TIERÄRZTLICHE HOCHSCHULE HANNOVER · STIFTUNG ·
17 78

HELMHOLTZ
ZENTRUM FÜR
INFEKTIONSFORSCHUNG

57th Annual Meeting of the European Association for Animal Production

Antalya, Turkey, 17-20 September 2006

Development of a BAC-based physical map of the horse genome

Session H36.11

O. Distl, T. Leeb, H. Blöcker

ottmar.distl@tiho-hannover.de

VolkswagenStiftung

Objectives

Facilitating genome navigation

150,000 CHORI-241 clones

(11x coverage, average insert size: 171 kb)

- → **BAC end sequences (BESs)**
- → **fluorescent BAC fingerprints (4 colors, ABI 3730xl)**
  - → **data collection in FPC software**
  - → **genome-wide BAC contigs**
  - → **anchoring FPC-contigs to RH-map and WGS horse assembly**

Horse BES Analysis

Sequences in EMBL:	110,299	(plates 1-168)
Sequences done:	~150,000	
Sequences in TraceArchive:	~40,000	

Preliminary analysis of 70,299 BES:

total bases:	43.8 Mb
average read length:	618 bp
clones attempted:	41,472
clones with read pairs:	34,094 (82%)
clones with single BES:	3,993 (10%)
overall sequencing success rate:	85.5%
BLAST hits ($E \leq 10^{-5}$):	30,894 (44%)

Time Schedule

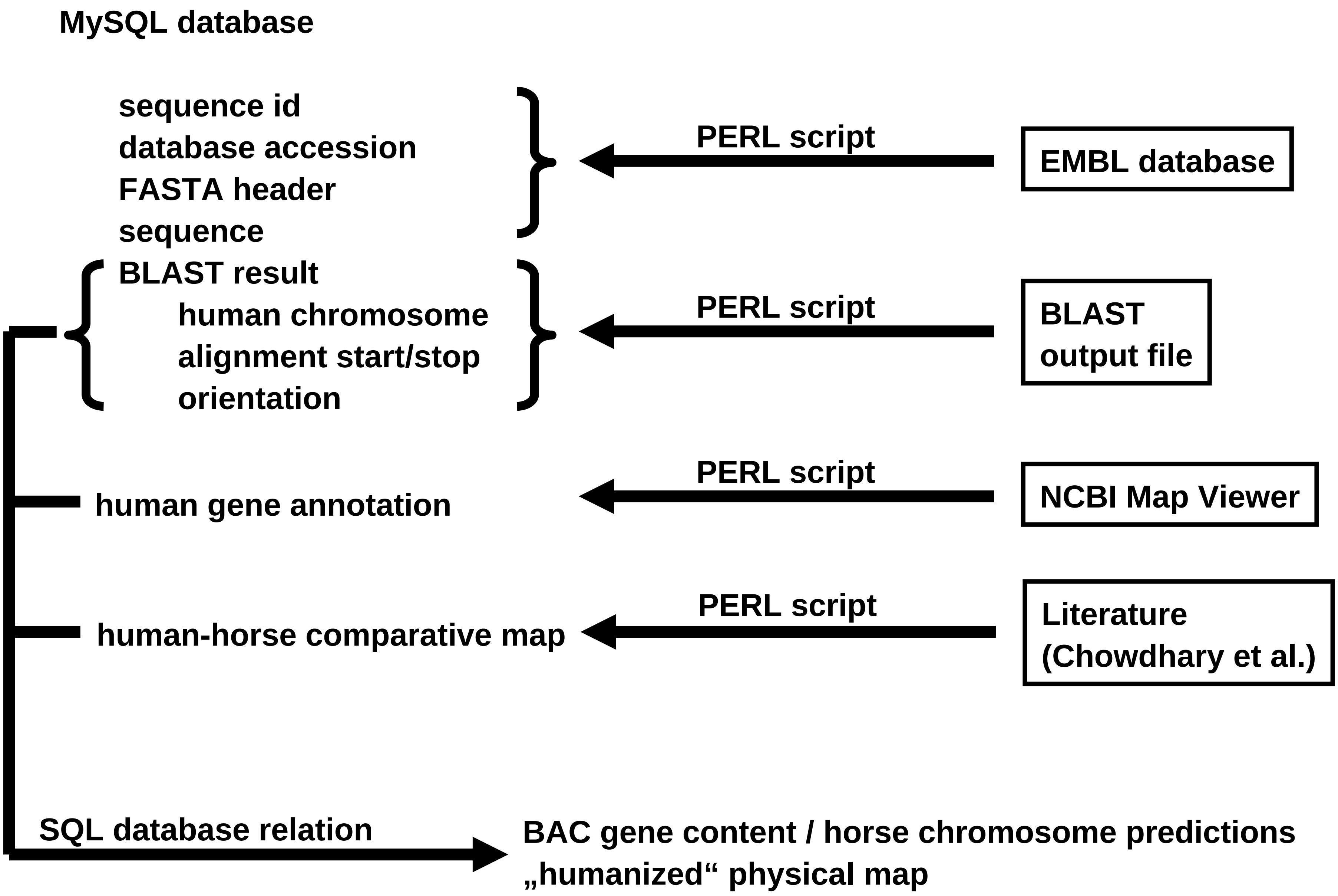
current BES capacity at HZI:

25 plates per week (384 well)

estimated finishing date (400 plates):

October 2006 (current rate)

Bioinformatics



MySQL database

sequence id
database accession
FASTA header
sequence
BLAST result
human chromosome
alignment start/stop
orientation

human gene annotation

human-horse comparative map

PERL script

EMBL database

PERL script

BLAST output file

PERL script

NCBI Map Viewer

PERL script

Literature (Chowdhary et al.)

SQL database relation

BAC gene content / horse chromosome predictions
„humanized“ physical map

Comparative Results (I)

data based on 70,299 BES

BACs with at least one BES	36,823
BACs with 1 matching BES	15,979 (43%)
BACs with 2 matching BES	6,841 (19%)

comparatively mapped BACs total	22,325 (61%)

