Origin of the Y Chromosome by Craig Paardekooper

Are we really from Chimpanzees, or are we more like Gorillas or Orangutans? Out of all our chromosomes, the most divergent from chimpanzees is the Y chromosome. Here I use codon frequencies as a finger print to find out where we got our Y chromosome from.

CHIMPS

I used the latest versions of the human and chimp genomes. Here are the current codon counts for the Y chromosome in each species -

Codon	Human Y	Chimp Y	Abs Diff	% Diff
TTT	505517	323622	181895	56.21
TTC	287487	174763	112724	64.50
TTA	382248	165830	216418	130.51
TTG	324313	166557	157756	94.72
TCT	274494	188254	86240	45.81
TCC	222646	131468	91178	69.35
TCA	308661	173549	135112	77.85
TCG	154285	16582	137703	830.44
CTT	235834	167780	68054	40.56
CTC	228150	134597	93553	69.51
CTA	195394	106949	88445	82.70
CTG	278330	169624	108706	64.09
CCT	209264	144372	64892	44.95
CCC	135841	99541	36300	36.47
CCA	257862	156675	101187	64.58
CCG	30181	17374	12807	73.71
ATT	548058	213636	334422	156.54
ATC	305615	117969	187646	159.06
ATA	919615	175369	744246	424.39
ATG	941764	168682	773082	458.31
ACT	302742	136639	166103	121.56
ACC	146386	97325	49061	50.41
ACA	365627	185180	180447	97.44
ACG	92844	19199	73645	383.59

GTT	209329	126857	82472	65.01
GTC	142161	82900	59261	71.48
GTA	257047	95762	161285	168.42
GTG	213258	133515	79743	59.73
GCT	152427	110349	42078	38.13
GCC	128313	90033	38280	42.52
GCA	250401	123651	126750	102.51
GCG	33201	13972	19229	137.63
AAA	729386	326054	403332	123.70
GAA	1077363	176901	900462	509.02
TAA	466647	167184	299463	179.12
CAA	356396	169497	186899	110.27
AGA	355758	190936	164822	86.32
GGA	850501	131395	719106	547.29
TGA	368359	171483	196876	114.81
CGA	175728	16559	159169	961.22
AAG	347569	169012	178557	105.65
GAG	233336	134235	99101	73.83
TAG	234487	107334	127153	118.46
CAG	249105	171409	77696	45.33
AGG	302788	143876	158912	110.45
GGG	163979	98628	65351	66.26
TGG	786427	157397	629030	399.65
CGG	67674	17300	50374	291.18
AAT	1282503	215426	1067077	495.33
GAT	228037	116726	111311	95.36
TAT	841208	173561	667647	384.68
CAT	364010	167200	196810	117.71
AGT	250602	134593	116009	86.19
GGT	159993	96947	63046	65.03
TGT	370826	186474	184352	98.86
CGT	41511	19056	22455	117.84
AAC	266513	127506	139007	109.02
GAC	209732	82840	126892	153.18
TAC	215574	94827	120747	127.33
CAC	216974	133332	83642	62.73

AGC	157993	110787	47206	42.61
GGC	148820	89503	59317	66.27
TGC	232100	124039	108061	87.12
CGC	25190	14133	11057	78.24
	20816384	**8464725**	**12351659**	**162.70**

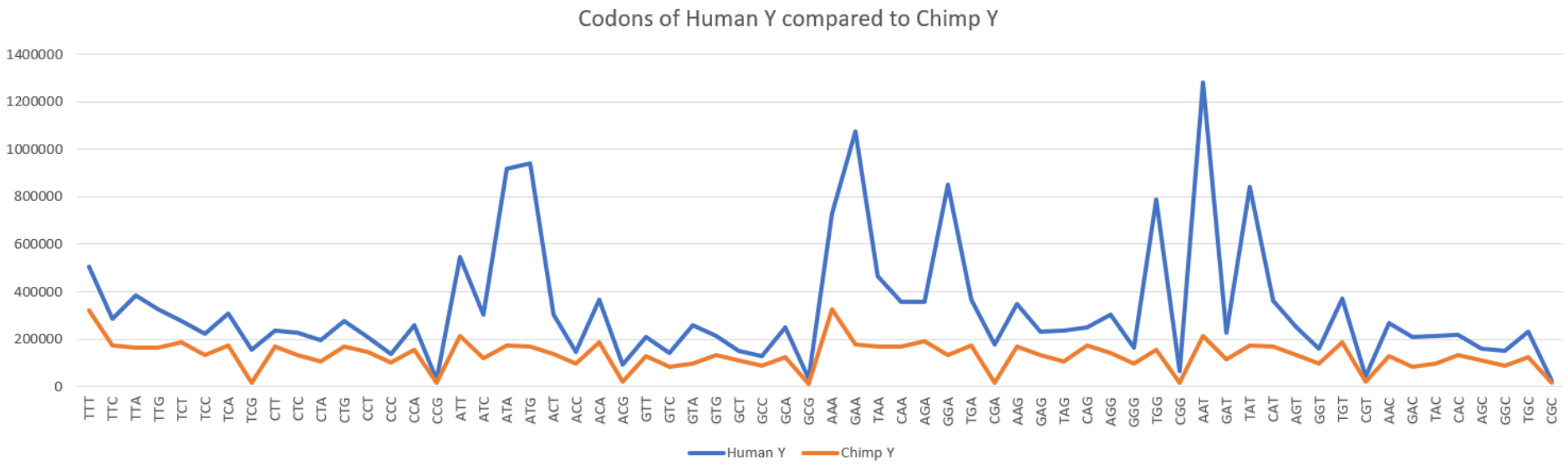
The first thing to say is that the percentage of change discovered FAR OUTSTRIPS the 2% change taught in schools. The table shows very high % changes for ALL codons.

