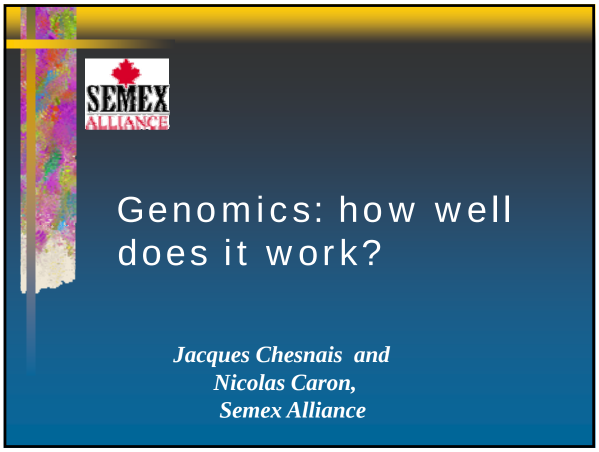
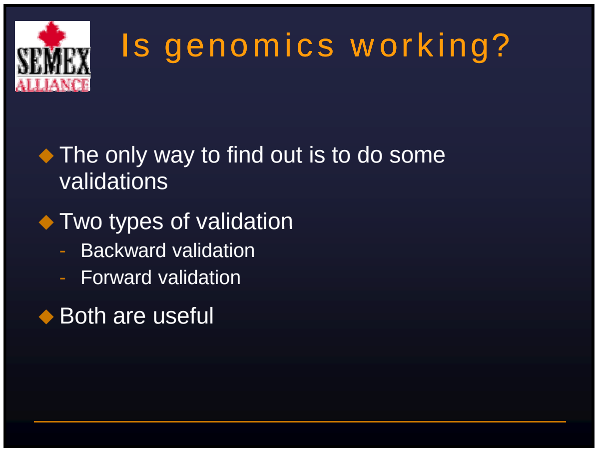


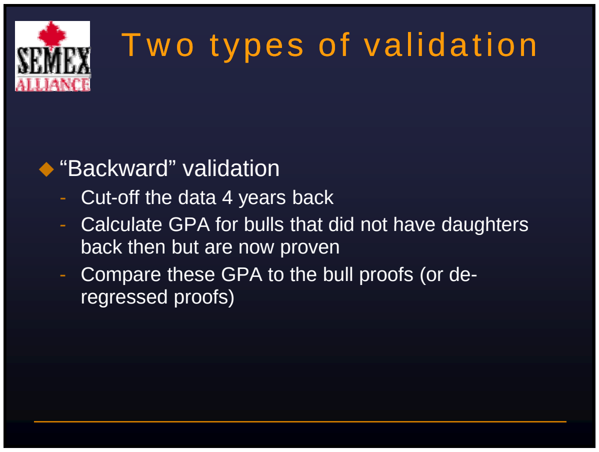
Genomics: how well does it work?

Jacques Chesnais and
Nicolas Caron,
Semex Alliance



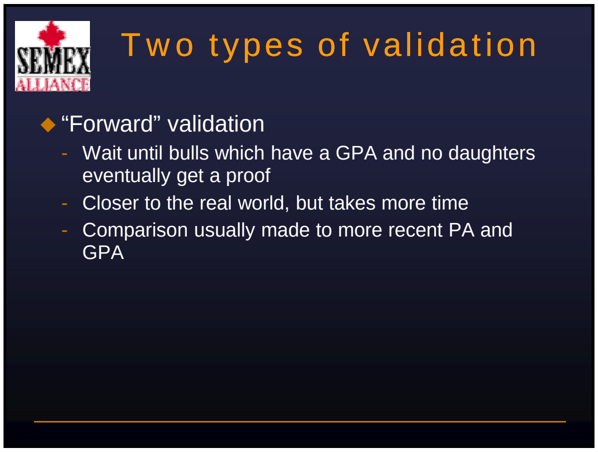
Is genomics working?

- The only way to find out is to do some validations
- Two types of validation
  - Backward validation
  - Forward validation
- Both are useful



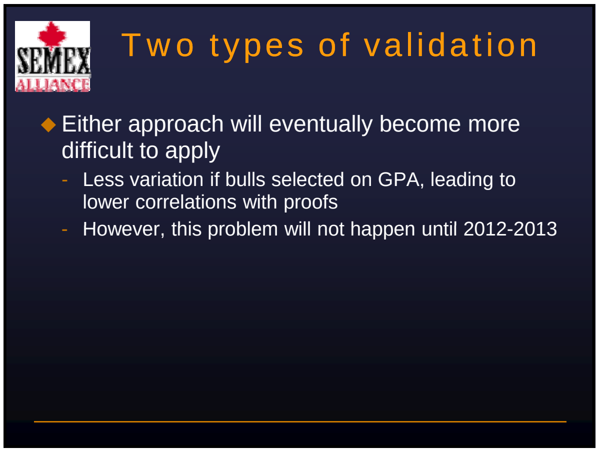
Two types of validation

- "Backward" validation
  - Cut-off the data 4 years back
  - Calculate GPA for bulls that did not have daughters back then but are now proven
  - Compare these GPA to the bull proofs (or de-regressed proofs)



