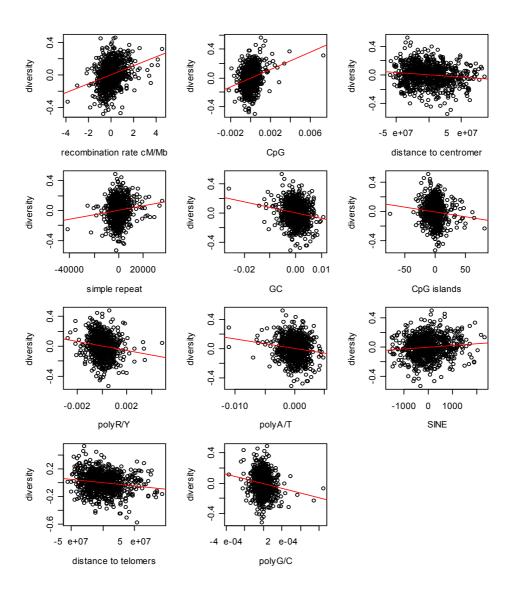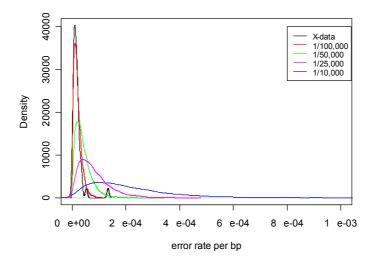**Supplementary Figure 1.** Scatterplots of the residuals from the full multiple linear regression of non-repeat human diversity on each predictor variable.

Supplementary Figure 2. Various distributions of the sequencing error rate. The black curve is the sequencing error rate as estimated from Clint's X-chromosome (see Methods). The red curve is a Γ-distribution meant to approximate the observed distribution, while the remaining curves are for higher error rates (see Methods and Table 3).

Supplementary Figure 1

Supplementary Figure 2

Supplementary Table 1. All pairwise rank-correlation coefficients for the putative predictors of diversity and divergence.

τ	recombination rate (cM/Mb)	LINE-count	SINE-count	CpG islands	AluY-count	CpG content	GC-content	Poly A/T	gene content	simple repeat content	Poly G/C	expression strength	expression breadth	fraction of genes expressed in testes	fraction of genes expressed in ovaries	distance to telomeres	distance to centromere	Poly (CA)	Poly R/Y
p-values																			
recombination rate (cM/Mb)		-0.123	0.079	0.197	0.127	0.321	0.288	-0.305	0.082	0.193	0.231	0.099	0.000	0.079	0.076	-0.341	0.074	0.266	0.261
LINE-count	0.0000		0.057	-0.049	-0.026	-0.073	-0.039	0.037	-0.024	-0.040	-0.039	0.007	0.030	-0.012	-0.015	0.233	-0.099	-0.072	-0.036
SINE-count	0.0006	0.0131		0.122	0.108	0.153	0.155	-0.155	0.080	0.010	0.161	0.074	0.020	0.105	0.086	-0.098	0.027	0.060	0.109
CpG islands	0.0000	0.0336	0.0000		0.412	0.717	0.709	-0.631	0.389	0.017	0.706	0.357	0.157	0.570	0.478	-0.197	-0.001	0.213	0.401
AluY-count	0.0000	0.2615	0.0000	0.0000		0.485	0.402	-0.312	0.315	0.019	0.431	0.244	0.126	0.406	0.343	-0.125	-0.068	0.091	0.399
CpG content	0.0000	0.0016	0.0000	0.0000	0.0000		0.840	-0.734	0.348	0.121	0.709	0.335	0.130	0.472	0.394	-0.323	0.018	0.295	0.517
GC-content	0.0000	0.0872	0.0000	0.0000	0.0000	0.0000		-0.869	0.352	0.101	0.690	0.359	0.137	0.476	0.402	-0.279	0.013	0.298	0.497
polyA/T	0.0000	0.1061	0.0000	0.0000	0.0000	0.0000	0.0000		-0.308	-0.141	-0.621	-0.344	-0.117	-0.415	-0.356	0.297	-0.034	-0.336	-0.461
gene content	0.0004	0.3105	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000		-0.115	0.397	0.229	0.171	0.423	0.378	-0.078	0.018	0.051	0.265
simple repeat content	0.0000	0.0807	0.6727	0.4577	0.4100	0.0000	0.0000	0.0000	0.0000		0.060	-0.003	-0.075	-0.128	-0.134	-0.296	0.038	0.461	0.189
polyG/C	0.0000	0.0940	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0096		0.341	0.169	0.517	0.437	-0.233	0.014	0.236	0.411
expression strength	0.0000	0.7738	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8798	0.0000		0.233	0.374	0.369	-0.087	0.001	0.112	0.223
expression breadth	0.9943	0.1908	0.3775	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0012	0.0000	0.0000		0.284	0.263	0.047	-0.013	-0.025	0.084
fraction of genes expressed in testes	0.0006	0.6132	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		0.707	-0.045	-0.043	0.039	0.330
fraction of genes expressed in ovaries	0.0009	0.5247	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		-0.040	-0.024	0.024	0.301
distance to telomeres	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0008	0.0000	0.0000	0.0002	0.0422	0.0505	0.0831		-0.271	-0.287	-0.228
distance to centromere	0.0012	0.0000	0.2481	0.9598	0.0034	0.4440	0.5664	0.1419	0.4397	0.1013	0.5341	0.9669	0.5868	0.0598	0.2893	0.0000		0.054	-0.028
poly(CA)	0.0000	0.0017	0.0090	0.0000	0.0001	0.0000	0.0000	0.0000	0.0283	0.0000	0.0000	0.0000	0.2809	0.0869	0.3016	0.0000	0.0198		0.248
polyR/Y	0.0000	0.1211	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.2198	0.0000	
# of significant (p<0.05/171)	**12**	**3**	**10**	**15**	**15**	**16**	**16**	**16**	**12**	**10**	**15**	**14**	**11**	**13**	**13**	**14**	**2**	**11**	**16**

