

Grundlagen der genomische Selektion

- Was sagt ein genomischer Zuchtwert aus?

Dr. Christian Edel, Dr. Reiner Emmerling und Dr. Kay-Uwe Götz
Institut für Tierzucht der bayerischen Landesanstalt für Landwirtschaft



Was ist genomische Selektion/ZWS?

- Zuchtwertschätzung unter Zuhilfenahme genomischer Information
 - ◆ Phänotypische (Leistungs-) Information: Ergebnisse der ‚normalen‘ Routinezuchtwertschätzung
 - ◆ Genomische Information: Ergebnisse einer molekulargenetischen Untersuchung



Gliederung

- Was ist genomische Selektion/ZWS?
- Versuch einer Begriffserklärung
- Wie weiß man, ob das System funktioniert?
- Einflussgrößen auf die Güte der GS



Was ist genomische Selektion/ZWS?

- Phänotypische Information
 - ◆ Grundlage sind die Ergebnisse einer ordentlichen, flächendeckend durchgeführten Leistungsprüfung und
 - ◆ die darauf basierenden Ergebnisse einer soliden konventionellen Zuchtwertschätzung auf dem aktuellen Stand der Technik
 - In Form von Zuchtwerten oder Töchterabweichungen
 - Kontinuierliche Weiterentwicklung der konventionellen ZWS als Konsequenz



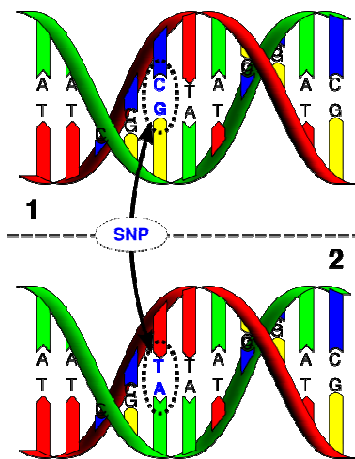
Was ist genomische Selektion/ZWS?

- Zuchtwertschätzung unter Zuhilfenahme genomischer Information
 - ◆ Der technologische Aspekt: Leistungsfähige und kostengünstige Feststellung des Genotyps an sehr vielen definierten Genorten
 - ◆ Der statistische Aspekt: Schätzen der Zuchtwerte dieser vielen Gensegmente und Kombination zum genomischen Zuchtwert

Was ist genomische Selektion/ZWS?

- Genomische Information
 - ◆ Grundlage bisheriger Entwicklungen weltweit, sog. 50k SNP-Chip (Illumina, bildet 54.001 SNP-Marker ab)
 - Dichtes Netz aus genetischen Markern überzieht das gesamte Genom des Rindes
 - ◆ Zukünftig: 600k (-850k) SNP-Chip
 - Abdeckung wird dichter und vollständiger

Was ist genomische Selektion/ZWS?



Was ist genomische Selektion/ZWS?

- Single Nucleotide Polymorphisms (SNPs, 'snips')
 - ◆ Einzelbasenaustausch in der DNA
AAGCCTA -> AAGCTTA
 - ◆ Aufgrund der Besonderheit des genetischen Codes häufig 'still', selektiv neutral und in kodierenden Bereichen (Genen) vorkommend
 - ◆ Offensichtlich an sehr vielen Stellen im Erbgut vorkommend (> 500.000)

Was ist genomische Selektion/ZWS?

- Single Nucleotide Polymorphisms (SNPs, 'snips')
- ◆ Gut für Hochdurchsatztypisierungen geeignet (SNP-Chips)
- ◆ Kostengünstig: Typisierungskosten für 54.000 Marker < 200 €
- ◆ Hohe Abdeckung und damit viel nutzbares Kopplungsungleichgewicht zu erwarten
- ◆ Durchbruch für ernsthafte Überlegungen zu genomischer Selektion



Was ist genomische Selektion/ZWS?

