

Table S1. Sequence representation on the X chromosome by species.

| X chromosome | | | | |
|-----------------------------|--------|--------------|----------------|----------------------|
| Species | Sex | Gapped bases | Ungapped bases | Total scaffold count |
| <i>Pan troglodytes</i> | Male | 156Mb | 135Mb | 1,310 |
| <i>Macaca mulatta</i> | Female | 148Mb | 144Mb | 238 |
| <i>Chlorocebus sabaesus</i> | Male | 130Mb | 127Mb | 123 |

Table S2. Assembly metrics for sequenced primate genomes

| Common Name | Assembled version | N50 contig (kb) | N50 scaffold (Mb) | Assembled Bases | Gene Count | Scaffold Gaps (Mb) |
|--------------------|-------------------|-----------------|-------------------|-----------------|------------|--------------------|
| Chimpanzee | Pan_tro 2.1.4 | 50 | 9 | 3.3 | 18,759 | 407 |
| Gorilla | gorGor3.1 | 12 | .9 | 3 | 20,962 | 206 |
| Orangutan | P_pygmaeus2.0.2 | 16 | .7 | 3.4 | 20,424 | 347 |
| Gibbon | Nleu1.0 | 35 | 52 | 2.9 | 18,575 | 205 |
| Rhesus macaque | MacaMv7 | 54 | 3 | 3.2 | 21,905 | 117 |
| Cynomolgus macaque | MFas 5.0 | 86 | 88 | 2.9 | NA | 142 |
| Vervet | ChlSab1.0 | 90 | 81 | 2.8 | 21,128 | 27 |
| Marmoset | C_jacchus3.2.1 | 29 | 5 | 2.9 | 20,993 | 162 |
| Bushbaby | OtoGar3.0 | 27 | 13 | 2.5 | 19,506 | 160 |
| Baboon | PapAnu2.0 | 40 | .5 | 2.9 | 19,210 | 55 |
| Squirrel monkey | saiBol1 | 38 | 18 | 2.6 | NA | 131 |
| Snub-nose monkey | Rrox V1 | 77 | 1.5 | 2.9 | NA | 43 |
| Tarsier | TSyr2.0.1 | 38 | .4 | 3.4 | 20,820 | 48 |
| Bonobo | Pan_pan | 66 | 10 | 2.9 | 20,826 | 143 |

Table S3. Total interspersed repeats in vervet. Unique vervet sequences, defined by comparison of vervet and rhesus presence absence patterns.

| <i>Chlorocebus aethiops</i> (ChlSab1.1) 2,751,988,575 bp | | no. of families | no. of copies | no. of full-length | total length (kb) | % of genome | vervet (vs. rhesus macaque) | |
|---|-----------|-----------------|---------------|--------------------|----------------------|----------------|-----------------------------|-----------------|
| | | | x1000 | | | | lineage-specific | active families |
| LINEs | | 183 | 986.66 | 1,567 | 568952.6 | 20.67 | 2309 | 8 |
| LINE1 | L1 RS | 9 | 40.14 | 0.427 | 35693.5 | 1.30 | 2136 | 6 |
| | L1P | 34 | 98.95 | 0.888 | 122214.5 | 4.44 | 173 | 2 |
| | L1M | 94 | 416.30 | 0.048 | 295998.8 | 10.76 | | |
| LINE2 | | 11 | 363.88 | 0.095 | 100294.3 | 3.64 | | |
| LINE3/CR1 | | 26 | 47.94 | 0.107 | 10465.3 | 0.38 | | |
| RTE | | 6 | 17.71 | 0.000 | 4073.8 | 0.15 | | |
| Other | | 3 | 1.74 | 0.002 | 212.5 | 0.01 | | |
| SINEs | | 55 | 1687.56 | 793.889 | 382519.7 | 13.90 | 10612 | 14 |
| Alu | AluYR | 13 | 80.12 | 46.976 | 19220.9 | 0.70 | 9170 | 11 |
| | AluY | 5 | 108.25 | 83.643 | 30643.9 | 1.11 | 1442 | 3 |
| | AluS | 19 | 630.68 | 457.251 | 176695.3 | 6.42 | | |
| | AluJ | 4 | 269.54 | 128.745 | 66896.2 | 2.43 | | |
| | monomers | 4 | 58.18 | 36.677 | 7513.1 | 0.27 | | |
| SVA | | 1 | 0.14 | 0.016 | 47.6 | 0.00 | | |
| MIR | MIR | 3 | 452.29 | 36.536 | 71024.1 | 2.58 | | |
| | MIR3 | 1 | 77.95 | 3.910 | 9206.0 | 0.33 | | |
| others | | 5 | 10.42 | 0.135 | 1272.5 | 0.05 | | |
| LTR-retrotransposons | | 542 | 499.63 | 0.302 | 247507.34 | 8.99 | 140 | 3 |
| ERV class I | | 284 | 109.27 | 0.004 | 75202.9 | 2.73 | | |
| ERV(K) class II | | 33 | 8.46 | 0.000 | 8954.5 | 0.33 | 140 | 3 |
| ERV(L) class III | | 138 | 109.13 | 0.009 | 52896.9 | 1.92 | | |
| ERV(L)-MaLR | | 45 | 246.70 | 0.289 | 104010.4 | 3.78 | | |
| Gypsy | | 26 | 15.70 | 0.000 | 3897.2 | 0.14 | | |
| others | | 16 | 10.38 | 0.000 | 2545.4 | 0.09 | | |
| DNA-transposons | | 327 | 415.30 | 104.893 | 99844.0 | 3.63 | | |
| MuDR | | 5 | 1.49 | 0.023 | 653.1 | 0.02 | | |
| PiggyBac | | 6 | 2.01 | 1.128 | 513.8 | 0.02 | | |
| hAT | Ac | 3 | 4.89 | 0.008 | 739.2 | 0.03 | | |
| | Blackjack | 8 | 15.68 | 6.188 | 3108.3 | 0.11 | | |
| | Charlie | 82 | 214.96 | 61.867 | 43022.1 | 1.56 | | |
| | hAT | 11 | 9.02 | 2.669 | 1389.7 | 0.05 | | |
| | Tag1 | 2 | 1.55 | 0.000 | 394.4 | 0.01 | | |
| Tip100 | | 48 | 42.83 | 4.670 | 9137.6 | 0.33 | | |
| Tc1 | Mariner | 6 | 13.60 | 6.217 | 2621.8 | 0.10 | | |
| | TcMar | 9 | 1.37 | 0.065 | 229.2 | 0.01 | | |
| | Tc1 | 4 | 0.47 | 0.032 | 86.0 | 0.00 | | |
| | Tc2 | 8 | 6.05 | 0.947 | 1476.6 | 0.05 | | |
| | Tigger | 75 | 91.77 | 20.783 | 34907.0 | 1.27 | | |
| Helitron | | 4 | 1.49 | 0.042 | 366.3 | 0.01 | | |
| others | | 56 | 8.13 | 0.254 | 1198.9 | 0.04 | | |
| small RNA pseudogenes | | 83 | 7.842 | 2.763 | 880.5 | 0.03 | 25 | 1 |
| tRNA | | 61 | 0.83 | 0.388 | 45.5 | 0.00 | | |
| U-RNA | | 12 | 2.51 | 1.245 | 197.5 | 0.01 | | |
| others | | 10 | 4.51 | 1.130 | 637.5 | 0.02 | 25 | 1 |
| unclassified | | 60 | 5.05 | 0.168 | 766.0 | 0.03 | | |
| satellite | | | | | 10453.3 | 0.38 | | |
| simple/low complexity | | | | | 14044.8 | 0.51 | | |
| satellite, simple/low complexity | | | | | 24498.1 | 0.89 | | |

