Genome-wide association study of survival in acoustically tagged steelhead smolts in Puget Sound

Kenneth I. Warheit[1], Megan E. Moore[2], Barry A. Berejikian[2]

[1]WDFW, Olympia WA, [2]NMFS NOAA, Port Orchard, WA

2016 Pacific Coast Steelhead Management Meeting

ASILOMAR CONFERENCE GROUNDS
PACIFIC GROVE, CALIFORNIA
MARCH 9, 2016

Washington Department of FISH and WILDLIFE

SALISH SEA
MARINE SURVIVAL PROJECT
www.marinesurvivalproject.com

NOAA FISHERIES

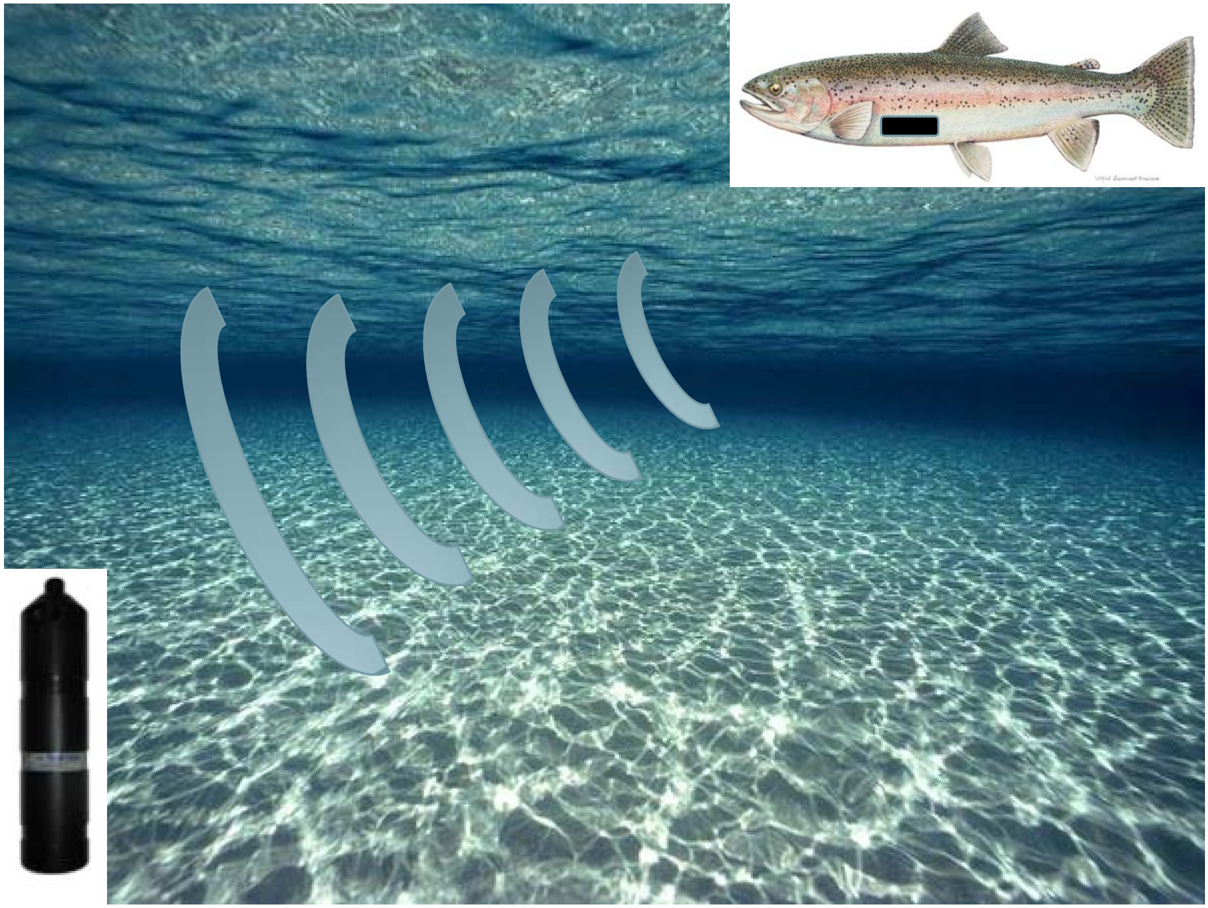
Definitions

- Genome-wide association study (GWAS)
  - Correlation between genotype – genetic “fingerprint” (e.g., SNPs) and phenotype (e.g., QTLs, fate – survival)
- Steelhead
- Smolt