- Das Modell: Kopplungsungleichgewicht (linkage disequilibrium, LD) zwischen Marker und einem merkmalsbestimmenden Gen
- ◆ Im Dorf gibt es die Dorfschönheit Monika und vier junge Burschen (Anton, Bert, Charlie, Detlef)
- ◆ Bei verschiedenen Gelegenheiten sieht man Monika mit jeweils einem von diesen vieren, sie gibt dabei keinem den Vorzug – sie und die Burschen sind im Kopplungsgleichgewicht
- ◆ Sieht man Monika häufiger mit Anton als mit den anderen dreien, dann sind sie und Anton im Kopplungsungleichgewicht



Was ist genomische Selektion/ZWS?

- ◆ Gehen beide miteinander und man sieht Monika nur noch mit dem Anton dann ist das Kopplungsungleichgewicht vollständig

Erster Grad der Abstraktion

- ◆ Man sieht Monika und denkt: „bestimmt ist der Anton in der Nähe“, bzw. „die ist doch meistens mit dem Anton zusammen“

Zweiter Grad der Abstraktion

- ◆ Man weiß überhaupt nichts von einem Burschen Namens Anton, aber die Monika blüht plötzlich sichtlich auf – man vermutet: „wahrscheinlich ist sie in Kopplungsungleichgewicht mit einem, der einen guten Einfluss auf sie hat!“



Was ist genomische Selektion/ZWS?

- Klassische Modellvorstellung
- Alternative Modellvorstellung



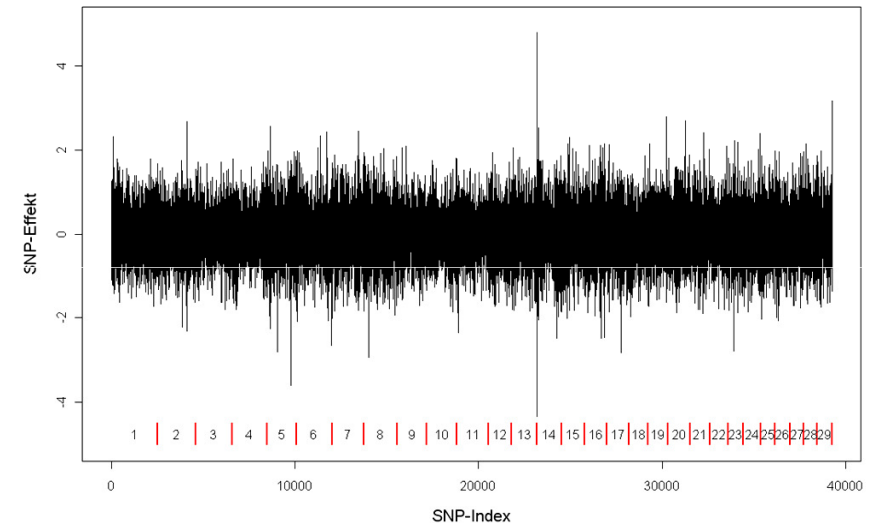
Was ist genomische Selektion/ZWS?

□ Klassische Modellvorstellung genomische ZWS:

- ◆ Der Zuchtwert eines Tieres ist die Summe der Zuchtwerte aller einzelnen Gene bzw. Gensegmente die dieses Tier trägt
- ◆ Wir kennen die Gene bzw. Gensegmente nicht, aber wir haben das dichte Netz aus Markern, somit erwarten wir ein Kopplungsungleichgewicht zwischen potentiellen Genen und benachbarten Markern
- ◆ Wir schätzen den ‚Markereffekt‘ als kombinierten Effekt eines unbekanntes Gens und des Kopplungsungleichgewichts zum Marker
- ◆ Wir nutzen diese ‚Markereffekte‘ zur Schätzung/Vorhersage des wahren Zuchtwerts



Was ist genomische Selektion/ZWS?



Was ist genomische Selektion/ZWS?

