

Genomic selection for bacterial cold water disease resistance in rainbow trout using SNP Chip and RAD genotyping platforms with single-step GBLUP and Bayesian variable selection models



Roger Vallejo

**USDA-ARS, National Center for Cool and Cold Water
Aquaculture, Leetown, West Virginia, USA**



Bacterial cold water disease (BCWD)

- High priority disease problem in US trout aquaculture
- Caused by *Flavobacterium psychrophilum* (Fp)
- There is no licensed vaccine available
- Treatment is with antibiotics: Florfenicol & Oxytetracycline
- Antimicrobial resistance is a growing concern



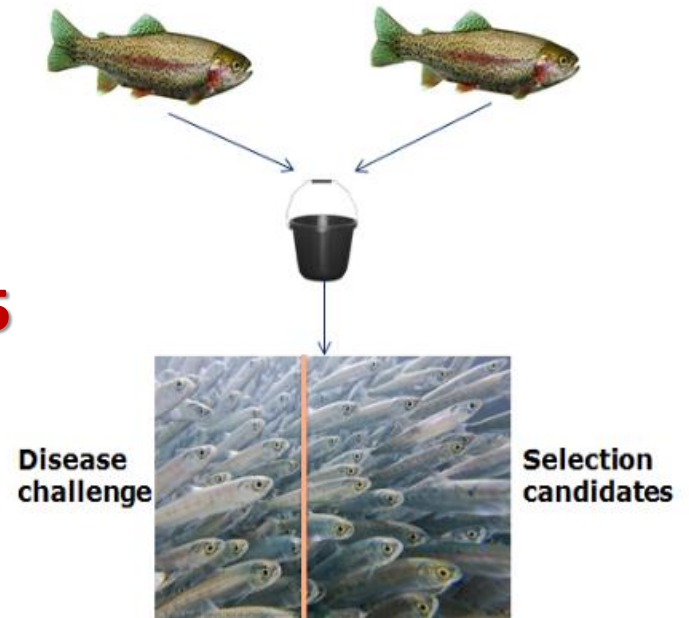
Symptoms of BCWD

- Fry-lethargy, lack of feeding, darkened skin, enlarged spleen, anemia and high mortality



Sib selection in rainbow trout

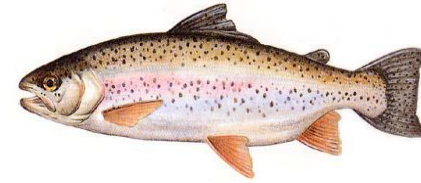
- Many traits selected for breeding cannot be recorded in selection candidates
 - Disease/parasite resistance (BCWD)
 - Meat quality traits (fillet yield)
- Sib selection for such traits is common (**Family selection**)
 - Estimated Breeding Value (EBV) has **limited reliability <0.5**
 - Low selection differential (families \ll individuals)
 - Strong co-selection of family members (Inbreeding)



Genomic selection (GS)

- Large family size is an advantage for GS (>1,000 offspring / family)
- A FS Family split into = TRAINING & TESTING (**High relationship!**)
- Within-family individual selection is possible with GEBVs (**Mendelian sampling**)
- **Reliability of GEBV can be >0.5 !!**

Objectives



1) Predict genomic breeding value (GEBV) for BCWD resistance using 4 GS models:

- ❖ **BayesB**
- ❖ **BayesC**
- ❖ **ssGBLUP**
- ❖ **WssGBLUP**

2) Compare reliability:

- **EBVs vs. GEBVs**
- **Among 4 GS models**
- **Chip vs. RAD**

GS for BCWD resistance in NCCCWA odd-year breeding line

BayesB & BayesC

(Matched sibs with phenotype & genotype records)

ssGBLUP & WssGBLUP

(All sibs and parents with phenotype & genotype records)

Training

(2005 families)

10 fam x 39-80 progeny
 $n = 583$

Genotyped = 583
Phenotyped = 583

10 fam x 39-80 progeny
 $n = 658$

Genotyped = 658
Phenotyped = 627

Validation

(2007 Breeders)