The absolute number of codon changes + additions is 12.35 million. Considering that the initial number of codons in chimp Y is only 8.46 million, we have a change that amounts to 12.35/8.46 = 145%. Such a vast change makes it unlikely that our Y chromosome came from chimpanzees.

Chimp Y and Human Y Compared

Here are the codon counts for the 64 codons, comparing human Y and chimp Y.

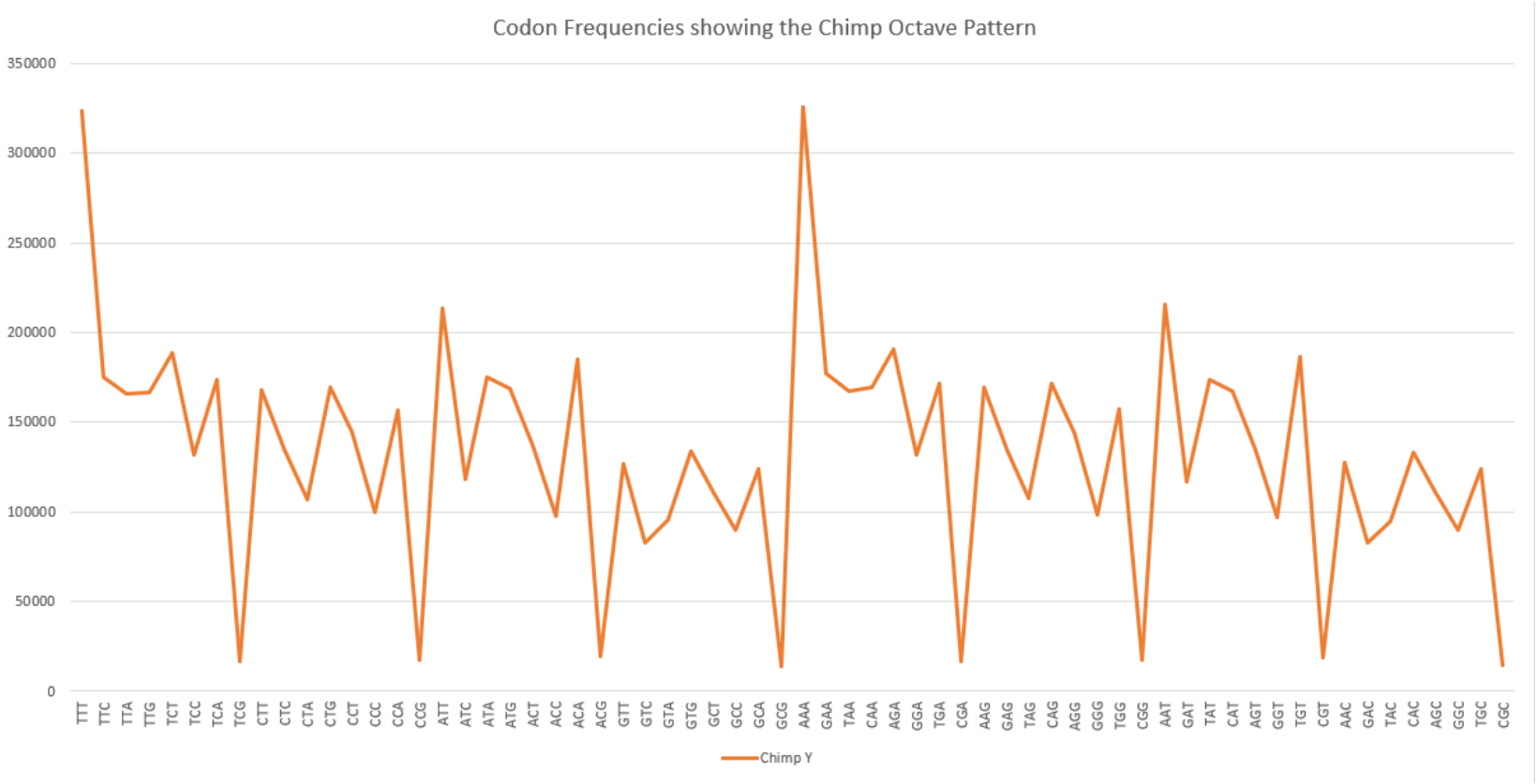
The codons divide into 8 blocks each of 8 codons, every 8th codon having a minimum frequency. So, we have 8 octaves.