average spacing of comparatively anchored BACs in the human genome: 0.0836 Mb

Distribution of BLAST hits over human chromosomes

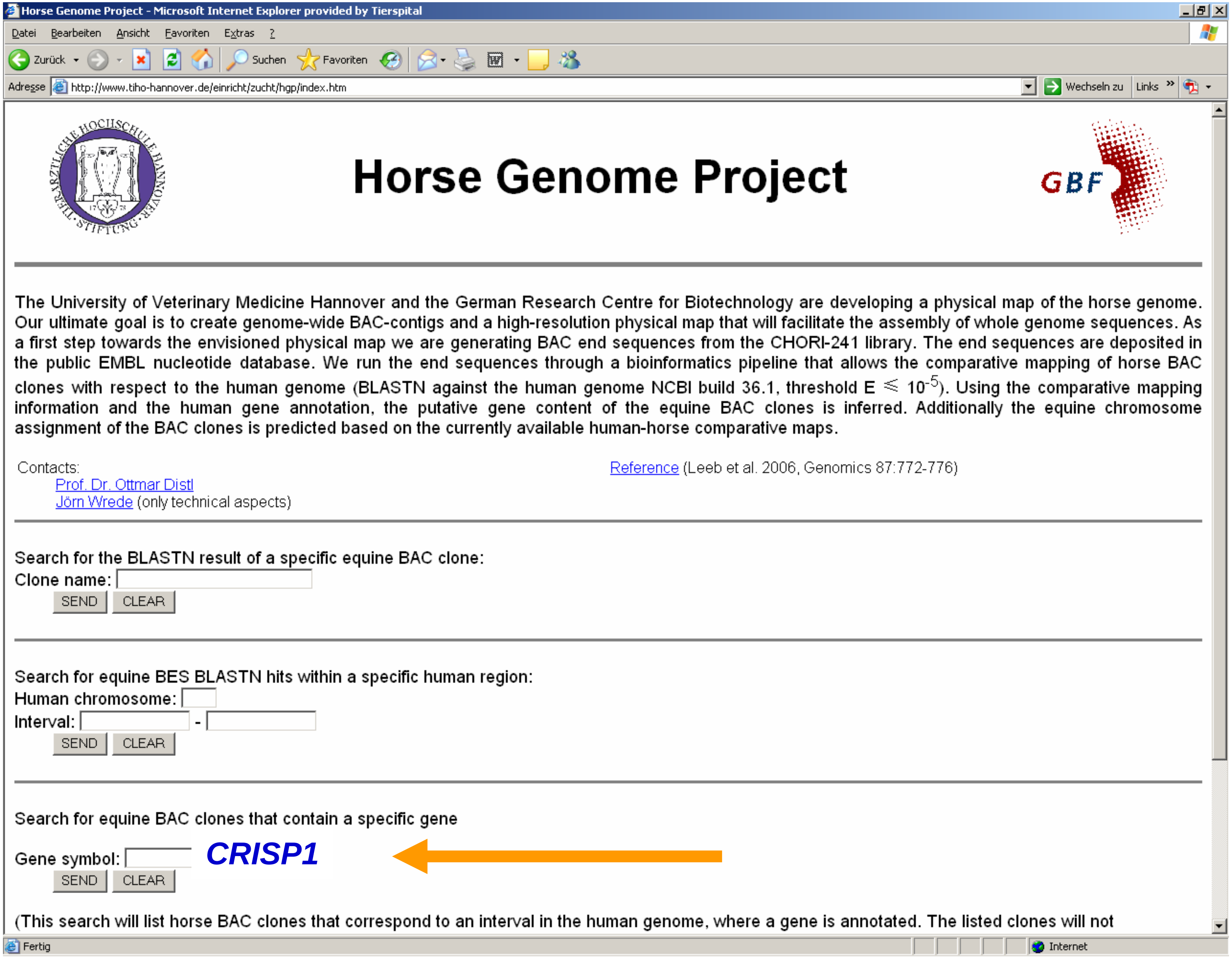
HSA	Size (Mb)	Percentage of hits	Average spacing (Mb)
1	245	8.37	0.084
10	135	4.55	0.084
22	50	0.88	0.109
X	155	2.41	0.180

Comparative Results (II)

annotated genes in NCBI build 36.1 28,914

no. of horse BACs	human genes with predicted horse BAC clones
0	9.167 (32%)
>0	19.747 (68%)
1	7.238 (25%)
2	5.263 (18%)
3	3.291 (11%)
>3	3.955 (14%)

www.tiho-hannover.de/einricht/zucht/hgp/index.htm



Horse Genome Project

The University of Veterinary Medicine Hannover and the German Research Centre for Biotechnology are developing a physical map of the horse genome. Our ultimate goal is to create genome-wide BAC-contigs and a high-resolution physical map that will facilitate the assembly of whole genome sequences. As a first step towards the envisioned physical map we are generating BAC end sequences from the CHORI-241 library. The end sequences are deposited in the public EMBL nucleotide database. We run the end sequences through a bioinformatics pipeline that allows the comparative mapping of horse BAC clones with respect to the human genome (BLASTN against the human genome NCBI build 36.1, threshold $E \leqslant 10^{-5}$). Using the comparative mapping information and the human gene annotation, the putative gene content of the equine BAC clones is inferred. Additionally the equine chromosome assignment of the BAC clones is predicted based on the currently available human-horse comparative maps.

Contacts:
Prof. Dr. Ottmar Distl
Jörn Wrede (only technical aspects)

Reference (Leeb et al. 2006, Genomics 87:772-776)

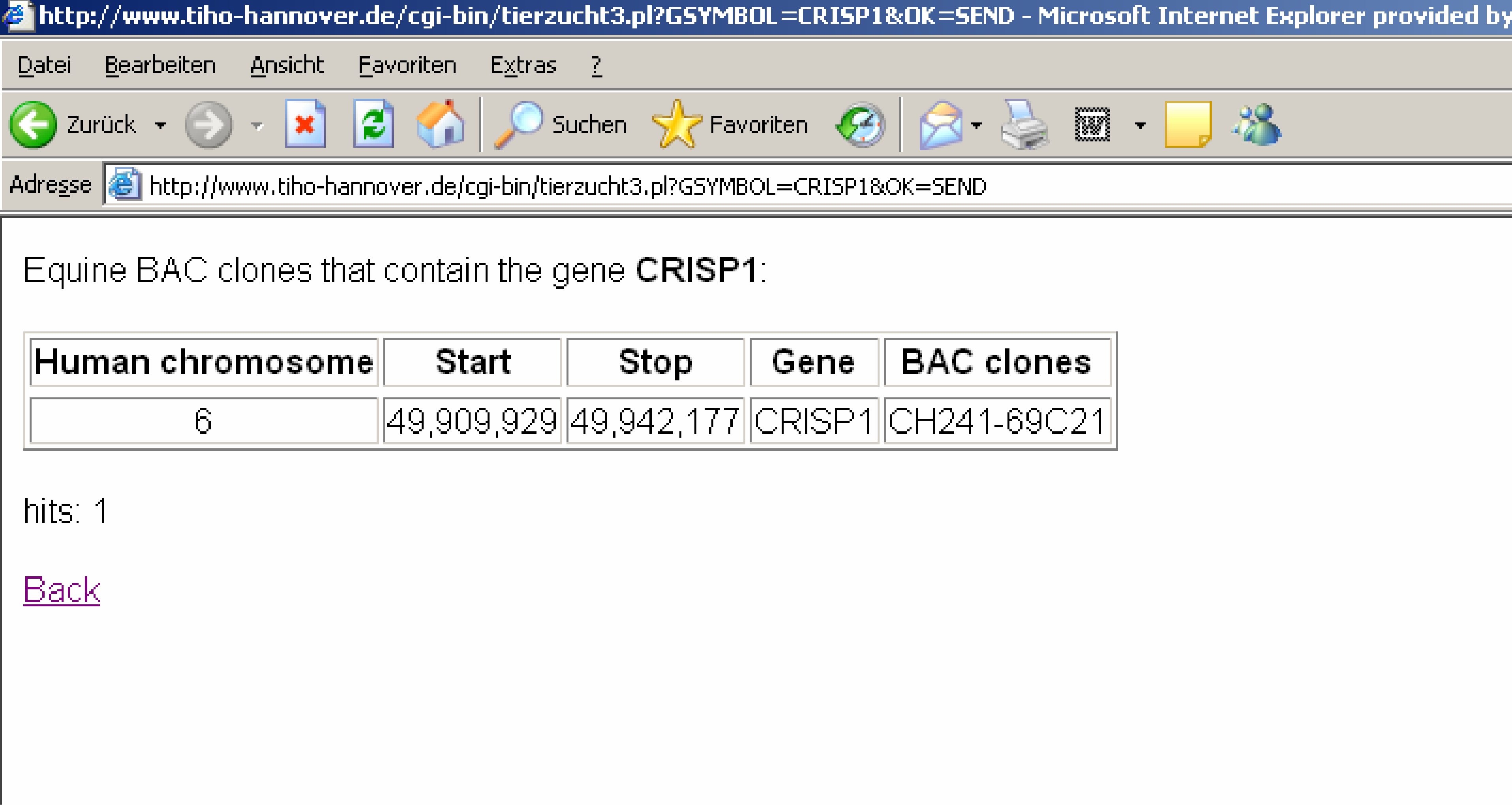
Search for the BLASTN result of a specific equine BAC clone:
Clone name:
SEND CLEAR

