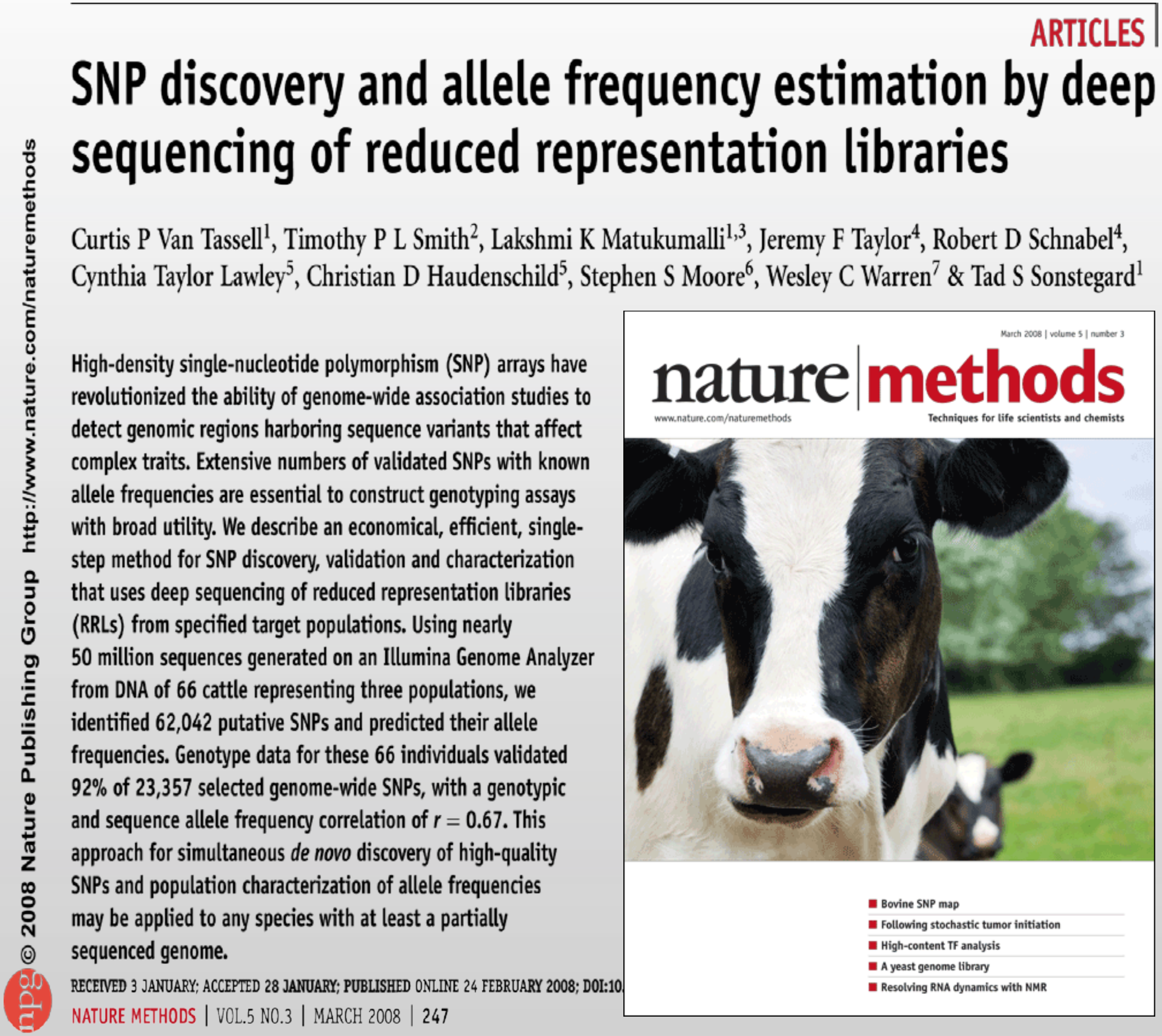
High Throughput Genomic Technologies for Bison

Robert Schnabel

Tulsa, OK. March 24, 2011

BovineSNP50 Assay



ARTICLES

SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries

Curtis P Van Tassell[1], Timothy P L Smith[2], Lakshmi K Matukumalli[1,3], Jeremy F Taylor[4], Robert D Schnabel[4], Cynthia Taylor Lawley[5], Christian D Haudenschild[5], Stephen S Moore[6], Wesley C Warren[7] & Tad S Sonstegard[1]

High-density single-nucleotide polymorphism (SNP) arrays have revolutionized the ability of genome-wide association studies to detect genomic regions harboring sequence variants that affect complex traits. Extensive numbers of validated SNPs with known allele frequencies are essential to construct genotyping assays with broad utility. We describe an economical, efficient, single-step method for SNP discovery, validation and characterization that uses deep sequencing of reduced representation libraries (RRLs) from specified target populations. Using nearly 50 million sequences generated on an Illumina Genome Analyzer from DNA of 66 cattle representing three populations, we identified 62,042 putative SNPs and predicted their allele frequencies. Genotype data for these 66 individuals validated 92% of 23,357 selected genome-wide SNPs, with a genotypic and sequence allele frequency correlation of $r = 0.67$. This approach for simultaneous *de novo* discovery of high-quality SNPs and population characterization of allele frequencies may be applied to any species with at least a partially sequenced genome.