10 fam x 2-11 progeny
 $n = 53$

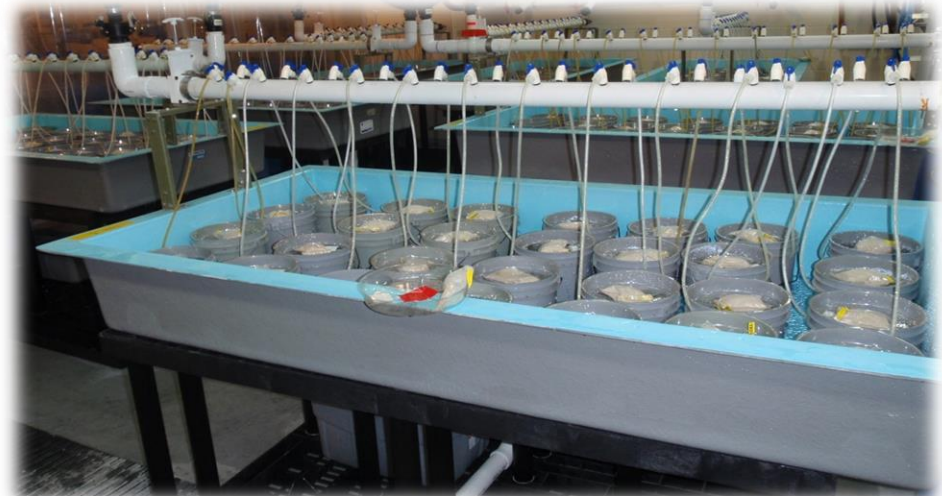
Genotype & EBV
records

10 fam x 2-11 progeny
 $n = 53$

Genotype & EBV
records

BCWD resistance phenotypes

- **Survival days (DAYS)**
 - Number of days to death post-challenge with Fp
 - Fish survival was evaluated for 21 days post-challenge
- **Survival status (STATUS)**
 - 1= fish died during the 21 d post-challenge evaluation
 - 2= fish was alive on day 22 post-challenge





Genotyping Platforms



- SNP 57K chip - Affymetrix: Axiom[®] Trout Genotyping Array (**Chip**)
- SNP 24K Restriction-site associated DNA (**RAD**)

Genotyped SNPs and Animals

SNPs and Animals	Chip	RAD
Total genotyped SNPs	49,468	24,465
Effective SNPs (after QC)	40,744	10,048
Total genotyped animals	658	658
Effective animals (after QC)	658	654
SNP/Sample Calling rate	0.90	0.70



Pedigree & GS models



- Pedigree-based model (**PED**) (MTDFREML; Boldman, 1995)
- Bayesian method **BayesB** (GENSEL; Fernando & Garrick, 2009)
- Bayesian method **BayesC** (GENSEL; Fernando & Garrick, 2009)
- Single-step GBLUP (**ssGBLUP**) (BLUPF90; Misztal et al. 2014)
- Weighted single-step GBLUP (**WssGBLUP**) (BLUPF90; Misztal et al. 2014)

Animal mixed linear model: $y = \mathbf{1}\mu + X\beta + Z\alpha + e$

where y is $n \times 1$ vector of phenotypic values; $\mathbf{1}$ is a vector of ones relating the overall mean μ to the phenotypic records; X is $n \times p$ incidence matrix relating vector β of non-genetic fixed effects to y ; Z is $n \times k$ matrix of genotype covariates (coded as -10, 0, or 10) for k SNP markers, α is $k \times 1$ vector of random partial regression coefficients of k SNPs (additive marker effects), and e is a vector of residuals.

