Supplementary materials for:

Stabilizing selection on Atlantic cod supergenes through a millennium of extensive exploitation

Marte Sodeland, Sissel Jentoft, Per Erik Jorde, Morten Mattingsdal, Jon Albretsen, Alf Ring Kleiven, Ann-Elin Wårøy Synnes, Sigurd Heiberg Espeland, Esben Moland Olsen, Carl Andrè, Nils Chr. Stenseth and Halvor Knutsen
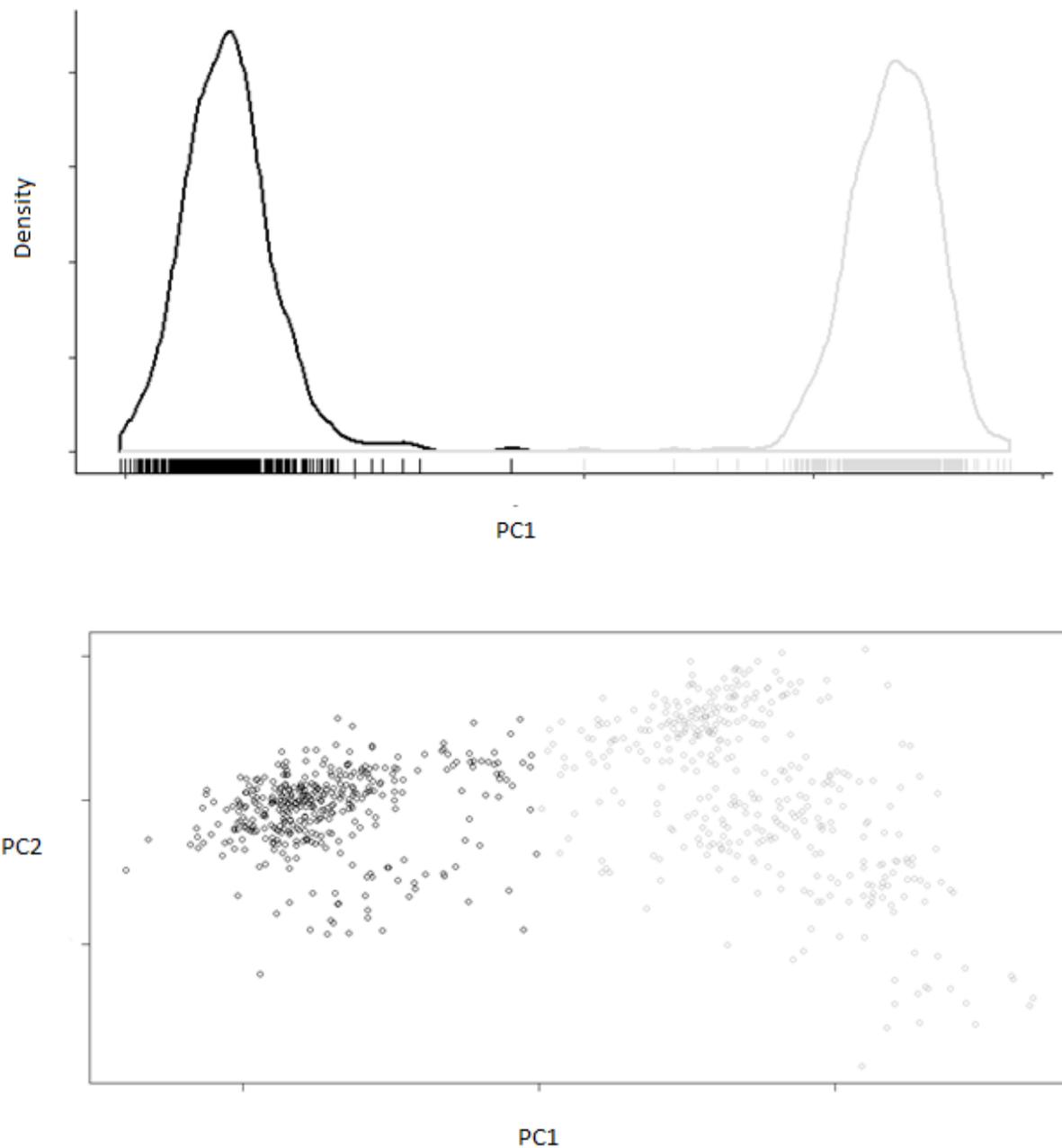
CONTENTS:

I: Supplementary Figures

II: Supplementary Tables

III: Supplementary Notes
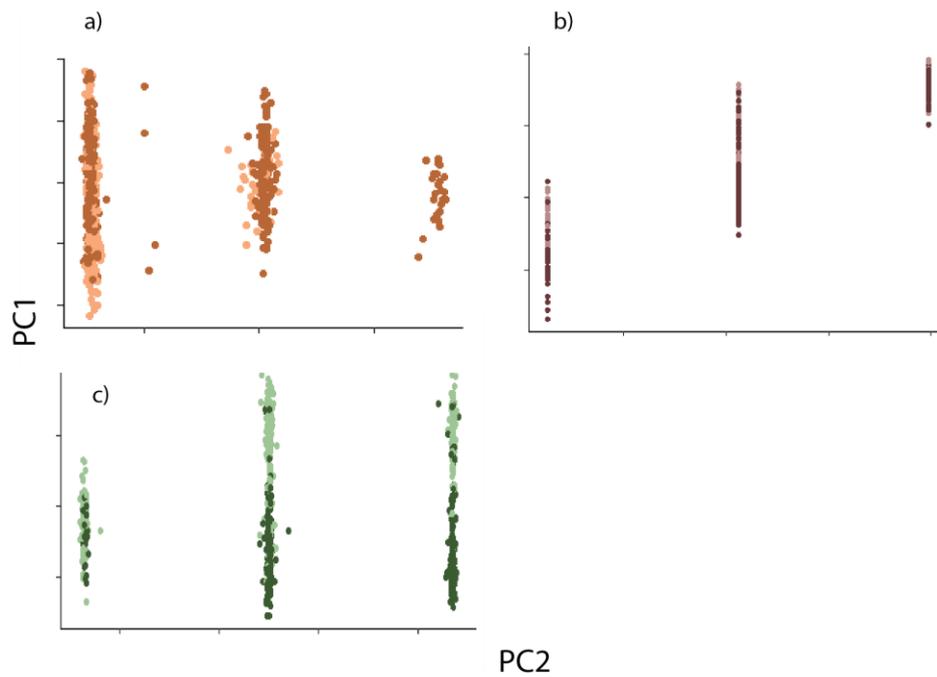
I: Supplementary Figures