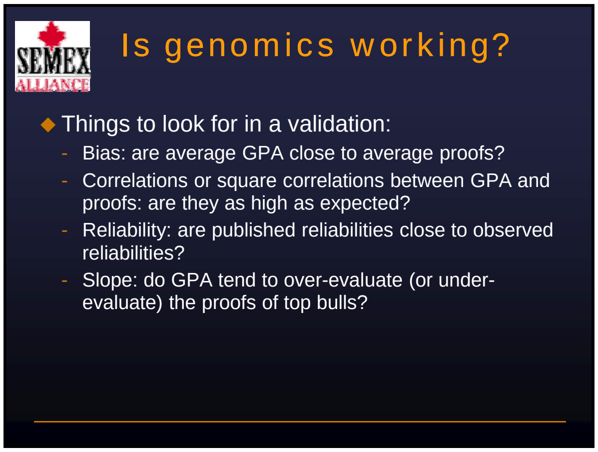
Two types of validation

- "Forward" validation
  - Wait until bulls which have a GPA and no daughters eventually get a proof
  - Closer to the real world, but takes more time
  - Comparison usually made to more recent PA and GPA



Two types of validation

- Either approach will eventually become more difficult to apply
  - Less variation if bulls selected on GPA, leading to lower correlations with proofs
  - However, this problem will not happen until 2012-2013



Is genomics working?

- Things to look for in a validation:
  - Bias: are average GPA close to average proofs?
  - Correlations or square correlations between GPA and proofs: are they as high as expected?
  - Reliability: are published reliabilities close to observed reliabilities?
  - Slope: do GPA tend to over-evaluate (or under-evaluate) the proofs of top bulls?

Bias: are average GPA close to average EBV?

- The average GPA and the average proof of validation bulls should be the same
- Otherwise, difficult to compare the GPA of young genotyped animals with the EBV or GEBV of older animals

Slope

- The regression of EBV on GPA should be close to 1 for each trait
- If it is less than 1, top GPA bulls are over-predicted compared to proven bulls (and bottom GPA under-predicted)
- If it is more than 1, top GPA bulls are under-predicted compared to proven bulls (and bottom GPA over-predicted)

Squared correlations

- Indicate the amount of variation in proofs that is explained by the GPA
- Vary between 0 and 1
- The higher the better
- Usually lower by a few points for de-regressed proofs than for proofs, especially for low heritability traits

Reliability

- Published reliability
  - computed from the pedigree information and genotypes of each animal and its relatives
  - separate reliability for each evaluated animal and trait
- Observed reliability
  - from validation studies
  - the most important one (is it working or not?)
  - computed for each trait from correlations between GPA and either proofs or de-regressed proofs (DYD)
  - one REL per trait for all bulls in validation group

Computation of observed reliabilities

- In validation studies, observed GPA reliabilities are computed from:
  - the squared correlation between the GPA and the proofs or de-regressed proofs of validation bulls
  - the accuracy (REL) of the proofs or de-regressed proofs of validation bulls
  - an adjustment to account for PA selection
- Issues with the computation of PA or their reliabilities can affect the computation of observed GPA reliabilities

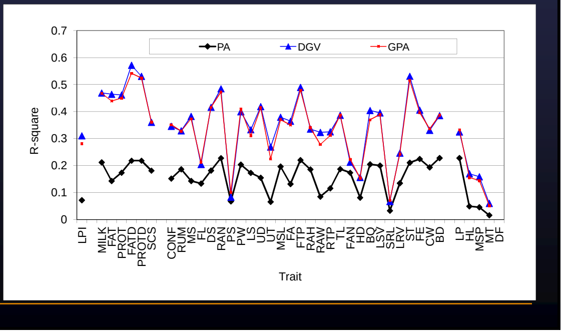
Backward validation

- The latest official backward validation results reported by CDN used April 2010 proofs
- Presented at CDN Open Industry Session in April 2010

CDN 2010 validation

- Calculate DGVs and GPAs using February 2006 proofs
- Determine how well these predict the April 2010 domestic bull proofs or de-regressed proofs
- 1,186 Holstein bulls with only a PA and GPA in 2006 and a proof in April 2010
- Results only shown in terms of R2: not easy to see if observed reliabilities correspond to published reliabilities

Young bulls – R-square – Holstein (Kistemaker, CDN, April 2010)



Squared correlations between PA, GPA and EBV and resulting gain in reliability of GPA over PA