RECEIVED 3 JANUARY; ACCEPTED 28 JANUARY; PUBLISHED ONLINE 24 FEBRUARY 2008; DOI:10

NATURE METHODS | VOL.5 NO.3 | MARCH 2008 | 247

March 2008 | volume 5 | number 3

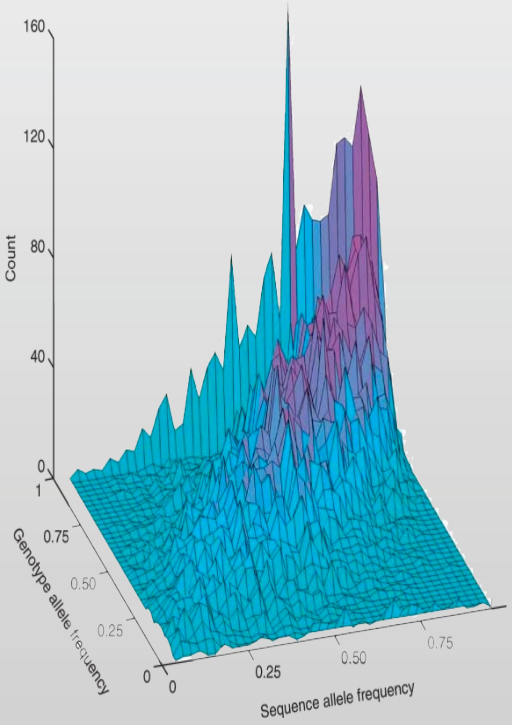
nature methods

www.nature.com/naturemethods

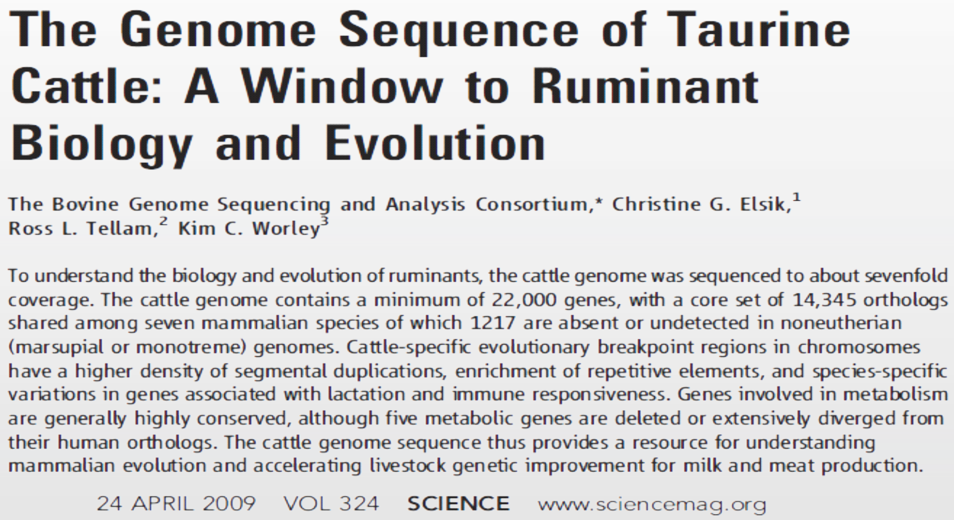
Techniques for life scientists and chemists

- Bovine SNP map
- Following stochastic tumor initiation
- High-content TF analysis
- A yeast genome library
- Resolving RNA dynamics with NMR

62,042 SNPs



Genome Sequenced...

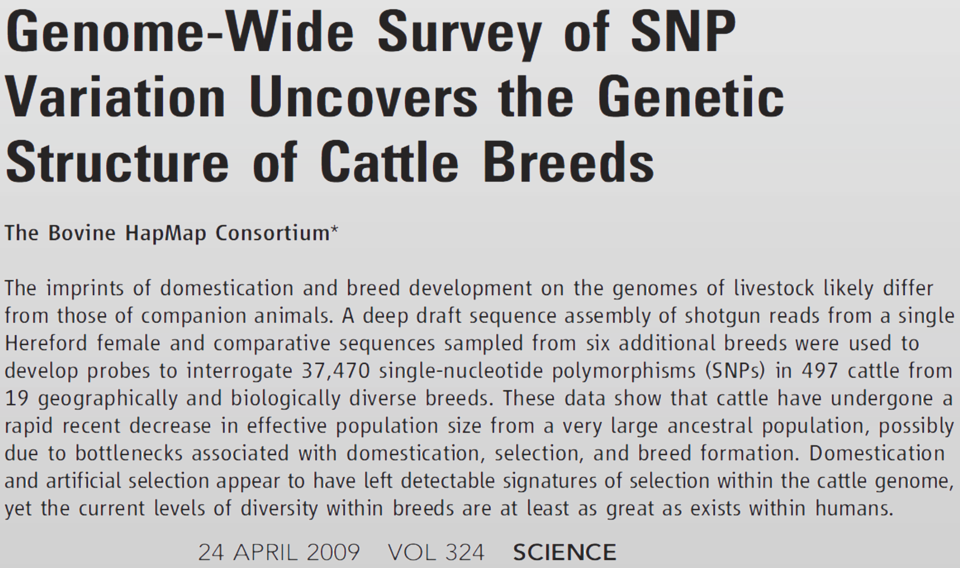


The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution

The Bovine Genome Sequencing and Analysis Consortium,* Christine G. Elsik,[1] Ross L. Tellam,[2] Kim C. Worley[3]

To understand the biology and evolution of ruminants, the cattle genome was sequenced to about sevenfold coverage. The cattle genome contains a minimum of 22,000 genes, with a core set of 14,345 orthologs shared among seven mammalian species of which 1217 are absent or undetected in noneutherian (marsupial or monotreme) genomes. Cattle-specific evolutionary breakpoint regions in chromosomes have a higher density of segmental duplications, enrichment of repetitive elements, and species-specific variations in genes associated with lactation and immune responsiveness. Genes involved in metabolism are generally highly conserved, although five metabolic genes are deleted or extensively diverged from their human orthologs. The cattle genome sequence thus provides a resource for understanding mammalian evolution and accelerating livestock genetic improvement for milk and meat production.

