Supplement to: Nucleotide Variation and Balancing Natural Selection at the *Ckma* gene in Atlantic cod: Analysis with multiple merger coalescent models

Einar Árnason and Katrín Halldórsdóttir

Institute of Biology, University of Iceland,
Sturlugata 7, 101 Reykjavík, Iceland

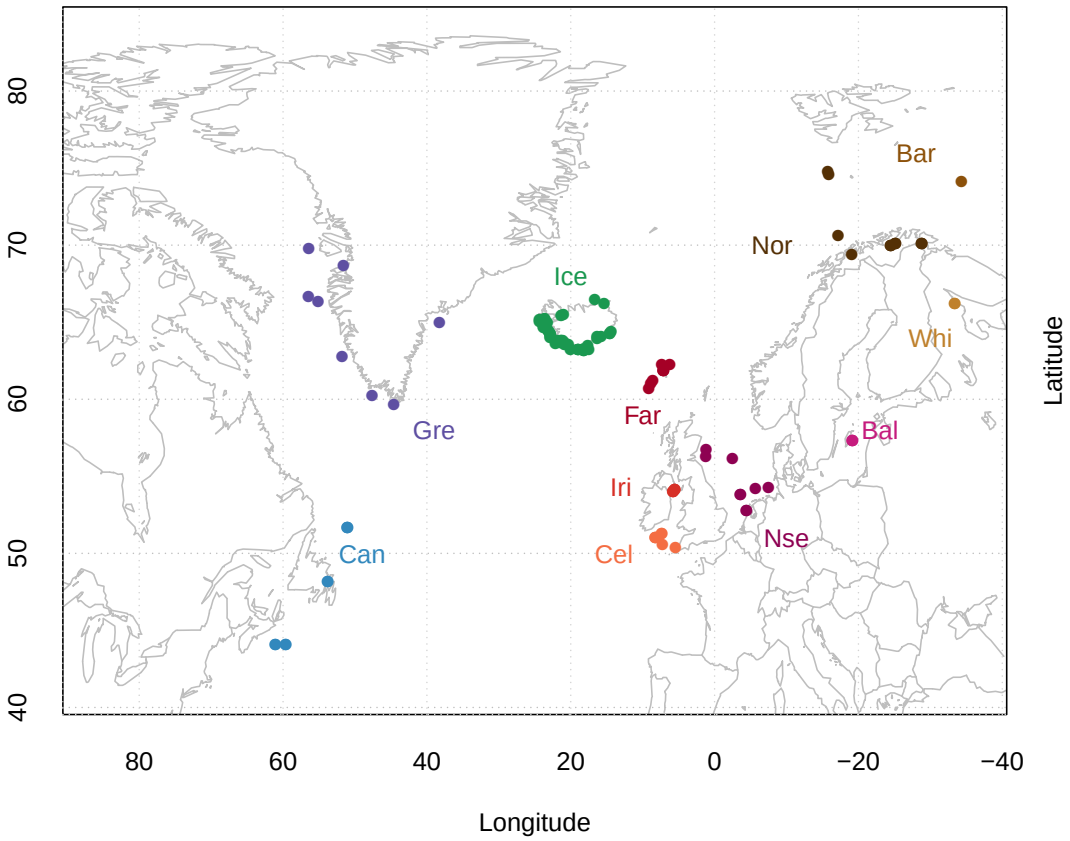

Figure S1. Map of sampling localities For Atlantic cod in the North Atlantic. Sampling localities in the waters of Canada (Nova Scotia and Newfoundland) Can, Greenland Gre, Iceland Ice, Norway Nor, Faroe Islands Far, and from the Barents Sea Bar, White Sea Whi, North Sea Nor, Baltic Sea Bal, Celtic Sea Cel, and Irish Sea

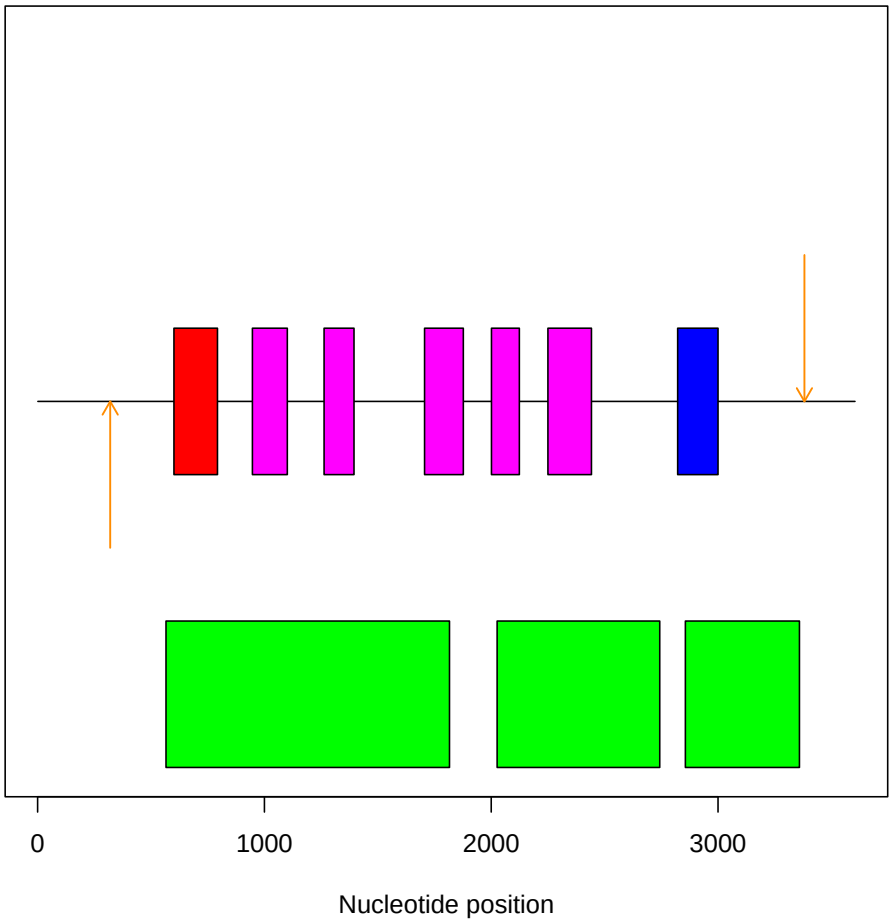

Figure S2. Structure of the *Ckma* gene and sequenced parts. Boxes represent exons, start (red), internal (magenta) and terminal (blue). Green boxes represent sequenced fragments trimmed to Phred score of at least 30. Up and down arrows mark TATA box and poly A signal starts respectively.

