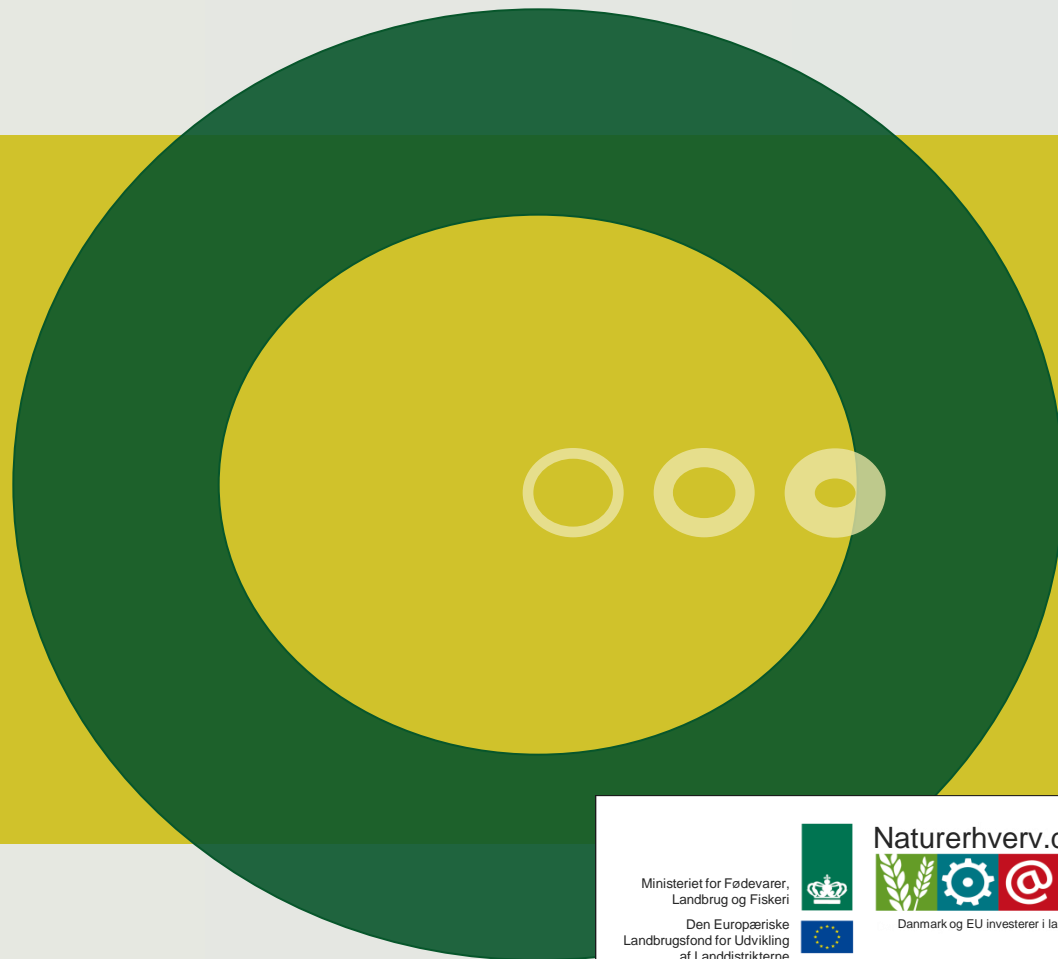




Imputation

Kevin Byskov og
Line Hjortø

Videncentret for
Landbrug
30-10-2013



Imputation: Hvad og Hvorfor

- **Impute = Tillægge**
- **Dvs, metode til at estimere manglende SNPer**
- **Øget markørtæthed → højere sikkerhed på GEBV**

Relevans for os

- **Illumina Bovine 54K chip**
 - **Version 1 og 2**
- **Eurogenomics (Holstein) og GENO (RDC)**
 - **Holland anvendte tidligere deres egen chip**
 - **GENO har også tidligere anvendt en anden chip**
- **Kører i referencegruppen (RDC og Jersey)**
 - **Low density chip**

NAV Missing calls



Ordforklaring

- SNP = Variation i et enkelt basepar
- Pr. dyr

2	2	2	2	2	1	2	0	0	2	1	1	2	2	1	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

- Homozygot for den ene allel = 0
- Heterozygot = 1
- Homozygot for den anden allel = 2