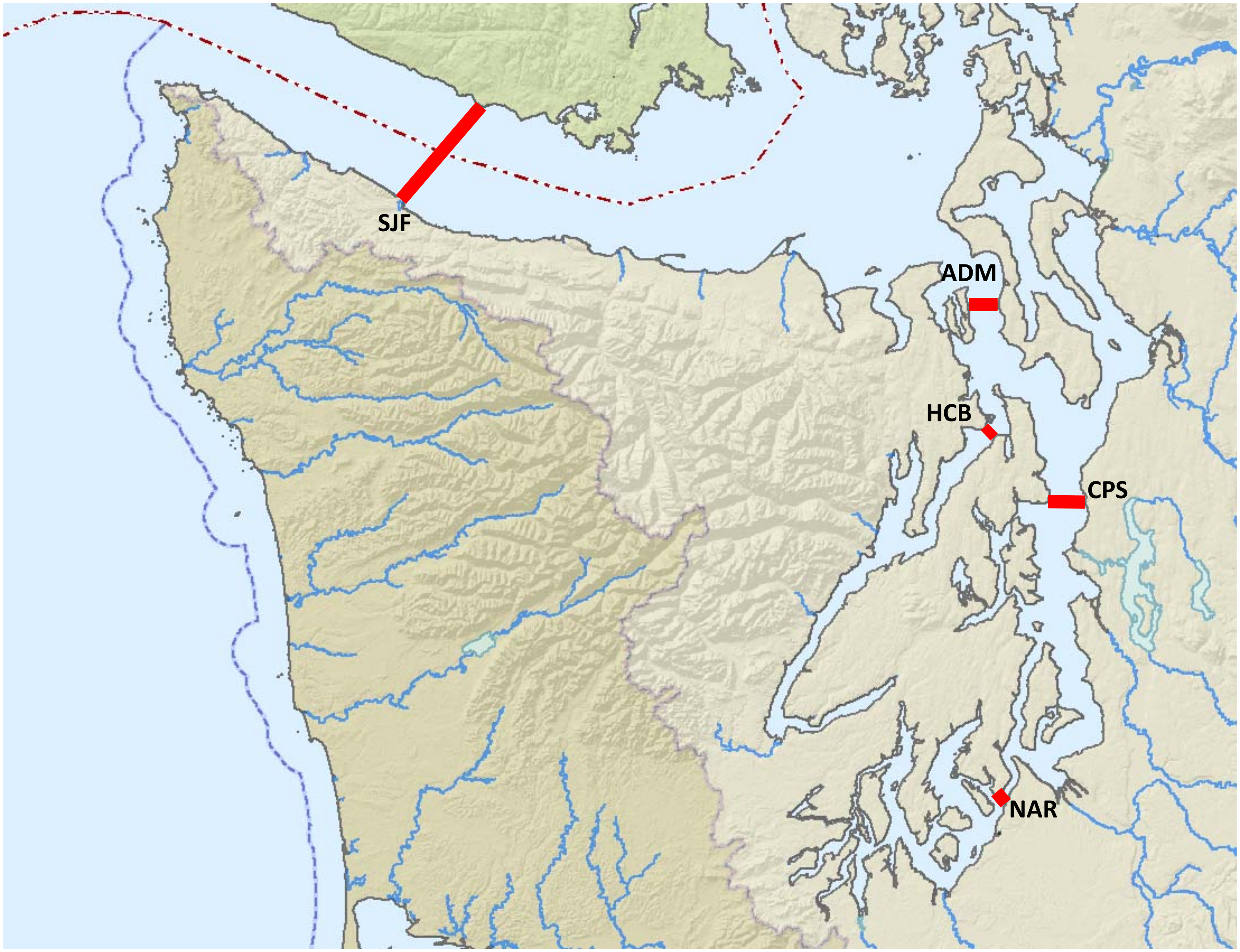
GWAS between smolt survival while out-migrating through Puget Sound, WA and smolt’s genome (reduced-representation of genome)

