Supporting Information

Hvilsom et al. 10.1073/pnas.1106877109

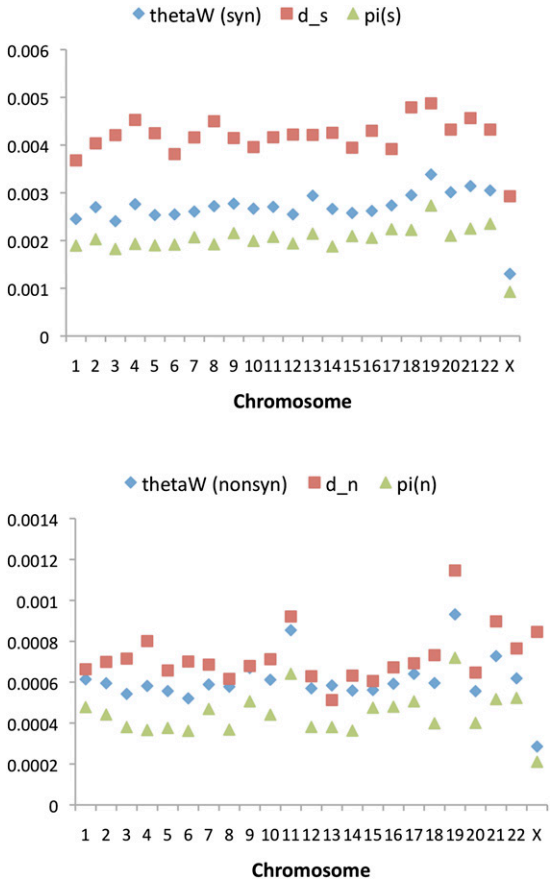

Fig. S1. Synonymous and nonsynonymous diversity and divergence per chromosome. Watterson's θ for synonymous sites was calculated as the number of synonymous SNPs divided by the number of synonymous sites for each chromosome, and then divided by $\sum_{i=1}^{23} \frac{1}{i}$ for autosomes and $\sum_{i=1}^{17} \frac{1}{i}$ for the X chromosome. Synonymous π was calculated as the average synonymous heterozygosity for each chromosome. Synonymous divergence was calculated as the number of fixed synonymous substitutions on the chimp branch divided by the number of synonymous sites. For diversity and divergence the same set of exons was used. Watterson's θ is shown in general higher than π, consistent with a growing population scenario. Watterson θ and π both increase with chromosome number, and both decrease dramatically on the X chromosome. Synonymous divergence shows a similar pattern, but for nonsynonymous divergence, the X chromosome is not reduced compared with the other chromosomes.

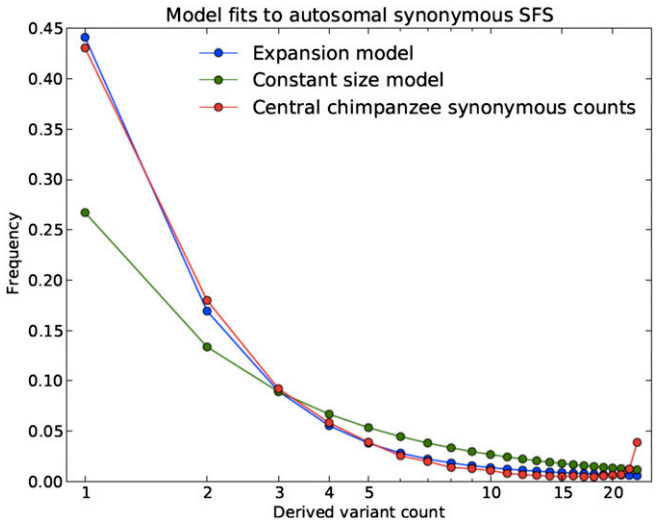

Fig. S2. Expansion model fit to the synonymous site frequency spectrum. The autosomal SFS data for synonymous mutations were fitted by different demographic model using DaDi. The best fitting model is shown, which is an expansion model of an increase in population size by a factor of ~4 about 150 generations ago.