| | | | | | | | |
|---------------|------|---------|---------|-----------|-------|-------|----|
| total repeats | 1250 | 3602.04 | 903.582 | 1324968.3 | 48.15 | 13086 | 26 |
|---------------|------|---------|---------|-----------|-------|-------|----|

Table S4. Chromosomal distribution of SINEs, LINEs, LTRs, and DNA transposons in vervet. The percentages in red indicate significant overrepresentations and those in blue underrepresentations of specific elements from the expected chromosomal distribution patterns ($p < 0.05$, two-sided confidential intervals).

| Chr No. | Size | SINE | LINE | LTR | DNA | Sum |
|---------|-----------|---------------|---------------|--------------|--------------|---------------|
| Chr 1 | 124826116 | 14.14% | 20.79% | 8.46% | 3.47% | 46.86% |
| Chr 2 | 90125767 | 14.87% | 17.91% | 9.30% | 3.91% | 45.99% |
| Chr 3 | 91104820 | 10.15% | 20.71% | 9.81% | 3.59% | 44.27% |
| Chr 4 | 90061126 | 10.09% | 21.30% | 10.26% | 3.77% | 45.41% |
| Chr 5 | 74295089 | 20.41% | 14.87% | 7.61% | 3.42% | 46.31% |
| Chr 6 | 50078605 | 29.44% | 12.33% | 7.60% | 2.10% | 51.46% |
| Chr 7 | 134930623 | 9.92% | 23.21% | 10.65% | 3.75% | 47.53% |
| Chr 8 | 138476607 | 12.46% | 21.15% | 9.42% | 3.66% | 46.69% |
| Chr 9 | 127496822 | 14.29% | 19.15% | 8.01% | 3.65% | 45.11% |
| Chr 10 | 128157101 | 11.49% | 20.78% | 8.93% | 3.83% | 45.03% |
| Chr 11 | 127462865 | 15.65% | 20.08% | 8.93% | 3.74% | 48.40% |
| Chr 12 | 107525478 | 14.59% | 20.24% | 8.22% | 3.61% | 46.66% |
| Chr 13 | 96997123 | 10.65% | 22.06% | 9.47% | 3.85% | 46.03% |
| Chr 14 | 106501582 | 13.12% | 20.48% | 8.57% | 3.80% | 45.97% |
| Chr 15 | 90889772 | 12.06% | 22.74% | 9.19% | 4.13% | 48.12% |
| Chr 16 | 74266894 | 22.61% | 14.43% | 6.23% | 3.23% | 46.50% |
| Chr 17 | 71084521 | 13.63% | 20.14% | 8.89% | 3.82% | 46.49% |
| Chr 18 | 71316943 | 11.26% | 20.32% | 8.66% | 3.71% | 43.95% |
| Chr 19 | 32379275 | 23.04% | 14.69% | 6.17% | 2.70% | 46.60% |
| Chr 20 | 132518397 | 17.65% | 18.24% | 8.37% | 3.21% | 47.47% |
| Chr 21 | 126617791 | 11.39% | 21.97% | 8.98% | 3.78% | 46.13% |
| Chr 22 | 100281677 | 12.79% | 18.88% | 9.26% | 3.70% | 44.63% |
| Chr 23 | 82197124 | 13.31% | 22.77% | 8.40% | 3.93% | 48.41% |
| Chr 24 | 83880503 | 13.86% | 20.46% | 9.37% | 3.66% | 47.35% |
| Chr 25 | 84816203 | 13.38% | 19.92% | 8.61% | 3.66% | 45.56% |
| Chr 26 | 57111635 | 16.39% | 20.27% | 7.24% | 3.83% | 47.73% |

| | | | | | | |
|--------|-----------|---------------|---------------|---------------|--------------|---------------|
| Chr 27 | 47839205 | 12.20% | 18.60% | 11.34% | 3.38% | 45.52% |
| Chr 28 | 20925048 | 28.83% | 12.25% | 6.55% | 2.51% | 50.15% |
| Chr 29 | 23194370 | 15.08% | 19.22% | 7.69% | 3.60% | 45.58% |
| Chr X | 141552053 | 10.75% | 31.95% | 11.22% | 3.44% | 57.36% |
| Chr Y | 7215306 | 8.78% | 24.87% | 21.22% | 1.77% | 56.63% |

Table S5. Chromosomal distribution of SINEs, LINEs, LTRs, and DNA transposons in human. The percentages in red indicate significant overrepresentations and those in blue underrepresentations of specific elements from the expected chromosomal distribution patterns ($p < 0.05$, two-sided confidential intervals).

