Moore's Law, Genomics and the Future of Animal Breeding

David MacHugh, Animal Genomics Lab., UCD
www.animalgenomics.ucd.ie/dmachugh.php

Chris Creevey, Teagasc Animal Bioscience Centre
www.agresearch.teagasc.ie/animalbioscience/people/ccreavy.asp

Moore's Law

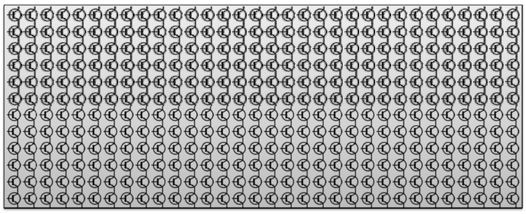
=

Cray supercomputer
1979

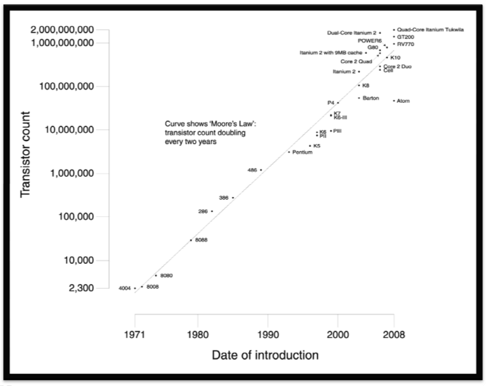
Apple iPhone
2009

Moore's Law

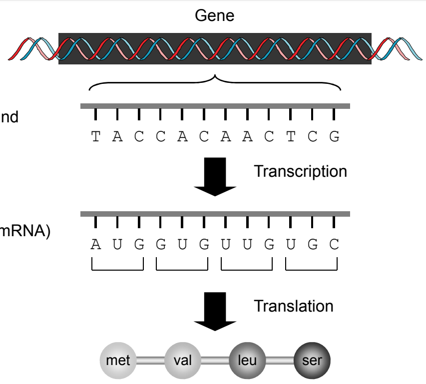
384 transistors....



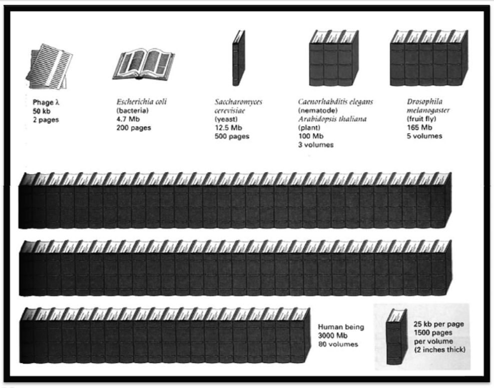
Moore's Law and CPU transistor counts: 1971-2008 doubling time ~ 24 months



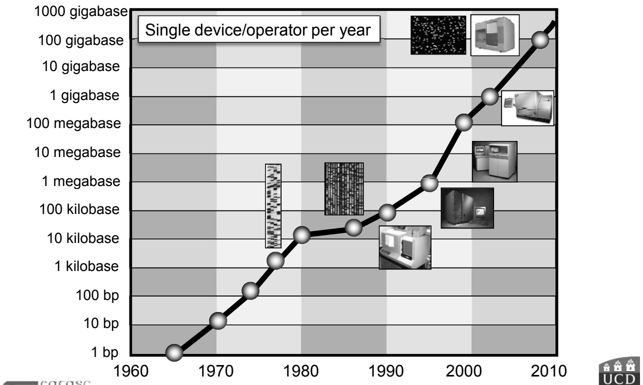
The flow of genetic information: gene expression



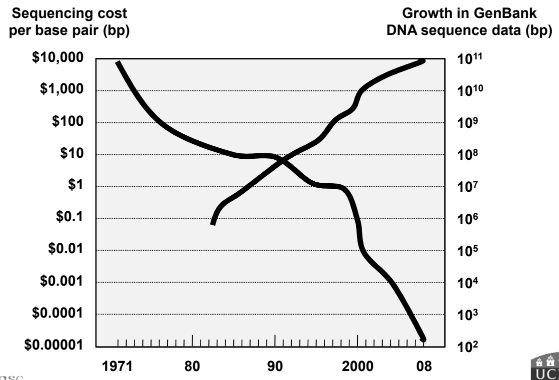
Genome size for various organisms



Moore's Law for genomics: DNA sequencing capacity doubling time ~15 months



DNA sequencing costs and the growth of DNA sequence data



Modified from *The Economist* (April 16th, 2009)

