Supplementary Material

Determinants of genetic variation across eco-evolutionary scales in pinnipeds

Claire R. Peart*, Sergio Tusso*, Saurabh D. Pophaly*, Fidel Botero-Castro, Chi-Chih Wu, David Aurioles-Gamboa, Amy B. Baird, John W. Bickham, Jaume Forcada, Filippo Galimberti, Neil J. Gemmell, Joseph I. Hoffman, Kit M. Kovacs, Mervi Kunnasranta, Christian Lydersen, Tommi Nyman, Larissa Rosa de Oliviera, Anthony J. Orr, Simona Sanvito, Mia Valtonen, Aaron B. A. Shafer[†], Jochen B. W. Wolf[†]

*Contributed equally;

†Correspondence: aaronshafer@trentu.ca (https://orcid.org/0000-0001-7652-225X); j.wolf@biologie.uni-muenchen.de (https://orcid.org/0000-0002-2958-51830)

Supplementary Methods

ClustRAD wrapper

ClustRAD is largely based on BEDtools[1] and defines genome intervals with valid ddRAD fragments adhering to predefined criteria. First, any valid fragment should contain the expected cut sites from both enzymes used for the double digest at either end (here MseI and SbfI). Any genomic region not meeting this assumption was therefore excluded. This was necessary as visual inspection of .bam files in a locally installed UCSC genome browser[2] further revealed several instances where sequencing reads extended beyond the enzyme cut sites (**Extended Data 10**). The extended regions are likely due to 'ghost PCR extensions' with low base quality scores and highly elevated mismatch and indel rates, and were likewise excluded from further analysis. A further source of complication, likely introduced by incomplete digestion, was partial overlap of fragments sharing the cut site of the rare cutter (SbfI), yet differing in neighbouring cut sites of the frequent cutter (MseI). While this does not introduce downstream errors *per se*, it complicates the definition of a valid locus and introduces heterogeneity in coverage. To make best use of the data, we first located all the MseI and SbfI sites in the reference genome using utilities available from the UCSC genome browser (http://hgdownload.soe.ucsc.edu/admin/exe/). Then we created genomic intervals between a SbfI and MseI sites (intervals either starting in SbfI and MseI or vice versa) up to 500 bp in length. As SbfI sites are rare in the genome compared to MseI, these intervals make clusters along the genome with SbfI cut site as the focal point and several MseI sites around it (Figure S8). Intervals in a cluster either start from SbFi and end in MseI (towards the right of the SbfI site) or they start in MseI and end in SbfI (towards the left of the SbfI site). For each cluster, only intervals with minimum read coverage (MAPQ ≥ 70 as given by STAMPY v1.0.23) of 3x in at least three individuals, each covering at least 99% of the interval were kept. Then the shortest left and the shortest right interval for each cluster (but still greater than 50bp) were selected for further analysis. The number of these intervals was determined for each individual to get the individual with the maximum number; individuals with fewer than half of the maximum number of intervals were then removed from further analysis. The source code used to generate clusters and filtering steps are available at https://github.com/EvoBioWolf/ClustRAD. For this study, the genome coordinates selected by ClustRAD were used and in addition only sites with less than 20% missing data were included.

Supplementary Tables

Supplementary Table 1. Branch-specific estimates of neutral substitution rates as inferred from synonymous mutations in coding sequence of 8,864 genes for five pinniped species (see **Extended Data 8**). Abbreviations: my = million years.

Species	Clade	Branch length [my]	Number of synonymous substitutions (8,864 genes)	Number of synonymous sites	dS (Synonymous substitutions [site $^{-1}$])	Substitution rate (Substitutions [site^{-1}*my^{-1}])
Arctocephalus gazella	Otariidae	9.16	43270.407	3918725.270	0.01104	0.00121
Zalophus californianus	Otariidae	9.16	17849.757	4204833.561	0.00424	0.00046
Neomonachus schauinslandi	Phocidae	18.16	41512.988	4261459.342	0.00974	0.00054
Leptonychotes weddellii	Phocidae	18.16	39907.235	4212734.594	0.00947	0.00052
Odobenus rosmarus	Odobenidae	21.67	50404.901	4588914.783	0.01098	0.00051

Supplementary Table 2. List of population samples including information on taxonomic status, sampling location, permits, population characteristics, genetic data, demographic reconstruction and life-history traits[4]. The table is available online.

Supplementary Table 3. Compilation of measures indicating the degree of population differentiation for the different species. The table is available online.

Supplementary Table 4. Summary of the best statistical models exploring the relationship between Tajima's D and various life-history parameters. Estimates of the parameters included in a specific model and model selection statistics are given for all models sorted by ΔAIC_c. AIC_c = Akaike's information criterion adjusted for small sample sizes; ΔAIC_c = difference in AIC to best model; $wAIC_c$ = Akaike weights; R^2_{adj}= squared correlation coefficient adjusted for number of parameters; $wAIC_c$: Akaike weights; BIC: Bayesian Information Criterion.