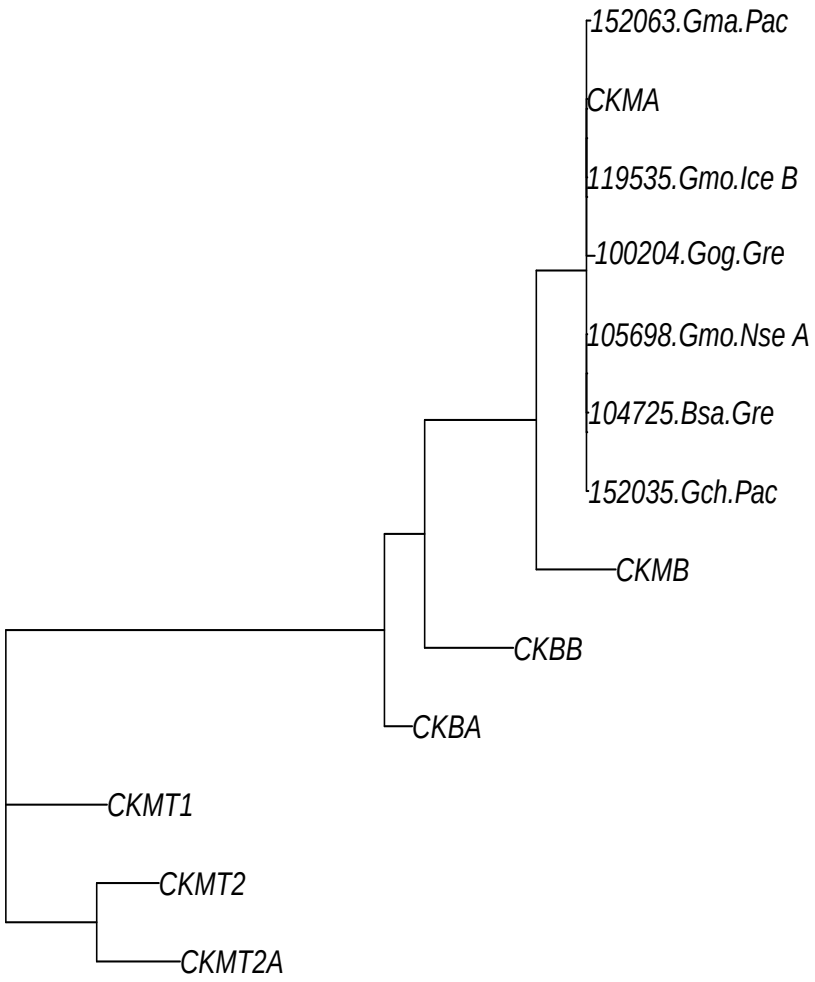

Figure S3. Maximum likelihood tree of creatin kinase proteins of paralogous genes in the Atlantic cod genome, and *Ckma* orthologs from this study the two alleles *A* (*Gmo.Nse A*) and *B* (*Gmo.Ice B*) in Atlantic cod and representatives from sister taxa. Predicted protein isoforms from mitochondria (CKMT), brain (CKB) and muscle (CKM). Sister taxa are *Boreogadus saida* (Bsa), *Gadus macrocephalus* (Gma), *Gadus ogac* (Gog), and *Gadus chalcogrammus* (Gch)

.

