Lessons learned in pooling data for reference populations

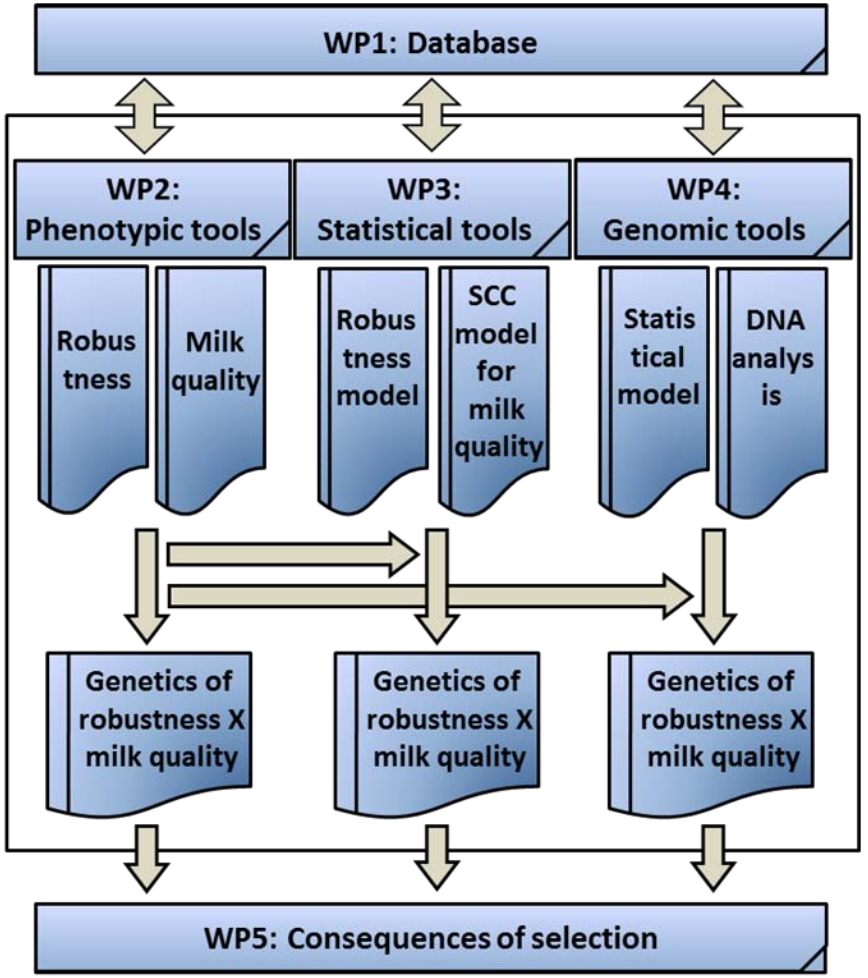
Eileen Wall, Mike Coffey, Roel Veerkamp, Sinead McParland and Georgios Banos

ROBUSTMILK

ROBUSTMILK

SEVENTH FRAMEWORK PROGRAMME

Develop innovative and practical breeding tools for improved dairy products from more robust dairy cows