Supplementary Figure 1: Above: density plot on principal component 1 (PC1) for 752 Atlantic cod specimens using 6'683 SNPs from a from a 10k Atlantic cod SNP array. By Discriminant analysis on principle components two main clusters were identified: offshore cod (black) and coastal cod (gray). PC1 for each specimen is indicated as a bar below the density plot for the identified clusters. Below: PC1 and PC2 are plotted together, also here with coastal cod (gray) and offshore cod (black) indicated.
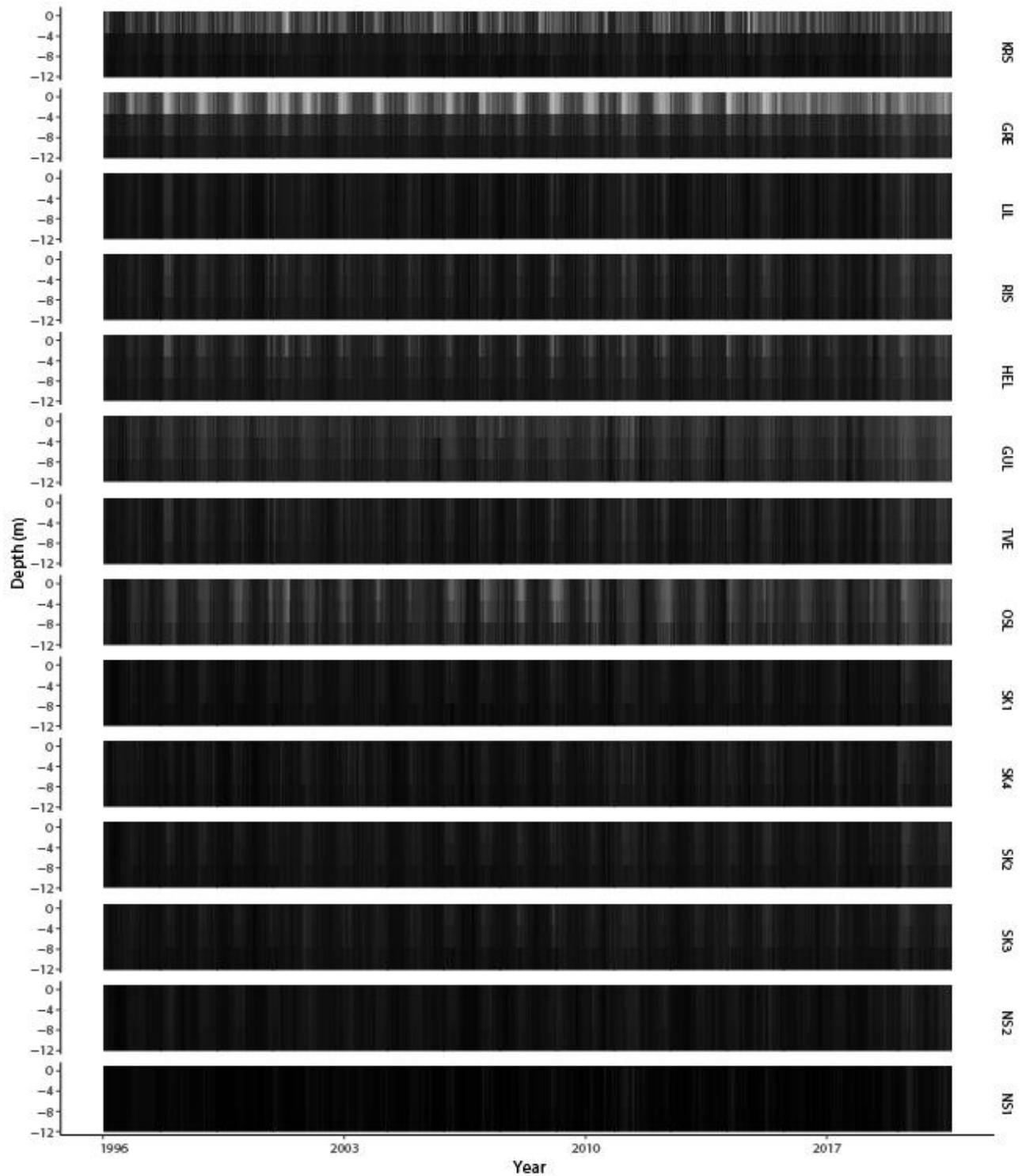
Supplementary Figure 2: Predicted daily salinity (lighter color is lower salinity) at three depths (0,5 and 10 below surface) over a twenty-three-year period (1996-2019) for fourteen sampling locations in Skagerrak and the North Sea. Sampling locations are ordered from bottom to top by decreasing fraction of coastal cod specimens relative to offshore cod specimens. Year is on the x-axis with one datapoint for each day, depth in meters from surface is on the y-axis.
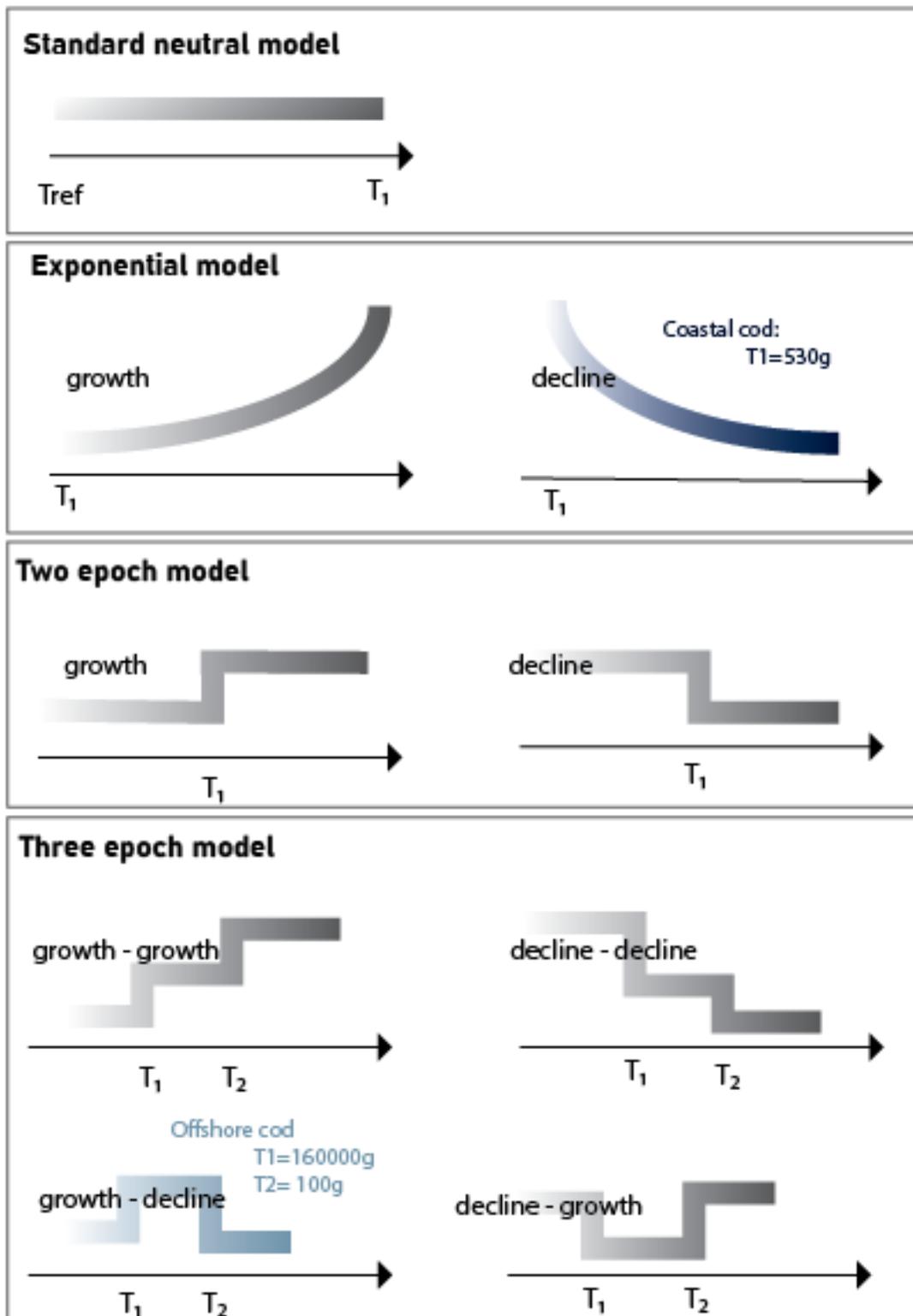
Supplementary Figure 3: Illustrations of four models of historic demography compared by diffusion approximation. The models were a standard neutral model (snm), an exponential model (exp), a two epoch model (two_epoch) and a three epoch model (three_epoch). For the latter three models alternative scenarios for positive and negative growth are illustrated. The model giving highest Aikake criterion (AIC) for Coastal cod (dark blue) and offshore cod (lighter blue) are indicated together with parameter estimates (see Supplementary Table 6 for more detailed results, and Supplementary Note 1 for python implementations).