$\alpha \sim N(\mathbf{0}, A\sigma_a^2)$ (PED)

$\alpha \sim N(\mathbf{0}, H\sigma_a^2); H = A + G$ (ssGBLUP)

$\alpha \sim \text{IIM}$ (point mass 0 with π , Univar- t with $1-\pi$ with $\mu = 0$, scale par S_α^2 , \cup df) (BayesB)

Reliability of genomic predictions

Reliability of GEBVs assessed with **EBVs estimated in validation animals** (sibs and progeny records)

Predictive ability: $PA_{GEBV} = r(EBV, GEBV)$

Accuracy: $R_{GEBV} = r(EBV, GEBV) / R_{EBV}$

Reliability: R^2_{GEBV}

Bias: $\beta_{EBV.GEBV}$

($\beta < 1$ up-biased; $\beta > 1$ down-biased GEBV)

GEBVs for DAYS and STATUS using GS models with Chip and RAD

Validation animal	PED-EBV		RAD								Chip							
			DAYS				STATUS				DAYS				STATUS			
	DAYS	STATUS	BayesB	BayesC	ssGBLUP	WssGBLUP	BayesB	BayesC	ssGBLUP	WssGBLUP	BayesB	BayesC	ssGBLUP	WssGBLUP	BayesB	BayesC	ssGBLUP	WssGBLUP
9659	2.58	25.18	0.39	0.06	16.05	-0.86	0.29	0.26	0.50	0.21	1.56	1.33	17.09	0.87	0.61	0.58	0.60	0.54
9665	1.26	17.63	2.10	2.37	19.40	3.76	0.55	0.72	0.85	0.78	2.33	2.01	18.53	2.87	0.73	0.65	0.58	0.75
9675	3.11	27.39	0.98	0.86	14.38	-0.65	0.38	0.34	0.45	0.25	1.26	1.02	14.41	-0.56	0.57	0.50	0.52	0.25
9682	4.97	38.45	2.16	2.25	19.78	3.76	0.25	0.30	0.60	0.55	2.59	2.44	19.04	3.53	0.47	0.42	0.52	0.63
9731	3.32	31.88	0.51	0.47	14.98	0.77	-0.03	-0.18	-0.15	-0.05	1.31	1.07	14.82	0.47	0.16	0.00	0.08	-0.16
9792	3.52	29.59	-0.76	-0.79	12.80	-1.72	-0.05	-0.17	-0.01	-0.23	-1.12	-0.94	11.95	-1.90	-0.21	-0.19	0.01	-0.39
9794	4.43	37.32	2.19	2.59	17.82	3.05	0.30	0.34	0.42	0.36	2.09	2.05	18.52	3.26	0.40	0.37	0.44	0.52
9811	2.83	26.98	-0.58	-0.82	15.43	-1.61	0.01	-0.05	0.18	-0.12	-0.42	-0.33	14.24	-0.60	0.02	0.04	0.15	0.10
9821	5.08	41.11	0.46	0.38	13.66	-0.17	-0.01	-0.28	-0.24	-0.13	1.02	0.85	14.66	0.58	0.21	0.04	0.07	-0.06
9868	2.36	25.6	0.15	0.21	12.99	-1.73	0.19	0.18	0.06	0.13	-0.05	-0.13	13.00	-0.70	0.11	0.02	0.09	-0.04
9968	2.27	23.83	1.06	0.91	16.98	0.31	0.31	0.34	0.43	0.05	1.59	1.34	17.45	0.58	0.36	0.29	0.45	0.13
10035	2.04	11.27	1.85	1.83	18.85	2.47	0.09	0.16	0.48	0.15	2.60	2.32	18.97	2.53	0.33	0.25	0.36	0.28
10069	1.93	21.82	-0.58	-0.77	16.55	-1.00	0.16	0.31	0.54	0.32	-0.55	-0.51	16.17	-1.38	0.11	0.17	0.36	0.23
10104	2.6	17.49	2.46	2.61	19.71	3.72	0.21	0.27	0.33	0.41	2.72	2.92	20.71	3.53	0.25	0.34	0.44	0.49
10110	2.52	19.8	1.80	1.61	17.75	1.18	-0.03	-0.13	0.13	-0.23	1.73	1.94	17.75	1.53	-0.19	-0.09	0.18	-0.34
10183	4.88	41.41	3.46	3.98	21.69	5.65	0.80	1.13	1.13	1.31	3.25	3.13	21.74	5.50	0.88	0.87	0.87	1.49
10188	3.2	30.24	1.71	1.71	17.62	2.30	0.58	0.73	0.79	0.85	2.25	2.17	18.62	2.27	0.62	0.63	0.74	0.88
10189	3.18	26.76	1.65	1.46	18.11	2.11	0.27	0.41	0.65	0.47	1.00	1.01	15.82	0.70	0.22	0.21	0.33	0.22
10192	2.88	28.04	1.97	2.45	17.51	3.40	0.53	0.66	0.44	0.88	1.28	1.12	16.73	1.42	0.36	0.28	0.34	0.36
10201	3.24	31.02	1.36	1.40	19.02	2.54	0.18	0.25	0.53	0.27	1.12	0.87	17.74	1.27	0.23	0.16	0.24	0.32
10212	5.61	45.29	2.07	2.07	19.93	3.91	0.34	0.44	0.76	0.76	2.17	2.34	18.72	3.69	0.35	0.40	0.52	0.67
10216	4.88	42.1	0.95	0.94	15.60	0.71	0.08	-0.01	0.29	0.13	1.06	0.82	15.68	0.86	0.30	0.18	0.25	0.25
10220	4.15	39.94	2.70	2.61	19.99	3.04	0.73	0.85	0.97	0.92	2.48	2.21	18.60	1.76	0.56	0.51	0.63	0.55
10302	2.48	23.05	2.65	2.94	21.27	4.76	0.70	0.98	1.05	0.90	2.10	2.25	19.91	2.85	0.62	0.73	0.73	0.83

