Erfðanefnd landbúnaðarins

Áslaug Helgadóttir, formaður
Rannsóknastofnun landbúnaðarins

Hólum í Hjaltadal, 1. desember 2004

Lokaskýrsla vegna verkefnisins:
Kortlagning erfðamarka sem tengjast sumarexemi í hrossum.

Síðastliðið sumar var safnað sýnum úr um 300 íslenskum hrossum sem flutt höfðu verið úr landi fyrir a.m.k. tveimur árum, flest til Danmerkur, Noregs og Svíþjóðar. Hrossin voru sjúkdómsgreind m.t.t. sumarexems og spurningalisti um ætterni, umhverfisaðstæður, fyrirbyggjandi aðgerðir og sjúkdómssögu var lagður fyrir eigendur hrossanna. Hrossin lifðu öll á svæðum þar sem mikið er um fluguna sem veldur sumarexemi, *Culicoides spp.*
Blóðsýnin voru send til erfðagreiningar hjá erfðafræðideil Animal Health Trust í Newmarket. Með þeim fylgdu upplýsingar um sjúkdómsgreininguna, ætterni og umhverfisaðstæður. Meðfylgjandi er skýrsla um niðurstöður greiningar á þessum gögnum.

Hér með er óskað eftir því að síðari hluti styrksins verði greiddur inn á reikning yfirdýralæknis með sama hætti og áður.

Virðingarfyllst,

Sigríður Björnsdóttir
Dýralæknir hrossasjúkdóma
Yfirdýralæknisembættinu

Genetic Linkage Mapping of Loci contributing to Sweet Itch (Recurrent Seasonal Pruritus) in the Horse

Supporting Detailed Final Report Document

Sample Collection

Blood samples were taken by our named collaborators in Iceland, and shipped on to us for genetic analysis. These fell into two consignments, both consisting of Icelandic Horses which had been exported from Iceland into mainland Europe. *Culicoides* species, which cause Sweet Itch, are not found in Iceland but Icelandic horses are particularly prone to Sweet Itch after exportation, and this makes them a suitable target for investigation.

The first consignment of blood samples consisted of several affected stallions and their affected and unaffected offspring. Firstly their paternity was double-checked using the standard paternity testing panel and any samples which disagreed with their alleged sire were abandoned. Subsequently these samples were genotyped using a whole genome scanning (WGS) panel of 152 markers (Batch 1, n=98).

The second consignment (Batch 2, n=190) of blood samples was collected from Icelandic horses which had been born in Iceland and then imported into either Sweden or Norway between one and twenty-four years ago at the age of one to twelve years. The current ages of the horses range from two to twenty-eight years. A clinical examination was performed on each horse at the time of sampling and each horse was diagnosed as either free from Sweet Itch (unaffected horses numbered 101) or affected on a grade of 1 [mild] to 3 [severe] (affected horses numbered 89). A questionnaire was also completed which noted relevant information such as the conditions in which the horse was kept and whether the horse suffered from any other allergic diseases such as COPD. It was also confirmed that all sampled horses did live in a region known to suffer from *Culicoides* infestation. These samples were subject to a WGS consisting of 100 markers.

The WGS markers are assembled into efficient multiplex groups, in which 3 markers amplify together. We are using an M13-tail approach (Oetting *et al.*,1995) in which the forward primer for each equine microsatellite has been tailed with an M13 sequence. A fluorescently-labelled M13 primer is then added to the PCR reaction, and in a multiplex the decreased complexity of the primers has been found to greatly facilitate multiplexing.

Candidate Genes

Twenty candidate genes, which were chosen due to their contribution to human allergic disease, were investigated for their involvement in Sweet Itch. These genes are located on twelve human chromosome regions. For each gene the precise location on the human genome was identified by referring to the Ensembl Human Genome Browser on the Wellcome Trust Sanger Institute web site (http://www.ensembl.org/Homo_sapiens/), and the corresponding horse genome location was estimated by referring to the most current horse/human comparative map (Chowdhary et al., 2003). An attempt was made to characterise markers linked to each of these candidate genes by identifying BAC (bacterial artificial chromosomes) clones which carried each horse gene, followed by the sequencing of sub-clones (**Method 1**). In practice this approach proved inefficient and disappointing since the second stage usually failed to provide microsatellites even when as many as 1000 random sub-clones were sequenced. Despite the previous success using this method with the dog genome at the AHT, the horse genome has yielded very few microsatellites from sub-cloned BAC clones. This is most probably due to a scarcity of microsatellites in the horse genome with respect to the dog. This difficulty led us to investigate other methods to identify markers in the vicinity of candidate genes. These other methods are referred to as **Methods 2-4** and all are described in more detail below:

