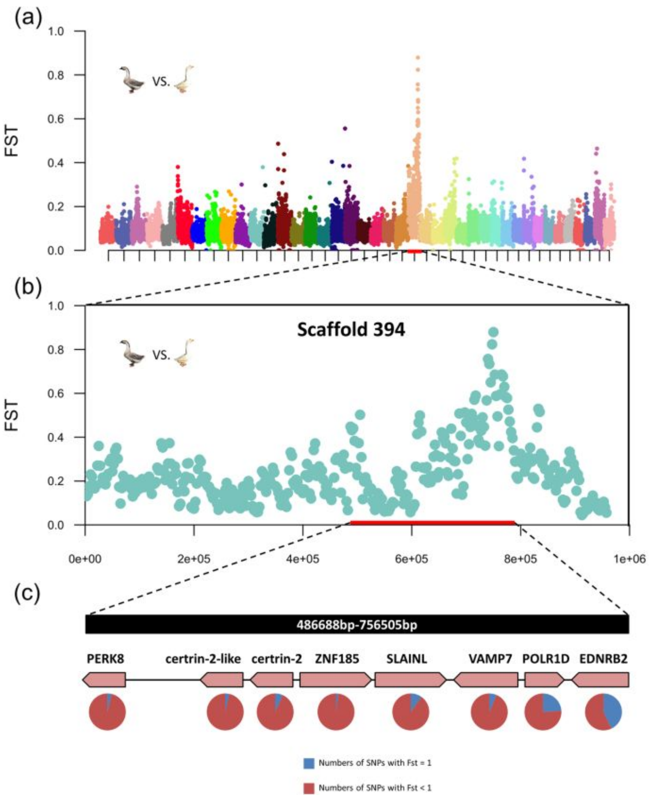
Detecting Selection Using Genomic Data: Methods and Characteristics

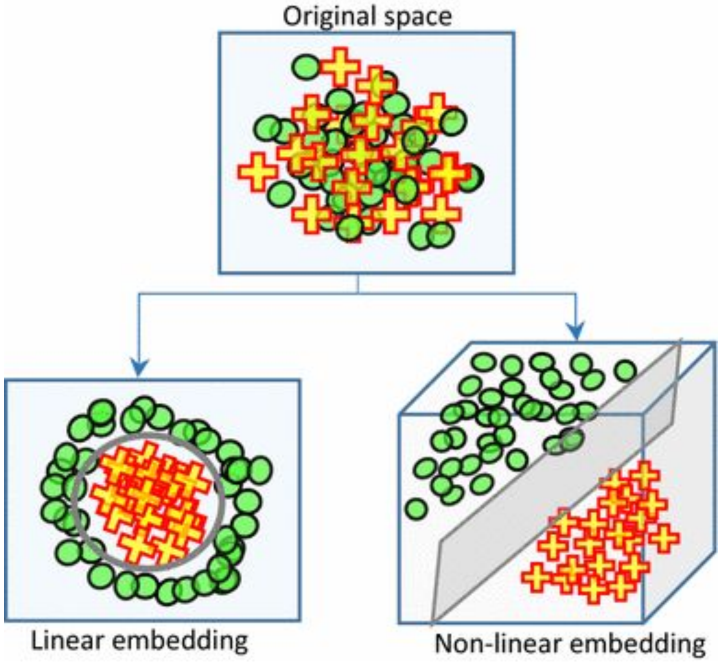
November 8, 2023

Fst statistics



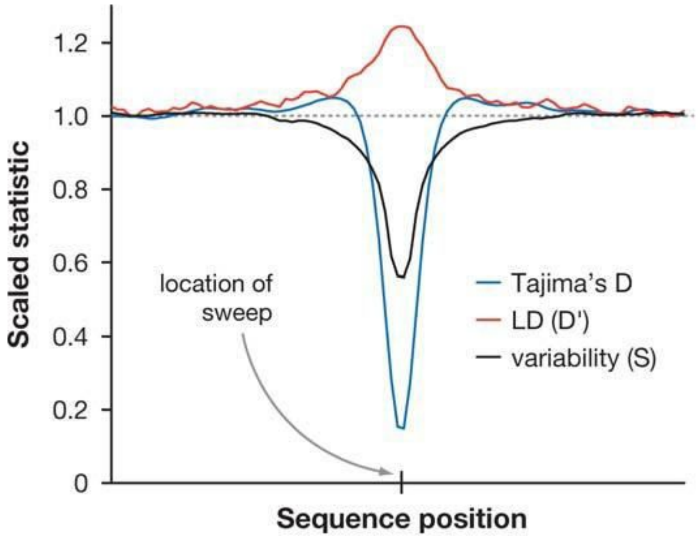
- **Ideal Situation:** Well-suited for studying populations with clear geographical or ecological boundaries.
- **Strengths:** Provides a straightforward measure of genetic differentiation between populations. Robust and widely applicable.
- **Weaknesses:** Sensitive to marker choice, may miss subtle population structure nuances.

Spatial FST (SFS)



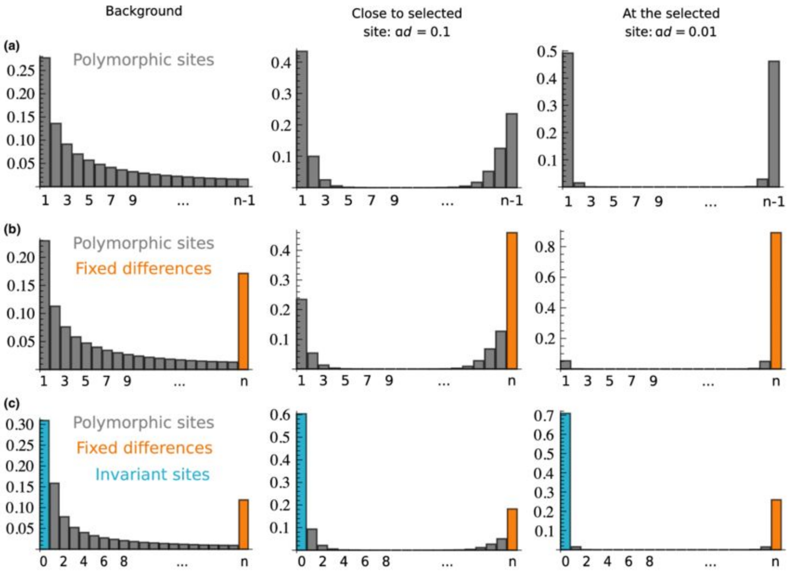
- **Ideal Situation:** Investigating the geographical patterns of genetic variation in a species with known migration corridors or barriers, such as a river or mountain range.
- **Strengths:** Incorporates geographical information, providing insights into spatial patterns of genetic differentiation.
- **Weaknesses:** Sensitivity to the choice of the kernel and potential bias introduced by uneven sampling across space.