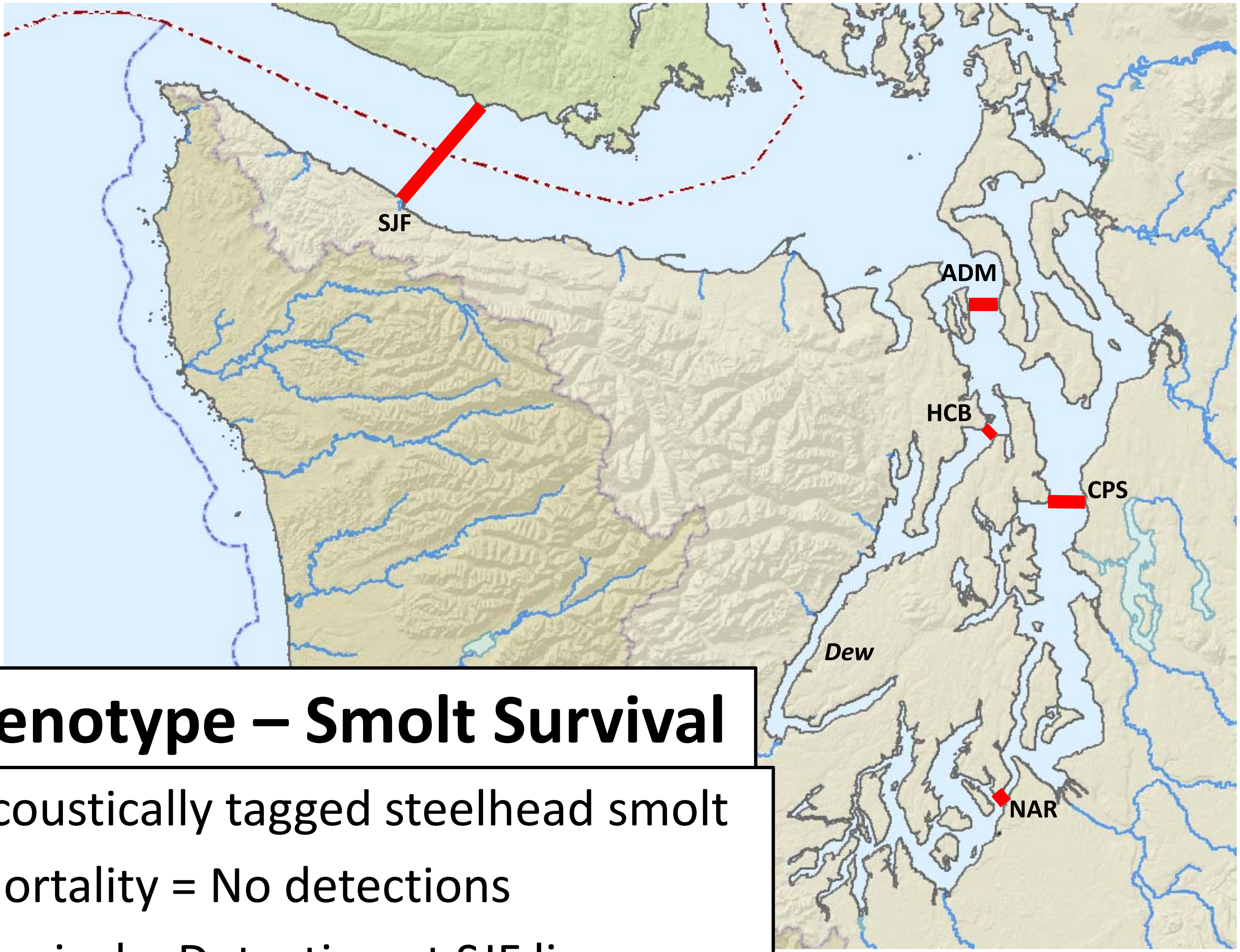
Phenotype – Smolt Survival

- Acoustically tagged steelhead smolt



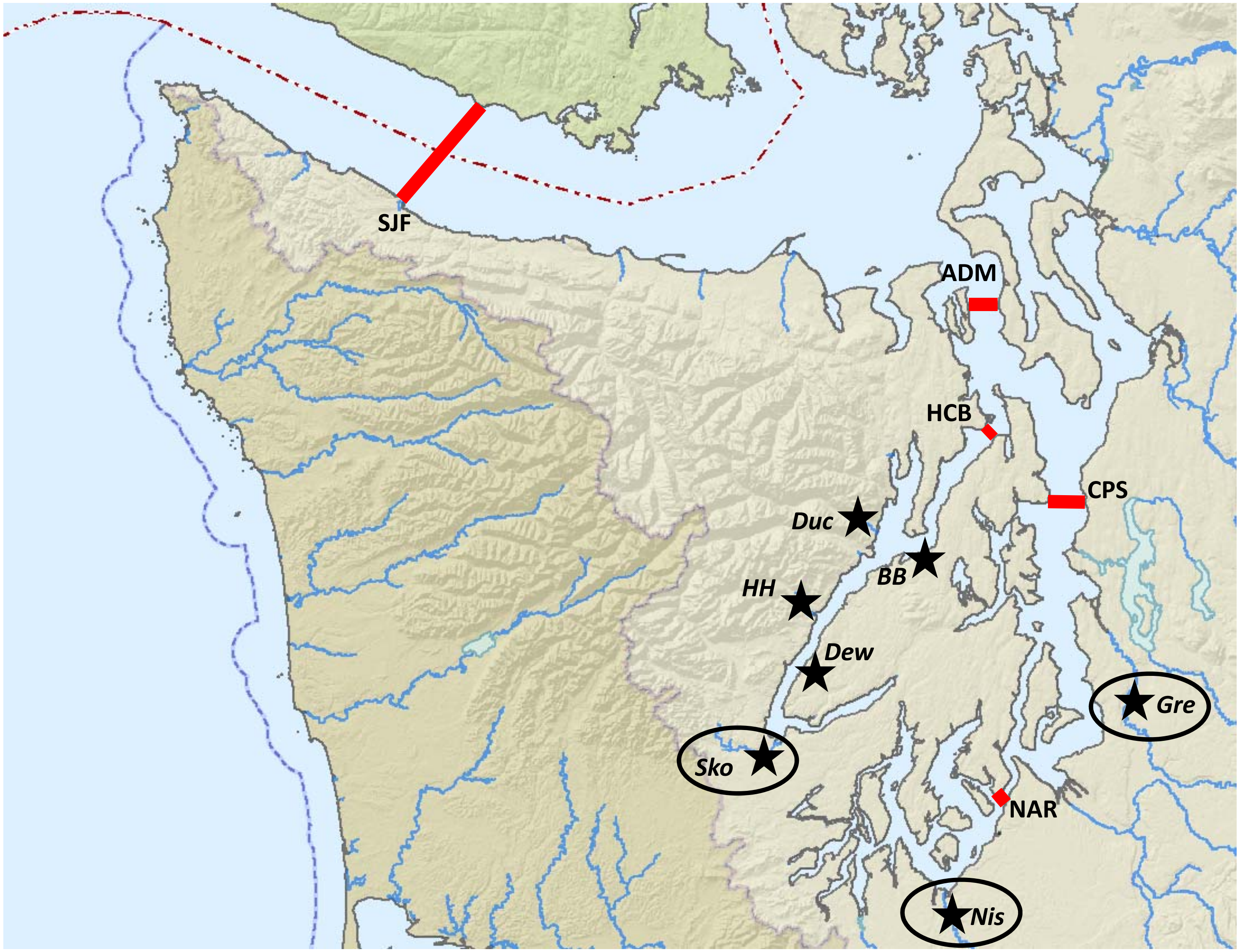
7mm and 9 mm transmitters
@ 69kHz, 136 db