Supplementary Table 2 Pairwise rank-correlation of human diversity and human-chimpanzee divergence with the 6 predictors which we investigated but did not contribute significantly to our full model. The divergence and diversity data come from non-repetitive intergenic regions.

	chimp-human divergence	human diversity
	τ	τ
LINE-count	-0.145***	-0.111***
AluY-count	-0.084***	-0.021
expression strength	-0.060**	-0.007
expression breadth	-0.135***	-0.076***
fraction of genes expressed in testes	-0.119***	-0.058*
fraction of genes expressed in ovaries	-0.127***	-0.042
* significant at the 5%, ** 1% and ***< 0.1 % level		

Supplementary Table 3. Regression of diversity on recombination rate for simulated data under three modes of selection: recurrent selective sweeps (i), background selection (ii), and both (iii). We varied the selection coefficients for deleterious (s_D) and advantageous mutations (s_A) as well as the rate of their occurrence (μ_D ; μ_A). Those data were binned according to their recombination rate, where each bin contained 25 data-points for which the mean recombination rate was regressed on the mean diversity. Listed are the resulting slope, and p-value from the regression model.

background		sweep			
s_D	μ_D	s_A	μ_A	**slope**	**p-value**
		data		0.620	1.54E-04
0.002	*2.00E-10*			0.742	1.17E-06
0.02	*2.00E-10*			0.910	5.16E-13
0.002	*2.00E-09*			0.809	2.13E-08
0.02	*2.00E-09*			0.906	1.02E-12
		0.002	*2.00E-11*	0.203	2.66E-01
		0.02	*2.00E-11*	0.444	1.09E-02
		0.002	*2.00E-10*	0.394	2.55E-02
		0.02	*2.00E-10*	0.640	8.07E-05
0.002	*2.00E-10*	*0.002*	*2.00E-11*	0.715	4.36E-06
0.02	*2.00E-10*	*0.002*	*2.00E-11*	0.892	7.06E-12
0.002	*2.00E-09*	*0.002*	*2.00E-11*	0.801	3.77E-08
0.02	*2.00E-09*	*0.002*	*2.00E-11*	0.910	5.46E-13
0.002	*2.00E-10*	*0.02*	*2.00E-11*	0.629	1.16E-04
0.02	*2.00E-10*	*0.02*	*2.00E-11*	0.650	5.71E-05
0.002	*2.00E-09*	*0.02*	*2.00E-11*	0.811	1.84E-08
0.02	*2.00E-09*	*0.02*	*2.00E-11*	0.877	4.85E-11
0.002	*2.00E-10*	*0.002*	*2.00E-10*	0.696	9.84E-06
0.02	*2.00E-10*	*0.002*	*2.00E-10*	0.840	1.81E-09
0.002	*2.00E-09*	*0.002*	*2.00E-10*	0.808	2.32E-08
0.02	*2.00E-09*	*0.002*	*2.00E-10*	0.895	4.61E-12
0.002	*2.00E-10*	*0.02*	*2.00E-10*	0.730	2.15E-06
0.02	*2.00E-10*	*0.02*	*2.00E-10*	0.729	2.27E-06
0.002	*2.00E-09*	*0.02*	*2.00E-10*	0.844	1.27E-09
0.02	*2.00E-09*	*0.02*	*2.00E-10*	0.922	6.94E-14

Supplementary Table 4. Linear regression models for chimpanzee-human divergence, where divergence was measured in intergenic, non-repeat sequences, excluding CpGs. R^2=0.522 for the model with three predictors and for the full model, i.e. all 12 predictors. The slope estimates are for the full model. Also listed are the pair-wise rank-correlation coefficients based on Kendall's τ. Although the pair-wise correlations change slightly, leaving out CpGs to estimate divergence does not change the overall model.

	slope	τ
CpG content	1.581***	-0.035
CpG islands	-0.433***	-0.151***
distance to centromere	-0.093**	0.040
distance to telomeres	-0.154***	-0.290***
GC-content	-2.190***	-0.071**
gene content	-0.099***	-0.238***
poly(CA)	-0.103**	0.152***
polyA/T	-1.150***	0.029
polyG/C	-0.204***	-0.139***
polyR/Y	-0.181***	-0.089***
recombination rate (cM/Mb)	0.189***	0.252
simple repeat content	0.311***	0.325***