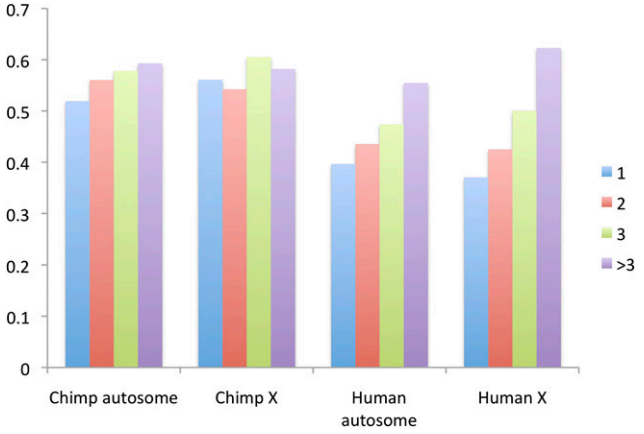

Fig. S3. More nonsynonymous rare SNPs in humans. The ratio of synonymous to nonsynonymous SNPs (*y* axis) for the three lowest frequency classes in the SFS (variant observed in chromosomes 1, 2, and 3) as well as for all higher frequencies pooled. There is an enrichment of rare nonsynonymous mutations for the human autosomes and in particular for the human X chromosome. For chimpanzee, a smaller enrichment is visible for the autosomes but cannot be seen for the X chromosome.

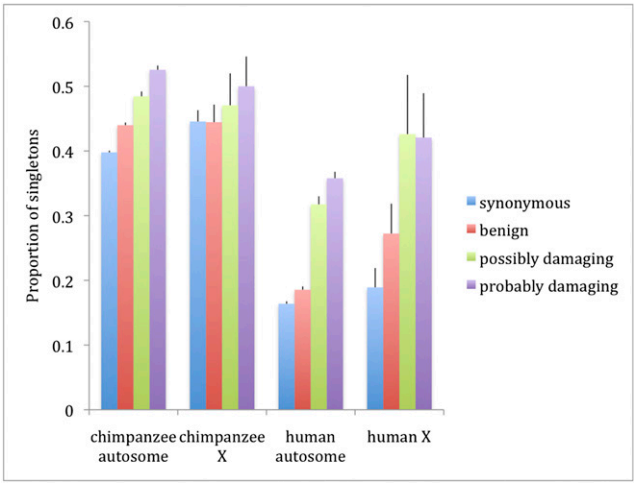

Fig. S4. The proportion of singletons within different functional classes of SNPs for human and chimpanzees divided into autosomes and X chromosome. The classification of SNPs is into synonymous and nonsynonymous SNPs. Within the nonsynonymous SNP, Polyphen 2.0 was used to classify each SNP as being benign, possibly damaging, or probably damaging. SEs are based on the binomial variance.