Search for equine BES BLASTN hits within a specific human region:
Human chromosome:
Interval: -
SEND CLEAR

Search for equine BAC clones that contain a specific gene

Gene symbol: **CRISP1**
SEND CLEAR

(This search will list horse BAC clones that correspond to an interval in the human genome, where a gene is annotated. The listed clones will not



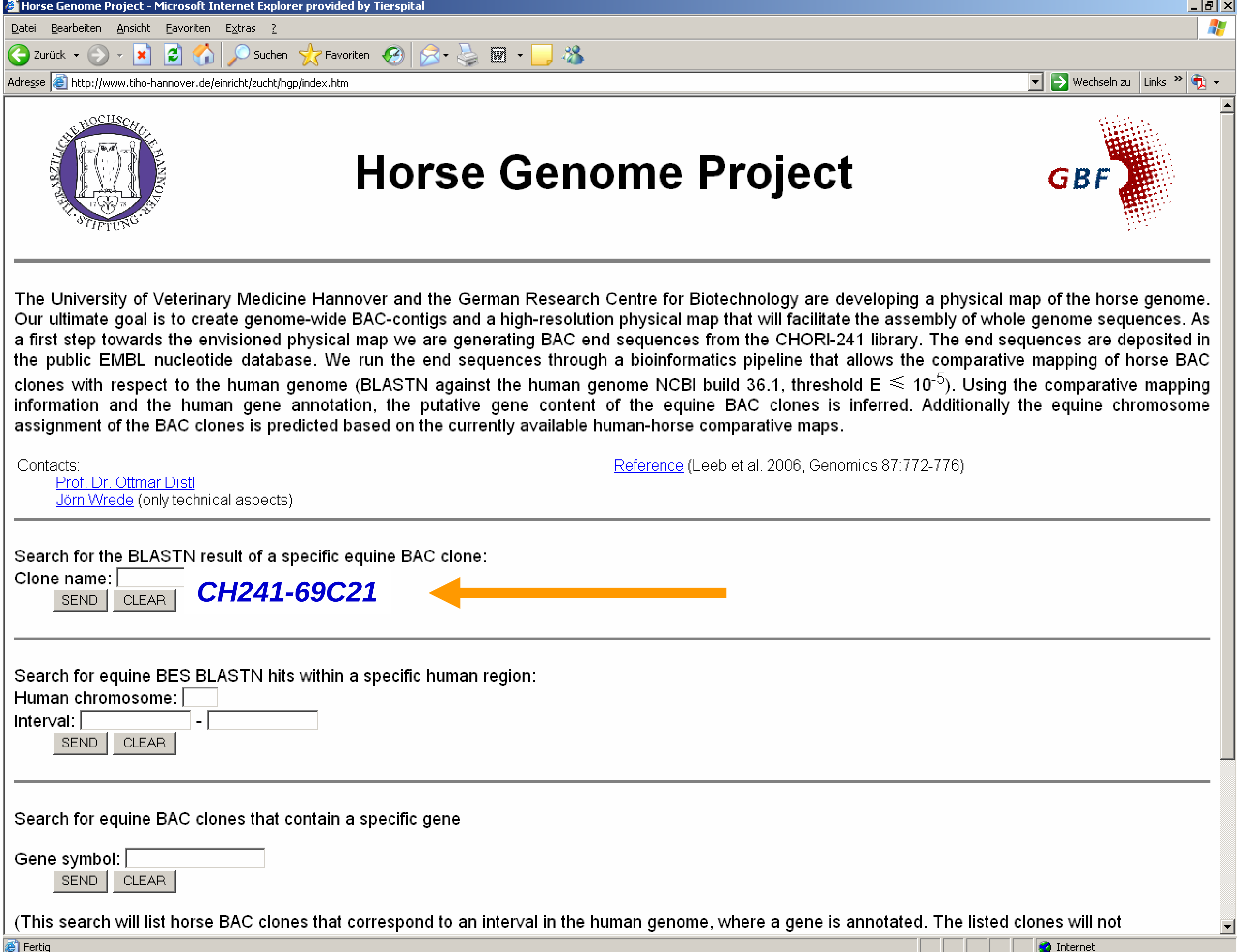
Adresse: http://www.tiho-hannover.de/cgi-bin/tierzucht3.pl?GSYMBOL=CRISP1&OK=SEND

Equine BAC clones that contain the gene **CRISP1**:

Human chromosome	Start	Stop	Gene	BAC clones
6	49,909,929	49,942,177	CRISP1	CH241-69C21

hits: 1

Back



http://www.tiho-hannover.de/einricht/zucht/hgp/index.htm

Horse Genome Project

The University of Veterinary Medicine Hannover and the German Research Centre for Biotechnology are developing a physical map of the horse genome. Our ultimate goal is to create genome-wide BAC-contigs and a high-resolution physical map that will facilitate the assembly of whole genome sequences. As a first step towards the envisioned physical map we are generating BAC end sequences from the CHORI-241 library. The end sequences are deposited in the public EMBL nucleotide database. We run the end sequences through a bioinformatics pipeline that allows the comparative mapping of horse BAC clones with respect to the human genome (BLASTN against the human genome NCBI build 36.1, threshold $E \leqslant 10^{-5}$). Using the comparative mapping information and the human gene annotation, the putative gene content of the equine BAC clones is inferred. Additionally the equine chromosome assignment of the BAC clones is predicted based on the currently available human-horse comparative maps.

