Casein milk proteins - novel genetic variation and haplotype structure

Siham A. Rahmatalla, Danny Arends, Monika Reissmann, Stefan Krebs, Gudrun A. Brockmann
Composition of goat milk

* Protein fraction can be divided into:
  * Insoluble: **Casein** (α$_{s1}$-casein, α$_{s2}$-casein, β-casein and κ-casein) encoded by *CSN1S1*, *CSN1S2*, *CSN2* and *CSN3* genes
  * Soluble: **Whey** (α-lactalbumin and β-lactoglobulin) encoded by *LALBA* and *LGB* genes
Casein gene cluster

- The four casein genes mapped to chromosome 6 in a tightly linked 250-kb cluster at around 86 Mb

Chromosome 6
Importance of goat milk proteins

* Proteins from goat milk are easier digestible\(^1\)

* \(\alpha_s\) -casein is one of the major milk allergens\(^2\)

* Goat milk has higher levels of \(\beta\)-casein and a lower level of \(\alpha_s\) -casein\(^3\)
  * More similar to human breast milk

* Cheese from goat milk is softer and more fragile compared to cheese made from cow milk\(^4\)

---

\(^1\) Park (2010) *Encyclopedia of Animal Science*

\(^2\) Ballabio et al. (2004) *Dairy Science*

\(^3\) Wang et al. (2017) *Scientific Reports*

\(^4\) Yangilar (2013) *Food and Nutrition Research*
Objective

Assess the allelic variation in casein genes from different rarely studied breeds by using high density capture sequencing

* Focus on indigenous Sudanese goat breeds
* Compared to commonly studied breeds

Identification of genetic variation is necessary for association analysis, breed improvement, and breed conservation

Different protein variants could affect human digestion and properties of goat cheese
M&M: Animals and samples

Nubian (7)  Desert (5)  Taggar (7)  Nilotic (7)

Nubian ibex (2)  Saanen (2)  Alpine ibex* (1)  Bezoar ibex* (2)

* Kindly provided by Leibniz institute for Zoo and Wildlife research, Berlin, Germany
M&M: Sequencing casein genes

* Goat reference sequences were obtained from the *Capra hircus* LWT01 genome version (NCBI)
* Array capture was used to obtain the casein cluster
* Sequencing using high density capture sequencing

<table>
<thead>
<tr>
<th>Gene</th>
<th>Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSN1S1</td>
<td>16,808</td>
</tr>
<tr>
<td>CSN1S2</td>
<td>16,695</td>
</tr>
<tr>
<td>CSN2</td>
<td>9,072</td>
</tr>
<tr>
<td>CSN3</td>
<td>14,114</td>
</tr>
</tbody>
</table>
M&M: Data analysis

- Variants were called using VarScan\(^1\), optimized to detect SNPs

- Polymorphisms were identified manually using the Integrative Genomics Viewer (IGV)\(^2\)

- Haplotypes were reconstructed from nonsynonomous SNPs using the `haplo.stats` package\(^3\) for R

---

\(^1\) Koboldt et al. (2012) *Genome Research*

\(^2\) Robinson et al. (2011) *Nature Biotechnology*

\(^3\) Sinnwell et al. (2013) *CRAN*
Results: Novel nonsynonymous SNPs

* **CSN1S1** - 6 SNPs, 2 Novel

CHR6: 85981710_C>A (Saanen), Histidine to Asparagine  
CHR6: 85984154_A>G (Alpine ibex), Isoleucine to Valine

* **CSN1S2** - 6 SNPs, 4 Novel

CHR6: 86079098_T>C (Saanen & Taggar), Phenylalanine to Serine  
CHR6: 86081790_T>C (Nubian ibex), Phenylalanine to Serine  
CHR6: 86081887_T>C (Nubian ibex), Isoleucine to Threonine  
CHR6: 86089407_G>A (Nubian ibex), Serine to Asparagine
Results: Novel nonsynonymous SNPs

* **CSN2 - 5 SNPs, 3 Novel**
  
  CHR6:86008103_G>A (Alpine ibex), Proline to Leucine  
  CHR6:86015259_T>C (Alpine & Nubian ibex), Histidine to Arginine  
  CHR6:86015278_G>C (Alpine & Nubian ibex), Leucine to Valine

* **CSN3 - 5 SNPs, 2 Novel**
  
  CHR6:86208927_G>A (Alpine ibex), Serine to Asparagine  
  CHR6:86208939_G>C (Alpine ibex), Serine to Threonine

A large number of **novel nonsynonymous SNPs** are only found in the **Alpine/Nubian Ibex**
Results: Constructed haplotypes

<table>
<thead>
<tr>
<th>Chr685881710 C&gt;A</th>
<th>rs65256221</th>
<th>rs64189353</th>
<th>Chr68598454 A&gt;G</th>
<th>rs15555556</th>
<th>rs28293072</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td>A</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>Haplotype B</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>Haplotype C</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>Haplotype D</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Haplotype E</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>Haplotype F</td>
<td>C</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Saanen</th>
<th>Desert</th>
<th>Nubian</th>
<th>Nilotic</th>
<th>Tagger</th>
<th>Ilex</th>
<th>Bezoar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td>0,25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype B</td>
<td>0,50</td>
<td>0,50</td>
<td>0,57</td>
<td>0,57</td>
<td>0,64</td>
<td></td>
</tr>
<tr>
<td>Haplotype C</td>
<td>0,25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype D</td>
<td>0,50</td>
<td>0,43</td>
<td>0,43</td>
<td>0,36</td>
<td>0,83</td>
<td>0,50</td>
</tr>
<tr>
<td>Haplotype E</td>
<td>0,50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,50</td>
</tr>
<tr>
<td>Haplotype F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0,17</td>
<td></td>
</tr>
</tbody>
</table>

**CSN1S1**: 6 haplotypes
**CSN1S2**: 6 haplotypes
**CSN2**: 6 haplotypes
**CSN3**: 4 haplotypes
Results: Constructed haplotypes

<table>
<thead>
<tr>
<th></th>
<th>CHR683208927_G&gt;A</th>
<th>CHR683208939_G&gt;C</th>
<th>rs26829109</th>
<th>rs26829113</th>
<th>rs65105868</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>Haplotype B</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>Haplotype C</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>Haplotype D</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Saanen</th>
<th>Desert</th>
<th>Nubian</th>
<th>Nilotic</th>
<th>Tagger</th>
<th>Ibex</th>
<th>Bezoar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haplotype A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype B</td>
<td>0.10</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haplotype C</td>
<td>0.50</td>
<td>0.90</td>
<td>0.86</td>
<td>1.00</td>
<td>0.93</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Haplotype D</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*CSN1S1*: 6 haplotypes  
*CSN1S2*: 6 haplotypes  
*CSN2*: 6 haplotypes  
*CSN3*: 4 haplotypes
Summary

**Novel nonsynonymous SNPs** in the casein genes were found in Alpine ibex and **critically endangered Nubian ibex**

This highlights the importance of **studying and preservation** these rare and/or endangered breeds
Acknowledgments

Humboldt university, Berlin
Prof. Dr. Gudrun Brockmann
Dr. Danny Arends
Dr. Monika Reissmann

Special thanks for the efforts of the goat owners and Research stations in Sudan. Leibniz institute for Zoo and Wildlife research in Berlin.
Thank you for your attention

I have a question