

BIS101 F2013 Lecture 7: Genomics and Phylogenetics

Reading

Genomics

what is "the genome"?

what's it good for:

- reference for understanding function, mechanism, evolution
- a "map" on which you place other information
- show browser

sequencing costs. human genome cost \$3 billion dollars

- now to **resequence** genome only \$4K.
- by time any of you is a Dr. human genom seq. < \$1K (cheaper than any genetic test)

What can you do w/ sequencing ?

- learn the sequence: detect **SNPs** and other variants
- bisulfite seq. - methylated C's -> T seq. compare to normal
- transcriptome
- proteome
- nuclease digestion
- CHip-seq seq. all DNA bound by protein

I can sequence your entire genome in 27 hours. In very near future, you will go to Dr. and to diagnose they may:

- sequence your transcriptome (what is?) at time series along the day to identify genes regulated oddly or responding to infection etc.
- multiple tissues: sequencing tumors to figure out what is wrong or find mutation
- methylome, etc. etc.
- proteome to see if some protein level is off.

good example of why programming important for biology -- genetic analysis is a computational discipline

- i sequence your genome, I get 90Gb of A,C,T,G plus related information -- 200Gb of data to analyze. can't do that in excel.

- show fastq file.

How to sequence a genome

You can't just sequence end to end each chromosome.

Whole-genome shotgun -- draw chromosome, sequences, assembly.

Resequencing is easier (but still tricky) as you are mapping to "known" reference.

- as we saw w/ Liu et al. paper, ref. is not static!
Annotation

so how do i know where the genes are? how to annotate ?

- RNA sequence, etc.

Phylogenetics

Study of evolutionary relationships among species. Will not get into much detail here.

Usually use gene sequences to determine relatedness. For example:

Raccoon

ATGGGA

Bear

ATGGGA

Cat

AACGGT

Mongoose

AACGGA

Platypus

GTCTTA

Parsimony method, but really people use whole genomes and very complex statistical methods.

Draw tree, label nodes. Define **outgroup** allowing you to put an ancestral root on the tree.

Does this mean platypus is "basal" or "ancestral" to all mammals ? No.

Useful for understanding when things evolved, for identifying conservation, for studying gene origin and evolution.

Gene duplication

Homolog closely related gene.

Two kinds.

ortholog **paralog**

separated by speciation vs. separated by duplication

draw examples.

duplication important form of change.

define **synteny**

draw how synteny can inform which gene is ortholog.

fates of a new gene duplicate: Sub vs. neo vs. pseudo

what would happen to a pseudogene over time?

Gene evolution

Purifying selection removes new deleterious mutations. Important genes with functions will be under stronger ? or weaker purifying selection?

How to identify.

Define **nonsynonymous** and **synonymous** changes.

If both have no effect on function, expect equal proportions. Can compare gene sequences along a phylogeny and count nonsyn and syn changes. if $dN/dS \ll 1$, constraint! if $dN/dS \gg 1$, rapid adaptation!