- Pr. streng

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1

- Den ene allel = 0
- Den anden allel = 1

Referencedyr genotypet med 54K

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation

Modificeret efter Marchini og Howie, 2010

Kandidater genotypet med LD

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

1	?	?	?	1	?	1	?	0	1	0	?	?	1	?	0
1	?	?	?	1	?	1	?	0	1	1	?	?	1	?	0

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation

Modificeret efter Marchini og Howie, 2010

Anvender information fra referencedyr

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

1	?	?	?	1	?	1	?	0	1	0	?	?	1	?	0
1	?	?	?	1	?	1	?	0	1	1	?	?	1	?	0

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation

Modificeret efter Marchini og Howie, 2010

Imputerer manglende SNP'er

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	0	1
1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0
0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0
1	1	1	0	0	1	0	0	1	1	1	0	1	1	1	0

1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0
1	1	1	1	1	0	1	0	0	1	1	1	1	1	1	0

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation

Modificeret efter Marchini og Howie, 2010

Fimpute

- **Udviklet til produktionsdyr**
- **Familiebaseret metode**
 - **Bruger slægtskabsinformation**
- **Giver præcise estimer for:**
 - **Homogene populationer**
 - **Dyr med genotypede slægtninge**
- **Forholdsvis kort beregningstid**
- **Bruges til Holstein**

NAV



Beagle

- **Udviklet til mennesker**
- **Populationsbaseret metode**
 - **Bruger ikke slægtskabsinformation**
- **Giver præcise estimater for:**
 - **Heterogene populationer**
 - **Ubeslægtede dyr**
 - **Dyr hvis nærmeste slægtninge ikke er genotypedede**
- **Meget robust metode**
- **Lang beregningstid**

NAV Bruges til RDC og Jersey



Test af imputationsmetoder

- Alle dyr er genotypede med 54K
- Datasættet opdeles:
 - Ældste dyr er referencedyr
 - Yngste dyr er kandidater
- De SNPer, der ikke er på LD-chippen, maskeres hos kandidaterne
- Imputation af maskerede SNPer
- Sammenligning af imputerede SNPer

År	Genotype
2011	1100010101
2011	0011100101
2012	1011110110
2012	0110000110

2013	1110001100
2013	0001100101

2013	1??0???1?0
2013	0??1???1?1

2013	1110001100
2013	0001101101

2013	1110001100
2013	0001100101

NAV og SNPer fra 54K



Imputation fra LD-chip til 54K

Holstein, Fimpute

Genotypedede forældre	# dyr	Imputationsfejl, %	Std. afvigelse	Max. # fejl
Ingen	67	3,4	1,9	8
Mor	14	1,7	1,1	5
Far	1213	3,2	1,8	18
Begge	432	0,7	0,6	5

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation

Imputation fra LD-chip til 54K

Holstein, Fimpute

Genotypedede forældre	# dyr	Imputationsfejl, %	Std. afvigelse	Max. # fejl
Ingen	67	3,4	1,9	8
Mor	14	1,7	1,1	5
Far	1213	3,2	1,8	18
Begge	432	0,7	0,6	5

RDC, Fimpute

Genotypedede forældre	# dyr	Imputationsfejl, %	Std. afvigelse	Max. # fejl
Ingen	9	10,3	4,0	14
Mor	3	3,7	0,6	4
Far	1147	4,3	2,7	19
Begge	234	0,7	0,5	2

