Recent discoveries in two milk proteins genes (*CSN1S2* and *LALBA*) revealed useful markers for gene's assisted selection in camels

N. Letaief, G. Gaspa, **A. Pauciullo**

UNIVERSITÀ DI TORINO

Department of Agricultural, Forest and Food Sciences
Grugliasco, Turin (Italy)

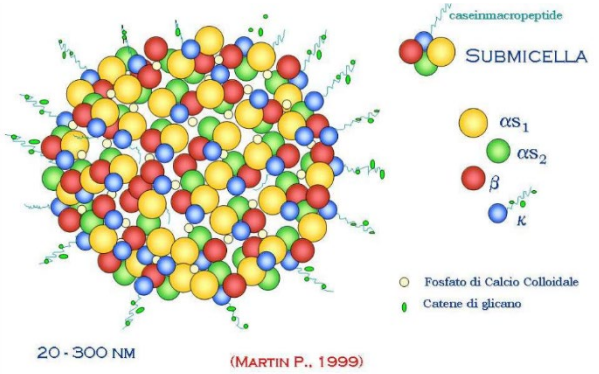
Outline

- **Why αs2-casein and α-lactalbumin?**
- **How was analysis done?**
- **Results and effects on gene expression**
- **Conclusion**

Why the αs2-casein (*CSN1S2* gene)?

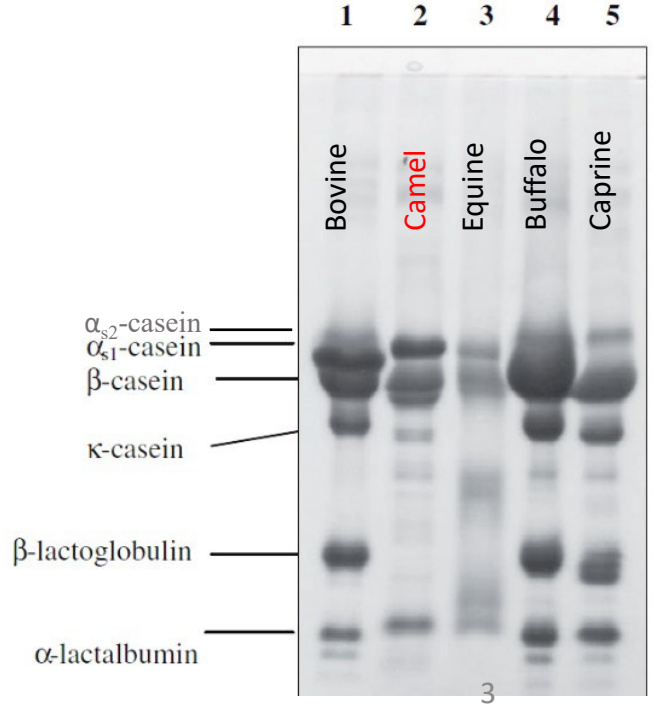
The αs2-casein (CN) is one of the three calcium-sensitive phosphoproteins secreted in milk together with the αs1- and β-CN and it is the most hydrophilic of all caseins.

Important for the stability of milk micelles and for dairy processing.

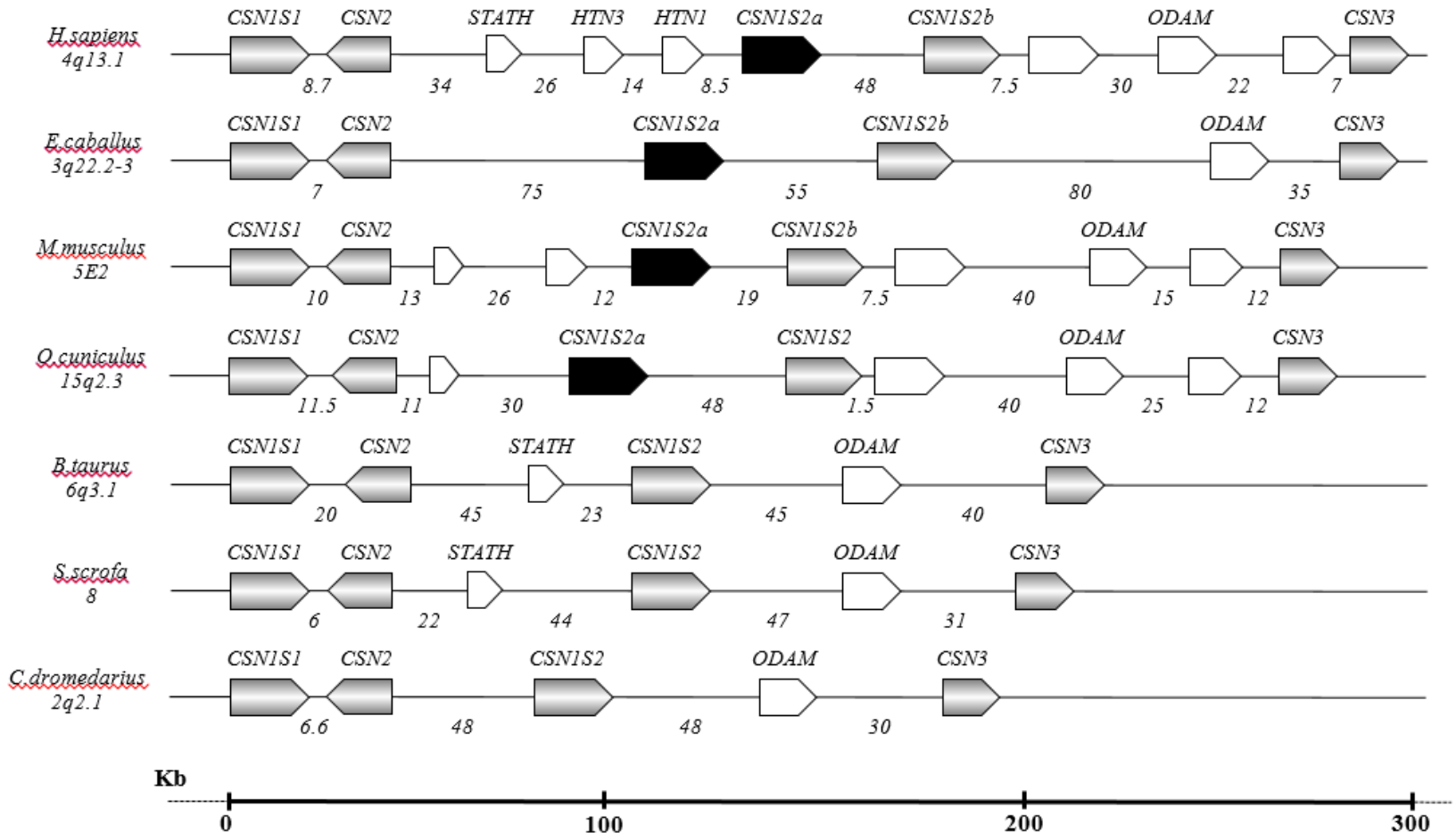


	Milk protein	Camel	Cow	Human
Ca^{2+} sensible caseins	α_{S1}-Casein	5000	12,000	minute
	α_{S2}-Casein	2200	3000	minute
	β-Casein	15,000	10,000	4670
	κ-Casein	800	3500	minute
Whey-proteins 12%	α-Lactalbumin	3500	1260	3400
	β-Lactoglobulin	nd	3500	nd
	Whey acidic protein	157	nd	nd

(Kappeler et al., 2003)



Caseins have been recognized as a powerful molecular model for evolutionary studies and the knowledge of tag markers within casein genes might play a fundamental role in the understanding of species diversification.

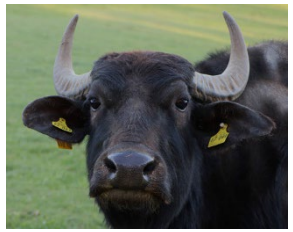


Martin et al., 2013
Pauciullo et al., 2019

How much do we know about genetic diversity of *CSN1S2* (αs2-CN) in the other farm animals?


7 alleles associated with three different αs2-CN levels (Boisnard et al., 1991; Ramunno et al., 2001a,b; Giambra and Erhardt, 2011).

4 variants A, B, C, and D have been found (Farrell et al., 2004; Ibeagha-Awemu et al., 2007).



8 alleles have been identified (Cosenza et al., 2021).


In dromedary and bactrian camels, a series of alternative splicing variants with **2 non allelic variants** (UP1 and UP2) have been recently described by Ryskaliyeva et al. (2019) that reported a deep characterization of milk protein in Old World camelids accomplished by LC-ESI-MS.

Why α-lactalbumin (*LALBA* gene)?

