PRIFYSGOL ABERYSTWYTH UNIVERSITY

Aberystwyth University

A preliminary examination of the genetic variation within and between the Improvement Society herds of Welsh Mountain ponies

McMahon, Robert; Debbonaire, Aliyah; McEwan, Neil; Nash, Deborah; Davies-Morel, Mina; Winton, Clare Louise; Blackmore, Tina; Hegarty, Matthew

Publication date:
2015

Citation for published version (APA):
McMahon, R., Debbonaire, A., McEwan, N., Nash, D., Davies-Morel, M., Winton, C. L., Blackmore, T., & Hegarty, M. (2015). *A preliminary examination of the genetic variation within and between the Improvement Society herds of Welsh Mountain ponies*.

Document License
CC BY

General rights
Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400
email: is@aber.ac.uk

Download date: 09. Jul. 2020

PRIFYSGOL ABERYSTWYTH UNIVERSITY

IBERS
Athrofa y Gwyddorau Biolegol, Amgylcheddol a Gwledig
Institute of Biological, Environmental and Rural Sciences

Report prepared for the WPCS- 2015

A preliminary examination of the genetic variation within and between the Improvement Society herds of Welsh Mountain ponies

Rob McMahon, Aliyah Debbonaire, Neil McEwan, Debbie Nash, Mina Davies-Morel, Clare Winton, Tina Blackmore and Matthew Hegarty

Nov. 2015 - Draft

Report to the Welsh Pony and Cob Society, Nov. 2015

1.1 Executive summary

In order to understand the within and between herd diversity in the Upland Hill ponies of Wales and their relationships to their pedigree relatives and other native pony breeds a preliminary examination of a small sample of animals from a group of upland herds was carried out between 2014 and 2015.

Samples of DNA were obtained from 16 herds of Welsh Mountain Ponies from the Upland Improvement Societies. Around 130 animals were tested for genetic variation at 17 highly variable sites, 172 single base pair polymorphisms and by sequencing across 540 base pairs of the mitochondrial Control Region. These results were compared to a body of data obtained from Welsh Section A-D, and the Carneddau population and mixed groups of pedigree and non-pedigree UK ponies.

We have established that, while clearly related to the pedigree Section-A animals, it is possible to perform a genetic assignment of unknown Upland animals to their correct herd of origin in the majority of cases, and that the native Upland animals can be distinguished from several common 'pony crosses' of the sort that might be found 'abandoned' on Welsh common land.

We would argue that despite the presence of Upland blood-lines in the pedigree animals there are unique patterns of mitochondrial diversity within some upland herds that argues for there being a long term local stability to these populations. Our findings suggest that further work is required to ascertain to what extent the upland populations retain a unique genetic signature of Natural Selection *in situ*, and whether it is possible to exploit this, if present, to understand the particular adaptive complexes that make the Welsh Hill pony unique.

1.2 Background

The Welsh Pony and Cob Society (WPCS) is the official registry of UK native Welsh ponies and cobs. These animals are categorised into 4 sections according to height and conformation characteristics: Section A is a Welsh Mountain Pony, <121.9 cm; Section B is a Welsh Riding Pony of fine-built morphology, <137.2 cm; Section C is a Welsh Pony of Cob Type of stocky morphology <137.2 cm and Section D a Welsh Cob >137.2 cm. The stud book is closed, meaning that only animals bred from registered parents can be registered themselves.

The majority of these animals are kept, bred, produced and competed by dedicated breeders and owners with controlled management, often in lowland studs/farms. However, several geographically distinct, upland populations exist that, although managed, are semi-feral. Each of these populations is overseen by an equivalent WPCS-administered Welsh Mountain Hill Pony Improvement Society. For example, the pony population in the Brecon Beacons is overseen by the designated Brecon Beacons Hill Pony Improvement Society. There are 27 Hill Pony Improvement Societies that include populations that reside in harsh upland environments, such as the Brecon Beacons, as well as salt marsh environments on the coast such as the Gower.

The Hill Improvement Societies represent ponies that potentially provide a highly unique genetic resource, including hardy adaptations to specific environmental features which domestic horses and ponies may not express. Hill ponies may be adapted to harsh upland life, grazing and nutritional pressures or even extremes of salt exposure which may require specialised homeostasis. Obtaining an understanding of the gene complexes that contribute to these characteristics could provide a scientific basis for selection for improved survival of equine and possibly other upland grazing species and as a platform for a planned sustainable intensification of upland agriculture. Furthermore, the locales where these animals live are also directly dependant on the ponies because their interaction with the environment is integral for maintaining unique habitat ecology.

The importance of these animals as a genetic resource is substantiated by the fact that breeders of registered Welsh Ponies draw upon the hill ponies to improve their stock, using them to breed in ‘hardiness’ otherwise assumed to have been bred out over time in lowland populations.

A ‘premium’ scheme has run previously whereby stallions representing Hill Pony Societies are assessed annually at Glanusk Stallion Show. Stallions judged to be worthy of a premium are approved to run out on the hills with populations included in the scheme. It is often the case that stallions are moved from hill population to hill population so that more the one group of hill ponies benefits from his ‘approved’ genetic input. Each year in the autumn, Hill Improvement Societies collect (round up) the animals to remove the colts, before returning the remaining individuals to the hillside.

The premium scheme is financially supported charitably by the Horseracing Betting Levy Board (HBLB) and qualifies because the ponies are recognised by the Rare Breeds Survival

Trust (RBST). However, the RBST have removed their recognition and without further evidence that the hill ponies are genetically distinct from registered Welsh Ponies and Cobs, this cannot be reinstated. As such the HBLB funding will no longer be available to maintain the premium scheme and encourage breeders/owners to continue The RBST based the decision to delist the Welsh Mountain Pony Section A (semi-feral) as a separate breed on the premise that the continued use of upland animals as breeding stock intended to input 'hardiness' to the lowland stud pedigree animals, effectively means that semi-feral blood lines are well represented amongst pedigree Section A Welsh ponies.

This study is intended as a preliminary investigation of the genetic diversity of current Hill Pony Improvement Society herds and their relationships to the Section A Welsh ponies, the semi-feral ponies of the Carneddau plateau and other Native British pony breeds.

1.3 Methodology overview.

Samples of hair roots were collected from each of 16 Hill pony improvement Society herds from across Wales. The locations of the sampled were approximately as indicated on the map in Figure 1. Herds 1-10 were sampled by Emily Ham as part of her MSc study in 2013/14 and the rest were collected by members of the WPCS in the 2014 roundups. Around 20-40 individual hairs were pulled from the mane or tail of animals as they were handled during the annual roundups and inspections. These were placed into small bags and transported to the laboratories at Aberystwyth for storage.

DNA was extracted from a subsample of these hair-pulls using standard techniques and quantified for further analysis and long term storage.

Figure 1. Location of Improvement Society herds sampled in this study. The numbers are representative of Hill Pony Improvement Societies from which samples were collected; 1. Begwyns, 2.Black Mountain, 3.Brecon Beacons, 4. Drum Hill, 5. Hergest Hill, 6. Llanafan & Llanwrthwl, 7. Llandefalle, 8. Llangoed, 9. Llynyfan(Gwynfe), 10. Presili. 11. Mynydd Trefil Ddu and Las, Cefn Edmwnt , Pontlottyn 12 Dowlais Hill, 13 Llanrhidian Marsh, Cenydd Gwyr The 'X' donates university and laboratory facilities. (adapted from Ham 2014)

Four approaches to examining the genetic diversity in the populations were investigated.

1. Pedigree data were collected and analysed for levels of predicted consanguinity in the animals present in the database.
2. A sample of animals from several herds were genetically typed for the Stockmark panel of 17 Simple Sequence Repeats (SSRs, also known as Variable Number of Tandem Reat, VNTR, markers).
3. A sample of animals were tested for a panel of Single Nucleotide Polymorphisms (SNPs) determined from previous studies to have a high decrimination ability between the Welsh pedigree Sections and other horse breeds. The panel is a custom made Illumina infinium array, the AberBeef Chip designed to detect and discriminate between multiple species/breeds concurrently.
4. A 550 basepair fragment of the mitochondrial control region was amplified by PCR and unidirectionally sequenced on the Forward (Heavy) strand using standard Sanger Sequencing methodology on an ABI 3730 analyser.

Detailed methodologies used and protocols for these tests are available on request.

Section 2 Results.

2.1 Pedigree analysis.

Data supplied by the WPCS was entered into an Excel workbook and uploaded to the pedigree handling software *Pedigree ViewerVersion 6.5b* freely available from http://metz.une.edu.au/~bkinghor/pedigree.htm. The general output of this program is shown in Figure 2.1. demonstrating the multigenerational and interconnected nature of the upland ponies.

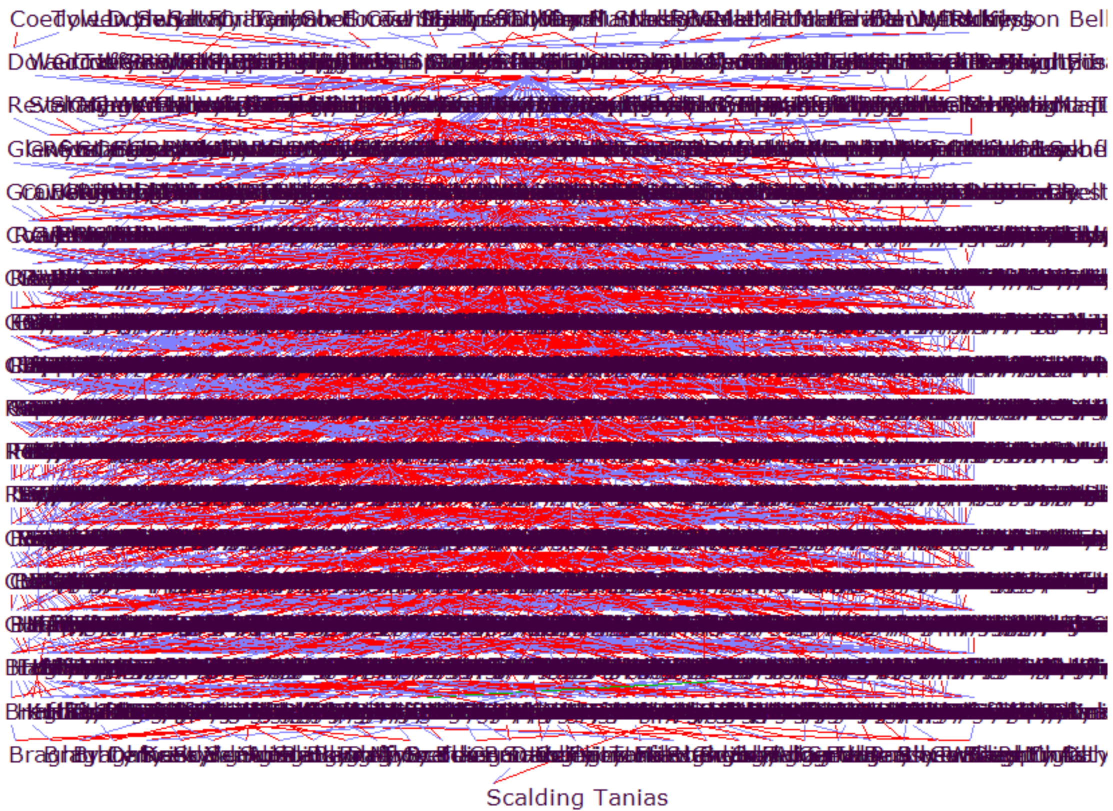


Figure 2.1 Overview of pedigrees for all Upland ponies.