SJF
ADM
HCB
CPS
NAR



Phenotype – Smolt Survival

- Acoustically tagged steelhead smolt
- Mortality = No detections
- Survival = Detection at SJF line


SJF
ADM
HCB
CPS
Duc
BB
HH
Dew
Gre
Sko
NAR
Nis

Source	Year	Origin	Release	Mortality	Survival	Total
Skokomish	2006	N	Skokomish	5	6	11
Skokomish	2007	N	Skokomish	5	5	10
Skokomish	2008	N	Skokomish	8	4	12
Skokomish	2009	N	Skokomish	4	3	7
Skokomish	2010	N	Skokomish	7	0	7
Green	2014	U	Nisqually	4	2	6
Green	2014	U	Green	16	7	23
Green	2008	N	Green	7	2	9
Nisqually	2014	U	Nisqually	2	2	4
Nisqually	2014	U	Green	12	3	15
				70	34	104

RAD-seq

- Restriction-site associated DNA (RAD) (sbfI)
- Used to discover 1000s – 10,000s SNPs
- Distributed throughout the genome
- 80 bp fragment

Category	**# Loci**
Initial set of sequences	~650 million
Catalog 1 (m5M2)	109,571
Catalog 2 (m3M4)	137,521
Overlap, filters	5,702

GWAS Methods

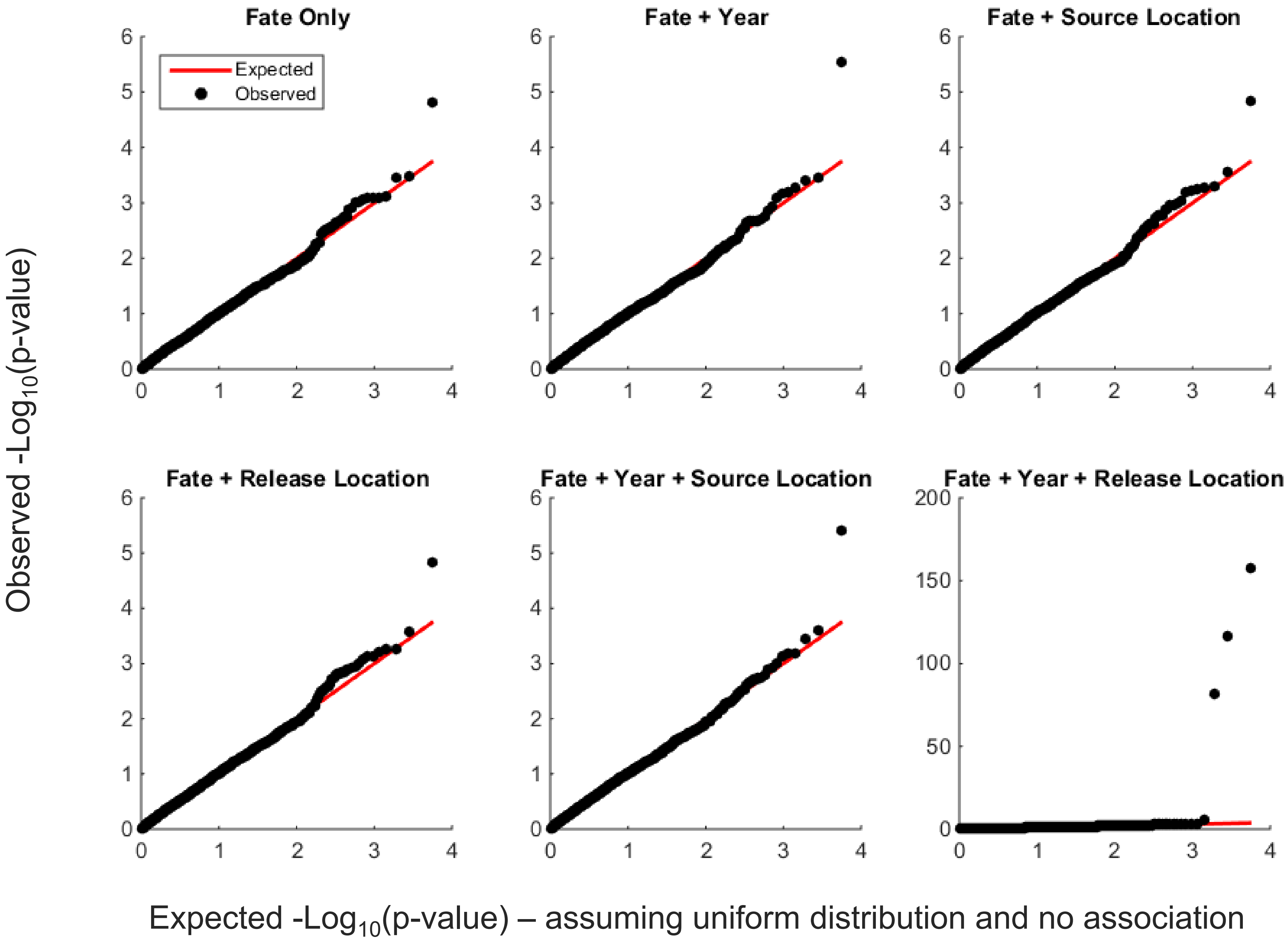
- Mixed linear model (MLM) implemented in program *TASSEL*
- Population structure and kinship included to eliminate false association and reduce Type 1 error
  - Population structure: program *STRUCTURE*
  - Kinship (relatedness): *TASSEL*
- Grouping factors as covariates
  - Year, source location, release location
- Genotype: 5702 SNP loci
- Phenotype: Fate (survival v. mortality)