Supplementary Figure 4: Principal component analysis for 752 Atlantic cod specimens using a) 83 SNPs in supergene SG02 (orange), b) 187 SNPs in supergene SG07 (burgundy) and 199 SNPs in supergene SG12 (green). Principal component 1 (PC1) is shown on the x-axis and principal component 2 (PC2) is shown on the y-axis. Coastal cod specimens are shown in light color and offshore cod specimens are shown in dark color.
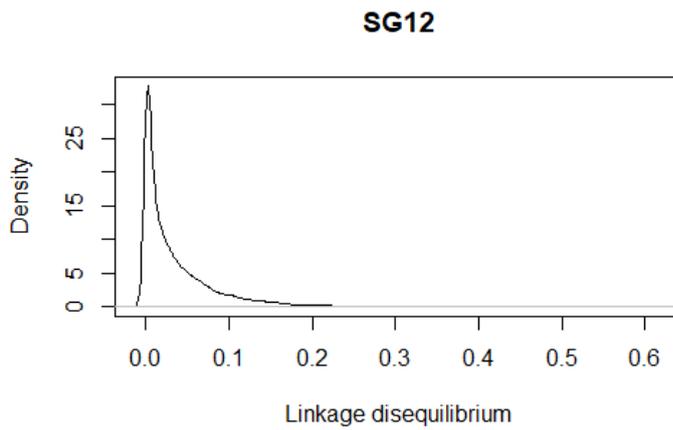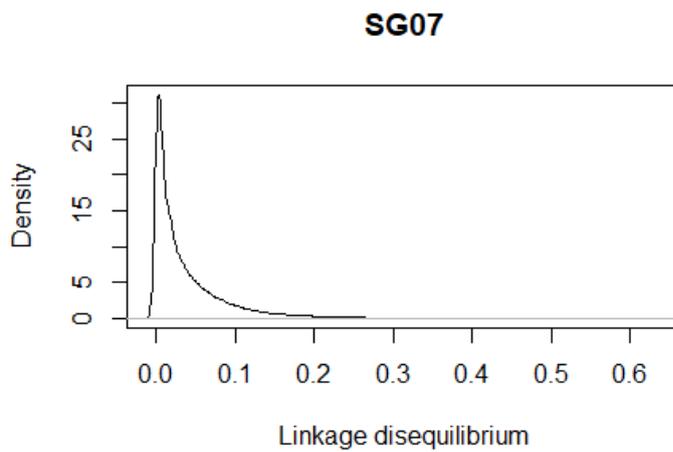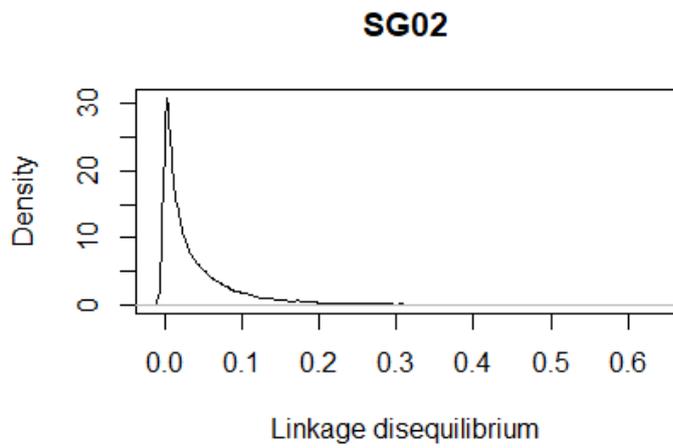
Supplementary Figure 3: Distribution of inter-chromosomal linkage disequilibrium between the supergenes SG02 (top), S07 (middle) and SG12 (bottom), and the 42'788 SNPs used for demographic inference.