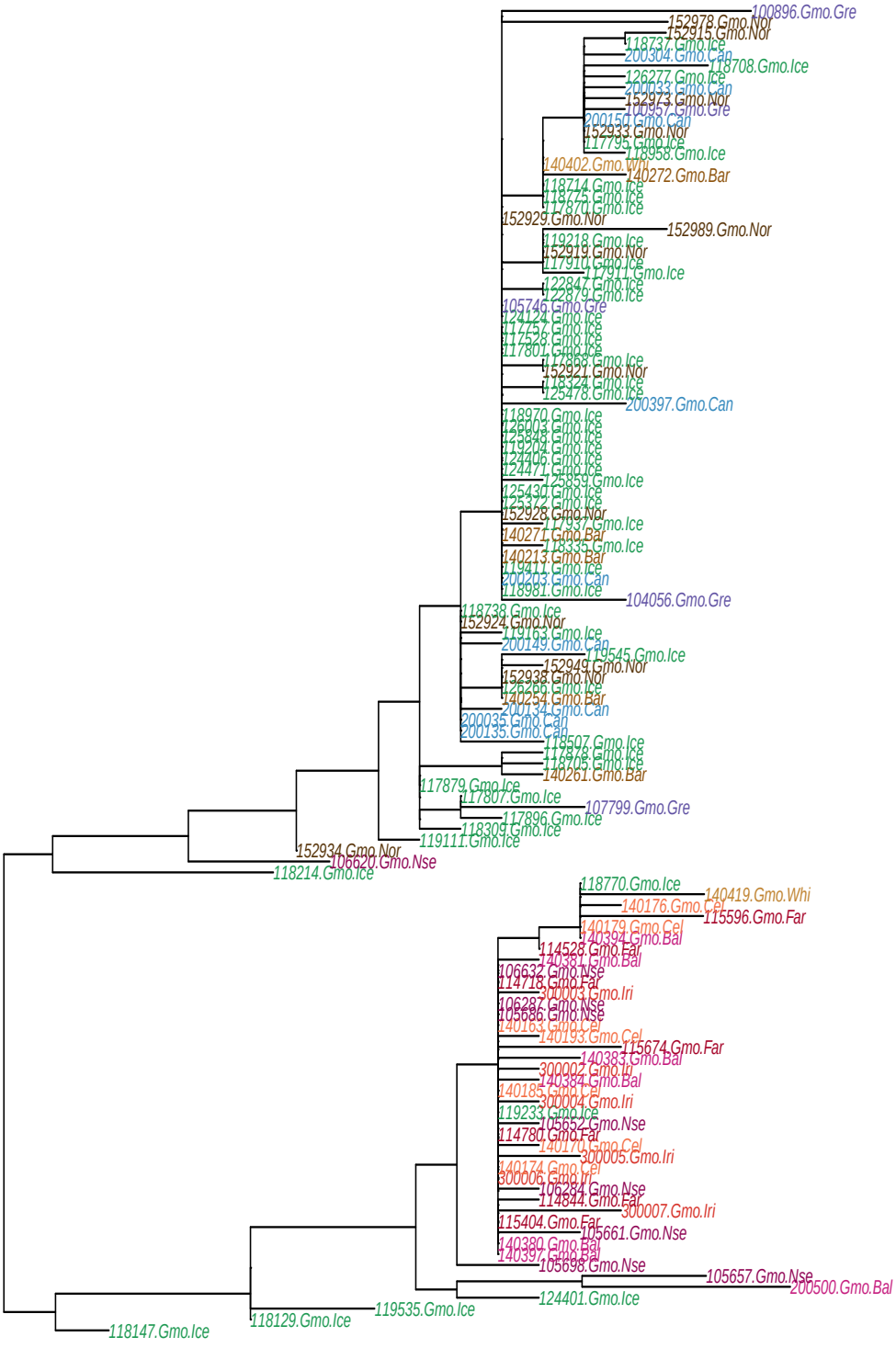

Figure S4. Maximum likelihood tree of variation of *Ckma* among 122 Atlantic cod individuals. Color codes for localities same as in Figure S1.

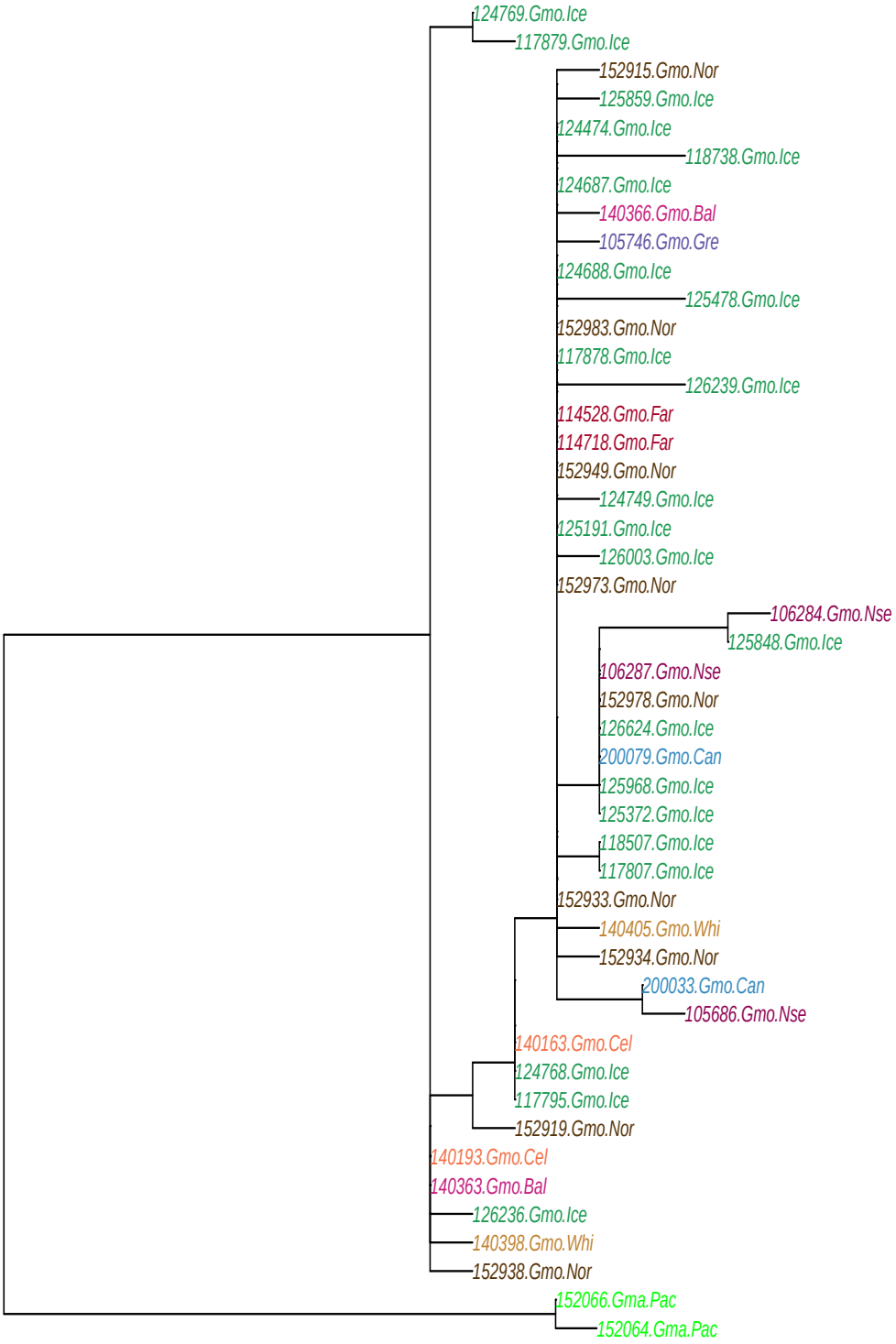

Figure S5. Maximum likelihood tree of nucleotide variation of *Myg* among 45 Atlantic cod and two *Gadus macrocephalus* individuals. Color codes for species and localities same as in Figure S1.