24 APRIL 2009 VOL 324 SCIENCE www.sciencemag.org



Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds

The Bovine HapMap Consortium*

The imprints of domestication and breed development on the genomes of livestock likely differ from those of companion animals. A deep draft sequence assembly of shotgun reads from a single Hereford female and comparative sequences sampled from six additional breeds were used to develop probes to interrogate 37,470 single-nucleotide polymorphisms (SNPs) in 497 cattle from 19 geographically and biologically diverse breeds. These data show that cattle have undergone a rapid recent decrease in effective population size from a very large ancestral population, possibly due to bottlenecks associated with domestication, selection, and breed formation. Domestication and artificial selection appear to have left detectable signatures of selection within the cattle genome, yet the current levels of diversity within breeds are at least as great as exists within humans.

24 APRIL 2009 VOL 324 SCIENCE

PNAS

Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics

Jared E. Decker[a], J. Chris Pires[b], Gavin C. Conant[a], Stephanie D. McKay[a], Michael P. Heaton[c], Kefei Chen[d], Alan Cooper[d], Johanna Vilkki[e], Christopher M. Seabury[f], Alexandre R. Caetano[g], Gary S. Johnson[h], Rick A. Brenneman[i], Olivier Hanotte[j,k], Lori S. Eggert[b], Pamela Wiener[l], Jong-Joo Kim[m], Kwan Suk Kim[n], Tad S. Sonstegard[o], Curt P. Van Tassell[o], Holly L. Neibergs[p], John C. McEwan[q], Rudiger Brauning[q], Luiz L. Coutinho[r], Masroor E. Babar[s], Gregory A. Wilson[t], Matthew C. McClure[a], Megan M. Rolf[a], JaeWoo Kim[a], Robert D. Schnabel[a], and Jeremy F. Taylor[a,1]

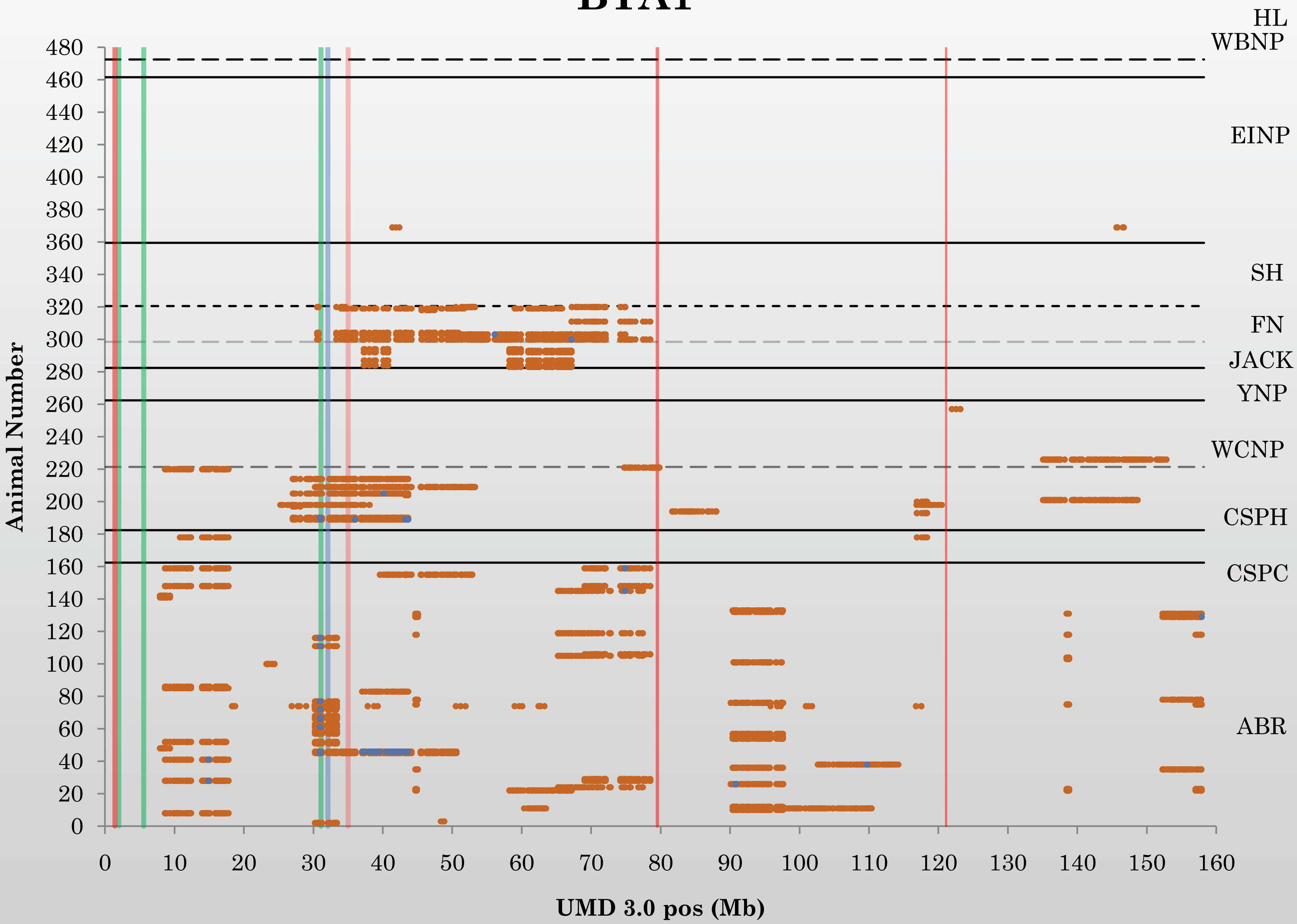
18644–18649 | PNAS | November 3, 2009 | vol. 106 | no. 44   www.pnas.org/cgi/doi/10.1073/pnas.0904691106

- 678 individuals from 61 Pecoran species
  - 48 breeds of cattle
- 40,843 SNPs… 39,695 parsimony informative characters