Tajima's D



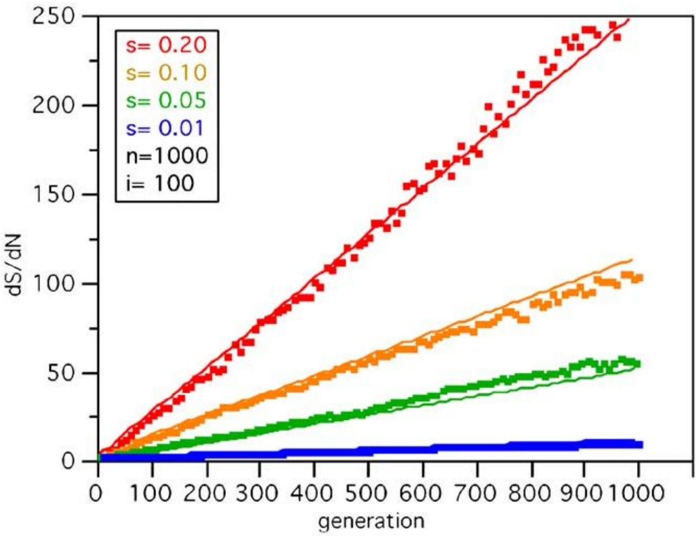
- **Ideal Situation:** Assessing signatures of natural selection in a population with a recent demographic change, like a population recovering from a bottleneck.
- **Strengths:** Sensitive to selection and demographic events, provides information about the frequency spectrum of mutations.
- **Weaknesses:** Susceptible to population structure and demographic history effects.

Site Frequency Spectrum (SFS)



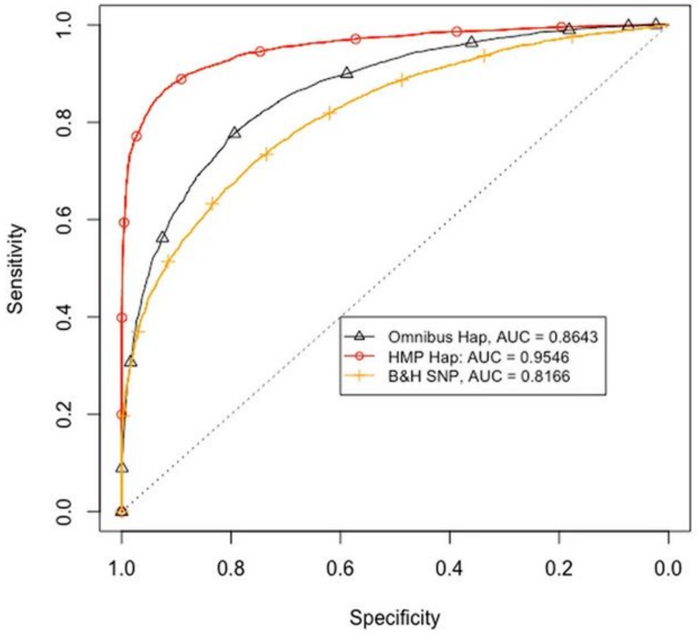
- **Ideal Situation:** Analyzing a population with a known demographic history, such as recent expansion or contraction, to understand the impact on allele frequencies.
- **Strengths:** Directly examines allele frequency distributions, useful for detecting demographic changes.
- **Weaknesses:** May not distinguish between different evolutionary processes affecting allele frequencies.

dN/dS Ratio



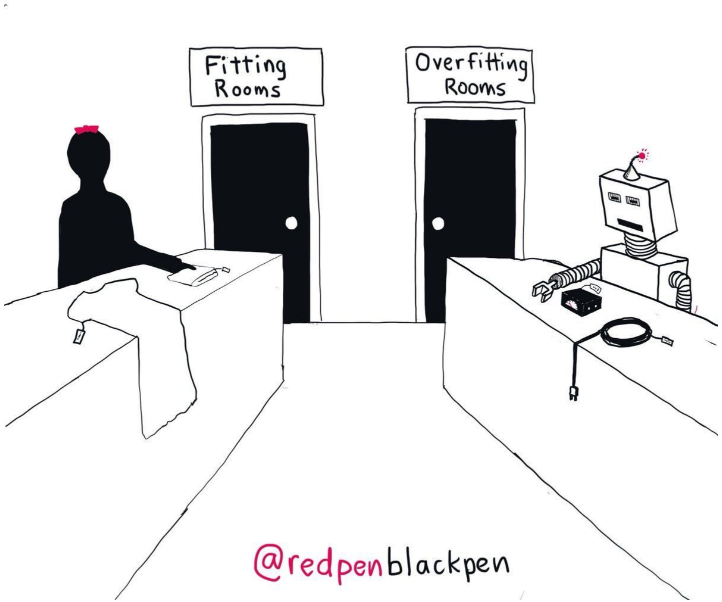
- **Ideal Situation:** Investigating positive selection on protein-coding genes in a population facing strong selective pressures, such as a pathogen-host arms race.
- **Strengths:** Focuses on protein-coding genes, highlighting signatures of positive selection.
- **Weaknesses:** Relies on accurate gene annotation and assumptions about synonymous mutations.

Haplotype-based Tests



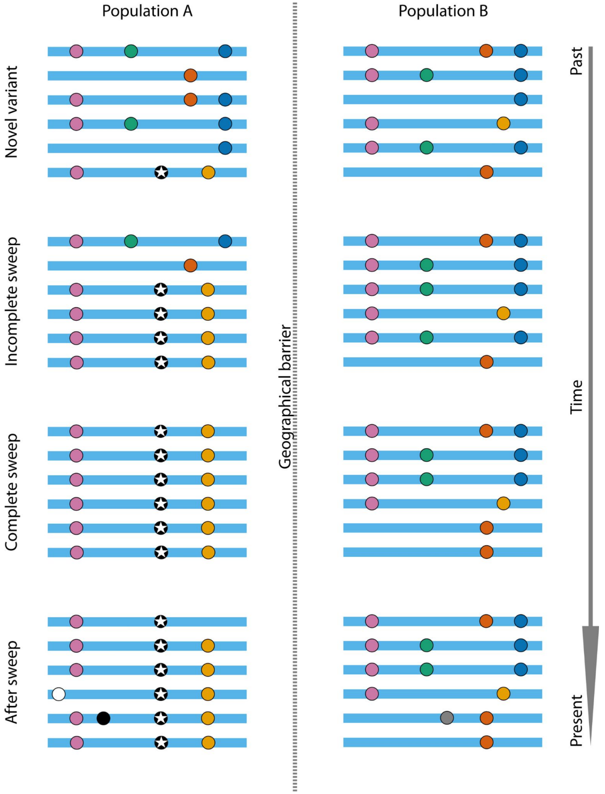
- **Ideal Situation:** Exploring patterns of selection in regions with complex haplotype structures, like genomic regions associated with disease resistance.
- **Strengths:** Captures information on linkage disequilibrium and haplotype structure.
- **Weaknesses:** Sensitive to recombination rates and complex demographic scenarios.

Machine learning based approaches



- **Ideal Situation:** Employing machine learning techniques like neural networks to recognize subtle patterns indicative of positive selection in genomic sequences, especially in non-coding regions.
- **Strengths:** Can capture intricate patterns and dependencies in data. Suitable for high-dimensional and non-linear relationships.
- **Weaknesses:** Requires substantial computational resources and large training datasets. Interpretability may be challenging.

Selective Sweep Analysis



- **Ideal Situation:** Detecting selective sweeps in a population exposed to a sudden environmental change, such as a change in climate or the introduction of a new predator.
- **Strengths:** Detects regions where selected alleles rapidly rise in frequency.
- **Weaknesses:** May miss soft sweeps and is sensitive to demographic history and recombination rates.

