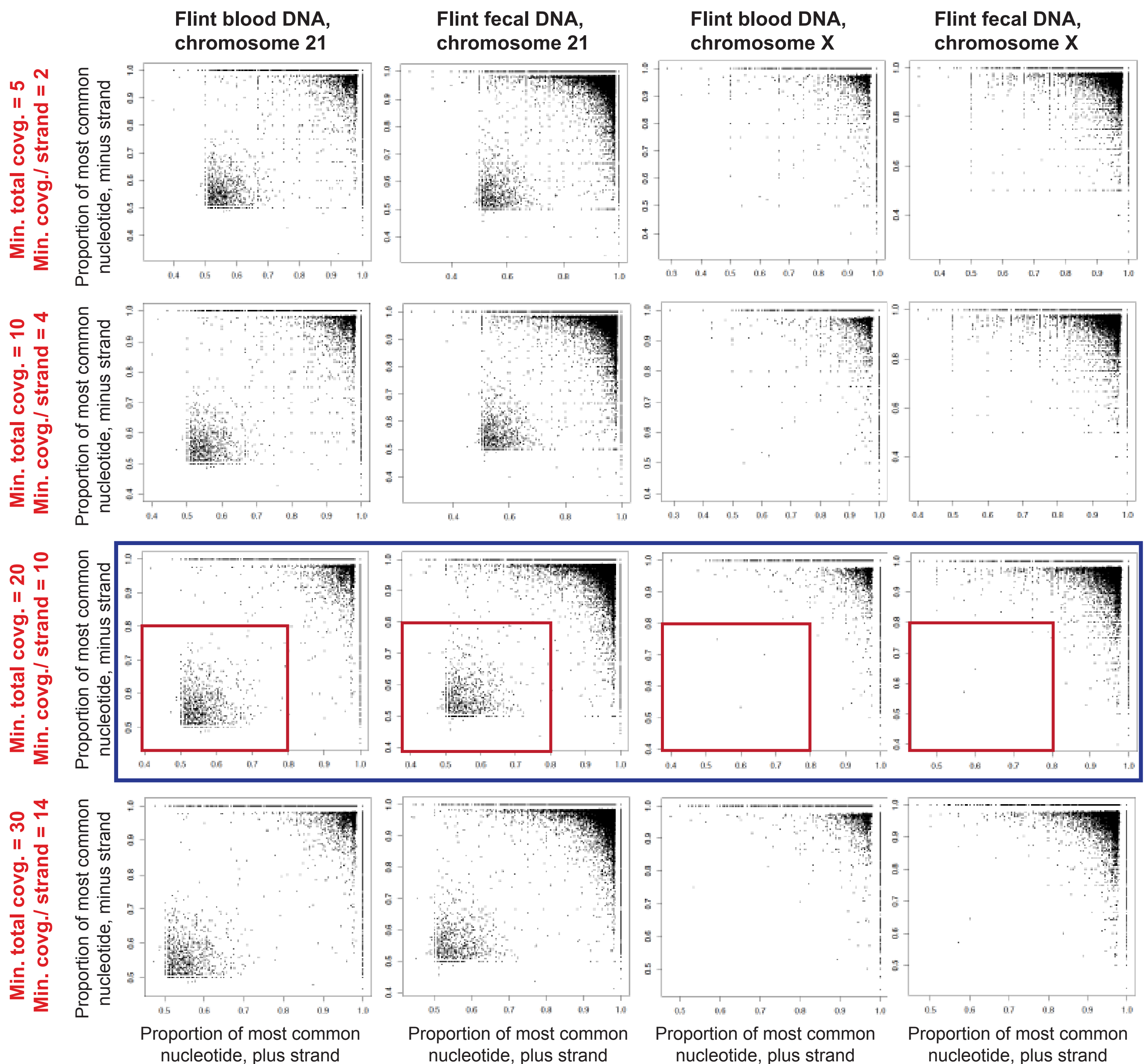

Fig S1. Heterozygous site single nucleotide polymorphism identification criteria.
After sub-sampling the reads to include only one read mapped to each start position per strand to avoid post-adapter ligation and post-capture amplification biases, we evaluated the SNP calling properties of different read coverage criteria, using iterations of total filtered read coverage per base, strand-specific filtered read coverage per base, and proportion of the most common nucleotide per site on each strand. Shown are plots of the most common nucleotide proportion on the forward strand (X axis) and reverse strand (Y axis) at each site for a selected sample of the iterated coverage cutoffs (the overwhelming majority of sites have most common nucleotide proportions of 1.0 on each strand), on chromosome 21 and X from the blood and fecal DNA sequence data from Flint, 93A009, a male chimpanzee. We sought to maximize the accuracy in SNP calling while minimizing the numbers of sites that could not be analyzed due to insufficient coverage. For our final heterozygous site SNP identification critera (enclosed by blue box), we considered sites covered by a minimum of 10 reads from each strand (total minimum 20), and identified heterozygous SNPs as the sites where the most common nucleotide proportion was ≤ 0.8 on both strands (red box).

