1 Representing Trees

2 Tree Building

3 Positive Seletion

(a)

i

Long haplotypes indicate that a derived locus has risen to prominence in a population on a shorter time scale than recombinative disruption. For a large population this is unlikely to happen unless selection is at work.

ii

Because in a long haplotype many individual polymorphisms may have been derived and hitchihiked to prominence by chance alone.

iii

cM is preferable here because rates of recombination can vary across the chromosome - recombination frequency measure a genetic distance with recombination taken into account. (see figs/3aiii.png)

iv

C, Europe appears to be under selection as it scores high in EHH in comparison with each population.

v

79 snps appear above 2.0.

(b)

(i)

Back of the envelope: the ancestor is about halfway between chimp and human thus .66% of the traits that we label "ancestral" will have in fact been derived by the chimp. For these we will usually observe the true ancestral trait in humans which we will misidentify as derived.

ii

See figs/3bii.png

iii

See figs/3biii.png

(c)

i

$$\begin{aligned} H_S &= \overline{p \times (1-p)} \\ &= \frac{(p+d) - (p+d)^2 + (p-d) - (p-d)^2}{2} \\ \implies F_{ST} &= \frac{H_T - H_S}{H_T} \\ &= \frac{p(1-p) - (p - p^2 - d^2)}{p(1-p)} \\ &= \frac{d^2}{p(1-p)} \end{aligned}$$

4 Variant Discovery

ii

Can estimate H_T from $\frac{\#\text{heterozygotes}}{\#\text{total}}$ and estimate H_P over subpopulations. Using these numbers, can compute F_{ST}. Given H_T compute p:

$$2p^2 + 2p + H_T = 0 \implies p = \frac{-2 \pm \sqrt{4 - 4(2)(H_T)}}{4}$$

and compute d:

$$d = \sqrt{p(1-p)F_{ST}} \tag{1}$$

5 Coalescence