SAC

SLU

gembloux agro bio tech

Université de Liège

WAGENINGEN UNIVERSITY WAGENINGEN UR

ANIMAL SCIENCES GROUP WAGENINGEN UR

eatasc

www.robustmilk.eu

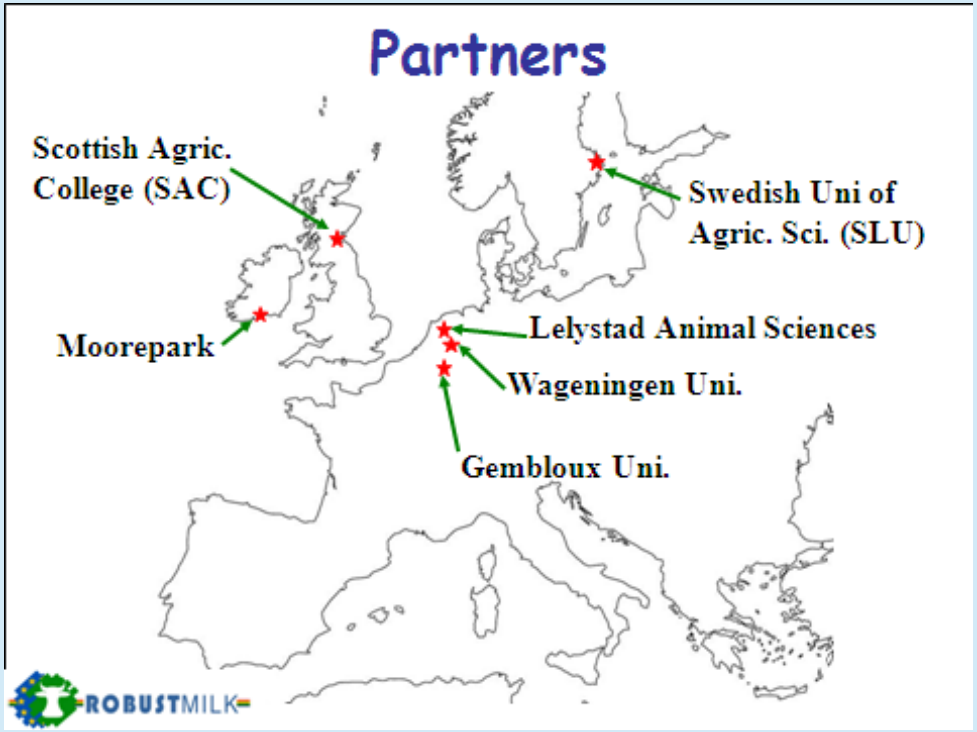
Pooling data – central database

- In the age of the genotype…the phenotype is king!
  - Currently genomics is largely based on data from national monitoring schemes and genetic evaluations
  - Extensively recorded resource populations with “villages” of data on novel/ expensive traits
    - Need to be linked to create the phenotypic “kingdom” for to allow a more rigorous genetic and genomic analysis
- The objectives of this study were to
  1. Demonstrate feasibility of merging phenotypic data from different resources populations
  2. Characterise merged database and its suitability for a joint genetic analysis

Data sources

- Key resource population data included experiemental/test herds in Holland (2), Ireland, Sweden and the UK

- For example, across 4 herds there are ~ 2,000 cows with repeated feed intake data



Database

- Cross institutional database of dairy herd animal identities with links to performance and genomic information where available.
  - SQL Server 2005, with a web-interface developed within VPN (secure!)
  - QA control, error checking and back-up storage
  - Common protocols for preparing and storing of data related to animals, pedigree, genotypes and performance records (user up/downloads)

Explorer Screenshots

Database Explorer

Database	Format	Institute	Show Tables
Langhill	MSSQL 2000	SAC-Edinburgh	->->->->->
T'Gen	MS Access 2003	ASG-Netherlands	->->->->->
fUSAGx-AWE	Files	Gembloux-Belgium	->->->->->
GeHuDB	MYSQL	SLU-Sweden	->->->->->
Teagasc	Oracle 10.1.0.4.0	Ireland	->->->->->

Database Explorer

<-<-<-<-Back to Database List->->->->

Database
T'Gen

Tables

Table	Description	View Table Fields
Dbgen	This table shows some general information of each animal. Unique records can be identified by the primary key: DNR (or LVN)	->->->->
genped	This table shows the pedigree information with sire and dam. Each unique record/animal can be identified by the primary key: Anim (or Lvnan)	->->->->
HGENERG	This table shows required information for the energy balance of the cows. Unique records can be identified by the combination of DNR (animal) and DAT (date of measuring)	->->->->
spotsamples	This table shows the milk components per measurement on 781 cows. Unique records can be identified by the combination of DNR (animal) and DAT (date of measuring)	->->->->
Akoe	This table shows information from the calvings of 2857 cows. Unique records can be identified by the combination of DNR (animal id) and DOC (date of calving).	->->->->
CLA98	This table shows information from the õcorpus luteum activityö and can be identified by DNR (animal id).	->->->->
Progesterone	This table shows information from progesterone measurements on several stages of lactation and can be identified by DNR (animal id) and DIM (days in milk).	->->->->
Leptine	This table shows leptin-concentrations for 412 cows. Unique records can be identified by the combination of DIER (Animal id) and DATUM (date).	->->->->
LiefersData	This table shows a combination of traits in previous tables and traits combining information in other tables. Unique records can be identified by the combination of SNR (sire id) and DATUMMETING (date of measuring).	->->->->
Lep011106ib	This table shows a combination of traits in previous tables and traits combining information in other tables. Unique records can be identified by the combination of SNR (sire id) and DATUMMETING (date of measuring).	->->->->
Sire	This table shows the genotypes of all sires with more than 10 daughters in the dataset. Unique records can be identified by SIRE (or SNR).	->->->->
HULP_OPN	This table shows records from daily feed intake with all the energy-contents of the food. Unique records can be identified by the combination of DNR (animal id) and DAT (date).	->->->->

