IRISH CATTLE BREEDING FEDERATION

International Genomic Partnership (IGenoP) An Irish Implementation

By Karl O'Connell

NDP National Development Plan Transforming Ireland

ICAR2012

Background

- Irish requirement for facility to manage, storage and share both national and international genotype information.
- Initial proof of concept, consultation with other IGenoP members and then prototype built.
- Developed a standard for all input & output files.

Basic Genotype Upload Process

- Genotype files are uploaded to secure ftp server
- Automated process:
  - Un-compress and inspect files
  - database utility loads file to staging table
  - Identification, and quality assurance checks are performed.
    - Incoming animal number replaced with Interbull number
  - Valid entries are prepared for loading to Genotype database
  - Invalid entries are flagged and reports

Genotype Loading (cont)

- Individual Genotype files created, compressed and load as BLOB
- Genotype details loaded to database table
- Parentage SNP's extracted
- Call Rates Checked
- Parentage verification performed where possible
- SNP_Map checked and logged where available

File	Delimiter	Default File Naming Convention
Genotype/Final Report File	Tab	ContributorCode_final_report_01052012_Chip.txt
Animal Summary File	Comma	ContributorCode_dna_summary_01052012_Chip.csv
Sentrix/Summary File	Tab	ContributorCode_summary_01052012_Chip.txt
SNP_Map File	Tab	ContributorCode_snp_map_01052012_Chip.txt

Genotype Files

```
[Header]
GSGT Version    1.9.4
Processing Date 10/13/2011 11:03 AM
Content         Bovine3K OPA Manifest GS0012180-OPA.opa
Num SNPs        2900
Total SNPs      2900
Num Samples     40
Total Samples   41
[Data]
SNP Name        Sample ID       Allele1 - Forward       Allele2 - Forward       Allele1 - AB    Allele2 - AB    Allele1 - Top   Allele2 - Top   GC Score
ARS-BFGL-NGS-116814     HOLIRLF211085292048     G       G       B       B       C       C       0.7563  0.057   0.570
ARS-BFGL-NGS-119107     HOLIRLF211085292048     T       T       A       A       A       A       0.5499  1.357   0.030
UA-IFASA-1397   HOLIRLF211085292048     C       C       B       B       G       G       0.9279  0.022   0.274
ARS-USMARC-Parent-DQ647186-rs29014143   HOLIRLF211085292048     G       G       B       B       G       G       0.8282  0.055   0.354
UA-IFASA-8520   HOLIRLF211085292048     A       T       A       B       A       T       0.8009  0.482   0.331
ARS-BFGL-NGS-13878      HOLIRLF211085292048     A       G       A       B       A       G       0.8558  0.870   0.688
ARS-BFGL-NGS-32909      HOLIRLF211085292048     T       G       A       B       A       C       0.8639  0.346   0.430
BTB-00581295    HOLIRLF211085292048     T       T       A       A       A       A       0.8583  1.127   0.193
ARS-BFGL-NGS-57054      HOLIRLF211085292048     C       C       B       B       C       C       0.6651  0.207   0.854
Hapmap48689-BTA-23260   HOLIRLF211085292048     A       A       A       A       A       A       0.6183  1.113   0.005
```

	A	B	C	D	E	F	G	H	I	J	K	L
1	DNA Report on P:\Illumina Project Data TEMP!!!!\Irish 30jun2010\Irish Bovine 30jun2010\Irish Bovine 30jun2010_DNAReport.csv											
2		# LOCI = 54001	# DNAs = 14	ProjectName	GenCall Vers	Low GenCall Score Cutoff = 0.1500						
3	Row	DNA_Name	#No_Calls	#Calls	Call_Rate	A/A_Freq	A/B_Freq	B/B_Freq	Minor_Freq	50%_GC_Score	10%_GC_Score	0/1
4	1	IE211110641955	696	53305	0.9871	0.3092	0.2906	0.4002	0.4545	0.8494	0.4657	1
5	2	IE141360920741	516	53485	0.9904	0.3105	0.2947	0.3948	0.4578	0.85	0.4658	1
6	3	IE351066180935	297	53704	0.9945	0.3071	0.3059	0.3869	0.4601	0.8502	0.4658	1
7	4	IE371405790646	239	53762	0.9956	0.3091	0.3014	0.3894	0.4599	0.8503	0.4658	1