II: Supplementary Tables

Supplementary Table 1: The table present the number of specimens, sampling years, number of coastal cod, longitude and latitude for each sampling location.

Location	Specimens (n)	Years	Latitude	Longitude
GRE	19	2001,2012	59.05	9.65
GUL	158	2005	58.4	11.6
HEL	29	2003,2010,2011	58.88	9.38
KAT	48	2004	56.90	12.15
KRS	48	2000, 2005, 2008	58.16	8.06
LIL	48	2004, 2005, 2010	58.24	8.38
NS1	48	2002	58.20	5.66
NS2	48	2012	55.60	5.85
ORE	48	2003	55.95	12.70
OSL	40	2001	59.81	10.56
RIS	48	1997, 1998, 2004, 2005	58.70	9.25
SK1	32	2012	57.80	8.26
SK2	32	2012	58.05	8.62
SK3	32	2012	58.39	9.39
SK4	18	2004	58.56	10.83
TVE	47	2004	58.59	8.94

Supplementary Table 2*:* Overrepresentation analysis on gene ontologies within supergenes SG02, SG07 and SG12. For each supergene (SG) the number of annotated genes and gene ontologies (GOs) is presented. For each gene ontology overrepresented within a supergene (>=2 fold increase), number of occurrences genome-wide excluding supergenes (N_{GW}), number of occurrences within the supergene (N_{SG}), fold increase and gene ontology name is given. Only gene ontologies that had 5 or more occurrences within the supergene were included in the overrepresentation analysis.

