

The penalty is counted as a running total of rarity of an amino acid at some position. If this passes a threshold, then the species is flagged for that protein.

Penalty

R S R H I A A S R V

The run of SRHIAASR is too improbable, and gets flagged

Penalty

K E K V F I D G Q V

The amino acids vary from the consensus, but in common ways