Red and blue lines represent lines of descent paternal and maternal respectively

Several individuals have had a very widespread influence on the current animals – as just one example the diagram in Figure 2.2 demonstrates all the descendants within the overall pedigree that are related back to a single animal, 'Barley Sugar', who was born in 1937.

Figure 2.2 pedigree showing the descendants of Barley Sugar, an animal born in 1937.

In order to summarise these data we examined the degree of inbreeding within the pedigree as a whole assuming that all original animals were fully outbred and that any animals 'crossed into' the pedigree were fully outbred as well.

Figure 2.3 Showing individual estimates of level of inbreeding calculated from pedigree data for all upland horses include in the analysis.

Table 2.1 Estimated average inbreeding coefficient for animals born in each year from 1960-2014, based on pedigree data.

Average inbreeding	Year of birth
0.01786	1960
0.04791	1970
0.06068	1980
0.06088	1990
0.07574	1998
0.11030	1999
0.08856	2000
0.10052	2001
0.09171	2002
0.09591	2003
0.10534	2004
0.08394	2005
0.07391	2006
0.07578	2007
0.08960	2008
0.09751	2009
0.08672	2010
0.11022	2011
0.10291	2012
0.10137	2013
0.06129	2014

Although all animals born since 2008 are inbred to some degree there has been very little increase in the average level of inbreeding across the birth years since the 1970s.

Examining the average inbreeding of each of the studs represented in the analysis below indicates that although most herds have similar patterns of inbreeding there are some herd to herd differences, for example average predicted inbreeding amongst the 12 Llynyfan animals tested is 0.054 and amongst 12 Preseli animals 0.148.

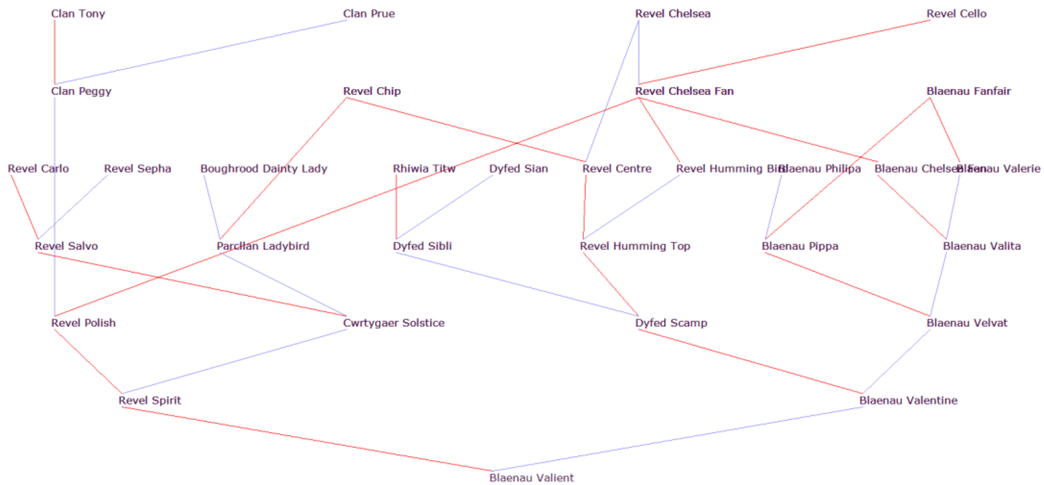


Figure 2.4 Showing a fairly typical (5 generation) pedigree for an inbreeding coefficient of 10%. Note the loops, through Revel Chelsea Fan and Revel Chip for example.

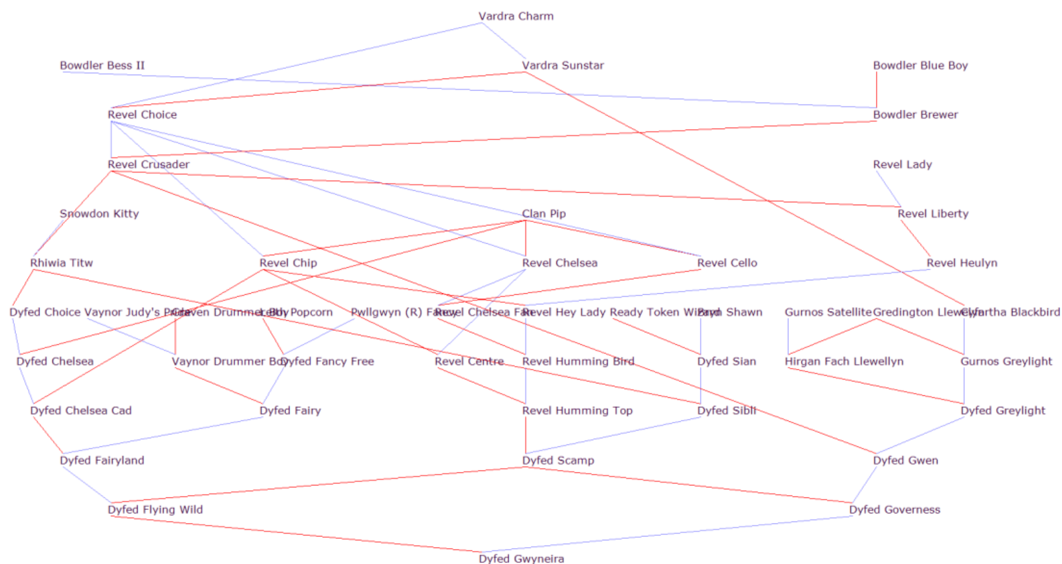


Figure 2.5 Showing a pedigree for an individual with inbreeding coefficient of 18.7%. Note again the number of loops involving individuals of the Revel blood line.

Figure 2.4 and 2.5 show relatively typical pedigrees involving inbred individuals from the Black Mountain (Blaenau) and Preseli Herd (Dyfed) respectively. These pedigrees demonstrate two things in particular. The first is the importance of the Revel stud in many of the individuals alive today, and the second is that as a consequence the individuals in different herds are predicted to share a degree of common genetic heritage, in this case between members of the Preseli and Black Mountain herds, mainly via the influence of stallions.

Section 2.2 Autosomal genetic variation – Variable Number of Tandem Repeats (VNTR) – the Stockmark markers

2.2.1 Methods

Initially 100 assays of the Stockmark-17 VNTR panel were purchased and applied across a group of sample DNA purified 10 herds by Emily Ham as part of her MSc project. These results were highly fragmentary and inconsistent and have been rejected as a source of reliable data for this study. Standard protocols were used according to the Standard Operating Procedures supplied with the kit.

It appeared that the most likely reason for the repeated failure of this procedure lay with the quality of the DNA since larger products were failing in the multiplex. It was felt unlikely that we would obtain higher quality DNA on a re-extraction of the remaining stored hair roots and it was decided to order individual primer pairs for each of the equine VNTR separately and to try amplification in smaller multiplex units.

2.2.2 Results

Despite repeated attempts the number of samples producing reliable results was somewhat disappointing. Results were obtained from 68 individuals spread across 13 herds, for between 4 and 15 SSR loci each. The inconsistent distribution of marker coverage and the small number of animals that successfully typed from some populations mean that the results from this section are extremely preliminary in nature. In addition the amplification in smaller multiplexes is known to affect the nature of the 'amplified products' and therefore in order to compare the actual allele calls obtained so far with those from previous testing with the stockmark kit we will need to run a set of control DNA of known genotype and use these to calibrate the absolute results obtained. However these results do allow comparison between the upland animals tested and they complement the results reported in section 2.3 below.

In general the results do demonstrate that there is a degree of autosomal differentiation between the upland herds, but that animals in some herds are more closely related than in others. This is illustrated in table 2.2 which shows that estimated probability of obtaining

two sample results as extreme as those found here by taking two random samples from the overall allele distributions in the data set.

Table 2.2
Probability results being drawn from a single hypothetical population, p-value for each population pair across all loci (Fisher's method). Highlighted values are significant at the p<0.05 level Data estimated from Genepop4.2 (Raymond & Rousset, 1995, Rousset, 2008).

Population pair		Chi2	df	P-Value
"Cui3"	& "CG5"	42.11353	28	0.042303
"Cui3"	& "PO4"	31.373627	26	0.214648
"CG5"	& "PO4"	38.901159	26	0.049825
"Cui3"	& "HH1"	21.308143	18	0.264153
"CG5"	& "HH1"	17.944492	18	0.459315
"PO4"	& "HH1"	15.208546	18	0.647601
"Cui3"	& " BEG28"	40.261218	22	0.010076
"CG5"	& " BEG28"	41.838576	22	0.006538
"PO4"	& " BEG28"	30.457481	20	0.062772
"HH1"	& " BEG28"	9.536029	16	0.889710
"Cui3"	& " BM5"	44.715937	28	0.023569
"CG5"	& " BM5"	50.818318	28	0.005234
"PO4"	& " BM5"	29.327598	26	0.296397
"HH1"	& " BM5"	12.217884	18	0.835787
" BEG28"	& " BM5"	27.261736	22	0.201454
"Cui3"	& " BRC9"	25.203885	24	0.394750
"CG5"	& " BRC9"	34.416939	24	0.077527
"PO4"	& " BRC9"	33.609691	24	0.091849
"HH1"	& " BRC9"	7.190076	16	0.969421
" BEG28"	& " BRC9"	22.158583	16	0.138145
" BM5"	& " BRC9"	25.596373	24	0.373958
"Cui3"	& " DH5"	28.612701	28	0.432363
"CG5"	& " DH5"	30.159240	28	0.355646
"PO4"	& " DH5"	26.620527	26	0.429436
"HH1"	& " DH5"	11.811101	16	0.756878
" BEG28"	& " DH5"	35.274245	20	0.018691
" BM5"	& " DH5"	25.329566	28	0.609845
" BRC9"	& " DH5"	18.996344	22	0.645554
"Cui3"	& " GWY9"	10.693952	6	0.098308
"CG5"	& " GWY9"	5.400308	6	0.493587
"PO4"	& " GWY9"	4.984682	6	0.545779
"HH1"	& " GWY9"	2.510597	6	0.867280
" BEG28"	& " GWY9"	17.705623	6	0.007012
" BM5"	& " GWY9"	5.323492	6	0.503041
" BRC9"	& " GWY9"	13.645551	6	0.033856
" DH5"	& " GWY9"	3.324848	6	0.767108
"Cui3"	& " LD9"	34.628645	16	0.004466
"CG5"	& " LD9"	48.845750	16	0.000035
"PO4"	& " LD9"	15.398035	16	0.495702
"HH1"	& " LD9"	15.583636	16	0.482378
" BEG28"	& " LD9"	19.864987	16	0.226363
" BM5"	& " LD9"	34.324437	16	0.004912
" BRC9"	& " LD9"	24.251909	16	0.084145
" DH5"	& " LD9"	38.730608	16	0.001189
" GWY9"	& " LD9"	1.953340	6	0.923939
"Cui3"	& " LF9"	80.838579	28	0.000001
"CG5"	& " LF9"	63.982617	28	0.000123
"PO4"	& " LF9"	Infinity	26	Highly sign.
"HH1"	& " LF9"	22.894674	18	0.194670

```
" BEG28"       & " LF9"        43.270535 22   0.004370
" BM5"         & " LF9"         Infinity 28   Highly sign.
" BRC9"        & " LF9"        37.954437 24   0.035047
" DH5"         & " LF9"        48.548876 28   0.009346
" GWY9"        & " LF9"        23.936514 6    0.000536
" LD9"         & " LF9"        30.756764 16   0.014451
"Cui3"         & " LL9"        49.906301 26   0.003226
"CG5"          & " LL9"         Infinity 26   Highly sign.
"PO4"          & " LL9"        30.520392 26   0.246615
"HH1"          & " LL9"        20.783555 18   0.290473
" BEG28"       & " LL9"        44.126324 18   0.000553
" BM5"         & " LL9"        36.377367 26   0.084944
" BRC9"        & " LL9"        34.055741 18   0.012397
" DH5"         & " LL9"        25.301993 22   0.282852
" GWY9"        & " LL9"        4.781323  6    0.572150
" LD9"         & " LL9"        38.239596 16   0.001399
" LF9"         & " LL9"         Infinity 26   Highly sign.
"Cui3"         & " PRE6"       22.784313 26   0.645119
"CG5"          & " PRE6"       24.347639 26   0.556094
"PO4"          & " PRE6"       27.363789 26   0.390454
"HH1"          & " PRE6"       6.876163  16   0.975575
" BEG28"       & " PRE6"       25.059088 18   0.123297
" BM5"         & " PRE6"       19.959294 24   0.699089
" BRC9"        & " PRE6"       11.786414 22   0.961634
" DH5"         & " PRE6"       12.663113 20   0.891366
" GWY9"        & " PRE6"       4.418833  6    0.620190
" LD9"         & " PRE6"       15.242529 16   0.506952
" LF9"         & " PRE6"       39.681760 26   0.041907
" LL9"         & " PRE6"       18.361029 22   0.684381
```