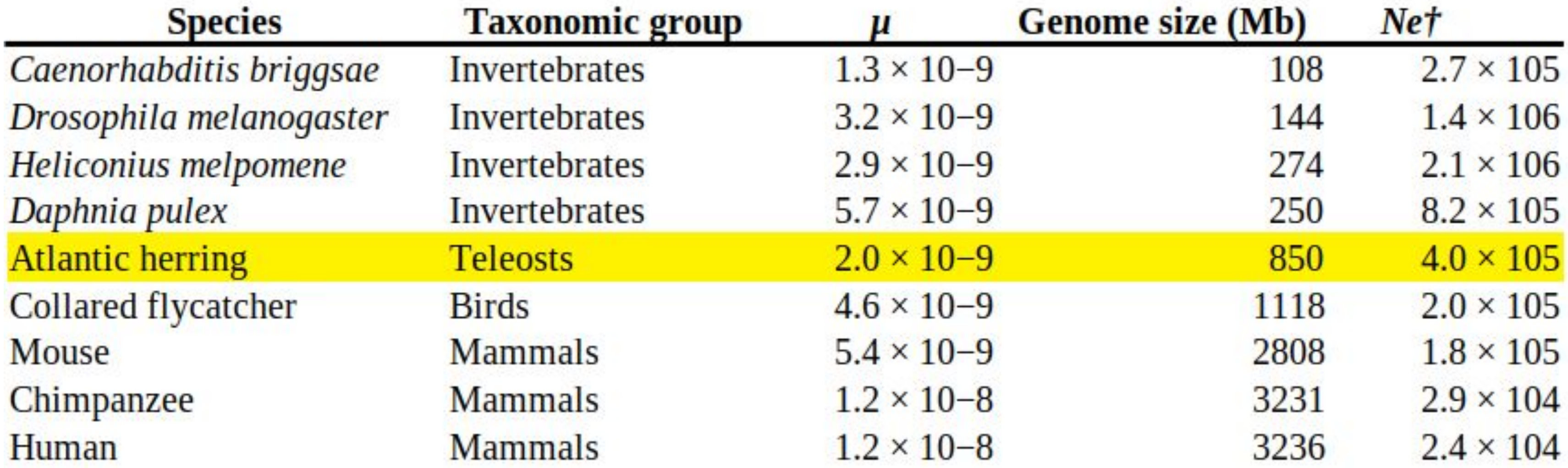
Detecting signatures of selection in Atlantic herring using whole genome sequencing

Photo Per Eide

http://sointularipple.ca

- A near ideal population structure
- huge population size (census population size about 10^{12})
- random mating
- high fecundity
- natural selection very effective
- genetic drift at selectively neutral loci is minute

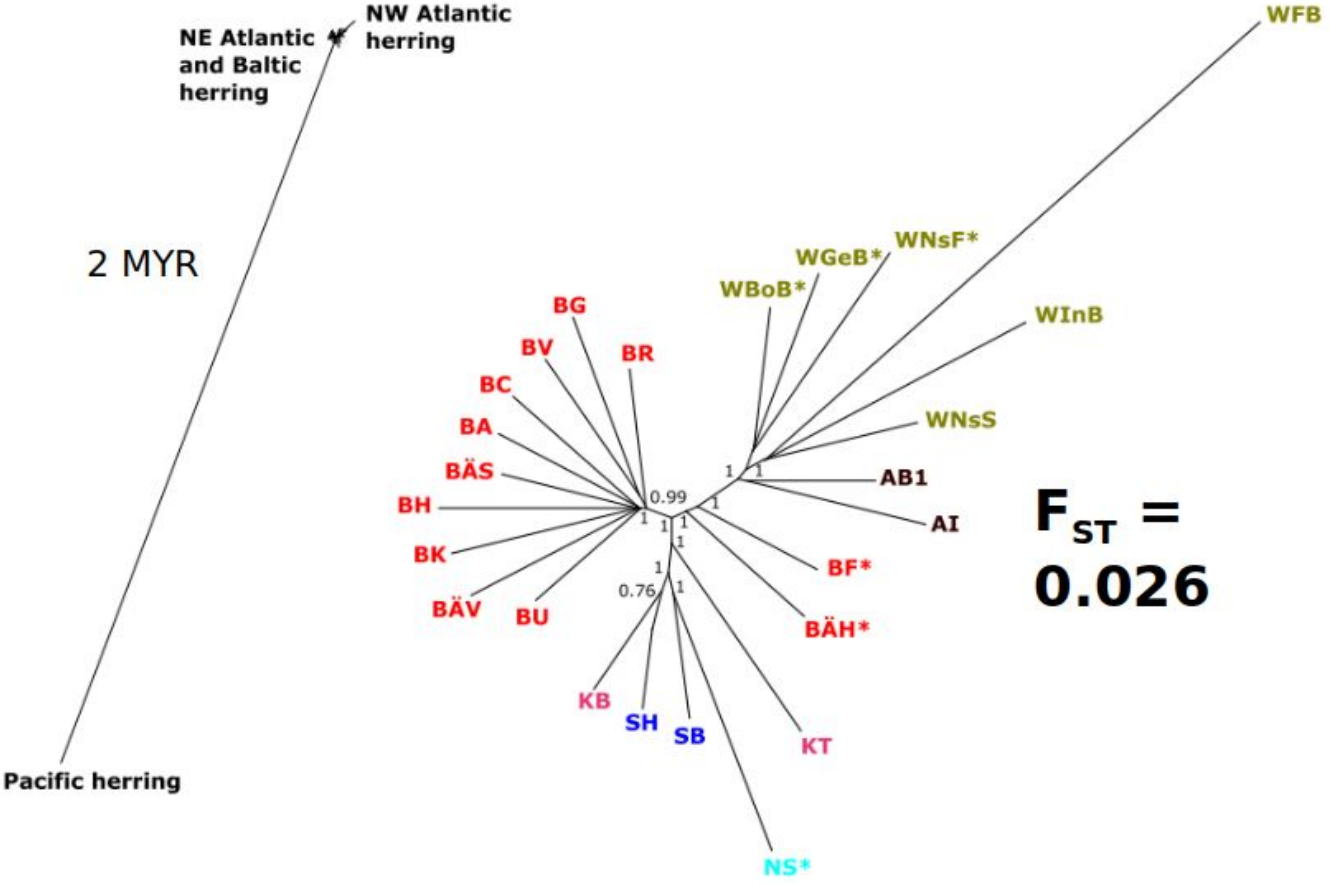
The herring has a very low mutation rate



Species	Taxonomic group	*μ*	Genome size (Mb)	*Ne†*
Caenorhabditis briggsae	Invertebrates	1.3 × 10−9	108	2.7 × 105
Drosophila melanogaster	Invertebrates	3.2 × 10−9	144	1.4 × 106
Heliconius melpomene	Invertebrates	2.9 × 10−9	274	2.1 × 106
Daphnia pulex	Invertebrates	5.7 × 10−9	250	8.2 × 105
Atlantic herring	Teleosts	2.0 × 10−9	850	4.0 × 105
Collared flycatcher	Birds	4.6 × 10−9	1118	2.0 × 105
Mouse	Mammals	5.4 × 10−9	2808	1.8 × 105
Chimpanzee	Mammals	1.2 × 10−8	3231	2.9 × 104
Human	Mammals	1.2 × 10−8	3236	2.4 × 104