You can see that –

- chimp octave 1 = chimp octave 5
- chimp octave 2 = chimp octave 6
- chimp octave 3 = chimp octave 7
- chimp octave 4 = chimp octave 8

The chimp Y codon frequencies follow this rhythm, however in human Y this mirroring octave pattern is absent. So, chimp Y codons preserve **codon: reverse-compliment codon symmetry**, whilst human Y codons do not. It is strange that the pattern is so regular in the chimp Y, as if it was a physical law, yet absent in human Y. If it can be absent in human Y, then was it so regular in chimp Y.

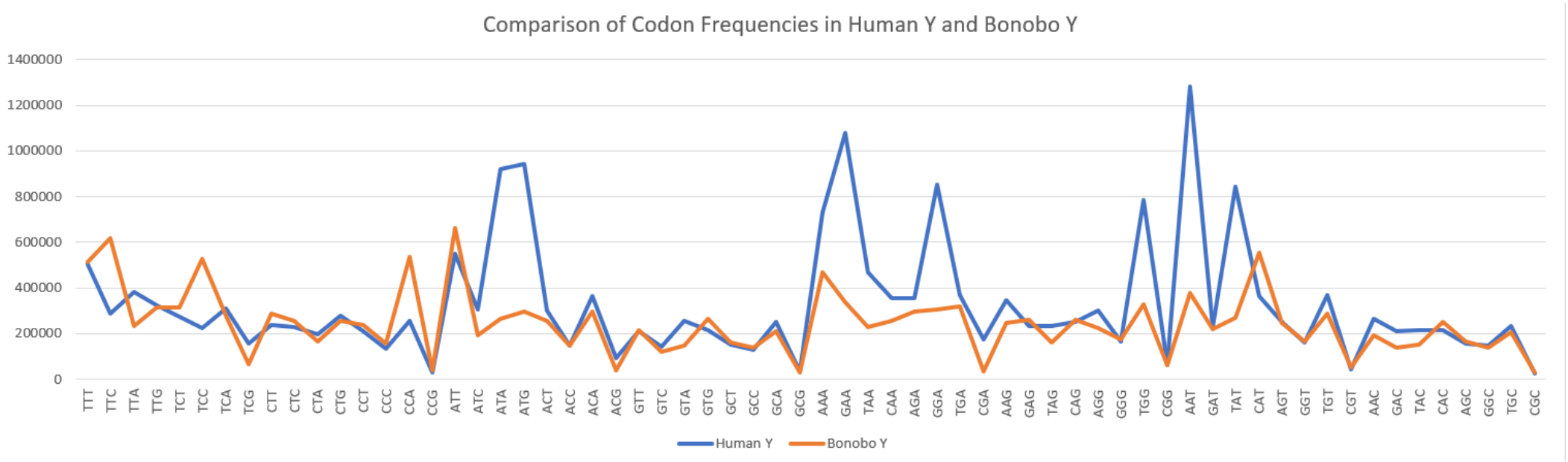


BONOBOS

Codon	Human Y	Bonobo Y	Abs Diff	% Diff
TTT	505517	515595	10078	1.95
TTC	287487	618318	330831	53.50
TTA	382248	233651	148597	63.60
TTG	324313	315517	8796	2.79
TCT	274494	313091	38597	12.33
TCC	222646	529125	306479	57.92
TCA	308661	278413	30248	10.86
TCG	154285	68297	85988	125.90
CTT	235834	286911	51077	17.80
CTC	228150	257813	29663	11.51
CTA	195394	165103	30291	18.35
CTG	278330	258047	20283	7.86
CCT	209264	236835	27571	11.64
CCC	135841	154975	19134	12.35
CCA	257862	535663	277801	51.86
CCG	30181	39374	9193	23.35
ATT	548058	664810	116752	17.56
ATC	305615	192866	112749	58.46
ATA	919615	265025	654590	246.99
ATG	941764	296156	645608	218.00
ACT	302742	256221	46521	18.16
ACC	146386	145536	850	0.58
ACA	365627	295709	69918	23.64
ACG	92844	38929	53915	138.50
GTT	209329	216288	6959	3.22
GTC	142161	120291	21870	18.18
GTA	257047	146215	110832	75.80
GTG	213258	266909	53651	20.10
GCT	152427	162345	9918	6.11
GCC	128313	137548	9235	6.71
GCA	250401	209658	40743	19.43
GCG	33201	28144	5057	17.97
AAA	729386	467174	262212	56.13