Contacts:
Prof. Dr. Ottmar Distl
Jörn Wrede (only technical aspects)

Reference (Leeb et al. 2006, Genomics 87:772-776)

Search for the BLASTN result of a specific equine BAC clone:
Clone name:
SEND CLEAR

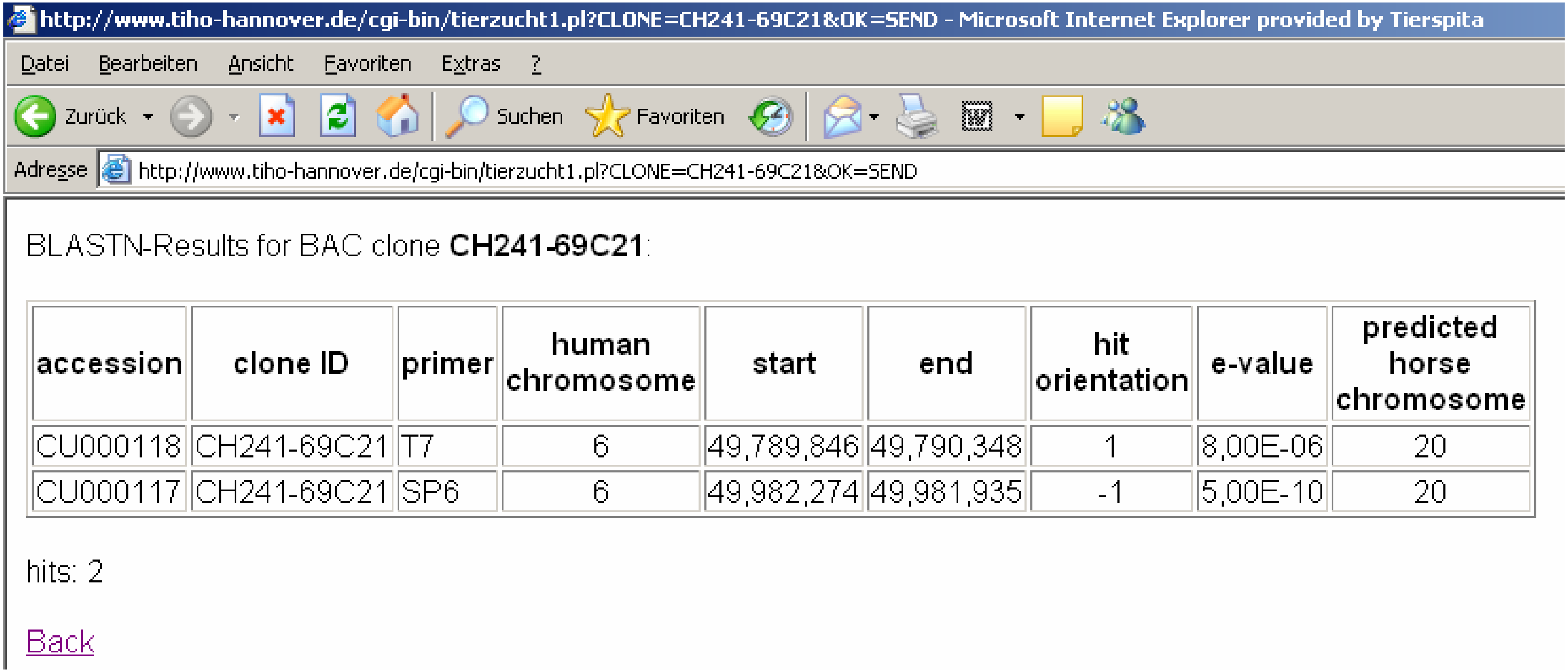
CH241-69C21

Search for equine BES BLASTN hits within a specific human region:
Human chromosome:
Interval: -
SEND CLEAR

Search for equine BAC clones that contain a specific gene

Gene symbol:
SEND CLEAR

(This search will list horse BAC clones that correspond to an interval in the human genome, where a gene is annotated. The listed clones will not

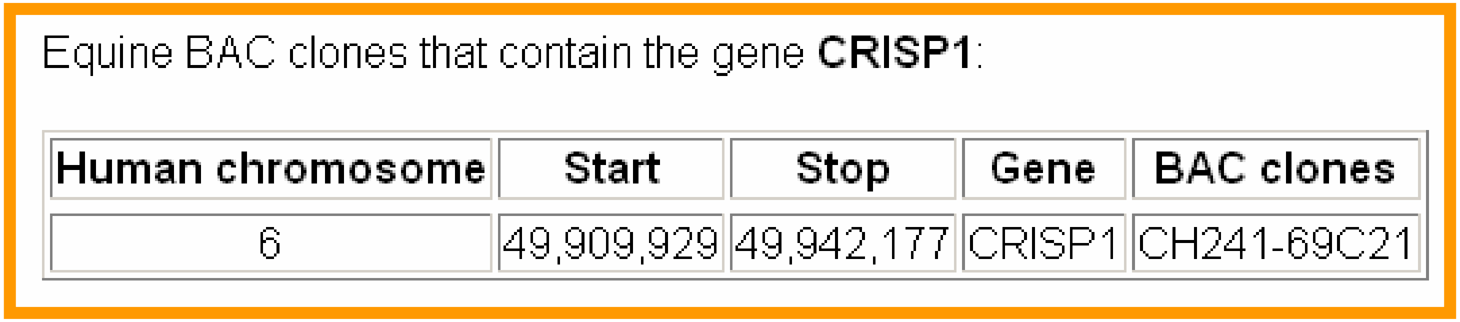


BLASTN-Results for BAC clone **CH241-69C21**:

accession	clone ID	primer	human chromosome	start	end	hit orientation	e-value	predicted horse chromosome
CU000118	CH241-69C21	T7	6	49,789,846	49,790,348	1	8,00E-06	20
CU000117	CH241-69C21	SP6	6	49,982,274	49,981,935	-1	5,00E-10	20