Index	Name	ID	Gender	Plate	Well	Group	Parent1	Parent2	Replicate	SentrixPosition
2	10511	HOLIRLM331504610762			Unknown	WG0097290		A02	Chip1	7757461035_R06C01
3	10512	HOLIRLM151623492558			Unknown	WG0097290		A03	Chip1	7757461035_R09C02
4	10513	HOLIRLM151549236627			Unknown	WG0097290		A04	Chip2	7757461038_R01C01
5	10514	HOLIRLM191613230849			Unknown	WG0097290		A05	Chip2	7757461038_R06C01
6	10515	HOLIRLM301441791396			Unknown	WG0097290		A06	Chip2	7757461038_R09C02
7	10516	HOLIRLM351307391146			Unknown	WG0097290		A07	Chip3	7757461030_R01C01
8	10517	HOLIRLM141893671985			Unknown	WG0097290		A08	Chip3	7757461030_R06C01
9	10518	HOLIRLM151549266662			Unknown	WG0097290		A09	Chip3	7757461030_R09C02
10	10519	HOLIRLM151006021829			Unknown	WG0097290		B01	Chip1	7757461035_R03C01

ICBF Database (cont)

Breakdown by Chip Type of Genotypes Loaded

Chip Type	Genotypes Loaded
3K	2,384
LD	7,384
50K	7,428
HD	2,606

Breakdown by Database table of Genotypes Loaded

Database Table	Total Records
Summary	21,205
Genotypes	21,129
Parentage	2,711,044
SNP Data	2,298,622,257

Typical File Size of Genotypes Received

File	96 LD Chips	48 50K Chips	48 HD Chips
Final Report File	45MB	170MB	3.0GB
Summary File	8KB	5KB	5KB
Sample Map File	7KB	4KB	4KB
SNP_Map File	410KB	3MB	46MB

ICBF Genotype Database

Chip Type	Perl Script & Oracle SQL Loader	Oracle External Tables & PL/SQL
96 LD	1.6 – Hours	5-10 – Minutes
48 50K	2.5 – Hours	5-10 – Minutes
48 HD	48 – Hours	2 – Hours

