

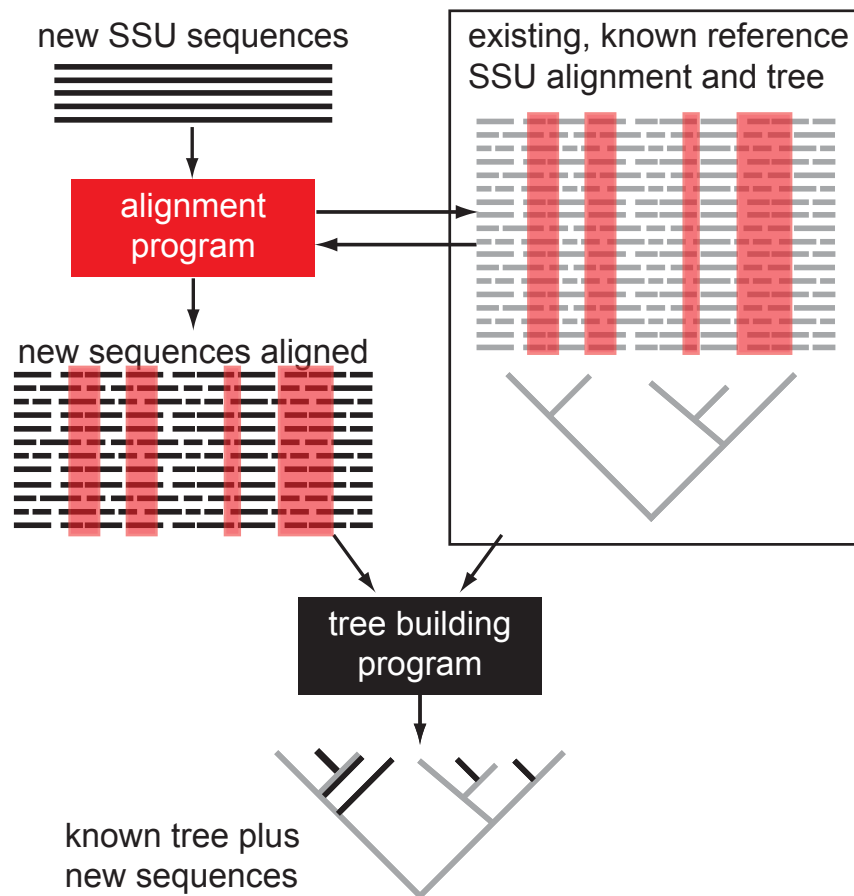
Two types of questions:

What organisms (known/unknown) are in my sample?

What is the phylogeny of a set of organisms?

Main assumption:

SSU gene tree approximates organismal tree.



Goals of the alignment program:

accurate: b/c alignment errors confound phylogenetic inference

fast and scalable: to handle up to millions of seqs

flexible: to be useful for all 3 domains

Sampling of recent SSU studies			
environment/ phylogeny	domain(s)	#seqs	year
soil, many others	bacteria	21,752	2007
cecal microbiota of mice	bacteria	5,088 4,157	2005 2006
Sargasso sea	all 3	1,164	2004
hydrothermal vents	eukarya	374	2002
endolithic environment (pore space of rocks)	archaea, bacteria	342 588	2005 2007
oxidized iron deposits, marine tidal mat, microbial steamers	bacteria	308	2004
soil & burrow casts of earthworms	archaea, bacteria	204	2002
tidal flat sediment	archaea	90	2005
salt marsh	eukarya	79	2003
dipteran hindgut	bacteria	59	2007
<hr/>			
bumble bee phylogeny	eukarya	~200	2007
anaplasma phylogeny	bacteria	21	2003
protostome phylogeny	eukarya	20	2002