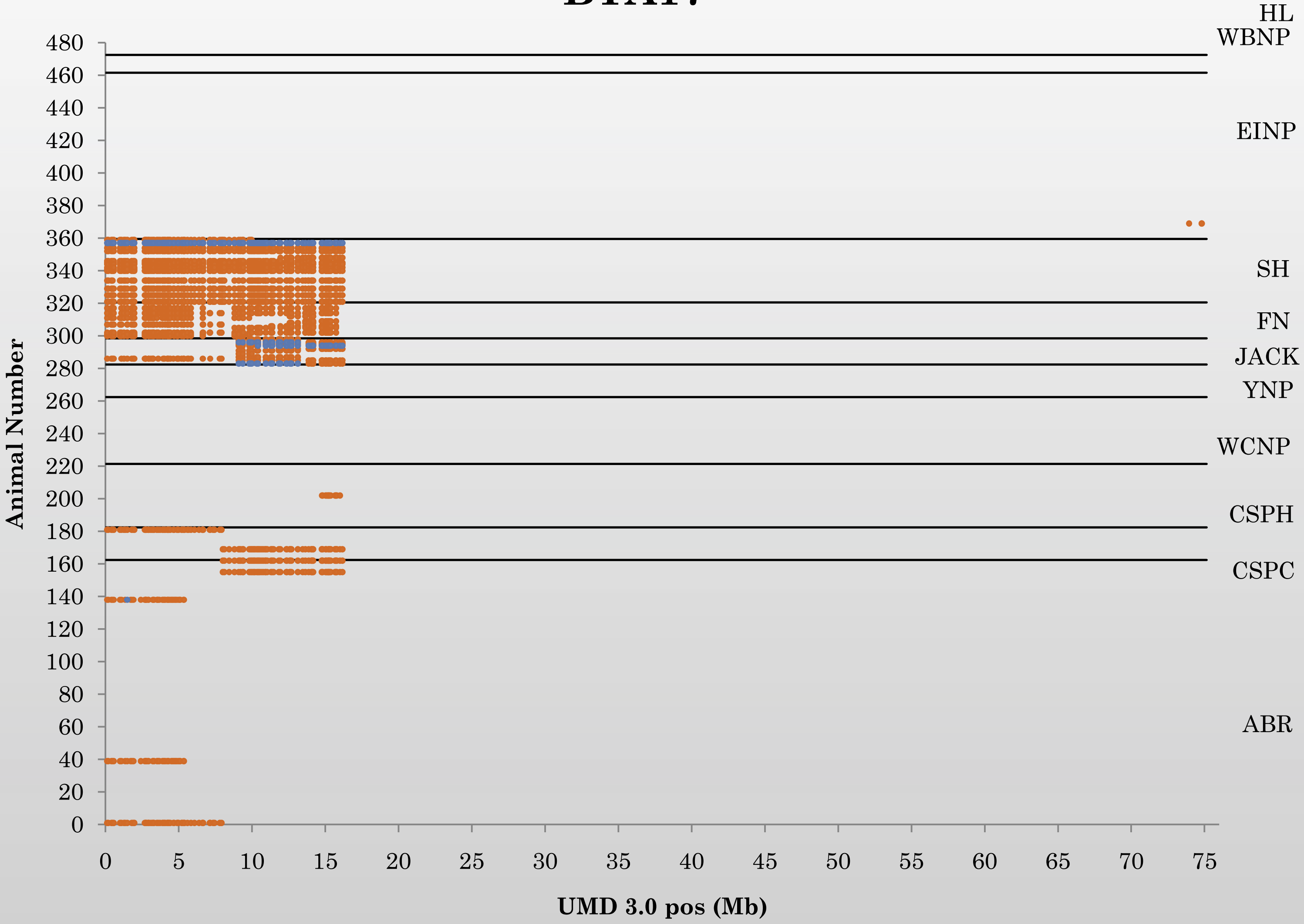
Populations

Breed Code	Breed/Population	N
AN	Angus	6,448
HFD	Hereford	1,521
LM	Limousin	2,604
CHA	Charolais	796
SIM	Simmental	824
TXLH	Texas longhorn	45
SH	Shorthorn	108
HO	Holstein	18,904
AYR	Finnish Ayrshire	516
JER	Jersey	93
BB	Arrowhead Ranch	169
CSP	Custer State Park	60
EINP	Elk Island National Park	106
HL	Hook Lake Wood Bison	11
WBNP	Wood Buffalo National Park	12
FNBB	Fort Niobara	27
SHBB	Sully's Hill	40
WCNP	Wind Cave National Park	44
YNP	Yellowstone National Park Bison	20
JACK	Jackson Bison	20
GAUR	Gaur	47
YAK	Yak	17
Bjav	Banteng	4
Bbub	Water buffalo	12
Scaf	Cape buffalo	8
Scan	Congo buffalo	3