Database Explorer

<-<-<-<-Back to Tables List->->->->

Database	Table
T'Gen	Dbgen

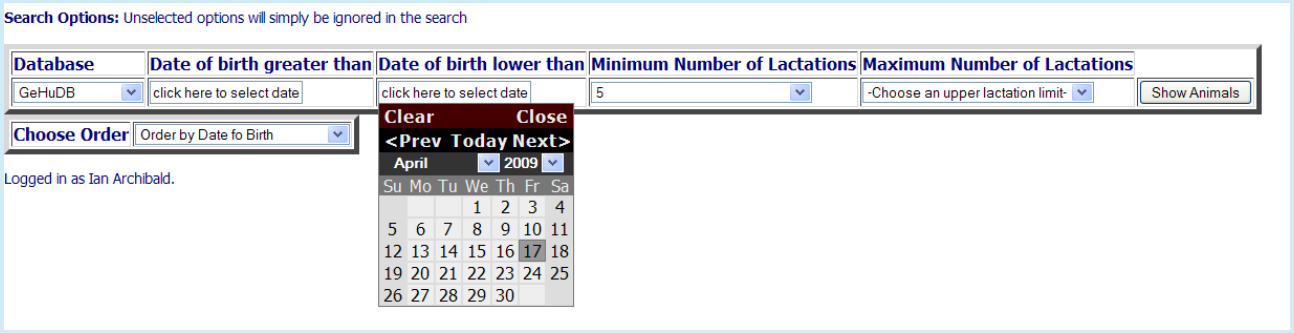
Fields

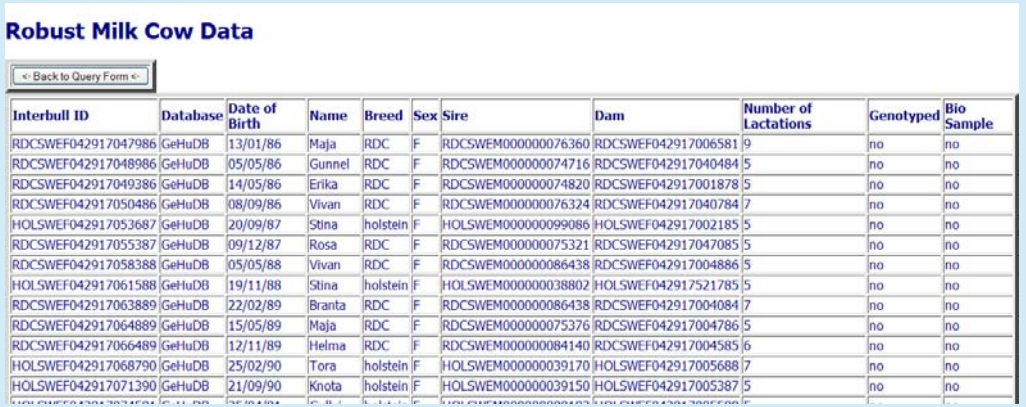
Field	Type	Description	Units(where applicable)
DNR	Character	Animal number (renumbered)	
LVN	Character	Unique animal herdbook id	
RAS	Integer	Breed	in classes and not in codes (37 classes)
SEXE	Character	Gender	M/V (male/female)
GBD	Date	Date of birth	d-m-yyyy
VDR	Character	Sire (father)	
MDR	Character	Dam (mother)	
ET	Character	Embryo transfer	J/N (Yes/No)
STAT	Character	Status	Afgevoerd (i.e. Transported) vs. Melkvee (i.e. Dairy cow)
GRP	Character	Group where animal belongs to	24 groups
AANK	Character	Bought	J/N (Yes/No)
AANGEV	Date	Arrival	dd-mm-yyyy
UBN	Integer	Herd number	
EIGEN	Character	Owner	
PROJ	Character	Project number	
BEDR	Character	Herd (name)	

Logged in as Ian Archibald.

Animal List

- Animal records can be searched with various options





Interbull ID	Database	Date of Birth	Name	Breed	Sex	Sire	Dam	Number of Lactations	Genotyped	Bio Sample
RDCSWEF042917047986	GeHuDB	13/01/86	Maja	RDC	F	RDCSWEM000000076360	RDCSWEF042917006581	9	no	no
RDCSWEF042917048986	GeHuDB	05/05/86	Gunnel	RDC	F	RDCSWEM000000074716	RDCSWEF042917040484	5	no	no
RDCSWEF042917049386	GeHuDB	14/05/86	Erika	RDC	F	RDCSWEM000000074820	RDCSWEF042917001878	5	no	no
RDCSWEF042917050486	GeHuDB	08/09/86	Vivan	RDC	F	RDCSWEM000000076324	RDCSWEF042917040784	7	no	no
HOLSWEF042917053687	GeHuDB	20/09/87	Stina	holstein	F	HOLSWEM000000099086	HOLSWEF042917002185	5	no	no
RDCSWEF042917055387	GeHuDB	09/12/87	Rosa	RDC	F	RDCSWEM000000075321	RDCSWEF042917047085	5	no	no
RDCSWEF042917058388	GeHuDB	05/05/88	Vivan	RDC	F	RDCSWEM000000086438	RDCSWEF042917004886	5	no	no
HOLSWEF042917061588	GeHuDB	19/11/88	Stina	holstein	F	HOLSWEM000000038802	HOLSWEF042917521785	5	no	no
RDCSWEF042917063889	GeHuDB	22/02/89	Branta	RDC	F	RDCSWEM000000086438	RDCSWEF042917004084	7	no	no
RDCSWEF042917064889	GeHuDB	15/05/89	Maja	RDC	F	RDCSWEM000000075376	RDCSWEF042917004786	5	no	no
RDCSWEF042917066489	GeHuDB	12/11/89	Helma	RDC	F	RDCSWEM000000084140	RDCSWEF042917004585	6	no	no
HOLSWEF042917068790	GeHuDB	25/02/90	Tora	holstein	F	HOLSWEM000000039170	HOLSWEF042917005688	7	no	no
HOLSWEF042917071390	GeHuDB	21/09/90	Knota	holstein	F	HOLSWEM000000039150	HOLSWEF042917005387	5	no	no

Home | Explore Database | Animal List | Upload | Phenotypic Data | Documents | Admin | Field Tags | Logout