Imputation fra LD-chip til 54K

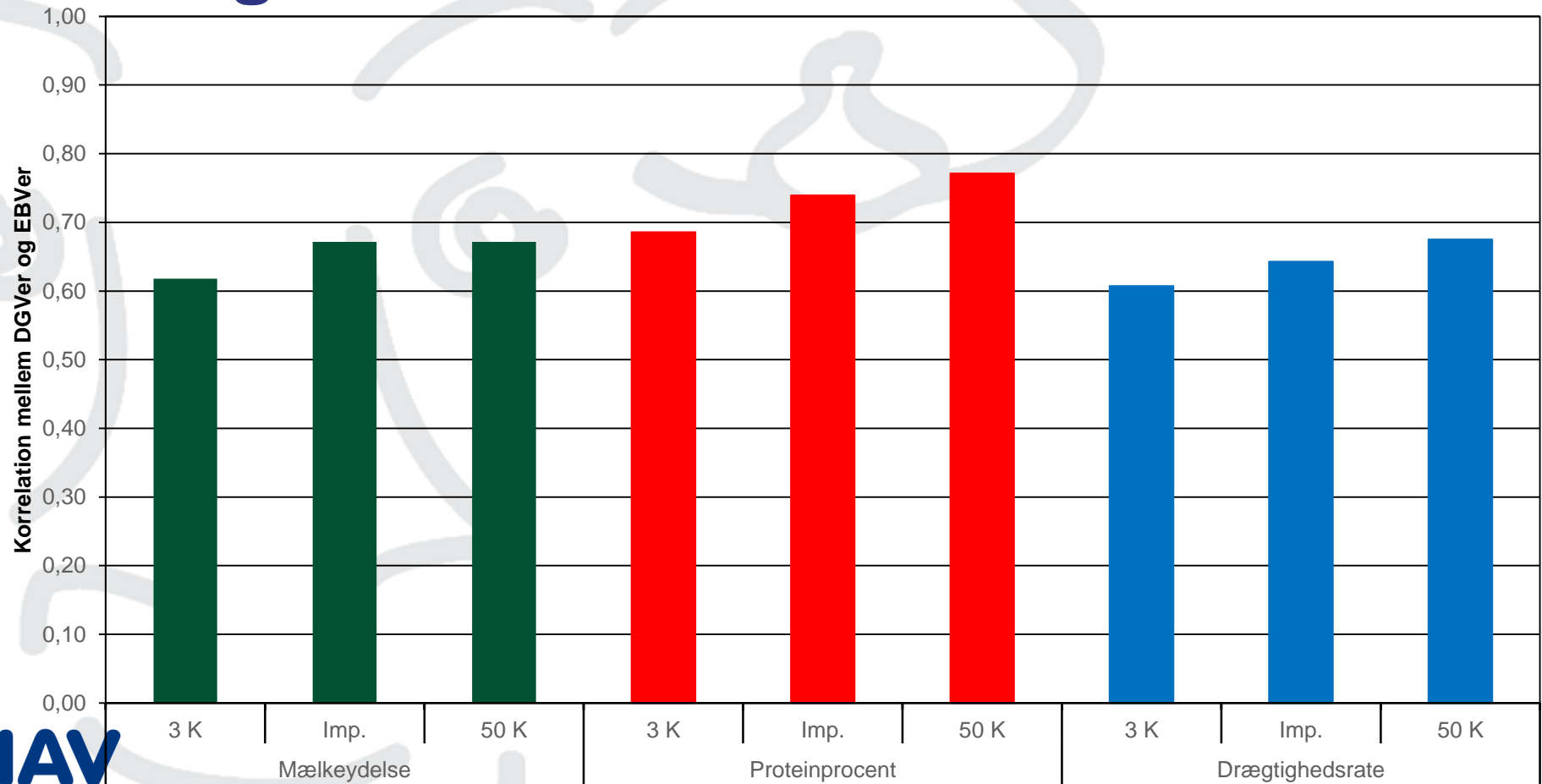
RDC, Beagle

Genotypedede forældre	# dyr	Imputationsfejl, %	Std. afvigelse	Max. # fejl
Ingen	9	2,3	1,2	5
Mor	3	1,0	0	1
Far	1147	1,3	0,6	6
Begge	234	1,0	0,1	2

RDC, Fimpute

Genotypedede forældre	# dyr	Imputationsfejl, %	Std. afvigelse	Max. # fejl
Ingen	9	10,3	4,0	14
Mor	3	3,7	0,6	4
Far	1147	4,3	2,7	19
Begge	234	0,7	0,5	2

Er imputerede SNPer bedre end manglende SNPer?



Faktorer der påvirker imputation

- **Referencegruppens størrelse**
- **Markørtæthed**
- **Effektiv populationsstørrelse**
- **Genetisk afstand mellem referencedyrene og kandidaterne**

Konklusion

- **Imputation øger sikkerheden på GEBV**
- **Forskellige metoder afhængigt af race:**
 - **Beagle til JER og RDC**
 - **Tidskrævende**
 - **Robust og præcis**
 - **Flmpute til HOL**
 - **Hurtig**
- **God alternativ metode - anvender slægtskab**

NAV



Nordisk Avlsværdi Vurdering • Nordic Cattle Genetic Evaluation