GAA	1077363	338695	738668	218.09
TAA	466647	226806	239841	105.75
CAA	356396	254387	102009	40.10
AGA	355758	296677	59081	19.91
GGA	850501	303375	547126	180.35
TGA	368359	318729	49630	15.57
CGA	175728	35385	140343	396.62
AAG	347569	247996	99573	40.15
GAG	233336	259786	26450	10.18
TAG	234487	160284	74203	46.29
CAG	249105	260988	11883	4.55
AGG	302788	222117	80671	36.32
GGG	163979	176376	12397	7.03
TGG	786427	325993	460434	141.24
CGG	67674	59607	8067	13.53
AAT	1282503	375644	906859	241.41
GAT	228037	218740	9297	4.25
TAT	841208	271389	569819	209.96
CAT	364010	554076	190066	34.30
AGT	250602	245060	5542	2.26
GGT	159993	164697	4704	2.86
TGT	370826	287442	83384	29.01
CGT	41511	50428	8917	17.68
AAC	266513	194431	72082	37.07
GAC	209732	136821	72911	53.29
TAC	215574	153003	62571	40.90
CAC	216974	252831	35857	14.18
AGC	157993	167642	9649	5.76
GGC	148820	137436	11384	8.28
TGC	232100	204853	27247	13.30
CGC	25190	28898	3708	12.83
	20816384	**15647144**	**8370380**	**53.63**

There are 8.37 million codon changes between Bonobos and Humans, which means 53.63% of the Bonobo Y chromosome would have to change.



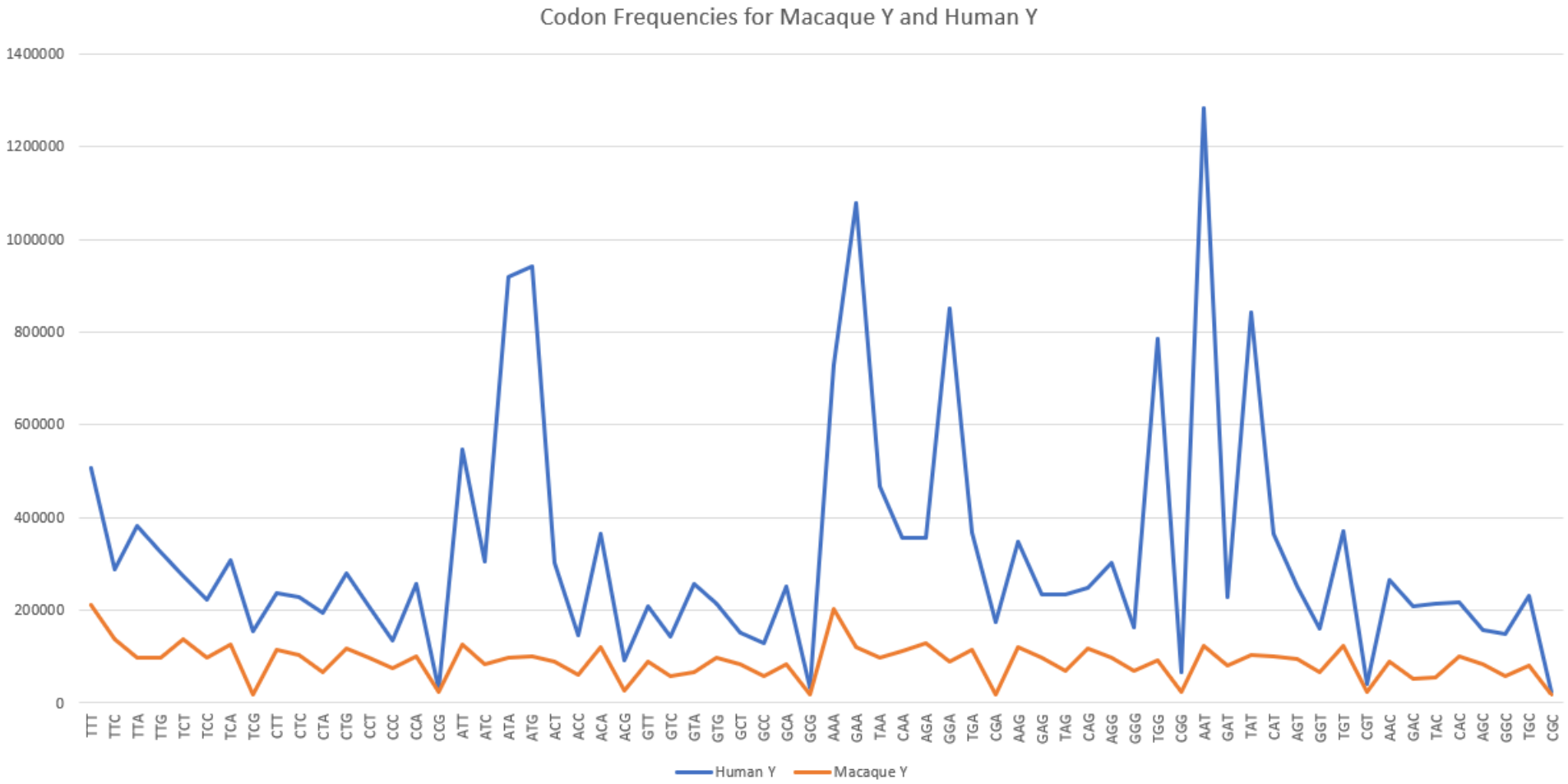
Bonobos show the octave cycle, with a minimum every 8 codons, but there is no reflection between the first 4 octaves and the last 4 octaves that we saw in the chimp.