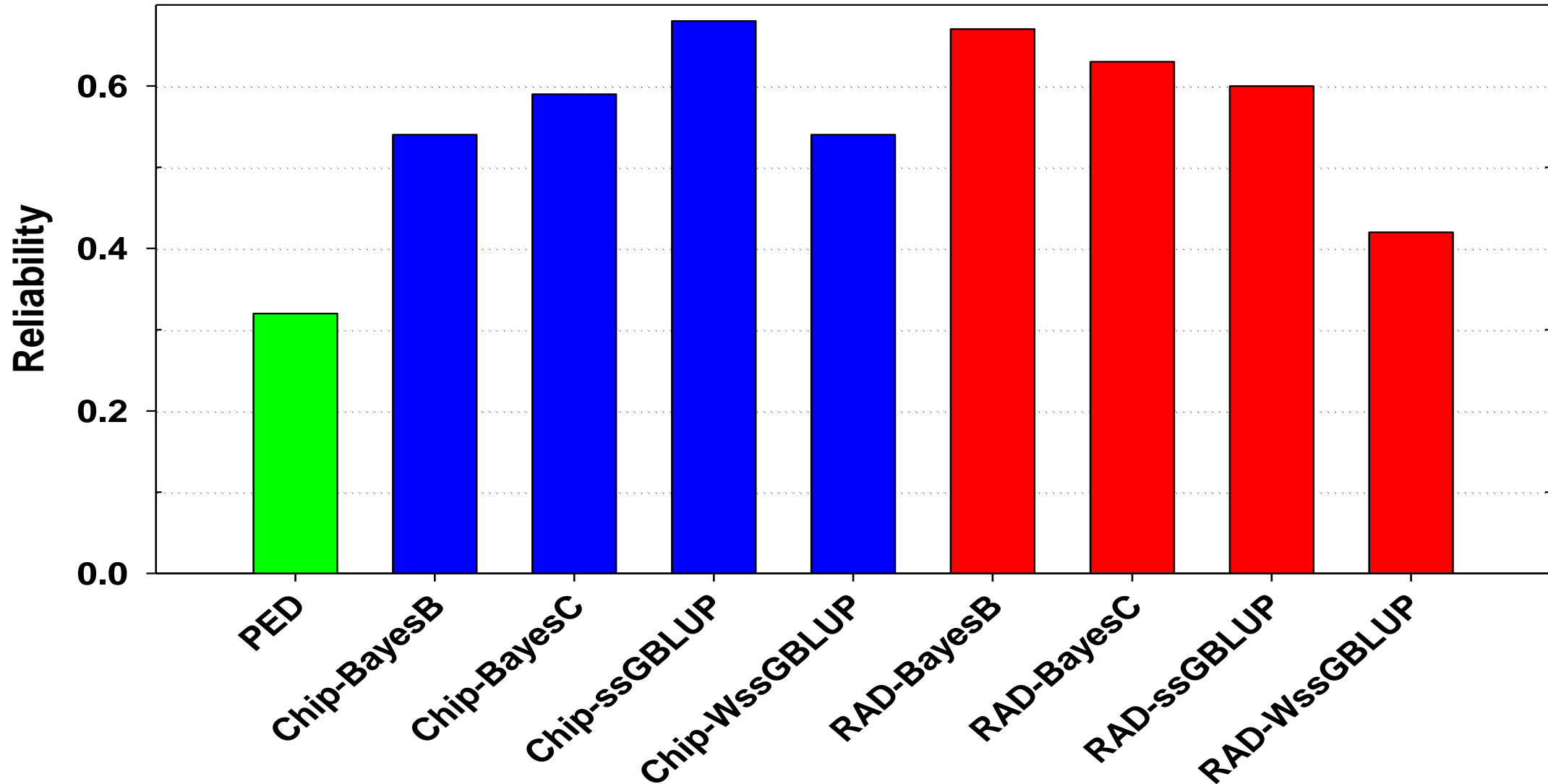
Correlation between GEBVs estimated with GS models