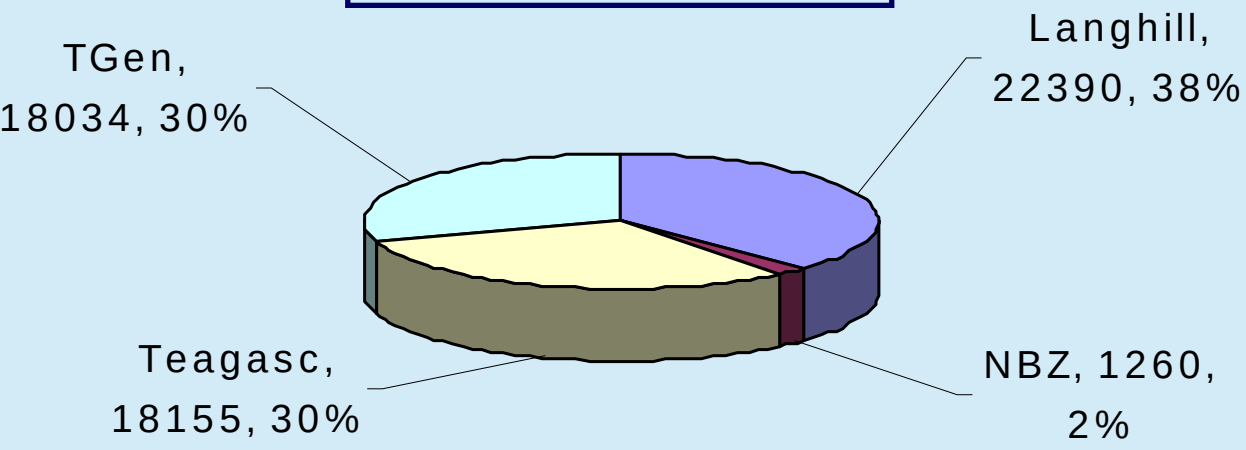
Database Name	Table Name	Description				
T'Gen	hollandPhenoTable1	Wageningen phenotypic breed table	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoTable2	Wageningen birthdate and genetic line	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoTable3	Wageningen milk sample analysis data	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbBreeds	Teagasc breed 1 and breed 2 composition of animals	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbExperiments	Teagasc experiment period detail of animals	View	Download	Upload	Get Upload Template
Langhill	langhillPhenoTable1	SAC Table of genetic line and feed group details	View	Download		
Langhill	langhillPhenoTable2	SAC Table of breeds and breed percentages	View	Download		
Langhill	langhillPhenoTable3Weekly	SAC Table of milk yields and content analysis	View	Download		
Langhill	langhillPhenoTable4	SAC Table of calving/service details	View	Download		
Langhill	langhillPhenoTable5	SAC Table of sporadic health events	View	Download		
T'Gen	hollandPhenoTable4	Wageningen calving/heat details	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbFixedEff	Teagasc fixed effects of animals	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbLactation	Teagasc lactation details of animals	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbRoutine	Teagasc routine milk sample details	View	Download	Upload	Get Upload Template
Teagasc	irelandPhenoRbSporadic	Teagasc sporadic codes and dates of events involving animals	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoNBZMaster	Wageningen extra animals master table	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoNBZPedBreed	Wageningen extra animals pedigree/breeds	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoNBZTable1	Wageningen extra animals breed compositions	View	Download	Upload	Get Upload Template
T'Gen	hollandPhenoNBZTable2	Wageningen extra animals genetic merit	View	Download	Upload	Get Upload Template
Langhill	langhillPhenoProgesterone	direct transfer from langhill Progesterone table	View	Download		
	phenotypicAnimalsUnion	Union of animals with phenotypic data from all partners	View	Download		
	phenotypicMilkUnion	union of milk sample data from partners(Langhill,ASG and MPK)	View	Download		
AUSAGx-AWE	belgiumLactationPheno	lactation numbers and dates	View	Download	Upload	Get Upload Template
AUSAGx-AWE	belgiumMilkCompositionPheno	milk composition data	View	Download	Upload	Get Upload Template

Click HERE to view upload logs.

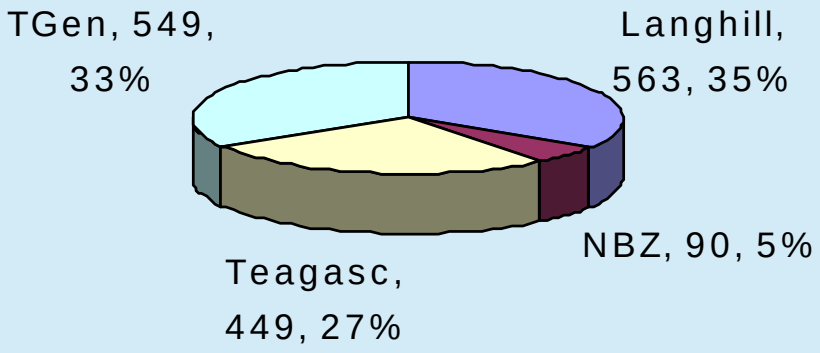
Merging - Data available

- 4 cow resource populations (herds)
  - Langhill, Holland (NBZ & TGen), Teagasc
- Weekly records
  - Milk, fat, protein yield
  - SCC
  - Live weight
  - Dry matter intake
  - Energy intake
- 1st lactation