In agreement with the evidence from the pedigree section above there was no evidence of significant levels of disruption from Hardy Weinberg (HW) expectations which would indicate severe inbreeding or extreme isolation of some herds (Genepop4.2 , HW probability test with 100 batches of 1000 sampling iterations over-all loci Chi square = 120, df = 212, probability of results being in HW p = 1, not significant). Comparing all loci and populations independently and specifically for evidence of heterozygote deficiency, (characteristic of inbreeding or a subdivided population) indicates that all populations appear to be in HW apart from the Black Mountain animals where the probability of HW equilibrium is highly significantly rejected (p = 0.0081) indicating a deficit in heterozygotes amongst the Black Mountain animals tested).

The data were used to calculate the degree of 'private' alleles in each population and using this to estimate the average level of migration between herds as being about 0.85 migrants per generation between populations determined using the method of implemented in Genepop4.2.

This suggests that there is a considerable amount of geneflow between the upland herds, but is an average based on several assumptions.   Using the distribution and size of alleles in each

population an estimate of the 'genetic differentiation' between herds can be obtained. This value, Rho_{ST}, is analogous to Sewell Wright's Fixation index (Fst), and takes account of the number of different alleles within and between herds and also differences in sizes of those alleles according to a model of the genetic mutation process. Table 2.3 provides a summary of this value across all upland population pairs tested. Rho_{ST} varies from 0 to 1, but in general, values that are negative or very close to zero are taken to be zero and indicate no evidence of genetic variation between populations. Values of around 0.01-0.05 would be found between European populations of humans, 0.2 would indicate the same degree of genetic differentiation as that seen between Sub-Saharan African populations and East Asian populations and values > 0.35 would be typical of geographically isolated populations of naturally subdivided species. Values less than 0.1 are indicated in green highlight in Table2.3. The general pattern is for herds to be relatively similar to each other with a few showing more divergent results.

Table 2.3 Estimates of Rho_{ST} for all loci (diploid):

Indices for populations: 1 = "Cefn Edmwnt-Cui", 2 = "Cenydd Gwyr-CG", 3 = "Pontlottyn-PO", 4 = "Hergest Hill-HH", 5 = " Begwns-BEG", 6 = " Black Mountain-BM", 7 = " Brecon Beacons-BB", 8 = " Drum Hill-DH", 9 = " llynyfan (Gwynfe)-GWY", 10 = "Llandefalle-LD", 11 =" Llanafan and Llanwrthwl-LF", 12 = " Llangoed-LL", 13 = " Preseli-Pre"

pop	1	2	3	4	5	6	7	8	9	10	11	12
2	-0.0288											
3	0.1112	0.2101										
4	0.0615	-0.0611	0.2235									
5	0.4844	0.4271	0.5057	0.1049								
6	0.0016	-0.0134	0.1059	-0.1218	0.3043							
7	0.2074	0.1393	0.3633	-0.0515	0.2473	0.0887						
8	-0.0792	0.0108	0.0276	-0.0680	0.2843	0.0094	0.1186					
9	0.2037	0.0042	-0.0008	-0.0735	0.2039	0.0446	0.0286	0.0446				
10	0.2833	0.0527	0.4749	0.1058	0.2230	0.0653	0.1597	0.1302	-0.0028			
11	0.4066	0.3904	0.3440	0.0230	0.1455	0.1786	0.0388	0.2041	0.0170	0.1870		
12	0.1262	0.0722	0.2011	-0.0726	0.2437	-0.0002	0.1650	0.1153	-0.0323	0.1067	0.1227	
13	0.0904	0.0904	0.2688	-0.0841	0.4461	0.0387	0.1171	0.0270	-0.0921	0.2968	0.1751	0.1096

These values should be taken with considerable caution as some populations, for example the BEGWNS herd, are represented by very few datapoints and none of these results should be considered significant.

Section 2.3 Autosomal genetic variation - reduced genomic SNP panel, (Illumina AberBeef Chip)

2.3.1: Method

Previous genotyping of native pony breeds using the illumina Infinium equine50k Chip had established a large collection of Single Nucleotide Polymorphisms (SNPs) that showed specifically high deviations in allele frequency among horse breeds and the Welsh pony sections and Carneddau population of Welsh Mountain ponies. As part of a project into tracing deliberate adulteration of meat in the beef industry a panel 180 markers were chosen from these 50K Chip results and added to a custom designed Chip (the AberBeef Chip) to test for the ability to perform concurrent multi-breed identification across multiple mixed species DNA on the same array. These markers were chosen as the best diagnostic markers for ‘upland pony’ phenotype relative to larger ponies and horse breeds. So we decided to add a sample of 75 individuals from 12 of the Upland pony herds to these test panels. These results have the two fold advantage of allowing us to examine the genetic relationships between the Upland animals and other pony/horse breeds, (Table 2.5) but also permit the development of a probability based estimate of herd origin for any individual animal tested (Table2.6). This is because the 172 markers that type reliably are scattered across the genome of the horse and give approximately independent estimates of the genetic history of the individual.

The software program GeneClass2 (Piry *et al* 2004) was used to determine the relative likelihood of each sample belonging to any of the tested populations. This program operates by calculating the allele frequencies in each ‘reference population’. These reference frequencies are then used to compare the genotypes obtained for each test individual. The probability of each genotype is then calculated across all reference populations and across all tested loci to obtain an aggregate likelihood of an individual coming from each reference

population. These composite likelihoods are then compared to determine the predicted origin of the test sample. The nature of this process means that the accuracy of the assignment and the 'correctness' of the judgement is determined by several things, but particularly 1. The relative differences in allele frequencies between populations. 2. The accuracy of the estimated allele frequencies in the reference populations. Due to the restricted time and resources available, the reference populations for the Upland herds in this study are of variable quality. Previous results have shown that between 10 and 20 individuals provide reasonable estimates for discriminatory purposes, with populations of fewer than 5 reference individuals giving less robust results. As Table 2.5 indicates, several of the upland populations consist of fewer than 5 results. To maximise the effectiveness of the assay the tests were performed by creating a separate reference population for each individual that contained all other upland animals minus the sample being tested. While this removal and replacement approach is the best option available it does mean that, for the smaller populations in particular, removing an individual can have significant effects on the allele frequencies in that individuals 'reference' population. Hence the absolute allocation of individuals must be taken with caution for the smaller groups.

2.3.2: Results and Discussion

In the first instance an estimate of genetic diversity within the pony populations under test was performed using the ANOVA model of Weir and Cockerham (1984) as implemented in Genepop4.2.

Table 2.4: Estimates of 'Observed (H_O)' and 'Expected (H_E)' heterozygocity in each of the tested pony populations. All populations demonstrated no significant evidence of inbreeding disturbance form HW but there was some indication of small but significant levels of Heterozygote excess in some highlighted populations (prob<0.05)

Population	1-Qintra H_O	1-Qinter H_E	HW U test for H1 = H_O>H_E P-val	S.E.	switches (ave.)
1=Llanrhidian Marsh	-				
2=Dowlais Hill	0.284	0.219	0.0001	0.0000	33332.68
3=Hergest Hill	0.316	0.275	0.0021	0.0001	25376.19
4=Llangoed	0.231	0.256	0.9998	0.0000	38870.24
5=Llandefalle	0.296	0.294	0.3169	0.0020	35310.97
6=Llanafan and Llanwrth	0.263	0.245	0.0299	0.0006	33575.97
7=Begwns	0.265	0.231	0.0007	0.0001	25478.31
8=Brecon Beacons	0.282	0.274	0.2435	0.0019	33136.33
9=Black Mountain	0.245	0.212	0.0000	0.0000	33897.45
10= Preseli	0.306	0.264	0.0073	0.0002	28780.17
11= Llynyfan	0.309	0.282	0.0154	0.0005	31697.77
12=Drum Hill	0.288	0.276	0.1491	0.0014	27683.28
13=Upland Dartmoor	0.357	0.313	0.0053	0.0002	30762.69
14=Section A	0.209	0.205	0.2207	0.0019	27211.37
15=Section D	0.239	0.255	0.9978	0.0001	38536.91
16=Pedigree Dartmoors	0.304	0.320	0.9827	0.0005	32914.98
17=Gypsy Cob	0.288	0.303	0.7989	0.0013	33288.81
18=Irish Draft (Cross)	0.356	0.324	0.0010	0.0001	35194.08
19=Connemara	0.258	0.261	0.8482	0.0015	37520.01
20=polo pony	-				
21=Carneddau	0.179	0.183	0.6959	0.0020	31642.47
22=Thoroughbred crosses	0.388	0.354	0.0028	0.0002	33644.48
23=Warmblood	0.346	0.373	0.9530	0.0007	29706.73

Overall, all populations were not disturbed from HW expectations for these 172 markers. However, a specific test for excess levels of diversity (Table 2.4) did indicate that several populations were significantly though slightly more heterozygous than predicted from allele frequencies present. Although Ho>He can indicate the influence of balancing selection, the most likely reason for such an observation is different allele frequencies in the male and female parents of the tested animals. This is almost certainly the reason in the Irish Draft and ThoroughBred animals where a significant number of those tested were chosen as deliberate

crosses between those breeds and 'pony breeds' to simulate the genotype of animals that might be 'discarded' on the uplands. Similarly in small populations like the Improvement Society herds, stochastic sampling of small parental populations can results in different male and female allele frequencies. However, we believe it is more likely we are seeing here the signature of the management practise of running selected stallions with a group of mares, where the stallion does not necessarily come from the same cohort of animals as the mares with whom he is mating. In effect we are seeing the genetic effect of overlapping generations and/or the transfer of stallions from one hill to another, as indeed was seen in the pedigree results above, as an increase in individual diversity in these populations. The level of genetic diversity reflects this management as it indicates similar levels of heterozygosity within each of the Improvement Society herds to that found in other native breeds. The Carneddau population has been shown elsewhere to have some evidence of long term inbreeding and this is reflected in the slightly lower diversity detected here.