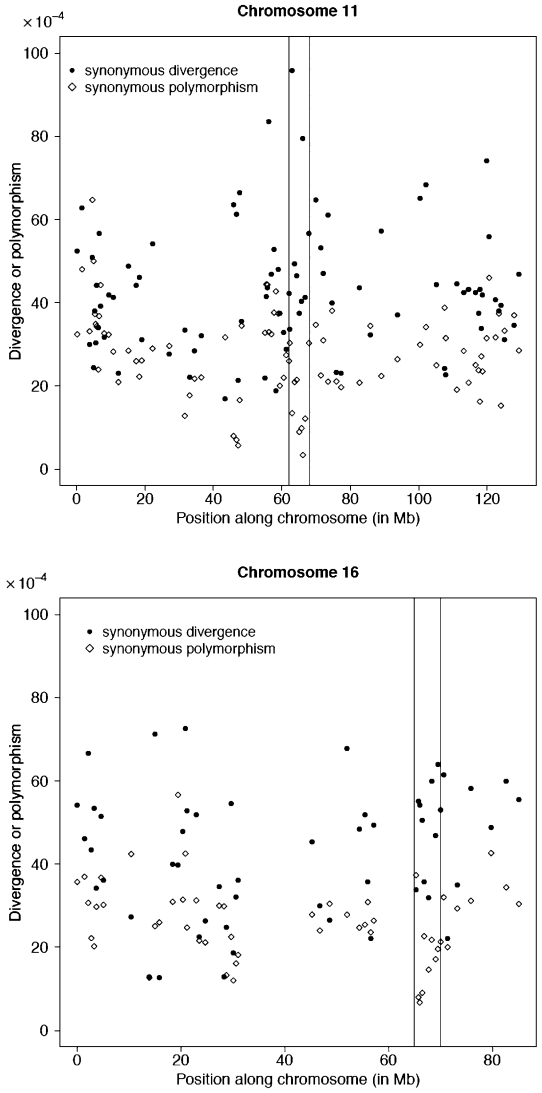

Fig. S5. Further examples of putative sweeps associated with immunity. Second to the chromosome 3 example (Fig. 3), chromosome 11 and chromosome 16 showed the strongest signals of reduced polymorphism marked in this figure. On chromosome 11, the marked region include the genes OTUB1, chr11:63510461–63522463, CLCF1 chr11:66888215–66897782, and PTPRCAP chr11:66959557–66961729. The region on chromosome 16 includes TRADD chr16:65745604–65751306, HSF4 chr16:65756217–65761347, EDC4 chr16:66464500–66475907, PSMB10 chr16:66525908–66528254, NFATC3 chr16:66676876–66818338, and NFAT5 chr16:68156498–68296054.

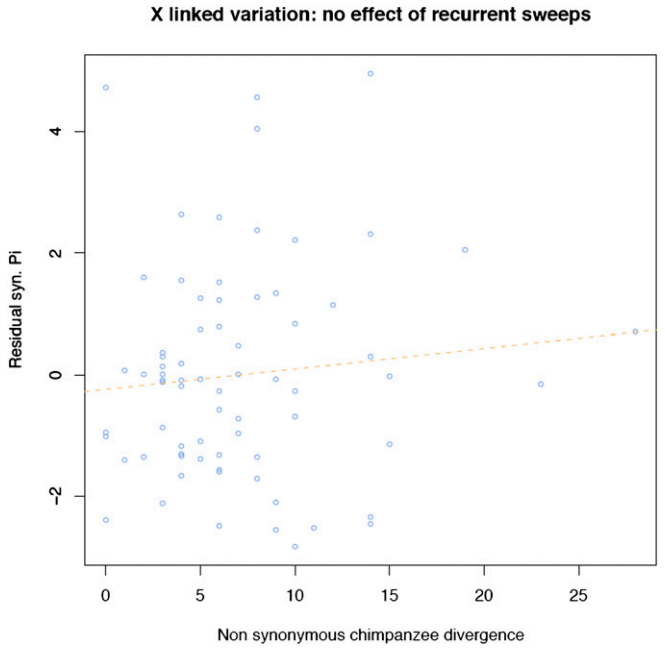

Fig. S6. No apparent effect of recurrent selective sweep on synonymous diversity in X-linked regions. Data on chromosome X are divided in 72 windows (blue open circle) comprising each 10 kb of exonic material. Nonsynonymous divergence is measured through the number of nonsynonymous substitutions affecting specifically the chimpanzee branch in each window. A corrected measure of synonymous nucleotide diversity is obtained after accounting for the level of synonymous divergence among windows (a regression model with synonymous divergence explains about 11% of the total variation in synonymous nucleotide diversity). The orange dotted line denotes the regression of corrected (residual) synonymous diversity against nonsynonymous divergence and is not significantly different from zero (*P* value >0.4).

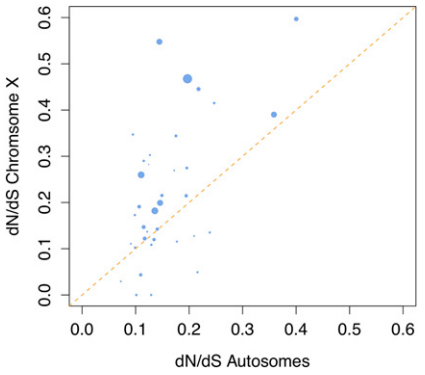

Fig. S7. Rates of evolution (d_N/d_S) are higher on the X chromosome than on autosomes. Rates of evolution (d_N/d_S) in autosomes versus the X chromsome are contrasted for the most abundant Gene Ontology categories of biological processes (n = 37). Taken together these 37 GO categories comprise 58 and 60%, respectively, of all autosomal exons and X-linked exons in Dataset S1. For each GO category, data from all pooled exons were used to estimate d_N/d_S using the Nei–Gojobori method. Each circle denotes the d_N/d_S estimates for a gene ontology category. The size of each circle is proportional to the abundance of each category (on autosomes). The diagonal (orange line) denotes the expectation under null hypothesis that within each GO category the d_N/d_S in autosomal versus X linked genes are identical. Of 37 GO categories considered, 27 exhibit higher d_N/d_S in X-linked genes (paired Wilcoxon rank test, *P* value <0.007). Accordingly, the predicted d_N/d_S, based on the relative abundance of GOs on the X for these 37 categories, is 0.155 (CI 95%: 0.147–0.163), whereas the observed d_N/d_S on the X is 0.211 (95% CI: 0.179–0.242).