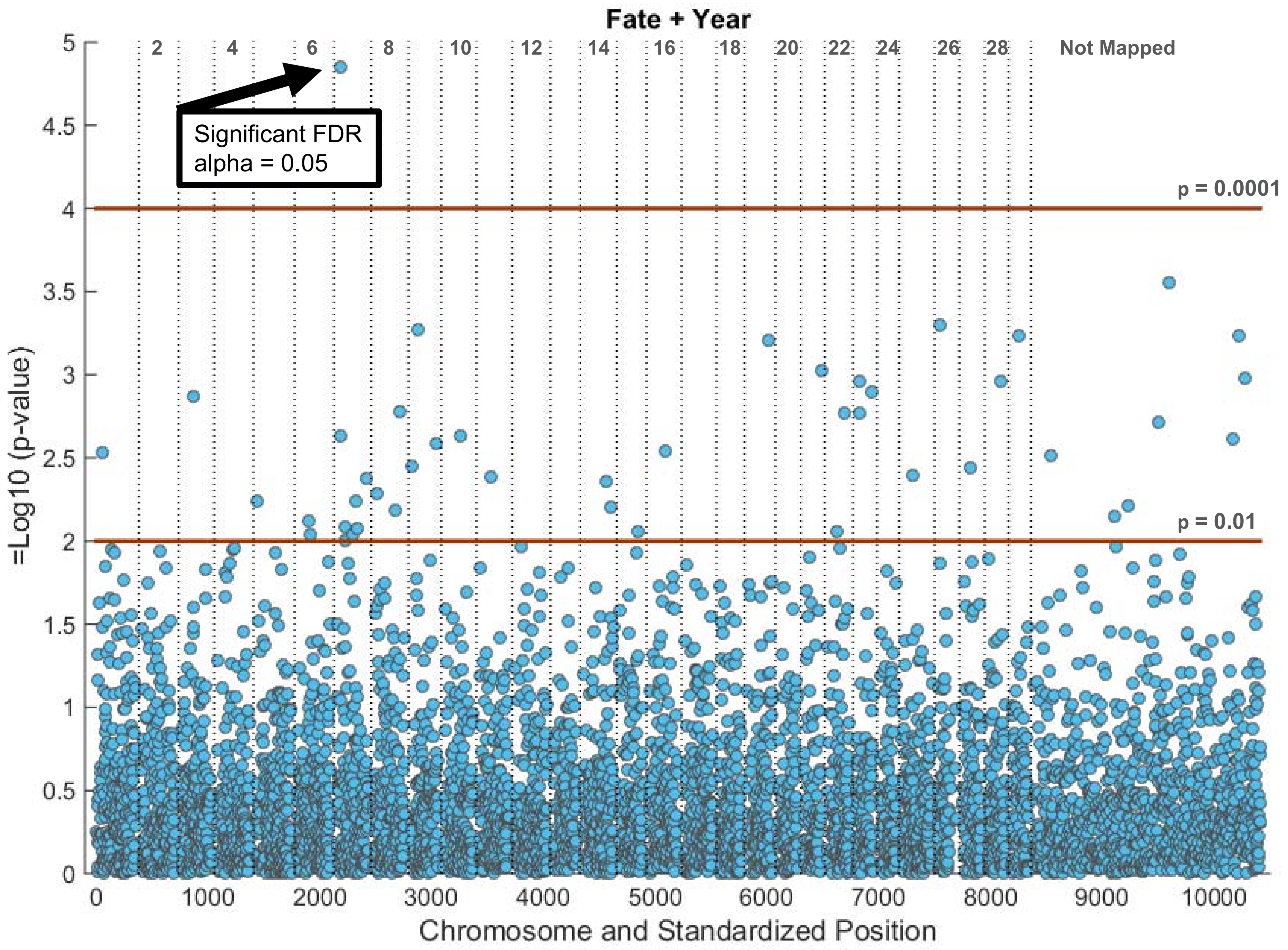
Calculate probability for each SNP H_o: no association

