J Mol Evol (1998) 46:256–260

JOURNAL OF MOLECULAR EVOLUTION

Microsatellite Evolution: Testing the Ascertainment Bias Hypothesis

Allan M. Crawford,[1] Steven M. Kappes,[2] Korena A. Paterson,[1] Mauricio J. deGotari,[2] Ken G. Dodds,[1] Brad A. Freking,[2] Roger T. Stone,[2] Craig W. Beattie[2]

[1] AgResearch Molecular Biology Unit, Centre for Gene Research and Department of Biochemistry, University of Otago, PO Box 56, Dunedin, New Zealand
[2] USDA ARS Meat Animal Research Center, PO Box 166, Clay Center, NE 68933, USA

Received: 6 January 1997 / Accepted: 19 May 1997

Abstract. Previous studies suggest the median allele length of microsatellites is longest in the species from which the markers were derived, suggesting that an ascertainment bias was operating. We have examined whether the size distribution of microsatellite alleles between sheep and cattle is source dependent using a set of 472 microsatellites that can be amplified in both species. For those markers that were polymorphic in both species we report a significantly greater number of markers ($P < 0.001$) with longer median allele sizes in sheep, regardless of microsatellite origin. This finding suggests that any ascertainment bias operating during microsatellite selection is only a minor contributor to the variation observed.

Key words: Microsatellite — Sheep — Cattle — Ascertainment bias

Introduction

When the size distribution of microsatellite alleles across different species is compared, the allele sizes in the species from which the microsatellite was derived are often greater than those found in closely related species. This observation has been reported for seven of eight domestic sheep-derived microsatellites when used in bighorn sheep (Forbes et al. 1995); 33 of 42 human-derived microsatellites when used in other primates (Rubinsztein et al. 1995); 10 of 14 dog microsatellites in foxes; and also amongst related species of swallows, cetaceans, ruminants, and turtles (Ellegren et al. 1995). These observations could result from either directional evolution occurring within different species (Rubinsztein et al. 1995, Amos and Rubinsztein 1996) or from an ascertainment bias in the selection of clones for sequencing and eventual primer pair development (Ellegren et al. 1995).

A critical test of the ascertainment bias hypothesis is to examine microsatellite repeat lengths following reciprocal amplification of microsatellites derived from both species. Sheep and cattle provide large numbers of dinucleotide repeat microsatellites (Bishop et al. 1994; Crawford et al. 1995; Stone et al. 1995; de Gortari et al. 1996 [in press]). Approximately one-third of microsatellites from either species are polymorphic in the other, providing a useful dataset in which to test this hypothesis.

Materials and Methods

Animal Populations. Two sheep populations were used in this study. The first comprised the 15 unrelated animals that are grandparents of the AgResearch International Mapping Flock (Crawford et al. 1995). These animals were derived from a variety of European sheep breeds: Texel (one animal), Coopworth (four), Perendale/Coopworth Cross (seven), and Merino/Romney Cross (three). The second population, unrelated parents of a USDA resource flock, comprised five Romanovs, two Rambouillet/Romanov crosses, and two Suffolk/Romanov crosses. We used unrelated beef (nine Angus/Hereford Cross) and dairy (three Friesian) cattle from New Zealand as well as the 28 un-

Correspondence to: A.M. Crawford; e-mail crawforda@agresearch.cri.nz

Table 1. Markers with significant allele size differences showing the origin of the marker and the species with the larger allele sizes

A. Makers polymorphic and with syntenic locations in both species

Origin of the marker	Larger alleles in Bovine	Larger alleles in Ovine	*P* value
Ovine	4	16	0.012
Bovine	84	159	<0.0001

B. Markers polymorphic in both species but unmapped in sheep

Origin of the marker	Larger alleles in Bovine	Larger alleles in Ovine	*P* value
Bovine	26	39	0.136

C. Markers polymorphic in the species of origin and monomorphic in the other

Origin of the marker	Larger alleles in Bovine	Larger alleles in Ovine	*P* value
Ovine	0	3	0.250
Bovine	109	22	<0.0001

D. Markers monomorphic in the species of origin and polymorphic in the other

Origin of the marker	Larger alleles in Bovine	Larger alleles in Ovine	*P* value
Ovine	1	0	1.000
Bovine	2	7	0.180

E. All markers that could be amplified in both species (polymorphic or monomorphic)

Origin of the marker	Larger alleles in Bovine	Larger alleles in Ovine	*P* value
Ovine	5	19	0.007
Bovine	221	227	0.813

related animals from the USDA linkage mapping pedigrees (Bishop et al. 1994) which included a variety of meat breeds from both *Bos taurus* and *Bos indicus* species.