It must be borne in mind that these markers are not a random selection, rather they had been selected as a maximally informative set for distinguishing between the Welsh sections and other Native pony breeds. These may therefore include loci under active selection in one or more populations so should not be used to determine phylogenetic relationships between the animals. However, having said that, Table 2.5 demonstrated the aggregate genetic distance between each of the tested groups expressed in terms of Wrights F_{ST}, which can be equated to the proportion of total genetic variation in a comparison of two populations that lies between those populations. The cells in this table are colour coded from bright green for no genetic difference between the two populations to bright red for all genetic markers being different. The pattern of genetic distance clearly indicates that the upland herds form a 'natural group'. Although they are definitely also close to the section A Welsh ponies. However, there is a significant signature of genetic difference between some individual upland herds.

This pattern is reflected in the probabilities of correct assignment of individual to designated herd illustrated in Table 2.6. If we consider only the Upland herds, the Carneddau and Section A ponies then 56/88 animals were correctly assigned to the population of origin with a probability of 99% or greater. Of the remaining 31, all were assigned to another Upland Herd, sometimes more than a single herd with intermediate probability, a pattern of assignment often seen in animals that result from the hybridization between populations. Two of the Upland animals were incorrectly diagnosed as being drawn from the Section A, while one of the pedigree Section A ponies came out as being related to the Brecon Beacons and

another was most likely to be Section A but had 40% chance of assignment to the Llangoed herd. None of the other ponies tested were assigned to the Upland or Section A animals (though one pedigree Dartmoor and 2 upland Dartmoors did have marginal assignments (none higher than 2%) to Upland Welsh herds, which may reflect the use of Welsh pony stallions in the post War recovery of the Dartmoor breed (Edwards, 1992). None of the Upland hill animals were characterised with any of the other mixed pony groups.

Table 2.5: Estimates of F_{ST} for all loci (diploid) based of 172 SNPs per animal, for 23 groups: Populations represented are (number of animals); 1=Llanrhidian Marsh (1), 2=Dowlais Hill (2), 3=Hergest Hill (3), 4=Llangoed (15), 5=Llandefalle (5), 6=Llanafan and Llanwrth (8), 7=Begwns (4), 8=Brecon Beacons (6), 9=Black Mountain (8), 10= Preseli (3), 11= Llynyfan (6), 12=Drum Hill (5), 13=Upland Dartmoor (6), 14=Section A (12), 15=Section D (12), 16=Pedigree Dartmoors (9), 17=Gypsy Cob (2), 18=Irish Draft (10), 19=Connemara (12), 20=polo pony (1), 21=Carneddau (12), 22=Thoroughbred crosses (7), 23=Warmblood (4).

pop	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	0.212																					
3	0.144	0.261																				
4	0.152	0.163	0.041																			
5	0.063	0.158	0.044	0.058																		
6	0.275	0.289	0.079	0.154	0.139																	
7	0.254	0.271	0.034	0.052	0.062	0.102																
8	0.194	0.223	0.009	0.057	0.056	0.077	0.046															
9	0.227	0.263	0.138	0.114	0.086	0.221	0.124	0.14														
10	0.181	0.222	0.062	-0.006	0.058	0.163	0.095	0.072	0.162													
11	0.139	0.18	0.032	0.04	0.035	0.115	0.05	0.038	0.076	0.043												
12	0.128	0.204	0.007	0.04	0.003	0.088	0.018	0.013	0.056	0.022	0.03											
13	0.289	0.33	0.184	0.237	0.198	0.167	0.245	0.158	0.329	0.216	0.191	0.165										
14	0.148	0.306	0.083	0.106	0.106	0.224	0.109	0.138	0.082	0.106	0.104	0.07	0.354									
15	0.503	0.486	0.429	0.449	0.419	0.42	0.477	0.402	0.517	0.446	0.423	0.433	0.327	0.552								
16	0.148	0.143	0.165	0.227	0.177	0.175	0.228	0.156	0.284	0.23	0.162	0.191	0.134	0.326	0.342							
17	0.256	0.318	0.183	0.235	0.158	0.19	0.259	0.14	0.341	0.28	0.162	0.191	0.126	0.382	0.31	0.098						
18	0.291	0.291	0.211	0.278	0.235	0.194	0.257	0.178	0.321	0.287	0.211	0.234	0.136	0.359	0.319	0.116	0.044					
19	0.446	0.426	0.38	0.418	0.377	0.348	0.406	0.345	0.459	0.45	0.353	0.391	0.29	0.492	0.461	0.254	0.211	0.18				
20	0.66	0.438	0.297	0.301	0.256	0.258	0.37	0.241	0.417	0.393	0.248	0.286	0.215	0.44	0.416	0.143	0.102	0.104	0.183			
21	0.5	0.426	0.244	0.262	0.269	0.212	0.291	0.158	0.358	0.319	0.252	0.231	0.312	0.352	0.529	0.218	0.321	0.26	0.427	0.423		
22	0.247	0.253	0.195	0.254	0.204	0.18	0.237	0.17	0.312	0.25	0.192	0.21	0.141	0.359	0.272	0.105	0.017	0.036	0.146	0.048	0.28	
23	0.204	0.248	0.198	0.278	0.212	0.194	0.251	0.162	0.348	0.269	0.213	0.215	0.084	0.389	0.262	0.084	-0.009	0.023	0.167	0.018	0.3	0.003

Table 2.6: Summary of the results of GeneClass 2 assignment of population to individual samples. The 'most likely population' is indicated. The 'expected' result of correct assignment to population of origin is shown by shaded boxes– see text for explanation.

Predicted most probable population

Origin of sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1=Llanrhidian Marsh (1)					1										
2=Dowlais Hill (2)		2													
3=Hergest Hill (3)							1	1					1		
4=Llangoed (15)				8			1	1	1	1	1		2		
5=Llandefalle (5)					4				1			1			
6=Llanafan and Llanwrth (8)						7		1							
7=Begwns (4)				1		1	2								
8=Brecon Beacons (6)				1				2			2	1			
9=Black Mountain (8)									7				1		
10= Preseli (3)				3											
11= Llynyfan (6)				1					1		4				
12=Drum Hill (5)								2			1	2			
13=Section A (12)				1				1					10		
14=Carneddau (12)				1				1						10	
15=Section D (12)															12

Section 2.4 The maternal story: -Mitochondrial DNA.

2.4.1: Method

The mitochondria are intracellular organelles that are found in the cytoplasm of the cells where they produce energy and contain their own approximately 16kb circular DNA molecule. During the formation of an embryo, the male parent's mitochondrial DNA is excluded from the fertilised egg so that mitochondrial DNA is inherited from mother to child only. Hence the mitochondrial pattern of inheritance is a marker for maternal history that is independent of the autosomal results.
We amplified the so called D-loop region of the ponies mitochondrial DNA using previously described primers under standard PCR conditions.

Primers used to sequence mt DNA
Forward - 5′-ATT TCT TCC CCT AAA CGA CAA C-3′
Reverse - 5′-CGT TCA ATT TAA GTC CAG CTT C-3′

The resulting PCR products were column purified and single direction sequences were obtained using the forward primer, to give a 540 base pair product. These sequences were aligned to previously obtained reference samples from various sources using *kalign* and checked in *bioedit* manually to ensure consistency of alignment.

Alignments were imported into *Mega4* and Phylogenetic trees constructed to visualise the relationships between individuals (Figure 2.7 below).

A fresh alignment was then prepared using a 220 base pair section of these samples and previously published sequences from European and American populations of several pony breeds and archaeological and reference type sequences to characterise the phylogenetic relationships between the sequences (Figure 2.6).

24.2: Results and Discussion

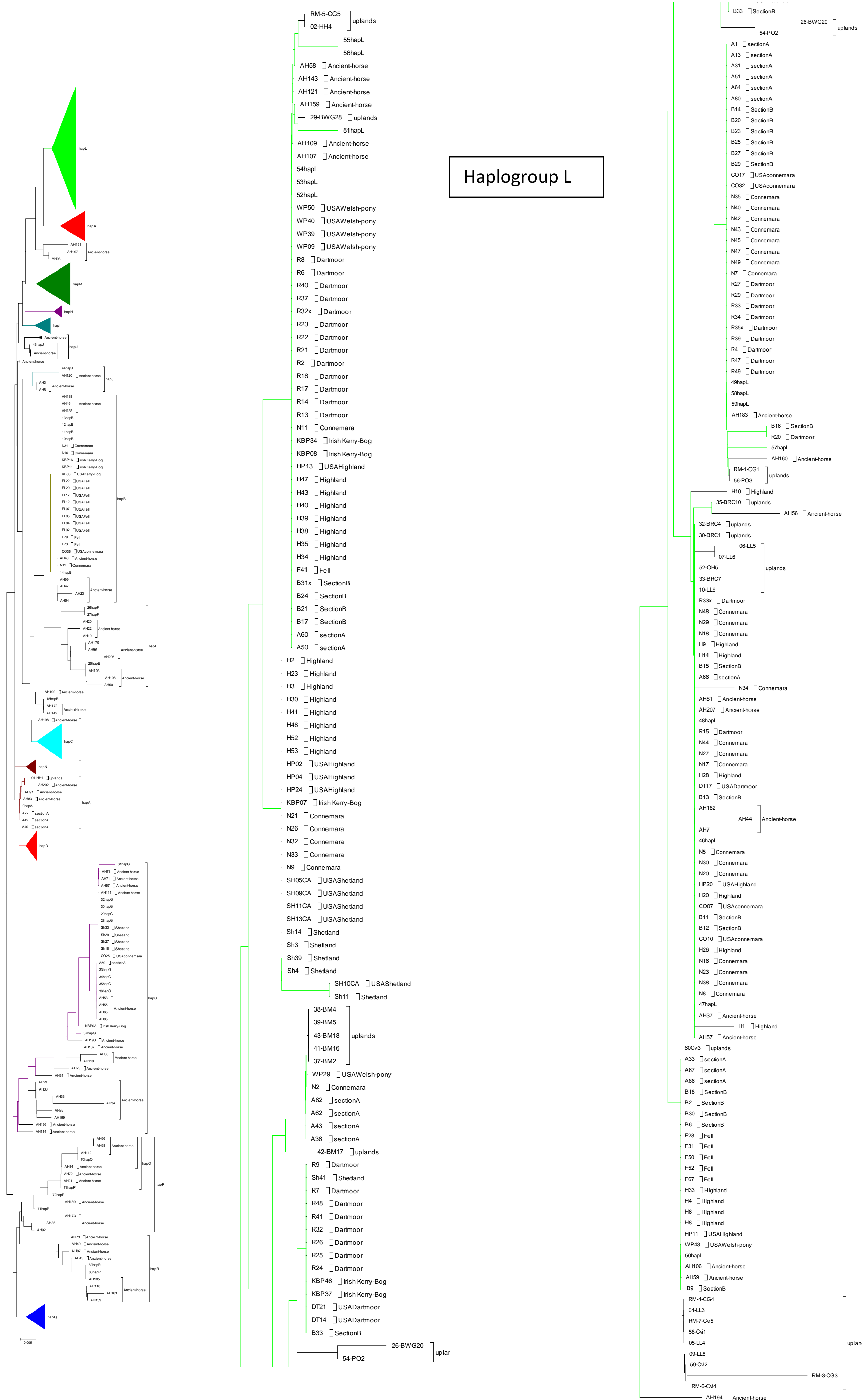
In general there is no clear distinction between the maternal haplotypes present in the Section A-D Welsh animals and those in the upland herds, when the data from all the herds are combined as inTable 2.7 below. However, there are distinctions between the haplotype