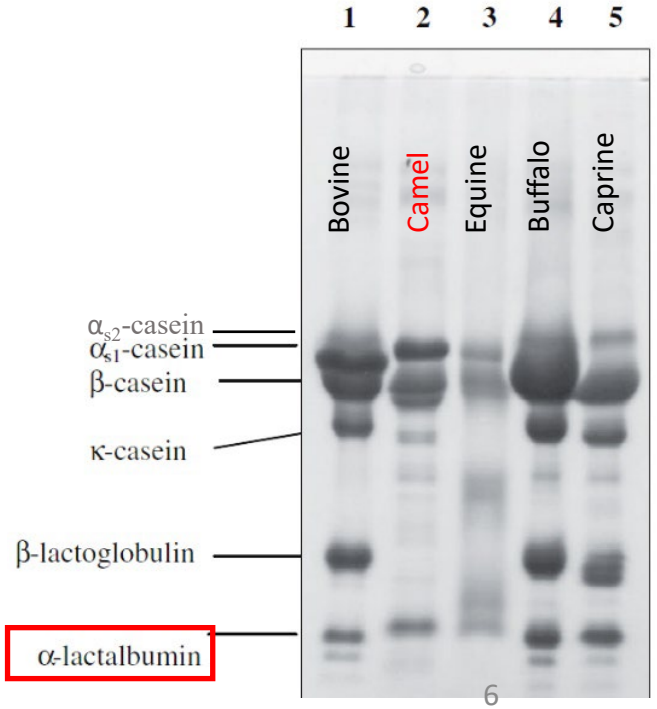
Almost 3 times higher than the concentration found in cow milk, but very similar to that in human milk

If we refer to complete protein pattern, camel milk is more similar to human milk than cow's milk:

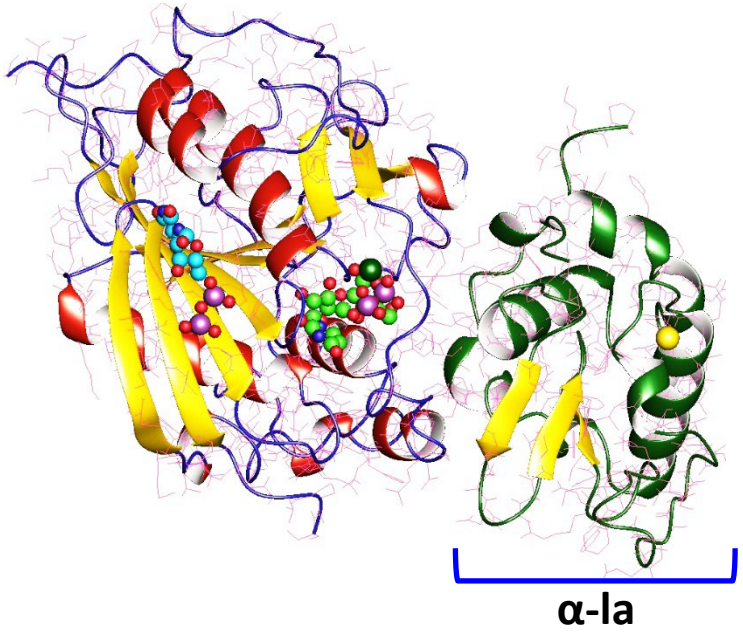
- Significantly higher concentrations of β-CN (~65%)
- Lower amounts of k-CN
- Lack of β-lactoglobulin

	Milk protein	Camel	Cow	Human
Ca^{2+} sensible caseins	α_{S1}-Casein	5000	12,000	minute
	α_{S2}-Casein	2200	3000	minute
	β-Casein	15,000	10,000	4670
	κ-Casein	800	3500	minute
Whey-proteins 12%	α-Lactalbumin	3500	1260	3400
	β-Lactoglobulin	nd	3500	nd
	Whey acidic protein	157	nd	nd

(Kappeler et al., 2003)

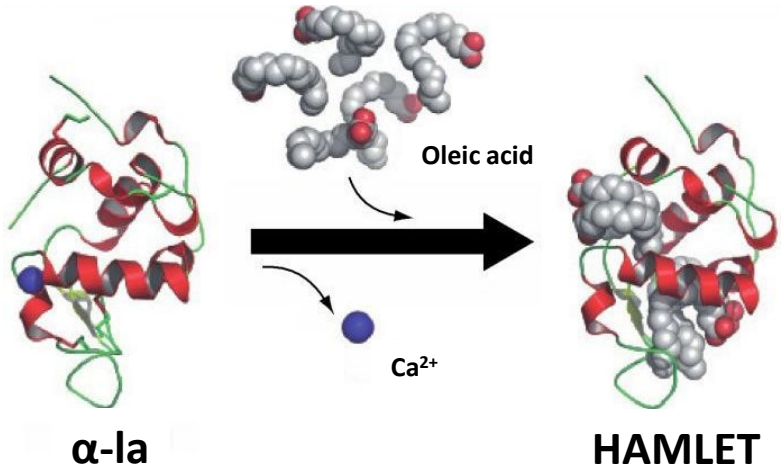


Why α-lactalbumin (*LALBA* gene)?



α-lactalbumin belongs to the lysozyme superfamily that also includes lysozymes c and calcium-binding lysozymes (Nitta et al., 1989)

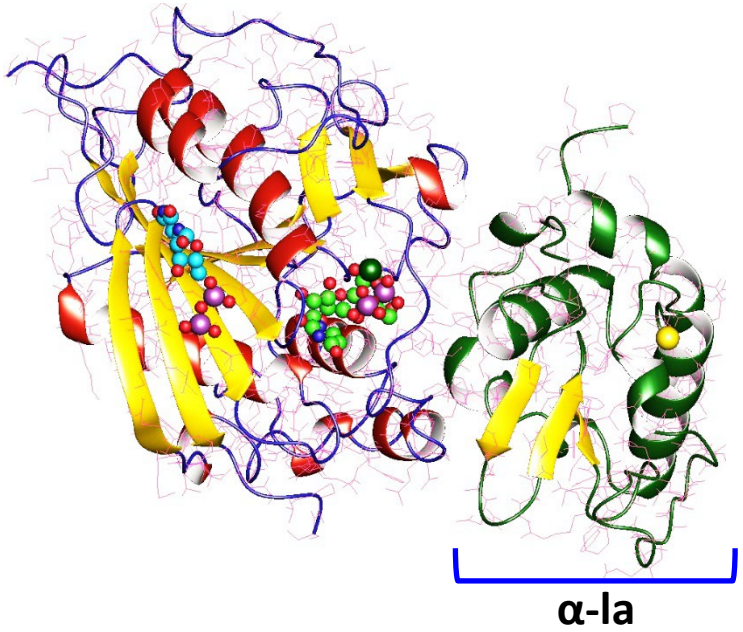
The key function of this protein is to **facilitate lactose synthesis** by the galactosyltransferase component of the enzyme system, serving as **regulatory subunit**



Other biological function have been demonstrated for α-La

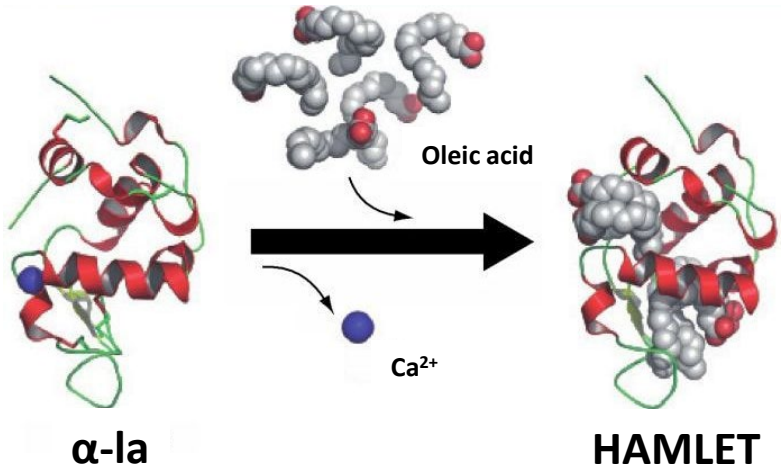
- Immune modulation (Cross and Gill, 2000)
- Gastric protection (Matsumoto et al., 2001)
- Stress reduction (Markus et al., 2002)
- Antimicrobial activity (Madureira et al., 2007)
- Human Alpha-lactalbumin Made LEthal to Tumor cells – HAMLET (Svensson et al., 1999; 2000; Pettersson et al., 2006; Uversky et al., 2017)

Why α-lactalbumin (*LALBA* gene)?



α-lactalbumin belongs to the lysozyme superfamily that also includes lysozymes c and calcium-binding lysozymes (Nitta et al., 1989)

The key function of this protein is to **facilitate lactose synthesis** by the galactosyltransferase component of the enzyme system, serving as **regulatory subunit**

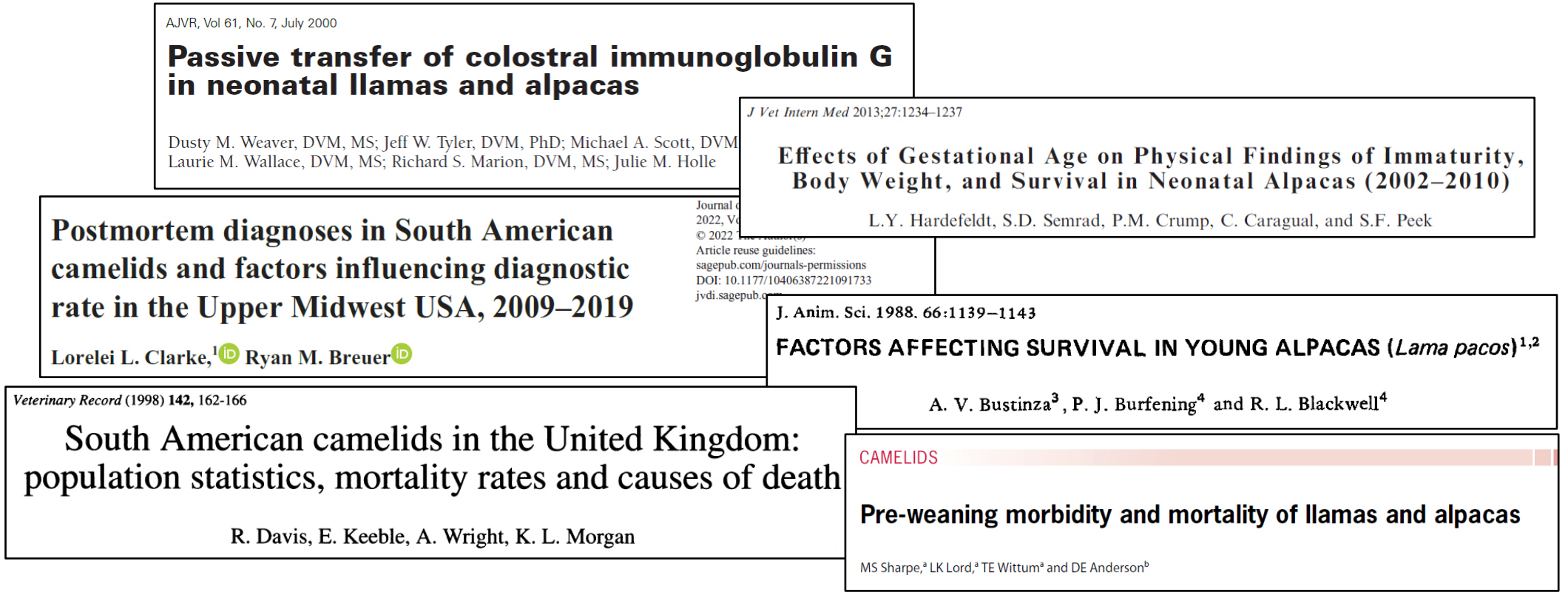


Other biological function have been demonstrated for α-La

- **Immune modulation** (Cross and Gill, 2000)
- Gastric protection (Matsumoto et al., 2001)
- Stress reduction (Markus et al., 2002)
- **Antimicrobial activity** (Madureira et al., 2007)
- Human Alpha-lactalbumin Made LEthal to Tumor cells – HAMLET (Svensson et al., 1999; 2000; Pettersson et al., 2006; Uversky et al., 2017)

Why α-lactalbumin (*LALBA* gene)?