Table S1. Origin of samples

ID	Sex	Origin
Aboume	M	Gabon
Amelie	F	Gabon, Haut-Ogooué
Ayrton	M	Gabon, Moanda
Bakoumba	M	Gabon
Benefice	F	Gabon (CNRS, Makokou)
Chiquita	F	Gabon
Cindy	F	Wild-caught Africa*
Lalala	F	Gabon
Makokou	F	Gabon, Ogooué-Ivindo
Masuku	F	Gabon, Haut-Ogooué
Noemie	F	Equatorial Guinea
Susi	F	Wild-caught Africa*

The 12 *Pan troglodytes troglodytes* samples are from wild-born unrelated individuals. F, female; M, male.
*Tested by mtDNA and microsatellites to be of Central African origin.

Table S2. Sequencing statistics for first six samples

Exon capture	ABOUME	AMELIE	AYRTON	BAKOUMBA	BENEFICE	CHIQUITA
Initial bases on target	37,806,033	37,806,033	37,806,033	37,806,033	37,806,033	37,806,033
Initial bases near target*	57,200,015	57,200,015	57,200,015	57,200,015	57,200,015	57,200,015
Initial bases on or near target	95,006,048	95,006,048	95,006,048	95,006,048	95,006,048	95,006,048
Total uniquely mapped reads	20,566,908	20,897,094	20,863,152	21,477,662	23,371,004	22,756,168
Total effective yield, Mb	1851.02	1880.74	1877.68	1932.99	2103.39	2048.06
Average read length, bp	90.00	90.00	90.00	90.00	90.00	90.00
Effective sequence on target, Mb	1329.39	1277.51	1332.36	1329.94	1434.26	1352.83
Effective sequence near target, Mb	219.01	290.06	241.94	280.91	336.41	316.13
Effective sequence on or near target, Mb	1548.40	1567.57	1574.30	1610.85	1770.68	1668.96
Fraction of effective bases on target, %	71.8	67.9	71.0	68.8	68.2	66.1
Fraction of effective bases on or near target, %	83.7	83.3	83.8	83.3	84.2	81.5
Average sequencing depth on target	**35.16**	**33.79**	**35.24**	**35.18**	**37.94**	**35.78**
Average sequencing depth near target	3.83	5.07	4.23	4.91	5.88	5.53
Mismatch rate in target region, %	0.62	0.62	0.60	0.64	0.60	0.60
Mismatch rate in all effective sequence, %	0.74	0.75	0.73	0.76	0.72	0.73
Base covered on target	36,691,325	36,665,142	36,628,348	36,759,294	36,696,323	36,712,506
Coverage of target region, %	**97.1**	**97.0**	**96.9**	**97.2**	**97.1**	**97.1**
Base covered near target	29582614	35146064	32302856	33562545	40692616	39546192
Coverage of flanking region, %	51.7	61.4	56.5	58.7	71.1	69.1
Fraction of target covered with at least 20×, %	59.0	58.5	60.4	58.4	65.0	63.2
Fraction of target covered with at least 10×, %	78.6	78.7	79.2	78.4	81.8	81.2
Fraction of target covered with at least 4×, %	91.0	91.2	91.0	91.2	91.9	91.8
Fraction of flanking region covered with at least 20×, %	4.7	6.6	5.3	6.4	7.7	7.1
Fraction of flanking region covered with at least 10×, %	11.6	16.0	13.2	15.1	18.9	17.8
Fraction of flanking region covered with at least 4×, %	24.9	32.8	28.0	30.9	38.9	37.2

*Flanking region within 200 bp of target regions.