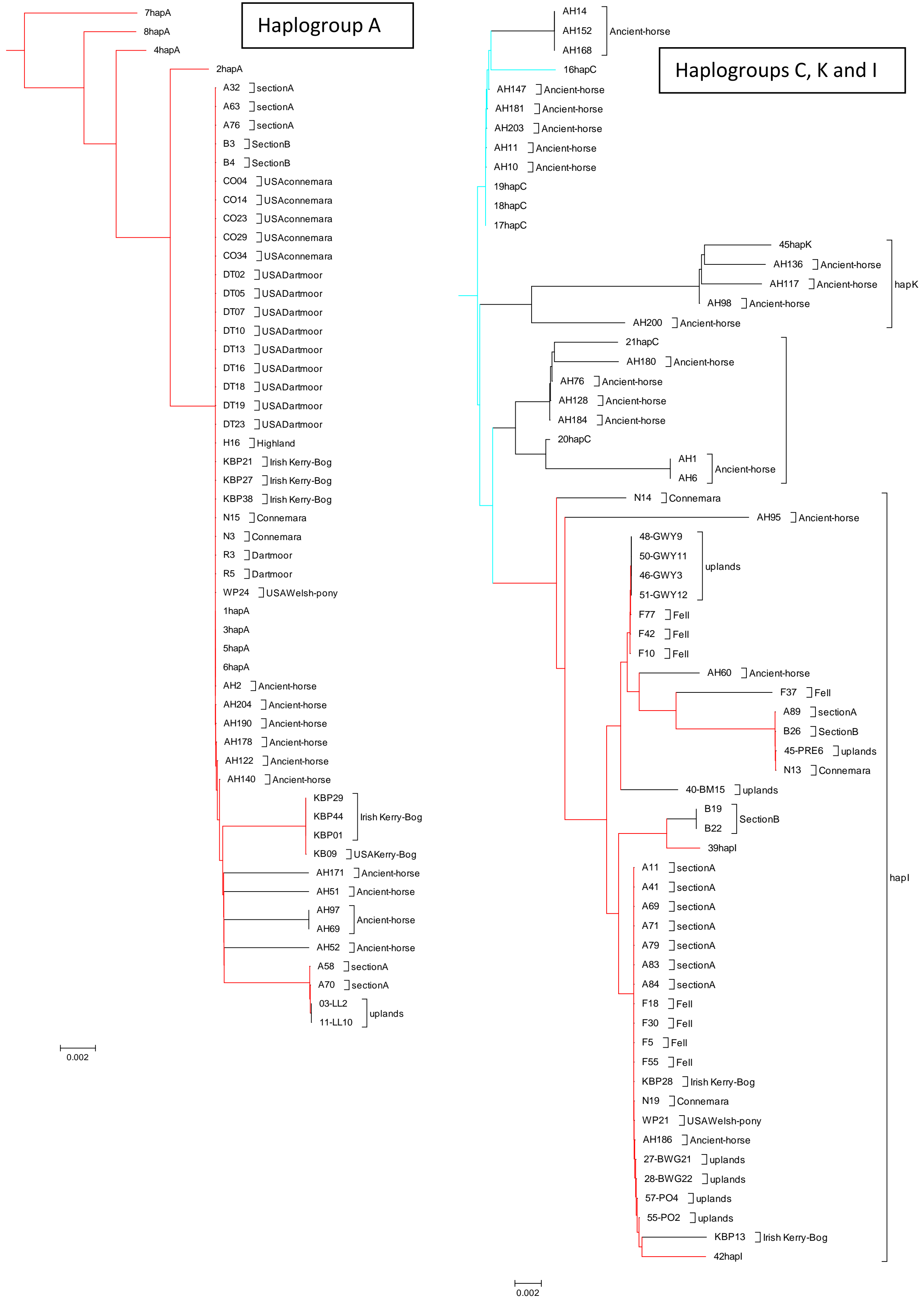
distributions between individual upland herds and between individual upland herds and the pedigree Section animals (see Figures 2.6 and 2.7). We would take this to be consistent with the know history of the pedigree animals which are believed have been founded from the upland groups by selection from within a mixed blood stock based on Welsh Mountain ponies and by crossing mares from these mixed stocks to other breeds such as Trotters and Arabs to create Section B-D Welsh ponies.

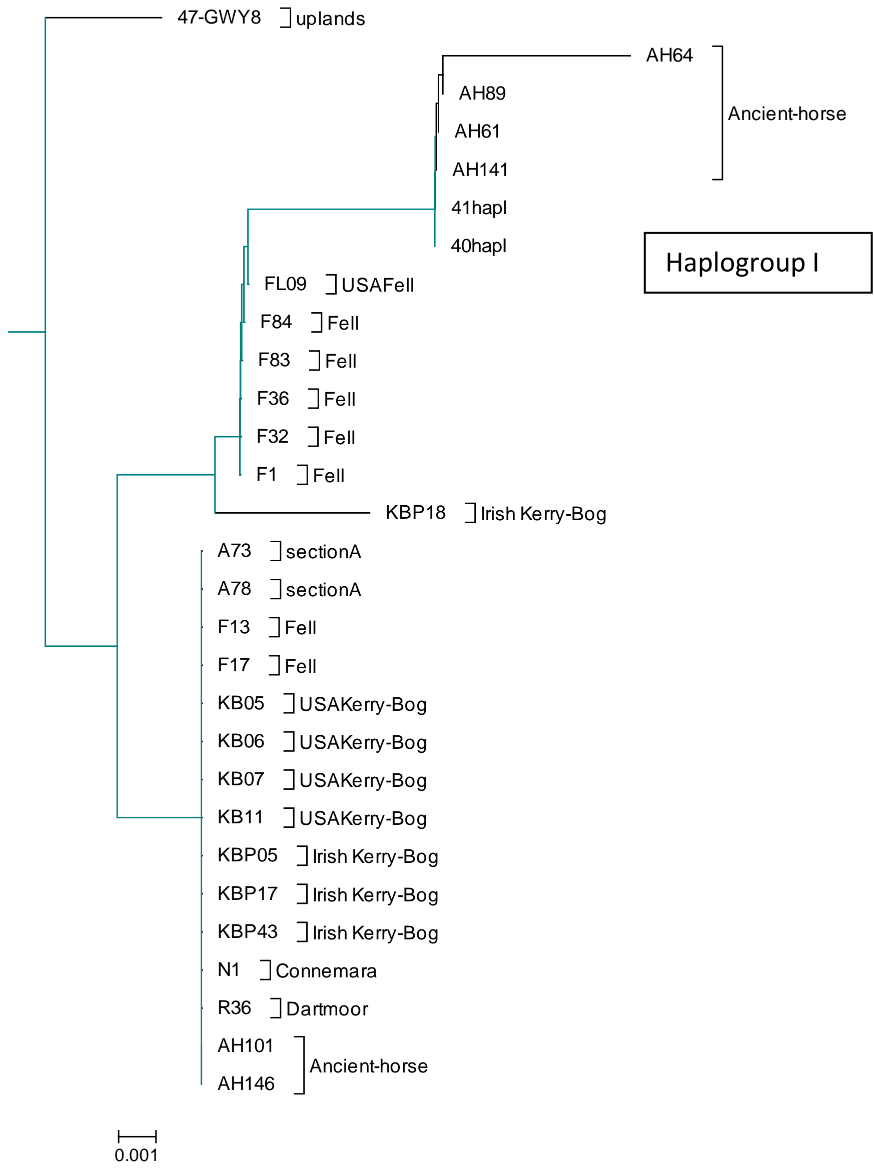
Figure 2.6. Phylogenetics relationships of 759 taxa

The full tree is shown in the first panel of the figure with major haplogroups collapsed into solid coloured triangles. Subsequent panels/pages show the subtrees representing the unique haplogroup branches characterised by Achili et al 2012 using the control region sequences and a Network analysis. Only Haplogroups containing Upland animals are expanded. A phylogenetic tree is used here to represent the results simply to allow the visualisation of individual animals and their relationships to each haplogroup, hence some haplogroups are 'split' across different tree branches as a consequence of the presence of recurrent mutations in the horse mitochondrial phylogeny.

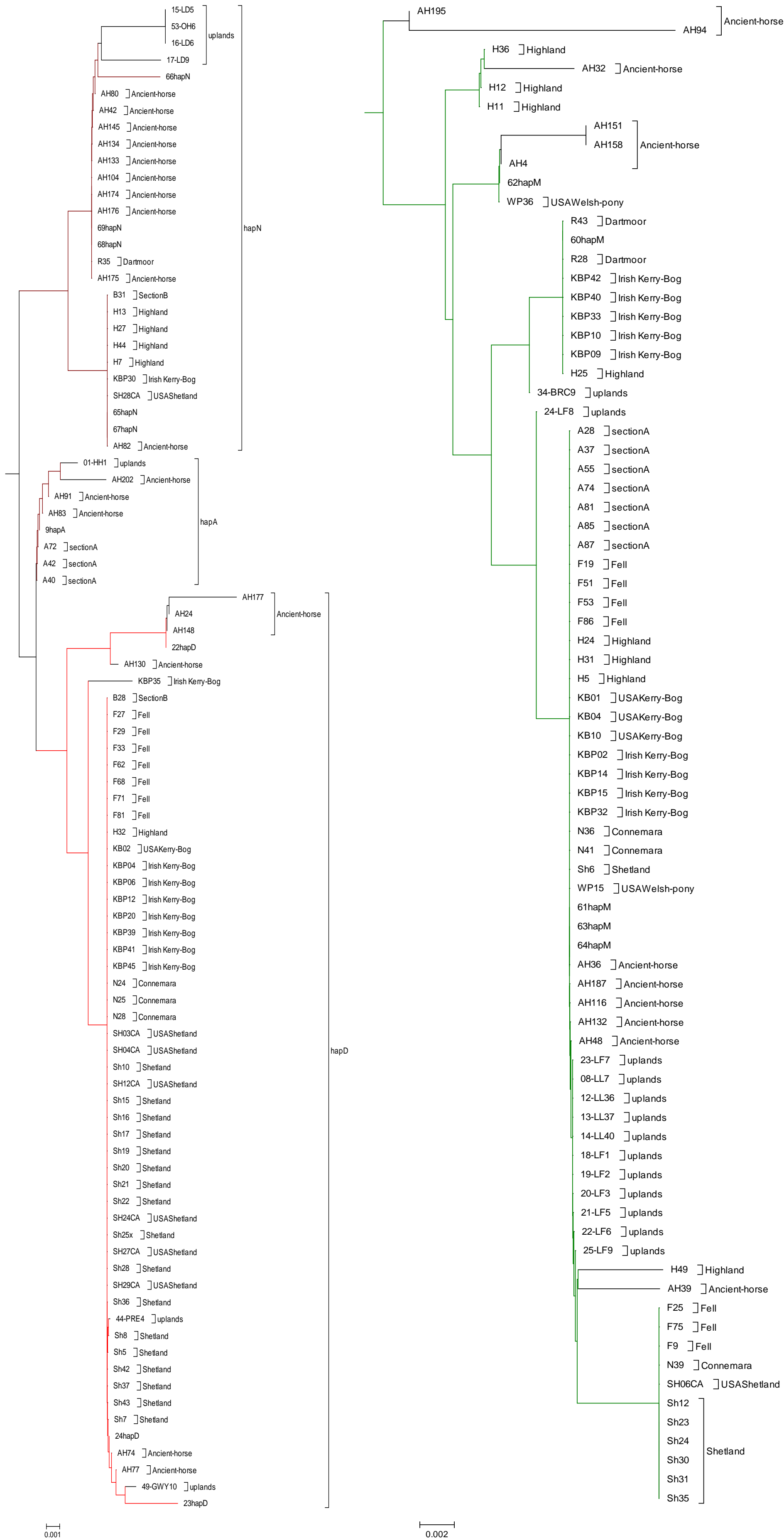
The evolutionary history was inferred using the Neighbor-Joining method [Saitou and Nei 1987]. The optimal tree with the sum of branch length = 0.897 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [Tamura, Nei and Kumaar 2004] and are in the units of the number of base substitutions per site. The rate variation among sites was modelled with a gamma distribution (shape parameter = 1). Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 240 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [Tamura, Dudley, Nei and Kumar 2007]. Data for mitochondria not sequenced in this project include sequences downloaded from NCBI and from our previous work, Winton 2013, Winton et al 2013, 2015.