Model	Intercept	Breeding habitat	Species range	Harem size	Reproductive Season length	R2adj	k	AIC	AICc	ΔAIC	ΔAICc	wAIC	wAICc	bic
lm2.6	0.78	0.27	-0.09			0.48	4	0.26	3.59	0.00	0.00	0.55	0.50	3.59
lm1.4	1.49		-0.11			0.34	3	3.71	5.56	3.46	1.97	0.10	0.19	6.21
lm2.3	1.36		-0.11	0.07		0.38	4	3.33	6.66	3.07	3.07	0.12	0.11	6.66
lm1.3	-0.75	0.34				0.24	3	5.96	7.81	5.71	4.22	0.03	0.06	8.46
lm2.5	0.71		-0.09		0.11	0.33	4	4.70	8.04	4.45	4.45	0.06	0.05	8.04
lm1.2	-1.15				0.25	0.19	3	7.09	8.94	6.84	5.35	0.02	0.03	9.59
lm2.4	-1.12	0.24			0.14	0.25	4	6.75	10.08	6.49	6.49	0.02	0.02	10.08
lm0	-0.17					0.00	2	9.80	10.66	9.54	7.07	0.01	0.01	11.47
lm2.2	-0.82	0.42		-0.04		0.21	4	7.60	10.94	7.35	7.35	0.01	0.01	10.94
lm1.1	-0.29			0.07		0.03	3	10.23	12.07	9.97	8.48	0.00	0.01	12.73
lm2.1	-1.10			0.03	0.23	0.15	4	8.82	12.16	8.57	8.57	0.01	0.01	12.16
m.full	0.94	0.30	-0.10	0.00	-0.04	0.40	6	4.15	12.55	3.90	8.97	0.08	0.01	9.15
wAICc		0.593	0.847	0.14	0.12									

Supplementary Table 5. Summary of the best statistical models exploring the relationship between the Ne/Nc ratio and various life-history parameters. Estimates of the parameters included in a specific model and model selection statistics are given for all models sorted by ΔAIC$_c$. AIC$_c$ = Akaike's information criterion adjusted for small sample sizes; ΔAIC$_c$ = difference in AIC to best model; wAIC$_c$ = Akaike weights; R$^2_{adj}$= squared correlation coefficient adjusted for number of parameters; wAIC$_c$: Akaike weights; BIC: Bayesian Information Criterion.

Model	Intercept	Breeding habitat	Breeding latitude	Harem size	Reproductive Season length	Tajima's jD	R2adj	k	AIC	AICc	ΔAIC	ΔAICc	wAIC	wAICc	bic
lm1.1	-1.52					2.34	0.26	3	54.80	56.64	0.74	0.00	0.14	0.27	57.30
lm1.3	-6.14				1.08		0.20	3	56.03	57.87	1.96	1.23	0.08	0.14	58.53
lm2.2	-4.19				0.66	1.70	0.28	4	55.00	58.33	0.94	1.69	0.13	0.11	58.33
lm2.3	-0.02	-0.81				3.03	0.28	4	55.14	58.48	1.08	1.83	0.12	0.11	58.48
lm2.4	-1.14		-0.01			2.01	0.23	4	56.29	59.62	2.23	2.98	0.07	0.06	59.62
lm0	-1.92						0.00	2	58.94	59.80	4.88	3.15	0.02	0.06	60.61
lm2.8	-6.24	-0.83			1.47		0.22	4	56.51	59.85	2.45	3.20	0.06	0.05	59.84
lm1.5	-0.90		-0.02				0.09	3	58.21	60.06	4.15	3.41	0.03	0.05	60.71
lm2.1	-1.42			-0.05	NA	2.40	0.21	4	56.74	60.07	2.68	3.43	0.05	0.05	60.07
lm2.9	-5.12		-0.01		0.91		0.16	4	57.77	61.10	3.71	4.46	0.03	0.03	61.10
lm2.5	-6.32			-0.10	1.17		0.15	4	57.83	61.17	3.77	4.52	0.03	0.03	61.16
lm1.2	-2.11			0.11			-0.05	3	60.71	62.55	6.65	5.91	0.01	0.01	63.21
lm1.4	-2.30	0.22					-0.06	3	60.82	62.67	6.76	6.03	0.01	0.01	63.32
lm2.10	0.17	-0.47	-0.03				0.05	4	59.76	63.09	5.70	6.45	0.01	0.01	63.09
lm2.7	-0.91		-0.02	0.01			0.03	4	60.21	63.54	6.15	6.90	0.01	0.01	63.54
lm2.6	-2.11	0.00		0.11			-0.13	4	62.71	66.04	8.65	9.40	0.00	0.00	66.04
m.full	-1.61	-2.08	-0.01	0.26	0.96	2.48	0.39	7	54.06	66.51	0.00	9.86	0.21	0.00	59.89
wAICc		0.19	0.16	0.10	0.42	0.60									

Supplementary Table 6. List of individual samples including information on population membership. For each sample, accession numbers to the Sequencing Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) are provided. The table is available online.

Bibliography

1. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

2. Kent, W. J. *et al.* The Human Genome Browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

3. Higdon, J. W., Bininda-Emonds, O. R. P., Beck, R. M. D. & Ferguson, S. H. Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset. *BMC Evol. Biol.* **7: 216**, (2007).

4. Krüger, O., Wolf, J. B. W., Jonker, R. M., Hoffman, J. I. & Trillmich, F. Disentangling the contribution of sexual selection and ecology to the evolution of size dimorphism in pinnipeds. *Evolution* **68**, 1485–1496 (2014).