Method 1: The identification of relevant BACs was attempted by probing the BAC library with overgo probes designed from the horse or human sequence for the gene. The BACs were then sub-cloned and screened or randomly sequenced for microsatellites. This approach potentially provides markers which are very tightly linked to the desired gene.
Method 2: The recent generation of data describing the cross-species homology between horse microsatellite flanking sequence or horse BAC end sequence with human genomic sequence (Farber and Medrano, 2004; Chowdhary et al., 2003; M.M. Binns and T. Leeb, unpublished data) has allowed the direct selection of markers or BAC clones in the correct genomic region, nominally defined as within 2Mb, as estimated on the human genome. This methodology is not consistently expected to provide markers which are as tightly linked to the gene as Method 1. Nevertheless, the flanking sequence for one identified marker (TKY458) directly matches the sequence for one of the candidate genes (nitric oxide synthase) (Farber and Medrano, 2004).
Method 3: In the interim we have also obtained BAC clones for several of these genes from collaborating colleagues at Cornell, NY, which they had previously identified during the course of their work. These were sub-cloned and screened as in Method 1.
Method 4: Additionally one candidate gene has a microsatellite in the promoter region, the details of which have already been published (Caetano and Bowling, 1998).

Table 1 describes the markers which were identified near candidate genes. These microsatellites were routinely mapped onto the genetic map using the Newmarket horse linkage-mapping reference family to confirm that they were located in the expected region. The microsatellites were then used to genotype the Sweet Itch samples and investigate association of the microsatellite, and hence the candidate gene, with Sweet Itch.
For all twenty genes either a BAC clone which contains the gene, or a linked marker were identified. For ten candidate genes a linked marker was obtained using one of the four methods described, and used to genotype the two sweet itch sample consignments. The number of alleles which was seen among the sweet itch samples is listed for each marker. Interestingly, when the microsatellite identified for interferon gamma was mapped onto the Newmarket reference family four distinct alleles were observed. However, when used with the Sweet Itch samples alleles which could not be categorised were observed, as their sizes fell in a continuum. This observation is reminiscent of many microsatellites identified in the hyper-variable MHC class II region in both the dog and horse genomes. MHC5 is one such marker, where it was unclear whether there were eight or nine alleles segregating in the Sweet Itch samples (Table 1).

Analysis
The genotyping data was subject to statistical examination to reveal associations of marker with clinical disease.
Samples in Batch 1 were from related individuals and initial examination has not revealed any association of a marker with the clinical disease.
The samples in Batch 2 were from unrelated individuals. Given the complex and unknown mode of inheritance, the chosen method of analysing associations between the markers and clinical disease was the chi-squared test of homogeneity of proportions (Sham, 1998). This approach involves testing each marker for unequal allele frequencies between the affected and unaffected horses. Due to sparseness of the contingency tables relating allele frequency to disease status, the significance of the standard Pearson chi-squared statistic was calculated by Monte Carlo simulation (Sham, 1998). Results were obtained for 113 markers and resulted in the identification of twelve markers with significant associations at the 5% level (i.e. $p<0.05$) (Table 2). Of these twelve, two mapped to the same chromosome. From our data-set of 113 markers, 5-6 would be expected by chance to display a p value of <0.05. One way of reducing the risk of false detection of a disease-marker association is to set the threshold at which the p
value is considered significant at <0.01 (where only 1% of markers would be expected to give this result by chance). Five markers from the data set of 113 displayed p values below 0.01.
In addition each marker was subjected to a pooling regime in which each allele was compared to a pool of the others. This protocol provided an indication of whether any single allele was responsible for an association, or whether several alleles were segregating with the condition. The most significant p value obtained for each marker from this protocol is referred to as p-max (after

adjustment for multiple testing, as described by Sham, 1998). Similarity between p-max and the overall p-value for a significant marker suggests that one allele is responsible for most or all of the association. In addition, where a non-significant p-max is obtained for a marker with a significant overall p value, this may suggest that multiple alleles contribute to the association which has been detected; in three cases further investigation has revealed which alleles these are likely to be and this is shown in Table 2. However, use of this pooled approach is less powerful than Monte Carlo simulation of the full contingency table (as described above), and non-significance of p-max may simply reflect the adjustment for multiple tests of uninformative alleles (eg allele 230 for marker TKY311 has a non-significant p-max even though it is the only allele with a substantial increase in frequency amongst the affected samples).