| Chr No. | Size | SINE | LINE | LTR | DNA | Sum |
|---------|-----------|---------------|---------------|---------------|--------------|---------------|
| Chr 1 | 249250621 | 14.24% | 19.53% | 7.87% | 3.17% | 44.82% |
| Chr 2 | 243199373 | 11.75% | 21.82% | 8.72% | 3.69% | 45.98% |
| Chr 3 | 198022430 | 11.99% | 22.84% | 9.40% | 3.88% | 48.11% |
| Chr 4 | 191154276 | 10.12% | 23.73% | 10.97% | 3.57% | 48.40% |
| Chr 5 | 180915260 | 11.19% | 23.17% | 9.53% | 3.69% | 47.59% |
| Chr 6 | 171115067 | 11.51% | 22.64% | 9.25% | 3.70% | 47.10% |
| Chr 7 | 159138663 | 13.76% | 21.07% | 8.49% | 3.41% | 46.73% |
| Chr 8 | 146364022 | 11.83% | 22.23% | 9.53% | 3.47% | 47.06% |
| Chr 9 | 141213431 | 11.94% | 18.76% | 7.30% | 2.94% | 40.94% |
| Chr 10 | 135534747 | 13.73% | 20.24% | 8.08% | 3.50% | 45.55% |
| Chr 11 | 135006516 | 13.48% | 22.62% | 8.73% | 3.32% | 48.15% |
| Chr 12 | 133851895 | 14.93% | 21.48% | 9.18% | 3.65% | 49.24% |
| Chr 13 | 115169878 | 8.52% | 18.78% | 8.61% | 2.99% | 38.90% |
| Chr 14 | 107349540 | 11.72% | 18.85% | 8.40% | 2.99% | 41.97% |
| Chr 15 | 102531392 | 13.50% | 18.89% | 6.51% | 3.13% | 42.03% |
| Chr 16 | 90354753 | 19.29% | 15.06% | 7.31% | 3.17% | 44.83% |
| Chr 17 | 81195210 | 24.29% | 16.60% | 6.72% | 3.51% | 51.11% |
| Chr 18 | 78077248 | 10.76% | 21.21% | 8.93% | 3.54% | 44.44% |
| Chr 19 | 59128983 | 31.48% | 16.28% | 10.26% | 2.38% | 60.40% |
| Chr 20 | 63025520 | 15.98% | 18.97% | 8.39% | 4.07% | 47.41% |
| Chr 21 | 48129895 | 8.98% | 15.67% | 9.39% | 2.54% | 36.59% |
| Chr 22 | 51304566 | 16.38% | 12.24% | 5.14% | 2.10% | 35.87% |
| Chr X | 155270560 | 10.35% | 33.02% | 10.87% | 3.16% | 57.41% |
| Chr Y | 59373566 | 4.47% | 10.95% | 7.29% | 0.76% | 23.46% |

Table S6. Chromosomal distribution of SINEs, LINEs, LTRs, and DNA transposons in rhesus macaque (rheMac7). The percentages in red indicate significant overrepresentations and those in blue underrepresentations of specific elements from the expected chromosomal distribution patterns ($p < 0.05$, two-sided confidential intervals).

| Chr No. | Size | SINE | LINE | LTR | DNA | Sum |
|---------|-----------|---------------|---------------|---------------|--------------|---------------|
| Chr 1 | 229590362 | 14.15% | 17.23% | 7.73% | 3.25% | 42.36% |
| Chr 2 | 192599291 | 11.19% | 18.88% | 8.64% | 3.81% | 42.53% |
| Chr 3 | 198365852 | 11.30% | 16.55% | 8.03% | 3.13% | 39.01% |
| Chr 4 | 169853907 | 10.90% | 18.45% | 8.40% | 3.63% | 41.38% |
| Chr 5 | 184069350 | 9.55% | 19.44% | 10.18% | 3.51% | 42.69% |
| Chr 6 | 180491593 | 10.48% | 18.70% | 8.62% | 3.55% | 41.35% |
| Chr 7 | 170124641 | 13.32% | 17.10% | 7.42% | 3.42% | 41.26% |
| Chr 8 | 150158102 | 11.08% | 17.83% | 8.33% | 3.30% | 40.55% |
| Chr 9 | 130959383 | 13.01% | 16.78% | 7.42% | 3.41% | 40.61% |
| Chr 10 | 95678458 | 15.69% | 14.74% | 6.77% | 3.32% | 40.52% |
| Chr 11 | 134985053 | 14.28% | 17.66% | 8.31% | 3.51% | 43.76% |
| Chr 12 | 106825570 | 10.69% | 17.94% | 8.28% | 3.73% | 40.64% |
| Chr 13 | 137160470 | 11.40% | 17.88% | 7.71% | 3.39% | 40.38% |
| Chr 14 | 135538710 | 11.99% | 17.17% | 7.33% | 3.06% | 39.54% |
| Chr 15 | 135538710 | 11.03% | 14.94% | 6.38% | 2.83% | 35.18% |
| Chr 16 | 80904711 | 19.22% | 12.62% | 5.49% | 2.92% | 40.26% |
| Chr 17 | 94676912 | 9.37% | 18.47% | 9.23% | 3.38% | 40.46% |
| Chr 18 | 74429130 | 10.07% | 17.77% | 8.09% | 3.43% | 39.36% |
| Chr 19 | 64156825 | 20.40% | 8.85% | 5.60% | 1.62% | 36.46% |
| Chr 20 | 88476065 | 15.89% | 11.53% | 5.93% | 2.78% | 36.13% |
| Chr X | 155416530 | 9.05% | 25.69% | 9.63% | 3.01% | 47.37% |

Table S7. Mapping statistics of VRC monkeys used for structural variation discovery

| Vervet ID | Aligned Reads | %Alignment | %Duplicated Reads | Estimated Coverage |
|-----------|---------------|------------|-------------------|--------------------|
| 1985088 | 998479432 | 97.7 | 18.9 | 36.1X |
| 1986014 | 1004994181 | 97.9 | 18.6 | 36.4X |
| 1987079 | 912967664 | 97.5 | 18.6 | 32.3X |
| 1988009 | 929826449 | 97.8 | 18.8 | 33.6X |
| 1990016 | 998180018 | 98.8 | 17.3 | 35.3X |
| 1996031 | 658571737 | 98.7 | 16.4 | 23.3X |