hits: 2

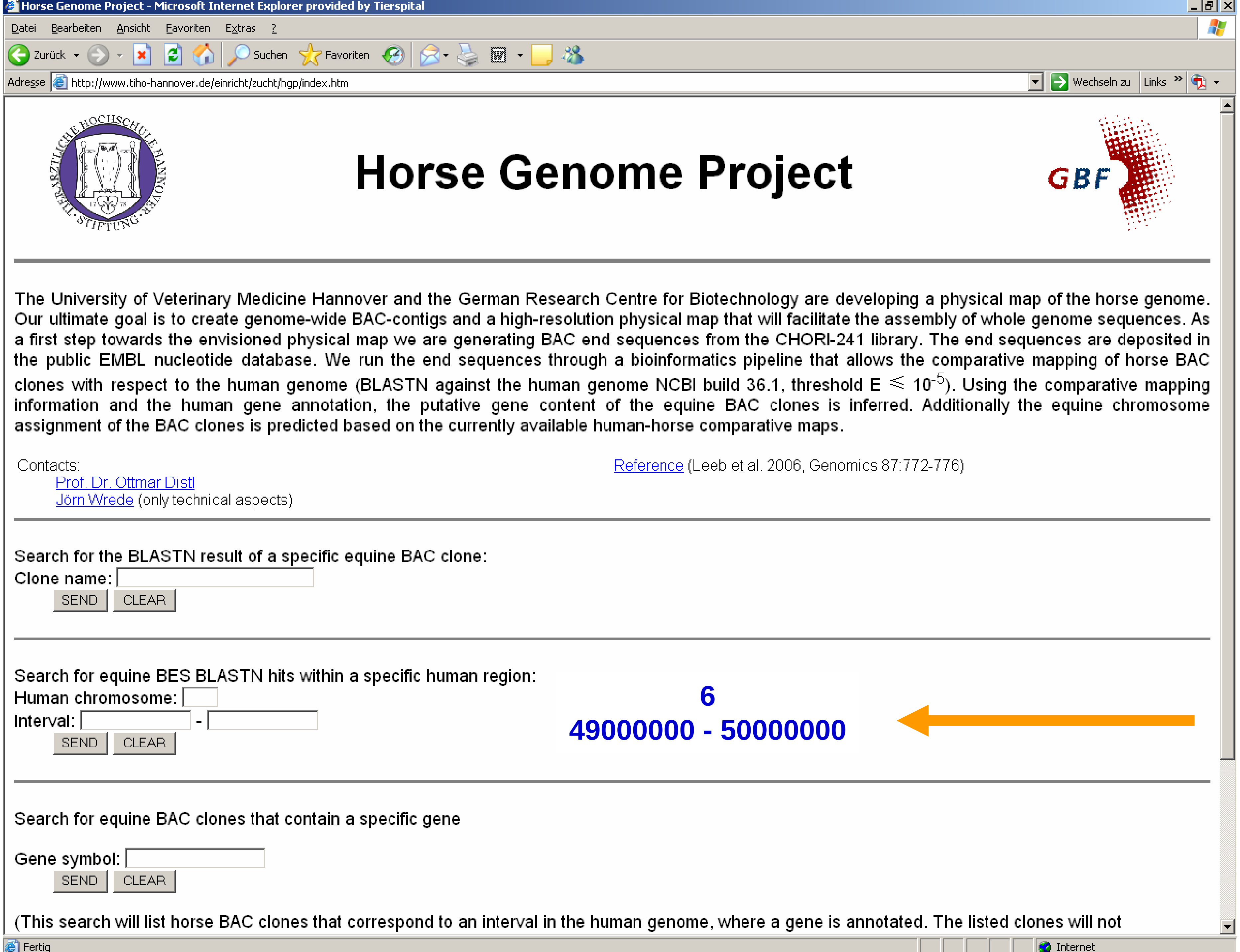
Back



Equine BAC clones that contain the gene **CRISP1**:

Human chromosome	Start	Stop	Gene	BAC clones
6	49,909,929	49,942,177	CRISP1	CH241-69C21

→ **complete *CRISP1* gene contained on CH241-69C21**



http://www.tiho-hannover.de/einricht/zucht/hgp/index.htm

Horse Genome Project

The University of Veterinary Medicine Hannover and the German Research Centre for Biotechnology are developing a physical map of the horse genome. Our ultimate goal is to create genome-wide BAC-contigs and a high-resolution physical map that will facilitate the assembly of whole genome sequences. As a first step towards the envisioned physical map we are generating BAC end sequences from the CHORI-241 library. The end sequences are deposited in the public EMBL nucleotide database. We run the end sequences through a bioinformatics pipeline that allows the comparative mapping of horse BAC clones with respect to the human genome (BLASTN against the human genome NCBI build 36.1, threshold $E \leqslant 10^{-5}$). Using the comparative mapping information and the human gene annotation, the putative gene content of the equine BAC clones is inferred. Additionally the equine chromosome assignment of the BAC clones is predicted based on the currently available human-horse comparative maps.

Contacts:
Prof. Dr. Ottmar Distl
Jörn Wrede (only technical aspects)

Reference (Leeb et al. 2006, Genomics 87:772-776)

Search for the BLASTN result of a specific equine BAC clone:
Clone name:

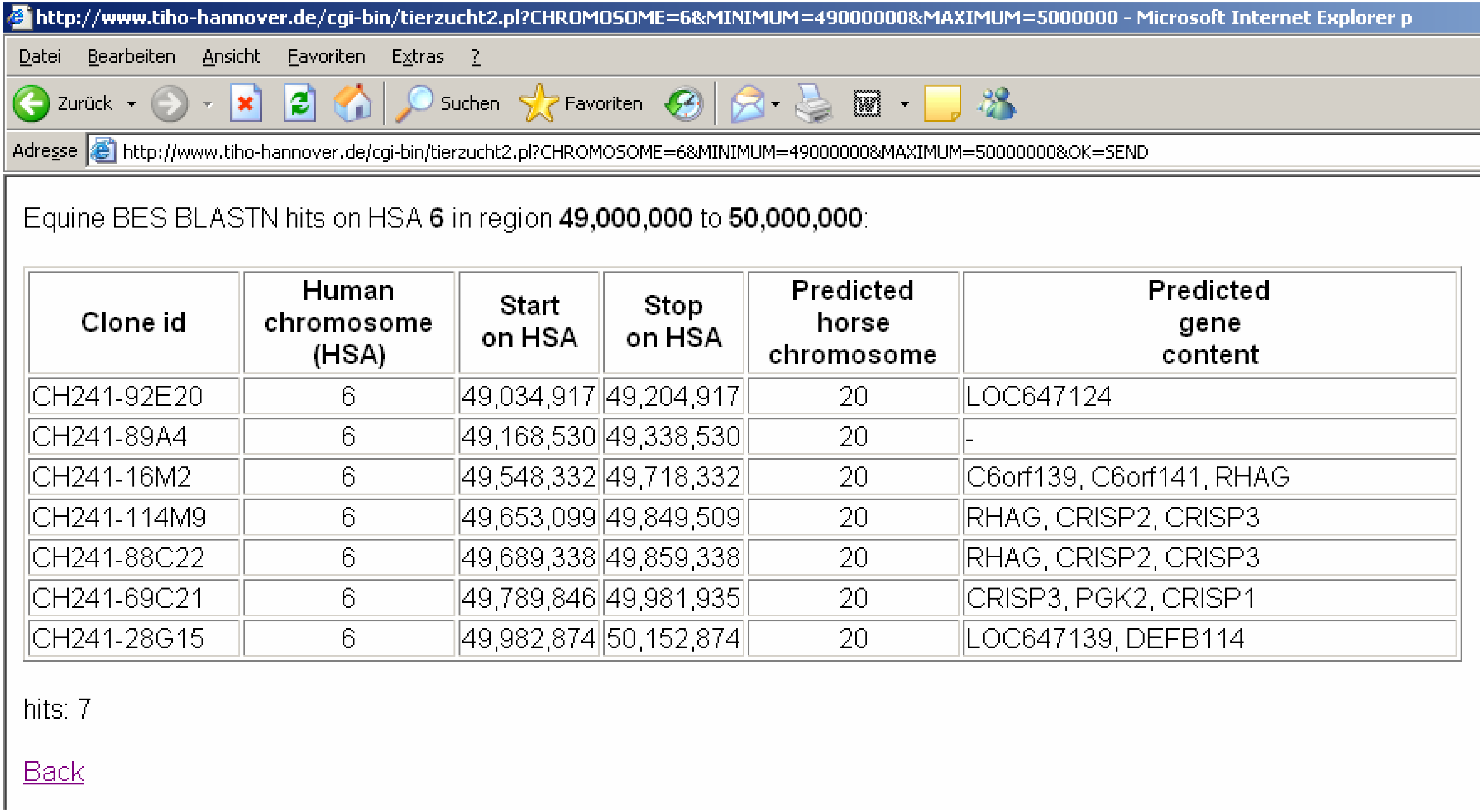
Search for equine BES BLASTN hits within a specific human region:
Human chromosome:
Interval: -