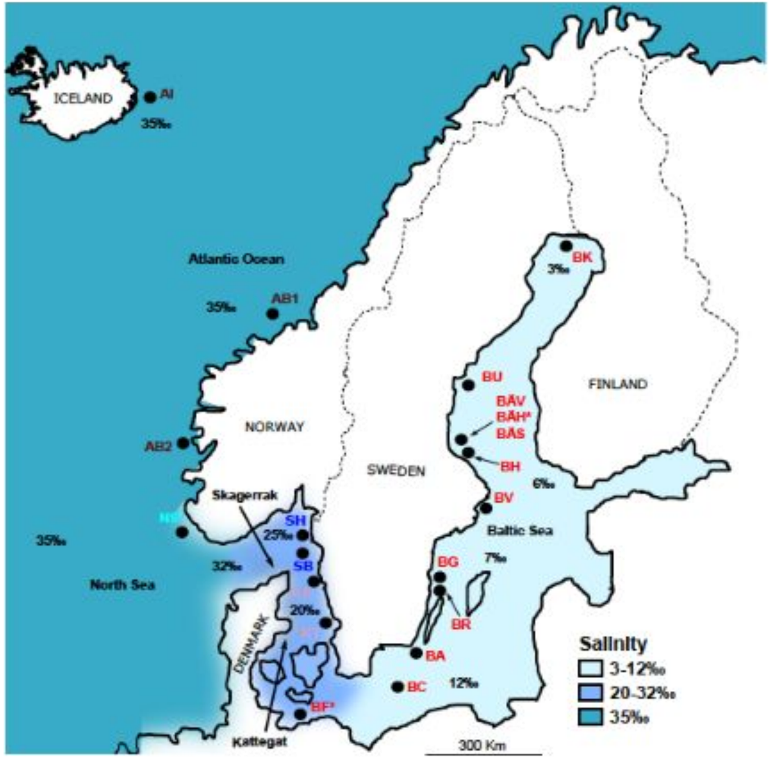
….but in each generation **each** nucleotide site mutates thousands of times!

Phylogenetic analysis



F_{ST} = 0.026

A genetic screen for loci underlying adaptation to the Baltic Sea vs. Atlantic Ocean



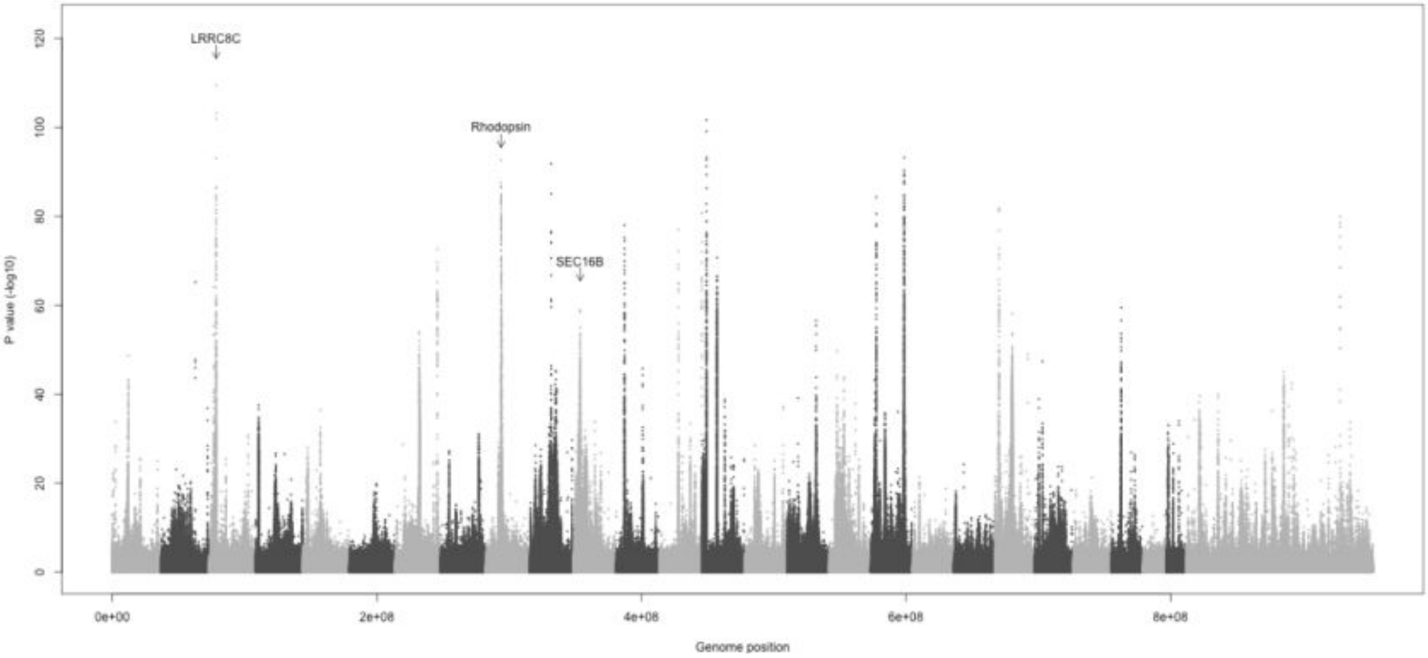
<u>Pool Atlantic</u>: All samples from Skagerrak, Kattegat, North Sea and Atlantic Ocean