Table S3. Sequencing statistics for last six samples

Exon capture	LALALA	MAKOKOU	MASU.K.U	NOEMIE	SUSI_11043	CINDY_11525
Initial bases on target	37,806,033	37,806,033	37,806,033	37,806,033	37,806,033	37,806,033
Initial bases near target*	57,200,015	57,200,015	57,200,015	57,200,015	57,200,015	57,200,015
Initial bases on or near target	95,006,048	95,006,048	95,006,048	95,006,048	95,006,048	95,006,048
Total uniquely mapped reads	17,275,141	22,450,448	20,576,915	20,036,672	21,436,483	26,448,088
Total effective yield, Mb	1554.76	2020.54	1851.92	1803.30	1929.28	2380.33
Average read length, bp	90.00	90.00	90.00	90.00	90.00	90.00
Effective sequence on target, Mb	1070.54	1389.72	1233.10	1237.95	1312.52	1655.74
Effective sequence near target, Mb	227.31	297.94	298.78	275.42	296.20	357.00
Effective sequence on or near target, Mb	1297.84	1687.66	1531.88	1513.37	1608.72	2012.73
Fraction of effective bases on target, %	68.9	68.8	66.6	68.6	68.0	69.6
Fraction of effective bases on or near target, %	83.5	83.5	82.7	83.9	83.4	84.6
Average sequencing depth on target	**28.32**	**36.76**	**32.62**	**32.74**	**34.72**	**43.80**
Average sequencing depth near target	3.97	5.21	5.22	4.82	5.18	6.24
Mismatch rate in target region, %	0.60	0.59	0.62	0.59	0.60	0.70
Mismatch rate in all effective sequence, %	0.72	0.72	0.75	0.72	0.73	0.81
Base covered on target	36,497,258	36,682,710	36,701,512	36,604,851	36,687,496	36,460,039
Coverage of target region, %	**96.5**	**97.0**	**97.1**	**96.8**	**97.0**	**96.4**
Base covered near target	34,592,251	37,139,854	36,594,614	38,247,253	38,809,387	32,979,857
Coverage of flanking region, %	60.5	64.9	64.0	66.9	67.8	57.7
Fraction of target covered with at least 20×, %	53.5	63.5	57.6	60.0	61.9	67.7
Fraction of target covered with at least 10×, %	75.7	81.1	78.4	79.2	80.6	81.8
Fraction of target covered with at least 4×, %	89.7	91.7	91.2	91.0	91.6	91.0
Fraction of flanking region covered with at least 20×, %	4.2	6.8	6.7	5.7	6.4	9.6
Fraction of flanking region covered with at least 10×, %	12.4	16.6	16.6	15.4	16.5	19.8
Fraction of flanking region covered with at least 4×, %	29.1	34.4	34.6	34.1	35.4	34.7

*Flanking region within 200 bp of target regions.

Table S4. Differences between using the hg18 and pantro2 for mapping of reads and identification of SNPs and the potential consequences of using either for the calculation of π_N/π_S

	Autosome	X chromosome
Exon not present in PanTro2 assembly*	1374	87
Synonymous (S) SNPs not polymorphic against panTro2[†]	1029	23
Nonsynonymous (NS) SNPs not polymorphic against panTro2[†]	1439	21
NS/S SNPs including not polymorphic against panTro2 SNPs[‡]	0.803	0.764
NS/S SNPs excluding not polymorphic against panTro2 SNPs[‡]	0.784	0.759

*Number of SNPs where the corresponding exon is absent from the panTro2 assembly.
[†]Number of SNPs called against hg18, which show no polymorphism when mapped against pantro2.
[‡]Consequences on the nonsynonymous to synonymous ratio of SNPs when keeping or leaving out these SNPs.