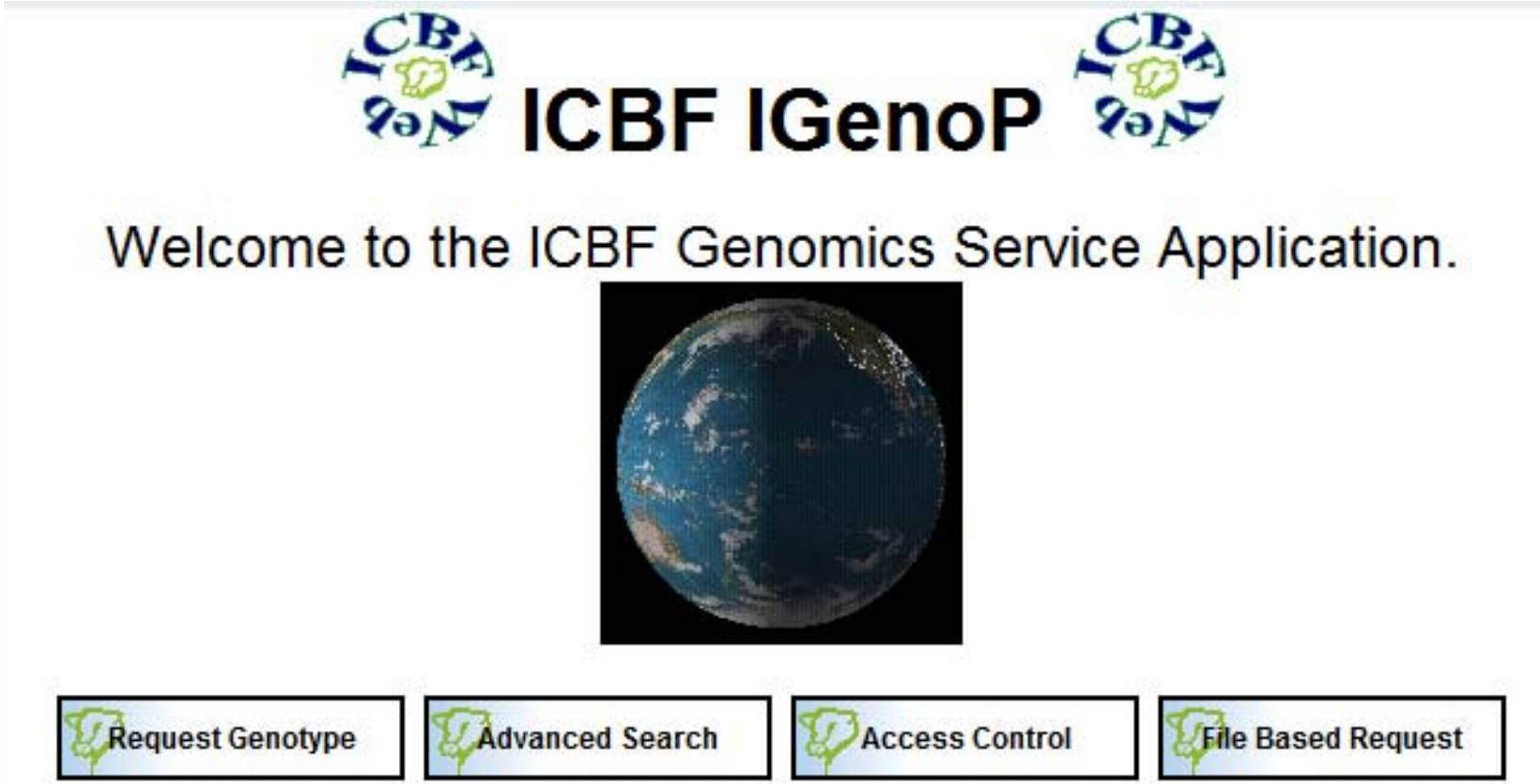
- Perl
  - Red Hat 6. 64bit CPU: 2 * Intel Xeon E5640 Quad Core - 64GB RAM
- Oracle
  - Oracle 10g R2 Enterprise Edition using a HP6100 SAN with 1TB storage allocated
  - Server HP rp5470 CPU – 2 * HP PA-RISC-8700 650MHz CPUs - 8GB RAM

