Comment on "Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage"

Shigeki Nakagome *et al.*

Science **339**, 1522 (2013);
DOI: 10.1126/science.1227339

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.

Comment on "Nuclear Genomic Sequences Reveal that Polar Bears Are an Old and Distinct Bear Lineage"

Shigeki Nakagome,[1]* Shuhei Mano,[1] Masami Hasegawa[1,2]

Based on nuclear and mitochondrial DNA, Hailer *et al*. (Reports, 20 April 2012, p. 344) suggested early divergence of polar bears from a common ancestor with brown bears and subsequent introgression. Our population genetic analysis that traces each of the genealogies in the independent nuclear loci does not support the evolutionary model proposed by the authors.

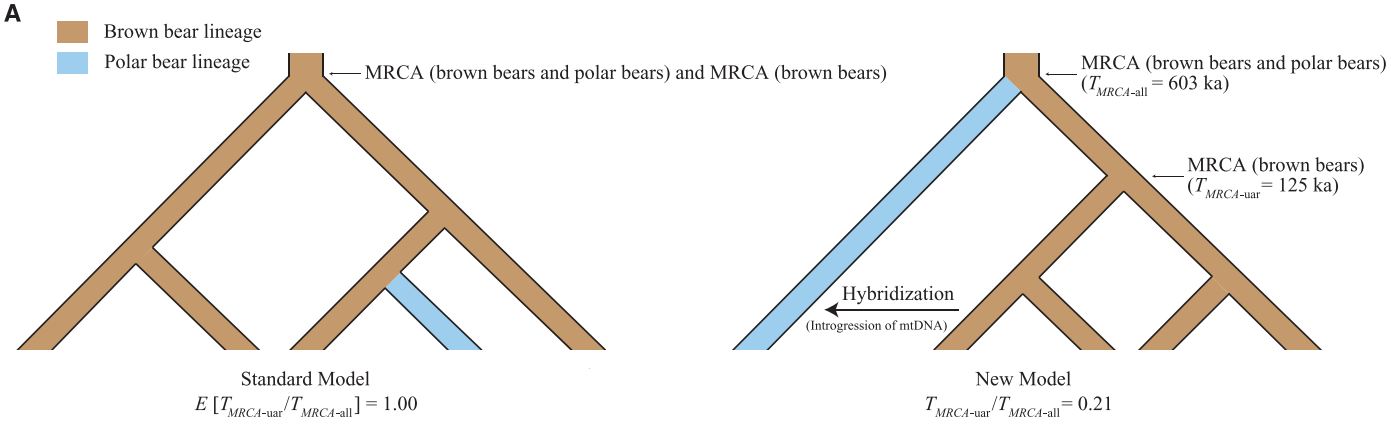
Hailer *et al*. (*1*) sequenced 14 independent nuclear loci across the genomes of polar (*Ursus maritimus*), brown (*U. arctos*), and American black bears (*U. americanus*). The Bayesian multilocus coalescent approach showed one species tree in which polar bears split from brown bears before the diversification of brown bear lineages. The divergence of polar bears was estimated to have occurred 603 thousand years ago (ka). These results conflict with the mitochondrial DNA phylogeny, which defines the root of the polar bear lineage within brown bear diversity (*2*, *3*) ("standard model" in Fig. 1A). Additional phylogenetic analysis using the concatenated sequences supported the sister lineage of polar bears to all brown bears. Based on these observations, the authors proposed a new evolutionary model in which polar bears diverged from the common ancestor of all extant brown bears, whereas hybridization with female brown bears facilitated introgression ("new model" in Fig. 1A).

Here, we focused on a unique genealogical history in each locus and applied population genetic analysis to the same data set evaluated in Hailer *et al*. (*1*), including 14 unlinked nuclear loci from brown bears ($2N = 36$) and polar bears ($2N = 36$). A crucial difference between the standard model and the new model is the time to the most recent common ancestor (MRCA) in brown bears ($T_{\text{MRCA-uar}}$) (Fig. 1A). In the standard model, the $T_{\text{MRCA-uar}}$ is the same for all of the samples from brown and polar bears ($T_{\text{MRCA-all}}$) ($T_{\text{MRCA-all}} = T_{\text{MRCA-uar}}$). The ratio of $T_{\text{MRCA-uar}}$ to $T_{\text{MRCA-all}}$ ($T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$) is expected to be 1.00 in this model. In contrast, the new model shows that the $T_{\text{MRCA-all}}$ is older than the $T_{\text{MRCA-uar}}$. From the estimates of $T_{\text{MRCA-all}}$ (603 ka) and $T_{\text{MRCA-uar}}$ (125 ka), Hailer *et al*. (*1*) estimated the ratio of 0.21. We used the $T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$ as a test statistic to ask whether the observed genetic variation significantly deviates from the null hypothesis of the standard model.