Camelids, in particular SAC, are **at increased risk of morbidity and mortality during the first months of life**. Pre-weaning mortality rates have been reported to range between **12-16%**.



AJVR, Vol 61, No. 7, July 2000

Passive transfer of colostral immunoglobulin G in neonatal llamas and alpacas

Dusty M. Weaver, DVM, MS; Jeff W. Tyler, DVM, PhD; Michael A. Scott, DVM
Laurie M. Wallace, DVM, MS; Richard S. Marion, DVM, MS; Julie M. Holle

J Vet Intern Med 2013;27:1234–1237

Effects of Gestational Age on Physical Findings of Immaturity, Body Weight, and Survival in Neonatal Alpacas (2002–2010)

L.Y. Hardefeldt, S.D. Semrad, P.M. Crump, C. Caragual, and S.F. Peek

Postmortem diagnoses in South American camelids and factors influencing diagnostic rate in the Upper Midwest USA, 2009–2019

Lorelei L. Clarke,[1] Ryan M. Breuer

Journal
2022, Vo
© 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10406387221091733
jvdi.sagepub.com

J. Anim. Sci. 1988. 66:1139–1143

FACTORS AFFECTING SURVIVAL IN YOUNG ALPACAS (*Lama pacos*)[1,2]

A. V. Bustinza[3], P. J. Burfening[4] and R. L. Blackwell[4]

Veterinary Record (1998) **142,** 162-166

South American camelids in the United Kingdom: population statistics, mortality rates and causes of death

R. Davis, E. Keeble, A. Wright, K. L. Morgan

CAMELIDS

Pre-weaning morbidity and mortality of llamas and alpacas

MS Sharpe,[a] LK Lord,[a] TE Wittum[a] and DE Anderson[b]

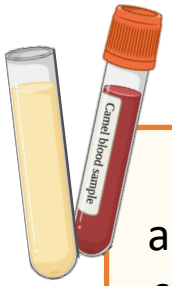
This is a **critical factor** determining survival of the newborn/cria

Major economic consideration for producers

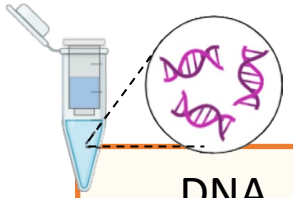
Aims

- Sequencing *CSN1S2* and *LALBA* genes and their flanking regions (promoter and 3'-end) in camels;
- Identifying **the regulatory elements** (Transcription factors binding sites, microRNA, etc.) that may affect **gene expression** or change the amino acid composition;
- Identifying **markers** potentially useful for the selective breeding.

How was analysis done?

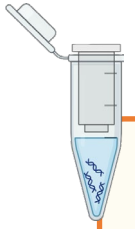


Blood and Milk samples

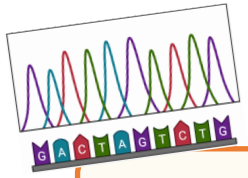


DNA extraction

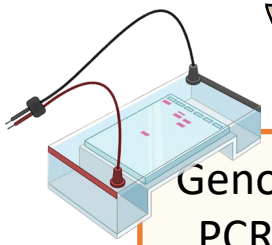
Polymerase Chain Reaction



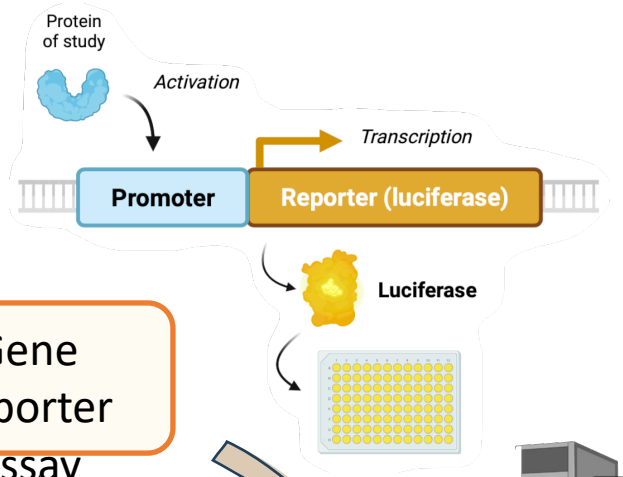
Purification of PCR amplicons



Sanger Sequencing



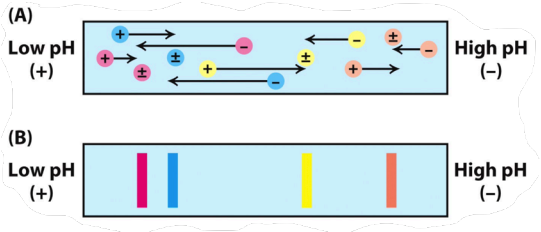
Genotyping PCR-RFLP



Gene Reporter Assay

Isoelectric focusing

Bioinformatics and statistics



Results *CSN1S2* (αs2-casein)

- **Bactrian camel: (EMBL ID: OQ720238)**

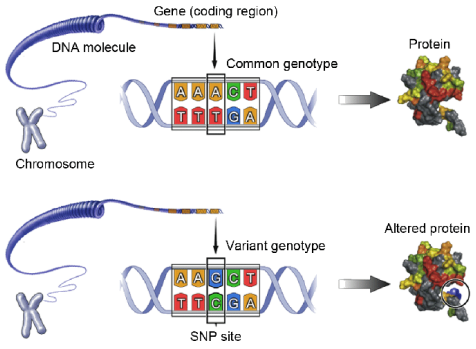
  18 SNPs (5 transversions and 13 transitions)

- **Dromedary camel: (EMBL ID: OQ720239)**

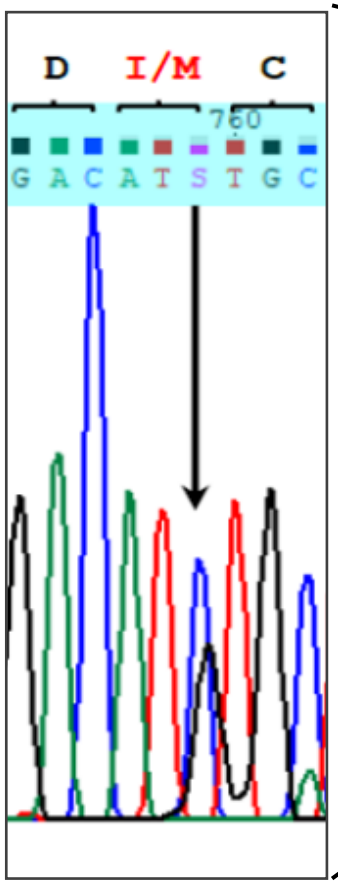
  16 SNPs (8 transversions and 8 transitions)

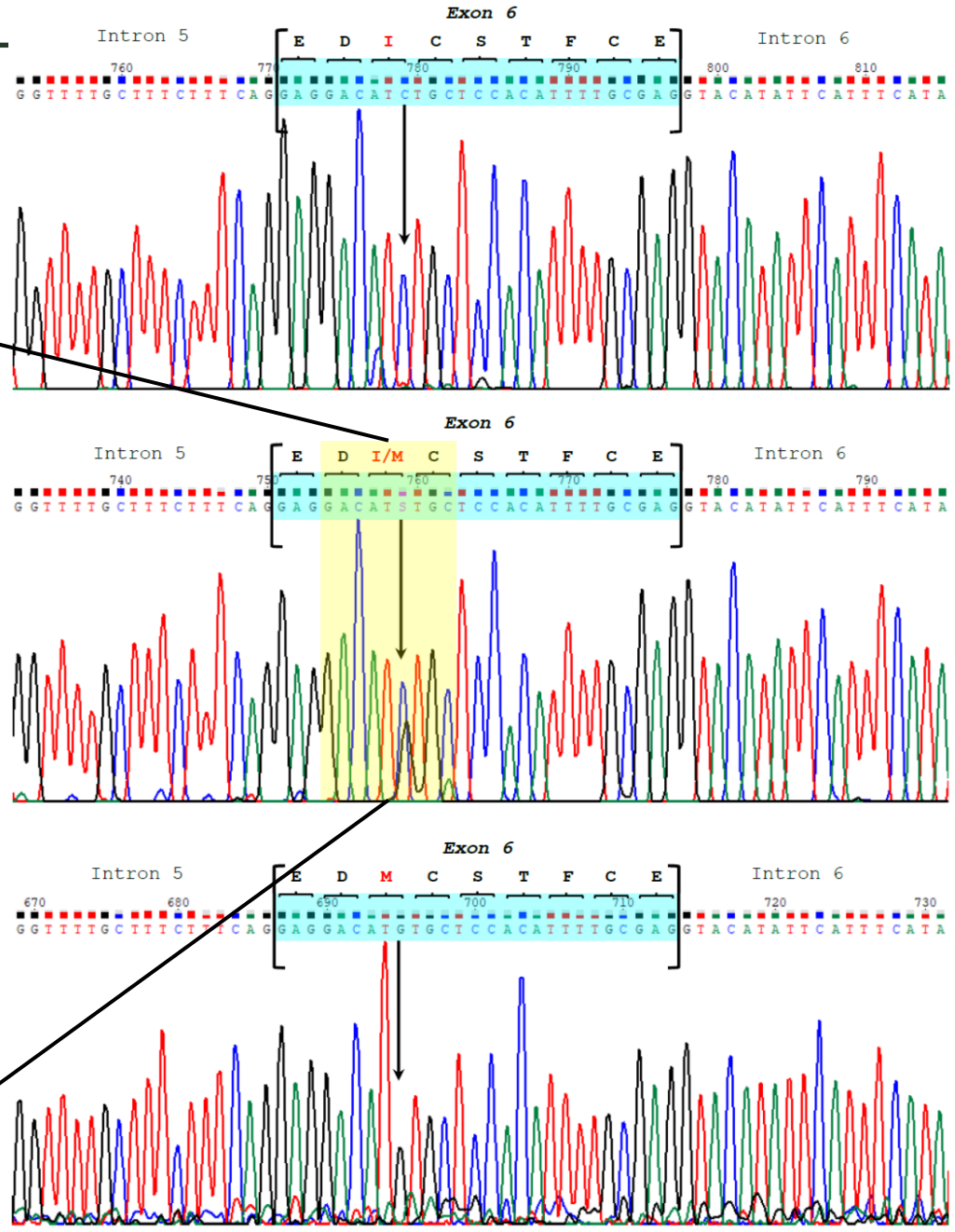
Results *CSN1S2* (αs2-casein)