High-throughput genome sequencers at the UCD Conway Institute

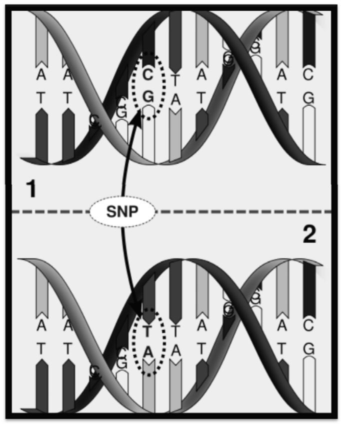
3rd generation DNA sequencing technologies: maintaining Moore's Law for genomics

- **In development – first commercial platforms expected between 2011-2013.**
  - **Pacific Biosciences.**
  - **Helicos Biosciences.**
  - **Oxford Nanopore Technologies.**
- **Driven by growth in personalised medicine and 'recreational' genomics.**
  - **Accurate mammalian genome readout in less than an hour**
  - **Complete genome sequence for a few hundred euros (or less).**
  - **Massive potential for scaling through miniaturisation and parallelization.**

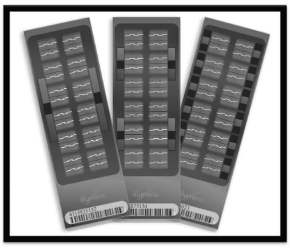
Oxford Nanopore Technologies

www.nanoporetech.com

A single nucleotide polymorphism (SNP)



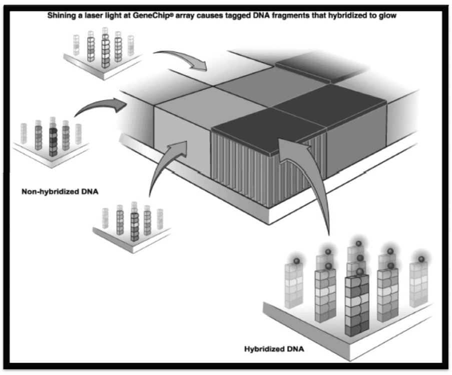
SNP genotyping platforms



Illumina® BeadChip®

Affymetrix® GeneChip®

GeneChip® SNP analysis is based on hybridisation of different SNP alleles and detection using fluorescence



An example of a simple 25-marker SNP-chip

	A	B	C	D	E
1	SNP 01	SNP 02	SNP 03	SNP 04	SNP 05
2	SNP 06	SNP 07	SNP 08	SNP 09	SNP 10
3	SNP 11	SNP 12	SNP 13	SNP 14	SNP 15
4	SNP 16	SNP 17	SNP 18	SNP 19	SNP 20
5	SNP 21	SNP 22	SNP 23	SNP 24	SNP 25

	A		B		C		D		E	
1	A	G	A	G	C	T	G	T	A	G
2	C	T	A	G	A	G	C	T	A	C
3	A	T	C	T	A	G	A	G	C	T
4	A	G	C	T	C	T	C	T	A	G
5	C	G	A	G	A	G	C	T	C	T

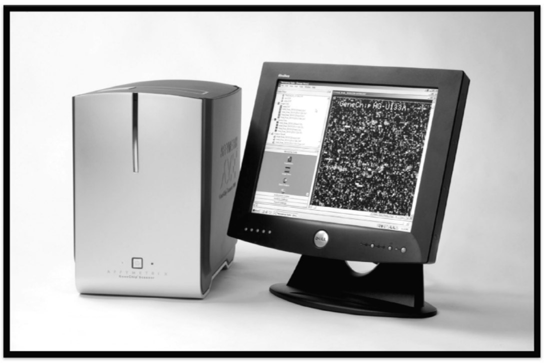
Animal 1

	A		B		C		D		E	
1	A	G	A	G	C	T	G	T	A	G
2	C	T	A	G	A	G	C	T	A	C
3	A	T	C	T	A	G	A	G	C	T
4	A	G	C	T	C	T	C	T	A	G
5	C	G	A	G	A	G	C	T	C	T