Trait	R²(PA,EBV)	R²(GPA,EBV)	ΔR²	Gain in REL
LPI	.09	.33	.24	28
Milk	.24	.51	.27	30
Fat	.17	.49	.32	36
Protein	.21	.50	.29	32
SCS	.26	.43	.17	21
Conf.	.23	.44	.21	25
MSys	.21	.46	.25	29
F&L	.28	.38	.10	13
HL*	.05	.16	.11	20
DF**	.005	.12	.115	17

*de-regressed proofs, Kistemaker, 2010 **2009 validation, Kistemaker

CDN April 2010 validation results in terms of REL

Trait	PA reliability*	Observed Gain in reliability	Observed GPA reliability	Published GPA reliability*	Difference observed vs publ.
LPI	36	28	64	64	0
Milk	38	30	68	68	0
Fat	38	36	74	68	+6
Protein	38	32	70	67	+3
SCS	37	21	58	65	-7
Conf.	36	25	61	63	-2
MSys	37	29	66	64	+2
F&L	35	13	48	59	-11
HL	31	20**	51	58	-7
DF	30	17***	47	54	-7

*average reliability reported by CDN in April 2010

**2010 CDN validation, de-regressed proofs, ***2009 CDN validation,

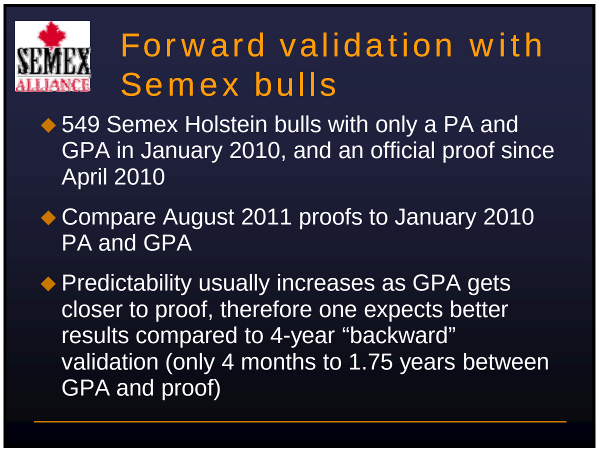
CDN Validation, April 2010 - Conclusions

- Reliability fairly accurate for production traits
- Observed REL were lower than published REL for SCS, Feet and Legs, Herd Life and Daughter Fertility
- No information available on bias or slopes
- Would be useful for CDN to do validations regularly as data accumulates, and report on bias, correlations, REL and slopes for each trait

CDN Forward validation, August 2010-CDN article

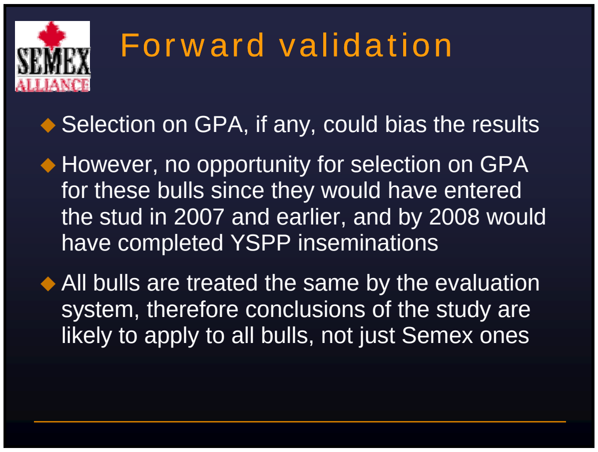
- 225 Holstein bulls with first official proofs in April or August 2010
- Average GPA bias: 149 points LPI

	12months	8 months	4 months
R²(PA,GLPI)	0.29	0.34	0.40
R²(GPA,GLPI)	0.49	0.53	0.59



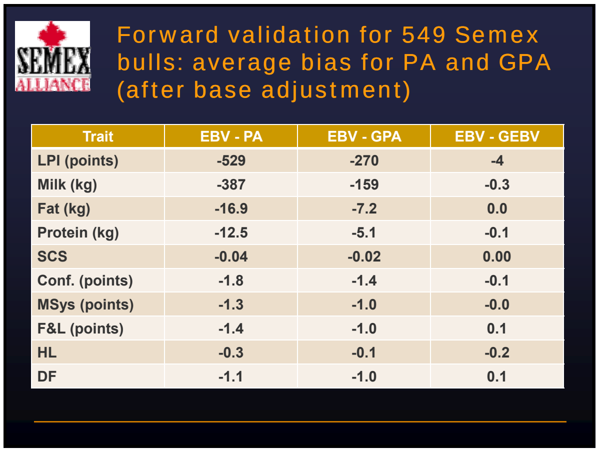
Forward validation with Semex bulls

- 549 Semex Holstein bulls with only a PA and GPA in January 2010, and an official proof since April 2010
- Compare August 2011 proofs to January 2010 PA and GPA
- Predictability usually increases as GPA gets closer to proof, therefore one expects better results compared to 4-year “backward” validation (only 4 months to 1.75 years between GPA and proof)



