What’s in a Mutt?
An Intro to Dog DNA Analysis

Lecture 6
Jan 18th, 2019
Tasha’s genome
Tasha’s genome
Tasha’s genome

Illumina HiSeq 2000
Sequencing by Synthesis

1. Attach
2. Amplify
3. Image

x ~100
Tasha’s genome

Billions of ~250 base reads

GATT    ACCG    TTAC
CGCA    CCGC    ATTA
GATT    GTTA    CGCA
Tasha’s genome

Billions of ~250 base reads

Assembled genome
Sequence Alignment

...CCATAG TATGCGCCC CGGAAATT TT CGGTATAC
...CCAT CTATATGC CG CGGAAATT TT CGGTATAC
...CCAT GGCTATATG CTAT CGGAAA AA GCGGTATA
...CCA AGGCTATAT CCTATCGGA TT TCGGTA C...
...CCA AGGCTATAT GCCCTATCG TTTGCGGT C...
...CC AGGCTATAT GCCCTATCG AAATTG GC ATAC...
...CC TAGGCTATA GCGCCCTA AAATTG GC GTATAC...
...CCAT AGGCTATAT GC GCGGCA GCTA TT GCGGTATA...
We found a SNP!
Finding the best alignment

How might we tell if one alignment is better than another?
Finding the best alignment

- Number of mismatches
- Type of difference
- Base “quality”
  - Sequencer reports how confident it is in the base “call”

1 mismatch
...TGATCATATA... → ...TGATcATAA...
GATCAAA
Mismatches are low quality (lowercase)

2 mismatches
...TGATATTA... → ...TGATCATA...
GATcAT
Mismatches are high quality bases

...TGATCATATA... → ...TGATcATAA...
GATCAAA
Mismatches are low quality (lowercase)

...TGATATTA... → ...TGATCATA...
GATcAT
Mismatches are high quality bases
# Alignment and phasing

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>A</th>
<th>T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tasha (Reference)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
Alignment and phasing

Tasha (Reference)

C A T

A

G

G

G

G

A

G

T

T

T

G

A

C

G

G

C G G

A A T

?
Alignment and phasing

Tasha (Reference)

Alignment and phasing
Alignment and phasing

Tasha (Reference)
Getting to 2.5 million SNPs
Which SNPs are interesting?

HCRTR2 gene

G

G

G

G

A

A

A

A
Discovering cause of fever illness in Shar-Peis

Creature Feature: Shar-Pei dog

By Alice McCarthy

Scientists at the Broad Institute and Uppsala University in Sweden have made a discovery in the Shar-Pei dog breed that may help explain the cause of repetitive human fever illnesses. In a recent paper, the team published that a genetic regulatory gene mutation causes the dogs to develop Familial...
Following strong selection for the “wrinkled” skin phenotype, Shar-Pei dogs in the western world most commonly present as the meatmouth type (A–C). The traditional type of Shar-Pei (D) is the ancestral version and is still common in China. The characteristic skin is a result of a deposition of mucin, mainly hyaluronic acid (HA), in the upper dermis of the skin. The deposit collects in certain areas of Shar-Pei skin and often as “socks” around the hocks (E). The meatmouth Shar-Pei (A–C) is also predisposed to a breed-specific periodic fever syndrome called Familial Shar-Pei Fever (FSF). (Olsson et al, 2011).
Genome wide association study (GWAS)

Regions with high homozygosity, when comparing 50 shar-peis to 24 other breeds (230 dogs). Regions of chromosome 13 in shar-peis have 10-fold less heterozygosity than other breeds.

Olsson et al, 2011
Genome wide association study (GWAS)

Genome wide SNP associations with Familiar Shar-Pei Fever. Strongest (and only statistically significant) association is on chromosome 13.

*Manhattan plot:* x-axis plots all SNPs as a point, ordered by chromosome position. y-axis plots $p$-value of significance (log scaled).

Olsson et al, 2011
GWAS in Shar-Peis; summary

“Shar-Pei dogs have two unique features: a breed defining “wrinkled” skin phenotype and a genetic disorder called Familial Shar-Pei Fever (FSF). The wrinkled phenotype is strongly selected for and is the result of excessive hyaluronan (HA) deposited in the skin. ... FSF is characterized by unprovoked episodes of fever and/or inflammation and resembles several human autoinflammatory syndromes. Here we show that the two features are connected and have the same genetic origin, a regulatory mutation located close to a HA synthesizing gene (HAS2). ... HAS2 was previously not known to associate with autoinflammatory disease, and this finding is of wide interest since approximately 60% of human patients with periodic fever syndrome remain genetically unexplained.” (Olsson et al 2011)
Genome wide association study (GWAS)

Genome wide association study (GWAS)

Associations with granulomatous colitis in boxers and bulldogs, Hayward et al 2016.

Associations with idiopathic epilepsy in Irish wolfhounds, Hayward et al 2016.