

The phyloseq package and analysis of high throughput amplicon sequencing data

Paul McMurdie
Statistics Department
Stanford University

BioC 2012

July 24-25, 2012 (Developer Day: July 23)

Fred Hutchinson Cancer Research Center - Seattle, WA



Workshop Outline:

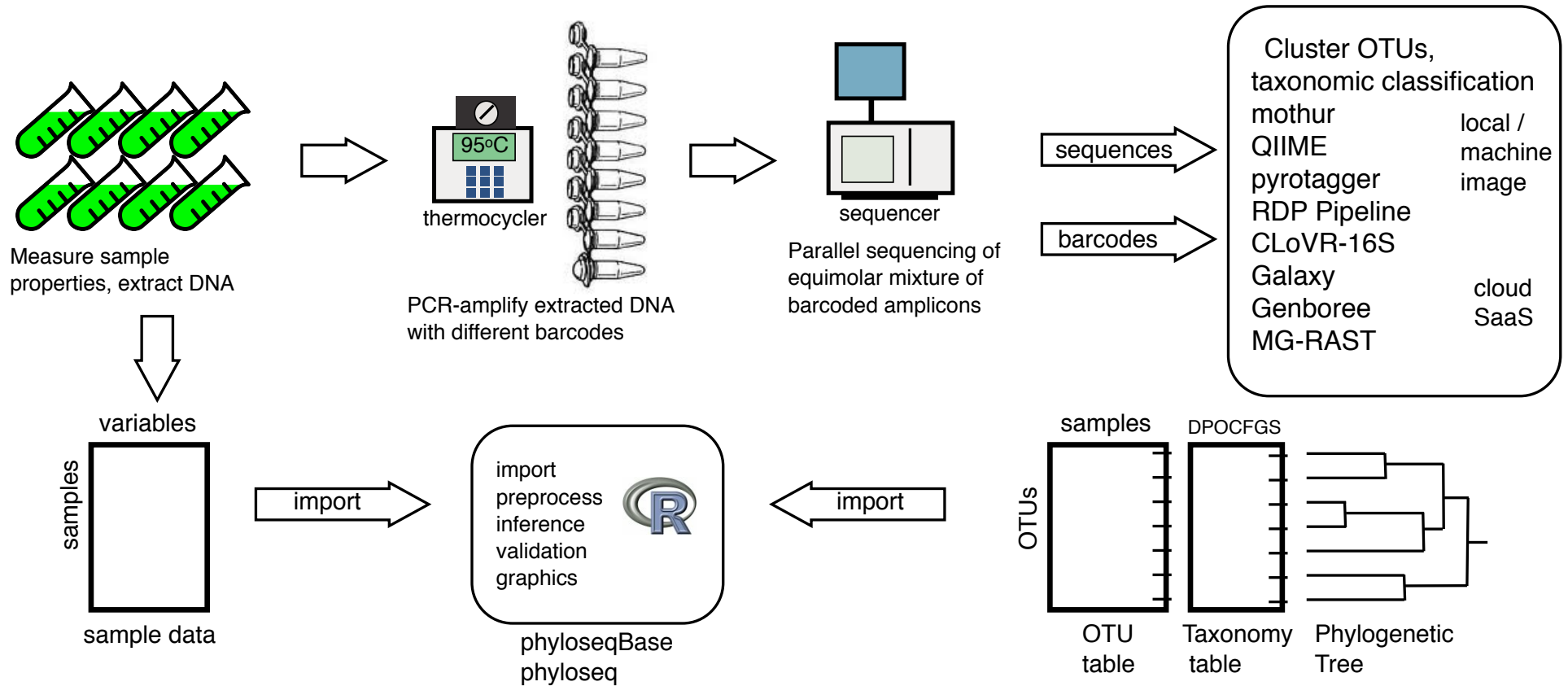
- Background of modern phylogenetic sequencing (if needed).
- Motivation, Design, and Philosophy of phyloseq

live
code

- How to import data with phyloseq
- Basic interaction with data and simple summary graphics
- Data preprocessing using phyloseq tools
- More complex exploratory/summary graphics, including ordination
- Validation tools supported in phyloseq
- Additional validation/testing using other R tools
(getting data components to other R functions)