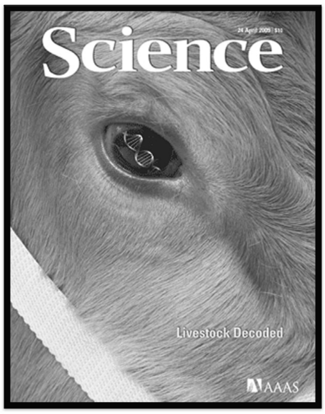
Animal 2

	A		B		C		D		E	
1	A	G	A	G	C	T	G	T	A	G
2	C	T	A	G	A	G	C	T	A	C
3	A	T	C	T	A	G	A	G	C	T
4	A	G	C	T	C	T	C	T	A	G
5	C	G	A	G	A	G	C	T	C	T

Thousands of individual SNPs can be analysed across the bovine genome



Bovine genome sequence and genome diversity: *Science,* vol 324, 24th April 2009



Bovine genome sequence

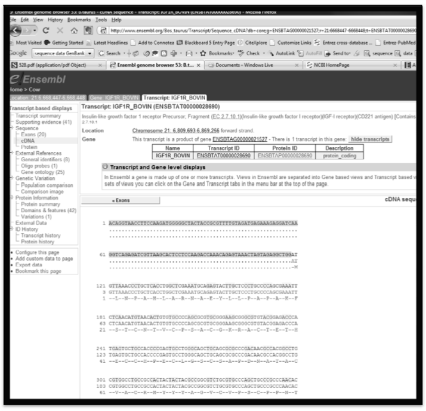
The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution

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

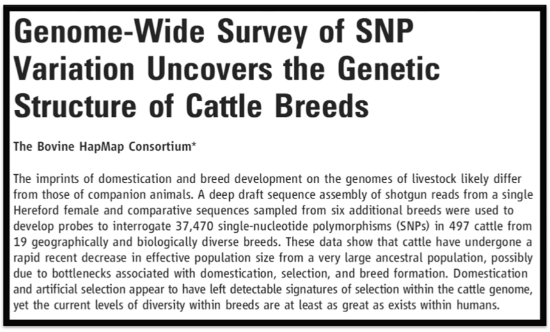
To understand the biology and evolution of ruminants, the cattle genome was sequenced to about sevenfold coverage. The cattle genome contains a minimum of 22,000 genes, with a core set of 14,345 orthologs shared among seven mammalian species of which 1217 are absent or undetected in noneutherian (marsupial 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.

Science, 24th April 2009, vol. **324**, 522-528.

Bovine genome, genes and genome variation can be browsed and analysed



Bovine genome diversity



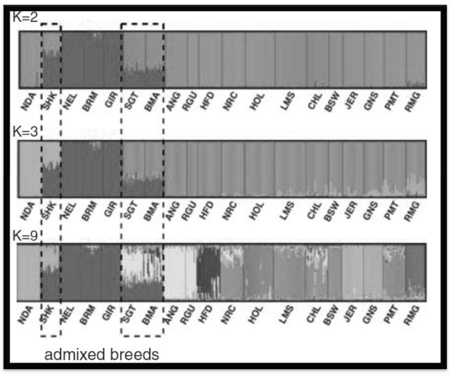
Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds

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.

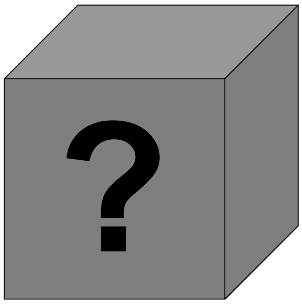
Science, 24th April 2009, vol. **324**, 528-532.

Discrete breed genetic structure revealed using SNP chip data



Science, 24th April 2009, vol. **324**, 528-532.