Table S8. Deletion variants defined in the vervet research colony population¹

| Vervet ID | Total Deletions | Depth Validated Deletions | Total Genes | Null Gene Deletions |
|-----------|--------------------|------------------------------|----------------|------------------------|
| 1985088 | 200 | 109 | 12 | 1 |
| 1986014 | 203 | 104 | 9 | 1 |
| 1987079 | 168 | 93 | 16 | 1 |
| 1988009 | 98 | 60 | 4 | 0 |
| 1990016 | 200 | 116 | 26 | 3 |
| 1996031 | 138 | 74 | 10 | 0 |
| Total | 1007 | 556 | 77 | 6 |
| Shared | 2201 | 672 | 57 | 5 |

¹Vervets housed at Wake Forest Primate Center

Table S9. The total estimated base loss events by length unique for each VRC individual.

| <u>Group</u> | <u>Count</u> | <u>Sum-Length</u> |
|-------------------|--------------|-------------------|
| >1,000,000 | 0 | 0 |
| 700,001-1,000,000 | 0 | 0 |
| 500,001-700,000 | 0 | 0 |
| 300,001-500,000 | 1 | 472,648 |
| 100,001-300,000 | 2 | 310,809 |
| 50,001-100,000 | 1 | 83,643 |
| 10,001-50,000 | 21 | 368,723 |
| 5,001-10,000 | 49 | 342,995 |
| 2,001-5,000 | 122 | 378,027 |
| 501-2,000 | 359 | 379,810 |
| 0-500 | 1 | 500 |
| Total | 556 | 2,337,155 |

Table S10. The total estimated base loss events shared by any 3 sequenced VRC individuals.

| Group | Count | Sum-Length |
|-------------------|------------|------------------|
| 0-500 | 1 | 500 |
| 501-2,000 | 410 | 414,568 |
| 2,001-5,000 | 132 | 412,944 |
| 5,001-10,000 | 71 | 490,966 |
| 10,001-50,000 | 41 | 837,098 |
| 50,001-100,000 | 8 | 537,143 |
| 100,001-300,000 | 4 | 670,164 |
| 300,001-500,000 | 5 | 2,021,740 |
| 500,001-700,000 | 0 | 0 |
| 700,001-1,000,000 | 0 | 0 |
| >1,000,000 | 0 | 0 |
| Total | 672 | 5,385,123 |

Table S11. Autosomes 1-29 segmental duplication base counts

| Sample | Total duplicated bp | Shared duplicated bp | Non sample-specific duplicated bp | Sample-specific duplicated bp | %Shared duplicated bp | %Sample-specific duplicated bp |
|---------|---------------------|----------------------|-----------------------------------|-------------------------------|-----------------------|--------------------------------|
| 1985088 | 45047327 | 35519790 | 8427020 | 1100517 | 0.79 | 0.02 |
| 1986014 | 43735125 | 35519790 | 7285538 | 929797 | 0.81 | 0.02 |
| 1987079 | 44203395 | 35519790 | 7994310 | 689295 | 0.8 | 0.02 |
| 1988009 | 43334974 | 35519790 | 7074055 | 741129 | 0.82 | 0.02 |
| 1990016 | 44081727 | 35519790 | 7222277 | 1339660 | 0.81 | 0.03 |
| 1996031 | 42470402 | 35519790 | 6384619 | 565993 | 0.84 | 0.01 |

Table S12. Genes residing in segmental duplication regions showing enrichment among canonical KEGG pathways

| Canonical pathway | P value | Genes |
|------------------------------|---------|--|
| Glutathione metabolism | .01 | GGT1, GSTM5, GSTA5, GSTM1 |
| Chemokine signalling | .02 | CXCL5, VAV3, JAK2, CCL3, CXCL3, DOCK2 |
| Basal transcription | .02 | TAF13, GTF2I |
| Drug metabolism | .04 | GUSB, CES5A, CES1 |
| Arginine/Proline metabolism | .04 | CKMT1B, ACY1, NOS2 |
| Metabolic pathways | .04 | GGT1, CKMT1B, DHRS4L2, ACY1, GUSB, CES5A, LPCAT4, CES1, MGAM, MGAT4C, NOS2, SDHA, POLR2B, PSPH, PFKP |
| Complementation pathways | .04 | C1R, CFH, CR1 |
| Leishmaniasis | .04 | JAK2, NOS2, CR1 |
| Natural killer cell activity | .04 | VAV3, RAET1G, TNFRSF10D, ULBP1 |

Table S13. Summary of gene gain and loss events inferred after correcting for annotation and assembly error across all 11 species. The number of rapidly evolving families is shown in parentheses for each type of change.

| | Expansions | | | Contractions | | | No Change | Avg. Expansion |
|------------|------------|-------|---------------------|--------------|-------|-----------------------|-----------|----------------|
| | Families | Genes | genes/ expansion | Families | Genes | genes/ contraction | | |
| Macaque | 1018 (153) | 1810 | 1.78 | 958 (17) | 995 | 1.04 | 10020 | 0.067939 |
| Tarsier | 1183 (30) | 1680 | 1.42 | 1797 (15) | 1941 | 1.08 | 9016 | -0.021757 |
| Gibbon | 313 (8) | 400 | 1.28 | 1357 (59) | 1621 | 1.19 | 10326 | -0.101784 |
| Human | 1225 (288) | 2727 | 2.23 | 234 (9) | 246 | 1.05 | 10537 | 0.206819 |
| Rat | 988 (108) | 1833 | 1.86 | 343 (20) | 412 | 1.20 | 10665 | 0.118456 |
| Chimpanzee | 162 (29) | 190 | 1.17 | 1027 (63) | 1218 | 1.19 | 10807 | -0.085695 |
| Marmoset | 991 (103) | 1912 | 1.93 | 1449 (17) | 1554 | 1.07 | 9556 | 0.029843 |
| Bushbaby | 641 (26) | 979 | 1.53 | 2081 (8) | 2275 | 1.09 | 9274 | -0.108036 |
| Vervet | 510 (38) | 654 | 1.28 | 535 (38) | 647 | 1.21 | 10951 | 0.000584 |
| Colugo | 2074 (40) | 3164 | 1.53 | 1507 (19) | 1685 | 1.12 | 8415 | 0.123291 |
| Mouse | 419 (60) | 931 | 2.22 | 433 (42) | 598 | 1.38 | 11144 | 0.027759 |

Table S14. A summary of sequencing measures for vervet subspecies.