□ Vorgehensweise genomische ZWS:

- ◆ Eine möglichst große Gruppe von Tieren mit bereits sicheren Zuchtwerten wird genotypisiert und als Lernstichprobe (=Kalibrierungsstichprobe) verwendet
- ◆ Anhand dieser Lernstichprobe werden die Zuchtwerte der einzelnen SNPs (SNP-Effekte, Markereffekte) geschätzt
- ◆ Der Katalog von SNP-Effekten wird für die Vorhersage der Zuchtwerte junger Tiere verwendet
- ◆ Von Zeit zu Zeit muss neu ‚kalibriert‘ werden, d.h. die SNP-Effekte werden neu geschätzt



Was ist genomische Selektion/ZWS?

□ Alternative Modellvorstellung genomische ZWS:

- ◆ Äquivalente Form: genomische Verwandtschaftsmatrix
 - Eine auf Markerinformationen beruhende Form der additiv-genetischen (Ko-)Varianzmatrix
 - Beschreibt die Ähnlichkeit von Tieren auf Markerebene
 - Bezug zur klassischen Verwandtschaftsmatrix
- ◆ Mögliche Schlussfolgerung:
 - Mit der genomischen Information werden tatsächlich in wesentlichem Umfang die ‚wahren‘ Verwandtschaftsbeziehungen zwischen Tieren abgebildet
 - Der Informationsbeitrag von LD zwischen Markern und Genen nach der klassischen Modellvorstellung ist unklar (merkmalsspezifisch?)



Was ist genomische Selektion/ZWS?

„Normale“ Verwandtschaftsmatrix (2 Halbgeschwisterfamilien)

1.00	0.25	0.25	0.00	0.00	0.00	0.00	0.00
0.25	1.00	0.25	0.00	0.00	0.00	0.00	0.00
0.25	0.25	1.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	1.00	0.25	0.25	0.25	0.25
0.00	0.00	0.00	0.25	1.00	0.25	0.25	0.25
0.00	0.00	0.00	0.25	0.25	1.00	0.25	0.25
0.00	0.00	0.00	0.25	0.25	0.25	1.00	0.25
0.00	0.00	0.00	0.25	0.25	0.25	0.25	1.00

- Verwandtschaftskoeffizient (=durchschnittliche, erwartete Beziehung): 0.25
- Keine Verwandtschaft der Gruppen im Beispiel

Was ist genomische Selektion/ZWS?

„Realisierte“ Verwandtschaftsmatrix (2 Halbgeschwisterfamilien)

1.00	0.30	0.22	0.00	0.00	0.00	0.00	0.00
0.30	1.00	0.29	0.00	0.00	0.00	0.00	0.00
0.22	0.29	1.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	1.00	0.24	0.27	0.22	0.31
0.00	0.00	0.00	0.24	1.00	0.29	0.22	0.24
0.00	0.00	0.00	0.27	0.29	1.00	0.23	0.31
0.00	0.00	0.00	0.22	0.22	0.23	1.00	0.27
0.00	0.00	0.00	0.31	0.24	0.31	0.27	1.00

- Verwandtschaftskoeffizienten streuen um durchschnittliche, erwartete Beziehung von 0.25 mit einer Standardabweichung von etwa 0.029

Was ist genomische Selektion/ZWS?

„Realisierte“ Verwandtschaftsmatrix (2 Halbgeschwisterfamilien)

1.00	0.30	0.22	0.02	0.01	0.07	0.02	0.08
0.30	1.00	0.29	0.03	0.06	0.05	0.02	0.02
0.22	0.29	1.00	0.01	0.00	0.04	0.01	0.09
0.02	0.03	0.01	1.00	0.24	0.27	0.22	0.31
0.01	0.06	0.00	0.24	1.00	0.29	0.22	0.24
0.07	0.05	0.04	0.27	0.29	1.00	0.23	0.31
0.02	0.02	0.01	0.22	0.22	0.23	1.00	0.27
0.08	0.02	0.09	0.31	0.24	0.31	0.27	1.00

- Verwandtschaftskoeffizienten streuen um durchschnittliche, erwartete Beziehung von 0.25 mit einer Standardabweichung von etwa 0.029
- Auch vorher nicht verwandte Tiere zeigen jetzt eine Beziehung

Was ist genomische Selektion/ZWS?