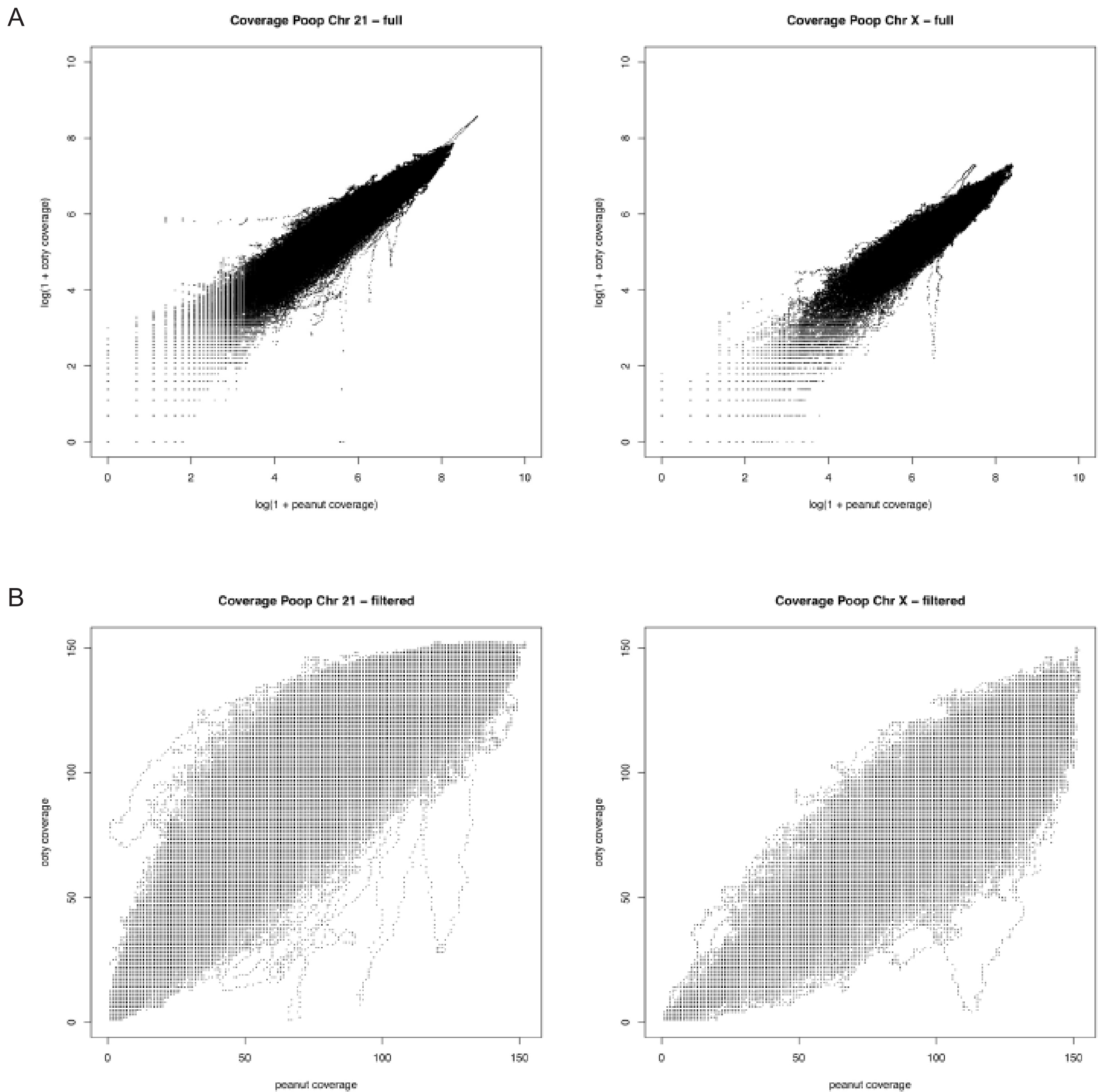

Fig S2. Sequence coverage correspondence.
Two sample per site sequence read coverage comparisons, for Peanut fecal DNA (*x*-axis) and Coty fecal DNA (*y*-axis), which exemplify the results observed across all samples. (A) Total coverage, without filtering. (B) Coverage after sub-sampling one read for each start position and strand (maximum filtered coverage = 152/bp; 76 bp reads x 2 strands) to avoid potential post-ligation and post-capture amplification biases. A small proportion of sites have low filtered coverage, but these positions tend to be the same across samples. Therefore, sites with insufficient coverage for SNP identification are often in common across samples rather than random.