<u>Pool Baltic</u>: All samples from the Baltic Sea

Compare allele frequencies between the two pools SNP by SNP

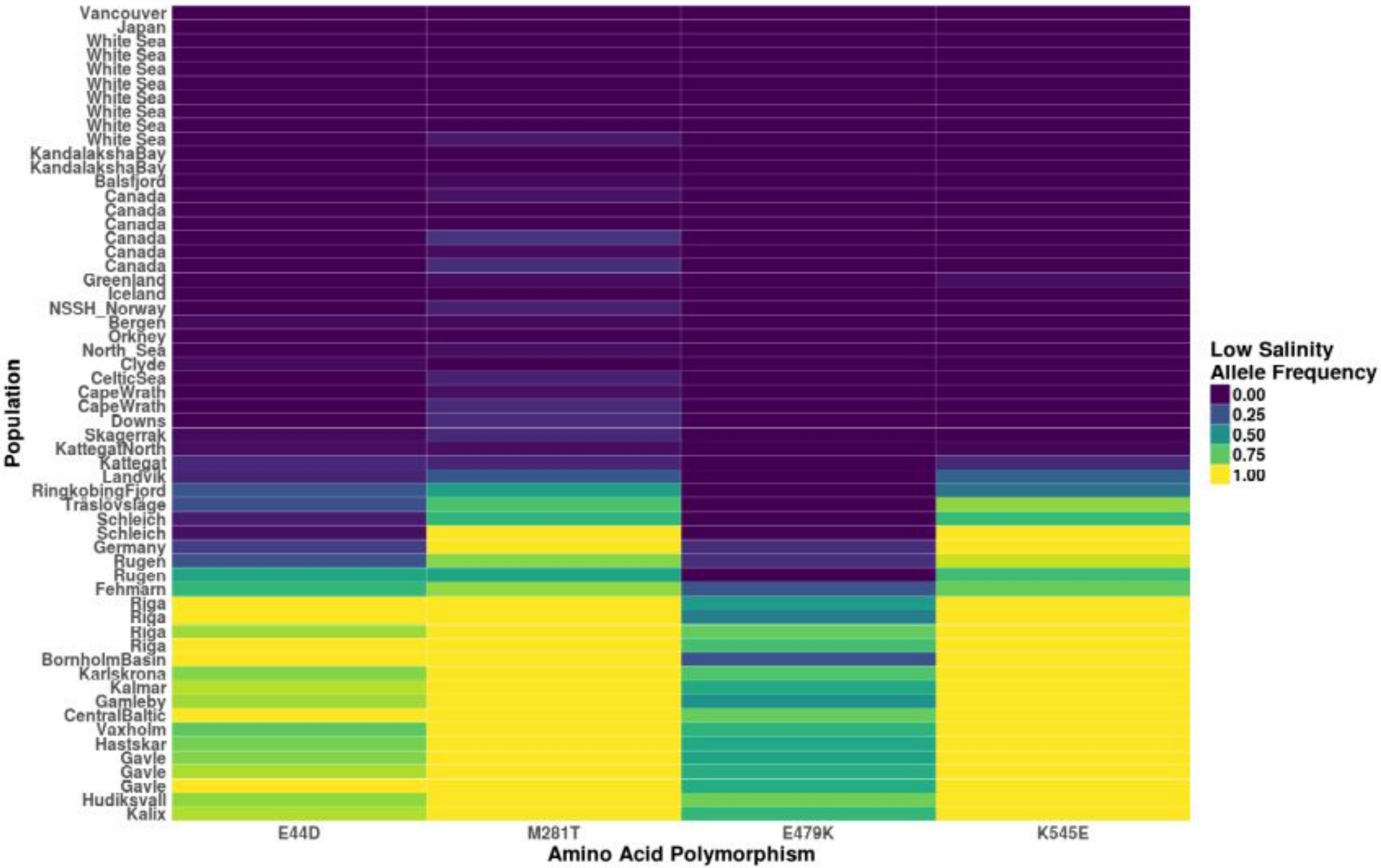
How polygenic is ecological adaptation in a species with a huge breeding population?