ICBF Web Portal

- Login

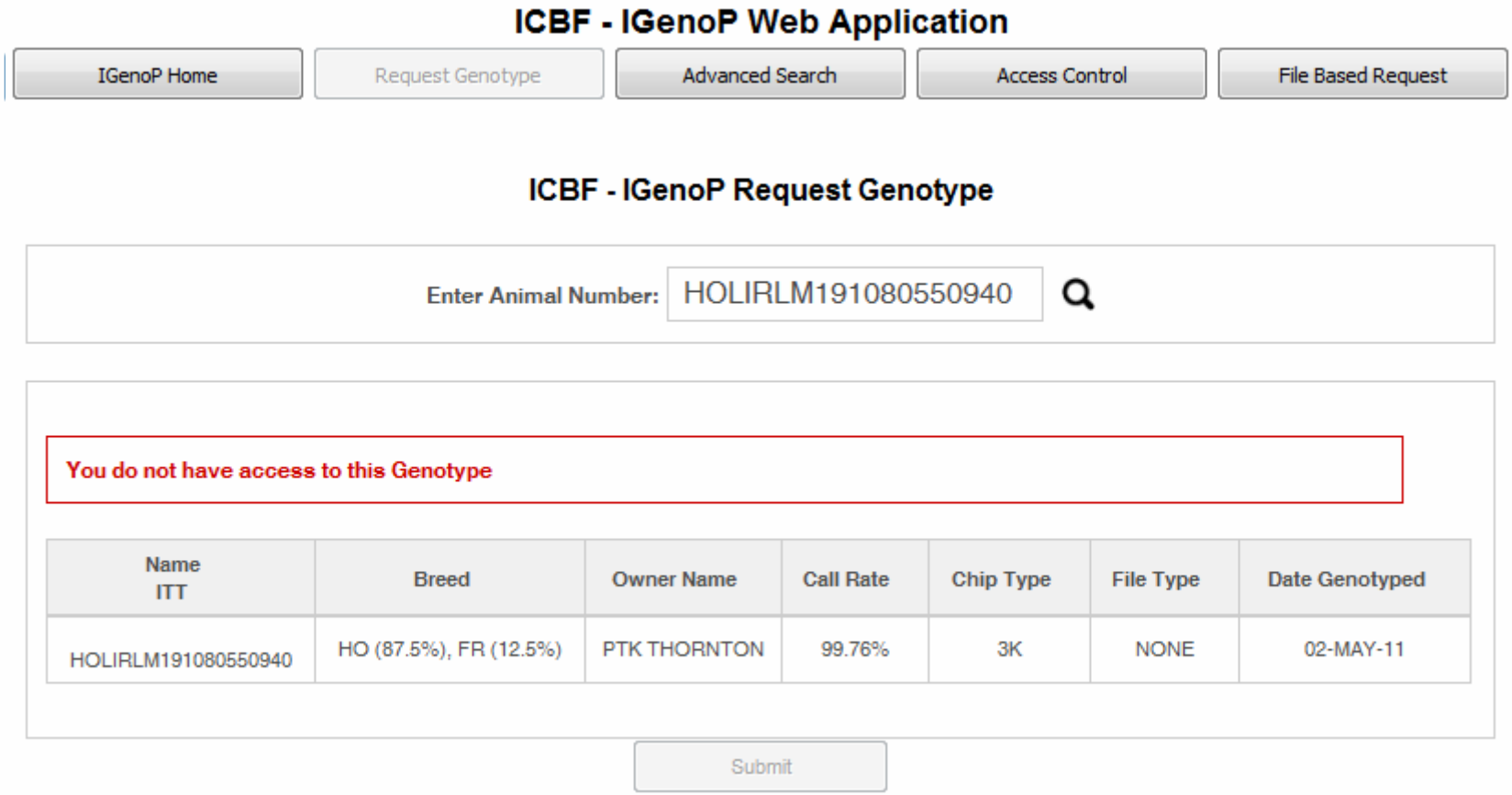
ICBF Web Portal (cont)