GS model 1 – GS model 2 ²	RAD ³		Chip ³	
	DAYS ⁴	STATUS ⁴	DAYS	STATUS
BayesB – BayesC	0.99	0.98	0.99	0.97
BayesB – ssGBLUP	0.85	0.91	0.84	0.91
BayesB – WssGBLUP	0.88	0.91	0.88	0.87
BayesC – ssGBLUP	0.85	0.95	0.87	0.98
BayesC – WssGBLUP	0.90	0.95	0.91	0.93
ssGBLUP – WssGBLUP	0.92	0.91	0.94	0.92
Mean	0.90	0.94	0.90	0.93

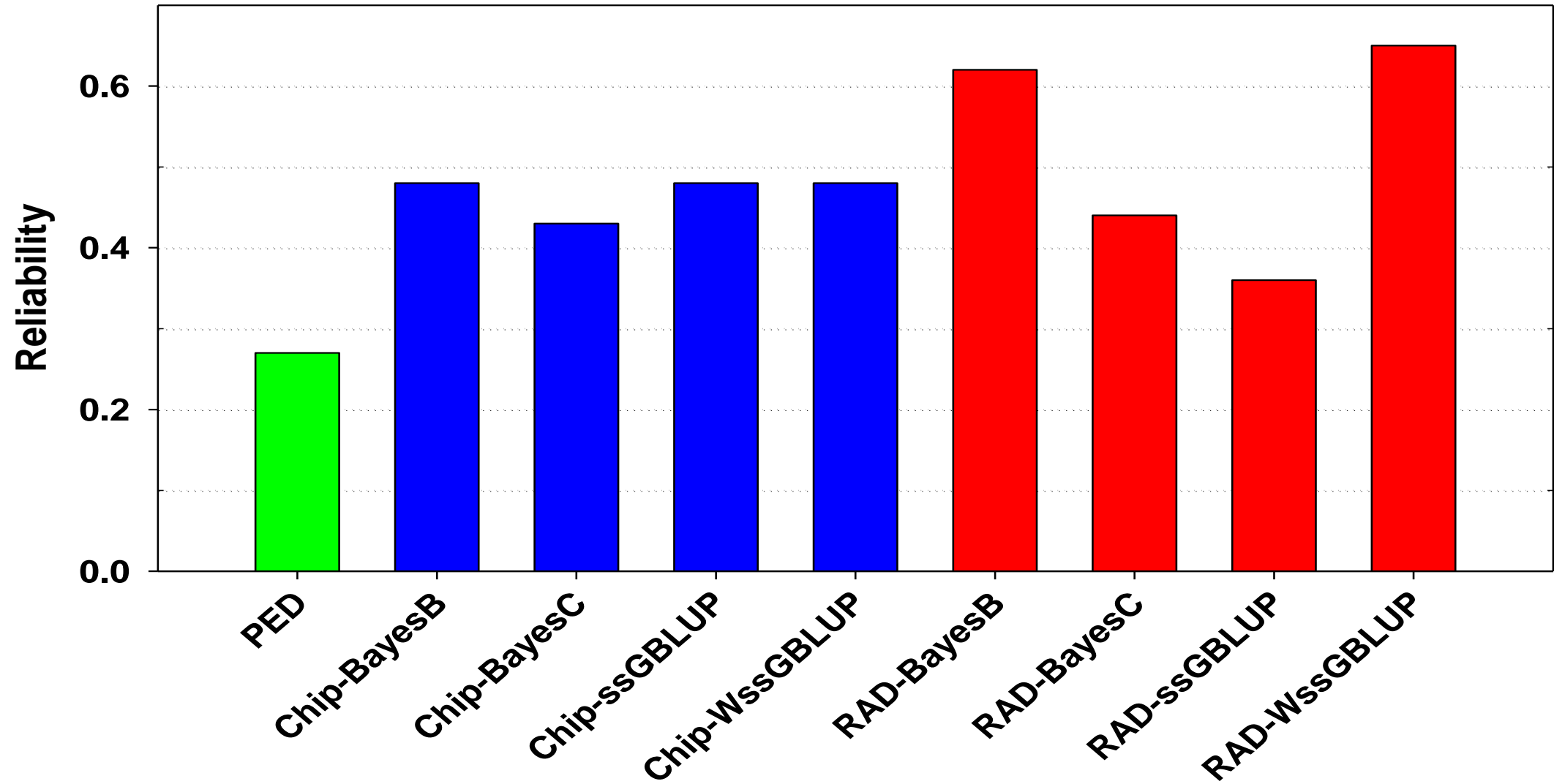
Correlation between EBVs and GEBVs (Predictive ability)