Genome-wide association analysis
Atlantic versus Baltic

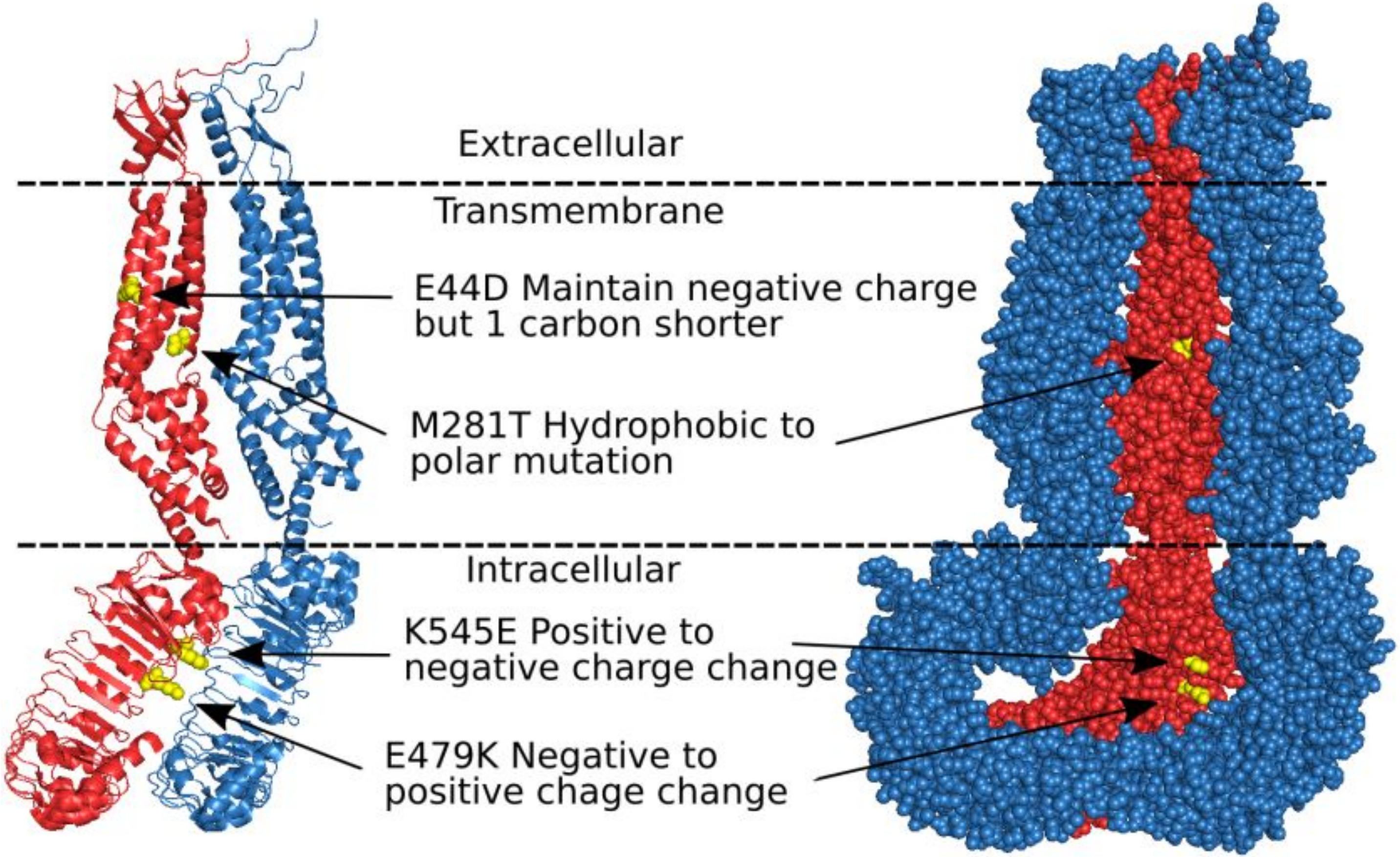


About 500 regions in the genome with highly differentiated SNPs $P<10^{-20}$

Frequency of Low-Salinity allele in LRRC8C



LRRC8C – AA sites > .65 ΔAF baltic/atlantic



Extracellular

Transmembrane

E44D Maintain negative charge
but 1 carbon shorter

M281T Hydrophobic to
polar mutation

Intracellular

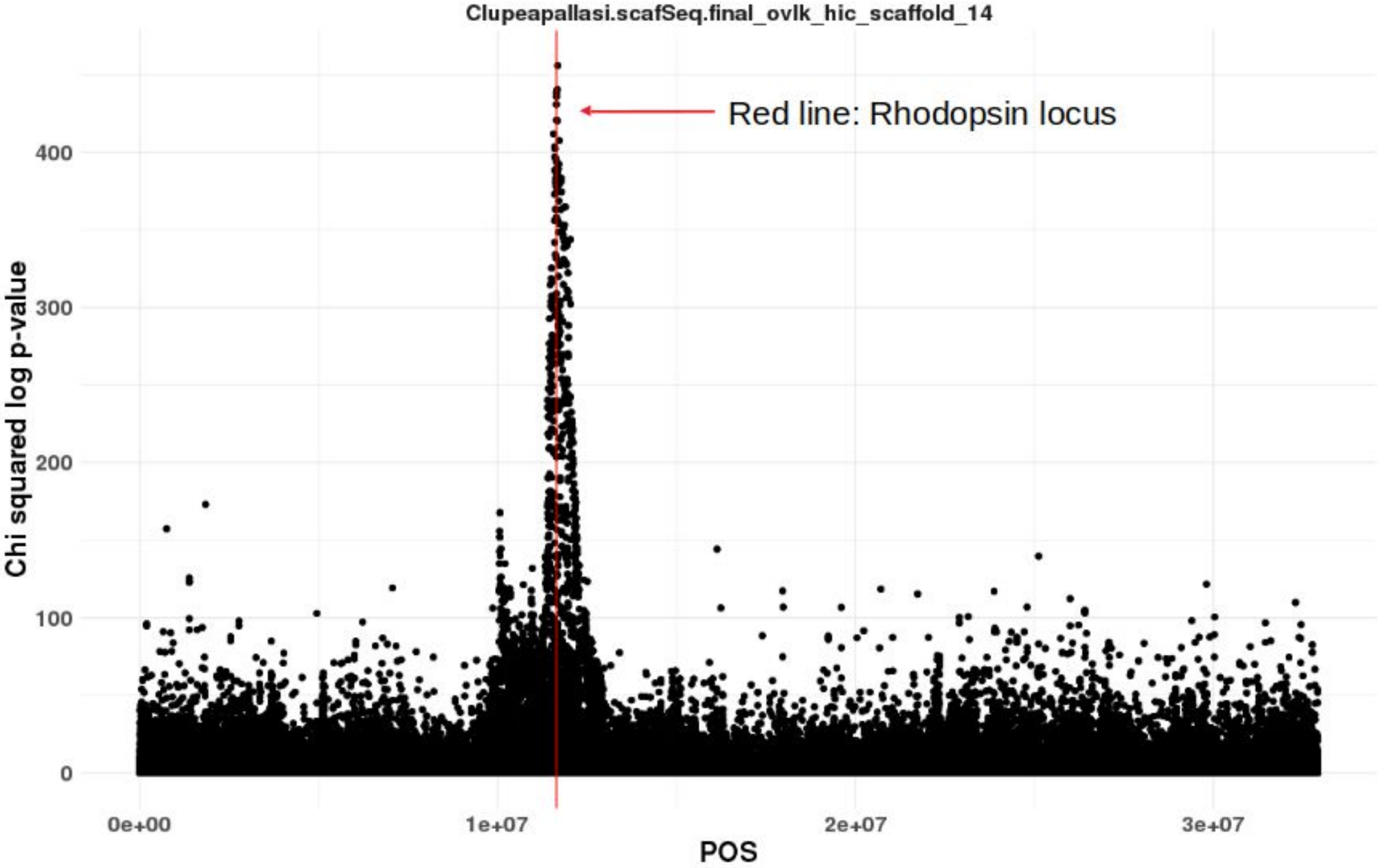
K545E Positive to
negative charge change

E479K Negative to
positive chage change

Exercise

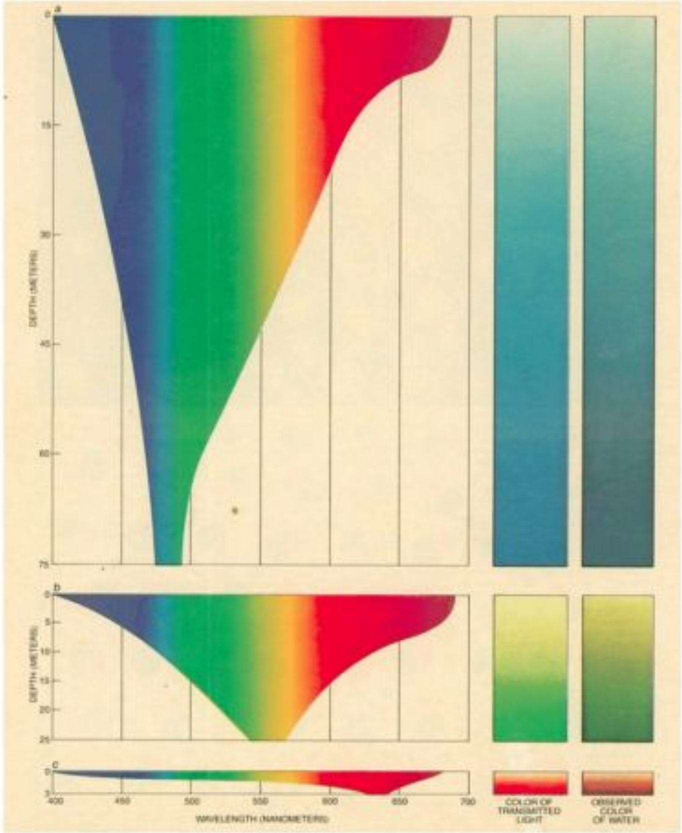
1) Inside your working directory: $ git clone https://github.com/JasonAnthonyHill/pgip
2) $ module load R/4.2.1 R_packages/4.2.1 RStudio
3) $ rstudio
4) Within rstudio load the file: herring_selection_scan.Rmd