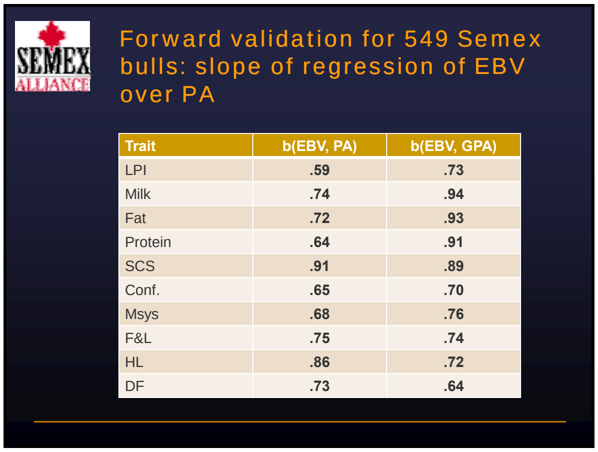
Forward validation

- Selection on GPA, if any, could bias the results
- However, no opportunity for selection on GPA for these bulls since they would have entered the stud in 2007 and earlier, and by 2008 would have completed YSPP inseminations
- All bulls are treated the same by the evaluation system, therefore conclusions of the study are likely to apply to all bulls, not just Semex ones



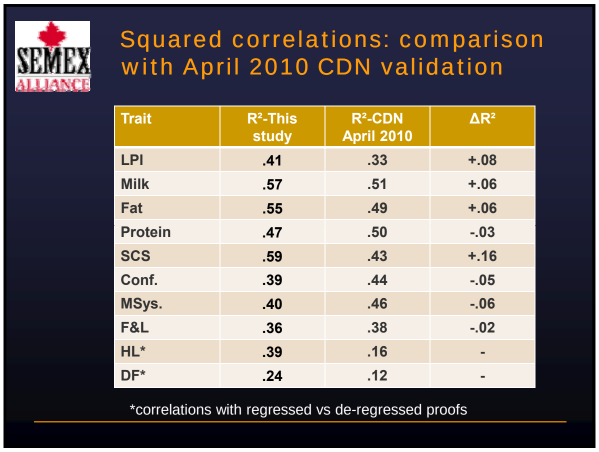
Forward validation for 549 Semex bulls: average bias for PA and GPA (after base adjustment)

Trait	EBV - PA	EBV - GPA	EBV - GEBV
LPI (points)	-529	-270	-4
Milk (kg)	-387	-159	-0.3
Fat (kg)	-16.9	-7.2	0.0
Protein (kg)	-12.5	-5.1	-0.1
SCS	-0.04	-0.02	0.00
Conf. (points)	-1.8	-1.4	-0.1
MSys (points)	-1.3	-1.0	-0.0
F&L (points)	-1.4	-1.0	0.1
HL	-0.3	-0.1	-0.2
DF	-1.1	-1.0	0.1



Forward validation for 549 Semex bulls: slope of regression of EBV over PA

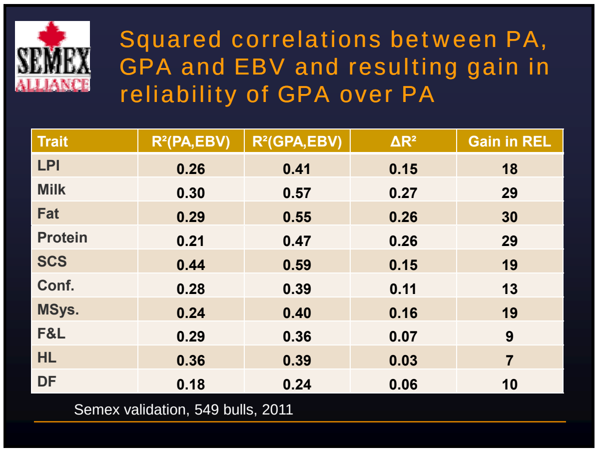
Trait	b(EBV, PA)	b(EBV, GPA)
LPI	.59	.73
Milk	.74	.94
Fat	.72	.93
Protein	.64	.91
SCS	.91	.89
Conf.	.65	.70
Msys	.68	.76
F&L	.75	.74
HL	.86	.72
DF	.73	.64



Squared correlations: comparison with April 2010 CDN validation

Trait	R^2-This study	R^2-CDN April 2010	ΔR^2
LPI	.41	.33	+.08
Milk	.57	.51	+.06
Fat	.55	.49	+.06
Protein	.47	.50	-.03
SCS	.59	.43	+.16
Conf.	.39	.44	-.05
MSys.	.40	.46	-.06
F&L	.36	.38	-.02
HL*	.39	.16	-
DF*	.24	.12	-