6
49000000 - 50000000

Search for equine BAC clones that contain a specific gene

Gene symbol:

(This search will list horse BAC clones that correspond to an interval in the human genome, where a gene is annotated. The listed clones will not

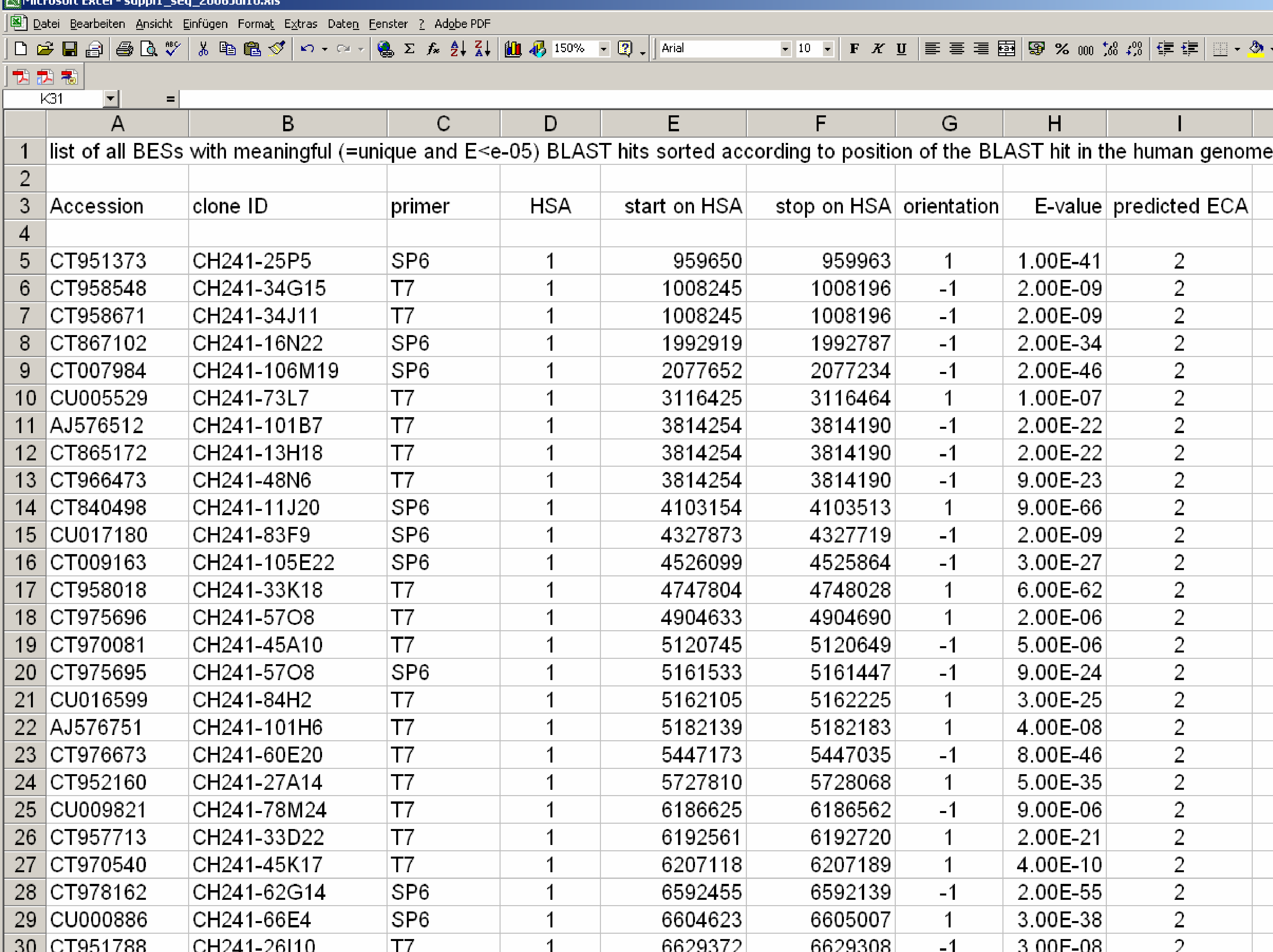


Equine BES BLASTN hits on HSA **6** in region **49,000,000** to **50,000,000**:

Clone id	Human chromosome (HSA)	Start on HSA	Stop on HSA	Predicted horse chromosome	Predicted gene content
CH241-92E20	6	49,034,917	49,204,917	20	LOC647124
CH241-89A4	6	49,168,530	49,338,530	20	-
CH241-16M2	6	49,548,332	49,718,332	20	C6orf139, C6orf141, RHAG
CH241-114M9	6	49,653,099	49,849,509	20	RHAG, CRISP2, CRISP3
CH241-88C22	6	49,689,338	49,859,338	20	RHAG, CRISP2, CRISP3
CH241-69C21	6	49,789,846	49,981,935	20	CRISP3, PGK2, CRISP1
CH241-28G15	6	49,982,874	50,152,874	20	LOC647139, DEFB114