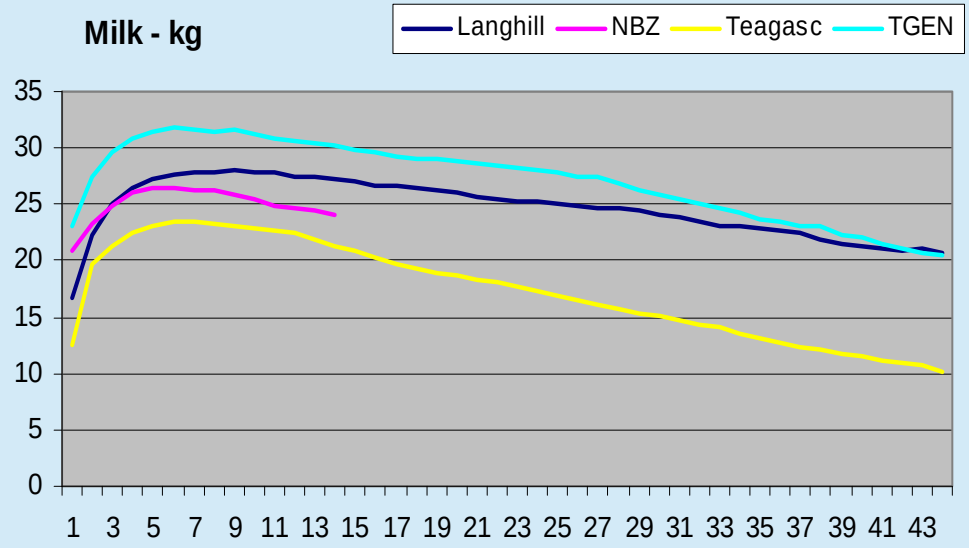
59,839 Weekly records

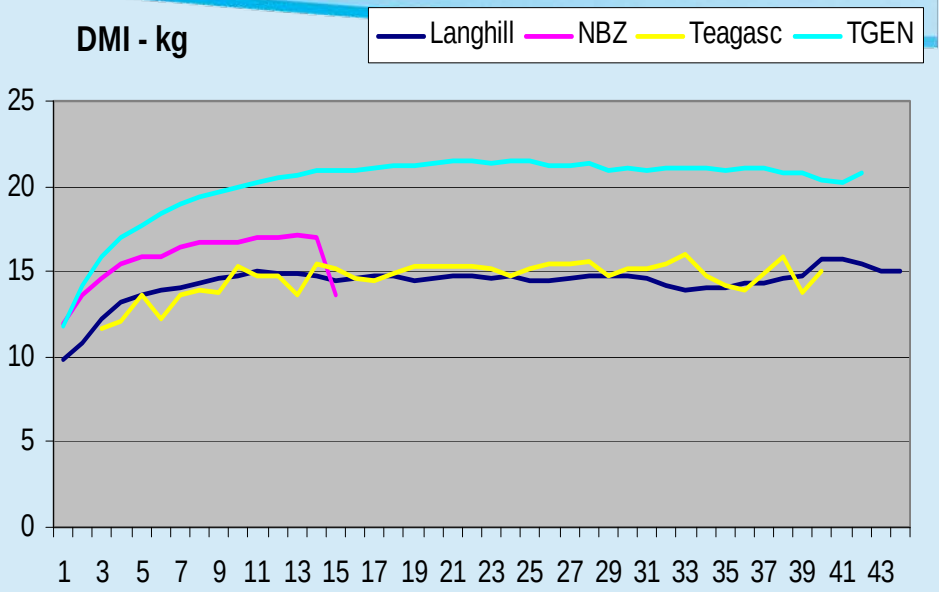


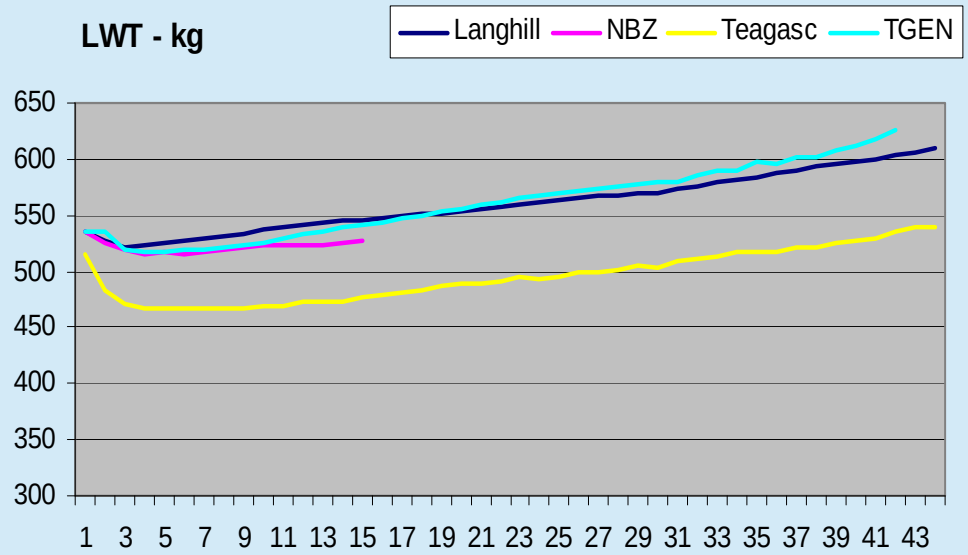
1651 Cows

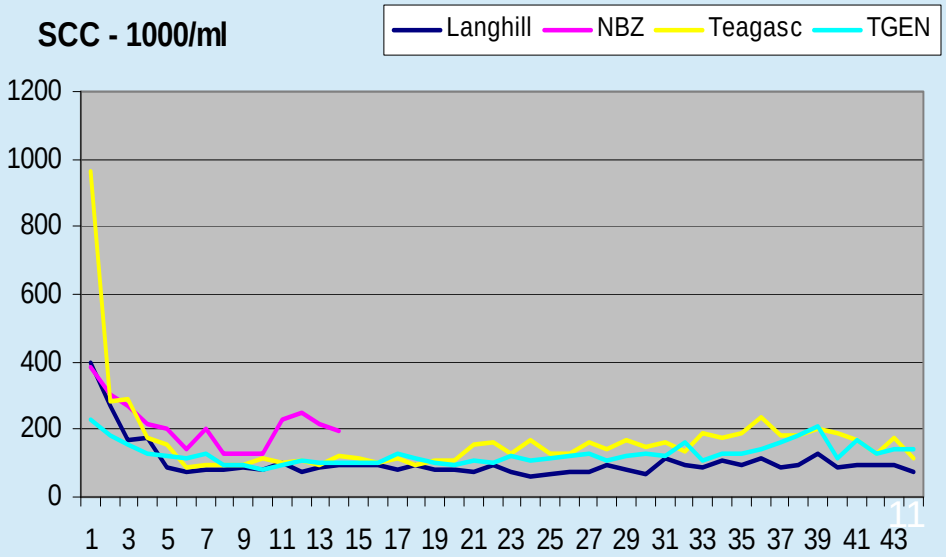


“Raw” data

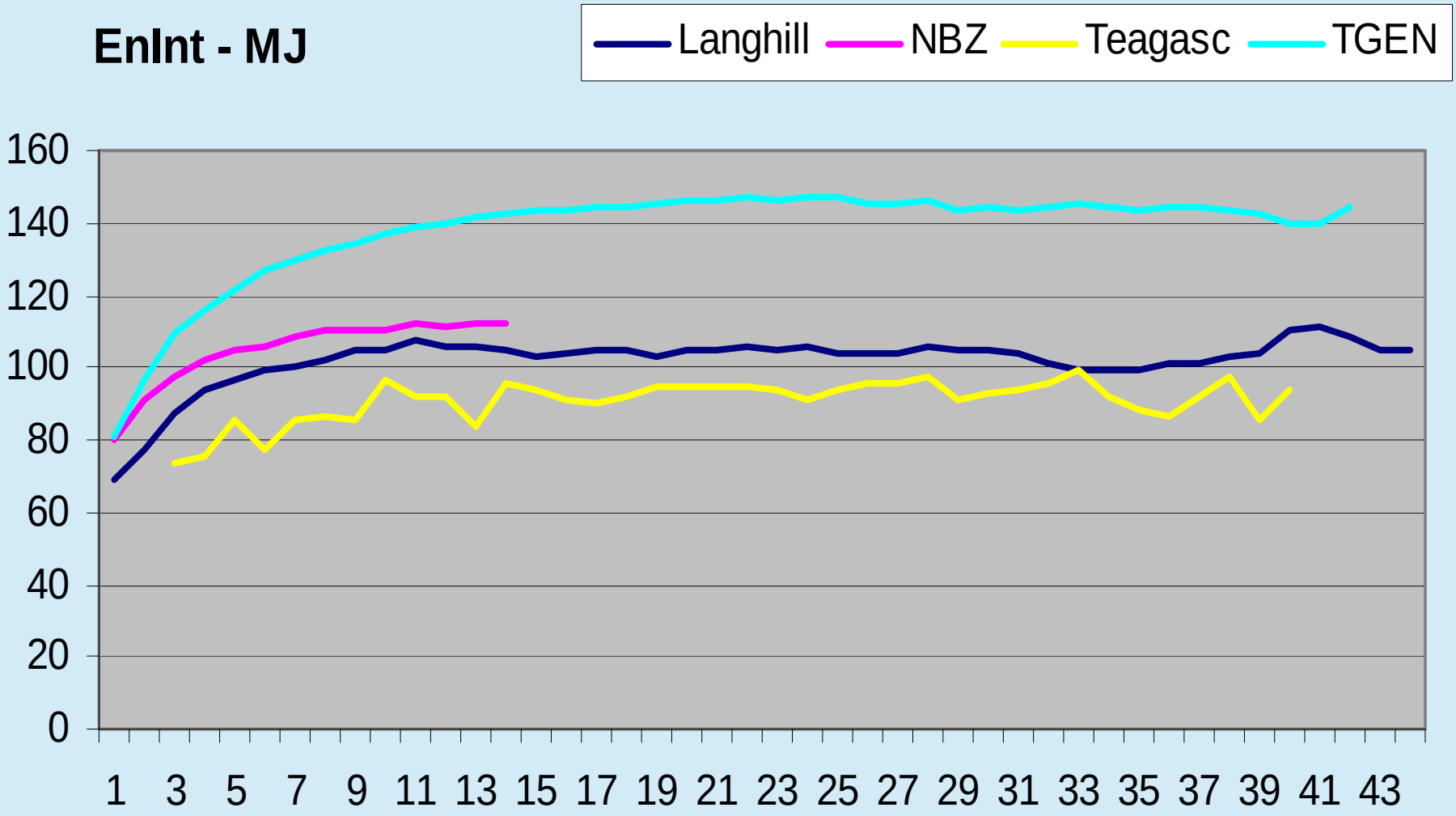

Milk - kg
Langhill
NBZ
Teagasc
TGEN
35
30
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43


DMI - kg
Langhill
NBZ
Teagasc
TGEN