□ Vorgehensweise genomische ZWS:

- ◆ Eine möglichst große Gruppe von Tieren mit bereits sicheren Zuchtwerten wird genotypisiert und als Eichprobe (=Kalibrierungstichprobe) verwendet
- ◆ Anhand der Zuchtwerte (bzw. Töchterabweichungen) und der Genotypen dieser Kalibrierungstichprobe werden die Zuchtwerte junger typisierter Tiere ohne eigene Leistung geschätzt
- ◆ Die Kalibrierungstichprobe wird sukzessive um Tiere mit Genotyp und Phänotyp erweitert, sobald solche Tiere vorliegen

Was ist genomische Selektion/ZWS?

□ Mögliche Implikationen

- ◆ Wann liegt wirklich ein Markereffekt im klassischen Sinn vor, wann nicht?
- ◆ Wie hoch ist die Schätzgenauigkeit aus reinem LD?
- ◆ Die Verwandtschaftsstruktur zwischen der Kalibrierung und der Validierung bestimmt zu einem guten Teil die Vorhersagekraft. Konsequenz:
 - Tiere mit vielen Verwandten in der Kalibrierung könnten sich deutlich besser schätzen lassen als Mitglieder seltener Linien
- ◆ Wie kann überhaupt vernünftig validiert werden?
- ◆ Müssen wir bei zukünftigen Genotypisierungen bzw. Phänotypisierungen diesem Umstand Rechnung tragen?

Versuch einer Begriffserklärung

- ◆ **DGV (DGBV)**: direct genomic (breeding) value – direkter genomischer Zuchtwert **dgW**
- ◆ **PA**: parent average – Elterndurchschnitt, besser: Pedigreezuchtwert, nur aus Pedigreeinformation geschätzter Zuchtwert **PI**
- ◆ **GEbv**: genomically enhanced breeding value – verrechneter Wert aus normalem Zuchtwert (PA) und dem direkten genomischen Zuchtwert (,blending`) **gZW**

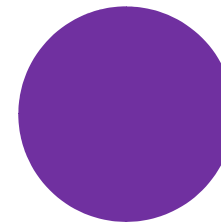
Wie weiß man ob das System funktioniert?

□ Kreuzvalidierung

- ◆ Gesamtheit aller typisierten Tiere wird in Kalibrierung und Validierungsgruppen aufgeteilt
- ◆ Von Kalibrierung und Validierung liegen genaue Zuchtwert und Töchterabweichungen vor
- ◆ Diese werden als ‚wahre‘ Zuchtwerte betrachtet

Wie weiß man ob das System funktioniert?

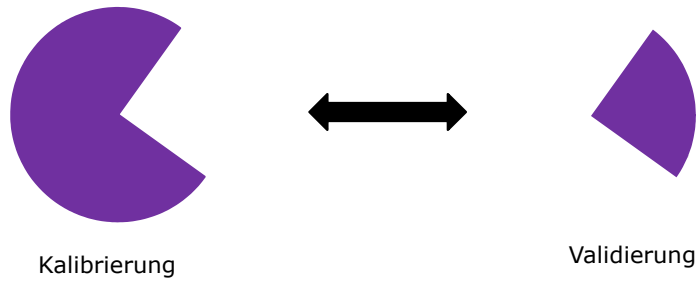
□ Kreuzvalidierung



Gesamtheit aller typisierten Tiere von denen **zusätzlich** bereits ein genauer Zuchtwertschätzer vorliegt

Wie weiß man ob das System funktioniert?

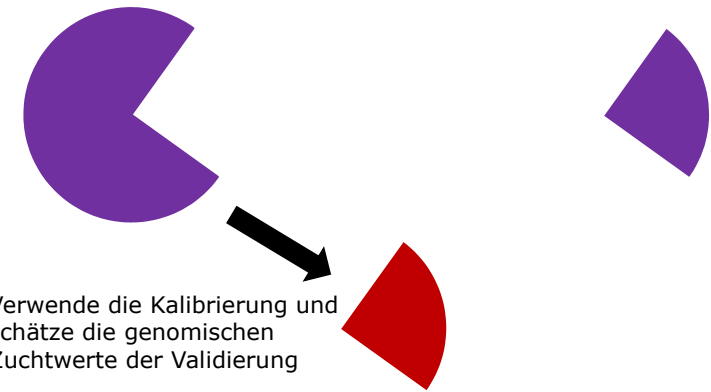
Kreuzvalidierung



Aufteilung **zufällig** (aber bestimmte Anforderungen) und **wiederholt**

Wie weiß man ob das System funktioniert?