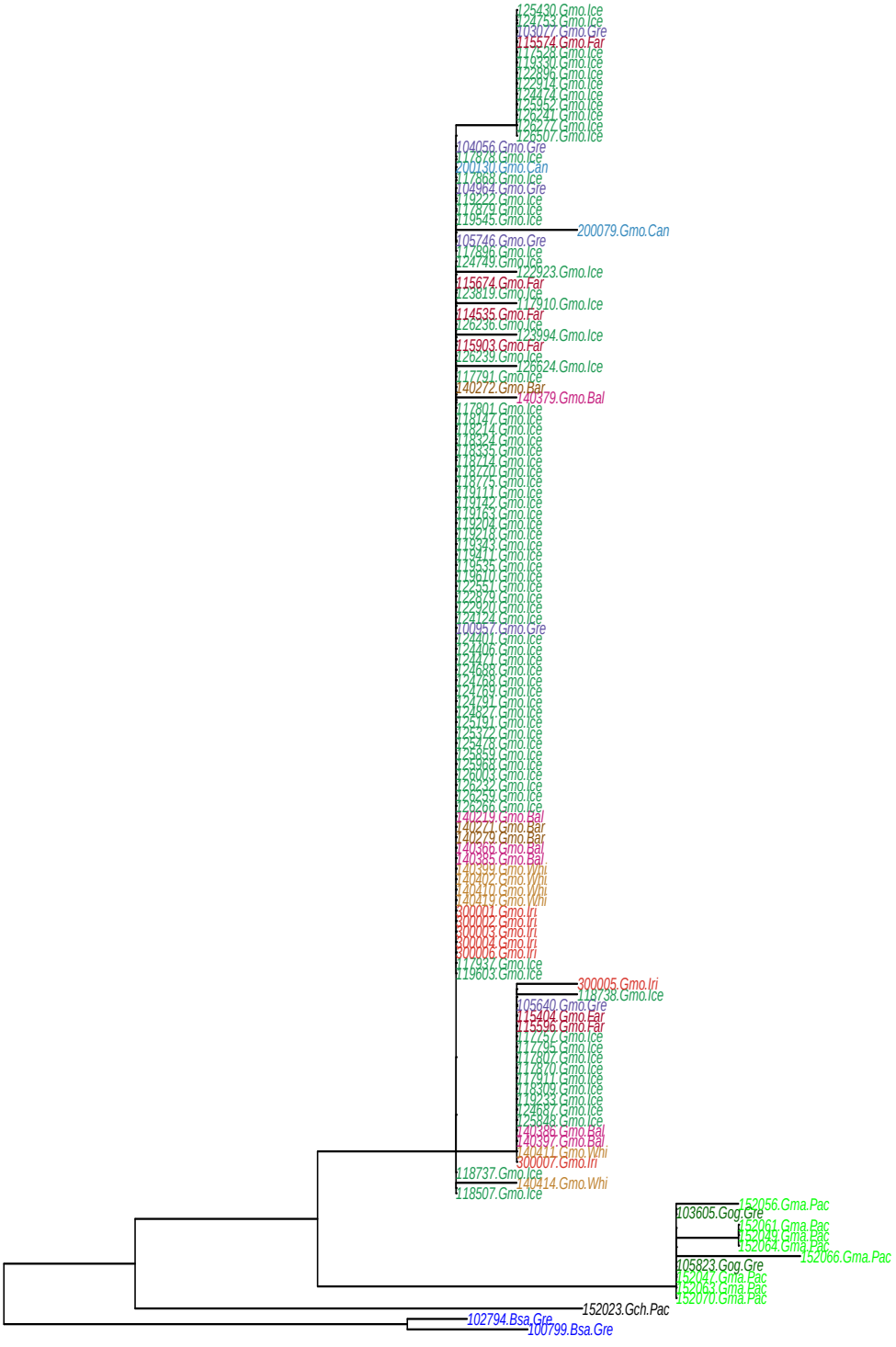

Figure S6. Maximum likelihood tree of nucleotide variation of *HbA2* gene among 113 Atlantic cod and 14 individuals of sister taxa. Color codes for species and localities same as in Figure S1.