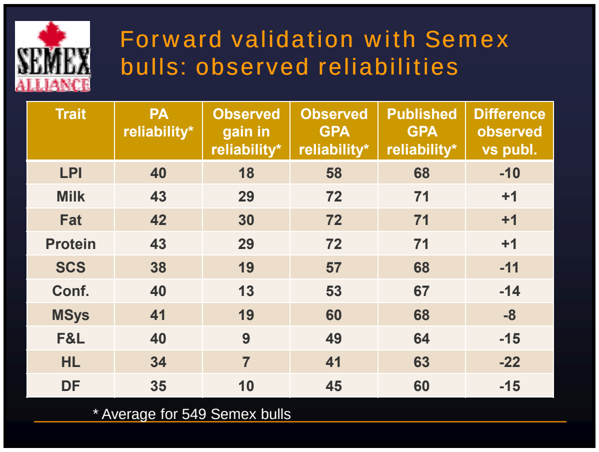
*correlations with regressed vs de-regressed proofs



Squared correlations between PA, GPA and EBV and resulting gain in reliability of GPA over PA

Trait	R^2(PA,EBV)	R^2(GPA,EBV)	ΔR^2	Gain in REL
LPI	0.26	0.41	0.15	18
Milk	0.30	0.57	0.27	29
Fat	0.29	0.55	0.26	30
Protein	0.21	0.47	0.26	29
SCS	0.44	0.59	0.15	19
Conf.	0.28	0.39	0.11	13
MSys.	0.24	0.40	0.16	19
F&L	0.29	0.36	0.07	9
HL	0.36	0.39	0.03	7
DF	0.18	0.24	0.06	10

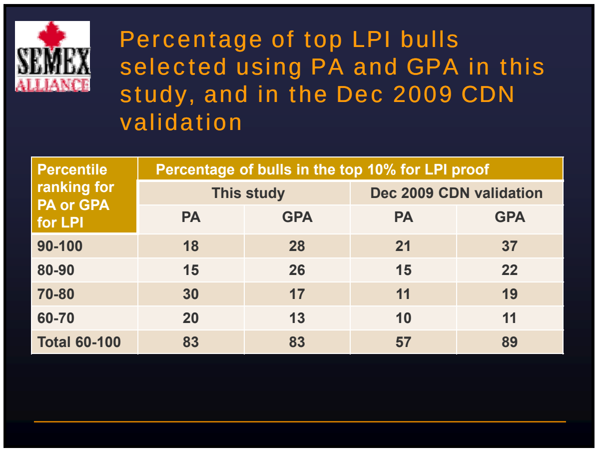
Semex validation, 549 bulls, 2011



Forward validation with Semex bulls: observed reliabilities

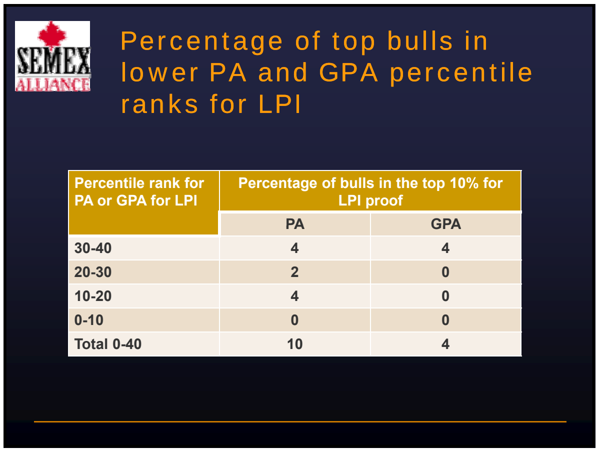
Trait	PA reliability*	Observed gain in reliability*	Observed GPA reliability*	Published GPA reliability*	Difference observed vs publ.
LPI	40	18	58	68	-10
Milk	43	29	72	71	+1
Fat	42	30	72	71	+1
Protein	43	29	72	71	+1
SCS	38	19	57	68	-11
Conf.	40	13	53	67	-14
MSys	41	19	60	68	-8
F&L	40	9	49	64	-15
HL	34	7	41	63	-22
DF	35	10	45	60	-15

* Average for 549 Semex bulls



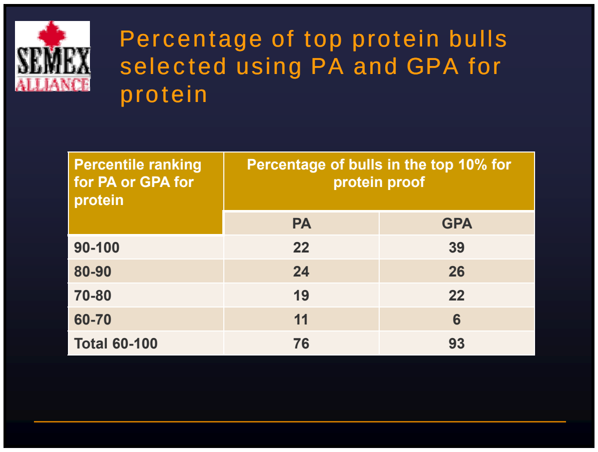
Percentage of top LPI bulls selected using PA and GPA in this study, and in the Dec 2009 CDN validation

