An update on ADAPTmap

Alessandra, on behalf of the consortium

What is ADAPTmap

An initiative, not a funded project:

Meaning that researchers voluntarily put data, expertise, money, time, facilities, etc... on the table to obtain its goals

Centralized data sharing: for greater understanding of the goat genome and its multi-purpose use with a strong emphasis on adaptation to local environments

Keywords: integration of people and data

ADAPTmap: contributing projects

• The beginning.
  • The consortium grows larger...and more to come!

ADAPTmap in 2015 (I)

Collecting genotypic and “phenotypic” (e.g. anything that is not genotypes) data

• Fine-tuning of the ADAPTmap exchange server
• Reception of genotypic data in any imaginable format.
• Reception of phenotypic data in un-imaginable formats. Semi-strict only on (even approx.) GIS coordinates, when possible. +63% of the dataset samples carries GIS coordinates.
• Few (e.g. Less 100) has pictures or other phenotypes available.

ADAPTmap in 2015 (II)

Finalizing legal aspects and publication issues with partners

• Solved mostly in 2014, only few details were required for new partners.
• Currently, PHASE 1 has 18 contributing projects (or group of projects) and +20 MTA’s signed.

ADAPTmap in 2015 (III)

Standardization of data and definition of a final dataset

• 98% of the coordination work in 2015
• Final dataset (Nov 1st 2015):
  • 4117 54k genotyped individuals from 127 populations sampled on 33 countries (5 continents)
  • PLINK (ped/map) format – Alleles in TOP strand. SNPs mapped on latest CHI 1.0 assembly (in collaboration with IGGC)
  • GIS coordinates on 2605 samples
ADAPTmap in 2015 (IV)

Definition of working groups and leaders

WG1 (D. Bickhart – USDA-ARS): “Improvement of genome assembly”
WG2 (B. Rosen – USDA-ARS): “Genome annotation”
WG3 (R. Sayre – VSU): “Comparative genomics (with other ruminants)”
WG4 (G. Tosser-Klopp – INRA + C. VanTassel – USDA-ARS): “Study of the possibility to improve the current SNP panel”
WG5 (E. Nicolazzi – PTP): “Identification of parentage SNP panel”
WG6 (NO LEADER): “Integration, standardization and visualization of genomic data”
WG7 (L. Colli – UNICATT-IGC): “Population genetics analyses and population history”
WG8 (P. Crepaldi – UNIMI-IGC + S. Joost – EPFL): “Selection signatures (incl. landscape genomics & visible genetic profile)”
WG9 (R. Rupp – INRA): “Breeding and genetic improvement”
WG10 (NO LEADER – ON HOLD): “Sequencing data analysis”

.. and a simple example of the lot of cool stuff to do...

<table>
<thead>
<tr>
<th>CHIR 1.0 annotation LiftOver</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Look at CHIR1.0-genes split in our assembly</td>
</tr>
</tbody>
</table>

Comparative mapping with sheep

<table>
<thead>
<tr>
<th>O. aries exons</th>
<th>Exclusive New</th>
<th>Exclusive CHIR1.0</th>
<th>Shared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Mappings</td>
<td>9,534</td>
<td>564</td>
<td>225,365</td>
</tr>
<tr>
<td>Unmapped</td>
<td>564</td>
<td>9,534</td>
<td>9,140</td>
</tr>
<tr>
<td>Split Exons</td>
<td>592</td>
<td>1,930</td>
<td>1,528</td>
</tr>
</tbody>
</table>

Annotation comparisons

Annotation pipeline

Just a quick peek on (not all) the data.. and a simple example of the lot of cool stuff to do…
ADAPTmap in 2015 (IV)
Definition of working groups and leaders

WG1 (D. Bickhart – USDA-ARS): “Improvement of genome assembly”
WG2 (B. Rosen – USDA-ARS): “Genome annotation”
WG3 (B. Sayre – USDA-ARS): “Comparative genomics (with other ruminants)”
WG4 (G. Tosser-Klopp – INRA & C. VanTassel – USDA-ARS): “Study of the possibility to improve the current SNP panel”
WG5 (E. Nicolazzi – PTP): “Identification of parentage SNP panel”
WG6 (NO LEADER): “Integration, standardization and visualization of genomic data”
WG7 (L. Colli – UNICATT-IGC): “Population genetics analyses and population history”
WG8 (P. Crepaldi – UNIMI-IGC + S. Joost - EPFL): “Selection signatures (incl. landscape genomics & visible genetic profile)”
WG9 (R. Rupp – INRA): “Breeding and genetic improvement”
WG10 (NO LEADER – ON HOLD): “Sequencing data analysis”

Kick off of activities
WG5 – Identification of parentage SNP panel

Objective: Design an international parentage panel – Needs to be accurate in the largest possible number of populations. ADAPTmap fits perfectly!

Virtual call on November 17th
• 2 groups (UNIMI-PTP and INRA) had developed 2 different approaches to identify the best parentage SNP panel.
• Comparison of performance on a common – reduced dataset (March 2016)
• Decision on the method to follow and final tests on full dataset (July 2016)
• Presentation by Gwenola

SNP selection for parentage assessment

Three-steps procedure to identify SNPs for parentage assessment.
This procedure is not relying on SNP positioning and on direct linkage disequilibrium exclusion.
Analysis were performed on pairwise comparisons among animals (a total of 193,131 pairs)

1. Parent-Offspring pairs identification
2. Pe/Pi and Multivariate SNP selection
3. Mendelian errors (ME) and Final reduction

SNP selection for parentage assessment

Preliminary results on the Italian Goat Consortium dataset;
Identified two panels:
– 68 SNPs with no false positive;
– 154 SNPs with both no false positive and negative.
Further analysis will be performed after the expansion of the dataset with other breeds

WG6 – Standardization, integration and visualization of genomic data

http://bioinformatics.tecnoparco.org/SNPChimp/

Standardization

All the software used for standardization and integration of ADAPTmap data is open-source and available online at:
https://github.com/nicolazzie/SNPChimpRepo.git

Conversion of software from Illumina ROW and MATRIX format to PLINK, conversion of allele coding from/to any existing coding, update of genomic map

Before the end of Jan, a GUI for WIN and MAC users will be available.

Integration and quality control of data received performed with Zanardi pipeline:
https://github.com/bioinformatics-tpg/Zanardi

Visualization Still under discussion
**ADAPTmap WG7 “Population genetics & history”**

- WG7 Population genetics and history” and WG8 “Selection signatures” will be coordinated jointly to avoid duplicating efforts.
- According to the results of WG7 and WG8 analyses, either one joint or two separate main papers will be published.
- Soon after PAG congress the final version of the dataset will be defined (also depending on the availability of the breeds from Burundi and Pakistan).
- WG7 and WGB coordinators will circulate a report on the preliminary descriptive statistics of the dataset (e.g. MAF distribution, genotype quality etc.).
- Thresholds for QC and working dataset content will be defined based on this report.
- A “cured” version of the dataset will be made available to the WG members.
- A core set of analyses has already been identified.
- Allocation of tasks to efficiently run the analyses is ongoing.
- We have set up a Google drive shared folder to facilitate data and file exchange.

**Conclusion**

- ADAPTmap is active and will produce in 2016 its first results
- Tools are being shared (assembly data, SNP chips, know-how...)
- News in the assembly (presented @PAG2016) will be integrated ASAP.
- Data & knowledge sharing in ADAPTMAP allowing meta-data analysis
  - Open for new participants (phase 2
    (alessandra.stella@ptp.it or ezequiel.nicolazzi@ptp.it)

Thank you! www.ptp.it