Baltic vs. Atlantic SNPS

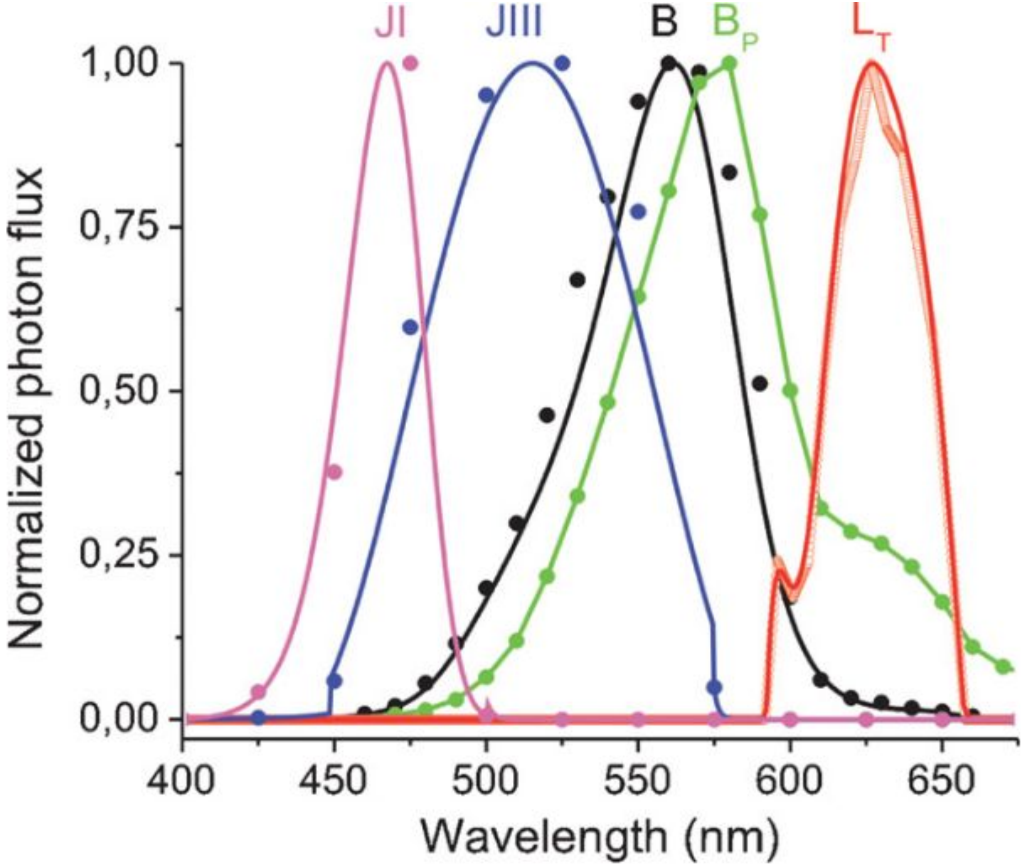


Clupeapallasi.scafSeq.final_ovlk_hic_scaffold_14

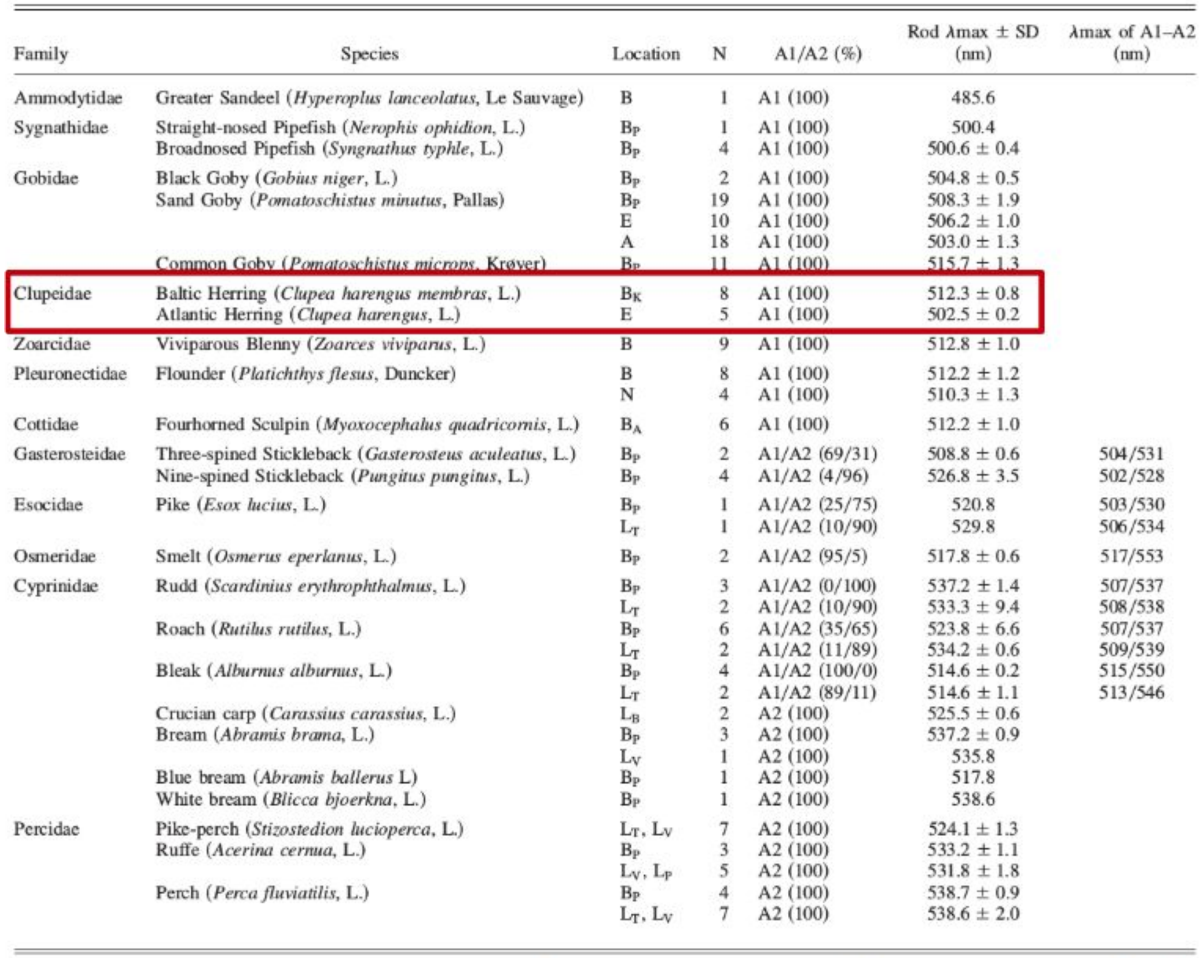
Red line: Rhodopsin locus



J. Pahlberg (2007)



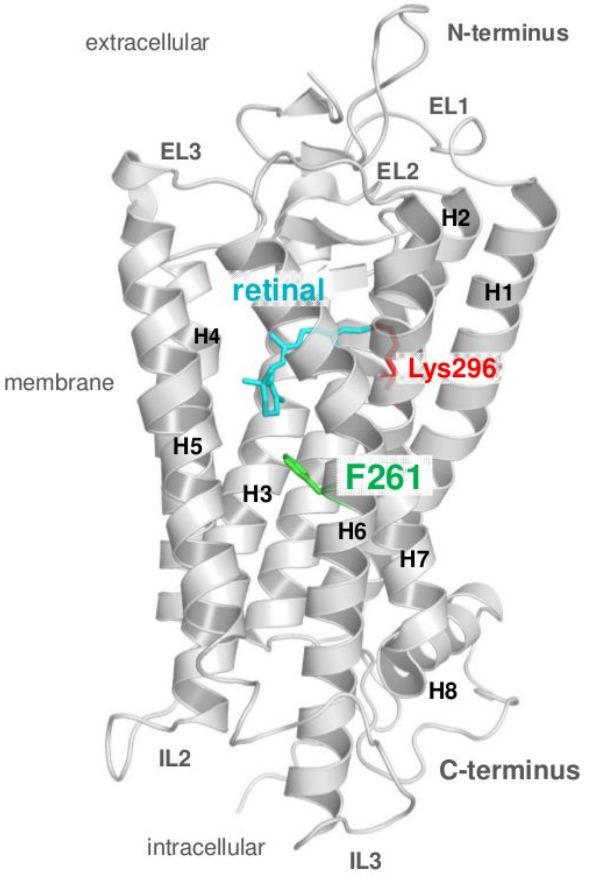
M. Jokela-Määttä *et al.* Visual Neuroscience (2007)