Percentile ranking for PA or GPA for LPI	Percentage of bulls in the top 10% for LPI proof			
	This study		Dec 2009 CDN validation	
	PA	GPA	PA	GPA
90-100	18	28	21	37
80-90	15	26	15	22
70-80	30	17	11	19
60-70	20	13	10	11
Total 60-100	83	83	57	89



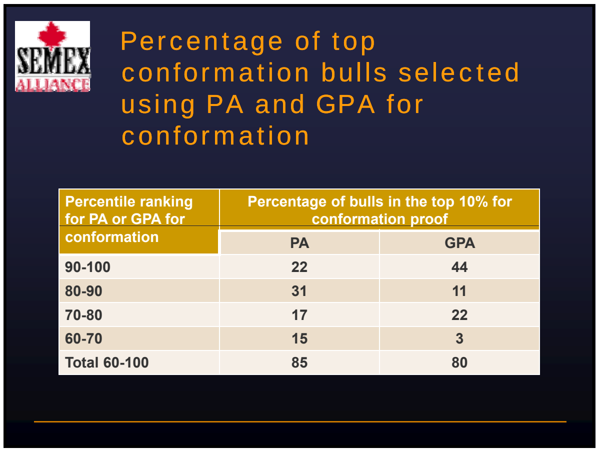
Percentage of top bulls in lower PA and GPA percentile ranks for LPI

Percentile rank for PA or GPA for LPI	Percentage of bulls in the top 10% for LPI proof	
	PA	GPA
30-40	4	4
20-30	2	0
10-20	4	0
0-10	0	0
Total 0-40	10	4



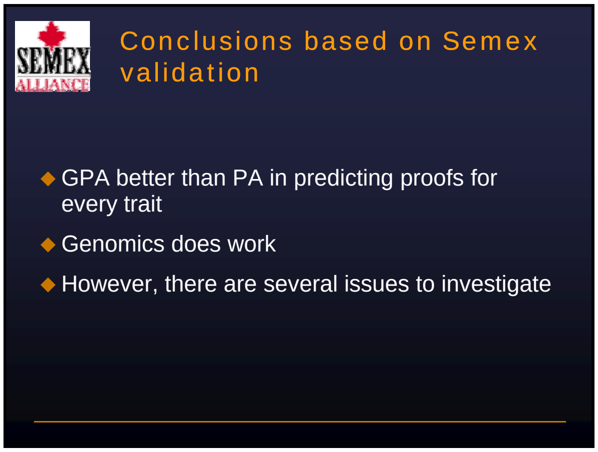
Percentage of top protein bulls selected using PA and GPA for protein

Percentile ranking for PA or GPA for protein	Percentage of bulls in the top 10% for protein proof	
	PA	GPA
90-100	22	39
80-90	24	26
70-80	19	22
60-70	11	6
Total 60-100	76	93



Percentage of top conformation bulls selected using PA and GPA for conformation

Percentile ranking for PA or GPA for conformation	Percentage of bulls in the top 10% for conformation proof	
	PA	GPA
90-100	22	44
80-90	31	11
70-80	17	22
60-70	15	3
Total 60-100	85	80



Conclusions based on Semex validation

- GPA better than PA in predicting proofs for every trait
- Genomics does work
- However, there are several issues to investigate

Average difference between GPA and EBV

- In this study, there was a significant bias in the GPA of young bulls, amounting to 270 points of LPI.
- This is more than what was reported by CDN for GPA calculated a year prior to the first proof (149 points). The GPA used in this study were computed between 4 months and 1.75 years before first proof.
- A large bias makes it more difficult to compare younger and older bulls.
- Reducing the size of this bias should be a priority, either by using a method similar to that of the USDA, or by developing a new Canadian method.

Regression of EBV over GPA (slope)

- In this study, slopes were lower than 0.80 for all traits except production traits and SCS
- In practice, this means the GPA of top young bulls may be over-estimated for these traits
- This again complicates any fair comparison with proven bulls

Squared correlations

- One would expect squared correlations to increase compared to those in a backward validation, because the time gap between GPA and proofs is shorter
- In this study, correlations for production traits increased, but those for type traits decreased
- Squared correlations for HL and DF are difficult to compare because they were computed on a different basis (proofs vs de-regressed proofs)