Microsatellite Analysis. Details of all microsatellites used in this study (Appendix I) are published (Bishop et al. 1994; Crawford et al. 1995; Stone et al. 1995; Kemp et al. 1996). This data can also be accessed at the following WWW sites:

http://sol.marc.usda.gov/genome/cattle/cattle.html
http://dirk.invermay.cri.nz/docs/sheepgbase/manager.html

The microsatellite markers were all obtained from small-insert genomic DNA libraries in either M13 or plasmids probed with (AC)n DNA probes. The microsatellites beginning BMS were derived from a library in which selection for (AC)n repeats had occurred (Stone et al. 1995) but the remainder were screened from unselected libraries of random genomic fragments. A DNA sequencing ladder was used to estimate the size of microsatellite alleles in cases where the original cloned microsatellite allele was not available to act as a standard for the amplified allele.

Statistical Analysis. The Mann Whitney U test was used to determine whether there was a significant species difference in allele size for each marker. The binomial test was used to assess whether the number of markers that had significantly larger median allele sizes in sheep vs the number of markers that had significantly larger median allele sizes in cattle were equally divided. Estimates of gene diversity (heterozygosity) were made using the method of Weir (1990).

Results and Discussion

With the exception of three centromeric fusions and one translocation (Crawford et al. 1995) the cattle and sheep linkage maps appear similar in marker order. To ensure that we were examining the same locus in both species we initially used only markers that were polymorphic in both species and mapped to a syntenic position in both species. Two hundred sixty-two of the 300 microsatellites polymorphic in both species showed a significant species difference in median microsatellite fragment size ($P < 0.05$). Regardless of species of origin of the marker, a significantly larger median fragment size was observed in sheep than in cattle (Table 1A). These data are not consistent with ascertainment bias being the predominant factor in determining species differences in microsatellite fragment length and may indicate that some form of directional evolution of microsatellites is occurring.

We also examined the mean of the median allele lengths and the mean of the ranges in fragment size (Table 3). Once again regardless of whether they were derived from cattle or sheep the mean of the median allele lengths and the mean of the ranges were larger in sheep. For the bovine derived markers this difference was significant for the range ($P = 0.02$).