BTA1

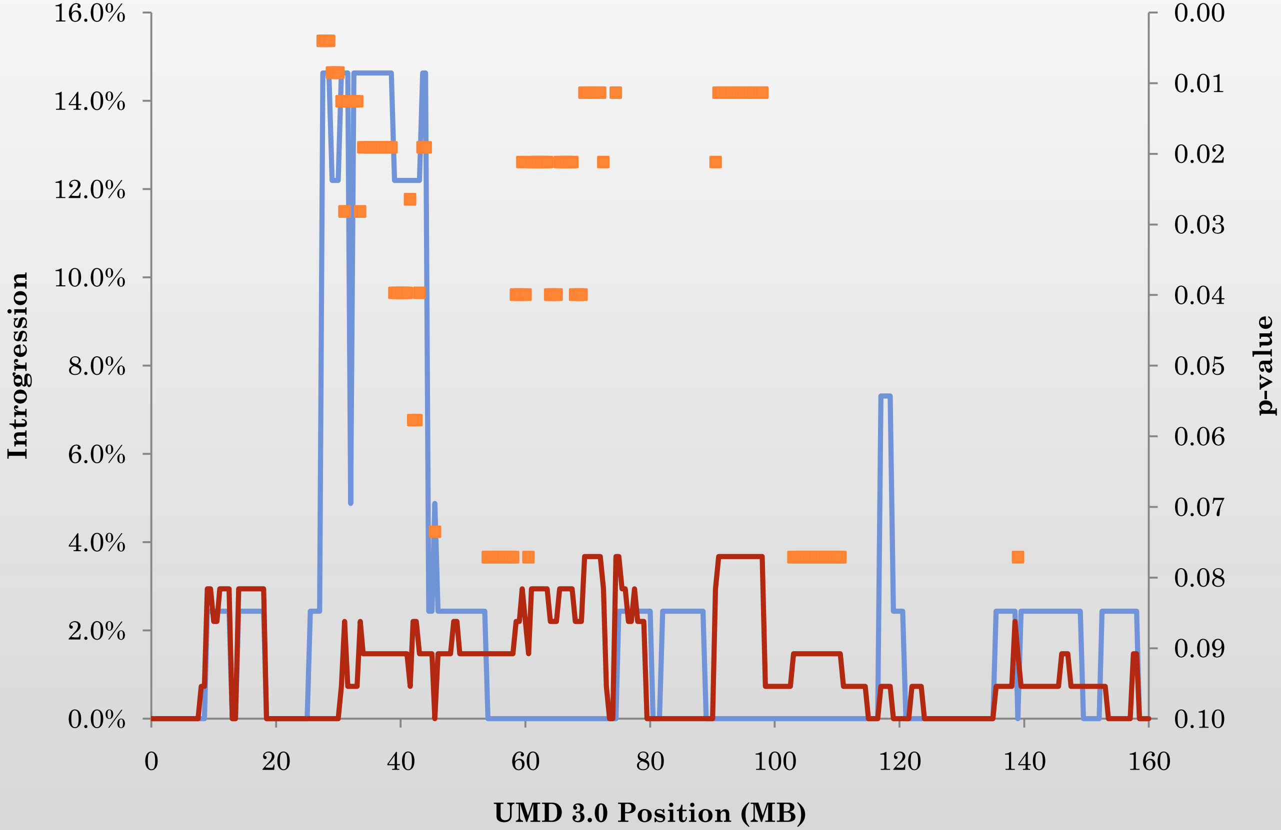

HL
WBNP
EINP
SH
FN
JACK
YNP
WCNP
CSPH
CSPC
ABR
Animal Number
480
460
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
UMD 3.0 pos (Mb)