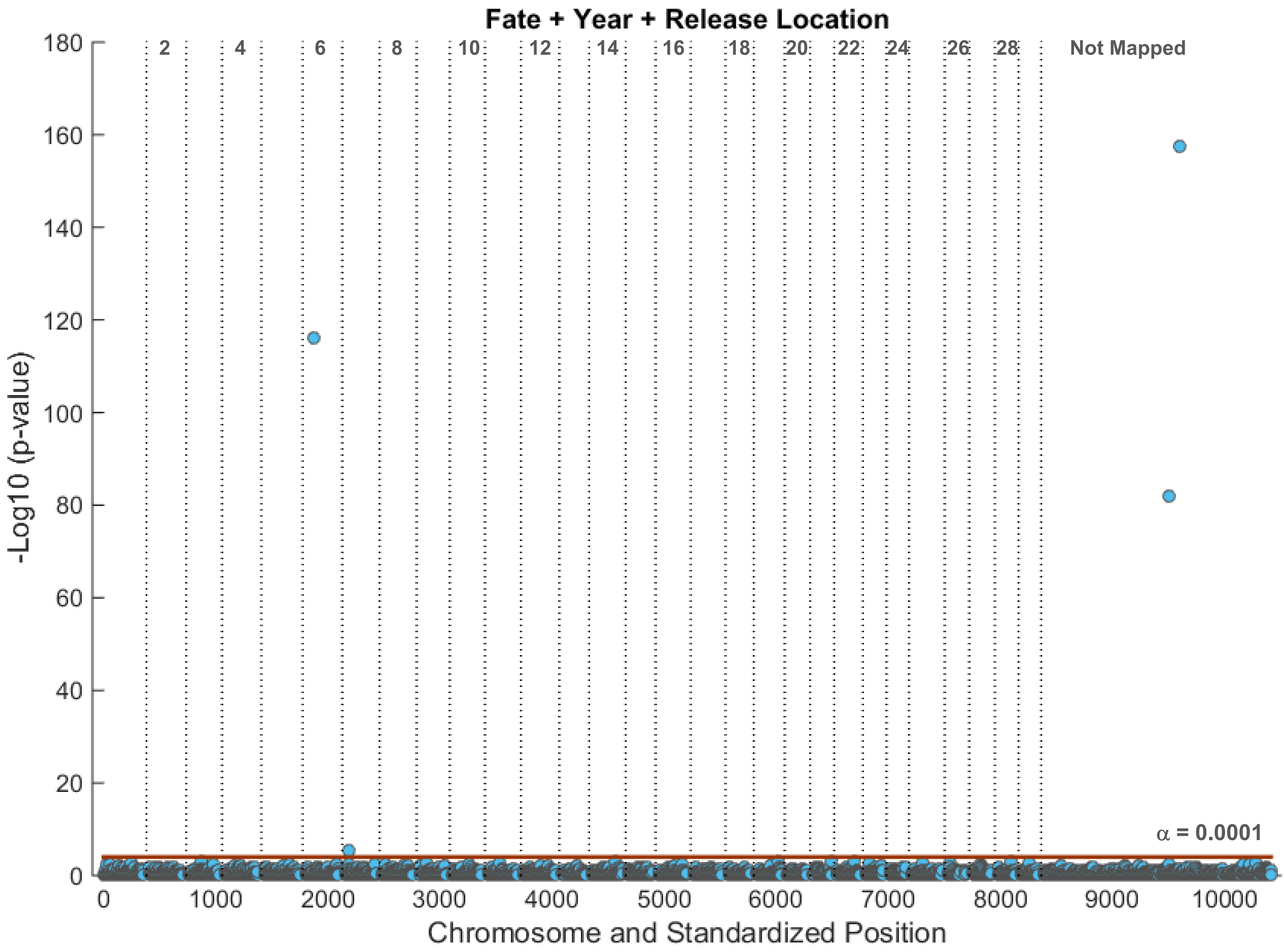
Grouping Factors

- Fate only (no grouping factor)
- Fate + Year
- Fate + Source
- Fate + Release
- Fate + Year + Source
- Fate + Year + Release
- Fate + Year + Source + Release[1]

[1] No results: design matrix not full column rank


Fate Only
Fate + Year
Fate + Source Location
Fate + Release Location
Fate + Year + Source Location
Fate + Year + Release Location
Expected
Observed
0
1
2
3
4
5
6
50
100
150
200
Observed -Log$_{10}$(p-value)
Expected -Log$_{10}$(p-value) – assuming uniform distribution and no association


Fate + Year
2
4
6
8
10
12
14
16
18
20
22
24
26
28
Not Mapped
Significant FDR
alpha = 0.05
p = 0.0001
p = 0.01
=Log10 (p-value)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Chromosome and Standardized Position



Fate + Year + Release Location
2
4
6
8
10
12
14
16
18
20
22
24
26
28
Not Mapped
α = 0.0001
-Log10 (p-value)
0
20
40
60
80
100
120
140
160
180
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000
Chromosome and Standardized Position