An additional 65 of 81 cattle microsatellites identified

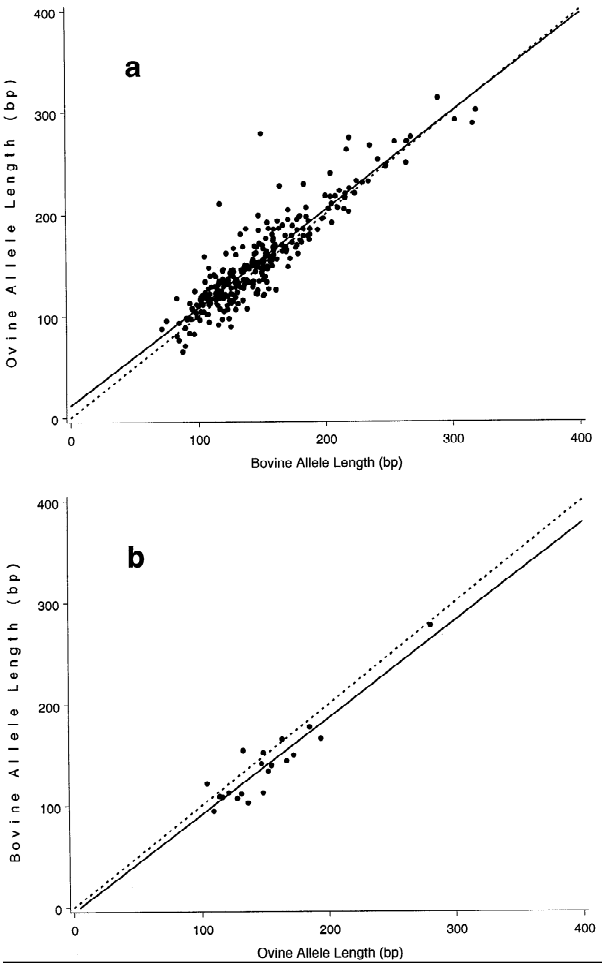

Fig. 1. The relationship between mean allele sizes in cattle and sheep; the mean allele size for the species in which the marker was derived is on the horizontal axes. The *dotted line* represents no difference, while the *solid line* shows the fitted regression. **A** Bovine markers. The regression line is (SE in *brackets*) ovine allele length = 11.7 (4.0) + 0.963 (0.026) × bovine allele length. **B** Ovine markers. The regression line is bovine allele length = −4.5 (14.0) + 0.955 (0.090) × ovine allele length.

as polymorphic in both sheep and cattle but not yet placed on the sheep linkage map were also significantly different in median allele length. Once again more of these microsatellites had significantly larger median allele sizes in sheep although the difference was not significant (Table 1B).

As an additional check for ascertainment bias, we compared the median allele sizes in the two species. Ascertainment bias would yield a regression slope less than 1 when the size for the new species is regressed against the size in the species in which the marker was derived (Fig. 1). Although both regression slopes were less than 1 neither was significantly so, again suggesting that ascertainment is not the main determinant of allele size differences between these two species.

To check the possibility that we were creating additional bias by choosing only those markers that have remained or become polymorphic in both species since

Table 2. Genetic diversity of microsatellites derived from cattle and sheep compared in cattle and sheep

		Ovine origin	SE	Bovine origin	SE	Total	SE
Mean	Ovine	0.649	0.043	0.640	0.009	0.640	0.011
	Bovine	0.547	0.050	0.619	0.010	0.615	0.010

sheep and cattle diverged, we identified a group of markers that were polymorphic in one species but monomorphic in the other with a typical ''stutter'' band appearance. We have assumed that the predominant fragment represented the correct locus but cannot independently verify this from a map location. The markers were divided into two groups. Group 1 (Table 1C) contained markers that were polymorphic in the species of origin and monomorphic in the other whereas group 2 (Table 1D) contained a small number of markers monomorphic in the species of origin and polymorphic in the other species. One hundred forty-four of 151 showed a significant difference in median fragment size, with significantly more bovine derived microsatellites fragments larger in cattle than in sheep (111 vs 29). Three of four ovine microsatellites were larger in sheep. When we combined all datasets significantly more sheep derived microsatellites are larger in sheep with no significant difference in microsatellites derived from cattle (Table 1E).

The recent publication (Amos et al. 1996) of additional human microsatellite mutations extends the observation that the majority of microsatellite mutations result from insertions of one repeat unit (Weber and Wong 1993). Recent studies in swallows (Primer et al. 1996) and sheep (Crawford and Cuthbertson 1996) also support an expansion model for microsatellite mutation.

Amos et al. (1996) also found that mutations tend to occur in individuals where the size difference between alleles is large and on that basis proposed a heterozygote expansion model to account for the difference in allele sizes between species. Amos et al. (1996) also suggest that the rate of microsatellite allele expansion may be related to the genetic diversity of a population, with the more diverse populations having longer microsatellites. If the population of a species is large and expanding, the proportion of heterozygous individuals and hence the probability of a mutation (which will most likely be an expansion) occurring and being maintained in the population is increased.

The mean gene diversity of the microsatellites polymorphic in both species, regardless of source, was higher in sheep compared to cattle although not significant ($P = 0.08$, Table 2). Early studies with human microsatellites (Weber 1990) showed that there is a positive relationship between allele length and gene diversity. This has also been shown for sheep (Buchanan et al. 1993). Measurements of gene diversity in any species are influ-

Table 3. Means of median allele lengths and ranges of microsatellites derived from cattle and sheep compared in cattle and sheep

		Ovine origin	SEM	Bovine origin	SEM
Mean of the medians	Ovine	150.4	9.1	153.8	2.8
	Bovine	139.3	9.0	147.0	2.6
Mean of the ranges	Ovine	18.5	2.5	18.9	0.9
	Bovine	12.5	2.3	16.3	0.6

enced greatly by the population sample used to assess the diversity. In this case we have made use of the founder animals from pedigrees designed as reference mapping populations. The reference mapping families have been designed to be genetically heterogeneous so the gene diversity measurements should be regarded as approaching the maximum for any population derived from either sheep or cattle. Different results could have been obtained by using within-breed measurements, and clearly a much wider study of gene diversity in sheep and cattle is required before solid comparisons of microsatellite diversity can be drawn between these two species. Perhaps the allele size differences found in our current study reflect predomestication differences in gene diversity that have been masked by the controlled breeding introduced to the species since domestication.