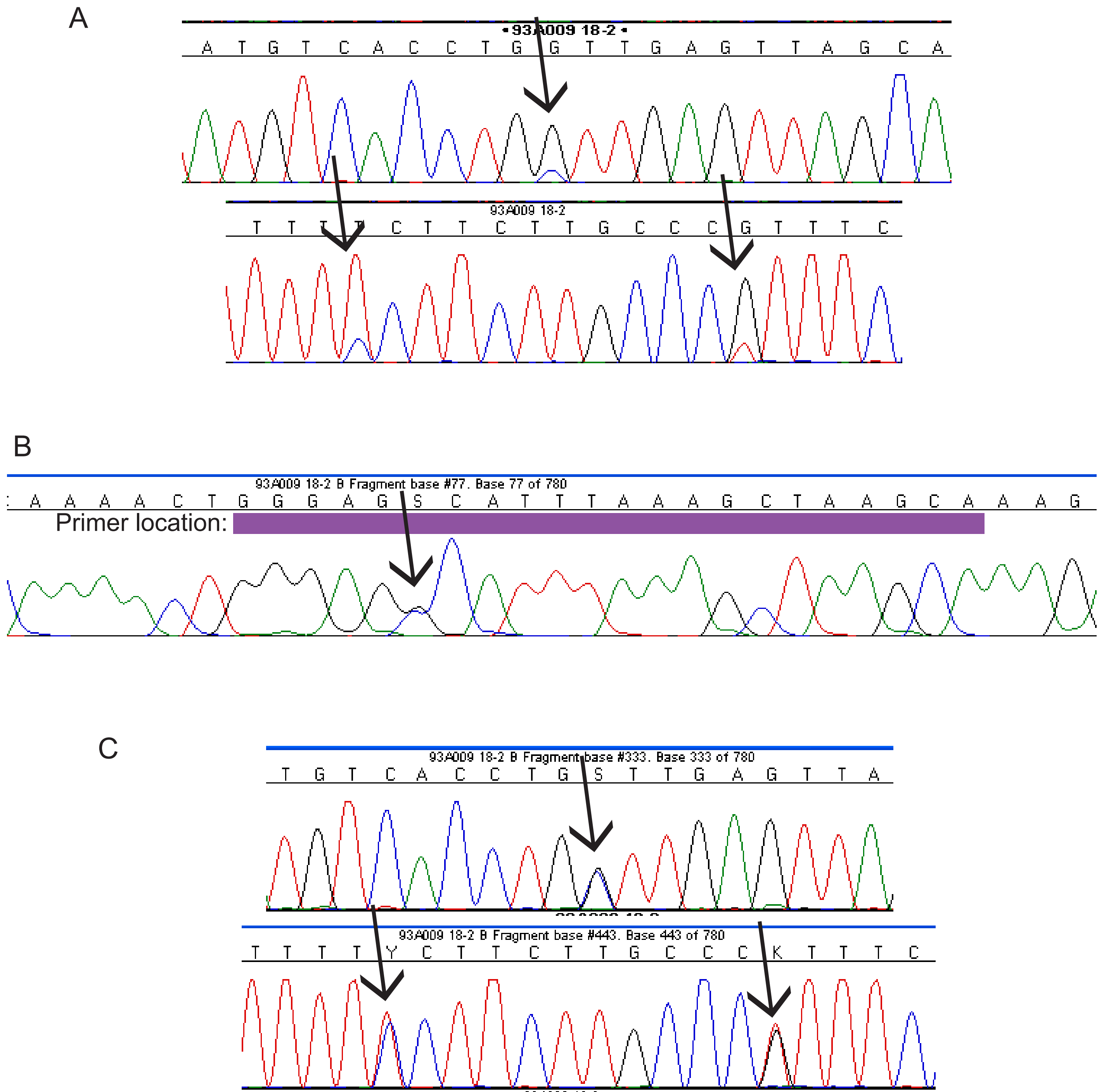

Fig S5. Validation of identified single nucleotide polymorphisms by PCR and Sanger sequencing. We used PCR and Sanger sequencing to analyze 20 of the regions we targeted with the SureSelect capture library and sequenced on the Illumina Genome Anaylzer IIx. (A) In one of the regions, we had identified heterozygous sites from analysis of the GAIIx data that were not confirmed by the Sanger sequencing data, indicated by arrows on chromatograms. However, the GAIIx data also indicated a SNP at the same genome position as the forward primer, suggesting the possibility of allele-specific PCR. We designed a different forward primer, upstream of the original forward primer. (B) Chromatogram from sequencing reaction using the second amplified product, with the original primer sequence indicated, and the heterozygous site suggested by the GAIIx data (arrow) confirmed. (C) Chromatograms from sequencing reactions that used the second amplified product, confirming the GAIIx-predicted heterozygous sites.

Table S1. Sequencing statistics summary.