Reliabilities

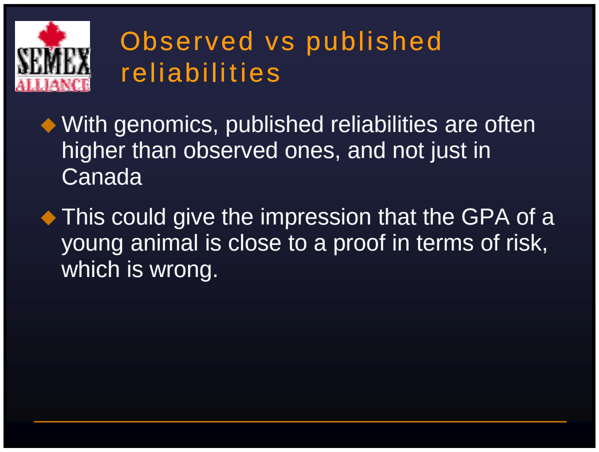
- Observed REL in this study were close to published ones for production traits but significantly lower for all other traits
- For these other traits, correlations between PA and proofs increased more than those between GPA and proofs. This reduced the increase in REL of the GPA over the PA.
- PA reliability increased, but not enough to compensate for the above reduction, which led to smaller observed reliabilities

Reliabilities

- The calculation of observed reliabilities used in backward validations may not be well suited to forward validations
- When bulls are selected, the available PA is the original one, not the more accurate one soon before the bull is proven
- It would be worth re-doing the study with de-regressed proofs to see the effect on observed reliabilities

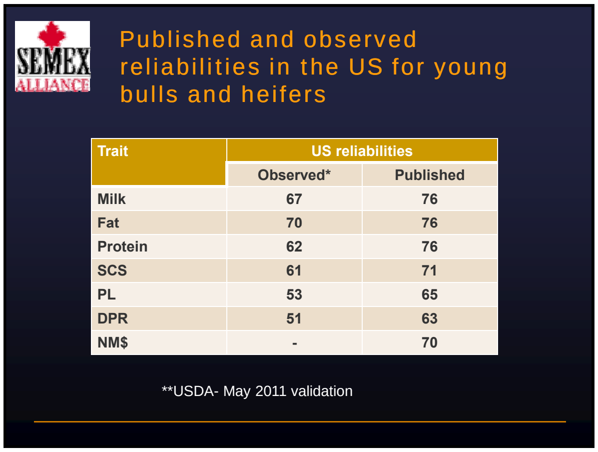
Reliabilities

- However, lower observed reliabilities for Feet and Legs, Herd Life and Daughter Fertility were also found in the CDN validation studies
- It seems composite traits tend to be more affected than single traits



Observed vs published reliabilities

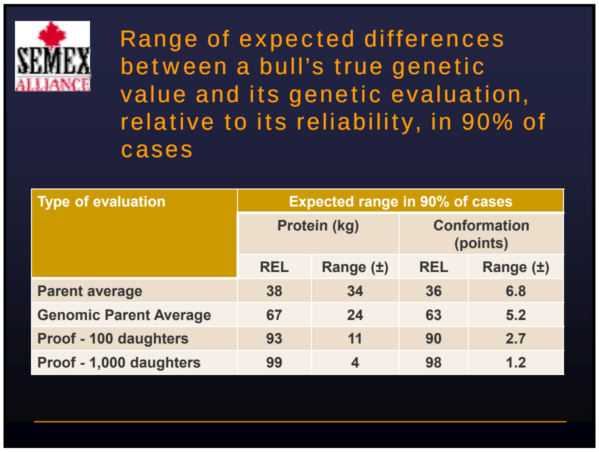
- With genomics, published reliabilities are often higher than observed ones, and not just in Canada
- This could give the impression that the GPA of a young animal is close to a proof in terms of risk, which is wrong.



Published and observed reliabilities in the US for young bulls and heifers

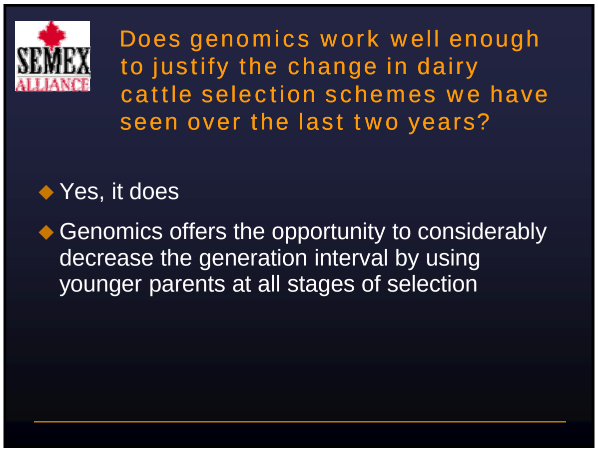
Trait	US reliabilities	
	Observed*	Published
Milk	67	76
Fat	70	76
Protein	62	76
SCS	61	71
PL	53	65
DPR	51	63
NM$	-	70