| VRC number | Subspecies | Country | Longitude/Latitude | % Aligned | Average coverage | Heterozygosity (%) | Derived alleles (million) |
|------------|--------------------------|--------------------------|--------------------|-----------|------------------|--------------------|---------------------------|
| VGA00006 | <i>C. a. sabaesus</i> | Gambia | -16.65/13.38 | 99.4 | 10.8 | 0.13 | 11.2 |
| VEC1016 | <i>C. a. aethiops</i> | Ethiopia | 37.40/11.59 | 99.3 | 9.1 | 0.08 | 11.4 |
| AGM141 | <i>C. a. tantalus</i> | Central African Republic | 18.58/4.36 | 99.4 | 10.0 | 0.17 | 11.3 |
| AG5417 | <i>C. a. pygerythrus</i> | Tanzania | 35.00/-7.00 | 99.5 | 10.2 | 0.18 | 11.4 |
| VZA3008 | <i>C. a. cynosurus</i> | Zambia | 25.99/-14.97 | 98.7 | 8.2 | 0.14 | 11.0 |

Table S15. Autosomal pairwise difference matrix across all subspecies

| | <i>C. a. pygerythrus</i> | <i>C. a. tantalus</i> | <i>C. a. aethiops</i> | <i>C. a. sabaesus</i> | <i>C. a. cynosurus</i> |
|--------------------------|--------------------------|-----------------------|-----------------------|-----------------------|------------------------|
| <i>C. a. pygerythrus</i> | 4073196 | 6030851 | 6496417 | 7564638 | 4702764 |
| <i>C. a. tantalus</i> | 6030851 | 3944089 | 6509592 | 7212378 | 5851541 |
| <i>C. a. aethiops</i> | 6496417 | 6509592 | 1911259 | 7738523 | 6296262 |
| <i>C. a. sabaesus</i> | 7564638 | 7212378 | 7738523 | 3042179 | 7332660 |
| <i>C. a. cynosurus</i> | 4702764 | 5851541 | 6296262 | 7332660 | 3239522 |

Table S16. Estimated subspecies split times in years.

| | <i>C. a.</i> <i>pygerythus</i> | <i>C.a.</i> <i>tantalus</i> | <i>C. a.</i> <i>aethiops</i> | <i>C. a.</i> <i>sabaeus</i> | <i>C. a.</i> <i>cynosurus</i> |
|-----------------------------------|-----------------------------------|--------------------------------|---------------------------------|--------------------------------|----------------------------------|
| <i>C. a.</i> <i>pygerythus</i> | 0 | 250146 | 433466 | 495657 | 129440 |
| <i>C. a.</i> <i>tantalus</i> | 250146 | 0 | 443081 | 460068 | 279528 |
| <i>C. a.</i> <i>aethiops</i> | 433466 | 443081 | 0 | 650882 | 460269 |
| <i>C. a.</i> <i>sabaeus</i> | 495657 | 460068 | 650882 | 0 | 518524 |
| <i>C. a.</i> <i>cynosurus</i> | 129440 | 279528 | 460269 | 518524 | 0 |

Table S17. Inferred average coalescent time in years for all subspecies pairs

| | <i>C.a.</i> <i>pygerythus</i> | <i>C. a.</i> <i>tantalus</i> | <i>C. a.</i> <i>aethiops</i> | <i>C. a.</i> <i>sabaeus</i> | <i>C. a.</i> <i>cynosurus</i> |
|-----------------------------------|----------------------------------|---------------------------------|---------------------------------|--------------------------------|----------------------------------|
| <i>C. a.</i> <i>pygerythus</i> | 503852 | 746012 | 803602 | 935741 | 581729 |
| <i>C. a.</i> <i>tantalus</i> | 746012 | 487881 | 805232 | 892166 | 723832 |
| <i>C. a.</i> <i>aethiops</i> | 803602 | 805232 | 236421 | 957250 | 778843 |
| <i>C. a.</i> <i>sabaeus</i> | 935741 | 892166 | 957250 | 376316 | 907045 |
| <i>C. a.</i> <i>cynosurus</i> | 581729 | 723832 | 778843 | 907045 | 400727 |

Note: To make our results more comparable with standard phylogenetic methods, we also report the phylogenetic tree estimated from the average coalescent time directly:

(sab:923.0505821,(aet:795.8926726,(tan:734.9220112,(pyn:581.728731,cyn:581.728731):153.1932802):60.97066143):127.1579095). We applied the same clustering procedure to the matrix of geographic distances of our samples in kilometers, leading to the following tree:

(sab:2705.662183,((pyn:661.7764172,cyn:661.7764172):565.0045939,(aet:1110.859628,tan:1110.859628):115.9213835):1478.881172);

Table S18. Sources of vervet (*C. a. sabaesus*) evaluated for MHC diversity

| Institution | # Animals | Origin |
|--|-----------|--|
| Tulane National Primate Research Center | 5 | St Kitts juveniles; Three Springs Scientific |
| Beth Israel Deaconess Medical Center, Harvard Medical School | 19 | New Iberia Research Center (16), NERPC (3) |
| Wake Forest University Primate Center | 20 | Vervet Research Colony |
| University of Pittsburgh | 6 | New Iberia Research Center |
| Wisconsin National Primate Research Center | 1 | Unknown |
| University of California, Los Angeles | 21 | St Kitts - Feral |
| University of California, Los Angeles | 11 | Ghana - Feral |