SG	Genes	GOs	GO Term	N_{GW}	N_{SG}	Increase	GO Name
SG02	145	194	GO:0046982	96	12	11.82759	protein heterodimerization activity
			GO:0000786	84	10	11.26437	nucleosome
			GO:0006334	89	10	10.63154	nucleosome assembly
SG07	172	214	GO:0007275	146	6	3.165393	multicellular organism development
			GO:0005216	186	6	2.484663	ion channel activity
			GO:0006811	228	7	2.364789	ion transport
			GO:0005215	186	5	2.070553	transporter activity
SG12	293	329	GO:0005230	60	6	4.621474	extracellular ligand-gated ion channel activity
			GO:0030170	53	5	4.359881	pyridoxal phosphate binding
			GO:0004713	75	5	3.080982	protein tyrosine kinase activity

Supplementary Table 3*:* Frequencies of supergene genotypes I/I, I/II and II/II in the 41 re-sequenced specimens from locations LIL (n=11), RIS (n=8) and NS1 (n=22) and from all genotyped specimens from the same locations (n=48 for all three locations).

		SG02			SG07			SG12		
		I/I	I/II	II/II	I/I	I/II	II/II	I/I	I/II	II/II
	LIL	0.27	0.45	0.27	0.14	0.35	0.50	0.27	0.36	0.36
Sequenced	**RIS**	0.00	0.25	0.75	0.13	0.13	0.75	0.00	0.25	0.75
	NS1	0.00	0.18	0.82	0.00	0.27	0.73	0.36	0.50	0.14
	LIL	0.15	0.50	0.35	0.06	0.48	0.46	0.06	0.29	0.65
Genotyped	**RIS**	0.04	0.31	0.65	0.08	0.34	0.58	0.04	0.31	0.65
	NS1	0.00	0.12	0.88	0.00	0.29	0.71	0.33	0.52	0.15

Supplementary Table 4: Four single-population demographic models were compared by diffusion approximation as implemented in the DADI python module. For each model and ecotype the Aikake criterion (AIC) and population (Theta) are presented. The model giving the best fit for each ecotype is indicated in italic, and parameter estimates are presented.

Model	Ecotype	AIC	Parameters
snm	Coastal	8985.34	
exp	*Coastal*	*1127.96*	*nref=26105,nu=261, T=527g*
two_epoch	Coastal	1191.24	
three_epoch	Coastal	1149.26	
snm	Offshore	14993.4	
exp	Offshore	3380.36	
two_epoch	Offshore	3581.42	
three_epoch	*Offshore*	*3357.84*	*nref=2903,nu1=30044,nu2=174,T1=161702g, ,T2=106g*

III: Supplementary Notes

Supplementary Note 1: Python code (blue) used to implement the standard neutral model (snm), the exponential development model (exp), the two-epoch development model (two_epoch) and the three-epoch development model (three_epoch).

```python
def snm(notused, ns, pts):

    xx = Numerics.default_grid(pts)

    phi = PhiManip.phi_1D(xx)

    fs = Spectrum.from_phi(phi, ns, (xx,))

    return fs


def exp(params, ns, pts):

    nu,T = params

    xx = Numerics.default_grid(pts)

    phi = PhiManip.phi_1D(xx)

    nu_func = lambda t: numpy.exp(numpy.log(nu) * t/T)

    phi = Integration.one_pop(phi, xx, T, nu_func)

    fs = Spectrum.from_phi(phi, ns, (xx,))

    return fs


def two_epoch(params, ns, pts):

    nu,T = params

    xx = Numerics.default_grid(pts)

    phi = PhiManip.phi_1D(xx)

    phi = Integration.one_pop(phi, xx, T, nu)

    fs = Spectrum.from_phi(phi, ns, (xx,))

    return fs


def three_epoch(params, ns, pts):

    nu1,nu2,TB,TF = params

    xx = Numerics.default_grid(pts)

    phi = PhiManip.phi_1D(xx)

    phi = Integration.one_pop(phi, xx, T1, nu1)

    phi = Integration.one_pop(phi, xx, T2, nu2)

    fs = Spectrum.from_phi(phi, ns, (xx,))

    return fs
```