Table S5. Exons and base pairs on the chip and included in the analysis

Chr	bp on chip	bp included in analysis	Exons on chip	Exons in analysis
Chr1	3,505,063	1,766,139	17,924	9,183
Chr2	2,504,138	1,102,350	11,567	5,820
Chr3	1,957,127	949,629	9,678	4,982
Chr4	1,372,542	662,912	6,300	3,181
Chr5	1,624,295	772,957	7,366	3,734
Chr6	1,750,838	803,917	8,486	3,904
Chr7	1,697,055	686,310	7,447	3,649
Chr8	1,218,967	475,993	5,218	2,451
Chr9	1,473,160	664,479	6,901	3,295
Chr10	1,431,287	693,089	7,229	3,706
Chr11	2,063,538	911,974	8,972	4,392
Chr12	1,776,009	836,062	9,132	4,721
Chr13	631,682	316,146	3,094	1,471
Chr14	1,090,664	490,301	5,032	2,458
Chr15	1,268,872	537,462	5,782	2,864
Chr16	1,510,643	578,315	7,153	3,287
Chr17	2,029,048	857,305	9,838	4,790
Chr18	544,231	267,124	2,472	1,320
Chr19	2,290,530	791,001	9,113	3,344
Chr20	870,891	364,430	4,380	2,094
Chr21	376,560	157,734	1,783	798
Chr22	793,609	263,734	3,566	1,520
ChrX	1,326,500	775,235	6,292	3,615
Autosome sum	35,107,249	15,724,598	164,852	8,0621

Table S6. Number of nonsynonymous SNPs per individual

	ABOUME	AYRTON	BAKOUMBA	AMELIE	BENEFICE	CHIQUITA	LALALA	MAKOKOU	MASU.K.U	NOEMIE	CINDY_11525	SUSI_11043	Nonsynonymous length
Chr1	486	668	732	656	677	655	617	691	697	668	603	623	1,355,702.92
Chr2	381	358	353	401	376	367	334	395	403	359	345	379	839,811.61
Chr3	296	270	308	256	288	266	251	300	260	257	288	284	728,525.9
Chr4	200	193	179	124	191	194	169	206	168	202	162	217	501,899.31
Chr5	233	216	213	215	205	241	209	240	247	250	178	224	591,837.88
Chr6	240	223	244	194	242	242	234	224	245	213	186	207	620,628.89
Chr7	239	252	238	224	244	267	226	255	252	234	274	224	520,704.03
Chr8	145	125	142	142	147	151	121	131	142	120	132	130	368,341.04
Chr9	253	246	254	256	277	262	254	249	226	258	255	248	500,575.51
Chr10	230	218	244	230	248	244	232	229	229	206	235	210	520,981.22
Chr11	471	374	393	494	443	447	448	490	472	443	455	442	698,178.96
Chr12	252	265	220	201	198	257	245	250	262	253	263	250	637,426.84
Chr13	89	87	88	83	96	81	82	107	116	87	93	77	237,950.36
Chr14	139	138	154	134	143	127	143	138	132	142	137	128	379,736.6
Chr15	196	227	172	199	223	192	190	186	214	180	172	204	412,794.94
Chr16	208	206	209	216	204	225	181	215	241	215	197	199	437,246.59
Chr17	335	315	369	326	328	337	300	310	359	317	324	309	647,024.96
Chr18	75	60	67	96	77	87	79	86	86	93	90	99	207,649.42
Chr19	391	436	425	407	461	395	407	414	416	402	381	428	574,978.99
Chr20	115	119	121	111	129	112	102	95	111	108	108	123	281,241.27
Chr21	76	62	61	72	63	65	65	60	41	60	78	51	121,469.96
Chr22	101	98	88	112	110	114	87	117	114	94	98	97	196,078.15
ChrX	0	0	0	83	86	86	76	72	93	73	85	95	394,956.52
Autosomal average	5,151	5,156	5,274	5,149	5,370	5,328	4,976	5,388	5,433	5,161	5,054	5,153	11,380,785.35