Haplogroup L
hapL
hapA
hapM
hapH
hapI
hapJ
hapB
hapF
hapC
hapN
hapD
hapG
hapO
hapP
hapR
hapQ
0.005
RM-5-CG5
02-HH4
uplands
55hapL
56hapL
AH58 Ancient-horse
AH143 Ancient-horse
AH121 Ancient-horse
AH159 Ancient-horse
29-BWG28 uplands
51hapL
AH109 Ancient-horse
AH107 Ancient-horse
54hapL
53hapL
52hapL
WP50 USAWelsh-pony
WP40 USAWelsh-pony
WP39 USAWelsh-pony
WP09 USAWelsh-pony
R8 Dartmoor
R6 Dartmoor
R40 Dartmoor
R37 Dartmoor
R32x Dartmoor
R23 Dartmoor
R22 Dartmoor
R21 Dartmoor
R2 Dartmoor
R18 Dartmoor
R17 Dartmoor
R14 Dartmoor
R13 Dartmoor
N11 Connemara
KBP34 Irish Kerry-Bog
KBP08 Irish Kerry-Bog
HP13 USAHighland
H47 Highland
H43 Highland
H40 Highland
H39 Highland
H38 Highland
H35 Highland
H34 Highland
F41 Fell
B31x SectionB
B24 SectionB
B21 SectionB
B17 SectionB
A60 sectionA
A50 sectionA
H2 Highland
H23 Highland
H3 Highland
H30 Highland
H41 Highland
H48 Highland
H52 Highland
H53 Highland
HP02 USAHighland
HP04 USAHighland
HP24 USAHighland
KBP07 Irish Kerry-Bog
N21 Connemara
N26 Connemara
N32 Connemara
N33 Connemara
N9 Connemara
SH05CA USAShetland
SH09CA USAShetland
SH11CA USAShetland
SH13CA USAShetland
Sh14 Shetland
Sh3 Shetland
Sh39 Shetland
Sh4 Shetland
SH10CA USAShetland
Sh11 Shetland
38-BM4
39-BM5
43-BM18
41-BM16
37-BM2
uplands
WP29 USAWelsh-pony
N2 Connemara
A82 sectionA
A62 sectionA
A43 sectionA
A36 sectionA
42-BM17 uplands
R9 Dartmoor
Sh41 Shetland
R7 Dartmoor
R48 Dartmoor
R41 Dartmoor
R32 Dartmoor
R26 Dartmoor
R25 Dartmoor
R24 Dartmoor
KBP46 Irish Kerry-Bog
KBP37 Irish Kerry-Bog
DT21 USADartmoor
DT14 USADartmoor
B33 SectionB
26-BWG20
54-PO2
uplar
B33 SectionB
26-BWG20
54-PO2
uplands
A1 sectionA
A13 sectionA
A31 sectionA
A51 sectionA
A64 sectionA
A80 sectionA
B14 SectionB
B20 SectionB
B23 SectionB
B25 SectionB
B27 SectionB
B29 SectionB
CO17 USAconnemara
CO32 USAconnemara
N35 Connemara
N40 Connemara
N42 Connemara
N43 Connemara
N45 Connemara
N47 Connemara
N49 Connemara
N7 Connemara
R27 Dartmoor
R29 Dartmoor
R33 Dartmoor
R34 Dartmoor
R35x Dartmoor
R39 Dartmoor
R4 Dartmoor
R47 Dartmoor
R49 Dartmoor
49hapL
58hapL
59hapL
AH183 Ancient-horse
B16 SectionB
R20 Dartmoor
57hapL
AH160 Ancient-horse
RM-1-CG1
56-PO3
uplands
H10 Highland
35-BRC10 uplands
AH56 Ancient-horse
32-BRC4 uplands
30-BRC1 uplands
06-LL5
07-LL6
52-OH5
33-BRC7
10-LL9
uplands
R33x Dartmoor
N48 Connemara
N29 Connemara
N18 Connemara
H9 Highland
H14 Highland
B15 SectionB
A66 sectionA
N34 Connemara
AH81 Ancient-horse
AH207 Ancient-horse
48hapL
R15 Dartmoor
N44 Connemara
N27 Connemara
N17 Connemara
H28 Highland
DT17 USADartmoor
B13 SectionB
AH182
AH44
AH7
Ancient-horse
46hapL
N5 Connemara
N30 Connemara
N20 Connemara
HP20 USAHighland
H20 Highland
CO07 USAconnemara
B11 SectionB
B12 SectionB
CO10 USAconnemara
H26 Highland
N16 Connemara
N23 Connemara
N38 Connemara
N8 Connemara
47hapL
AH37 Ancient-horse
H1 Highland
AH57 Ancient-horse
60Cvl3 uplands
A33 sectionA
A67 sectionA
A86 sectionA
B18 SectionB
B2 SectionB
B30 SectionB
B6 SectionB
F28 Fell
F31 Fell
F50 Fell
F52 Fell
F67 Fell
H33 Highland
H4 Highland
H6 Highland
H8 Highland
HP11 USAHighland
WP43 USAWelsh-pony
50hapL
AH106 Ancient-horse
AH59 Ancient-horse
B9 SectionB
RM-4-CG4
04-LL3
RM-7-Cvl5
58-Cvl1
05-LL4
09-LL8
59-Cvl2
RM-3-CG3
RM-6-Cvl4
uplan
AH194 Ancient-horse


Haplogroup A
7hapA
8hapA
4hapA
2hapA
A32 sectionA
A63 sectionA
A76 sectionA
B3 SectionB
B4 SectionB
CO04 USAconnemara
CO14 USAconnemara
CO23 USAconnemara
CO29 USAconnemara
CO34 USAconnemara
DT02 USADartmoor
DT05 USADartmoor
DT07 USADartmoor
DT10 USADartmoor
DT13 USADartmoor
DT16 USADartmoor
DT18 USADartmoor
DT19 USADartmoor
DT23 USADartmoor
H16 Highland
KBP21 Irish Kerry-Bog
KBP27 Irish Kerry-Bog
KBP38 Irish Kerry-Bog
N15 Connemara
N3 Connemara
R3 Dartmoor
R5 Dartmoor
WP24 USAWelsh-pony
1hapA
3hapA
5hapA
6hapA
AH2 Ancient-horse
AH204 Ancient-horse
AH190 Ancient-horse
AH178 Ancient-horse
AH122 Ancient-horse
AH140 Ancient-horse
KBP29
KBP44
KBP01
Irish Kerry-Bog
KB09 USAKerry-Bog
AH171 Ancient-horse
AH51 Ancient-horse
AH97
AH69
Ancient-horse
AH52 Ancient-horse
A58 sectionA
A70 sectionA
03-LL2
11-LL10
uplands
0.002
Haplogroups C, K and I
AH14
AH152
AH168
Ancient-horse
16hapC
AH147 Ancient-horse
AH181 Ancient-horse
AH203 Ancient-horse
AH11 Ancient-horse
AH10 Ancient-horse
19hapC
18hapC
17hapC
45hapK
AH136 Ancient-horse
AH117 Ancient-horse
AH98 Ancient-horse
AH200 Ancient-horse
hapK
21hapC
AH180 Ancient-horse
AH76 Ancient-horse
AH128 Ancient-horse
AH184 Ancient-horse
20hapC
AH1
AH6
Ancient-horse
N14 Connemara
AH95 Ancient-horse
48-GWY9
50-GWY11
46-GWY3
51-GWY12
uplands
F77 Fell
F42 Fell
F10 Fell
AH60 Ancient-horse
F37 Fell
A89 sectionA
B26 SectionB
45-PRE6 uplands
N13 Connemara
40-BM15 uplands
B19
B22
SectionB
39hapI
A11 sectionA
A41 sectionA
A69 sectionA
A71 sectionA
A79 sectionA
A83 sectionA
A84 sectionA
F18 Fell
F30 Fell
F5 Fell
F55 Fell
KBP28 Irish Kerry-Bog
N19 Connemara
WP21 USAWelsh-pony
AH186 Ancient-horse
27-BWG21 uplands
28-BWG22 uplands
57-PO4 uplands
55-PO2 uplands
KBP13 Irish Kerry-Bog
42hapI
hapI
0.002


47-GWY8
uplands
AH64
AH89
AH61
AH141
Ancient-horse
41hapl
40hapl
Haplogroup I
FL09
USAFell
F84
Fell
F83
Fell
F36
Fell
F32
Fell
F1
Fell
KBP18
Irish Kerry-Bog
A73
sectionA
A78
sectionA
F13
Fell
F17
Fell
KB05
USAKerry-Bog
KB06
USAKerry-Bog
KB07
USAKerry-Bog
KB11
USAKerry-Bog
KBP05
Irish Kerry-Bog
KBP17
Irish Kerry-Bog
KBP43
Irish Kerry-Bog
N1
Connemara
R36
Dartmoor
AH101
AH146
Ancient-horse
0.001

Haplogroup N, O and Haplotype M



Haplogroup Q
77hapQ
AH88 Ancient-horse
76hapQ
AH63 Ancient-horse
78hapQ
81hapQ
AH100 Ancient-horse
80hapQ
RM-2-CG2 uplands
AH27 Ancient-horse
CO35 USAconnemara
B7 SectionB
AH90 Ancient-horse
79hapQ
75hapQ
AH41 Ancient-horse
AH26
AH75
Ancient-horse
A27 sectionA
F21 Fell
F22 Fell
F4 Fell
F44 Fell
F69 Fell
HP10 USAHighland
HP16 USAHighland
HP18 USAHighland
HP22 USAHighland
KBP22 Irish Kerry-Bog
N37 Connemara
SH07CA USAShetland
SH08CA USAShetland
SH23CA USAShetland
Sh25 Shetland
SH25CA USAShetland
Sh26 Shetland
Sh32 Shetland
Sh34 Shetland
Sh40 Shetland
AH144 Ancient-horse
AH43 Ancient-horse
AH115 Ancient-horse
31-BRC3 uplands
F40
F64
Fell
74hapQ
AH70 Ancient-horse
0.001