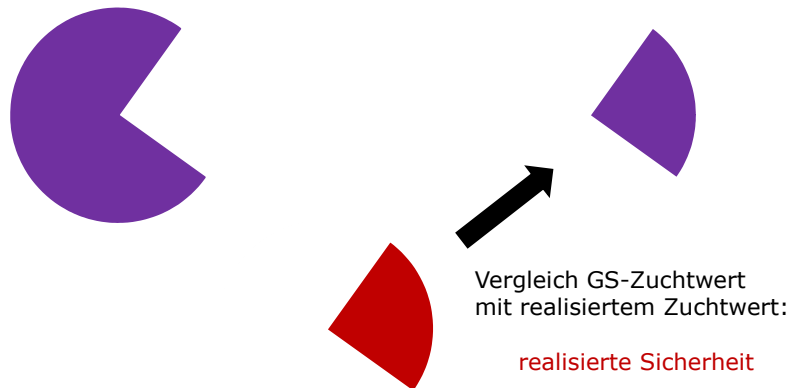
Kreuzvalidierung



Verwende die Kalibrierung und schätze die genomischen Zuchtwerte der Validierung

Wie weiß man ob das System funktioniert?

Kreuzvalidierung

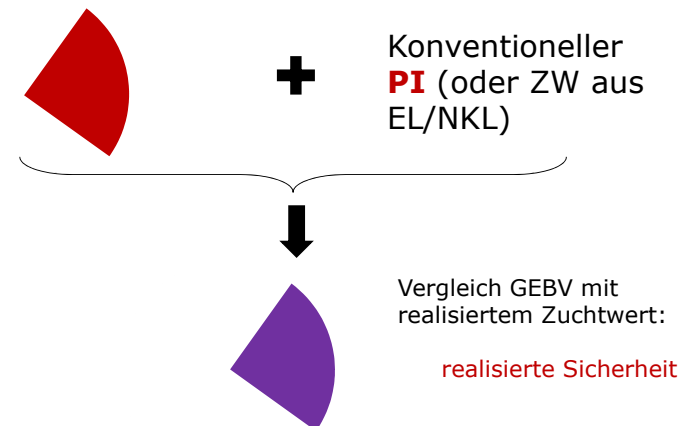


Vergleich GS-Zuchtwert mit realisiertem Zuchtwert:

realisierte Sicherheit

Wie weiß man ob das System funktioniert?

‚Blending‘



Konventioneller **PI** (oder ZW aus EL/NKL)

Vergleich GEBV mit realisiertem Zuchtwert:

realisierte Sicherheit

Wie weiß man ob das System funktioniert?

- Kreuzvalidierung: Aufteilung auf Kalibrierung und Validierung
 - ◆ Zufällig
 - ◆ Alt nach Jung (strikt/weniger strikt)
 - ◆ Mirror- und Forwardprediction
 - ◆ Nach Verwandtschaft gestaffelt
 - ◆ ...

Wie weiß man ob das System funktioniert?