We estimated the $T_{\text{MRCA-all}}$, $T_{\text{MRCA-uar}}$, and T_{MRCA} in polar bears ($T_{\text{MRCA-uma}}$) for 13 loci using the GENETREE program (*4*) (Table 1).

[1]Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 180-8562, Japan. [2]School of Life Sciences, Fudan University, Shanghai 200433, China.

*Corresponding author. E-mail: nakagome@ism.ac.jp



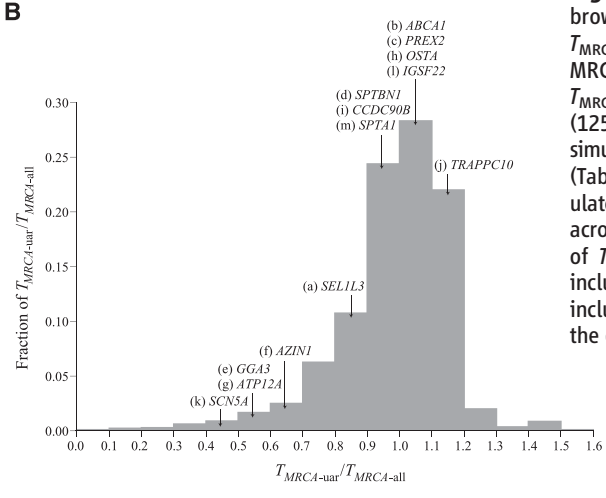


Fig. 1. Testing two evolutionary models. (**A**) In the standard model, the point of MRCA in brown bears is the same for all of the samples from brown and polar bears. The expected $T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$ ratio is 1.00. In contrast, the new model suggests that the points of MRCA are different between them and that the $T_{\text{MRCA-all}}$ is expected to be older than the $T_{\text{MRCA-uar}}$. Hailer *et al*. (*1*) estimated the ratio of 0.21 from $T_{\text{MRCA-all}}$ (603 ka) and $T_{\text{MRCA-uar}}$ (125 ka). (**B**) Histogram of the $T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$ ratio under the standard model. We generated simulated samples ($2N = 72$) with the program *ms* (*8*) using an average value of $\theta_{\text{ML}} = 1.41$ (Table 1). The T_{MRCA} in the samples and subsamples ($2N = 36$) was estimated for the simulated data using GENETREE, assuming a mutation rate of 7.921×10^{-6} (the average value across the 13 loci). These steps were repeated for 10,000 iterations to obtain the distribution of $T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$. The horizontal axis indicates the values of the $T_{\text{MRCA-uar}}/T_{\text{MRCA-all}}$, including the ratios binned with 0.1, while the vertical axis indicates the fraction of the ratios included in each bin. Each of the 13 loci was assigned into one of the bins corresponding to the observed values.

Table 1. GENETREE analysis for brown and polar bears, brown bears, and polar bears. Insertion and deletion polymorphisms were excluded in the GENETREE analysis. Recombinant sequences were also excluded in this analysis based on the infinite-site model. We confirmed that most of the recombinants were rare in a population (n = 1, 2, or 3) and that the segregating sites involved in the exclusion were also rare, which were thought to be recent mutations. These data are unlikely to affect our estimation of the T_{MRCA} or alter the overall results. The *LRGUK* locus was not included in this analysis because there were many recombinants with high frequencies in brown bears. N.A. denotes that coalescent times could not be assessed due to the lack of a mutation. μ, mutation rate per locus per generation; N_e, effective population size.