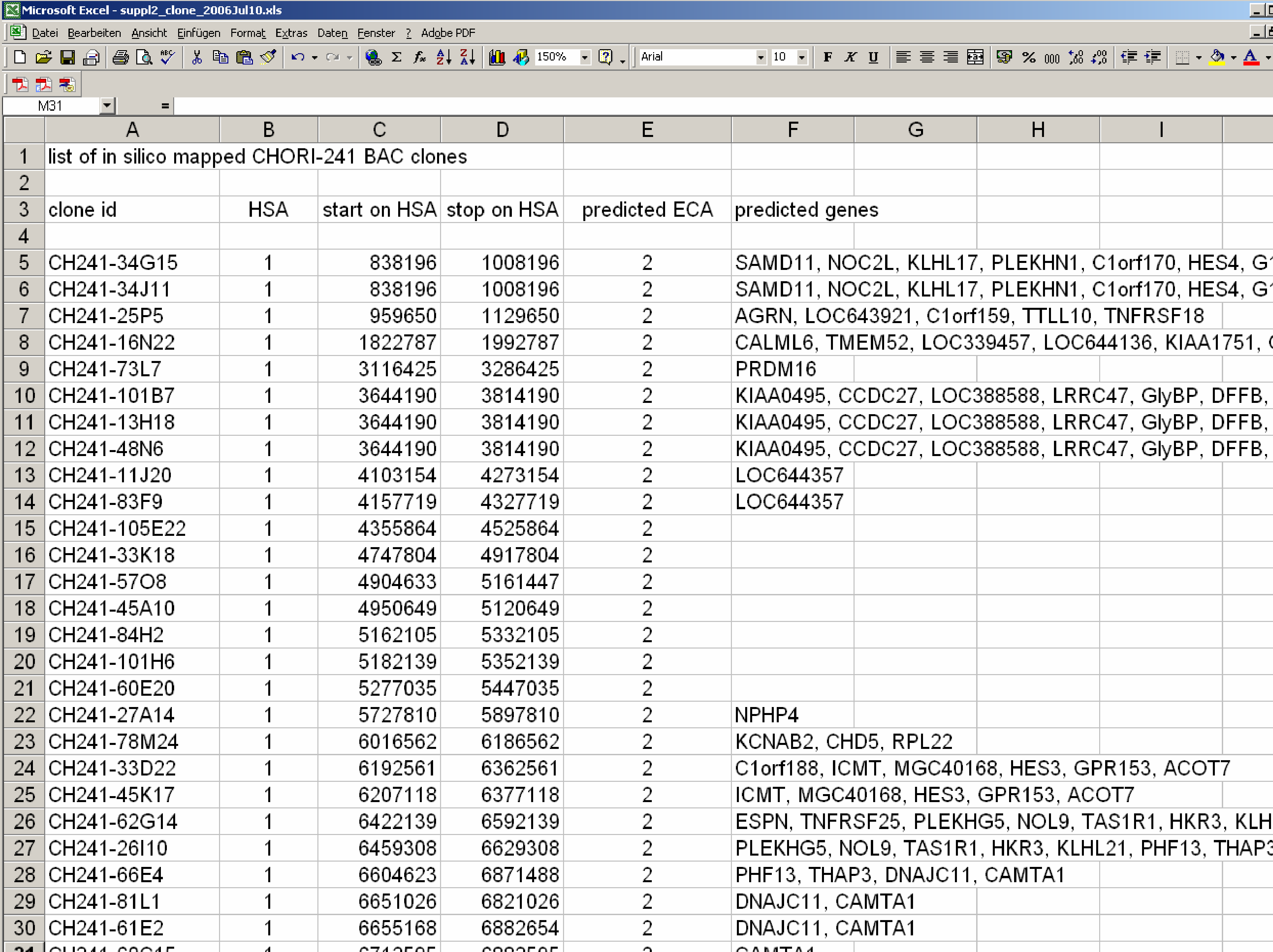
hits: 7

Back



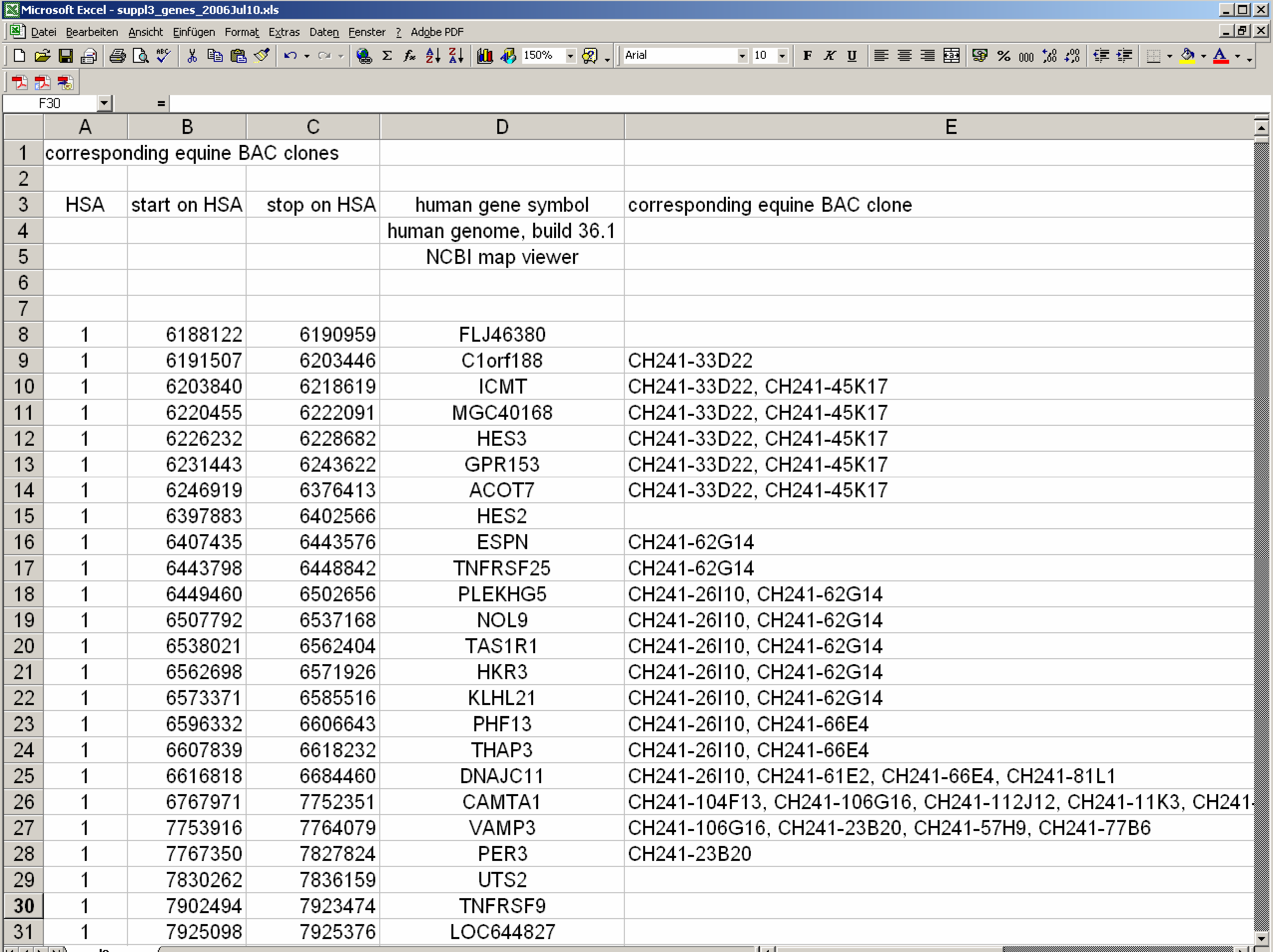
list of all BESs with meaningful (=unique and E<e-05) BLAST hits sorted according to position of the BLAST hit in the human genome