25
20
15
10
5
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43


LWT - kg
Langhill
NBZ
Teagasc
TGEN
650
600
550
500
450
400
350
300
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43


SCC - 1000/ml
Langhill
NBZ
Teagasc
TGEN
1200
1000
800
600
400
200
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43

Energy Intake



RRM to derive phenotypes across lactation

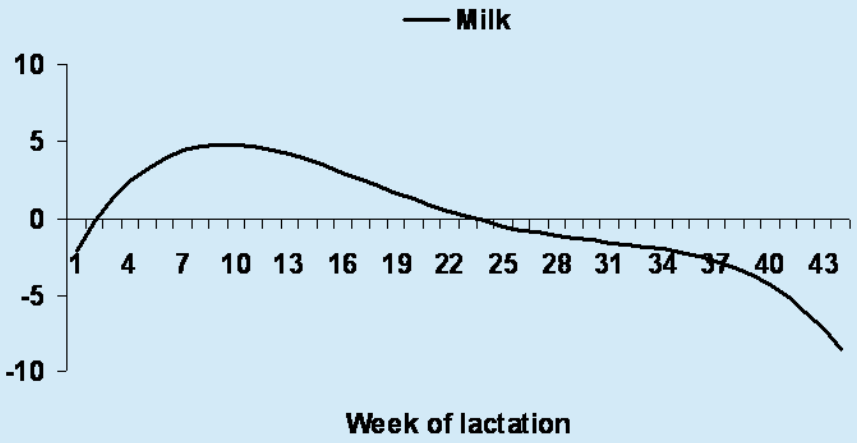
FIXED EFFECTS

- Genetic groups
  - 2 in each herd
- Feeding groups
  - 2 Langhill
  - 2 NBZ
  - 18 “Irish treatments”
- Milking frequency
  - 2x, 3x

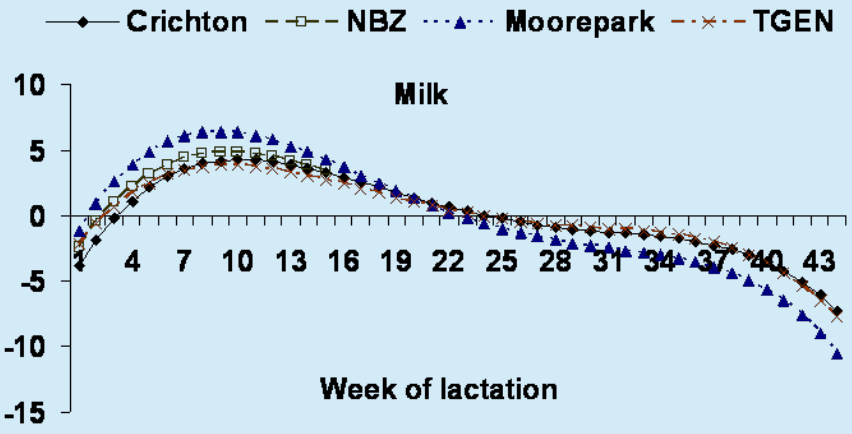
Y = herd*yr-month of test
yr-month of calving
calving age
genetic group
feeding group
Irish treatment
milking frequency
wk of test (poly 4)
animal*wk of test (poly 4)