- Erste inoffizielle Ergebnisse beim deutschen Holstein

Merkmal	berechnete Si. gZW	Realisierte Si. gZW
RZM	75 %	68 %
RZS	71 %	54 %
RZE	69 %	40 %
RZN	61 %	51 %
RZR	57 %	41 %
KVm	63 %	42 %
RZG	70 %	54 %

Quelle: VIT Verden, 2009

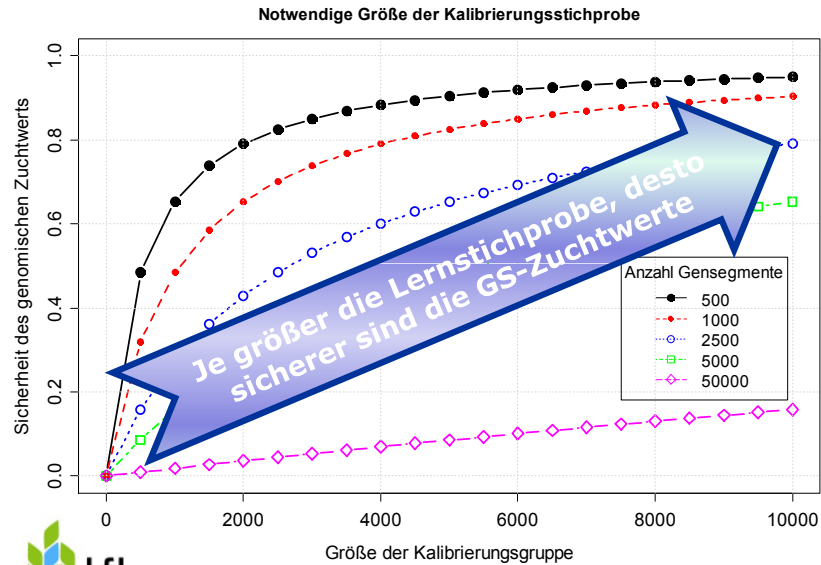
Wie weiß man ob das System funktioniert?

- Fazit
 - ◆ Es gibt bisher kein allgemein anerkanntes Verfahren zur Validierung der Ergebnisse einer genomischen Zuchtwertschätzung
- Beim Vergleich veröffentlichter Zahlen ist zu beachten:
 - ◆ Handelt es sich um theoretisch abgeleitete oder realisierte Werte?
 - ◆ Handelt es sich bei realisierten Werten um Ergebnisse einer Stichprobe (z.B. eines spezifischen Jahrgangs) oder einer wiederholten Stichprobenziehung?
 - ◆ Welche Einteilungskriterien für die Aufteilung in Kalibrierung- und Validierung wurden angelegt?

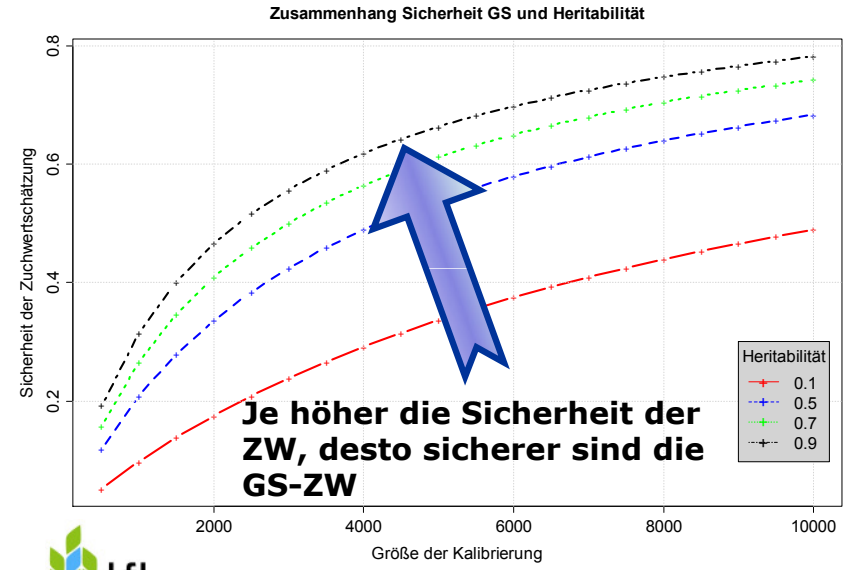
Einflussgrößen auf Güte der GS

- Größe der Kalibrierung
- Effektive Anzahl von Gensegmenten mit Einfluss auf das Merkmal (N_e , genetische Homogenität bzw. Diversität der Population)
- Vorherrschende Familienstrukturen in der Population (-> Zusammensetzung der Lernstichprobe)
- Heritabilität des Phänotypen in der Kalibrierung (=> Sicherheit des Zuchtwertes/der Tochterabweichung)