First allelic variant found in the Bactrian αs2-casein



SNP g.3639C>G in exon 6 resulted in an **amino acid change** at position 36 of the mature protein (**p.36Ile>Met**)





Results *CSN1S2* (αs2-casein)

First exonic SNP found in the Dromedary αs2-casein

SNP g. 15110G>T in exon 17

Analysis of the 3'UTR region of the dromedary *CSN1S2* gene (from the first nucleotide of exon 17) to test whether SNP g.15110G>T could influence miRNA binding sites

miRNA Name	miRNA sequence	Seed location	Custom target sequence	Target Score	
				T/T	G/G
hsa-miR-298	AGCAGAAGCAGGGAGGUUCUCCCA	20	CTTCTGCA	93	87
hsa-miR-4418	CACUGCAGGACUCAGCAG	23	CTGCAGT	83	76
hsa-miR-3158-5p	CCUGCAGAGAGGAAGCCCUUC	22	TCTGCAG	81	65
hsa-miR-548av-3p	AAAACUGCAGUUACUUUUGC	25	GCAGTTT	66	-
hsa-miR-4662a-3p	AAAGAUAGACAAUUGGCUAAAU	16	CTATCTT	52	-

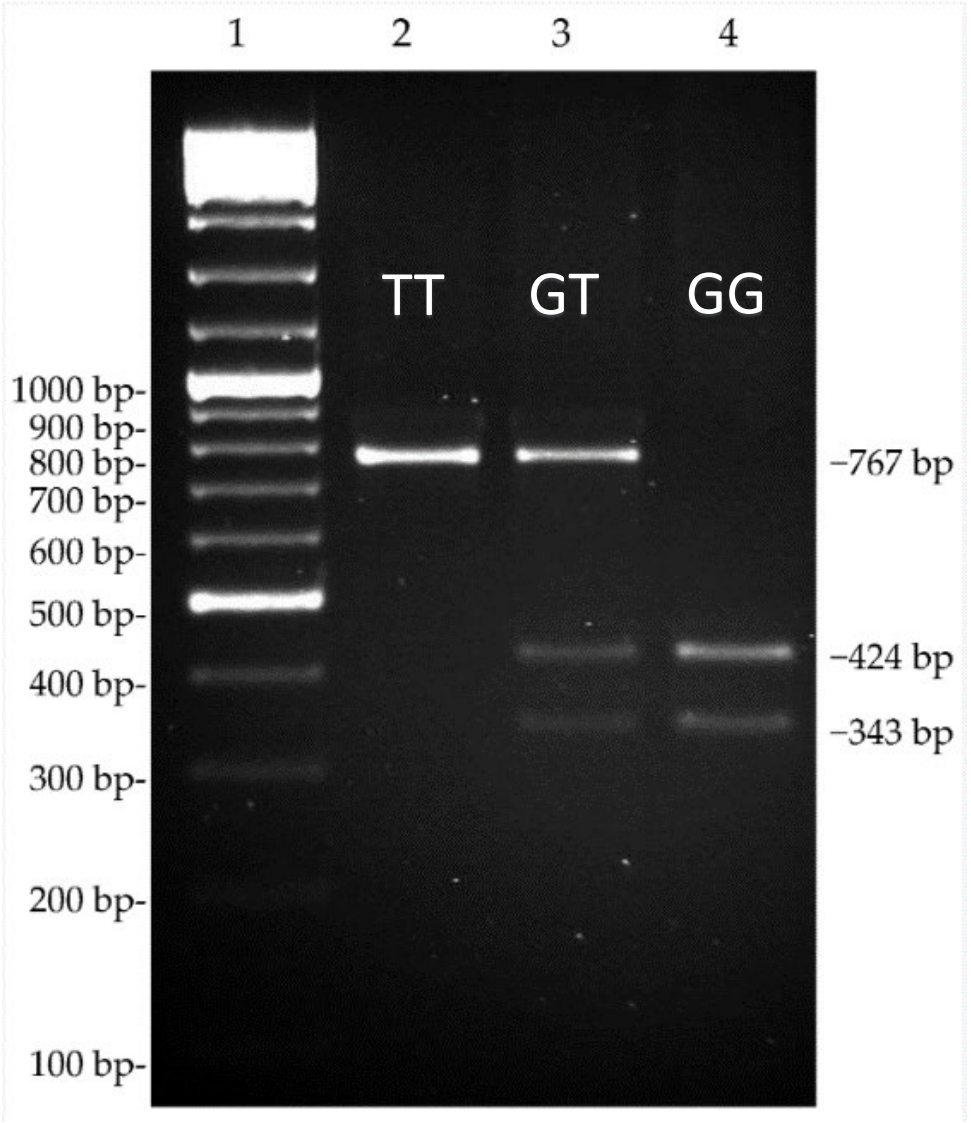
Additional miRNA target sites for the T allele → higher down regulation of *CSN1S2* gene (instead no effect from the G allele).

g.15110G>T mutation affected the seed sequence (CTATCTT) of miRNA 4662a-3p, which is the short sequence that allows the miRNA to act on the gene.

Results *CSN1S2* (αs2-casein)

	Genotypes			Allele frequency		
	TT	GT	GG	Total	T	G
Observed	83	61	13	157	0.723	0.277
Expected	82.05	62.89	12.05			

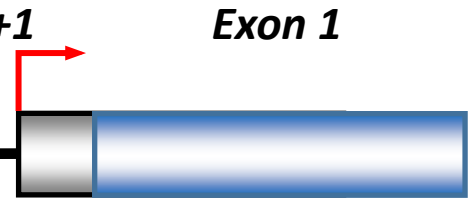
Lower frequency the G allele → stronger chances for a directional molecular selection in favor of this allele that could lead to a more rapid improvement in αs2-casein yield.



Results *CSN1S2* (αs2-casein)

VS

26 nucleotide differences were found in introns, but none of them appeared to affect the canonical sites of the splicing mechanism.



Promoter: *24 vs 22 Transcription Factors*

TRANSFAC database

In silico analysis

Transcription Factor	Consensus Motif	Signal Sequence	Strand	Score	*C. bactrianus*	*C. dromedarius*
AP-1	NNTGACTCANN	CCTGACTCCCT	+	0.913	−677/−667	−677/−667
TEC1	TNCATTCYWW	TTCATTCCAT	+	0.985	−620/−629	-
AP-4	NNCAGCTGNN	CACAGCTGGT	+	0.989	−591/−582	−591/−582
Oct-1	CWNAWTKWSATRYN	CACAATTAAATATG	+	0.946	−573/−560	−573/−560
Pit-1a	NNGAATATKCANNNN	AATATGAATATTATT	–	0.944	−565/−551	−565/−551
C/EBP-β	RNRTKNNGMAAKNN	AAGTTAAGAAAGTA	+	0.908	−527/−514	−527/−514
AP-4	NNCAGCTGNN	GAGAGCTGAG	–	0.934	−482/−473	−482/−473
C/EBP-β *	RNRTKNNGMAAKNN	GACTTGCATAAGACT	–	0.909	−453/−439	-
YY1	CCATNTWNNNW	CCATATTTTA	+	0.899	−436/−426	−436/−426
Pit-1a	TGAATAWNWA	TGAATATGAA	+	0.859	−404/−395	−404/−395
Oct-1	CWNAWTKWSATRYN	AATATGAAAAATGT	–	0.847	−402/−389	−402/−389
Oct-1	NNNRTAATNANNN	GTATTAATGAAAT	+	0.870	−378/−366	−378/−366
Oct-1	CWNAWTKWSATRYN	CACATCCAAAATAT	–	0.890	−356/−343	−356/−343
C/EBP-α	NNTKTGGWNANNN	TATTTGTTTAAAG	+	0.901	−333/−321	−333/−321

Full table available at: https://doi.org/10.3390/ani13172805

CSN1S2 gene expression could be different between Bactrians and dromedaries due to differences in the gene promoter.

Results

The intra-species comparison of sequences showed a total of 36 SNP, 26 of which detected within the LALBA gene.