Table S19. Summary of Chsa MHC class I sequences identified

| Allele Name | Chsa Haplotype | Origin | Genbank ID | IPD ID | Length | Reference Animal(s) |
|-------------------------|----------------|-----------|------------|----------|--------|-----------------------|
| <i>Chsa-A*01:01</i> | A1 | Caribbean | HQ686146 | 70007085 | 1108 | V002, V005 AGM074, |
| <i>Chsa-A*01:nov:01</i> | | Ghana | HQ686170 | | 530 | AGM077 |
| <i>Chsa-A*02:01</i> | A2 | Caribbean | HQ686144 | 70007082 | 1108 | V001, V004 |
| <i>Chsa-A*03:01</i> | A3 | Caribbean | HQ686151 | 70007083 | 693 | V004 |
| <i>Chsa-A*03:02</i> | A1 | Caribbean | HQ686145 | 70007084 | 1108 | V002, V005 |
| <i>Chsa-A*03:nov:01</i> | | Ghana | HQ686169 | | 530 | AGM080 |
| <i>Chsa-A*04:01</i> | A4 | Caribbean | HQ686143 | 70007081 | 1093 | V003, V005 |
| <i>Chsa-A*05:01</i> | A3 | Caribbean | HQ686166 | 70007086 | 659 | V004 |
| <i>Chsa-A*06:01</i> | A2 | Caribbean | HQ686147 | 70007087 | 1108 | V001, V004 AGM083, |
| <i>Chsa-A*06:nov:01</i> | A5 | Caribbean | HQ686172 | | 530 | AGM037 AGM071, |
| <i>Chsa-A*06:nov:02</i> | | Ghana | HQ686173 | | 530 | AGM075 AGM072, |
| <i>Chsa-A*06:nov:03</i> | | Ghana | HQ686174 | | 530 | AGM073 |
| <i>Chsa-A*06:nov:04</i> | | Ghana | HQ686175 | | 530 | AGM079 |
| <i>Chsa-A*06:nov:05</i> | | Ghana | HQ686176 | | 530 | AGM073 AGM083, |
| <i>Chsa-A*07:nov:01</i> | A5 | Caribbean | HQ686171 | | 530 | AGM037 AGM049, |
| <i>Chsa-A*08:nov:01</i> | A6 | Caribbean | HQ686177 | | 530 | AGM069 |
| <i>Chsa-A*10:nov:01</i> | | Ghana | HQ686178 | | 530 | AGM070 |
| <i>Chsa-A*11:nov:01</i> | | Ghana | HQ686179 | | 530 | AGM070 |
| <i>Chsa-A*12:nov:01</i> | | Ghana | HQ686180 | | 530 | AGM071 |
| <i>Chsa-A*13:nov:01</i> | | Ghana | HQ686181 | | 530 | AGM071 AGM074, |
| <i>Chsa-A*14:nov:01</i> | | Ghana | HQ686182 | | 530 | AGM076 |
| <i>Chsa-A*15:nov:01</i> | | Ghana | HQ686183 | | 530 | AGM078 |
| <i>Chsa-A*16:nov:01</i> | | Ghana | HQ686184 | | 530 | AGM078 AGM072, |
| <i>Chsa-A*17:nov:01</i> | | Ghana | HQ686185 | | 530 | AGM073 |
| <i>Chsa-A*18:nov:01</i> | | Ghana | HQ686186 | | 530 | AGM072 |

| | | | | | | |
|-------------------------|--------------------|------------------|----------|----------|------|-----------------------|
| <i>Chsa-A*18:nov:02</i> | | Ghana | HQ686187 | | 530 | AGM076 |
| <i>Chsa-A*19:nov:01</i> | | Ghana | HQ686188 | | 530 | AGM073 |
| <i>Chsa-A*20:nov:01</i> | | Ghana | HQ686189 | | 530 | AGM077 |
| <i>Chsa-A*21:nov:01</i> | | Ghana | HQ686190 | | 530 | AGM075 |
| <i>Chsa-A*22:nov:01</i> | | Ghana | HQ686191 | | 530 | AGM075 |
| <i>Chsa-A*23:nov:01</i> | | Ghana | HQ686192 | | 530 | AGM079 |
| <i>Chsa-A*24:nov:01</i> | | Ghana | HQ686193 | | 530 | AGM079 |
| <i>Chsa-A*25:nov:01</i> | | Ghana | HQ686194 | | 530 | AGM080 |
| <i>Chsa-A*26:nov:01</i> | | Ghana | HQ686195 | | 530 | AGM080 |
| <i>Chsa-B*01:01</i> | B4a, B4b, B4d | Caribbean, Ghana | HQ686159 | 70007100 | 692 | V003, V005 |
| <i>Chsa-B*02:01</i> | B2 | Caribbean | HQ686153 | 70007093 | 692 | V001, V002 |
| <i>Chsa-B*03:01</i> | B2 | Caribbean | HQ686157 | 70007098 | 974 | V001, V002 AGM034, |
| <i>Chsa-B*03:nov:01</i> | B5 | Caribbean, Ghana | HQ686206 | | 530 | AGM093 |
| <i>Chsa-B*04:01</i> | B1 | Caribbean | HQ686150 | 70007094 | 1107 | V005 |
| <i>Chsa-B*05:01</i> | B1, B11 | Caribbean, Ghana | HQ686167 | 70007103 | 659 | V005 |
| <i>Chsa-B*05:nov:01</i> | B13 | Ghana | HQ686208 | | 530 | AGM077 |
| <i>Chsa-B*05:nov:02</i> | B15 | Ghana | HQ686209 | | 530 | AGM075 |
| <i>Chsa-B*07:01</i> | B4a, B4b, B4c, B4d | Caribbean, Ghana | HQ686158 | 70007099 | 692 | V003, V005 |
| <i>Chsa-B*08:01</i> | B2 | Caribbean | HQ686148 | 70007088 | 1107 | V001, V002 |
| <i>Chsa-B*08:02</i> | B3a, B3b | Caribbean, Ghana | HQ686149 | 70007089 | 1107 | V004, AGM074 |
| <i>Chsa-B*09:01</i> | B4a, B4b, B4c | Caribbean, Ghana | HQ686163 | 70007102 | 693 | V003, V005 AGM071, |
| <i>Chsa-B*09:nov:01</i> | B4d | Ghana | HQ686207 | | 530 | AGM076 |
| <i>Chsa-B*10:01</i> | B3a, B3b | Caribbean, Ghana | HQ686164 | 70007101 | 680 | V004, AGM074 |
| <i>Chsa-B*10:nov:01</i> | B12 | Ghana | HQ686165 | | 551 | AGM080 |
| <i>Chsa-B*11:01</i> | B2 | Caribbean | HQ686161 | 70007090 | 693 | V001, V002 |
| <i>Chsa-B*11:02</i> | B1 | Caribbean | HQ686162 | 70007091 | 693 | V005 |
| <i>Chsa-B*11:nov:01</i> | B10a | Ghana | HQ686196 | | 530 | AGM075 |
| <i>Chsa-B*11:nov:02</i> | B10b | Ghana | HQ686197 | | 530 | AGM079 |
| <i>Chsa-B*12:01</i> | B1 | Caribbean | HQ686168 | 70007104 | 659 | V005 |
| <i>Chsa-B*13:01</i> | B3a | Caribbean | HQ686156 | 70007097 | 692 | V004 |
| <i>Chsa-B*13:nov:01</i> | B3b | Ghana | HQ686201 | | 530 | AGM074 AGM074, |
| <i>Chsa-B*13:nov:02</i> | B7a, B7b | Ghana | HQ686202 | | 530 | AGM080 |
| <i>Chsa-B*14:01</i> | B1 | Caribbean | HQ686152 | 70007092 | 692 | V005 |