Milk yield curves

- Population curve

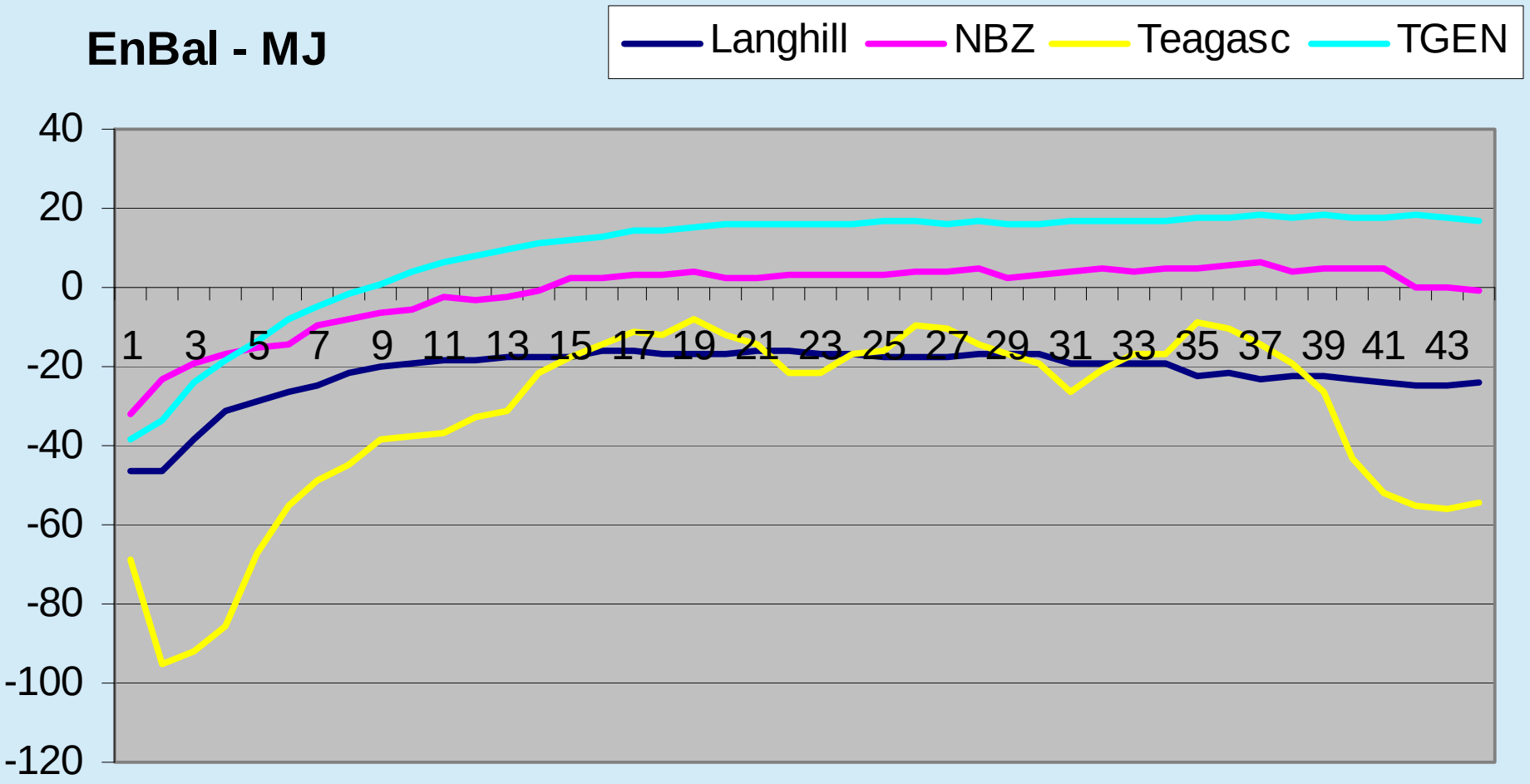


- Within herd curve



- Similar curves across and within herds
  - combining records across herds did not change the lactation profile of the trait.
  - consistent for all traits in the study.

Energy balance curves across lactation in the 4 herds



Heritability with/without sire*herd effect fitted

• Total milk yld (44 wks)	0.22 (0.07)	0.17 (0.08)
• Total fat yld (44 wks)	0.20 (0.07)	0.16 (0.08)
• Total protein yld (44 wks)	0.16 (0.07)	0.12 (0.07)
• Av fat % (44 wks)	0.68 (0.07)	0.66 (0.08)
• Av protein % (44 wks)	0.55 (0.07)	0.49 (0.08)
• Av fat:prot ratio (44 wks)	0.66 (0.07)	0.66 (0.08)
• Av SCC (wks 1-15)	0.14 (0.06)	0.10 (0.07)
• Av LWT	0.35 (0.07)	0.30 (0.08)
• Total DMI (44 wks)	0.15 (0.07)	0.15 (0.09)
• Total DMI (15 wks)	0.22 (0.08)	0.17 (0.09)
• Av energy balance (44 wks)	0.17 (0.08)	0.13 (0.09)
• Av energy balance (15 wks)	0.27 (0.08)	0.27 (0.08)

Conclusions

- Data from potentially very different experimental herds can be merged for larger genetic/genomic studies
- In this study, no evidence of sire*herd interaction
  - Good genetic links, numbers of animals
- Developing central and ongoing databases increases the value of the data, but it takes time and interaction!

Acknowledgements

- This research receives a financial support from the European Commission, Directorate-General for Agriculture and Rural Development, under Grand Agreement 211708 and form the Commission of the European Communities, FP7, KBBE-2007-1. This paper does not necessarily reflect the view of these institutions and in no way anticipates the Commission's future policy in this area.