		C. dromedarius		*C. bactrianus*		*V. pacos*		*L. glama*	
		Position	**Nucleotide (amino acid)**	**Position**	**Nucleotide (amino acid)**	**Position**	**Nucleotide (amino acid)**	**Position**	**Nucleotide (amino acid)**
LALBA	**Exon 1**	680	C	680	C	679	S	679	S
			(p.5Val)		(p.5Val)		(p.5Val)		(p.5Val)
	Intron 1	836	T	836	T	835	Y	835	T
	Exon 2	1160	T	1161	T	1159	Y	1159	T
			(p.54Gly)		(p.54Gly)		(p.54Gly)		(p.54Gly)
		1217	R	1218	G	1216	G	1216	G
			(p.73Gln)		(p.73Gln)		(p.73Gln)		(p.73Gln)
		1230	A	1231	A	1229	R	1229	R
			(*p.78Ile*)		(*p.78Ile*)		(*p.78Ile>Val*)		(*p.78Ile>Val*)
	Intron 2	1532	C	1533	C	1518	Y	1518	Y
		1562	T	1563	T	1548	T	1548	Y
		1636	C	1637	C	1622	Y	1622	Y
		1654	T	1655	T	1640	Y	1640	C
		1778	T	1779	T	1764	Y	1764	T
	Intron 3	1942	T	1943	T	1928	T	1928	Y
		2060	G	2061	G	2046	R	2046	G
		2063	T	2064	T	2049	Y	2049	T
		2065	C	2066	C	2051	C	2051	S
		2153	C	2154	C	2139	Y	2139	C
		2168	C	2169	C	2154	Y	2154	T
		2237	G	2238	G	2223	K	2223	G
		2284	G	2285	G	2270	R	2270	G
	Exon 4	2524	T	2525	T	2510	W	2510	W
		2642	C	2642	C	2628	Y	2628	Y

Note: W=A/T, Y=C/T; R=A/G; S=C/G, K=G/T

- **1 SNP (3.84%) was detected in dromedary camel**

Results

The intra-species comparison of sequences showed a total of 36 SNP, 26 of which detected within the LALBA gene.

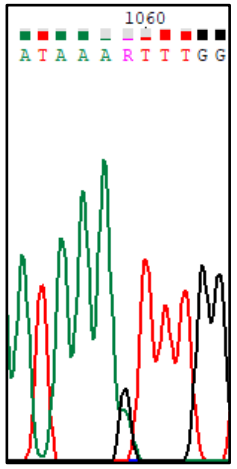
		C. dromedarius		*C. bactrianus*		*V. pacos*		*L. glama*	
		Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)
LALBA	Exon 1	680	C	680	C	679	S	679	S
			(p.5Val)		(p.5Val)		(p.5Val)		(p.5Val)
	Intron 1	836	T	836	T	835	Y	835	T
	Exon 2	1160	T	1161	T	1159	Y	1159	T
			(p.54Gly)		(p.54Gly)		(p.54Gly)		(p.54Gly)
		1217	R	1218	G	1216	G	1216	G
			(p.73Gln)		(p.73Gln)		(p.73Gln)		(p.73Gln)
		1230	A	1231	A	1229	R	1229	R
			(*p.78Ile*)		(*p.78Ile*)		(*p.78Ile>Val*)		(*p.78Ile>Val*)
	Intron 2	1532	C	1533	C	1518	Y	1518	Y
		1562	T	1563	T	1548	T	1548	Y
		1636	C	1637	C	1622	Y	1622	Y
		1654	T	1655	T	1640	Y	1640	C
		1778	T	1779	T	1764	Y	1764	T
	Intron 3	1942	T	1943	T	1928	T	1928	Y
		2060	G	2061	G	2046	R	2046	G
		2063	T	2064	T	2049	Y	2049	T
		2065	C	2066	C	2051	C	2051	S
		2153	C	2154	C	2139	Y	2139	C
		2168	C	2169	C	2154	Y	2154	T
		2237	G	2238	G	2223	K	2223	G
		2284	G	2285	G	2270	R	2270	G
	Exon 4	2524	T	2525	T	2510	W	2510	W
		2642	C	2642	C	2628	Y	2628	Y

Note: W=A/T, Y=C/T; R=A/G; S=C/G, K=G/T

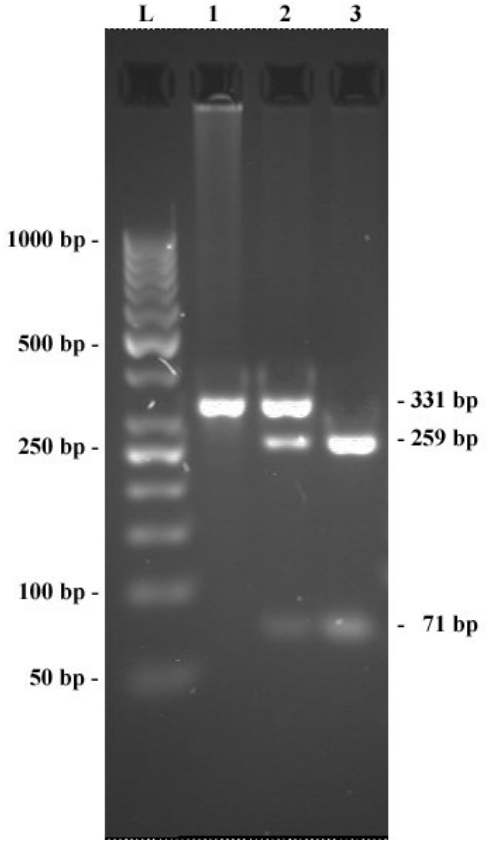
- **25 SNP (96.15%) were found in SAC**

g.1229A>G
First allelic variant
p.78Ile>Val

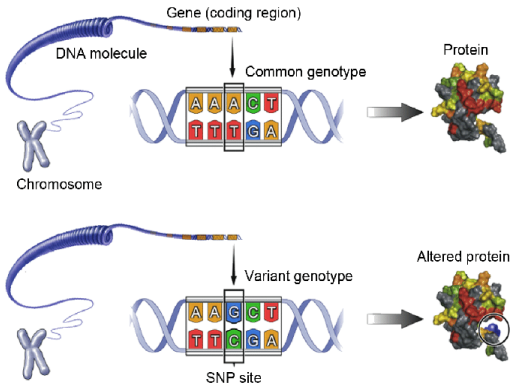
Genotyping for the Exon 2



1229A>G



Species		Exon 2				
		Genotypes g.1229A>G			Allelic frequency	
	N.	A/A	A/G	G/G	A	G
Alpaca	80 (100)	36 (45)	38 (48)	6 (8)	0.69	0.31
Llama	45 (100)	36 (80)	9 (20)	0 (0)	0.90	0.10
Total	**125 (100)**	**72 (57.6)**	**47 (37.6)**	**6 (4.8)**	**0.76**	**0.24**



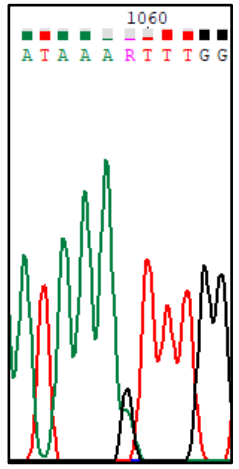
Missense mutation

ATT > GTT

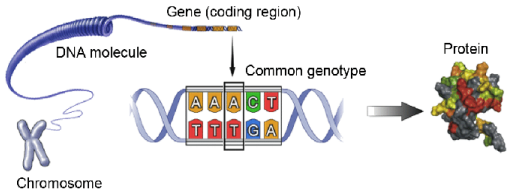
p.78Ile>Val

MAF was 0.31 and 0.1 in the Alpaca and llama, respectively

Iso-Electro focusing of skimmed alpaca milk confirmed the polymorphis at the protein level