PED – GS model ²	RAD ³		Chip ³	
	DAYS ⁴	STATUS ⁴	DAYS	STATUS
EBV – BayesB	0.47	0.41	0.42	0.36
EBV – BayesC	0.45	0.34	0.44	0.34
EBV – ssGBLUP	0.44	0.31	0.47	0.36
EBV – WssGBLUP	0.37	0.42	0.42	0.36
Mean	0.43	0.37	0.44	0.36

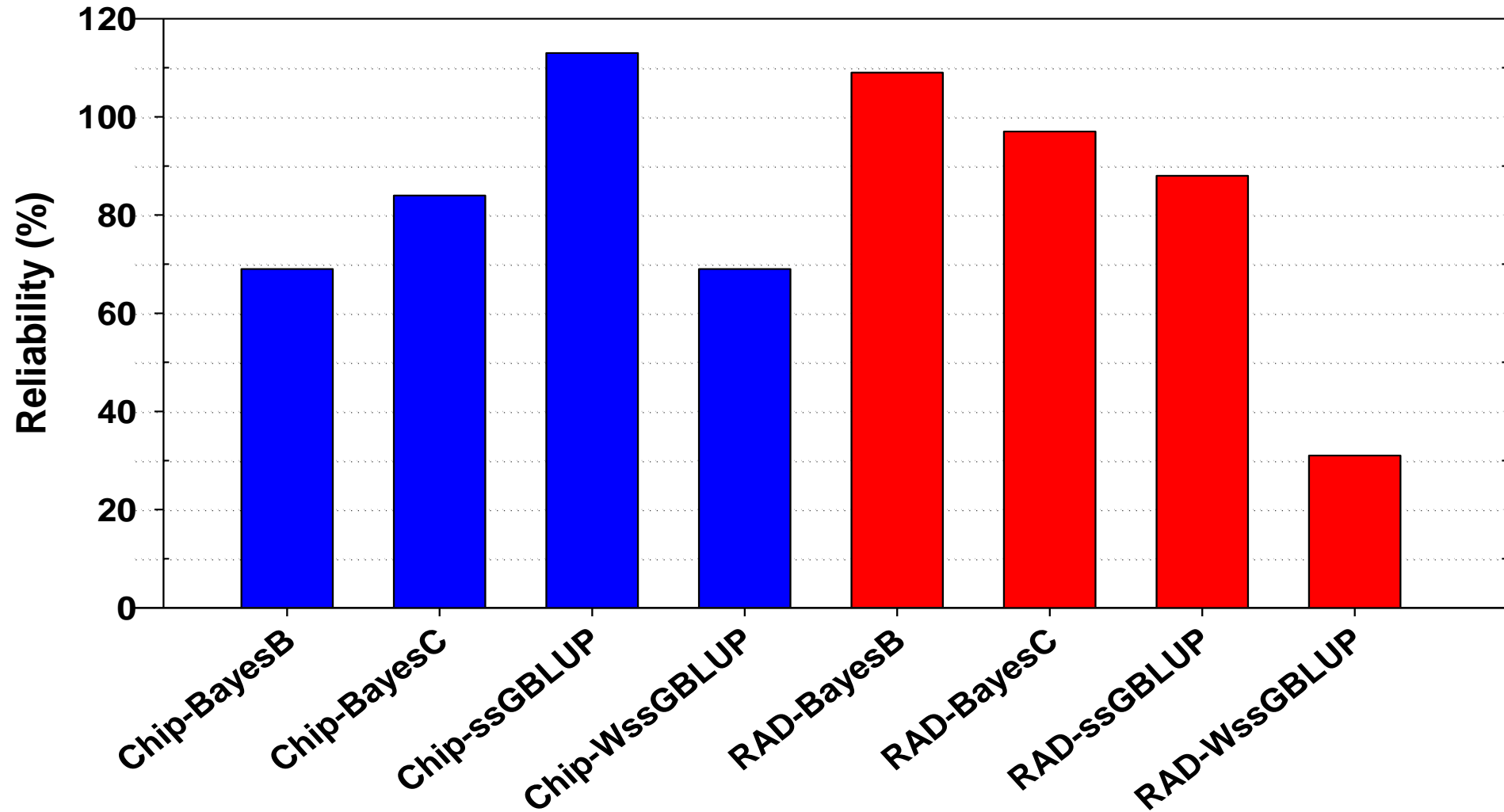
Reliability of genomic predictions for survival DAYS