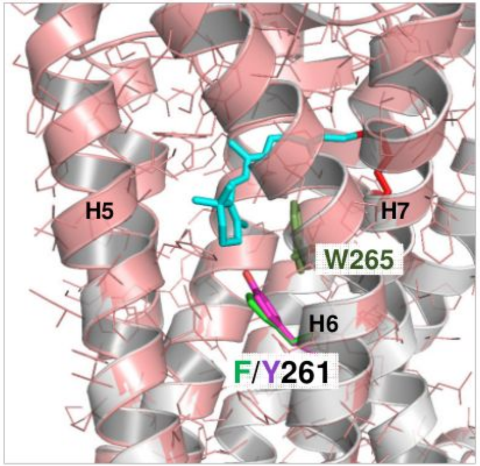
Family	Species	Location	N	A1/A2 (%)	Rod λmax ± SD (nm)	λmax of A1–A2 (nm)
Ammodytidae	Greater Sandeel (*Hyperoplus lanceolatus*, Le Sauvage)	B	1	A1 (100)	485.6	
Sygnathidae	Straight-nosed Pipefish (*Nerophis ophidion*, L.)	B_P	1	A1 (100)	500.4	
	Broadnosed Pipefish (*Syngnathus typhle*, L.)	B_P	4	A1 (100)	500.6 ± 0.4	
Gobidae	Black Goby (*Gobius niger*, L.)	B_P	2	A1 (100)	504.8 ± 0.5	
	Sand Goby (*Pomatoschistus minutus*, Pallas)	B_P	19	A1 (100)	508.3 ± 1.9	
		E	10	A1 (100)	506.2 ± 1.0	
		A	18	A1 (100)	503.0 ± 1.3	
	Common Goby (*Pomatoschistus microps*, Krøyer)	B_P	11	A1 (100)	515.7 ± 1.3	
Clupeidae	Baltic Herring (*Clupea harengus membras*, L.)	B_K	8	A1 (100)	512.3 ± 0.8	
	Atlantic Herring (*Clupea harengus*, L.)	E	5	A1 (100)	502.5 ± 0.2	
Zoarcidae	Viviparous Blenny (*Zoarces viviparus*, L.)	B	9	A1 (100)	512.8 ± 1.0	
Pleuronectidae	Flounder (*Platichthys flesus*, Duncker)	B	8	A1 (100)	512.2 ± 1.2	
		N	4	A1 (100)	510.3 ± 1.3	
Cottidae	Fourhorned Sculpin (*Myoxocephalus quadricornis*, L.)	B_A	6	A1 (100)	512.2 ± 1.0	
Gasterosteidae	Three-spined Stickleback (*Gasterosteus aculeatus*, L.)	B_P	2	A1/A2 (69/31)	508.8 ± 0.6	504/531
	Nine-spined Stickleback (*Pungitus pungitus*, L.)	B_P	4	A1/A2 (4/96)	526.8 ± 3.5	502/528
Esocidae	Pike (*Esox lucius*, L.)	B_P	1	A1/A2 (25/75)	520.8	503/530
		L_T	1	A1/A2 (10/90)	529.8	506/534
Osmeridae	Smelt (*Osmerus eperlanus*, L.)	B_P	2	A1/A2 (95/5)	517.8 ± 0.6	517/553
Cyprinidae	Rudd (*Scardinius erythrophthalmus*, L.)	B_P	3	A1/A2 (0/100)	537.2 ± 1.4	507/537
		L_T	2	A1/A2 (10/90)	533.3 ± 9.4	508/538
	Roach (*Rutilus rutilus*, L.)	B_P	6	A1/A2 (35/65)	523.8 ± 6.6	507/537
		L_T	2	A1/A2 (11/89)	534.2 ± 0.6	509/539
	Bleak (*Alburnus alburnus*, L.)	B_P	4	A1/A2 (100/0)	514.6 ± 0.2	515/550
		L_T	2	A1/A2 (89/11)	514.6 ± 1.1	513/546
	Crucian carp (*Carassius carassius*, L.)	L_B	2	A2 (100)	525.5 ± 0.6	
	Bream (*Abramis brama*, L.)	B_P	3	A2 (100)	537.2 ± 0.9	
		L_V	1	A2 (100)	535.8	
	Blue bream (*Abramis ballerus* L)	B_P	1	A2 (100)	517.8	
	White bream (*Blicca bjoerkna*, L.)	B_P	1	A2 (100)	538.6	
Percidae	Pike-perch (*Stizostedion lucioperca*, L.)	L_T, L_V	7	A2 (100)	524.1 ± 1.3	
	Ruffe (*Acerina cernua*, L.)	B_P	3	A2 (100)	533.2 ± 1.1	
		L_V, L_P	5	A2 (100)	531.8 ± 1.8	
	Perch (*Perca fluviatilis*, L.)	B_P	4	A2 (100)	538.7 ± 0.9	
		L_T, L_V	7	A2 (100)	538.6 ± 2.0	

M. Jokela-Määttä *et al.* Visual Neuroscience (2007)

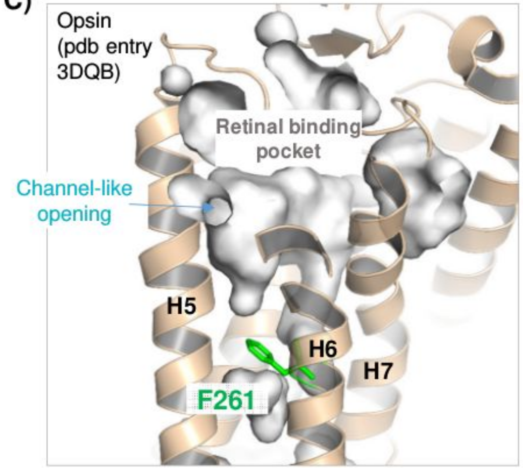
A) **Autumn spawning herring rhodopsin model, Meta-II state**



B) **Spring spawning rhodopsin model, Meta-II state**



C)



Insights from herring Rhodopsin models:

F261 is located in TMH6, close to the retinal binding site (**A**).

In consequence variant Y261 will be different to F261 in:

1. The electrostatics at this retinal binding region (additional OH group) (**B**)

2. The contact and interplay with Trp265 that is essentially involved in the retinal β-ionone ring/rhodopsin interplay and thereby activation of the receptor (**B**)

3. The variant could have impact on the channel-like opening for retinal release/entry as observed in the opsin structure (**C**)

- Park JH§, Scheerer P§, et al., Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 2008, 454(7201):183-7.
- Piechnick R, Ritter E, Hildebrand PW, Ernst OP, Scheerer P, Hofmann KP, Heck M. Effect of channel mutations on the uptake and release of the retinal ligand in opsin. *PNAS* 2012;109(14):5247-52

Kleinau, Scheerer, Charité, Berlin, 03. Oct. 2018

Multiple observations of Tyr 261 involved in fish vision adaptation

Spectral Tuning and Molecular Evolution of Rod Visual Pigments in the Species Flock of Cottoid Fish in Lake Baikal

DAVID M. HUNT,*† JUDE FITZGIBBON,† SERGEY J. SLOBODYANYUK,‡ JAMES K. BOWMAKER§

Received 28 March 1995; in revised form 19 June 1995; in final form 14 August 1995

Rhodopsin From the Fish, *Astyanax:* Role of Tyrosine 261 in the Red Shift

Ruth Yokoyama, Barry E. Knox,† and Shozo Yokoyama**

The Journal of Experimental Biology 212, 3415-3421
Published by The Company of Biologists 2009
doi:10.1242/jeb.031344

Individual variation in rod absorbance spectra correlated with opsin gene polymorphism in sand goby (*Pomatoschistus minutus*)

Mirka Jokela-Määttä[1,*], Annika Vartio[1], Lars Paulin[2] and Kristian Donner[1]
[1]Department of Biological and Environmental Sciences and [2]Institute of Biotechnology, University of Helsinki, Helsinki, Finland
*Author for correspondence (mirka.jokela@helsinki.fi)

Accepted 28 July 2009

Summary

- Many methods exist for detecting selection using genomic data
- Selecting the right method for the right species can yield powerful results
- Carry through to structure and function