| | | | | | | |
|---------------------------|-------------|------------------|----------|----------|-----|-------------------|
| <i>Chsa-B*15:01</i> | B3a, B3b | Caribbean, Ghana | HQ686154 | 70007095 | 692 | V004, AGM074 |
| <i>Chsa-B*15:02</i> | B1 | Caribbean | HQ686155 | 70007096 | 692 | V005 |
| <i>Chsa-B*15:nov:01</i> | B14 | Ghana | HQ686198 | | 530 | AGM076 |
| <i>Chsa-B*15:nov:02</i> | B16 | Ghana | HQ686199 | | 530 | AGM079 |
| <i>Chsa-B*15:nov:03</i> | B12 | Ghana | HQ686200 | | 530 | AGM080 |
| <i>Chsa-B*16:nov:01</i> | B5, B8, B12 | Caribbean, Ghana | HQ686210 | | 530 | AGM083, AGM037 |
| <i>Chsa-B*18:nov:01</i> | B5 | Caribbean, Ghana | HQ686211 | | 530 | AGM083, AGM037 |
| <i>Chsa-B*19:nov:01</i> | B5 | Caribbean, Ghana | HQ686212 | | 530 | AGM083, AGM037 |
| <i>Chsa-B*20:nov:01</i> | B5 | Caribbean, Ghana | HQ686213 | | 530 | AGM037 |
| <i>Chsa-B*21:nov:01</i> | B4a | Caribbean | HQ686203 | | 529 | V003, V005 |
| <i>Chsa-B*21:nov:02</i> | B7a | Ghana | HQ686204 | | 530 | AGM074, AGM080 |
| <i>Chsa-B*21:nov:03</i> | B7b | Ghana | HQ686205 | | 530 | AGM078 |
| <i>Chsa-B*23:nov:01</i> | B4a | Caribbean | HQ686214 | | 530 | V003, V005 |
| <i>Chsa-B*25:nov:01</i> | B5 | Caribbean, Ghana | HQ686215 | | 530 | AGM083, AGM037 |
| <i>Chsa-B*25:nov:02</i> | B7a | Ghana | HQ686216 | | 530 | AGM074, AGM080 |
| <i>Chsa-B*25:nov:03</i> | B7b | Ghana | HQ686217 | | 530 | AGM078 |
| <i>Chsa-B*26:nov:01</i> | B6 | Caribbean | HQ686218 | | 530 | AGM049, AGM069 |
| <i>Chsa-B*27:nov:01</i> | B6 | Caribbean | HQ686219 | | 530 | AGM049, AGM069 |
| <i>Chsa-B*27:nov:01sv</i> | B6 | Caribbean | HQ686220 | | 467 | AGM069 |
| <i>Chsa-B*28:nov:01</i> | B11 | Ghana | HQ686221 | | 530 | AGM070 |
| <i>Chsa-B*29:nov:01</i> | B11 | Ghana | HQ686222 | | 530 | AGM070 |
| <i>Chsa-B*30:nov:01</i> | B4b | Ghana | HQ686223 | | 530 | AGM071, AGM073 |
| <i>Chsa-B*30:nov:02</i> | B4c | Ghana | HQ686224 | | 530 | AGM072 |
| <i>Chsa-B*31:nov:01</i> | B4d | Ghana | HQ686225 | | 530 | AGM071, AGM076 |
| <i>Chsa-B*32:nov:01</i> | B4d | Ghana | HQ686226 | | 530 | AGM071, AGM076 |
| <i>Chsa-B*33:nov:01</i> | B7a, B7b | Ghana | HQ686227 | | 530 | AGM074, |

| | | | | | |
|-------------------------|------------|-----------|----------|-----|-------------------|
| | | | | | AGM080 |
| <i>Chsa-B*34:nov:01</i> | B8 | Ghana | HQ686228 | 530 | AGM077, AGM078 |
| <i>Chsa-B*35:nov:01</i> | B8 | Ghana | HQ686229 | 530 | AGM077, AGM078 |
| <i>Chsa-B*36:nov:01</i> | B8 | Ghana | HQ686230 | 530 | AGM072, AGM073 |
| <i>Chsa-B*38:nov:01</i> | B9 | Ghana | HQ686231 | 530 | AGM072, AGM073 |
| <i>Chsa-B*39:nov:01</i> | B9 | Ghana | HQ686232 | 530 | AGM072, AGM073 |
| <i>Chsa-B*40:nov:01</i> | B9 | Ghana | HQ686233 | 530 | AGM077 |
| <i>Chsa-B*41:nov:01</i> | B13 | Ghana | HQ686234 | 530 | AGM077 |
| <i>Chsa-B*42:nov:01</i> | B13 | Ghana | HQ686235 | 530 | AGM075 |
| <i>Chsa-B*42:nov:02</i> | B10a | Ghana | HQ686236 | 530 | AGM079 |
| <i>Chsa-B*42:nov:03</i> | B10b | Ghana | HQ686237 | 530 | AGM077 |
| <i>Chsa-B*43:nov:01</i> | B13 | Ghana | HQ686238 | 530 | AGM076 |
| <i>Chsa-B*44:nov:01</i> | B13 | Ghana | HQ686239 | 530 | AGM076 |
| <i>Chsa-B*45:nov:01</i> | B14 | Ghana | HQ686240 | 530 | AGM076 |
| <i>Chsa-B*46:nov:01</i> | B14 | Ghana | HQ686241 | 530 | AGM076 |
| <i>Chsa-B*47:nov:01</i> | B14 | Ghana | HQ686242 | 530 | AGM070 |
| <i>Chsa-B*47:nov:02</i> | B11 | Ghana | HQ686243 | 530 | AGM076 |
| <i>Chsa-B*48:nov:01</i> | B14 | Ghana | HQ686244 | 530 | AGM075 |
| <i>Chsa-B*49:nov:01</i> | B15 | Ghana | HQ686245 | 530 | AGM075 |
| <i>Chsa-B*50:nov:01</i> | B15 | Ghana | HQ686246 | 530 | AGM075 |
| <i>Chsa-B*51:nov:01</i> | B15 | Ghana | HQ686247 | 530 | AGM075, AGM079 |
| <i>Chsa-B*52:nov:01</i> | B10a, B10b | Ghana | HQ686248 | 530 | AGM079 |
| <i>Chsa-B*53:nov:01</i> | B16 | Ghana | HQ686249 | 530 | AGM079 |
| <i>Chsa-B*54:nov:01</i> | B16 | Ghana | HQ686250 | 530 | AGM079 |
| <i>Chsa-B*55:nov:01</i> | B16 | Ghana | HQ686251 | 530 | AGM080 |
| <i>Chsa-B*56:nov:01</i> | B12 | Ghana | HQ686252 | 530 | AGM085, V001 |
| <i>Chsa-E*01:nov:01</i> | E1, E2 | Caribbean | HQ686254 | 530 | AGM037 |
| <i>Chsa-E*01:nov:02</i> | E4 | Caribbean | HQ686253 | 530 | AGM070 |
| <i>Chsa-E*01:nov:04</i> | | Ghana | HQ686255 | 530 | AGM074 |
| <i>Chsa-E*01:nov:05</i> | | Ghana | HQ686256 | 530 | |