The sheep and cattle data presented here do not support ascertainment bias as the major reason for allele length differences except where the microsatellite is monomorphic in the heterozygous species. All our results, however, could be explained by the heterozygote expansion model. The larger allele sizes of cattle derived microsatellites that were monomorphic in sheep provide support for the expansion of polymorphic microsatellite loci compared with regions of the genome that remain monomorphic. The finding that markers polymorphic in both species had larger fragment sizes in sheep regardless of origin is consistent with the genetic diversity of sheep being higher than that of cattle, although further study is needed to verify that this is so.

This heterozygote expansion model could also explain the early results (Forbes et al. 1995; Rubinsztein et al. 1995; Ellegren et al. 1995) which suggested that ascertainment bias might be occurring. Most of the mammalian microsatellites used were from species that are very abundant and easily collected, such as rodents, domestic animals, and humans. These microsatellites were then tested in species with lower effective population sizes which survive in geographically limited habitats such as primates other than humans, foxes, and wild sheep species. As a consequence they are likely to be genetically less heterogeneous. According to the model, the mean allele sizes would be smaller in these species, and this has led to the conclusion that ascertainment bias was the major determinant of mean allele size.

In summary, therefore, our data from this species comparison suggest ascertainment bias during the cloning and characterisation of microsatellites is not having a large effect on their allele sizes. Some evidence for the directionality of microsatellite evolution is provided by the significantly greater number of microsatellites with larger median alleles sizes in sheep compared to cattle. What drives this evolution remains unclear but the heterozygote expansion model of Amos et al. (1996) is not excluded by our data.

References

Amos W, Rubinsztein DC (1996) Microsatellites are subject to directional evolution. Nat Genet 12:13–14

Amos W, Sawcer SJ, Feakes RW, Rubinsztein DC (1996) Microsatellites show mutational bias and heterozygote instability. Nat Genet 13:390–391

Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins GA, Toldo SS, Fries R, Grosz MD, Yoo J, Beattie CW (1994) A genetic linkage map for cattle. Genetics 136:619–639

Buchanan FC, Littlejohn RP, Galloway SM, Crawford AM (1993) Microsatellites and associated repetitive elements in the sheep genome. Mamm Genome 4:258–264

Crawford AM, Cuthbertson RP (1996) Mutations in sheep microsatellites. Genome Res 6:876–880

Crawford AM, Dodds KG, Ede AJ, Pierson CA, Montgomery GW, Garmonsway HG, Beattie AE, Davies K, Maddox JF, Kappes SW, Stone RT, Nguyen TC, Penty JM, Lord EA, Broom JE, Buitkamp J, Schwaiger W, Epplen JT, Matthew P, Matthews ME, Hulme DJ, Beh KJ, Mcgraw RA, Beattie CW (1995) An autosomal genetic linkage map of the sheep genome. Genetics 140:703–724

de Gortari MJ, Freking BA, Kappes SM, Leymaster KA, Crawford AM, Stone RT, Beattie CW (1996) Extensive genomic conservation of cattle microsatellite heterozygosity in sheep. Anim Genet (in press)

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias. Nat Genet 11:360–362

Forbes SH, Hogg JT, Buchanan FC, Crawford AM, Allendorf FW (1995) Microsatellite evolution in congeneric mammals. Mol Biol Evol 12:1106–1113

Kemp SJ, Hishida O, Rink A, Longeri ML, Ma RZ, Da Y, Lewin HA, Barendse W, Teale AJ (1995) A panel of polymorphic bovine, ovine and caprine microsatellite markers. Anim Genet 26:299–306

Primmer CR, Ellegren H, Saino N, Moller AP (1996) Directional evolution in germline microsatellite mutations. Nat Genet 13:391–393