The 'grey box' of genomic selection: towards genomic selection incorporating genome content



Implications and opportunities for animal breeding and genetic improvement of cattle

- **Currently, genomic selection is 'partial'.**
  - **Information derived from a subset of genomic variation**
  - **Genes and cellular networks underlying favourable genomic segments are ignored – the 'grey box'**
- **Sequencing costs will reach the point where it is feasible to sequence individual animals routinely (~20X, ~120 Gb).**
  - **'*Genometypes*' will emerge as fundamental units for genomic selection 'version 2.0'**
  - **Significant bottleneck is bioinformatics and computing infrastructure required for data handling**
- **Genomic analyses of economic trait loci (ETLs) will accelerate to identify the multiple causative genes and molecular genetic variation underlying ETLs.**

Implications and opportunities for animal breeding and genetic improvement of cattle

- **Information from basic functional and structural genomics research will inform genomic selection:**
  - **Lactation biology (complex suite of lactation genes in cattle)**
  - **Developmental biology (*e.g.* analyses of growth genes)**
  - **Systems biology (*i.e.* cellular pathways and networks)**
  - **Immunobiology (*e.g.* disease resistance genes/alleles)**
  - **Epigenomics (*e.g.* analyses of imprinted genes)**
- **Genomic selection will become more precise:**
  - **Environmentally-sensitive traits (*e.g.* methane production)**
  - **Behavioural traits (*e.g.* docility and pain sensitivity)**
  - **Nutritional traits (*e.g.* milk and meat composition)**
  - **Animal health traits (*e.g.* disease resistance)**
  - **Genomic information for crossbreeding and heterosis**

Implications and opportunities for animal breeding and genetic improvement of cattle

- Genomic ‘health’ of breeds and populations can be tracked and monitored in great detail:
  - Inbreeding avoidance – reduce genomic similarity between sires and dams (whole genome and specific genomic regions, *e.g.* MHC – immune genes)
  - Single gene genetic disorders arising from intense artificial selection for production traits will be erasable from populations immediately (*e.g.* BLAD, CVM, DUMPS)
  - Genetic architecture and cellular networks underlying complex genetic disorders will be revealed (relevant to human biomedicine)
- Traceability of cattle products will be enhanced:
  - Breed and source population
  - Mixed products derived from multiple animals

Current bovine genomics work at UCD/Teagasc

- Analyses of candidate genes in beef and dairy cattle.
  - Focus on imprinted genes related to growth and development
  - Development of a custom SNP chip for beef and dairy traits in Irish cattle populations
- Immunogenomics of bovine TB.
  - Identification of a gene expression biosignature of TB infection in cattle
  - Functional genomics and systems biology of macrophage/mycobacteria interactions: how does bovine TB evade the immune system?

Current bovine genomics work at UCD/Teagasc

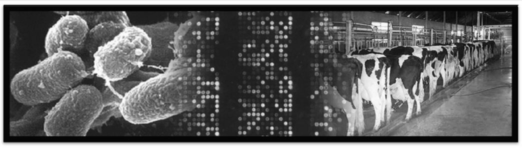
- Pan-genomic SNP chip analyses of beef cattle with detailed production data and biochemical phenotypes.
  - Identification of genomic regions containing genes for ETLs
  - Functional information for advanced genomic selection
- Genome sequencing of modern and ancient cattle.
  - Illumina® sequencing of extinct aurochs and extant modern cattle
  - Exon-capture sequencing of functional genes (~1.5% of genome)
- Physiological genomics of bovine reproduction (Prof. Alex Evans).