Table S7. Number of synonymous SNPs per individual

	ABOUME	AYRTON	BAKOUMBA	AMELIE	BENEFICE	CHIQUITA	LALALA	MAKOKOU	MASU.K.U	NOEMIE	CINDY_11525	SUSI_11043	Synonymous length
Chr1	660	759	784	774	771	806	701	784	780	746	665	755	395,421.08
Chr2	508	438	439	501	506	497	489	502	512	455	502	446	238,286.39
Chr3	399	359	389	373	406	386	374	382	359	391	383	380	209,358.1
Chr4	269	292	278	209	284	287	265	288	249	269	269	278	139,815.69
Chr5	333	320	338	340	287	334	306	343	346	342	263	332	170,423.12
Chr6	332	316	349	328	353	373	354	361	362	306	298	337	177,131.11
Chr7	314	317	306	290	319	328	300	321	345	296	309	293	150,446.97
Chr8	178	218	211	200	206	197	200	206	184	202	192	214	104,437.96
Chr9	311	321	329	322	309	324	311	303	299	333	287	297	144,877.49
Chr10	269	295	321	313	296	306	284	327	284	297	268	290	148,537.78
Chr11	452	382	438	458	412	396	437	453	439	414	419	402	204,590.04
Chr12	393	390	327	322	300	360	336	365	393	366	365	378	184,564.16
Chr13	138	126	162	151	135	151	149	158	157	122	153	132	67,398.64
Chr14	215	189	201	224	209	232	206	200	202	209	166	199	108,993.4
Chr15	242	243	253	278	272	245	227	260	283	243	206	248	119,387.06
Chr16	273	263	278	281	251	279	260	294	283	274	232	258	130,710.41
Chr17	439	417	437	418	457	456	417	382	448	387	437	435	191,001.04
Chr18	132	120	127	115	142	128	145	129	125	138	150	125	59,074.58
Chr19	467	506	456	445	458	444	433	460	481	457	388	463	166,396.01
Chr20	176	166	204	162	164	193	153	180	190	157	176	165	82,775.73
Chr21	90	91	86	65	81	84	93	72	48	83	83	70	35,043.04
Chr22	128	134	117	144	130	146	122	135	152	145	140	165	58,744.85
ChrX	0	0	0	89	118	114	78	105	104	106	121	107	113,291.48
Autosomal average	6,718	6,662	6,830	6,713	6,748	6,952	6,562	6,905	6,921	6,632	6,351	6,662	3,287,414.65

Table S8. Number of synonymous (π_S) and nonsynonymous (π_N) SNPs in central chimps and the number of synonymous (d_S) and nonsynonymous (d_N) changes on the chimpanzee branch, estimated from human–chimp divergence using orangutan as an outgroup for assigning events to branches

Chromosome	d_N	d_S	d_N/d_S	π_N	π_S	π_N/π_S	NI	Nonsyn sites	Syn sites
1	899	1,455	0.618	3,108	3,620	0.859	1.390	1,355,702.92	395,421.08
2	587	962	0.610	1,865	2,402	0.776	1.272	839,811.61	238,286.39
3	521	881	0.591	1,476	1,881	0.785	1.327	728,525.9	209,358.1
4	402	633	0.635	1,090	1,442	0.756	1.190	501,899.31	139,815.69
5	389	724	0.537	1,230	1,613	0.763	1.419	591,837.88	170,423.12
6	435	675	0.644	1,207	1,685	0.716	1.112	620,628.89	177,131.11
7	357	626	0.570	1,145	1,465	0.782	1.370	520,704.03	150,446.97
8	227	470	0.483	795	1,061	0.749	1.551	368,341.04	104,437.96
9	340	601	0.566	1,251	1,500	0.834	1.474	500,575.51	144,877.49
10	371	588	0.631	1,190	1,480	0.804	1.274	520,981.22	148,537.78
11	643	852	0.755	2,227	2,067	1.077	1.428	698,178.96	204,590.04
12	401	779	0.515	1,356	1,757	0.772	1.499	637,426.84	184,564.16
13	122	284	0.430	519	740	0.701	1.633	237,950.36	67,398.64
14	240	464	0.517	793	1,084	0.732	1.414	379,736.6	108,993.4
15	250	471	0.531	866	1,149	0.754	1.420	412,794.94	119,387.06
16	294	562	0.523	967	1,279	0.756	1.445	437,246.59	130,710.41
17	448	748	0.599	1548	1,952	0.793	1.324	647,024.96	191,001.04
18	152	283	0.537	462	651	0.710	1.321	207,649.42	59,074.58
19	659	811	0.813	2,000	2,103	0.951	1.170	574,978.99	166,396.01
20	182	358	0.508	584	931	0.627	1.234	281,241.27	82,775.73
21	109	160	0.681	330	411	0.803	1.179	121,469.96	35,043.04
22	150	254	0.591	453	669	0.677	1.147	196,078.15	58,744.85
X	508	505	1.006	617	808	0.764	0.759	600,624.31	172,475.69

The total number of syn and nonsyn sites are also shown.

Other Supporting Information Files

Dataset S1 (RTF)