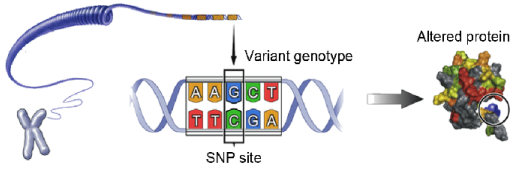
1229A>G

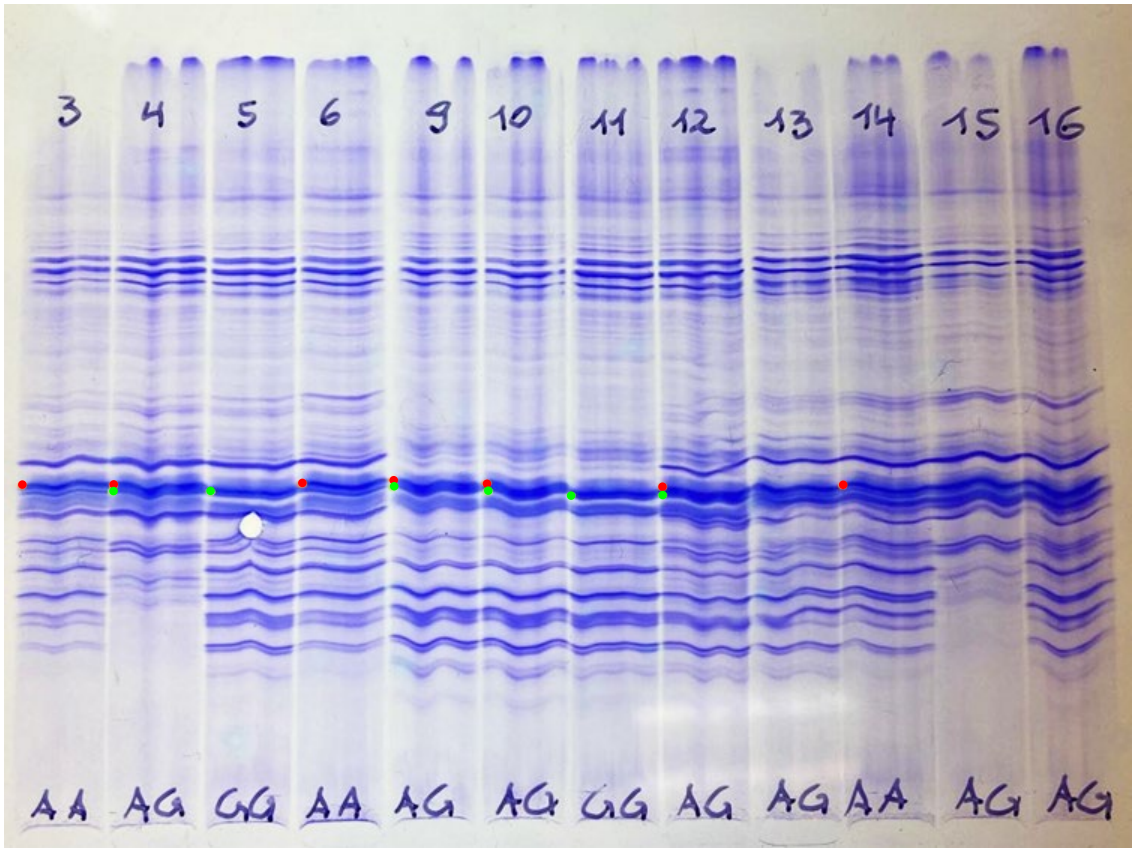


Missense mutation

ATT > GTT

p.78Ile>Val





The inter-species comparison showed 86 additional markers as showed in the table

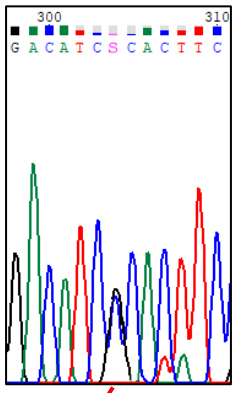
		C. dromedarius		*C. bactrianus*		*V. pacos*		*L. glama*	
		Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)
Promoter		24	T	24	T	24	G	24	G
		31	A	31	A	31	G	31	G
		50	C	50	C	50	T	50	T
		58	T	58	T	58	C	58	C
		85	A	85	A	85	G	85	G
		94	T	94	T	94	C	94	C
		134	G	134	A	134	A	134	A
		140	T	140	T	140	C	140	C
		205-207	TGC	205-207	TGC	205-207	CAT	205-207	CAT
		211	A	211	A	211	T	211	T
		223	A	223	A	223	C	223	C
		245	T	245	T	245	A	245	A
		275	T	275	T	275	C	275	C
		280	C	280	C	280	T	280	T
		358-359	TG	358-359	TG	358-359	CA	358-359	CA
		394	T	394	T	394	C	394	C
		462	T	462	T	462	G	462	G
		489	A	489	A	489	G	489	G
		492-493	AA	492-493	AA	492-493	C-	492-493	C-
		503	G	503	G	502	A	502	A
		507	T	507	T	506	C	506	C
		525	C	525	C	524	T	524	T
		573	A	573	A	572	G	572	G
		585	T	585	T	584	G	584	G
		655	G	655	G	654	A	654	A
	Exon 1	763	A	763	A	762	G	762	G
			(p.33Asp)		*(p.33Asp)*		*(p.33Gly)*		*(p.33Gly)*
		770	T (p.35Asn)	770	T (p.35Asn)	769	C (p.35Asn)	769	C (p.35Asn)
	Intron 1	801	A	801	A	800	G	800	G
		825	T	825	T	824	C	824	C
		830	A	830	A	830	G	830	G
		838	C	838	C	837	T	837	T
		966	C	966	C	965	T	965	T
		968	C	968	C	967	T	967	T
		1036/1037	-	1037	A	1035/1036	-	1035/1036	-
		1087	G	1088	G	1086	T	1086	T
		1118	A	1119	A	1117	C	1117	C
		1123	T	1124	T	1122	C	1122	C
		1126	C	1127	C	1125	T	1125	T
	Exon 2	1240	G	1241	G	1239	T	1239	T
			(p.81Arg)		*(p.81Arg)*		*(p.81Ile)*		*(p.81Ile)*
		1274	C (p.92Cys)	1275	C (p.92Cys)	1273	T (p.92Cys)	1273	T (p.92Cys)
	Intron 2	1326-1327	CC	1327-1328	CC	1325-1326	TT	1325-1326	TT
		1332/1333	-	1333/1334	-	1332	T	1332	T
		1339	G	1340	G	1339	C	1339	C
		1370	G	1371	G	1370	A	1370	A
		1396	A	1397	A	1396	G	1396	G
		1417	T	1418	T	1417	C	1417	C
		1432	C	1433	C	1432	G	1432	G
		1436-1449	CTAAGAGG CTGTTA	1437-1450	CTAAGAG GCTGTTA	1435/1436	-	1435/1436	-
		1452	G	1453	G	1438	A	1438	A
		1570	A	1571	A	1556	G	1556	G
		1603	C	1604	C	1589	T	1589	T
		1610	A	1611	A	1596	G	1596	G
		1651	G	1652	G	1637	A	1637	A
		1700	T	1701	T	1686	C	1686	C
		1751	C	1752	A	1737	C	1737	C
	Exon 3	1809	C (p.106Asp)	1810	C (p.106Asp)	1795	T (p.106Asp)	1795	T (p.106Asp)
		1814	A *(p.108Lys)*	1815	A *(p.108Lys)*	1800	T *(p.108Ile)*	1800	T *(p.108Ile)*
	Intron 3	1875	A	1876	A	1861	G	1861	G
		1909	C	1910	T	1895	T	1895	T
		1943	A	1944	G	1929	G	1929	G
		2008	T	2009	T	1994	C	1994	C
		2025	T	2026	T	2011	C	2011	C
		2140	C	2141	C	2126	T	2126	T
		2154	T	2155	T	2140	C	2140	C
		2302	T	2303	T	2288	G	2288	G
		2314	A	2315	G	2300	A	2300	A
	Exon 4	2535	A	2536	C	2521	C	2521	C
		2565	T	2566	T	2551	C	2551	C
		2622	T	2622/2623	-	2608	T	2608	T
		2643	A	2643	A	2629	G	2629	G
		2672	G	2672	G	2658	A	2658	A
		2676-2677	TG	2676-2677	TG	2662-2663	CT	2662-2663	CT
3' flanking region		2686	T	2686	T	2672	C	2672	C
		2696	A	2696	A	2682	T	2682	T
		2730	A	2730	A	2716	G	2716	G
		2784	A	2784	A	2770	G	2770	G
		2787	C	2787	C	2773	G	2773	G
		2836	A	2836	G	2822	A	2822	A
		2845-2859	AACCACCT GACTTAT	2844/2845	-	2831-2845	AACCACCT GACTTAT	2831-2845	AACCACCT GACTTAT
		2886	A	2871	A	2872	G	2872	G
		2896-2900	AGTAA	2881-2885	AGTAA	2881/2882	-	2881/2882	-
		2925	T	2910	T	2906	A	2906	A
		3033	T	3018	G	3014	T	3014	T
		3039	A	3024	G	3020	A	3020	A
		3071	A	3056	G	3052	A	3052	A
		3100	A	3085	G	3081	A	3081	A

Note: W=A/T, Y=C/T; R=A/G; S=C/G, K=G/T

9 SNP found in the promoter and one in the 3’ flanking region

	C. dromedarius		*C. bactrianus*		*V. pacos*		*L. glama*	
	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)	Position	Nucleotide (amino acid)
Promoter	24	T	24	T	24	G	24	G
	112	T	112	T	112	R	112	R
	237	T	237	T	237	Y	237	Y
	357	C	357	C	357	S	357	C
	366	R	366	A	366	A	366	A
	429	T	429	T	429	W	429	T
	593	G	593	G	592	S	592	G
	615	G	615	G	614	R	614	A
3’ flanking region	2699	T	2699	T	2685	T	2685	K

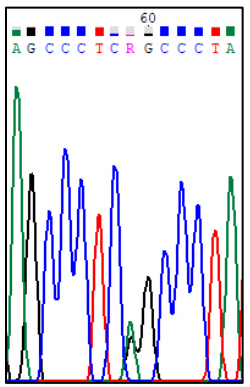
Note: W=A/T, Y=C/T; R=A/G; S=C/G, K=G/T



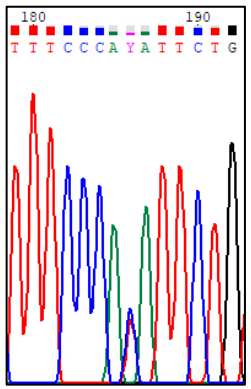
Promoter

TRANSFAC database

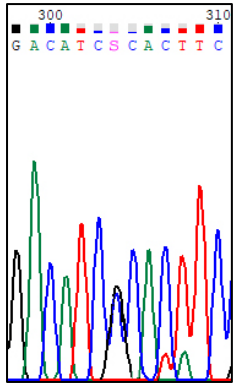
+1 **Exon 1**



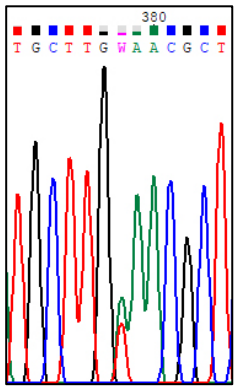
-553A>G



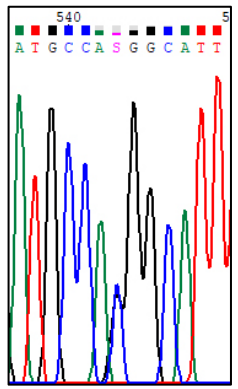
-428C>T



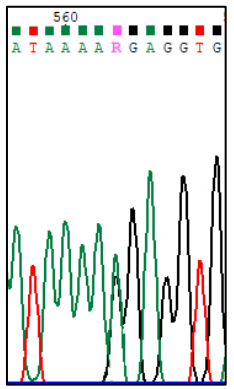
-308C>G



-236A>T



-73C>G



-51C>G

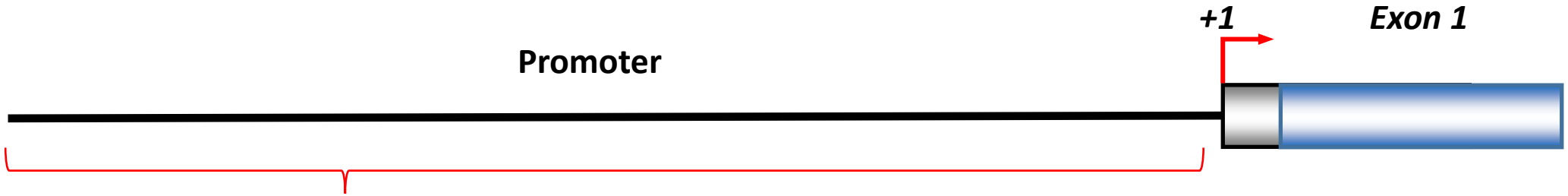


Table 5. List of transcription factors binding sites detected in the 5′-flanking regions of *LALBA* in domestic camelids[1]