BTA17

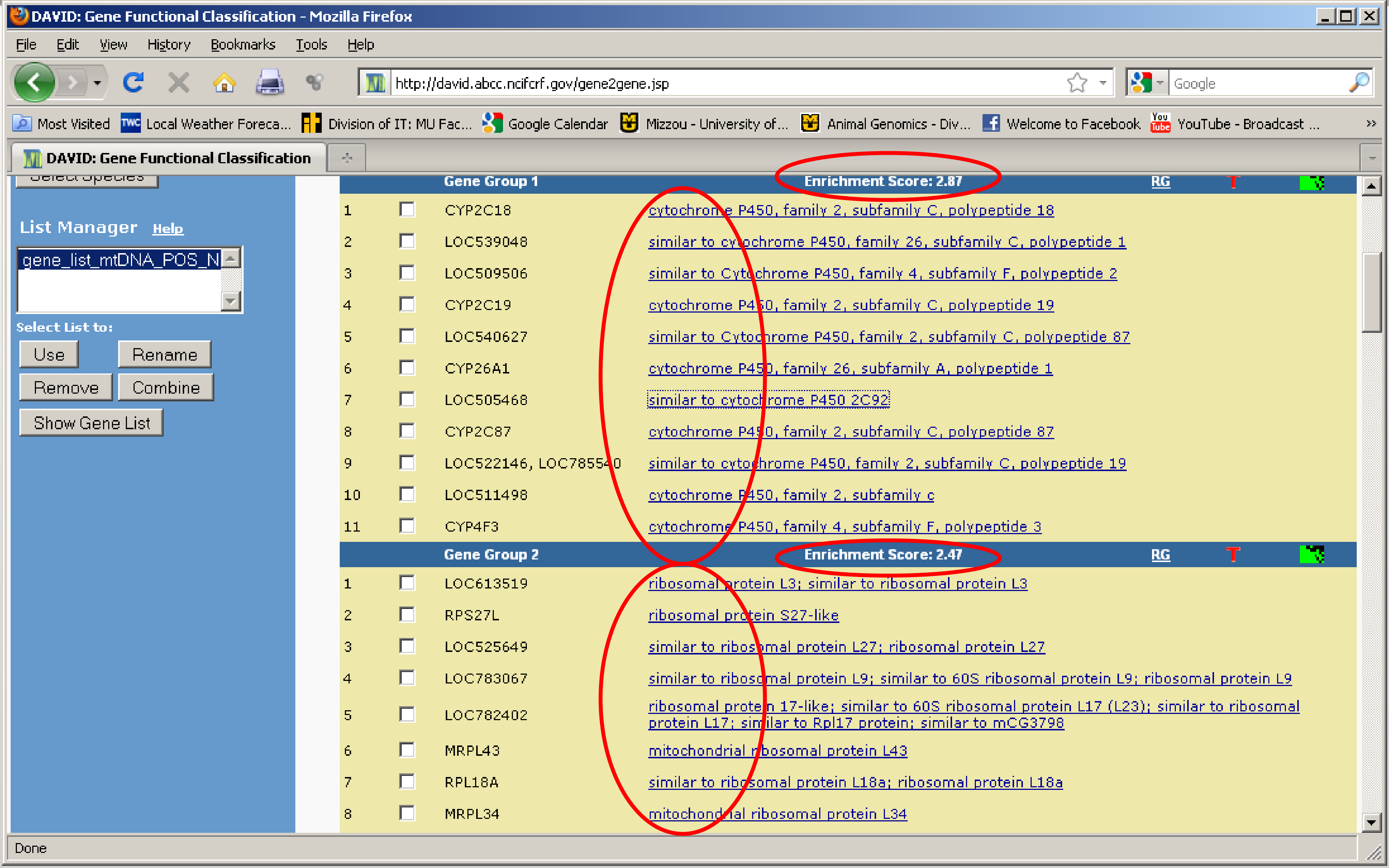


480
460
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
100
80
60
40
20
0
Animal Number
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
UMD 3.0 pos (Mb)
HL
WBNP
EINP
SH
FN
JACK
YNP
WCNP
CSPH
CSPC
ABR

BTA1

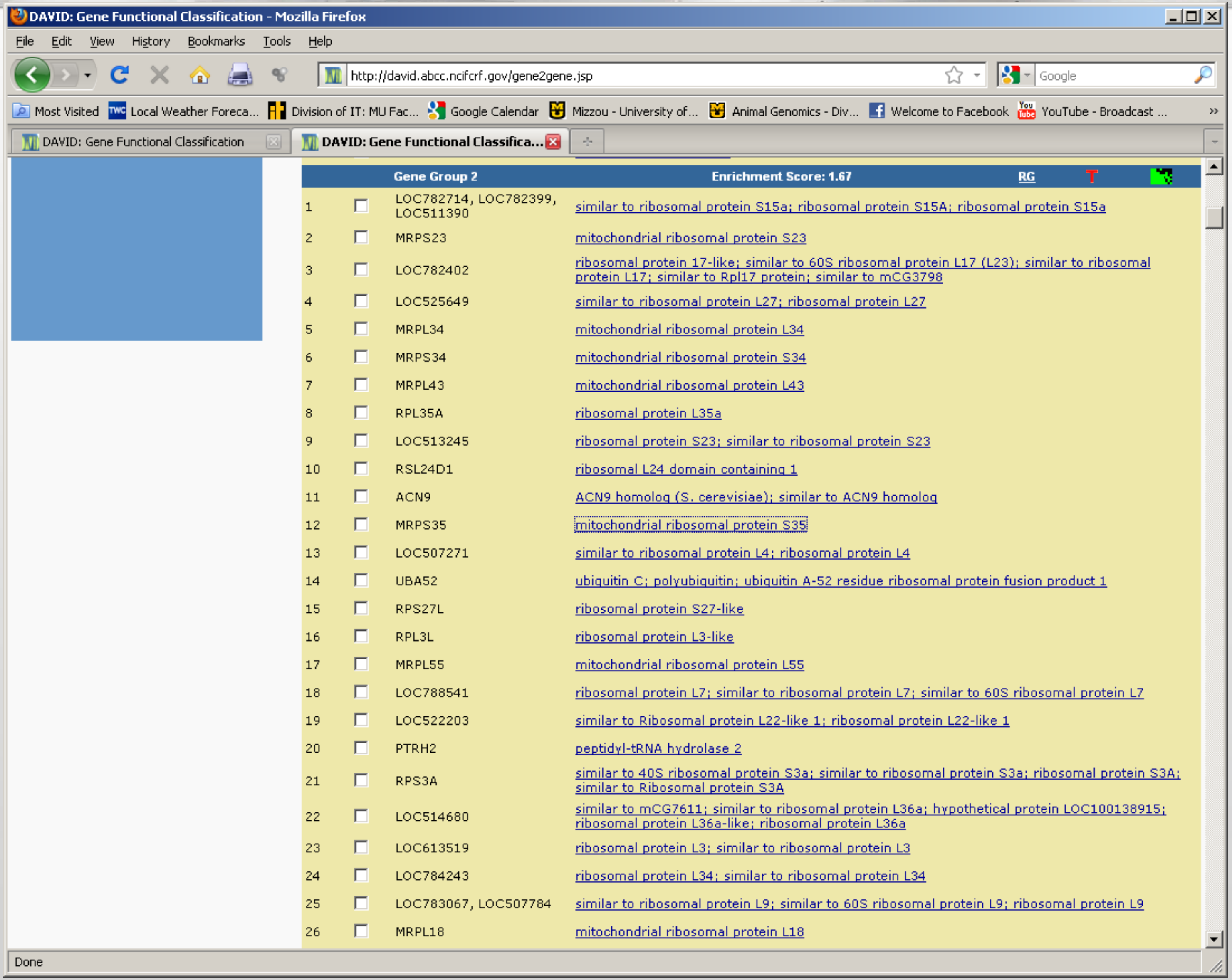

16.0%
14.0%
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
Introgression
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
p-value
0
20
40
60
80
100
120
140
160
UMD 3.0 Position (MB)