- Home Page

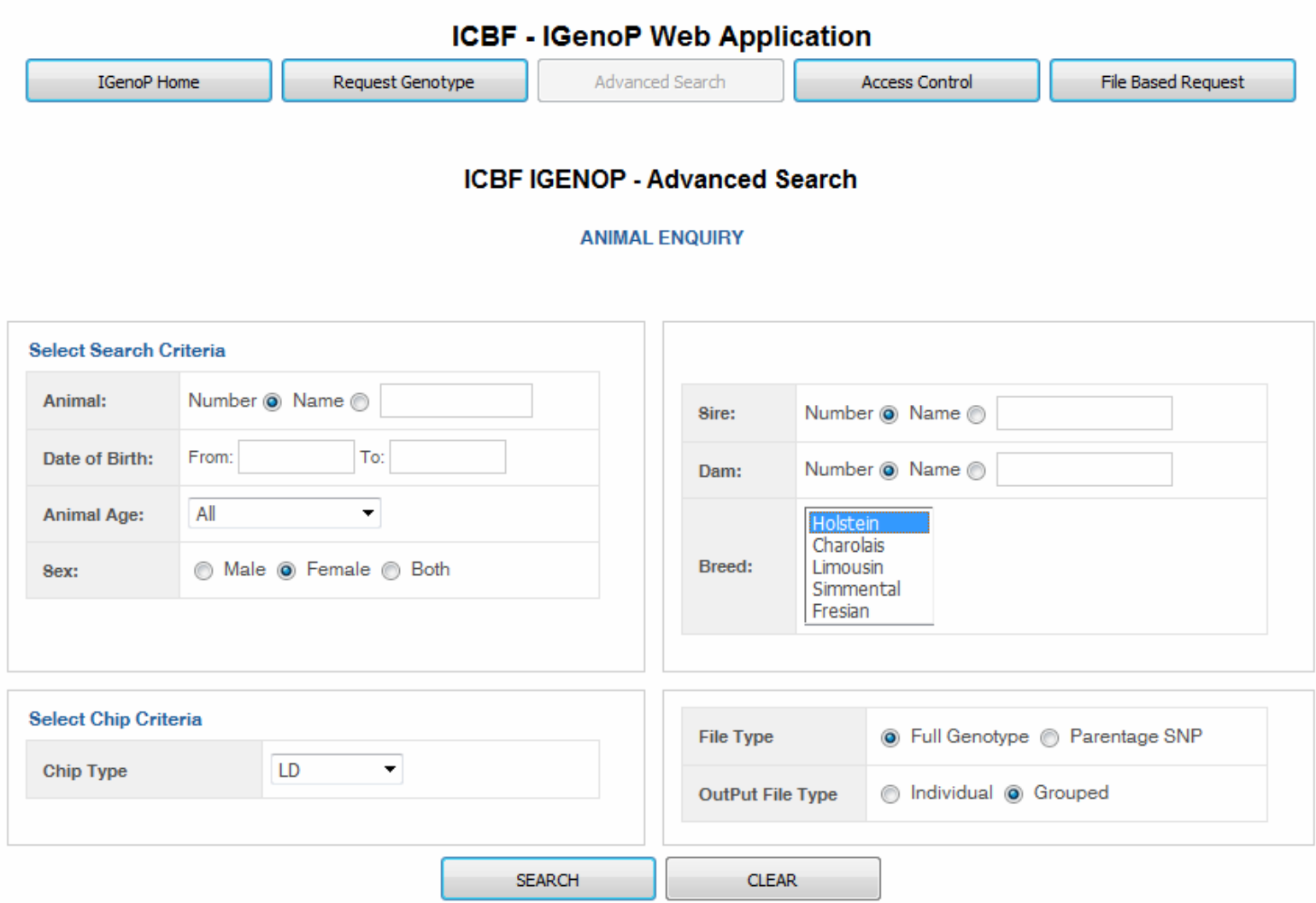


ICBF Web Portal (cont)