Individual	Source	Prop. Endog.[1]	Total reads	Chromosome 21				Chromosome X				Mitochondrial genome	
				Reads mapped to targets[2]	Enrich.[3]	Effective enrich.[4]	Mean(±sd) filtered coverage[5]	Reads mapped to targets[2]	Enrich.[3]	Effective enrich.[4]	Mean(±sd) filtered coverage[5]	Reads mapped to targets[2]	Mean(±sd) filtered coverage[5]
93A009 Flint (male)	Blood	NA	31.3M	9,563,180	3,454	(3,454)	85 (15.9)	2,604,801	2,753	(2,753)	55 (12.5)	6,699,099	147 (20.7)
	Fecal	0.009	27.5M	6,596,350	2,238,909	20,150	88 (23.7)	2,052,414	401,293	3,612	60 (17.1)	15,319,420	146 (22.6)
91A010 Peanut (female)	Blood	NA	30.1M	9,430,361	5,198	(5,198)	113 (20.4)	5,131,528	4,307	(4,307)	115 (23.1)	7,117,730	146 (21.4)
	Fecal	0.052	31.8M	9,027,939	680,940	35,409	93 (33.2)	7,006,967	434,395	22,589	110 (28.7)	14,190,809	148 (17.9)
A2A009 Sopulu (male)	Blood	NA	26.9M	9,717,238	5,783	(5,783)	110 (17.1)	2,573,133	4,356	(4,356)	76 (15.7)	7,581,089	149 (16.5)
	Fecal	0.006	31.7M	8,195,055	1,390,768	8,345	67 (15.8)	2,698,919	395,773	2,375	42 (10.4)	19,366,060	149 (16.3)
X161 Judd (male)	Blood	NA	26.6M	9,714,466	4,784	(4,784)	103 (15.3)	2,514,313	3,483	(3,483)	67 (13.7)	6,433,830	149 (16.3)
	Fecal	0.015	31.9M	11,288,033	1,915,672	28,735	114 (26.3)	3,832,523	347,908	5,219	90 (20.9)	15,567,341	149 (15.2)
91A016 Coty (male)	Blood	NA	30.8M	11,274,746	5,044	(5,044)	110 (16.7)	3,047,723	4,149	(4,149)	76 (15.2)	7,415,030	146 (22.8)
	Fecal	0.021	30.8M	7,516,434	1,133,867	23,811	113 (31.3)	2,510,573	368,809	7,745	95 (26.4)	19,894,407	148 (21.2)
A1A005 Kierra (female)	Blood	NA	27.4M	8,054,414	6,048	(6,048)	103 (16.7)	4,272,446	4,473	(4,473)	103 (19.4)	8,797,339	147 (21.7)
	Fecal	0.005	28.3M	3,978,701	1,353,830	6,769	54 (13.7)	2,454,901	479,987	2,400	58 (13.3)	20,923,666	147 (23.2)

Notes:

[1] Estimated proportion of endogenous DNA out of total in fecal DNA extract, based on quantitative PCR, using primers unique in the chimpanzee nuclear genome (forward 5'-3' CAATCAAGACGTCCAGCTCA and reverse 5'-3' TAGAACTGCTGCCCCACTTT), evaluated against a standard curve constructed from the blood DNA of one individual (Flint, 93A009). The samples were run in 25 µL reactions using iQ SYBR Green Supermix (Bio-Rad) on a Bio-Rad iCycler Thermal Cycler with an initial denaturation of 95ºC for 7 min, followed by 40 cycles of 95ºC for 30 sec and 60ºC for 45 sec. Test samples were run in triplicate and standards run in duplicate.

[2] Number of reads mapped uniquely to targeted regions on chimpanzee chromosome 21 (total size of targeted regions = 1,052,310 bp), chromosome X (total size of targeted regions = 550,471 bp), and the mitochondrial genome (total size of targeted region = 16,554 bp). The mapped read data for the mitochondrial genome are provided for informational purposes but these results were not used for analyses. Instead, due to the extreme variability in the hypervariable region, we performed *de novo* assembly of the complete mitochondrial genome sequence for each individual (see *Materials and Methods*).

[3] Enrichment of targeted regions in the sequence data. The number of reads mapped to targeted regions, per base, compared to the number of reads mapped to regions of the same chromosomes that were not targeted but that met the same filtering criteria for region selection (i.e., regions that were not targeted only because we were limited to 55,000 SureSelect baits), per base. Total size of chromosome 21 non-targeted regions = 1,428,685 bp. Total size of chromosome X non-targeted regions = 4,197,550 bp.

[4] Effective enrichment for the fecal DNA samples, accounting for the fact that endogenous DNA is overwhelmed by DNA from exogenous sources in these samples. Calculated as the estimated proportion of endogenous DNA times the enrichment of targeted regions in the sequence data.

[5] To avoid the possibility of analyzing multiple reads representing copies of the same original DNA fragment from amplification following adapter ligation and capture, we sub-sampled one read mapped to each position on each strand. After filtering, maximum coverage is 152x (76 bp x 2 directions).

Table S2. PCR amplifcation and sequencing primers.