**USDA- May 2011 validation



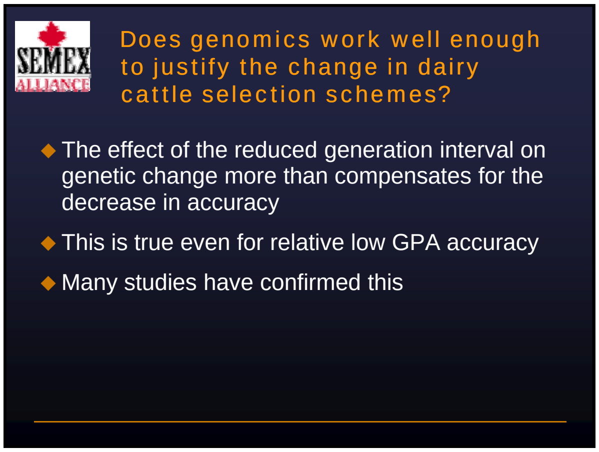
Range of expected differences between a bull's true genetic value and its genetic evaluation, relative to its reliability, in 90% of cases

Type of evaluation	Expected range in 90% of cases			
	Protein (kg)		Conformation (points)	
	REL	Range (±)	REL	Range (±)
Parent average	38	34	36	6.8
Genomic Parent Average	67	24	63	5.2
Proof - 100 daughters	93	11	90	2.7
Proof - 1,000 daughters	99	4	98	1.2



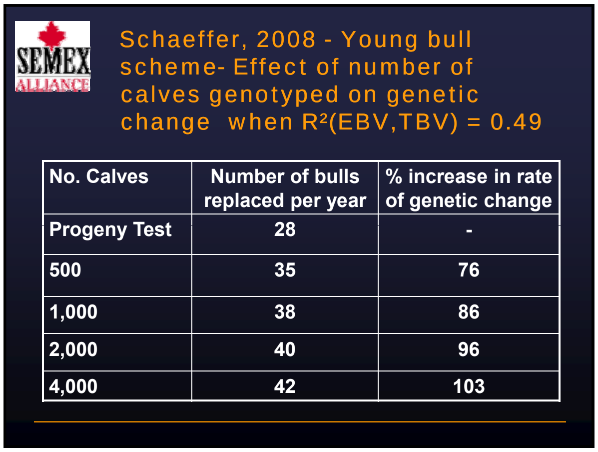
Does genomics work well enough to justify the change in dairy cattle selection schemes we have seen over the last two years?

- Yes, it does
- Genomics offers the opportunity to considerably decrease the generation interval by using younger parents at all stages of selection



Does genomics work well enough to justify the change in dairy cattle selection schemes?

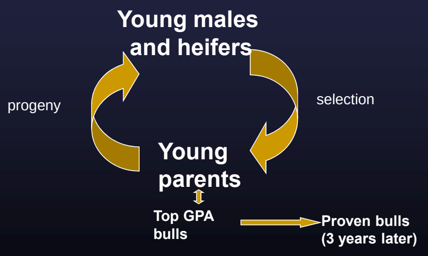
- The effect of the reduced generation interval on genetic change more than compensates for the decrease in accuracy
- This is true even for relative low GPA accuracy
- Many studies have confirmed this



Schaeffer, 2008 - Young bull scheme- Effect of number of calves genotyped on genetic change when $R^2(EBV,TBV) = 0.49$

No. Calves	Number of bulls replaced per year	% increase in rate of genetic change
Progeny Test	28	-
500	35	76
1,000	38	86
2,000	40	96
4,000	42	103

The new paradigm



Young males and heifers

progeny

selection

Young parents

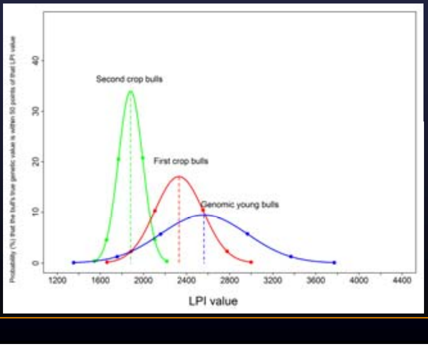
Top GPA bulls

Proven bulls (3 years later)

Does genomics work?

- Some of the issues found in this study could affect risk management for the producer
- Very high prices for heifers with top GPA, for example, are probably not warranted from an economic viewpoint (a heifer with a GPA LPI 100 or 200 points lower might be just as good genetically)
- The large bias, lower REL for some traits and low slope could also affect the way producers use young versus proven bulls

Top 30 young bulls versus top 10 proven bulls for LPI, April 2001



Conclusions

- GPA are always better predictors than PA, for all traits, in all validation studies so far
- Genomic selection works
- However, not everything works “as advertised”, as least based on this study
- CDN is in the best position to carry out validations (it has access to all bulls, complete data, etc…) and should report on them regularly

Conclusions

- For each validation, one should report bias, observed versus published REL, slopes, etc…
- In future (after 2012-13) , validation results will have to be adjusted for selection on GPA
- Potentially, methods should be developed to reduce the bias in GPA, and to tailor the computation or REL to individual traits
- The issues found in this study do not change the need to alter dairy cattle selection schemes in order to use genomics effectively