		Brown and polar bears				Brown bears				Polar bears					
Loci	μ*	S†	θ_{ML}‡	N_e§	$T_{MRCA\text{-}all}$‖	S	θ_{ML}	N_e	$T_{MRCA\text{-}uar}$	S	θ_{ML}	N_e	$T_{MRCA\text{-}uma}$	Ratio of $T_{MRCA\text{-}uar}$ / $T_{MRCA\text{-}all}$	P¶ (threshold: P = 0.004#)
(a) *SEL1L3***	6.869×10^{-6}	9	1.70	61,870	2,992,294	6	1.30	47,313	2,456,657	3	0.75	27,296	907,145	0.821	0.133
(b) *ABCA1*	6.366×10^{-6}	9	1.80	70,691	3,509,803	8	1.90	74,618	3,616,146	1	0.25	9,818	344,972	1.030	0.554
(c) *PREX2*	8.270×10^{-6}	8	1.30	39,300	2,631,307	8	1.60	48,369	2,818,974	0	N.A.	N.A.	N.A.	1.071	0.634
(d) *SPTBN1*	6.505×10^{-6}	10	2.10	80,712	2,566,795	7	1.70	65,338	2,486,114	2	0.50	19,217	695,582	0.969	0.369
(e) *GGA3*	6.921×10^{-6}	6	1.10	39,732	1,866,630	3	0.80	28,896	953,347	0	N.A.	N.A.	N.A.	0.511	0.018
(f) *AZIN1*	5.405×10^{-6}	5	1.10	50,878	1,760,479	3	0.80	37,002	1,090,971	1	0.22	10,176	486,190	0.620	0.036
(g) *ATP12A*††	8.015×10^{-6}	8	1.80	56,144	1,784,831	4	1.10	34,310	1,040,019	2	0.50	15,596	538,955	0.583	0.029
(h) *OSTA*	8.623×10^{-6}	6	1.10	31,893	1,626,396	5	1.10	31,893	1,677,361	1	0.25	7,248	260,084	1.031	0.557
(i) *CCDC90B*	6.053×10^{-6}	5	1.00	41,300	1,563,955	4	1.00	41,300	1,502,253	3	0.54	22,302	1,475,775	0.961	0.355
(j) *TRAPPC10*	1.025×10^{-5}	7	1.30	31,711	1,407,140	7	1.20	29,272	1,686,337	0	N.A.	N.A.	N.A.	1.198	0.968
(k) *SCN5A*	8.623×10^{-6}	5	0.90	26,094	1,366,802	2	0.44	12,757	578,660	0	N.A.	N.A.	N.A.	0.423	0.011
(l) *IGSF22*	8.131×10^{-6}	5	1.00	30,747	1,274,600	5	1.30	39,972	1,369,025	0	N.A.	N.A.	N.A.	1.074	0.645
(m) *SPTA1*	1.295×10^{-5}	10	2.10	40,554	1,370,006	8	2.00	38,623	1,309,866	2	0.50	9,656	318,873	0.956	0.345
Average	7.921×10^{-6}	7	1.41	46,279	1,978,541	5	1.25	40,743	1,737,364	1	0.44	15,164	628,447	0.865	

*Mutation rates were calculated from the number of substitutions between the giant panda and brown and/or polar bears, assuming the divergence of the giant panda and brown/polar bears 12 Ma. †The number of segregating sites in a population. ‡Maximum-likelihood estimates of the scaled population mutation rate, θ. §N_e was calculated from $\theta_{ML} = 4N_e\mu$. ‖T_{MRCA} was calculated from $T_{MRCA} = 2N_e t \times 10$ (years/generation), where t is the time in coalescent-time units estimated by GENETREE. ¶Empirical P values represent the fraction of ratios ($T_{MRCA\text{-}uar}/T_{MRCA\text{-}all}$) less than the observed ratio at each locus to 10,000 simulated points. #The significance threshold was corrected using the Bonferroni correction. **The total aligned sequences of 615 base pairs in the *SEL1L3* locus included recombinants from brown bears, and nucleotide positions from 1 to 395 were used for the GENETREE analysis. ††One polymorphic site at the *ATP12A* locus was excluded due to an unknown state of an ancestral allele (the giant panda with "A" and the brown and polar bears with "G/C").