| | | | | | |
|-------------------------|--------|-----------|----------|-----|------------------------------|
| <i>Chsa-E*01:nov:06</i> | | Ghana | HQ686257 | 530 | AGM078 |
| <i>Chsa-E*01:nov:07</i> | | Ghana | HQ686258 | 530 | AGM074 |
| <i>Chsa-E*01:nov:08</i> | | Ghana | HQ686259 | 530 | AGM073 |
| <i>Chsa-E*01:nov:09</i> | | Ghana | HQ686260 | 530 | AGM077 |
| <i>Chsa-E*02:nov:01</i> | E2, E5 | Caribbean | HQ686261 | 530 | V001, AGM083 |
| <i>Chsa-E*02:nov:02</i> | E6 | Caribbean | HQ686262 | 530 | AGM049, AGM069 AGM070, |
| <i>Chsa-E*02:nov:03</i> | | Ghana | HQ686263 | 530 | AGM077 |
| <i>Chsa-E*02:nov:04</i> | | Ghana | HQ686264 | 530 | AGM079 AGM071, |
| <i>Chsa-E*04:nov:01</i> | | Ghana | HQ686265 | 530 | AGM079 |
| <i>Chsa-E*05:nov:01</i> | | Ghana | HQ686266 | 530 | AGM078 AGM072, |
| <i>Chsa-E*05:nov:02</i> | | Ghana | HQ686267 | 530 | AGM076 AGM072, |
| <i>Chsa-E*06:nov:01</i> | | Ghana | HQ686268 | 530 | AGM073 |
| <i>Chsa-E*07:nov:01</i> | | Ghana | HQ686269 | 530 | AGM075 |

Table S20. The iterative masking steps for VRC segmental duplication discovery.

| Sample | Raw Reads | Raw Coverage | <u>BWA Mapping and PCR duplicate removal</u> | | | | <u>mFAST aligned</u> | |
|---------|------------|--------------|--|------------|-------------------|--------------------|----------------------|--------------------|
| | | | Aligned Reads | %Alignment | %Duplicated Reads | Estimated Coverage | Aligned Reads | Effective Coverage |
| 1985088 | 1255412834 | 45.7X | 998479432 | 97.7 | 18.9 | 36.1X | 923088520 | 23.78X |
| 1986014 | 1256976468 | 45.8X | 1004994181 | 97.9 | 18.6 | 36.4X | 936889027 | 24.14X |
| 1987079 | 1145288904 | 41.7X | 912967664 | 97.5 | 18.6 | 32.3X | 832534104 | 21.45X |
| 1988009 | 1166731378 | 42.5X | 929826449 | 97.8 | 18.8 | 33.6X | 865169866 | 22.29X |
| 1990016 | 1218844692 | 44.4X | 998180018 | 98.8 | 17.3 | 35.3X | 908056277 | 23.4X |
| 1996031 | 796267172 | 29.0X | 658571737 | 98.7 | 16.4 | 23.3X | 595190299 | 15.34X |

Table S21. Assembly and annotation error estimation and gene gain/loss rates in the 11 mammals included in this study compared to the same values for the 10 mammals used in Han et al, 2013.

| | λ (No Error Model) | ε (Estimated error) | λ (Error Model = ε) |
|-----------------|----------------------------|---------------------------------|--|
| This study | 0.00248 | 0.05171 | 0.00202 |
| Han et al, 2013 | 0.00238 | 0.07324 | 0.00186 |

Table S22. SNV filters applied to each subspecies aligned sequences.

| Filter name | GATK JEXL expression | Description |
|----------------|--|---|
| StrandBias | $FS > 60.0$ | Highly significant strand bias. |
| ABHet | $ABHet < 0.2 \parallel ABHet > 0.8$ | Allele ratio in heterozygotes unbalanced (away from 0.5). |
| MQ0_10pct | $MQ0 \geq 4 \ \&\& \ ((MQ0/(1.0*DP)) > 0.1)$ | Large amount of mapping quality 0 reads. |
| ExCov | - | Raw coverage over all individuals $> 1.5x$ median for this chromosome. |
| HaplotypeScore | $HaplotypeScore > 13.0$ | Low consistency of the site with two (and only two) segregating haplotypes. |
| QualityByDepth | $QD < 2.0$ | Low quality by depth. |
| 5bpIndel | - | Indels reported within $\pm 5bp$. |
| LowCov | - | Raw coverage over all individuals $< 1/2$ median for this chromosome. |
| ReadPos | $ReadPosRankSum < -8.0$ | Bias in read position. |
| LowQual | $QUAL < 50$ | Low quality. |
| LowMQ | $MQ < 25.0$ | Low mapping quality. |