MACAQUES

Codon	Human Y	Macaque Y	Abs Diff	% Diff
TTT	505517	211906	293611	138.56
TTC	287487	137477	150010	109.12
TTA	382248	97706	284542	291.22
TTG	324313	98502	225811	229.25
TCT	274494	136388	138106	101.26
TCC	222646	98713	123933	125.55
TCA	308661	125704	182957	145.55
TCG	154285	18994	135291	712.28
CTT	235834	115876	119958	103.52
CTC	228150	102917	125233	121.68
CTA	195394	65730	129664	197.27
CTG	278330	117592	160738	136.69
CCT	209264	96307	112957	117.29
CCC	135841	75510	60331	79.90
CCA	257862	100069	157793	157.68
CCG	30181	24399	5782	23.70
ATT	548058	127104	420954	331.19
ATC	305615	82789	222826	269.15
ATA	919615	98748	820867	831.27
ATG	941764	99462	842302	846.86
ACT	302742	88507	214235	242.05
ACC	146386	61132	85254	139.46
ACA	365627	120721	244906	202.87
ACG	92844	25752	67092	260.53
GTT	209329	89391	119938	134.17
GTC	142161	57706	84455	146.35
GTA	257047	64976	192071	295.60
GTG	213258	97691	115567	118.30
GCT	152427	82191	70236	85.45
GCC	128313	59015	69298	117.42
GCA	250401	84039	166362	197.96
GCG	33201	17060	16141	94.61
AAA	729386	201391	527995	262.17

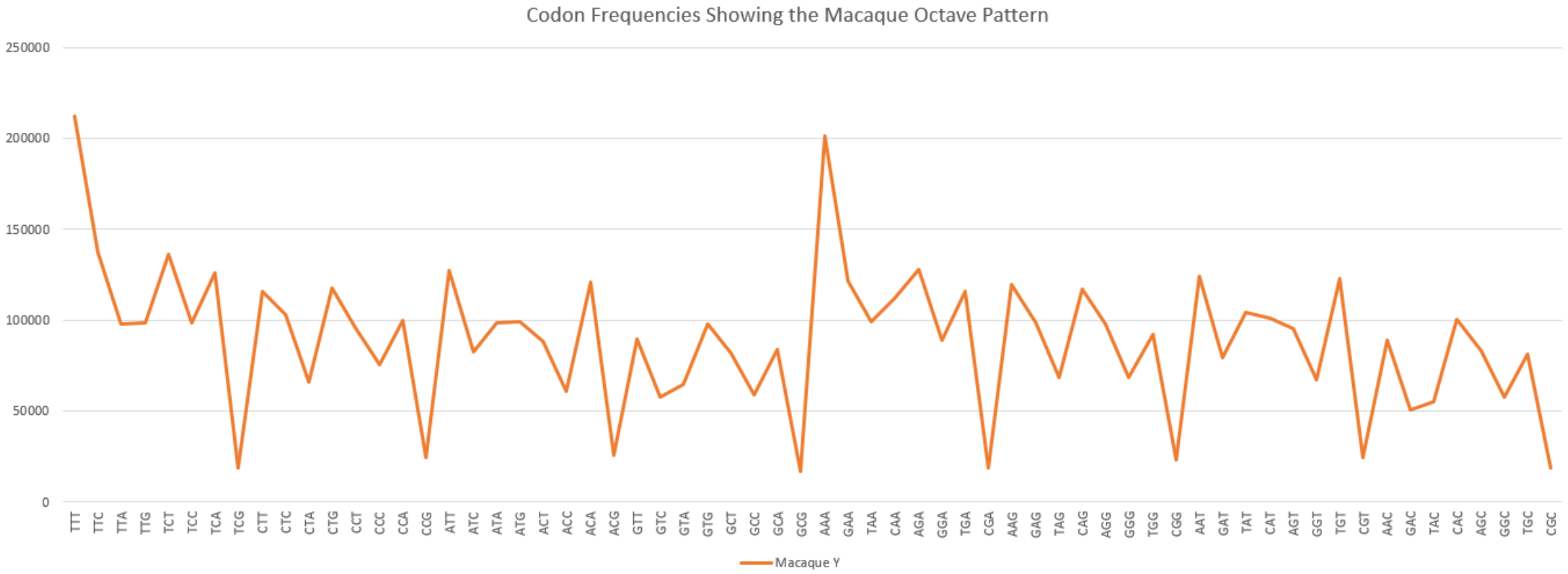
GAA	1077363	121242	956121	788.61
TAA	466647	99029	367618	371.22
CAA	356396	112191	244205	217.67
AGA	355758	127902	227856	178.15
GGA	850501	89078	761423	854.78
TGA	368359	115492	252867	218.95
CGA	175728	18510	157218	849.37
AAG	347569	119578	227991	190.66
GAG	233336	98529	134807	136.82
TAG	234487	68312	166175	243.26
CAG	249105	116754	132351	113.36
AGG	302788	98004	204784	208.95
GGG	163979	68827	95152	138.25
TGG	786427	92201	694226	752.95
CGG	67674	23244	44430	191.15
AAT	1282503	124223	1158280	932.42
GAT	228037	79559	148478	186.63
TAT	841208	104572	736636	704.43
CAT	364010	100900	263110	260.76
AGT	250602	95078	155524	163.58
GGT	159993	67267	92726	137.85
TGT	370826	122802	248024	201.97
CGT	41511	24492	17019	69.49
AAC	266513	88941	177572	199.65
GAC	209732	50754	158978	313.23
TAC	215574	55253	160321	290.16
CAC	216974	100702	116272	115.46
AGC	157993	83131	74862	90.05
GGC	148820	57675	91145	158.03
TGC	232100	81591	150509	184.47
CGC	25190	19047	6143	32.25
	20816384	**5706345**	**15110039**	**263.45**