Phenotype (MLM)	Locus			
	39529_18	55970_7	12301_21	51226_71
Fate	1.54E-05	3.49E-04	5.07E-01	9.14E-02
Fate + Year	**2.92E-06**	2.16E-03	3.96E-01	1.40E-01
Fate + Source	1.42E-05	2.80E-04	4.74E-01	1.51E-01
Fate + Release	1.46E-05	2.63E-04	5.00E-01	1.40E-01
Fate + Year + Source	**3.97E-06**	2.53E-03	3.74E-01	1.21E-01
Fate + Year + Release	**4.18E-06**	**3.62E-158**	**6.63E-117**	**9.44E-83**
Blast to N Sequences	2	0	5	14
Morphogenesis	1	0	0	1
Immunological	0	0	0	6
Other	1	0	5	7

Bold = significant using FDR (Benjamini – Hochberg) at alpha = 0.05

Bonferroni @ alpha 0.05 = 8.8E-6

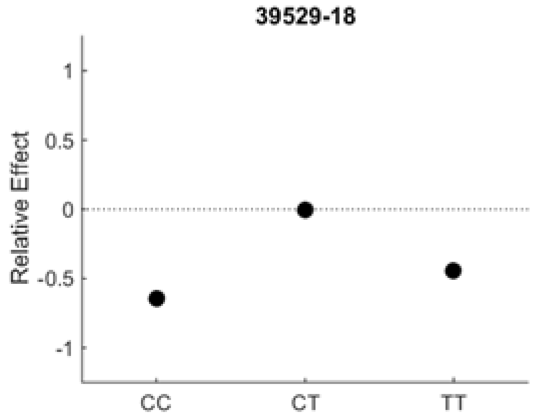
Bonferroni @ alpha 0.10 = 1.8E-5

Phenotype (MLM)	Locus			
	39529_18			51226_71
Fate	1.54E-05			9.14E-02
Fate + Year	**2.92E-06**			1.40E-01
Fate + Source	1.42E-05			1.51E-01
Fate + Release	1.46E-05			1.40E-01
Fate + Year + Source	**3.97E-06**			1.21E-01
Fate + Year + Release	**4.18E-06**			**9.44E-83**
Blast to N Sequences	2			14
Morphogenesis	1			1
Immunological	0			6
Other	1			7

Bold = significant using FDR (Benjamini – Hochberg) at alpha = 0.05

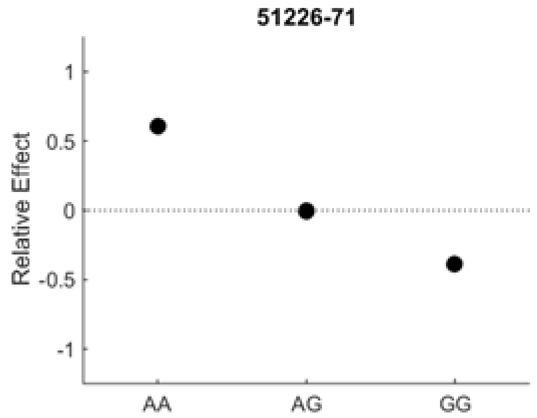
Bonferroni @ alpha 0.05 = 8.8E-6

Bonferroni @ alpha 0.10 = 1.8E-5



Blast to N Sequences	2
Morphogenesis	1
Immunological	0
Other	1

Basin	Fate	CC	CT	TT	N	*G*-test *P*-value
Green	Mortality	0.19	0.16	0.65	31	0.019 ✓
	Survival	0.00	0.80	0.20	10	
Nisqually	Mortality	0.50	0.17	0.33	6	0.017 ✓
	Survival	0.00	1.00	0.00	4	
Skokomish	Mortality	0.04	0.30	0.67	27	0.050
	Survival	0.00	0.56	0.44	18	



Blast to N Sequences	14
Morphogenesis	1
Immunological	6
Other	7

Basin	Fate	AA	AG	GG	N	*G*-test *P*-value
Green	Mortality	0.52	0.28	0.21	30	0.028 ✓
	Survival	0.78	0.22	0.00	9	
Nisqually	Mortality	0.57	0.00	0.43	6	0.028 ✓
	Survival	1.00	0.00	0.00	2	
Skokomish	Mortality	0.78	0.13	0.09	23	0.048
	Survival	0.83	0.17	0.00	12	

