

Reconstructing large regions of an ancestral mammalian genome in silico

Mathieu Blanchette, Eric D. Green, Webb Miller, and David Haussler

The numbers reported in Table 2 mistakenly refer to the differences between the genomes of living species and the reconstructed Euarchontoglires ancestor, not the reconstructed Boreoeutherian ancestor. The data for the Boreoeutherian ancestor are listed in the corrected Table 2, which is printed below. The authors apologize for any confusion this may have caused.

Table 2. Comparison of modern sequences to predicted ancestor

Species	Size of region (kb) (a)	Nonrepetitive %GC-content (b)	Deletions	Insertions	Substitutions
			% of ancestor lost (c)	% of extant species' bases acquired (nonrepetitive only) (d)	% of extant species' bases changed (expected # substitutions per site) (e)
Reconstructed Boreoeutherian ancestor	1105	37.5	N/A	N/A	N/A
Human	1296	37.3	16.2	28.6 (9.1)	8.7 (13.4)
Chimpanzee	1278	37.4	16.6	28.6 (9.0)	8.7 (13.4)
Gorilla	1264	37.4	17.6	28.0 (9.1)	8.8 (13.5)
Baboon	1267	37.5	17.0	27.2 (9.5)	9.2 (14.0)
Orangutan	1300	37.3	17.0	28.4 (8.9)	8.7 (13.5)
Vervet	1243	37.5	17.9	27.4 (9.5)	9.2 (14.1)
Macaque	1260	37.5	17.0	27.9 (9.9)	9.2 (14.1)
Lemur	1043	38.2	23.6	19.5 (11.3)	9.6 (13.1)
Mouse-lemur	1071	37.7	23.3	21.5 (12.3)	9.9 (13.9)
Mouse	1147	39.4	40.9	43.0 (24.1)	15.9 (35.6)
Rat	1277	39.6	40.9	49.0 (25.3)	14.4 (36.3)
Rabbit	1379	40.9	31.3	47.2 (31.1)	10.4 (23.2)
Cat	1217	38.2	19.8	27.2 (11.6)	9.7 (14.8)
Dog	1125	39.7	23.2	24.3 (12.6)	11.6 (17.4)
Cow	1317	37.4	20.8	33.9 (12.2)	10.0 (17.1)
Pig	1209	37.0	21.2	26.9 (12.2)	10.6 (16.3)
Horse	1133	38.7	15.0	17.9 (9.7)	8.7 (11.5)
Hedgehog	1545	39.8	46.9	63.2 (57.9)	8.2 (27.1)
Armadillo	1397	39.5	25.7	41.3 (37.0)	9.6 (18.7)

Listed are some properties of sequences of the extant species in the greater-*CFTR* locus and the predicted changes they incurred during evolution from the Boreoeutherian ancestral sequence. (a) Length of sequence. (b) Fraction of nonrepetitive bases that are G or C. (c) Deletions: percentage of the ancestral sequence lost in each species. (d) Insertions: percentage of extant species' sequence that was inserted since the reconstructed ancestor (in parentheses, percentage of extant species' sequence that resulted from insertions of nonrepetitive sequences, using RepeatMasker to identify repetitive sequences.) The high fraction of nonrepetitive inserted bases in rabbit and hedgehog is most likely due to lack of complete RepeatMasker libraries for the transposons specific to these species. (e) Substitutions: percentage of extant species' bases that were derived from an ancestral base but differ from that base (this is different from the standard percentage identity measure, where only aligned bases are considered). In parentheses, the expected number of substitutions per site under a Kimura 2-parameter model (Kimura 1980) is given, here using only the aligned bases.