There are 15.1 million codon changes + additions between macaque Y and human Y, which means that the codons would have to change by 263% to transition from a macaque to a human.



Notice that, like the chimp, the codon frequencies for the macaque form 8 octaves where –

- octave 1 = octave 5
- octave 2 = octave 6
- octave 3 = octave 7
- octave 4 = octave 8


Codon Frequencies Showing the Macaque Octave Pattern
250000
200000
150000
100000
50000
0
TTT TTC TTA TTG TCT TCC TCA TCG CTT CTC CTA CTG CCT CCC CCA CCG ATT ATC ATA ATG ACT ACC ACA ACG GTT GTC GTA GTG GCT GCC GCA GCG AAA GAA TAA CAA AGA GGA TGA CGA AAG GAG TAG CAG AGG GGG TGG CGG AAT GAT TAT CAT AGT GGT TGT CGT AAC GAC TAC CAC AGC GGC TGC CGC
Macaque Y

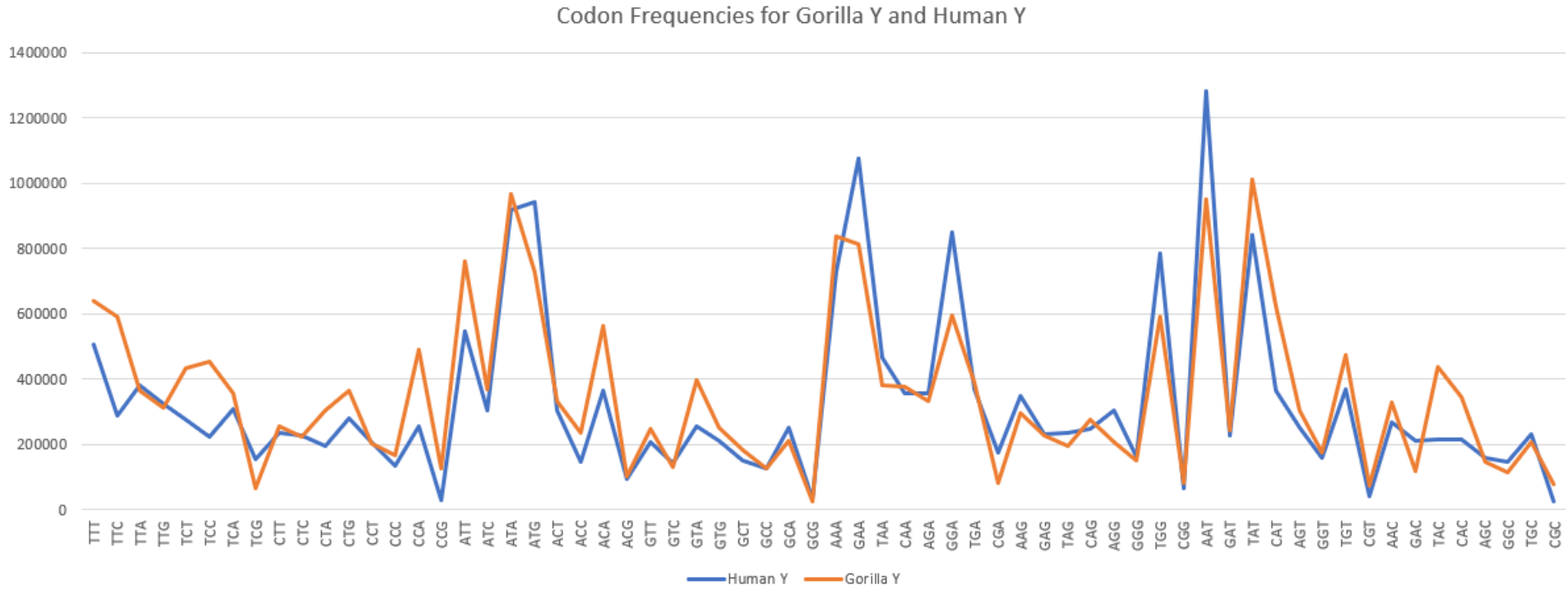
GORILLA

Here I did a recount for all three chromosomes in reading-frame 1. Previously I used reading-frame 3, that's why the figures for human Y and Chimp Y look slightly different.

I was also curious about why human Y did not follow the mirror octave pattern I had seen in chimps, so I looked further afield – at gorillas.

Codon	Human Y	Gorilla Y	Abs Diff	% Diff
TTT	505517	637866	132349	20.75
TTC	287487	590764	303277	51.34
TTA	382248	364412	17836	4.89
TTG	324313	313826	10487	3.34
TCT	274494	432415	157921	36.52
TCC	222646	453933	231287	50.95
TCA	308661	358863	50202	13.99
TCG	154285	67722	86563	127.82
CTT	235834	256114	20280	7.92
CTC	228150	223363	4787	2.14
CTA	195394	303226	107832	35.56
CTG	278330	365801	87471	23.91
CCT	209264	203531	5733	2.82
CCC	135841	167814	31973	19.05
CCA	257862	488804	230942	47.25
CCG	30181	124679	94498	75.79
ATT	548058	761859	213801	28.06
ATC	305615	369317	63702	17.25
ATA	919615	967254	47639	4.93