Amplicon Sequencing

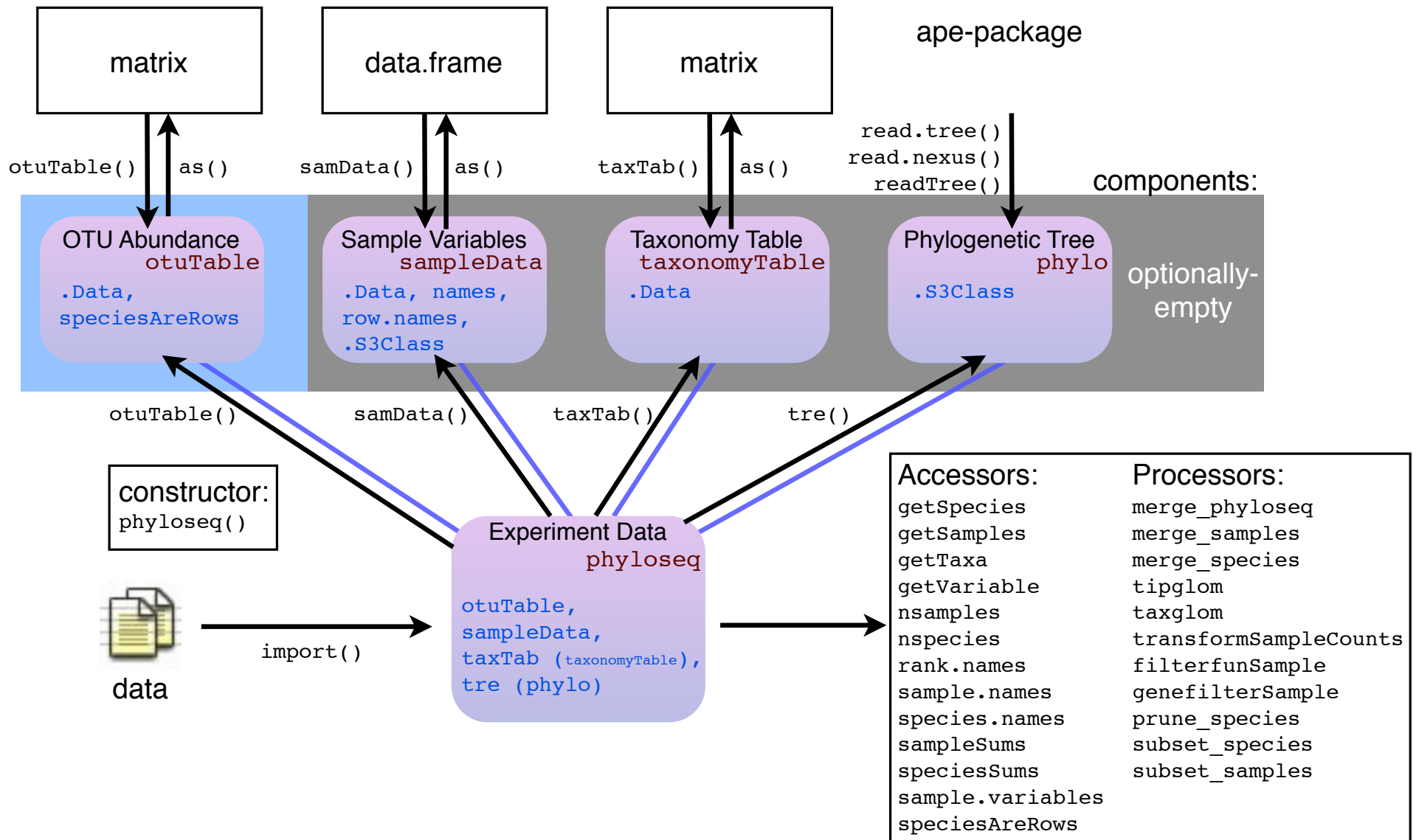
Overview of amplicon sequencing and analysis



phyloseq design and features

- A single, explicitly-defined S4 class that can store the different data types of a phylogenetic sequencing experiment in a single object.
- Importers all create this special “phyloseq” class
- Most phyloseq functions will act on this experiment-level object. It doesn't need to be diced-up to work. (Keep data together)
- Internal tools check validity and agreement among components of an experiment. Helps prevent mistakes.
- Plotting tools for creating quality graphics, built using [ggplot2](#).
- Example datasets from real published data, with references, documentation, and examples.
- Examples using other R tools after importing with phyloseq

phyloseq classes and “data infrastructure”



phyloseq accessors

Function	Description
[Standard extraction operator. works on otuTable, sampleData, and taxonomyTable
access	General slot accessor function for phyloseq-package
getslots.phyloseq	Return the slot names of phyloseq objects
getSpecies	Returns the abundance values of sample 'i' for all species in 'x'
getSamples	Returns the abundance values of species 'i' for all samples in 'x'
getTaxa	Get a unique vector of the observed taxa at a particular taxonomic rank
getVariable	Returns an individual sample variable vector/factor
nsamples	Get the number of samples described by an object
nspecies	Get the number of species (taxa) described by an object
otuTable	Build or access otuTable objects
rank.names	Get the names of the available taxonomic ranks
sampleData	Build or access sampleData objects
sample.names	Return the names of the samples described by an object
species.names	Return the names of the species described by an object
sampleSums	Returns the total number of individuals observed from each sample
sample.variables	Returns the names of sample variables in an object
speciesSums	Returns the total number of individuals observed from each species
speciesAreRows	Returns the orientation of the abundance table
taxTab	Build or access taxTab objects
tre	Access the tree contained in a phyloseq object

phyloseq constructors

Functions for building component data objects

Function	Input Class	Output Description
<code>otuTable</code>	numeric matrix	<code>otuTable</code> object storing taxa abundance
<code>otuTable</code>	<code>data.frame</code>	<code>otuTable</code> object storing taxa abundance
<code>sampleData</code>	<code>data.frame</code>	<code>sampleData</code> object storing sample variables
<code>taxTab</code>	character string	<code>taxonomyTable</code> object storing taxonomic identities
<code>tre</code>	file path char	phylo4-class tree, read from file
<code>tre</code>	phylo-class tree	phylo4-class tree, converted from argument
<code>read.table</code>	table file path	A matrix or <code>data.frame</code> (Std Rcore function)
<code>read.tree</code>	Newick file path	phylo-class tree object (ape)
<code>read.nexus</code>	Nexus file path	phylo-class tree object (ape)
<code>readNexus</code>	Nexus file path	phylo4-class tree object (phylobase)

Functions for building complex data objects

Function	Input Class	Output Description
<code>phyloseq</code>	2 or more component objects	phyloseq-class, “experiment-level” object
<code>merge_phyloseq</code>	2 or more component or phyloseq-class objects	Combined instance of phyloseq-class

Overview of analysis using phyloseq