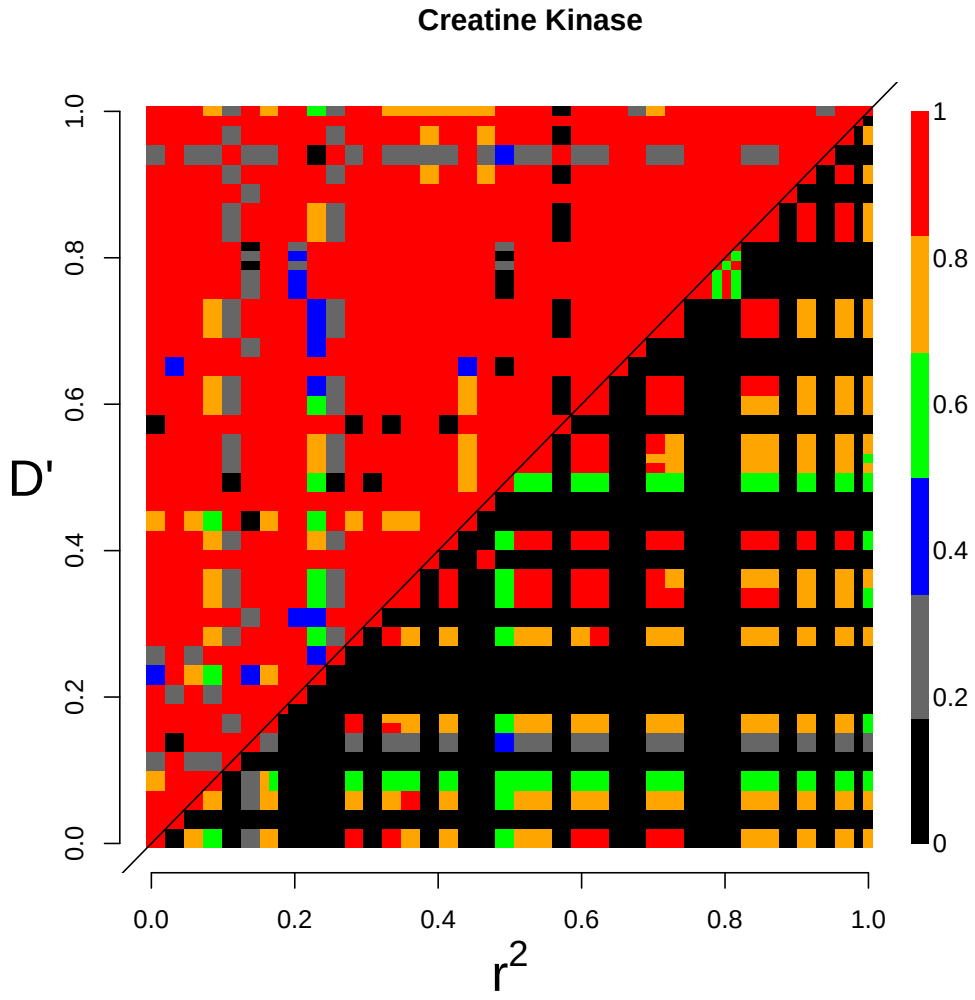

Figure S7. Linkage disequilibrium heatmap of D' and r^2 for 2500 bp fragment of the *Ckma* gene of Atlantic cod

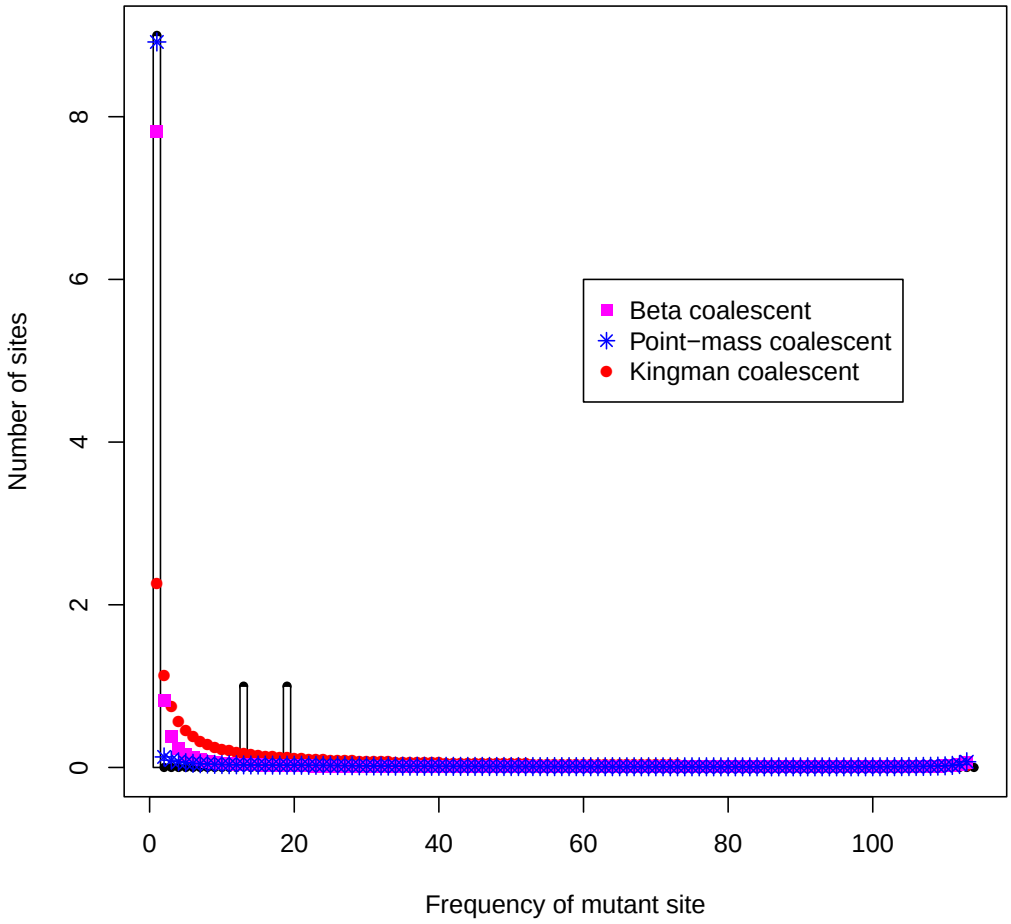

Figure S8. Unfolded site frequency spectrum of Atlantic cod *HbA2* gene. *Gadus macrocephalus* is outgroup. Theroretical expectation under Kingman coalescent (red dots), Beta$(2-\alpha, \alpha)$ coalescent (magenta squares), and point-mass coalescent (blue stars).