One locus, *LRGUK*, from Hailer *et al.* (*1*), was excluded in this analysis due to high frequencies of recombinants, whereas only recombinant sequences were excluded in the other loci. A genealogy of each locus was deduced based on the path of mutations to the MRCA under the infinitely many-site model. We computed the maximum likelihood estimates for the population mutation rate (θ_{ML}) under the constant size model in which we specified the range of θ_{ML} (*5*). The empirical distribution of the T_{MRCA} was obtained based on the likelihood estimated from each simulation run, conditional on the topology of the gene tree and θ_{ML}. The ancestral/derived state of an allele at a segregating site was inferred by alignment with the giant panda sequence (*6*). The genealogies were different for the 13 loci, and each locus had its own T_{MRCA}. Nine of the 13 loci showed that $T_{MRCA\text{-}uar}/T_{MRCA\text{-}all}$ ratios were 0.821 to 1.198. These results appear to support the standard model and the expected $T_{MRCA\text{-}uar}/T_{MRCA\text{-}all}$ value of 1.00.

To determine whether the observed values of the $T_{MRCA\text{-}uar}/T_{MRCA\text{-}all}$ could reject the null hypothesis, we generated a distribution of the ratios under the standard model (Fig. 1B). Empirical P values were calculated for the 13 loci (Table 1). The estimate in Hailer *et al.* (*1*) ($T_{MRCA\text{-}uar}/T_{MRCA\text{-}all}$ = 0.21) significantly deviates from the null distribution (P = 0.002). However, we found that all of the P values for the 13 loci were greater than the threshold for significance (corrected P = 0.004) and that the standard model could not be rejected. The estimates for the $T_{MRCA\text{-}uma}$ were consistently lower than the estimates for the $T_{MRCA\text{-}all}$ and $T_{MRCA\text{-}uar}$, which is consistent with the standard model because the $T_{MRCA\text{-}uma}$ is expected to be more recent than the $T_{MRCA\text{-}all}$ and $T_{MRCA\text{-}uar}$. These results provide more information on relevant arguments against the new model.

A recent study using a diploid genome pointed out that hundreds of thousands of independent loci within an individual have different T_{MRCA} (*7*). Our estimations of the $T_{MRCA\text{-}all}$ are 1.3 to 3.5 million years ago (Ma), whereas the $T_{MRCA\text{-}uma}$ is estimated to be 0.3 to 1.5 Ma (Table 1). Genealogies with $T_{MRCA\text{-}all}$ older than 1.5 Ma may be useful for tracing the population history before the divergence of brown and polar bears. The genealogies of the *SEL1L3*, *ABCA1*, *PREX2*, and *SPTBN1* loci indicate that the lineages leading to polar bears occurred during the diversification of brown bear lineages, which supports the standard model. Our population genetic analysis indicates that the observed patterns of genetic variation in nuclear loci can be explained by the recent origin of polar bears and ancestral polymorphisms.

In summary, we conclude that the 13 loci reported by Hailer *et al.* (*1*) fail to support the new model. The sequence data from the 13 loci are not sufficient to resolve the origin of the bears. Genome-scale sequence data are necessary to untangle the complex evolutionary history of brown bears and polar bears.

References and Notes

1. F. Hailer *et al.*, *Science* **336**, 344 (2012).
2. S. L. Talbot, G. F. Shields, *Mol. Phylogenet. Evol.* **5**, 477 (1996).
3. C. J. Edwards *et al.*, *Curr. Biol.* **21**, 1251 (2011).
4. R. C. Griffiths, www.stats.ox.ac.uk/~griff/software.html (2007).
5. R. C. Griffiths, S. Tavare, *Stat. Sci.* **9**, 307 (1994).
6. R. Li *et al.*, *Nature* **463**, 311 (2010).
7. H. Li, R. Durbin, *Nature* **475**, 493 (2011).
8. R. R. Hudson, *Bioinformatics* **18**, 337 (2002).

Acknowledgments: We thank T. Yonezawa for helpful comments on this manuscript. S.N. was supported by the Japan Society for the Promotion of Science (JSPS) Research Fellowship (24-3234). M.H. was supported by a Grant-in-Aid for Scientific Research C22570099 from JSPS.
Author contributions: S.N., S.M, and M.H. designed the study; S.N. performed the analysis; and S.N., S.M, and M.H. wrote the manuscript.

12 July 2012; accepted 22 February 2013
10.1126/science.1227339