Targeted Region	Forward primer	Reverse primer	Sequencing primer	Sequencing primer	Sequencing primer	Comments
chr21:14581134-14583181	TGGGTGTGATCAGAGAAATTCG	TTCAAATGCTGGAGGACAACAT	GTATTCATCAAAAAATTACT	AGATCAGAATATACCGTTAT	CTTAATGTGGAGTATACGTC	
chr21:14954854-14957395	CCCAAGTTGCACAGGAATGTAA	TTTGATGCTTTGCACTGTTTCC	TGAGCCTAAAAGTTGCCAGG	ATACATTTCCAATCTCCTGA	CTTTCCATTGCAGCGGTGCT	
chr21:15616780-15620018	GTTTCCCAAAGGTATGGTGGAG	GCAGCGAGCAGTTAAAGTGAAA	TGTAGAGATTAAATATGACC	GATAAATTGCTTTGAAATCC	TTTCTTGGCTTCTGAACTTT	
chr21:15808718-15810929	CAGCAGCTGTAACTCCCATTTG	TGCAAAATCAAGCCAGGAAATA	CAGCCTCATGGCCAACCTAT	TCACAAAAAGGGATTCATCC	AAACCCCAGCCGAAGAGATG	
chr21:16394928-16397206	CTCAGTGCCCTGAATTGTAGCC	ATGGGCCAGTTTACACTTTTCG	ACAACCATATGCAACATCCT	GCATAGCTGAGGTTTGTCTA	TTGGAAGCCTGTAAAGGGCA	
chr21:17372538-17376963	GACACACAACTTTAGGGGCACA	AGGACACAGCGGTAGGAAAGAG	AAAATTCAAACCACAATTCC	TGGTATTCTCACAATTTTTC	TTACCAAAATATAGGAATTC	
chr21:17379485-17381556	AAGATGGCCATTGATTACAATGAT	TGGAAGTCCCCTGTTCAGAAAT	TGATAGCATCAGACTGATAT	GAAGGTTAACCTAATAGGAG	TTCTAAAAGTAGAAATCATA	
chr21:18358191-18360757	TGGACCAAATTTGCACGATAAG	TTTGTTCACCTCTGCTCTCTGC	TGCATTTATTTTTTAATATA	AAACAGAATAGCCAGACCTT	CTGTTAATTTGTTCAGATTT	
chr21:18448806-18451503	GCAAAGACCCTCAGGAAGAGAA	CGATTCTCGCTTCCTGTCATAA	TCTCCATAATTTCCTTCCTT	TGTAGGAATCCAAACATTTT	AATATAGTGGACAGGACACT	
chr21:18838327-18840462	GACCATGAAAGTTTGGGCACTT	AAGCCCATTATTCCACTCCTGA	TACACATATATACACATAAG	AAAAATAAAATCTTCTGGAG	TTAATCTCTACAAACATAGA	