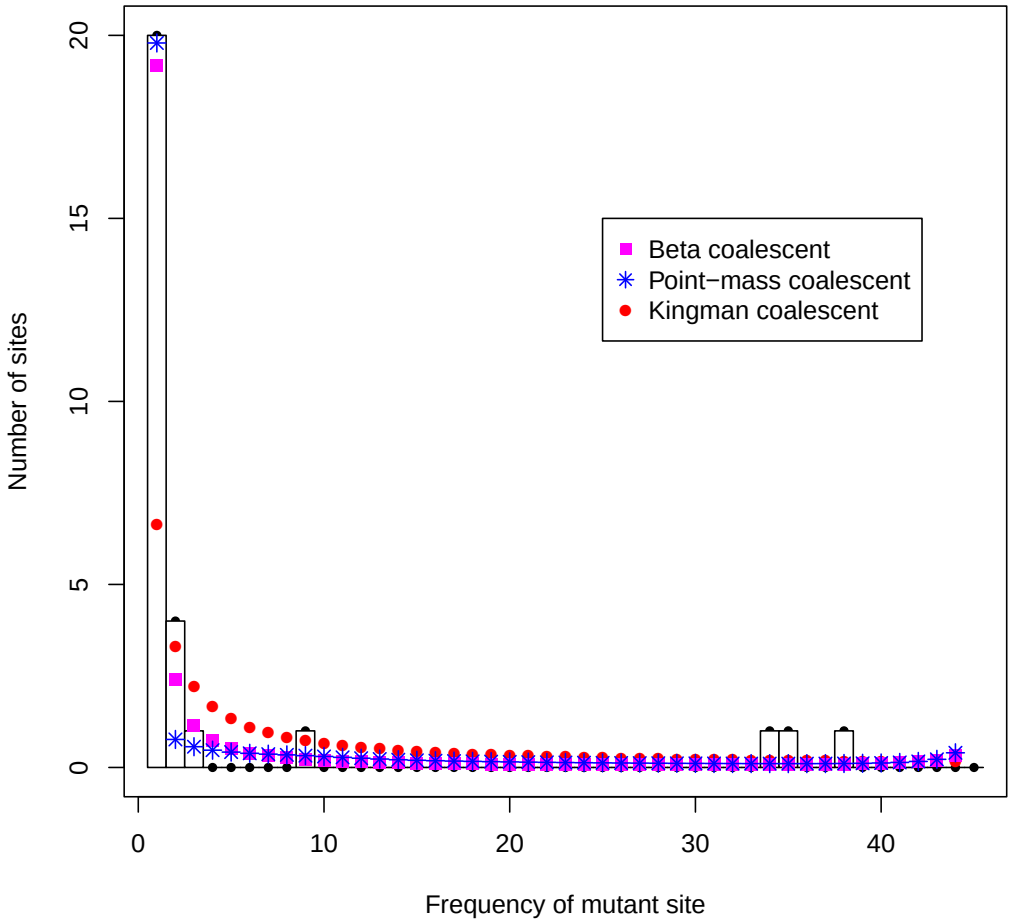

Figure S9. Unfolded site frequency spectrum of Atlantic cod *Myg* gene. *Gadus macrocephalus* is outgroup. Theroretical expectation under Kingman coalescent (red dots), Beta$(2-\alpha, \alpha)$ coalescent (magenta squares), and point-mass coalescent (blue stars).

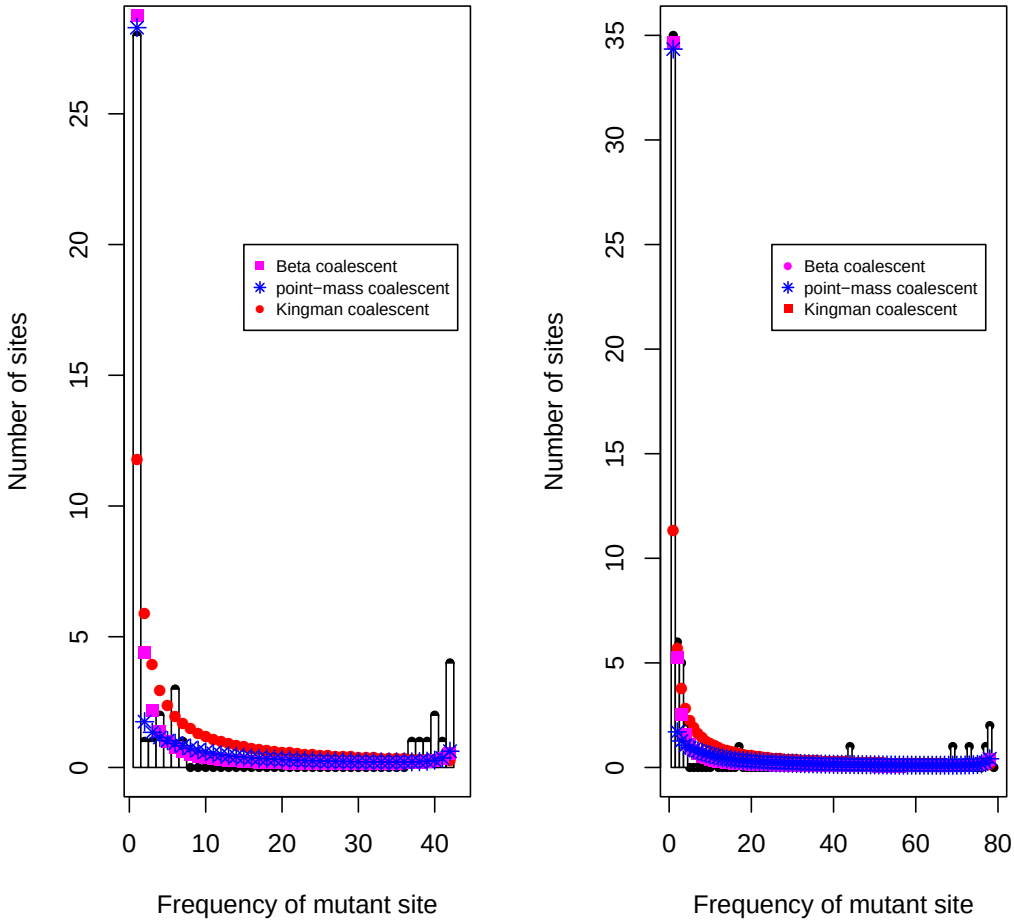

Figure S10. Unfolded site frequency spectrum of Atlantic cod *Ckma A* alleles (left) and *B* alleles (right). *Gadus macrocephalus* is outgroup. Number of individuals $n = 43$ and $n = 79$ respectively. Theroretical expectation under Kingman coalescent (red dots), Beta$(2-\alpha, \alpha)$ coalescent (magenta squares), and point-mass coalescent (blue stars).