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li S-H, Margolis RL, Ross CA, Ferguson-Smith A (1995) Microsatellite evolution—evidence for directionality and variation in rate between species. Nat Genet 10:337–343

Stone RT, Pulido JC, Duyk GM, Kappes SM, Keele JW, Beattie CW (1995) A small insert genomic library highly enriched for microsatellite repeat sequences. Mamm Genome 6:714–724

Weber JL (1990) Informativeness of human (dC-dA)n. (dG-dT)n polymorphisms. Genomics 7:523–230

Weber J, Wong C (1993) Mutation of short tandem repeats. Hum Mol Genet 8:1123–1128

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland, MA

Appendix I: List of Markers Used in the Analysis

A. Markers polymorphic and mapped in both species:

Markers showing no significant difference in median allele size
ADCY2, BL41, BM1520, BM1818, BM226, BM304, BM6404,

BM6444, BM6466, BM6506, BMS108, BMS1237, BMS1341, BMS1616, BMS2263, BMS356, BMS500, BMS501, BMS574, BMS710, BMS820, BMS882, BMS941, BRRIBO, CSSM032, FASMC2, HRH1, HUJ616, HUJII77, IDVGA46, ILSTS004, ILSTS070, INRA132, INRA144, INRA192, MAP2C, MB116.

Markers with the median allele size greater in cattle
ARO28, BL1095, BL42, BM17132, BM1861, BM1862, BM1905, BM2613, BM2830, BM302, BM4107, BM4129, BM4509, BM5004, BM6302, BM7234, BM757, BM8124, BM8151, BM8246, BM9289, BMC2228, BMS1048, BMS1120, BMS1185, BMS119, BMS1355, BMS1385, BMS1932, BMS1948, BMS2104, BMS2131, BMS2145, BMS2361, BMS2626, BMS2658, BMS390, BMS468, BMS522, BMS585, BMS651, BMS693, BMS695, BMS703, BMS745, BMS778, BMS861, BMS862, BMS938, BMS963, BOLA-DRB, BOLA-PSE, BP1, BP34, BP7, BR215, BR6504, CSSM041, CSSM047, CSSM065, ETH225, HEL11, ILSTS002, ILSTS008, ILSTS013, ILSTS027, ILSTS043, ILSTS049, ILSTS050, ILSTS059, ILSTS065, ILSTS102, INRA049, INRA063, INRA071, INRA131, INRA194, KRT10, MAF50, OarFCB11, OarFCB20, OarFCB48, RM067, RM150, TEXAN-10, TGLA122, TGLA261, TGLA337.

Markers with the median allele size greater in sheep
ACC08, AGLA269, AGLA29, BL1080, BL25, BL4, BL50, BL6-1, BM1225, BM1227, BM1258, BM1303, BM143, BM1577, BM2023, BM203, BM2113, BM2504, BM2901, BM2934, BM3011, BM3033, BM3205, BM3215, BM3412, BM3501, BM3509, BM4005, BM4006, BM4025, BM4208, BM4301, BM4621, BM4630, BM6041, BM6465, BM6526, BM7109, BM7144, BM7145, BM719, BM737, BM81124, BM8125, BM8225, BM8230, BM827, BMC1222, BMC5221, BMC6004, BMS1004, BMS1008, BMS1126, BMS1148, BMS1172, BMS1232, BMS1242, BMS1247, BMS1290, BMS1304, BMS1316, BMS1318, BMS1350, BMS1494, BMS1591, BMS1620, BMS1636, BMS1669, BMS1678, BMS1694, BMS1714, BMS1724, BMS1782, BMS1788, BMS1789, BMS1820, BMS1953, BMS2072, BMS2079, BMS2168, BMS2196, BMS2258, BMS2319, BMS2321, BMS2377, BMS2572, BMS2815, BMS332, BMS345, BMS357, BMS360, BMS397, BMS419, BMS424, BMS431, BMS434, BMS460, BMS482, BMS513, BMS517, BMS528, BMS538, BMS631, BMS648, BMS678, BMS740, BMS772, BMS792, BMS807, BMS812, BMS823, BMS835, BMS875, BMS887, BMS907, BMS975, BMS995, BP28, BP31, BP33, CSSM003, CSSM004, CSSM025, FCB193, FCB304, HBB, HEL10, HUJ246, IDVGA45, IL2RA, ILSTS005, ILSTS011, ILSTS017, ILSTS018, ILSTS019, ILSTS020, ILSTS022, ILSTS029, ILSTS030, ILSTS044, ILSTS056, ILSTS058, ILSTS087, INRA006, INRA011, INRA081, INRA111, INRA133, INRA135, INRA175, MAF23, MAF45, MAF65, MAF70, MAF92, MCM130, MCM58, MCM74, OarCP26, OarCP34, OarHH22, OarVH54, OCAM, OMHC1, POTCHA, RBP3, RM004, RM065, RM106, RM356, TEXAN-2, TGLA429, XBM11, XBM24.