Accession	clone ID	primer	HSA	start on HSA	stop on HSA	orientation	E-value	predicted ECA
CT951373	CH241-25P5	SP6	1	959650	959963	1	1.00E-41	2
CT958548	CH241-34G15	T7	1	1008245	1008196	-1	2.00E-09	2
CT958671	CH241-34J11	T7	1	1008245	1008196	-1	2.00E-09	2
CT867102	CH241-16N22	SP6	1	1992919	1992787	-1	2.00E-34	2
CT007984	CH241-106M19	SP6	1	2077652	2077234	-1	2.00E-46	2
CU005529	CH241-73L7	T7	1	3116425	3116464	1	1.00E-07	2
AJ576512	CH241-101B7	T7	1	3814254	3814190	-1	2.00E-22	2
CT865172	CH241-13H18	T7	1	3814254	3814190	-1	2.00E-22	2
CT966473	CH241-48N6	T7	1	3814254	3814190	-1	9.00E-23	2
CT840498	CH241-11J20	SP6	1	4103154	4103513	1	9.00E-66	2
CU017180	CH241-83F9	SP6	1	4327873	4327719	-1	2.00E-09	2
CT009163	CH241-105E22	SP6	1	4526099	4525864	-1	3.00E-27	2
CT958018	CH241-33K18	T7	1	4747804	4748028	1	6.00E-62	2
CT975696	CH241-57O8	T7	1	4904633	4904690	1	2.00E-06	2
CT970081	CH241-45A10	T7	1	5120745	5120649	-1	5.00E-06	2
CT975695	CH241-57O8	SP6	1	5161533	5161447	-1	9.00E-24	2
CU016599	CH241-84H2	T7	1	5162105	5162225	1	3.00E-25	2
AJ576751	CH241-101H6	T7	1	5182139	5182183	1	4.00E-08	2
CT976673	CH241-60E20	T7	1	5447173	5447035	-1	8.00E-46	2
CT952160	CH241-27A14	T7	1	5727810	5728068	1	5.00E-35	2
CU009821	CH241-78M24	T7	1	6186625	6186562	-1	9.00E-06	2
CT957713	CH241-33D22	T7	1	6192561	6192720	1	2.00E-21	2
CT970540	CH241-45K17	T7	1	6207118	6207189	1	4.00E-10	2
CT978162	CH241-62G14	SP6	1	6592455	6592139	-1	2.00E-55	2
CU000886	CH241-66E4	SP6	1	6604623	6605007	1	3.00E-38	2
CT951788	CH241-26I10	T7	1	6629372	6629308	-1	3.00E-08	2



list of in silico mapped CHORI-241 BAC clones

clone id	HSA	start on HSA	stop on HSA	predicted ECA	predicted genes
CH241-34G15	1	838196	1008196	2	SAMD11, NOC2L, KLHL17, PLEKHN1, C1orf170, HES4, G
CH241-34J11	1	838196	1008196	2	SAMD11, NOC2L, KLHL17, PLEKHN1, C1orf170, HES4, G
CH241-25P5	1	959650	1129650	2	AGRN, LOC643921, C1orf159, TTLL10, TNFRSF18
CH241-16N22	1	1822787	1992787	2	CALML6, TMEM52, LOC339457, LOC644136, KIAA1751,
CH241-73L7	1	3116425	3286425	2	PRDM16
CH241-101B7	1	3644190	3814190	2	KIAA0495, CCDC27, LOC388588, LRRC47, GlyBP, DFFB,
CH241-13H18	1	3644190	3814190	2	KIAA0495, CCDC27, LOC388588, LRRC47, GlyBP, DFFB,
CH241-48N6	1	3644190	3814190	2	KIAA0495, CCDC27, LOC388588, LRRC47, GlyBP, DFFB,
CH241-11J20	1	4103154	4273154	2	LOC644357
CH241-83F9	1	4157719	4327719	2	LOC644357
CH241-105E22	1	4355864	4525864	2	
CH241-33K18	1	4747804	4917804	2	
CH241-57O8	1	4904633	5161447	2	
CH241-45A10	1	4950649	5120649	2	
CH241-84H2	1	5162105	5332105	2	
CH241-101H6	1	5182139	5352139	2	
CH241-60E20	1	5277035	5447035	2	
CH241-27A14	1	5727810	5897810	2	NPHP4
CH241-78M24	1	6016562	6186562	2	KCNAB2, CHD5, RPL22
CH241-33D22	1	6192561	6362561	2	C1orf188, ICMT, MGC40168, HES3, GPR153, ACOT7
CH241-45K17	1	6207118	6377118	2	ICMT, MGC40168, HES3, GPR153, ACOT7
CH241-62G14	1	6422139	6592139	2	ESPN, TNFRSF25, PLEKHG5, NOL9, TAS1R1, HKR3, KLH
CH241-26I10	1	6459308	6629308	2	PLEKHG5, NOL9, TAS1R1, HKR3, KLHL21, PHF13, THAP
CH241-66E4	1	6604623	6871488	2	PHF13, THAP3, DNAJC11, CAMTA1
CH241-81L1	1	6651026	6821026	2	DNAJC11, CAMTA1
CH241-61E2	1	6655168	6882654	2	DNAJC11, CAMTA1



	A	B	C	D	E
1	corresponding equine BAC clones				
2					
3	HSA	start on HSA	stop on HSA	human gene symbol	corresponding equine BAC clone
4				human genome, build 36.1	
5				NCBI map viewer	
6					
7					
8	1	6188122	6190959	FLJ46380	
9	1	6191507	6203446	C1orf188	CH241-33D22
10	1	6203840	6218619	ICMT	CH241-33D22, CH241-45K17
11	1	6220455	6222091	MGC40168	CH241-33D22, CH241-45K17
12	1	6226232	6228682	HES3	CH241-33D22, CH241-45K17
13	1	6231443	6243622	GPR153	CH241-33D22, CH241-45K17
14	1	6246919	6376413	ACOT7	CH241-33D22, CH241-45K17
15	1	6397883	6402566	HES2	
16	1	6407435	6443576	ESPN	CH241-62G14
17	1	6443798	6448842	TNFRSF25	CH241-62G14
18	1	6449460	6502656	PLEKHG5	CH241-26I10, CH241-62G14
19	1	6507792	6537168	NOL9	CH241-26I10, CH241-62G14
20	1	6538021	6562404	TAS1R1	CH241-26I10, CH241-62G14
21	1	6562698	6571926	HKR3	CH241-26I10, CH241-62G14
22	1	6573371	6585516	KLHL21	CH241-26I10, CH241-62G14
23	1	6596332	6606643	PHF13	CH241-26I10, CH241-66E4
24	1	6607839	6618232	THAP3	CH241-26I10, CH241-66E4
25	1	6616818	6684460	DNAJC11	CH241-26I10, CH241-61E2, CH241-66E4, CH241-81L1
26	1	6767971	7752351	CAMTA1	CH241-104F13, CH241-106G16, CH241-112J12, CH241-11K3, CH241-
27	1	7753916	7764079	VAMP3	CH241-106G16, CH241-23B20, CH241-57H9, CH241-77B6
28	1	7767350	7827824	PER3	CH241-23B20
29	1	7830262	7836159	UTS2	
30	1	7902494	7923474	TNFRSF9	
31	1	7925098	7925376	LOC644827	

Conclusions

- **Creation of an ordered genome-wide BAC library**
- **Anchors on equine RH-map and WGS horse assembly**
- **Genome-wide and equally distributed genomic polymorphisms (genome-wide SNPs)**
- **Enhancement of functional horse genomics**

Acknowledgements

TiHo Hannover:

Tosso Leeb
Anne Wöhlke
Jörn Wrede

HZI Braunschweig:

Helmut Blöcker
Gabriele Nordsieck
Melanie Scharfe
Frank Schrader
Michael Jarek

Etienne Bäumle
Claus Vogl
Bhanu Chowdhary
Pieter de Jong

VolkswagenStiftung

Thanks for your attention !