Imprinted genes: ideal candidates for growth and development traits

Imprinted genes: ideal candidates for growth and development traits

Gene name	Location	Gene product	Gene product function	Expressed allele	No. of observed SNPs (Ensembl/*de novo*)	SNP density
IGF2	BTA 29	Insulin-like growth factor 2	Growth & development	Paternal	7 (7 / 0)	1 SNP / 847 bp
IGF2R	BTA 9	Insulin-like growth factor 2 receptor	Receptor of IGF2 protein (role in growth & development)	Maternal	12 (2 / 10)	1 SNP / 292 bp
H19	BTA 29	Untranslated RNA (BWS in humans)	Unknown (role in development)	Maternal	8 (4 / 4)	1 SNP / 293 bp
GRF1	BTA 21	Guanine nucleotide-releasing factor 1	Signal transduction	Paternal	18 (14 / 4)	1 SNP / 601 bp
GNAS	BTA 13	Guanine nucleotide-binding protein subunit A	Signal transduction (role in growth & development)	Maternal	9 (9 / 0)	1 SNP / 405 bp
GRB10 (MEG1)	BTA 26	Growth factor receptor-bound protein 10	Signal transduction (role in growth promotion)	Isoform-dependent	18 (7 / 11)	1 SNP / 103 bp
MEST (PEG1)	BTA 4	Mesoderm specific transcript homolog	Growth promoting effects	Paternal	0 (0 / 0)	-
PEG10	BTA 4	Paternally-expressed gene 10	Signal transduction (role in fetal gestation)	Paternal	1 (0 / 0)	1 SNP / 1,069 bp
PHLDA2	BTA 15	Pleckstrin homology-like domain, family A, member 2	Regulator of placental growth	Maternal	Under investigation	TBC
ZNF215	BTA 15	Zinc finger protein 215	Hydrolase protein (role in development)	Maternal	Under investigation	TBC
SGCE	BTA 4	Sarcoglycan (epsilon)	Transmembrane protein	Paternal	Under investigation	TBC
PEG3	BTA 18	Zinc finger protein	Role in cell division	Paternal	16 (13 / 3)	1 SNP / 243 bp
NNAT	BTA 13	Neuronatin	Iron channel regulator protolipid (brain development)	Paternal	1 (1 / 0)	1 SNP / 610 bp
DLK1	BTA 21	Delta-like 1 homolog (Drosophila)	Signal transduction (Callipyge in sheep)	Paternal	0 (0 / 0)	-
MEG3	BTA 21	Untranslated RNA	Regulation of DLK1 expression	Maternal	7 (5 / 2)	1 SNP / 193 bp
MEG8	BTA 21	Untranslated RNA	Potential regulation of DLK1 expression	Maternal	9 (8 / 1)	1 SNP / 222 bp
PEG11AS	BTA 21	Untranslated RNA	Potential regulation of DLK1 expression	Paternal	5 (5 / 0)	1 SNP / 270 bp
CLPG	BTA 21	Untranslated RNA	ICR in sheep (Callipyge)	N/A	1 (0 / 1)	1 SNP / 897 bp
D103	BTA 21	Deiodinase, iodothyronine type III	Thyroid hormone regulation	Paternal	0 (0 / 0)	-

Bovine tuberculosis research: host-pathogen interactions and development of new diagnostics



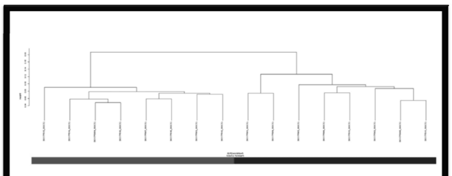
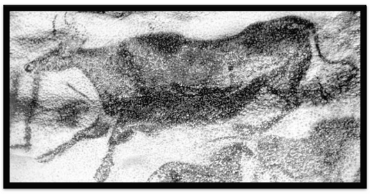
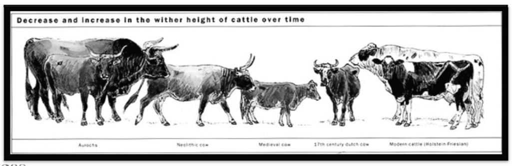


Fig. 2: Cluster dendrogram using gene expression data from 23,000 genes on the Affymetrix bovine GeneChip®. Shown are eight confirmed reactor BTB field-infected animals (red) and eight non-infected control animals (blue). All animals are unrelated female Holstein-Friesian dairy cattle.

Genome sequencing of aurochs





Acknowledgements

<u>Funding – Animal Genomics Laboratory, UCD</u>

- **Science Foundation Ireland (SFI)**
- **Enterprise Ireland (EI)**
- **Department of Agriculture, Fisheries and Food (DAFF)**
- **European Union Framework 7 (FP7)**