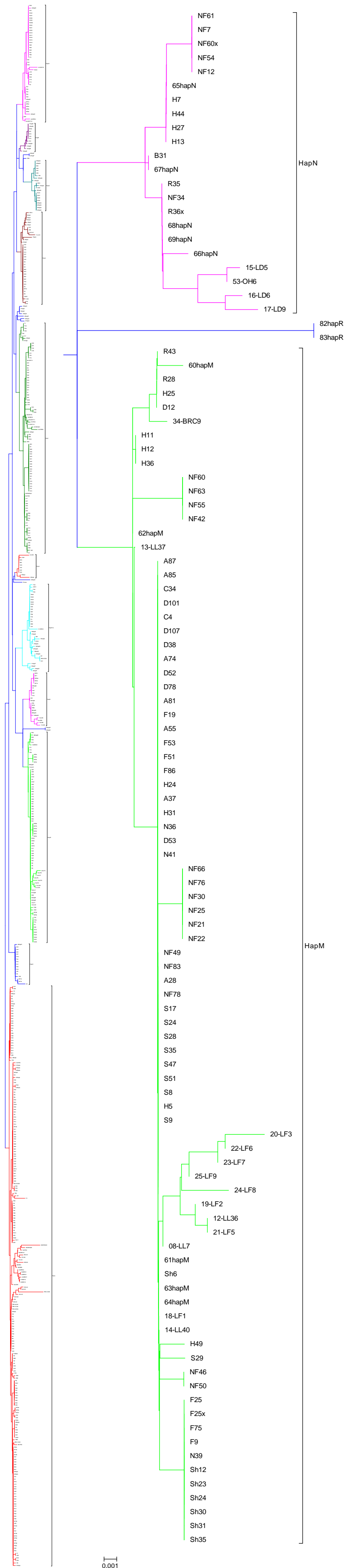


Figure2.7 Phylogenetic relationships of 634 taxa

The evolutionary history was inferred using the Minimum Evolution method [Rzhetsky and Nei, 1992]. Tree #1 out of 100 minimum evolution trees (sum of branch length = 0.554) is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method [Tamura et al, 2004] and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm [Nei and Kumar, 2000] at a search level of 1. The Neighbor-joining algorithm [Saitou and Nei, 1987] was used to generate the initial tree. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 534 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 [Tamura et al 2007].

Using the complete alignment it is possible to see further structure amongst the upland mitochondrial. It must be borne in mind when examining these results that the extended product included several 'hypervariable sites' where individual genotypes are known to be difficult to determine from unidirectional sequencing. Since this was an initial pilot investigation we did not perform bidirectional sequencing due to the additional costs involved. Hence some of the sequence variants within clusters may represent artifacts of the unidirectional sequencing and hence these results should not be used to date the clusters at this point.

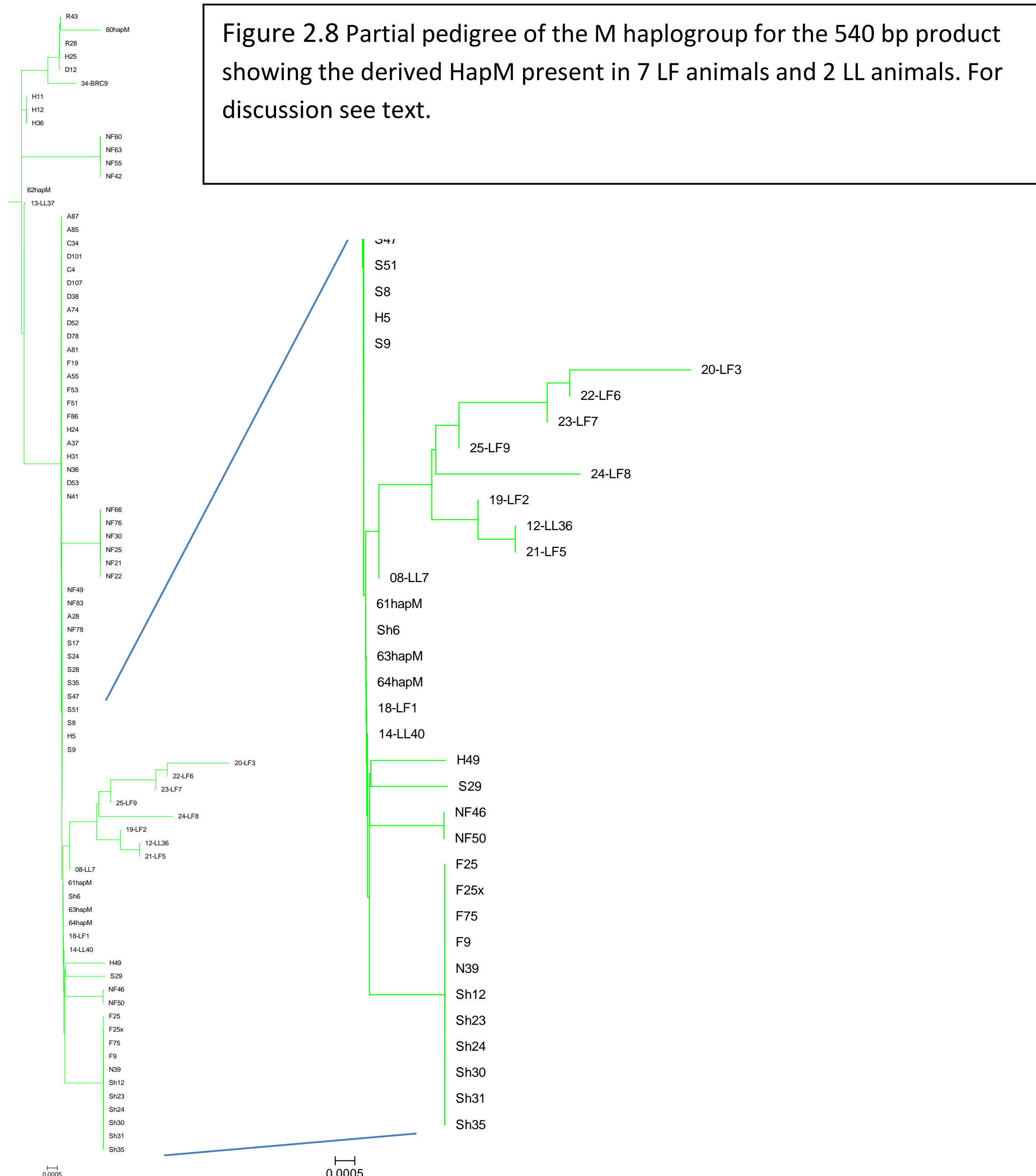


Figure 2.8 Partial pedigree of the M haplogroup for the 540 bp product showing the derived HapM present in 7 LF animals and 2 LL animals. For discussion see text.

What is apparent is that there are certainly groups of haplotypes that are population specific. Taking one example within the M haplogroup shown in Figure 2.8, here there is a small

cluster of haplotypes associated almost exclusively with animals from the Llanafan and Llanwrthwl herd (LF).

These animals share a common haplotype that distinguishes them from other M haplotype maternal lines. Examination of pedigree records indicate that these individuals do share a common maternal lineage centred on the mare Trawsnant Primrose born in 1950. We can assume that the two LL animals and similar clusters in other mitochondrial haplogroups also lead back to common individuals in the recent past, though most have 'unknown' mare at the head of their pedigrees. Since these genetic clusters link animals in different herds, we can assume the 'common ancestor' predates the isolation of those upland groups in some cases.

Figure 2.9 showing pedigree for descendants of the mare Trawsnant Primrose. individuals (LF) from the M-haplogroup illustrated above are indicated with green arrows. Blue lines indicate materal lines of inheritance for mitochondria, red paternal lines of inheritance. Only descendants of the focus mare are shown.

The pedigree shown in Figure 2.9 illustrates that in addition to common inheritance of several individuals in the Llanafan herd through the maternal line this particular mare has had significant influence on the maternal lines within the Hardre blood line and has also influenced the autosomal makeup of the Begwns herd through the stallion Hardre Testun. Hence again we have evidence of both the localised specialisation of each of the upland herds and their genetic interconnectedness.

Exploring the relationships between the herds we can approach this by initially looking at the genetic distance between the herds based on the number and distribution of sequence variants

present in each herd. These results are presented in Table 2.7 where it assumed that all of the Welsh ponies may have originated from a common source and therefore the ‘genetic distance’, in this case estimated by the average number of base substitutions per site, is related to the change in mitochondrial haplogroup distribution between groups resulting from genetic drift, and the creation of novel genetic variants by mutation. The general pattern is very similar, although the Uplands taken as a composite group show slightly more genetic distance, although the same patterns of relative distance, as expressed between the other populations. We would interpret this to suggest that the maternal ancestors of the Upland herds do not show a particularly closer relationship to the pedigree Section-A animals than they do to the other ‘Welsh’ populations. This suggests that, at least as far as the maternal component is concerned, the Upland haplotypes are as isolated from the Pedigree section-As as they are from the Carneddau and Section C animals.

Table 2.7 Estimates of Divergence over Sequence Pairs between Groups.

[1] #uplands

[2] #section_A
[3] #sectionB
[4] #sectionC
[5] #section_D
[6] #Carneddau

[	1	2	3	4	5]
[1]					
[2]	0.0226				
[3]	0.0210	0.0170			
[4]	0.0237	0.0187	0.0178		
[5]	0.0232	0.0186	0.0177	0.0192	
[6]	0.0253	0.0197	0.0212	0.0198	0.0205

The number of base substitutions per site from averaging over all sequence pairs between groups is shown . All results are based on the pairwise analysis of 279 sequences. Analyses were conducted using the Maximum Composite Likelihood method in MEGA4 [Tamura et al, 2004]. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (Pairwise deletion option). There were a total of 534 positions in the final dataset.

Taking this a step further we can examine the distribution of haplogroups across the different Upland herds. These results are shown relative to the Native pony and WorldWide data in Table 2.8 and Figure 2.10. These data clearly show that individual Upland groups appear to have distinct distributions of maternal contribution. In general there is no clear relationship between the genetic distance between herds estimated from the autosomal data in previous sections and the ‘maternal distance’ represented by these mitochondrial results. This suggests strongly that the continuity of the herds is being maintained via transfer of stallions, and not mares between populations.

Table 2.8 Haplogroup frequency (%) according to global region of horse breeds, and numbers sequenced in the upland herds

Equine mitochondrial haplogroups		A	B	C	D	E	F	G	H	I	J-K	L	M	N	O'P	Q	R	Sample nos
European		4.49	9.38	0.32	4.57	0.48	-	8.73	1.04	8.01	0.56	38.06	7.29	8.49	1.36	3.85	2.24	1249
Middle Eastern		7.81	10.94	3.13	2.08	0.52	-	9.09	3.03	15.15	-	24.24	3.03	3.03	9.09	9.09	-	192
Asian		11.93	1.70	3.58	2.90	2.21	3.07	16.35	1.36	6.13	6.47	13.46	4.09	2.90	6.64	13.80	1.87	587
Ancient European		11.43	-	-	1.43	-	-	10.00	2.86	8.57	2.86	21.43	17.14	10.00	2.86	4.29	7.14	70
Ancient Asian		8.82	-	-	10.29	-	1.47	22.06	5.88	2.94	2.94	10.29	2.94	4.41	10.29	10.29	7.35	68
Uplands total numbers per herd		4.41	-	-	4.41	-	-	-	-	17.65	-	45.59	19.12	5.88	-	2.94	-	68
Llangoed LL		2										6	4					12
Hergest Hill HH		1								1		1						3
Llandefalle LD														3				3
Llanafan and Llanwrthwl LF													8					8
Begwns BWG										2		2						4
Brecon Beacons BRC												1	1			1		3
Black Mountain BM										1		6						7
Preseli PRE					1					1								2
Drum Hill DH												1		1				2
LLynyfan GWY					2					5								7
PO										2		2						4
CG												7				1		8
CUi												5						5

Haplogroup frequency (%) according to geographical region. The first row identifiers “A” to “R” represent the major worldwide mtDNA haplogroups in horses.
Haplogroup classification is based upon control region motifs, as described by Achilli *et al*. (2012) and as shown in the tree diagrams above.
Geographic regions are designated according to sequence data collated by Achilli *et al*. (2012) and in total comprises of 2166 reference sequences, including ancient (fossil) DNA data.