B. Bovine markers polymorphic in both species but unmapped in sheep

Markers showing no significant difference in allele size
AGLA232, BM2814, BM3406, BMS1248, BMS1332, BMS1787, BMS1915, BMS2055, BMS2200, BMS2213, BMS2742, BMS4045, BMS462, BMS744, ILSTS026, ILSTS061

Markers with the median allele size greater in cattle
BL1022, BM148, BM6121, BM6507, BM720, BM7241, BM8118, BMS1145, BMS1878, BMS2, BMS2460, BMS2598, BMS362, BMS4011, BMS483, BMS689, BMS948, BR2936, BY10, BY5, INRA100, INRA122, INRA183, RM209, Z27075, Z27076

Markers with the median allele size greater in sheep
BL1009, BM121, BM7237, BM7247, BM746, BM856, BM9248, BMS1617, BMS1660, BMS1779, BMS1967, BMS2076, BMS2270, BMS2355, BMS2466, BMS2526, BMS2569, BMS2614, BMS2641, BMS2721, BMS2780, BMS2833, BMS2843, BMS4000, BMS4001, BMS4008, CSSM019, HAUT14, IGF-1, ILSTS028, ILSTS053, ILSTS060, INRA035, INRA107, INRA129, JAB1, RM024, UWCA28, UWCA46

C. Markers polymorphic in only one species

No significant difference in median allele size
BM1508, BM3010, BM315, BM7225, BMS904, ILSTS062, TEXAN-4

Markers with the median allele size greater in cattle
AGLA17, AGLA227, AGLA233, BL1103, BL28, BL37, BM103, BM1557, BM1832, BM1857, BM1864, BM2515, BM2607, BM3026, BM310, BM4307, BM4321, BM6026, BM6436, BM6458, BM7160, BM7207, BM7208, BM7228, BM733, BM741, BM8139, BM8247, BM888, BM9065, BM9138, BM9202, BM9208, BM9284, BMC5227, BMCR17A, BMS1074, BMS1192, BMS1231, BMS1282, BMS1296, BMS130, BMS1300, BMS1353, BMS1373, BMS1758, BMS1926, BMS1943, BMS1987, BMS2053, BMS2060, BMS2095, BMS2137, BMS2503, BMS2567, BMS2573, BMS2891, BMS382, BMS4018, BMS466, BMS490, BMS499, BMS504, BMS511, BMS518, BMS529, BMS803, BMS817, BMS911, BMS918, BMS929, BMS937, BOLA-DR2, BP38, CSSM033, CSSM039, ETH10, HAUT1, HU414, IAP, IL4, ILSTS012, ILSTS023, ILSTS035, ILSTS036, ILSTS037, ILSTS045, ILSTS068, ILSTS092, ILSTS104, INRA084, INRA112, INRA119, INRA120, INRA121, INRA177, JAB4, MAF46, RM012, RM019, RM033, RM074, RM088, RM090, RM137, RM330, TGLA170, TGLA179, TGLA245, TGLA28, UWCA19, UWCA20

Markers with the median allele size greater in sheep
AGLA280, BM1831, BM1834, BM3212, BM3507, BM4102, BM713, BM804, BM8129, BM861, BMS1101, BMS1315, BMS1561, BMS1580, BMS2914, BMS3002, BMS639, BMS657, BMS742, BMS960, BMS980, BR6027, OarCP16, HH41, ILSTS052, INRA090, MAF35, RM500, TEXAN-3 TEXAN-5, TGLA141, TGLA351