Further examination of these significant results show that five have data-sets which are less than 50% complete and it will therefore be worthwhile completing these data-sets.

For all markers listed except AHT010 and HMS041, the biased allele was present in >25% of affected animals – this would be an important pre-requisite for any marker which was to be used in a diagnostic test as the associated allele would need to identify a significant proportion of affected animals.

At first the most promising associations appear to be those demonstrated by VHL123A, where alleles 169 and 173 were twice as likely to be found in an affected horse. Similarly for marker TKY311, allele 230 is twice as likely to be found in an affected animal. However, perhaps the most suggestive marker appears to be COR018 where alleles 280 and 294 are found in 31% of affected animals, but in none of the unaffected animals.

References

Oetting,W.S. et al. 1995. *Genomics* **30**: 450-458.
Chowdhary et al. 2003. *Genome Research* **13**: 742-751.
Farber and Medrano, 2004. *Animal Genetics* **35**: 28-33.
Caetano and Bowling, 1998. *Genome* **41**: 70-73.
Sham, 1998. Statistics in Human Genetics. Arnold, London.

Expenditure on Project

Budget Headings*[1]	Budget (Yr 1)	Actual Or to date 2001	Budget (Yr 2)	Actual Or to date 2002	Budget (Yr 3)	Actual or to date 2003	TOTALS Budget	Actu
Staff costs	*21,369*	*24,879*	*23,187*	*17,214*	*25,196*	*20,413*	*69,752*	*62,*
Consumables	*23,200*	*27,492*	*10,800*	*13,947*	*12,000*	*13,219*	*46,000*	*54,*
Travel	*800*	*343*	*800*		*800*		*2,400*	
Equipment	*14,950*	*14,820*					*14,950*	*14,*
Totals	*60,319*	*67,534*	*34,787*	*31,161*	*37,996*	*33,632*	*133,102*	*132,*

Dr. Matthew Binns
Animal Health Trust
Newmarket

Table 1

Status of the Identification of Markers near Candidate Genes for Sweet Itch – September 2004