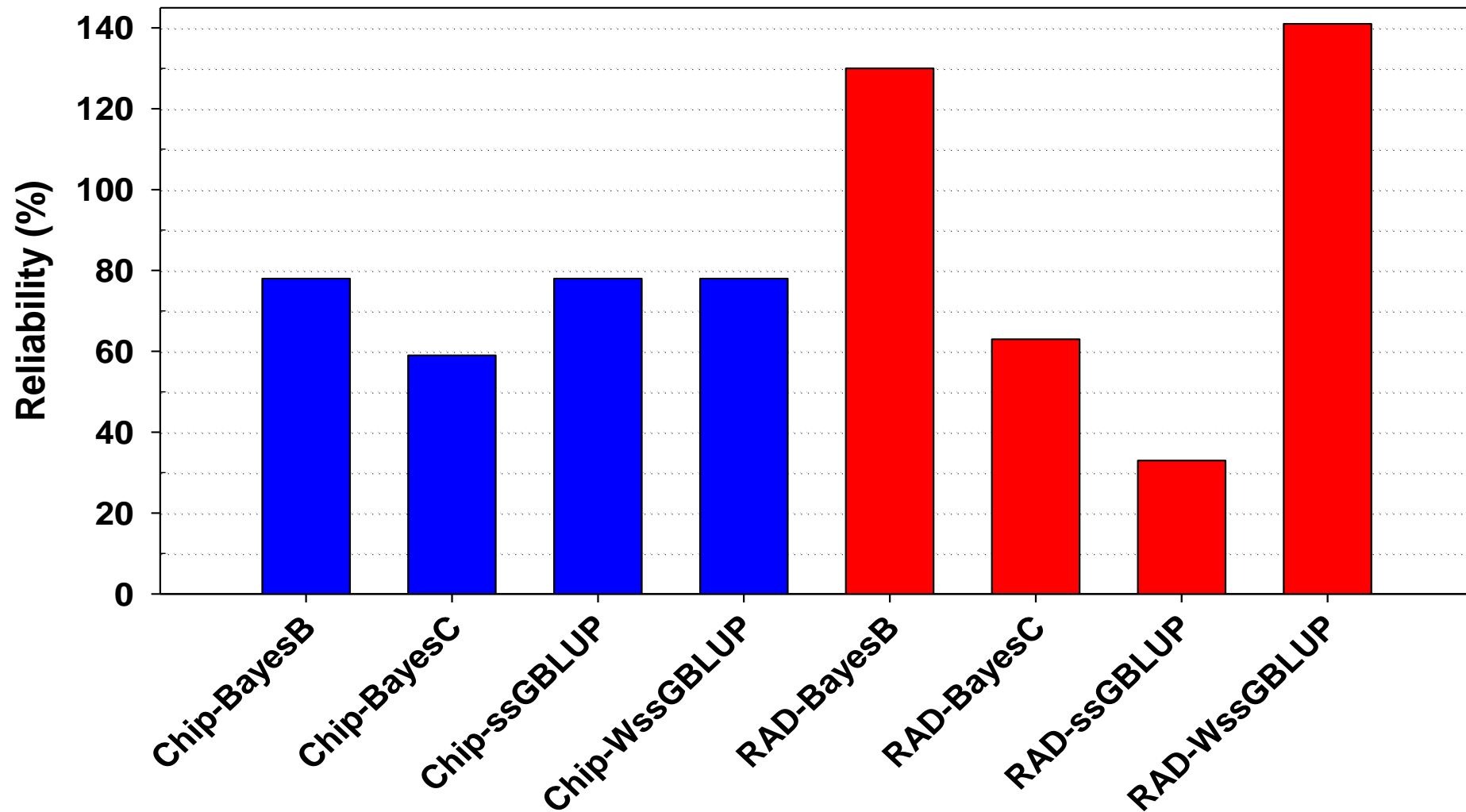
Reliability of genomic predictions for survival STATUS



Relative increase (%) in reliability of GEBVs for survival DAYS with reference to PED model



Relative increase (%) in reliability of GEBVs for survival STATUS with reference to PED model



Conclusions (1)

- 1. The correlation between GEBVs estimated with GS models was very high (0.84-0.99)**
- 2. The correlation between EBVs and GEBVs was moderate (0.31-0.47)**
- 3. The reliability of EBVs was low (DAYS 0.32; STATUS 0.27)**
- 4. The GEBVs for DAYS (.42-.68) had higher reliability than STATUS (.36-.65)**

Conclusions (2)

- 5)** The relative increase in reliability using GS models was highest for BayesB (97%) followed by WssGBLUP (80%)
- 6)** RAD platform (~10K SNPs) was as efficient as SNP Chip (~40K SNPS). However, SNP Chip is higher throughput and more practical for large scale GS studies and currently is as cost-effective as RADs
- 7)** The regression coefficients of EBVs on GEBVs were different than 1.0 suggesting the GEBVs are up- (DAYS) and down-biased (STATUS)
- 8)** Increasing training sample size would be expected to increase even further the reliability of GEBVs in rainbow trout

Acknowledgements

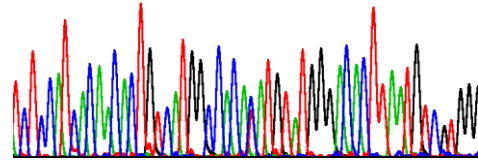
NCCCWA PIs

- Yniv Palti
- Tim Leeds
- Sixin Liu
- Guangtu Gao
- Tim Welch
- Greg Wiens

Laboratory Mol. Biol.

- Roseanna Long
- Kristy Shewbridge

20 220 240 250 260 270
TCTCCATCACACACACTGTGTCTGGCCACAAATGTATGGGACACCTTAAATGCTGTGGG



University of Georgia

- Ignacy Misztal
- Breno Fragomeni

Iowa State University

- Dorian Garrick
- Rohan Fernando



Fish Caretakers

- Jen Lipscomb
- James Everson
- David Payne
- Josh Kretzer