0.05 / 3 = 0.0167
0.10 / 3 = 0.0333

Conclusions

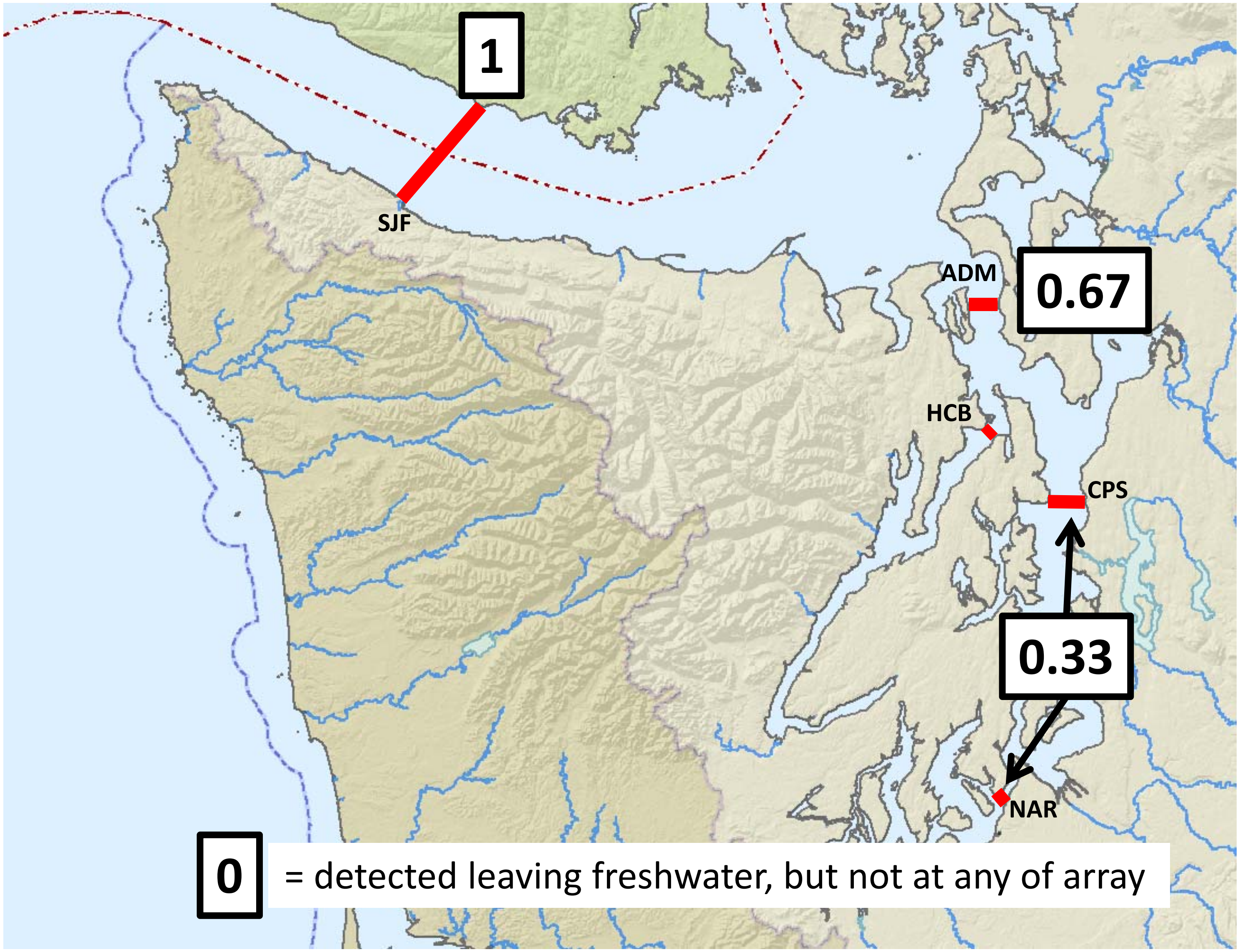
- The fate of out-migrating steelhead smolts is not independent of their genomes
- Two groups of genes appear to have significant association with survival
  - Immunological
  - Developmental
- May depends on time (year) and space (release location)
- Working hypothesis (w/ some arm-waving): Smolts with certain alleles may be compromised by their immunological response or fin development
- Predation, for example, maybe the proximate cause of mortality; but ultimate cause may be fish health and presence of pathogen(s)

Limitations

- Sample size (total and per stratum)
  - Power
  - Prevents post hoc tests
- Lack of independence:
  - between year and source location
  - Between source and release locations
- Design: model is for QTLs, but fate is categorical
- Short fragments (limits annotation)

Next Steps (2016 funding)

- Analysis based on 2014 & 2015 samples from Green and Nisqually rivers
- RAD-seq samples not originally sequenced
- Change fate from binary character to quasi-continuous



0 = detected leaving freshwater, but not at any of array

Next Steps (2016 funding)

- Analysis based on 2014 & 2015 samples from Green and Nisqually rivers
- RAD-seq samples not originally sequenced
- Change fate from binary character to quasi-continuous
- **Sample size increases from 59 to 282**
- Explore other analysis methods (e.g., Random Forest)
- GWAS with phenotype = # of *Nanophyetus* cysts

Acknowledgements

- This study is part of the Salish Sea Marine Survival Project. Funding was provided by Washington State
- Puget Sound Steelhead Marine Survival Workgroup for their support in development and implementation.
- Michael Schmidt (LLTK) for intrepid shepherding of the Salish Sea Marine Survival Project (and his herding of cats)
- Amelia Whitcomb (WDFW) for laboratory work and protocol development
- Sewall Young, Todd Seamons, Mo Small (WDFW) for bioinformatics and GWAS discussions
- Michael Schmidt, Todd Kassler (WDFW), Neala Kendall (WDFW) for contract management
- Field crews from NMFS, WDFW, USAC, Muckleshoot and Nisqually Tribes