Gene	Symbol	Human location -cytogenetic	Human location -genomic	Horse location[1]	BAC/Marker identified (position on human genome)		Method of Identification[2]	Number of Alleles in Sweet Itch Samples
Interleukin 10	IL10	1q32.1	204Mb	5p	100A3 (204.1Mb)	*- no marker found in 1000 sub-clones -*	Method 2	
					25C01	*- no marker found in 1000 sub-clones -*	Method 3	
Interleukin 3	IL3	5q23.3	131.4Mb	14p	101M11 (131.3Mb)	*- no marker found in 1000 sub-clones -*	Method 2	
					100K22 (130.1Mb)		Method 2	
					101M13 (131.2Mb)		Method 2	
Interleukin 5	IL5	5q23.3	131.9Mb	14p	287E11	}	Method 3	
Interleukin 4	IL4	5q23.3	132.0Mb	14p	287E11	} *- no marker found in 700 sub-clones -*	Method 3	
Interleukin 13	IL13	5q23.3	132.0Mb	14p	287E11	}	Method 3	
					100P23 (133.3Mb)		Method 2	
Interleukin 9	IL9	5q31.1	135.3Mb	14p	TKY844 (5q31.1: 135.8Mb)		Method 2	1
Beta-2 adrenergic receptor	ADRB2	5q33.1	148.2Mb	14p	101N12 (149.8Mb)	*- no marker found in 1000 sub-clones -*	Method 2	
Interleukin 12B	**IL12B**	**5q33.3**	**158.7Mb**	**14p**	**101E7 (159Mb)**		**Method 2**	**4**
					TKY1017 (159.0Mb)	*- no amplification -*	Method 2	
Major histocompatability complex	**MHC**	**6p21.32**	**~32.5Mb**	**20p**	**MHC1 from 288J19**		**Method 3**	**6**
					MHC4 from 288J19		**Method 3**	**8**
					MHC5 from 288J19		**Method 3**	**8/9**
Tumour necrosis factor a	**TNFA**	**6p21.33**	**31.65Mb**	**20p**	**15G08**		**Method 3**	**6**
High affinity IgE receptor B chain	FCER1B	11q12.1	59.6Mb	7q	100J24 (58.9Mb)	*- unscorable marker -*	Method 2	
					100N7 (58.8Mb)		Method 2	
					101B15 (58.9Mb)		Method 2	
Interferon gamma	IFNG	12q15	66.8Mb	6q	60D21	*- couldn't bin alleles -*	Method 1	
					35E01	*- no marker found in 300 sub-clones -*	Method 3	
Leukotriene A4 hydrolase	**LTA4H**	**12q23.1**	**94.9Mb**	**28p**	**TKY1153 (12q23.1:95.9Mb)**		**Method 2**	**5**
Insulin-like growth factor 1	**IGF1**	**12q23.2**	**101.3Mb**	**28p**	**IGF1**		**Method 4**	**5**
Nitric oxide synthase	**NOS1**	**12q24.22**	**116.1Mb**	**8p**	**TKY458 (116.2Mb)**		**Method 2**	**6**
Esterase D	**ESTD**	**13q14.2**	**45.1Mb**	**17p**	**COR067 (46.1Mb)**		**Method 2**	**2**
					HMS041 (13q14.2: 46.8Mb)		**Method 2**	**4**
Mast cell chymase	**CMA1**	**14q11.2**	**22.9Mb**	**24p**	**COR061 (21.0Mb)-**		**Method 2**	**13**
T cell receptor a chain	**TCRA**	**14q11.2**	**20.4Mb**	**1q**	**100C5 (21.3Mb)**		**Method 2**	**13**
Immunoglobulin E	IGHE	14q32.33	104.0Mb	?	67G21	*- no marker found in 1000 sub-clones -*	Method 3	
Interleukin 4 receptor a chain	**IL4R**	**16p12.1**	**27.3Mb**	**13q**	**mixed prep of 5 BACs**		**Method 3**	**8**

[1] Chowdhary et al., 2003

[2] **Method 1**: Probing the BAC library with overgo probes designed from the horse or human sequence for the gene. The BACs are then sub-cloned and screened for microsatellites.
Method 2: The recent generation of data describing the cross-species homology between horse microsatellite flanking sequence or horse BAC end sequence with human genomic sequence has allowed the direct selection of some markers or BAC clones in the correct region. The position of homology with the human genome is quoted.
Method 3: We obtained some BACs from collaborators in Cornell, NY which they had previously identified as containing the relevant gene during the course of their work.
Method 4: A microsatellite marker which lies within this gene has already been published (Caetano and Bowling, 1998).

Candidate genes for which a linked marker has been identified are shown in bold.
Other BACs to examine shown in blue

Table 2

ECA	Marker	Associated allele(s)	Number of animals with specified allele(s)		Total number of animals scored for this marker		Percentage of animals containing specified allele(s)		chisq	p	p-max
			Affected	Unaffected	Affected	Unaffected	Affected	Unaffected			
2	VHL123A	169 &173	32	19	78	92	41	21	14.27	0.009	0.06
5	AHT024	216	5	3	11	13	45	23	5.22	0.03	NA
		214	6	12	11	13	55	92			
9	COR008	269	5	1	13	19	38	5	14.91	0.039	0.114
9	LEX019	175 &180	26	13	62	69	42	19	12.58	0.008	0.106
16	TKY311	230	19	11	37	45	51	24	19.48	0.006	0.122
17	HMS041	128	12	2	89	101	13	2	11.13	0.021	0.007
20	MHC5	175	23	10	88	99	26	10	15.31	0.037	0.017
21	AHT010	138	0	3	78	88	0	3	6.43	0.038	0.095
23	AHT072	245	41	31	71	82	58	38	13.44	0.008	0.054
25	COR018	280 & 294	5	0	16	14	31	0	15.39	0.009	0.086
26	UM066	139	28	22	81	100	35	22	7.33	0.025	0.046
X	LEX010	221	8	6	21	62	38	10	12.1	0.029	0.048