chr21:19639278-19642238	AAAGTGTAAGCCCAGGTTTGGA	GATGCACCCTGATGAGTGAGAG	CTTTAAGTCACTAAACATCT	CTATTAAATAAAATACATAT	TTCATTGCAGAGCACATGGT	
chr21:20148439-20150927	GAGGGAAAGGTGCATTGTGAGT	TGCTGGACTTACTCTTGCTTCC	GTCTTGCTGAGATGTAAGAT	GACAGTTTAAGATTTGACAA	AAATATTGCTATTTTAAAGT	rev primer also used for seq
chr21:20274315-20276408	GCTCTTCCAACTTTTGGTTGCT	CAACGAACAGCATTTCCATAGG	TGAAAAGAAATTCCATTCTT	ACATTGGCAATAAGCACATC	TTTTTGCTATATTTTAAGTA	
chr21:23681637-23683647	GGCTGGGCAGCTTCATTATTTA	GCTCCTGATTTACCACAAATGG	TACTCACCAAGAAATGACAC	TGCAAATGTAAATACCTGTG	CATTCTGAGGAAAATCAGCA	
chr21:25129118-25131720	CACAGTGATTGGTTTCCTGCTC	GAAACCACCATTTTCTGGGAAG	TATCACCTTCGAAATAAGAC	TAAACAATATACTACATAGT	ATCACAGATCATTTTCCCAG	
chr21:26735598-26740725	TCAGCCACTATTTTTGGAAGCA	TGAGCCGTTTGGGACCTATACT	TGAAGGAGTTTGTAAGGTAA	AGGATAGAGCCAGGAGCTGA	GAGGGAAATTCAGCACTTGA	rev primer also used for seq
chr21:28246952-28250384	AGCAGGCCACAAACTGGTTATT	TGGACATTCCTAAGGAGGGAAA	ATAACTATATACATAGTTAT	TAAGTCTCTCATGCATCTTT	AATTCAACCAACTAAATTCT	
chr21:29277494-29281233	TGTTGGTTGCTTTGACCTCTGT	CTCACCTTTCTCACATGCCATC	ACTCTTCTCTCCTTGCTCAA	ATAATACCAGCATTAGACAT	TACTTCCTGTCTCCATCCTG	for & rev primers also used for seq
chr21:29376239-29378749	GCAAAGCTGAGAAACAATCAGG	TCCCCAGAATTACAAGGCAACT	CAAAGAAGTACTATATAGCT	CAGTCAAGAATAATTCCTGG	CCCCTGCTTCAACATACAGT	region highlighed in Fig S3
chr21:29407059-29409188	AGCTCATCACTTCCCTCTTTGG	CAAAGCCAGTGGGAGAAGAAAC	TTCTAATTAAGGGACTAATC	TTCTATGGTTACAATACAAA	GCAAACCACTAAGGTGGTAA	