Percentage of haplogroups in the upland animals are shown averaged across herds (%) and absolute numbers by herds below.

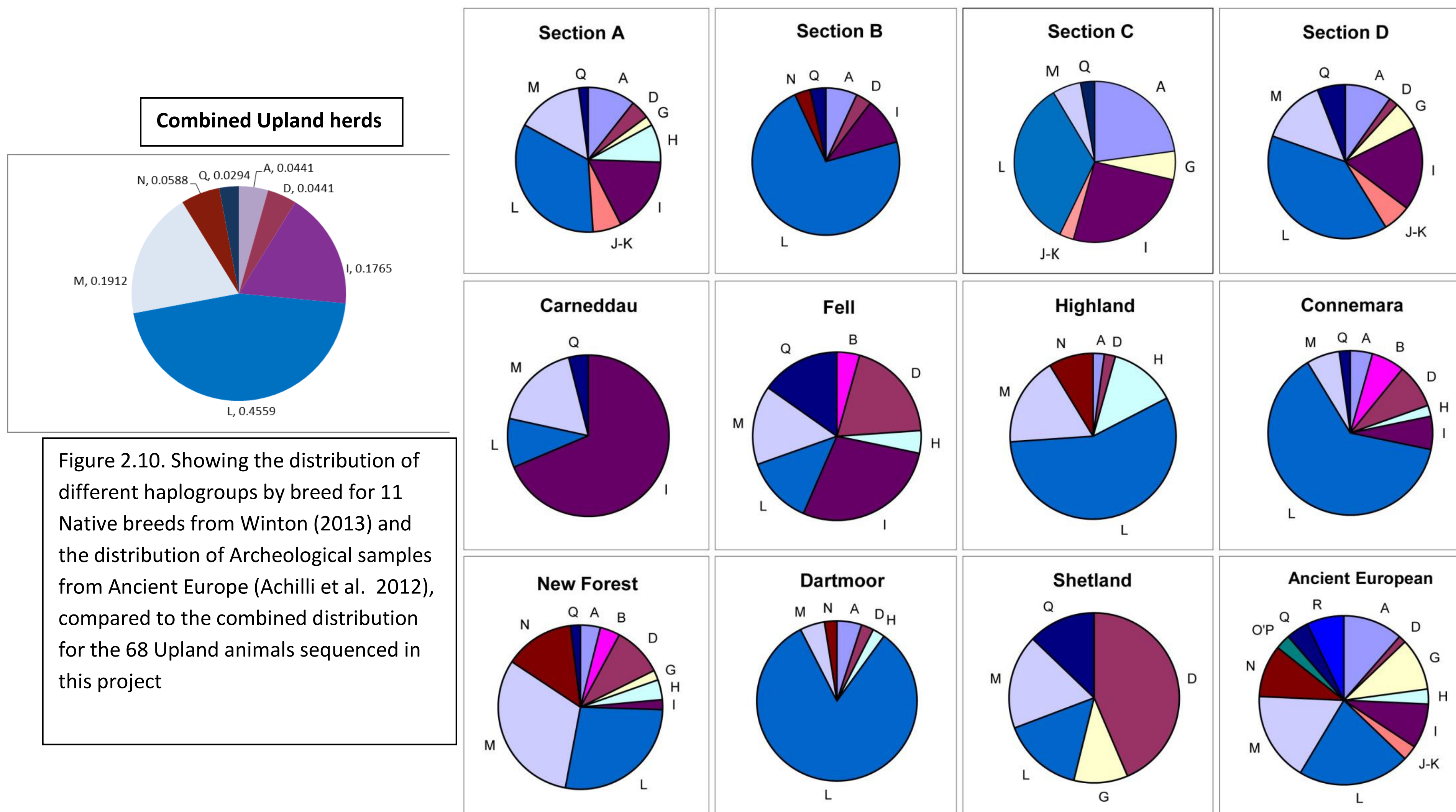


Figure 2.10. Showing the distribution of different haplogroups by breed for 11 Native breeds from Winton (2013) and the distribution of Archeological samples from Ancient Europe (Achilli et al. 2012), compared to the combined distribution for the 68 Upland animals sequenced in this project

Section 3.0: General Discussion and Conclusion

Taken as a whole the results above suggest that the 'genetic history' of the Welsh pony may be summarised by assuming an ancestral group of animals scattered across the Uplands and Lowlands of Wales. These populations may have been subject to local selection pressures by human or environmental influence for many years prior to the establishment of the stud books. As a consequence maternal genetic diversity has been differentially packaged into the resultant population, with subsequent mutation and drift further distancing these groups. There is no evidence from these data that the maternal blood-lines of the Upland animals are any closer to the pedigree Section-A animals than predicted in this simple picture of independent origin of the Welsh Sections, The Upland populations and the Carneddau population. In other words there is no evidence of regular pedigree section-A maternal introgression into the Upland herds after the initial split.
On the other hand molecular distance between autosomal data at the individual and population levels shows as much diversity within the individual herds as between them and the pedigree Section-A animals. With the Section-As being closer to the the Uplands than either are to the Section-Ds. Indeed the degree of genetic separation between the Upland herds and the pedigree animals is in the order of F_{ST} = 10% or roughly half that between the Ds and As. However, this level of 10% is comparable to that observed between other horse breeds for example in France (Leroy et al, 2009) and Poland (Stachurska et al, 2014).
These results could be taken to confirm the belief set out in the RBST letter to the Society of 29th January 2013 that 'the semi-feral bloodlines are well represented within the rest of the Section A stud book'. However, to take that to an absurd extreme, the same argument could be said to justify the non-preservation of wild Wolf populations on the basis that the genes of the Wolf are well represented in modern day dog breeds. The clear differentiation between the genetics of the individual upland herds suggest historical or Natural Selection based differences between these individual populations. While on the one hand history may not be of particular interest, an understanding of the genetic factors underpinning survival and thriftiness in marginal environments may be of profound importance to animal welfare and production efficiency in the Uplands. In this respect it should be noted that the Upland herds appear closer genetically to the small upland Dartmoor and pedigree Dartmoor samples than to the Section-D Welsh on basis of the 170 selected SNP markers illustrated in Table 2.5. These SNPs had been chosen to be diagnostic for the differences between Section-A,

Carneddau and Section-D ponies, and we cannot rule out that this group of selected SNPs may include markers for loci affected by Natural or human selection. Management practises of the Upland farmers have restricted the degree of predicted inbreeding to reasonable levels within herds but have retained a degree of inter-herd variation. Until this has been investigated further, we cannot exclude the possibility that this interherd variation has been assisted by local selection pressures and that loss of all the individuals from a particular herds will result in our losing the specific mixture of genetics that have contributed to the survival of those animals under semi-feral maintenance regimes. The genes involved may well exist in their lowland pedigree relatives, but we run the risk of being unable to identify which particular QTLs are involved in those survival characteristics and that might be usefully exploited to understand and manipulate upland biology.

Acknowledgements

We would like to thank the WPCS who funded this research through a grant from the Brecon Beacons National Park. The SNP assay was developed as part of an EU/Welsh Government funded project through an A4B collaborative project with Waitrose, Dovecot Park and Daleheads Foods, whose financial support is gratefully acknowledged by RMcM and AD in particular. The indispensable technical assistance of Charly Potter, Caron Evans and Kirsten Skott is also gratefully acknowledged in the generation of SSR and sequencing result.

References:

Achilli A, Olivieri A, Soares P, Lancioni H, Hooshiar Kashani B, Perego UA, Nergadze SG, Carossa V, Santagostino M, Capomaccio S, et al. (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proc. Natl. Acad. Sci. USA 109, 2449-2454.

Barton NH & M Slatkin, (1986) A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* 56:409-415

Edwards, E. (1992). Leading the Field: British Native Breeds of Horses and Ponies.

London, UK: Stanley Paul & Co. Ltd.

Grégoire Leroy, Lucille Callède, Etienne Verrier, JeanClaude Mériaux, Anne Ricard, Coralie Danchin-Burge and Xavier Rognon (2009) Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genetics Selection Evolution*, 41:5 doi:10.1186/1297-9686-41-5

Ham E, (2014) Genetic relatedness of section A Hill Pony Improvement Society animals to Lowland Section A Welsh ponies determined through microsatellite investigation. MSc thesis, IBERS, Aberystwyth University.

Nei M & Kumar S (**2000**) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.

Piry S, Alapetite A, Cornuet, J.-M., Paetkau D, Baudouin, L., Estoup, A. (2004) GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity* **95**:536-539

Raymond M. & Rousset F, (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86:248-249

Rousset, F., (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8: 103-106. Available from http://kimura.univ-montp2.fr/~rousset/Genepop.htm.

Rzhetsky A & Nei M (**1992**) A simple method for estimating and testing minimum evolution trees. *Molecular Biology and Evolution* **9**:945-967.

Saitou N & Nei M (**1987**) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.

A. Stachurska, A. Nogaj, A. Brodacki, J. Nogaj, and J. Batkowska (2014) Genetic distances between horse breeds in Poland estimated according to blood protein polymorphism. Czech J. Anim. Sci., 59, 2014 (6): 257–267.

Tamura K, Nei M & Kumar S (**2004**) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proceedings of the National Academy of Sciences* (*USA*) **101**:11030-11035.

Tamura K, Dudley J, Nei M & Kumar S (**2007**) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**:1596-1599.

Tajima F (**1989**) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.

B. S. Weir and C. Clark Cockerham (1984) Estimating F-Statistics for the Analysis of Population Structure. *Evolution,* **38** (6): 1358-1370.

C L Winton (2013) Characterisation of the phylogenetic relationships of native pony populations within the UK and Ireland, PhD Thesis, IBERS, Aberystwyth UNiversity

Clare L Winton, Matthew J Hegarty, Robert McMahon, Gancho T Slavov, Neil R McEwan, Mina CG Davies-Morel, Charly M Morgan, Wayne Powell, Deborah M Nash (2013) Genetic diversity and phylogenetic analysis of native mountain ponies of Britain and Ireland reveals a novel rare population. *Ecology and evolution* **3** (4): 934-947.

Clare L Winton, Yves Plante, Pamela Hind, Robert McMahon, Matthew J Hegarty, Neil R McEwan, Mina CG Davies-Morel, Charly M Morgan, Wayne Powell, Deborah M Nash (2015) Comparative genetic diversity in a sample of pony breeds from the UK and North America: a case study in the conservation of global genetic resources. *Ecology and Evolution* 5 (16): 3507-3522