DAVID POS/NEG $p<0.01$



Gene Group 1			Enrichment Score: 2.87
1	☐	CYP2C18	cytochrome P450, family 2, subfamily C, polypeptide 18
2	☐	LOC539048	similar to cytochrome P450, family 26, subfamily C, polypeptide 1
3	☐	LOC509506	similar to Cytochrome P450, family 4, subfamily F, polypeptide 2
4	☐	CYP2C19	cytochrome P450, family 2, subfamily C, polypeptide 19
5	☐	LOC540627	similar to Cytochrome P450, family 2, subfamily C, polypeptide 87
6	☐	CYP26A1	cytochrome P450, family 26, subfamily A, polypeptide 1
7	☐	LOC505468	similar to cytochrome P450 2C92
8	☐	CYP2C87	cytochrome P450, family 2, subfamily C, polypeptide 87
9	☐	LOC522146, LOC785540	similar to cytochrome P450, family 2, subfamily C, polypeptide 19
10	☐	LOC511498	cytochrome P450, family 2, subfamily c
11	☐	CYP4F3	cytochrome P450, family 4, subfamily F, polypeptide 3
Gene Group 2			**Enrichment Score: 2.47**
1	☐	LOC613519	ribosomal protein L3; similar to ribosomal protein L3
2	☐	RPS27L	ribosomal protein S27-like
3	☐	LOC525649	similar to ribosomal protein L27; ribosomal protein L27
4	☐	LOC783067	similar to ribosomal protein L9; similar to 60S ribosomal protein L9; ribosomal protein L9
5	☐	LOC782402	ribosomal protein 17-like; similar to 60S ribosomal protein L17 (L23); similar to ribosomal protein L17; similar to Rpl17 protein; similar to mCG3798
6	☐	MRPL43	mitochondrial ribosomal protein L43
7	☐	RPL18A	similar to ribosomal protein L18a; ribosomal protein L18a
8	☐	MRPL34	mitochondrial ribosomal protein L34

DAVID Hybrid/Non-hybrid



	Gene Group 2	Enrichment Score: 1.67	RG	T
1 ☐	LOC782714, LOC782399, LOC511390	similar to ribosomal protein S15a; ribosomal protein S15A; ribosomal protein S15a		
2 ☐	MRPS23	mitochondrial ribosomal protein S23		
3 ☐	LOC782402	ribosomal protein 17-like; similar to 60S ribosomal protein L17 (L23); similar to ribosomal protein L17; similar to Rpl17 protein; similar to mCG3798		
4 ☐	LOC525649	similar to ribosomal protein L27; ribosomal protein L27		
5 ☐	MRPL34	mitochondrial ribosomal protein L34		
6 ☐	MRPS34	mitochondrial ribosomal protein S34		
7 ☐	MRPL43	mitochondrial ribosomal protein L43		
8 ☐	RPL35A	ribosomal protein L35a		
9 ☐	LOC513245	ribosomal protein S23; similar to ribosomal protein S23		
10 ☐	RSL24D1	ribosomal L24 domain containing 1		
11 ☐	ACN9	ACN9 homolog (S. cerevisiae); similar to ACN9 homolog		
12 ☐	MRPS35	mitochondrial ribosomal protein S35		
13 ☐	LOC507271	similar to ribosomal protein L4; ribosomal protein L4		
14 ☐	UBA52	ubiquitin C; polyubiquitin; ubiquitin A-52 residue ribosomal protein fusion product 1		
15 ☐	RPS27L	ribosomal protein S27-like		
16 ☐	RPL3L	ribosomal protein L3-like		
17 ☐	MRPL55	mitochondrial ribosomal protein L55		
18 ☐	LOC788541	ribosomal protein L7; similar to ribosomal protein L7; similar to 60S ribosomal protein L7		
19 ☐	LOC522203	similar to Ribosomal protein L22-like 1; ribosomal protein L22-like 1		
20 ☐	PTRH2	peptidyl-tRNA hydrolase 2		
21 ☐	RPS3A	similar to 40S ribosomal protein S3a; similar to ribosomal protein S3a; ribosomal protein S3A; similar to Ribosomal protein S3A		
22 ☐	LOC514680	similar to mCG7611; similar to ribosomal protein L36a; hypothetical protein LOC100138915; ribosomal protein L36a-like; ribosomal protein L36a		
23 ☐	LOC613519	ribosomal protein L3; similar to ribosomal protein L3		
24 ☐	LOC784243	ribosomal protein L34; similar to ribosomal protein L34		
25 ☐	LOC783067, LOC507784	similar to ribosomal protein L9; similar to 60S ribosomal protein L9; ribosomal protein L9		
26 ☐	MRPL18	mitochondrial ribosomal protein L18		

Summary

Pop	Non introgressed	Introgressed	Percent Introgression	Avg Block Size	Avg Num Blocks	Avg Percent Cow DNA
CSPC	0	21	100.0%	4,324,735	5.33	0.43%
JACK	0	7	100.0%	4,024,869	3.86	0.25%
CSPH	1	38	97.4%	5,892,443	5.89	0.58%
ABR	12	144	92.3%	4,322,273	5.03	0.42%
SHBB	4	35	89.7%	4,466,352	2.91	0.25%
FNBB	2	9	81.8%	4,258,448	4.78	0.37%
WCNP	28	74	72.5%	4,412,274	2.53	0.19%
WBNP	10	1	9.1%	2,338,716	1.00	0.04%
EINP	74	3	3.9%	3,115,064	1.33	0.06%
HL	5	0	0.0%	0	0	0
YNP	19	0	0.0%	0	0	0