Notes:

All primers reported in 5' to 3' sequence

Amplification primers are located within the targeted regions.

PCR reactions: 25 ul with 25 ng blood DNA, 2 mM MgCl2, 0.2 uM dNTPs, 0.4 uM each primer, 1 unit Platinum Taq (Invitrogen), 1X Platinum Taq Buffer.

Conditions: 95C for 2 min followed by 40 cycles of 95C for 30 sec, 62C for 30 sec, and 72C for 2 min, in a Bio-Rad DNA Engine Peltier Thermal Cycler.

Table S3. Sample-level heterozygosity, for sites with coverage sufficient for SNP identification from both fecal and blood DNA results for each individual.

Individual	Source	Chromosome 21			Chromosome X		
		Sites (bp)[1]	Hets.[2]	π, %[3]	Sites (bp)[1]	Hets.[2]	π, %[3]
93A009 Flint (male)	Blood	921,130	788	0.086%	427,442	0	0.000%
	Fecal		775	0.084%		5	0.001%
A2A009 Sopulu (male)	Blood	929,796	919	0.099%	438,456	1	0.000%
	Fecal		914	0.098%		2	0.000%
X161 Judd (male)	Blood	946,724	779	0.082%	453,236	2	0.000%
	Fecal		776	0.082%		2	0.000%
91A016 Coty (male)	Blood	935,620	806	0.086%	450,731	3	0.001%
	Fecal		784	0.084%		1	0.000%
91A010 Peanut (female)	Blood	898,166	717	0.080%	457,102	172	0.038%
	Fecal		707	0.079%		161	0.035%
A1A005 Kierra (female)	Blood	901,558	893	0.099%	446,120	153	0.034%
	Fecal		885	0.098%		147	0.033%

[1] Number of sites with filtered read coverage sufficient for SNP identification from each of the fecal DNA and blood DNA sequencing data from each individual.

[2] Number of heterozygous sites in each sample identified using the criteria described in *Materials and Methods* (the X chromosome heterozygous sites in males are false positives).

[3] Pairwise nucleotide diversity, the percentage of heterozygous sites (π for the X chromosome in males is an approximation of the false-positive error rate).