		Transcription factor binding site		
Position	Consensus motif 5′→3′	Old World camel	New World camel	Shared
−580/−572	wwTyCCTAwT	MCM1		
−558/−549[2]	CCCkCsGCmy			SP1[3]
−323/−314[2]	rGsCTGGGGm			AP2-α
−314/−305	ACwTCCTsyk	PEA-3		
−183/−174	GGAAArkGrm		ICSBP/ISGF-3	
−150/−141	syTGGCmGsC	NF-1		
−102/−93	CCTCyTCCys		SP1	
−80/−71	rrkGsCAGGG			YY1
−62/−52	sATGmATrwA			PIT-1A
−57/−48	ATAAAArrnG			TATA
−11/−2	wksmAGCCAA		NF-1	

[1]k = G/T; m = A/C; n = A/C/G/T; r = A/G; s = C/G; w = A/T; y = C/T.
[2]SP1 and AP2-α positions in New World camels are −557/−548 and −313/−322, respectively.
[3]Presence of the SP1 binding site in New World camelids is genotype dependent due to the polymorphism g.112A>G.

At least 11 main consensus sequences for TF were found.

Five were motifs shared among the 4 investigated camelids and can be considered as essential for gene expression.

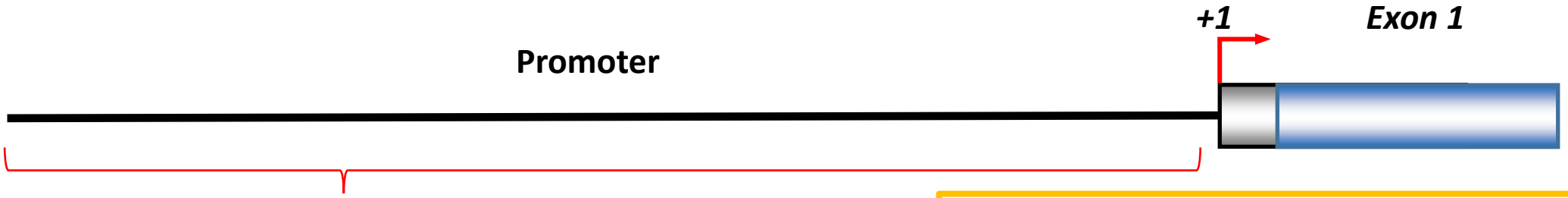


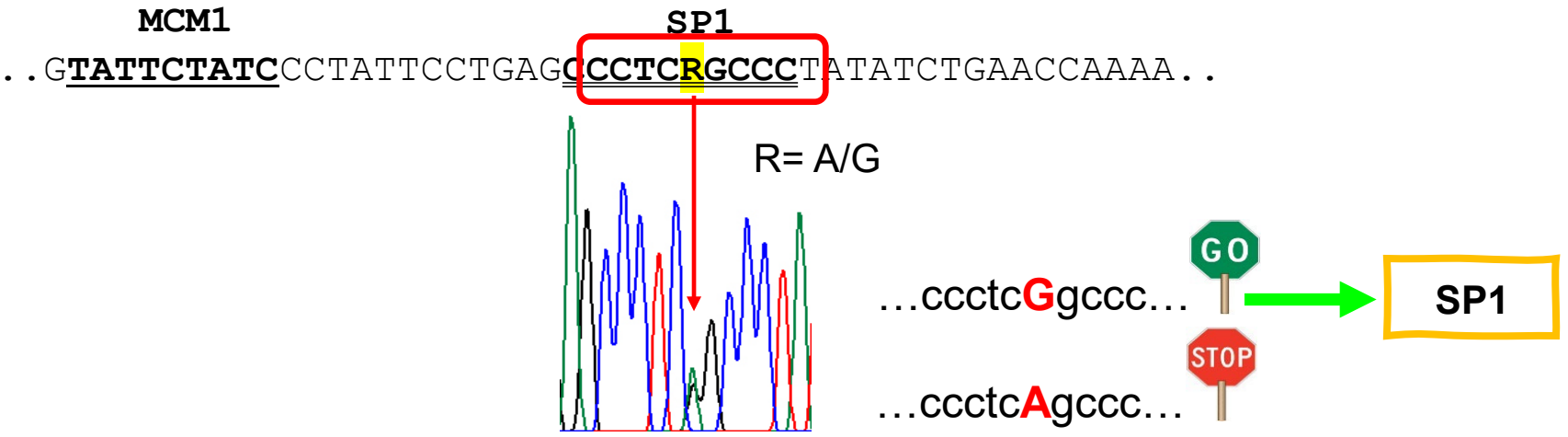
Table 5. List of transcription factors binding sites detected in the 5′-flanking regions of *LALBA* in domestic camelids[1]

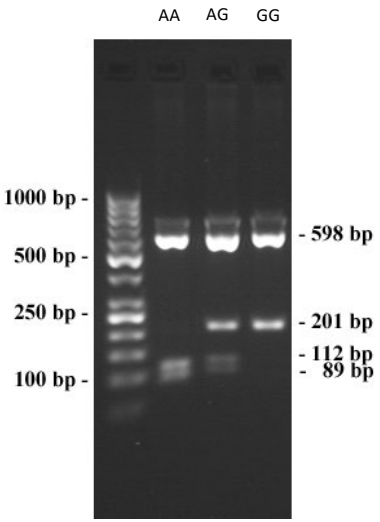
		Transcription factor binding site		
Position	Consensus motif 5′→3′	Old World camel	New World camel	Shared
−580/−572	wwTvCCTAwT	MCM1		
−558/−549[2]	CCCkCsGCmy			SP1[3]
−323/−314[2]	rGsCTGGGGm			AP2-α
−314/−305	ACwTCCTsyk	PEA-3		
−183/−174	GGAAArkGrm		ICSBP/ISGF-3	
−150/−141	syTGGCmGsC	NF-1		
−102/−93	CCTCyTCCys		SP1	
−80/−71	rrkGsCAGGG			YY1
−62/−52	sATGmATrwA			PIT-1A
−57/−48	ATAAAArrnG			TATA
−11/−2	wksmAGCCAA		NF-1	

[1]k = G/T; m = A/C; n = A/C/G/T; r = A/G; s = C/G; w = A/T; y = C/T.
[2]SP1 and AP2-α positions in New World camels are −557/−548 and −313/−322, respectively.
[3]Presence of the SP1 binding site in New World camelids is genotype dependent due to the polymorphism g.112A>G.

The SNP g.-553A>G creates a putative binding site for the *TF* SP1, genotype dependent in alpacas. This motif is a well-known enhancer element for the basal expression of many genes, including milk protein genes.

SNP g.-553A>G at alpaca LALBA				Allele frequency	
AA	AG	GG	TOT	A	G
41.25	41.25	17.50	100	0.62	0.38