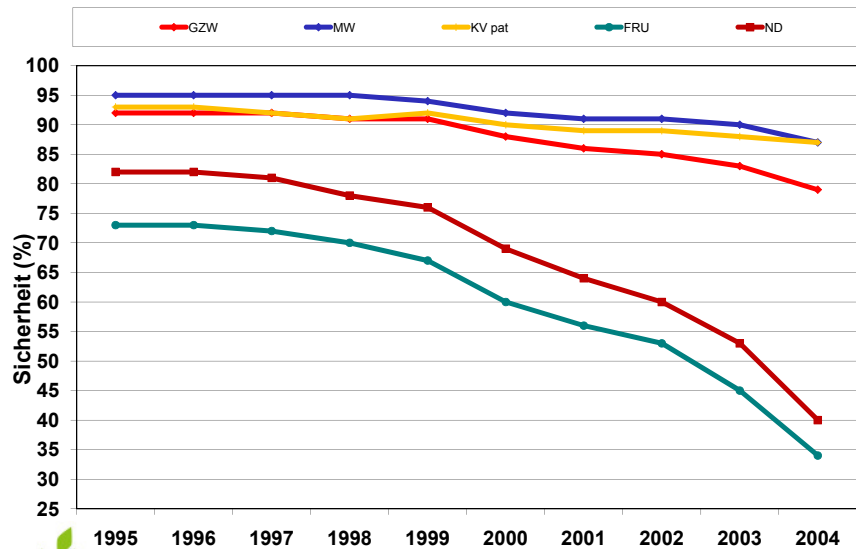
Einflussgrößen auf Güte der GS



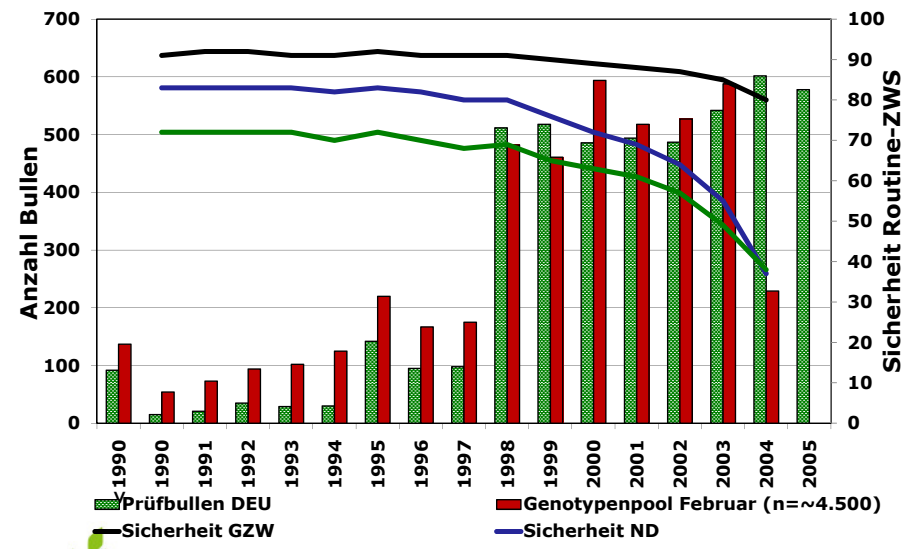
Einflussgrößen auf Güte der GS



Einflussgrößen auf Güte der GS



Einflussgrößen auf Güte der GS



Einflussgrößen auf Güte der GS

- Der von Deutschland und Österreich vereinbarte und nun realisierte Genotypenpool ist absolut notwendig
- Fleckvieh hat nach neueren Schätzungen eine 2 mal größere N_e als DH, welche Konsequenzen ergeben sich hieraus?
- Aus niedrigen Sicherheiten der Phänotypen ergeben sich weitere Probleme für eine Validierung: Gegen was wird verglichen?