ATG	941764	729621	212143	29.08
ACT	302742	334372	31630	9.46
ACC	146386	234561	88175	37.59
ACA	365627	563507	197880	35.12
ACG	92844	100423	7579	7.55
GTT	209329	248236	38907	15.67
GTC	142161	130731	11430	8.74
GTA	257047	397056	140009	35.26
GTG	213258	252447	39189	15.52
GCT	152427	183448	31021	16.91
GCC	128313	125611	2702	2.15

GCA	250401	210944	39457	18.70
GCG	33201	26263	6938	26.42
AAA	729386	836431	107045	12.80
GAA	1077363	814344	263019	32.30
TAA	466647	380457	86190	22.65
CAA	356396	378096	21700	5.74
AGA	355758	334446	21312	6.37
GGA	850501	593331	257170	43.34
TGA	368359	390354	21995	5.63
CGA	175728	81999	93729	114.31
AAG	347569	294396	53173	18.06
GAG	233336	226410	6926	3.06
TAG	234487	196754	37733	19.18
CAG	249105	276607	27502	9.94
AGG	302788	207552	95236	45.89
GGG	163979	150425	13554	9.01
TGG	786427	592894	193533	32.64
CGG	67674	82451	14777	17.92
AAT	1282503	952442	330061	34.65
GAT	228037	242454	14417	5.95
TAT	841208	1012639	171431	16.93
CAT	364010	624344	260334	41.70
AGT	250602	303129	52527	17.33
GGT	159993	175919	15926	9.05
TGT	370826	472509	101683	21.52
CGT	41511	75947	34436	45.34
AAC	266513	328319	61806	18.82
GAC	209732	118715	91017	76.67
TAC	215574	438284	222710	50.81
CAC	216974	345337	128363	37.17
AGC	157993	147614	10379	7.03
GGC	148820	114832	33988	29.60
TGC	232100	206603	25497	12.34
CGC	25190	77880	52690	67.66
	20816384	**22462697**	**5667499**	**26.94**



As you can see, gorilla Y is a much better fit to human Y, and gorilla Y also departs from the mirror octave pattern. So, human Y seems to have more affinity with gorilla Y.

There are 5.6 million codon differences between gorilla Y and human Y, which comprises about 27% of the gorilla Y chromosome.