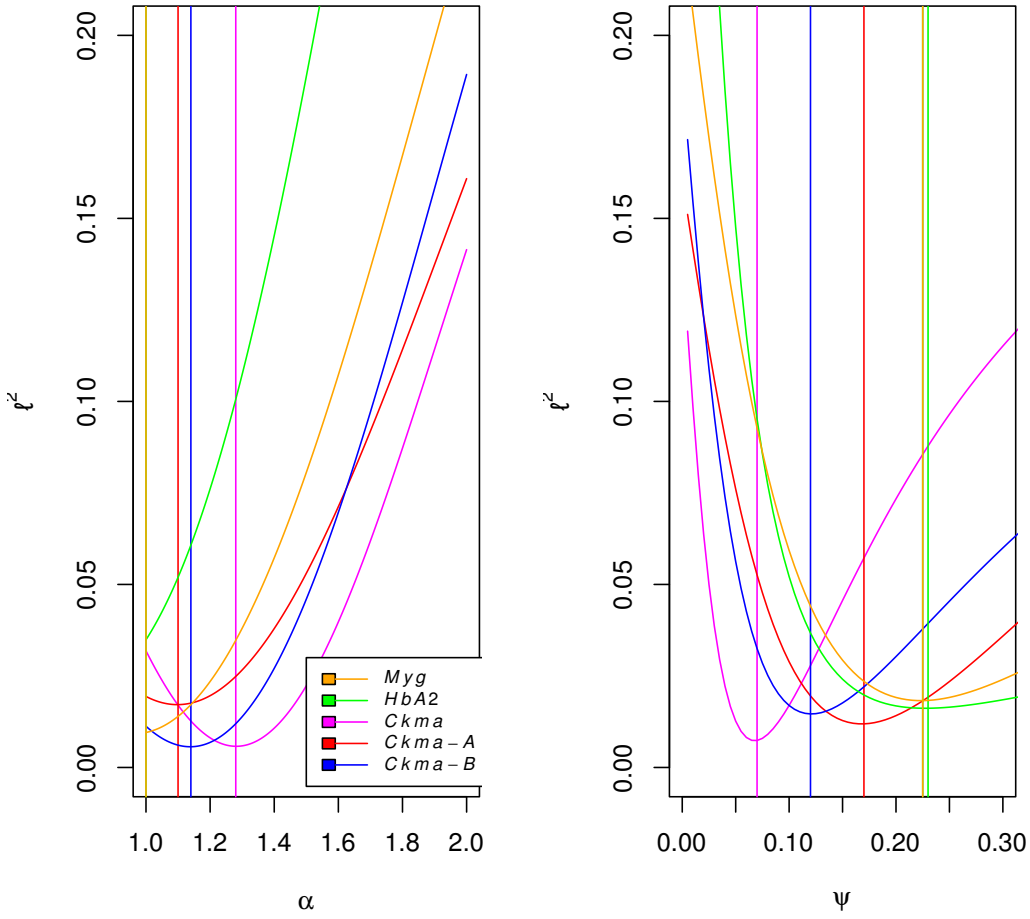

Figure S11. The ℓ^2 distance for the unfolded site frequency spectrum of the nuclear genes *Myg*, *Hb2A*, *Ckma*, and the *Ckma-A* and *Ckma-B* alleles of *Ckma* on the α parameter of the Beta$(2-\alpha, \alpha)$ coalescent (left panel) and the ψ parameter of the point-mass coalescent (right panel).

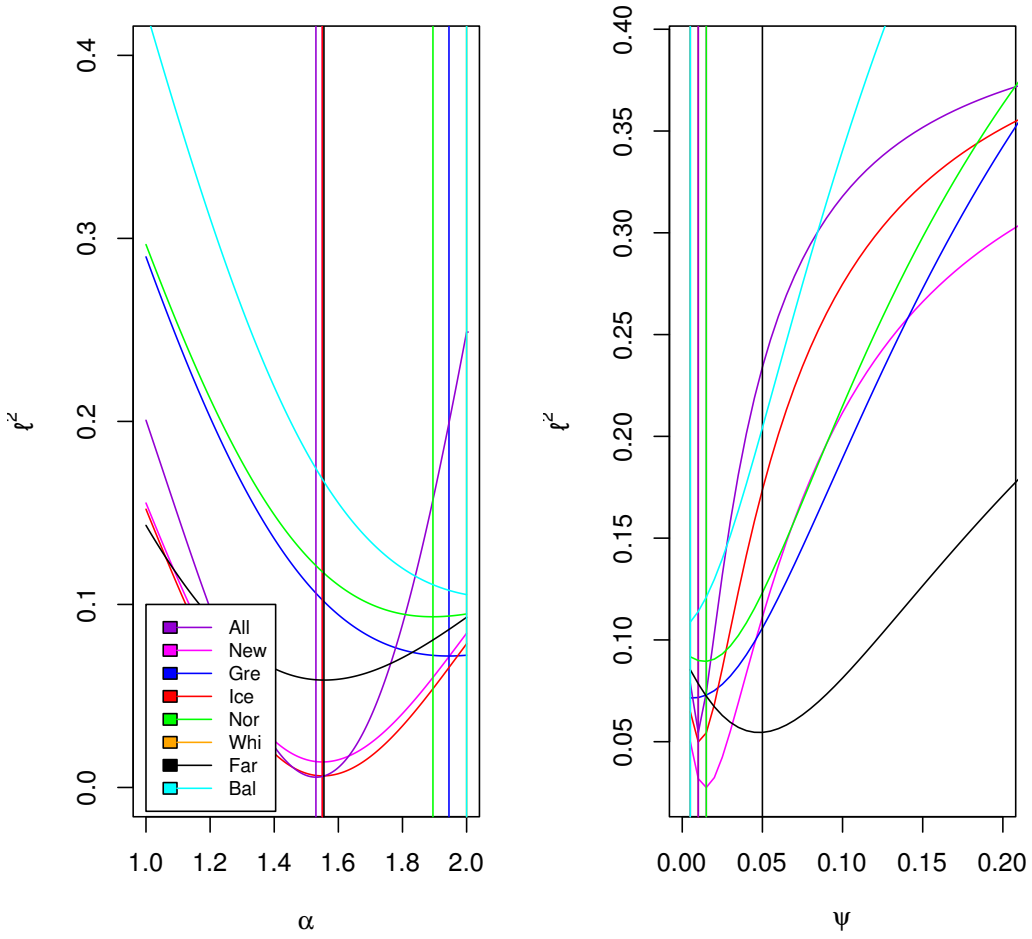

Figure S12. The ℓ^2 distance for the unfolded site frequency spectrum of mtDNA from various localities of the North Atlantic on the α parameter of the Beta$(2-\alpha, \alpha)$ coalescent (left panel) and the ψ parameter of the point-mass coalescent (right panel).