There is hope…

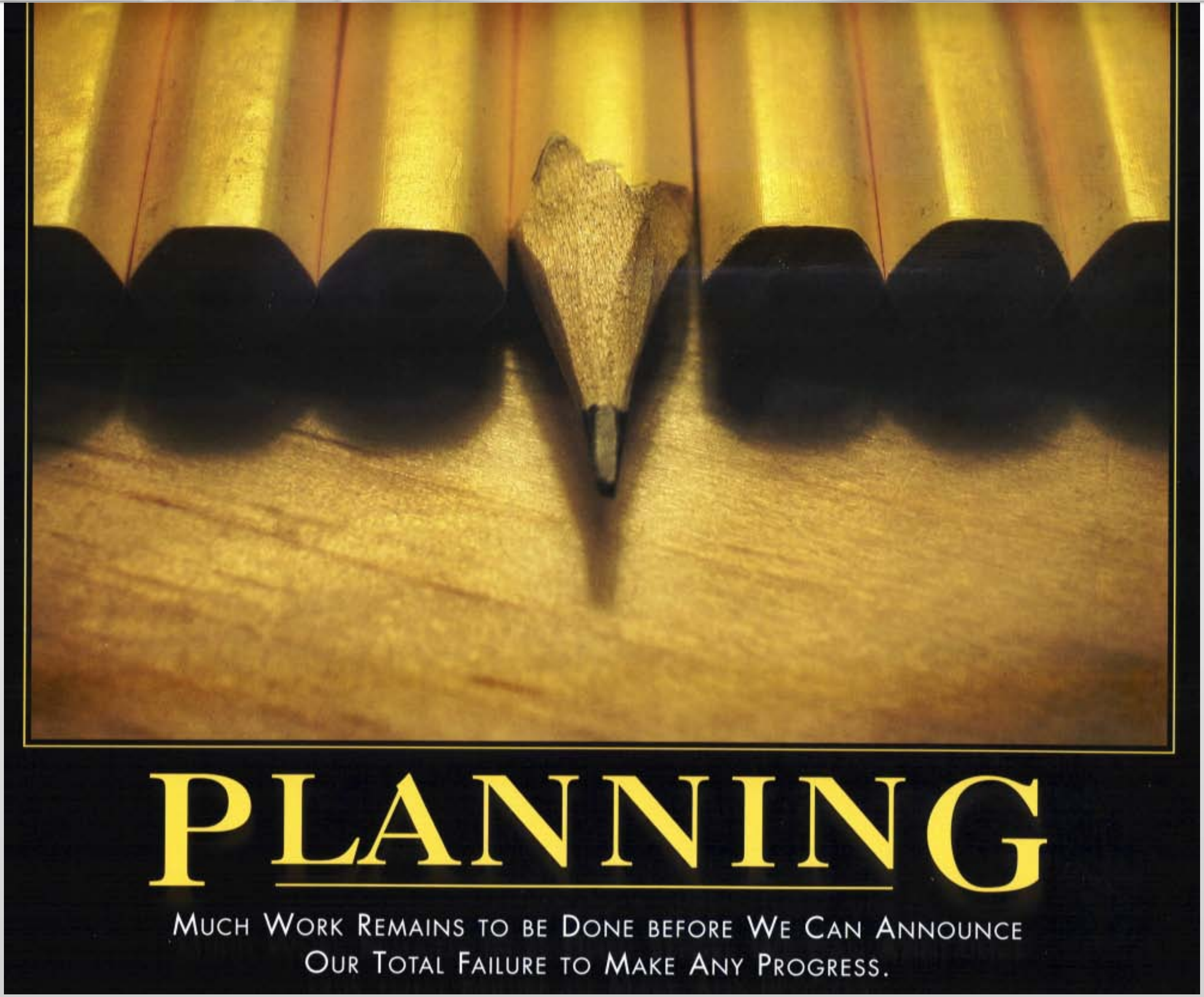
sample_id	reg	Num SNPs	Num Blocks	AVG Block Size (bp)	Total Introgressed Bp	parent1	parent2	Sire Num SNPs	Sire Num Blocks	Sire Percent Introgression	Dam Num SNPs	Dam Num Blocks	Dam Percent Introgression
400200870	8027		**0**			400024060	400200790	191	4	0.63%	214	8	0.72%
400202380	8231		**0**			400200830	400200750	9	1	0.02%	113	4	0.36%
400202390	8232		**0**			400200830	400200770	9	1	0.02%	192	6	0.65%
400202200	8213		**0**			400201280	400200700	70	3	0.23%	113	2	0.28%
400201550	8139		**0**			400201280	400201070	70	3	0.23%	125	6	0.35%

CYTB V98A

CSP
CYTB V98A
46 V, 36 A (44%)

N	Species	Location	Age
1	Pre-taurus	Yenisei, South Russia	12300 +/-40 yrs BP
1	Bison priscus	Alyoshkina Zaimka, Siberia, Russia	20k yrs BP
2	Bison priscus	Yukon, Canada	20k yrs BP
2	Bis. Bonasus	France	few hundred yrs BP
1	Bovid species X	Mezmaiskaya, North caucasus, Russia	>59400 yrs BP
1	Bovid species X	Rasik 1, Urals, Russia	12565 +/-55 yrs BP
1	Pre-taurus	Harbin, Northeast China	8260 +/-25 yrs BP
1	Pre-taurus	Yenisei, South Russia	12300 +/-40 yrs BP
1	Bubalus Bubalis	Yunnan, China	9k yrs BP

Questions?


PLANNING
Much Work Remains to be Done before We Can Announce
Our Total Failure to Make Any Progress.

Acknowledgements

Missouri
- Jerry Taylor
- Stephanie McKay (genotyping queen)
- Matt McClure
- Rich Chapple (*CYTB*)
- Jared Decker (phylogenomics)
- Megan Rolf (genomic selection)

BARC
- Curt Van Tassell
- Lakshmi Matukumalli
- Tad Sonstegard

USMARC
- Tim Smith
- Mark Allan

AIPL
- George Wiggans
- Paul VanRaden

Illumina
- Cindy Lawley

Affymetrix
- Chris Davies
- Robert Barkovich

International
- JongJoo Kim, S. Korea
- Andy Law, Alan Archibald Roslin Institute UK
- Harry Noyes Liverpool UK
- Numerous groups submitted DNA

Funding Agencies

USDA/NRI/CSREES
- 2006-35205-16888
- 2006-35616-16697
- 2006-35205-16701
- 2008-35205-18864
- 2008-35205-04687

USDA/ARS
- 1265-31000-081D
- 1265-31000-090-00D
- 5438-31000-073D

Merial
- Stewart Bauck

NAAB
- Godon Doak
- ABS Global
- Accelerated Genetics
- Alta Genetics
- CRI/Genex
- Select Sires
- Semex Alliance
- Taurus Service