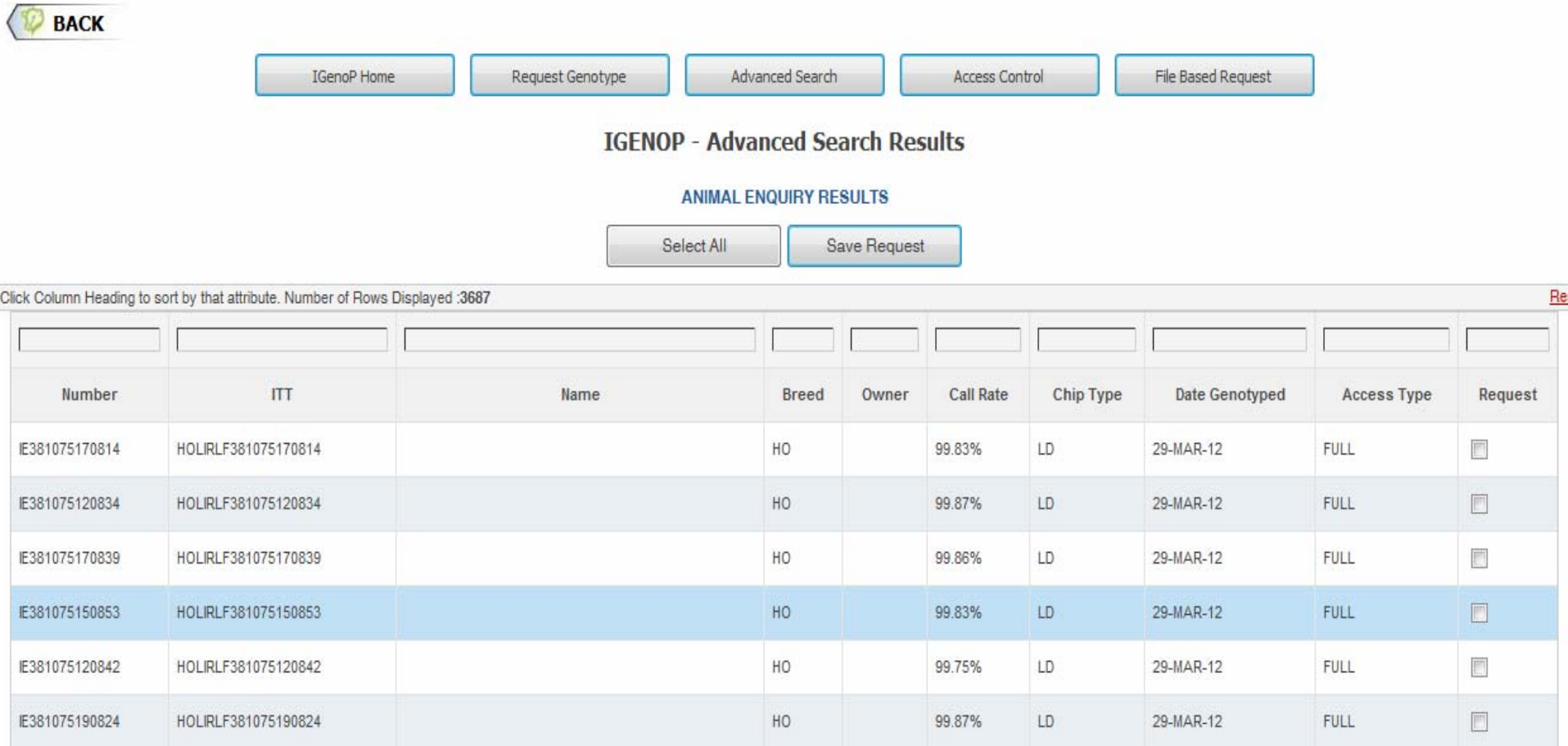
- Individual Animal Search



ICBF Web Portal (cont)



ICBF Web Portal (cont)



Number	ITT	Name	Breed	Owner	Call Rate	Chip Type	Date Genotyped	Access Type	Request
IE381075170814	HOLIRLF381075170814		HO		99.83%	LD	29-MAR-12	FULL	
IE381075120834	HOLIRLF381075120834		HO		99.87%	LD	29-MAR-12	FULL	
IE381075170839	HOLIRLF381075170839		HO		99.86%	LD	29-MAR-12	FULL	
IE381075150853	HOLIRLF381075150853		HO		99.83%	LD	29-MAR-12	FULL	
IE381075120842	HOLIRLF381075120842		HO		99.75%	LD	29-MAR-12	FULL	
IE381075190824	HOLIRLF381075190824		HO		99.87%	LD	29-MAR-12	FULL	

ICBF Web Portal (cont)

- File Upload Request

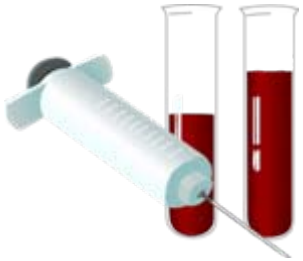
Genotype Extraction

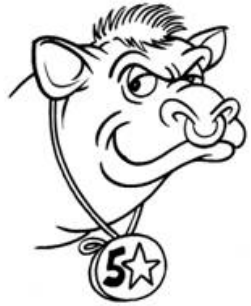
- Extract via Oracle Procedure to secure ftp directory
- For all requested extracts permissions are checked.
  - Genotype File
  - Summary File
  - Parentage File

Genotype Extraction (cont)

- Extraction Summary

Genotype Requested	Extraction Time
Full Genotype - 100 LD Chips	10 Seconds
Full Genotype - 100 50K Chips	20 Seconds
Full Genotype - 10 HD Chips	35 Seconds
Parentage SNP's - 100 LD Chips	5 Seconds





Thank you
&
Questions ?