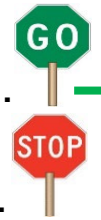




Results

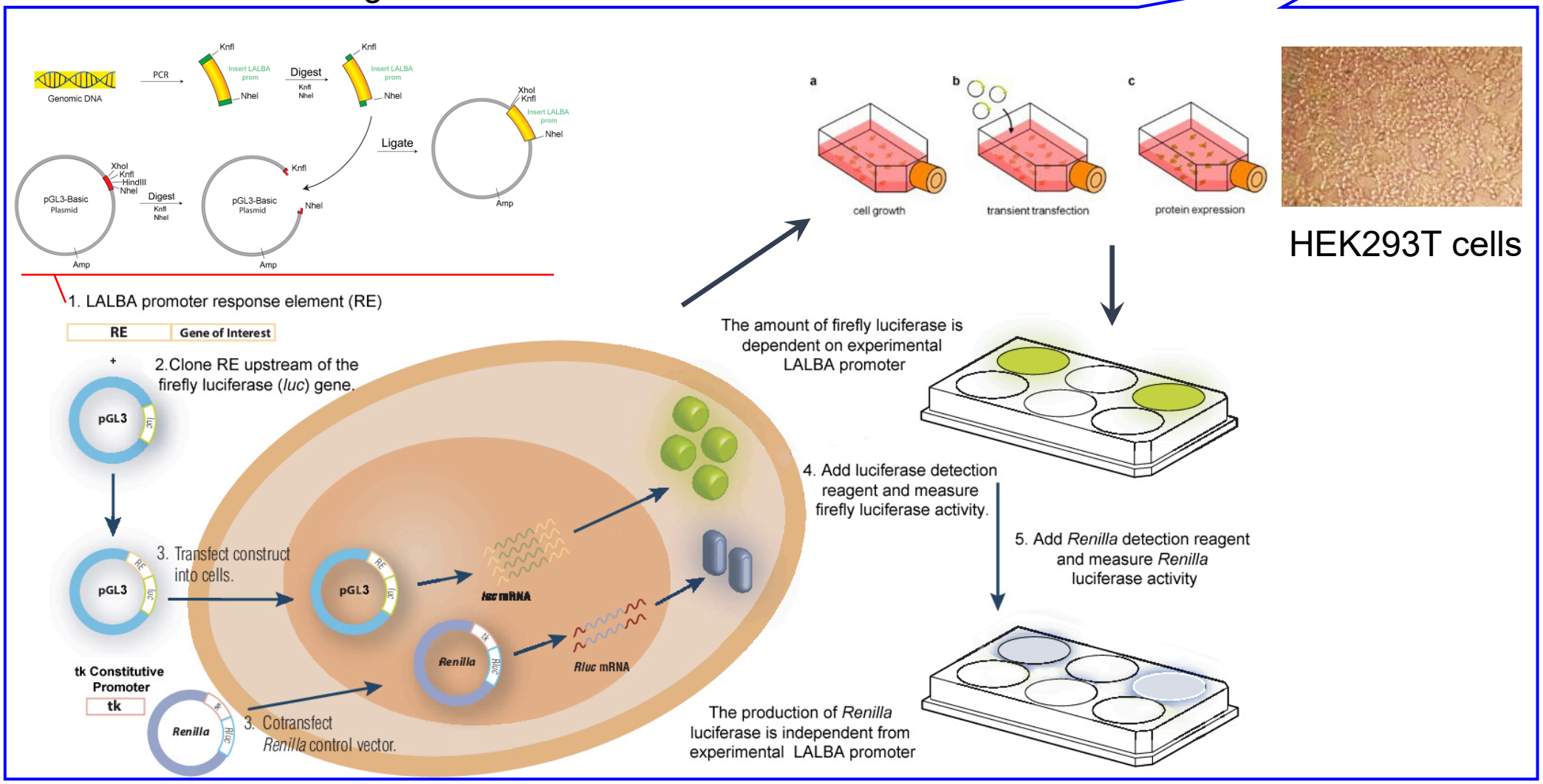
…ccctcGgccc…

…ccctcAgccc…



SP1 transcription factor

Gene Reporter Assay

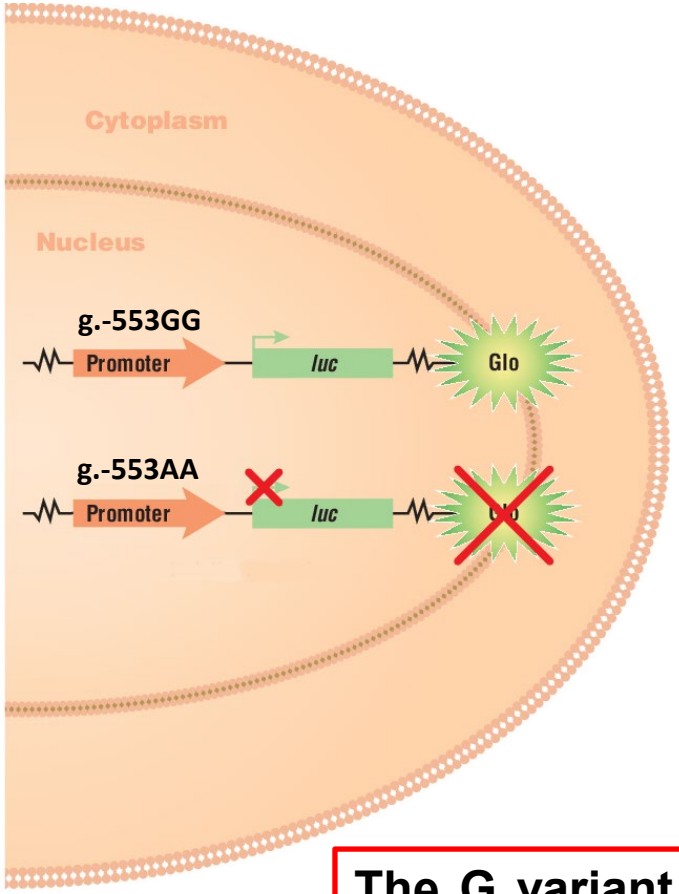


Results

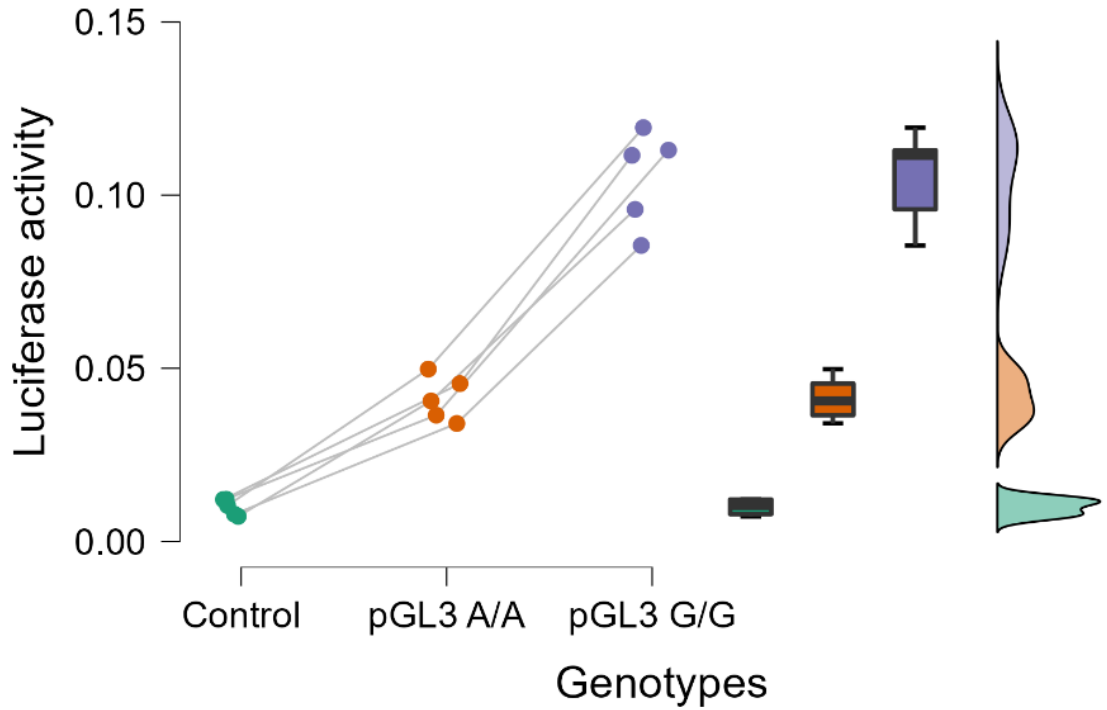
…ccctcGgccc…

…ccctcAgccc…

SP1
transcription factor



In vitro analysis



The G variant of this SNP enhances the promoter activity of the alpaca *LALBA* ($p<0.01$) and induces the SP1 binding, which consequently up-regulates *in vitro* the luciferase gene expression.

Conclusion

- ✓ A considerable number of SNPs has been found in *CSN1S2* and *LALBA* genes and their promoters, especially in the SAC that confirms an **higher level of diversity** of these genomes compared to those of Old World camels.
- ✓ We identified the **first allelic variants** at protein level (**p.36Ile>Met** in Bactrians and **p.78Ile>Val** in SAC) as a result of a SNPs (g.3639C>G and g.1229A>G) found in the exon 6 and exon 2 in *CSN1S2* and *LALBA* genes, respecitively.
- ✓ In the dromedary, the SNP g.15110G>T, found in the exon 17 of *CSN1S2*, shows promise to **functionally impact miRNA 4662a-3p**. The low frequency of the g.15110G allele offers an opportunity for genetic improvement through molecular-assisted selection.
- ✓ A gene reporter assay demonstrated **higher luciferase expression** as result of a **new SP1 binding** site generated by the **G variant** of the SNP g.-553A>G found in the LALBA promoter.
- ✓ The low frequency of both g.15110G and g.-553G alleles offers an opportunity for genetic improvement through **molecular-assisted selection**.
- ✓ The results here obtained provide profitable information for the **genetic progress** of domestic camels in milk proteins essential for **newborn/cria resilience, human nutrition and dairy purposes**.

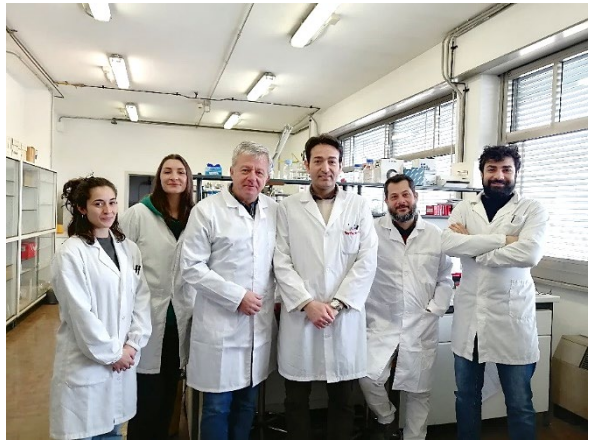
Conclusion

Full manuscripts are available:

Pauciullo, A.; Versace, C.; Gaspa, G.; Letaief, N.; Bedhiaf-Romdhani, S.; Fulgione, A.; Cosenza, G. Sequencing and characterization of αs2-casein gene (*CSN1S2*) in the Old-World camels have proven genetic variations useful for the understanding of species diversification. *Animals* 2023, 13, 2805. https://doi.org/10.3390/ani13172805

Pauciullo, A.; Versace, C.; Miretti, S.; Giambra I.J.; Gaspa, G.; Letaief, N.; Cosenza, G. Genetic Variability among and within domestic Old and New World camels at the α-lactalbumin gene (LALBA) reveals new alleles and polymorphisms responsible for differential expression. *Journal of Dairy Science* 2024, 107, 1068-1084. https://doi.org/10.3168/jds.2023-23813

My research group



Acknowledgments

This work was financially supported by: KING BAUDOUIN FOUNDATION UNITED STATES

Thank you very much for your attention

University of Turin
Dip. Agricultural, Forest and Food Sciences DISAFA

Prof. Dr. Alfredo Pauciullo
Genetics and Cytogenetics

Largo Paolo Braccini 2 - 10095
Grugliasco (Turin) - Italy
Tel. +39 011 6708598
alfredo.pauciullo@unito.it
www.disafa.unito.it
alfredo_pauciullo2022

Alma Universitas Taurinensis