Table S1. Primer sequences for amplification and sequencing fragments of *Ckma* gene from Atlantic cod and sister taxa.

Primer name	Use	Sequence
creL8945	Amplification	5′-GTT TAG GAA TCT ACG CCC ATC CAG AGA CA-3′
creR12945	Amplification	5′-TGG CTA TCA TGC ATT CCC AAT GTT C-3′
creseqR12388	Sequencing	5′-CAT GAC CGT TGG CTG CGT TG-3′
creseqL10486	Sequencing	5′-TCG AAC ACT CCA CCG ACG GA-3′
creseqR10602	Sequencing	5′-ACA GAT TTC GTC TGC CGA GT-3′

Table S2. Segregating sites of the *Ckma* gene among 122 Atlantic cod individuals and 10 individuals of sister taxa (see separate file).

Table S3. Gross D_{xy} and net D_a nucleotide divergence per site between *Gadus morhua* Gmo and *Gadus macrocephalus* Gma and *Gadus chalcogrammus* Gch and between *A* and *B* alleles of Atlantic cod.

Gene	Comparison	D_{xy}	$s_{D_{xy}}$	D_a	s_{D_a}
HbA2	Gmo vs Gma	0.013	0.004	0.012	0.004
HbA2	Gmo vs Tch	0.017	0.014	0.017	0.014
Myg	Gmo vs Gma	0.027	0.007	0.025	0.007
Ckma	Gmo vs Gma	0.014	0.001	0.011	0.001
Ckma	Gmo vs Tch	0.015	0.003	0.013	0.003
Ckma	*A* vs *B*	0.008	0.0005	0.006	0.0005

Divergence and standard deviation found using Jukes and Cantor correction.

Table S4. Non-synonymous changes within and between species.

Individual	amino acids	position
100896.Gmo.Gre	M ⇔ T	37
117937.Gmo.Ice	I ⇔ T	469
152915.Gmo.Nor	S ⇔ G	1158
118708.Gmo.Ice	R ⇔ G	1182
152978.Gmo.Nor	E ⇔ G	2068
152066.Gma.Pac	I ⇔ T	469
152047.Gma.Pac	V ⇔ I	2094
104474.Bsa.Gre	R ⇔ G	1182
104725.Bsa.Gre	Q ⇔ G	1305, 1306, 1307
104474.Bsa.Gre	Q ⇔ G	1305, 1306, 1307

Individuals with species and locality codes, first aa represents majority and the second the change, position refers to position in concatenated sequence in Table S2.

Table S5. Maximum likelihood analysis of a Kingman-coalescent HKA test of neutrality and selection at three genes in Atlantic cod.

					Hbα2		*Myg*		*Ckma*	
Description	$\ln L$	T	Test	*df*	θ	k	θ	k	θ	k
Neutral, all $k = 1$	-18.66	2.46			0.0035	1	0.0068	1	0.0056	1
Selection at *Ckma*	-17.47	3.86	2.38	1	0.0029	1	0.0054	1	0.0032	2.12

Test is twice the $\ln L$ difference of the two models, neutrality and selection at *Ckma*. Three loci are under test: Hemoglobin α 2 (*Hbα2*), Myoglobin (*Myg*), and Creatine Kinase Muscle (*Ckma*). θ is the scaled effective population size and the parameter k measures changes in diversity due to selection. Based on method of Wright and Charlesworth (2004).

Table S6. Frequency of A and B alleles in different localities.

	Can	Gre	Ice	Nor	Bar	Whi	Far	Nse	Bal	Cel	Iri	**Sum**
A allele	0	0	7	0	0	1	7	8	7	7	6	**43**
B allele	9	5	45	13	5	1	0	1	0	0	0	**79**
Sum	9	5	52	13	5	2	7	9	7	7	6	**122**

Table S7. Pairwise F_{ST} values (lower triangular) of population differentiation among localities.

	Can	Gre	Ice	Nor	Bar	Far	Nse	Bal	Cel	Iri
Can		0.78	0.06	0.11	0.14	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Gre	0.04		0.22	0.55	0.12	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Ice	0.08	0.01		0.21	0.53	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Nor	0.03	0.00	0.02		0.13	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Bar	0.01	-0.08	-0.02	-0.06		**0.00**	**0.00**	**0.00**	**0.00**	**0.00**
Far	0.88	0.81	0.77	0.84	0.84		0.41	0.88	0.21	0.22
Nse	0.78	0.71	0.65	0.74	0.74	0.04		0.93	0.48	0.16
Bal	0.82	0.76	0.71	0.78	0.78	-0.01	-0.03		0.36	0.82
Cel	0.91	0.84	0.80	0.87	0.87	-0.05	0.07	-0.01		0.09
Iri	0.88	0.82	0.78	0.84	0.84	0.02	0.04	0.01	0.04	

Probabilities in black on upper triangular, boldface are significant P values. North (blue) and South (red) defined ad hoc by results.

Table S8. Pairwise F_{ST} of *Ckma* gene between North and South (lower left corner) and between *A* and *B* alleles (upper right corner).

	North	*B* allele
A allele	0.738	0.804
South	0.763	0.828

North and South populations defined according to differentiation at *Ckma* locus.

Table S9. Pairwise F_{ST} of neutral genes between North and South.

	North	South
North		-0.029
South	0.004	

HbA2 locus on lower left corner, *Myg* locus on upper right corner. North and South populations defined according to differentiation at *Ckma* locus.

Table S10. Likelihood ratio test statistics G for observed site frequency spectra and expectation according to different coalescent models.

Model	G	Comparison	$2\Delta G$	df
I. Kingman	149.26			
II. Beta$(2-\alpha,\alpha)$	116.21	I vs II	66.10	1
III. point-mass	114.42	I vs III	69.69	1