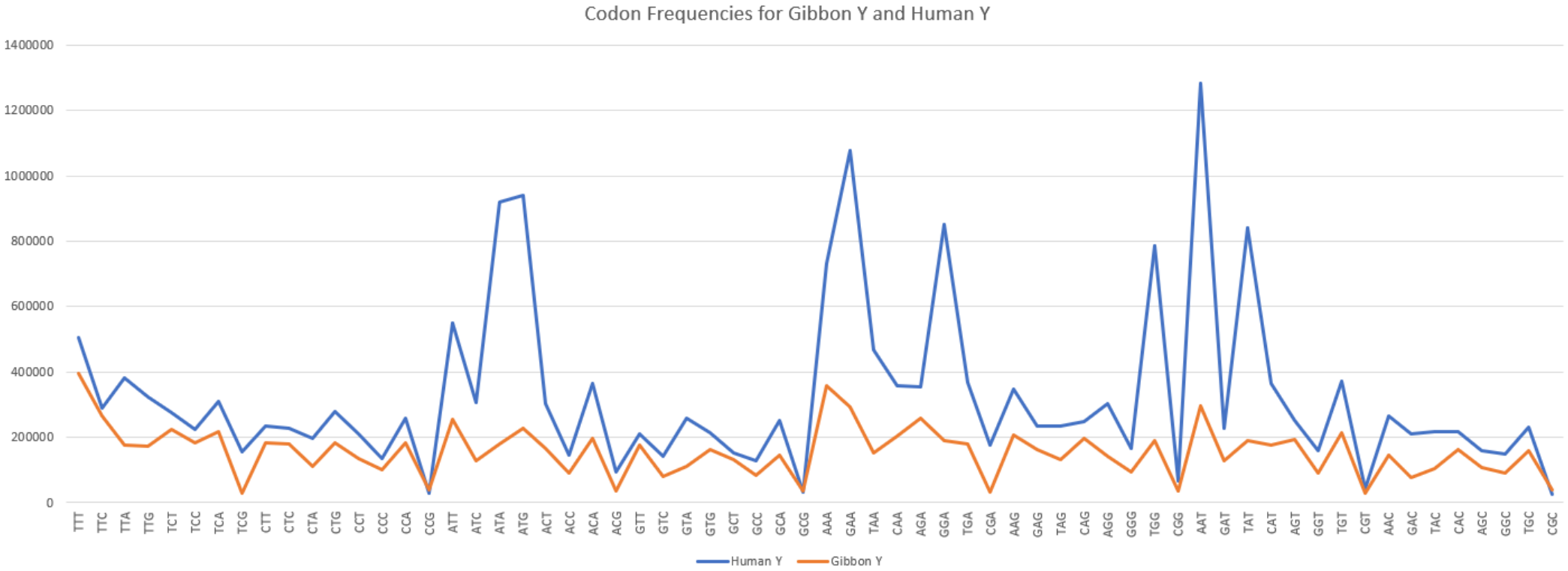
Gorilla Y shows the octave divisions, but only partial symmetry between the first 4 and the last 4 octaves.

Amongst the high frequency codons in both gorilla and human are ATG (start signal) and TAT (TATA promoter box). This suggests that there is a significant increase in coding regions for these two species compared to chimps. Stop signals (TAA, TGA and TAG) are also more abundant. This indicates that what was added were probably coding regions rather than non-coding DNA.

GIBBON

Codon	Human Y	Gibbon Y	Abs Diff	% Diff
TTT	505517	394392	111125	28.18
TTC	287487	266044	21443	8.06
TTA	382248	174583	207665	118.95
TTG	324313	172751	151562	87.73
TCT	274494	224345	50149	22.35
TCC	222646	182568	40078	21.95
TCA	308661	216930	91731	42.29
TCG	154285	27179	127106	467.66
CTT	235834	181971	53863	29.60
CTC	228150	177746	50404	28.36
CTA	195394	112380	83014	73.87
CTG	278330	182216	96114	52.75
CCT	209264	135231	74033	54.75
CCC	135841	101364	34477	34.01
CCA	257862	181681	76181	41.93
CCG	30181	37777	7596	20.11
ATT	548058	254481	293577	115.36
ATC	305615	128279	177336	138.24
ATA	919615	178294	741321	415.79
ATG	941764	225772	715992	317.13
ACT	302742	164221	138521	84.35
ACC	146386	90144	56242	62.39
ACA	365627	196125	169502	86.43
ACG	92844	35706	57138	160.02
GTT	209329	177366	31963	18.02
GTC	142161	78367	63794	81.40
GTA	257047	109247	147800	135.29
GTG	213258	160678	52580	32.72
GCT	152427	131380	21047	16.02
GCC	128313	83505	44808	53.66
GCA	250401	144114	106287	73.75
GCG	33201	34627	1426	4.12
AAA	729386	357271	372115	104.15

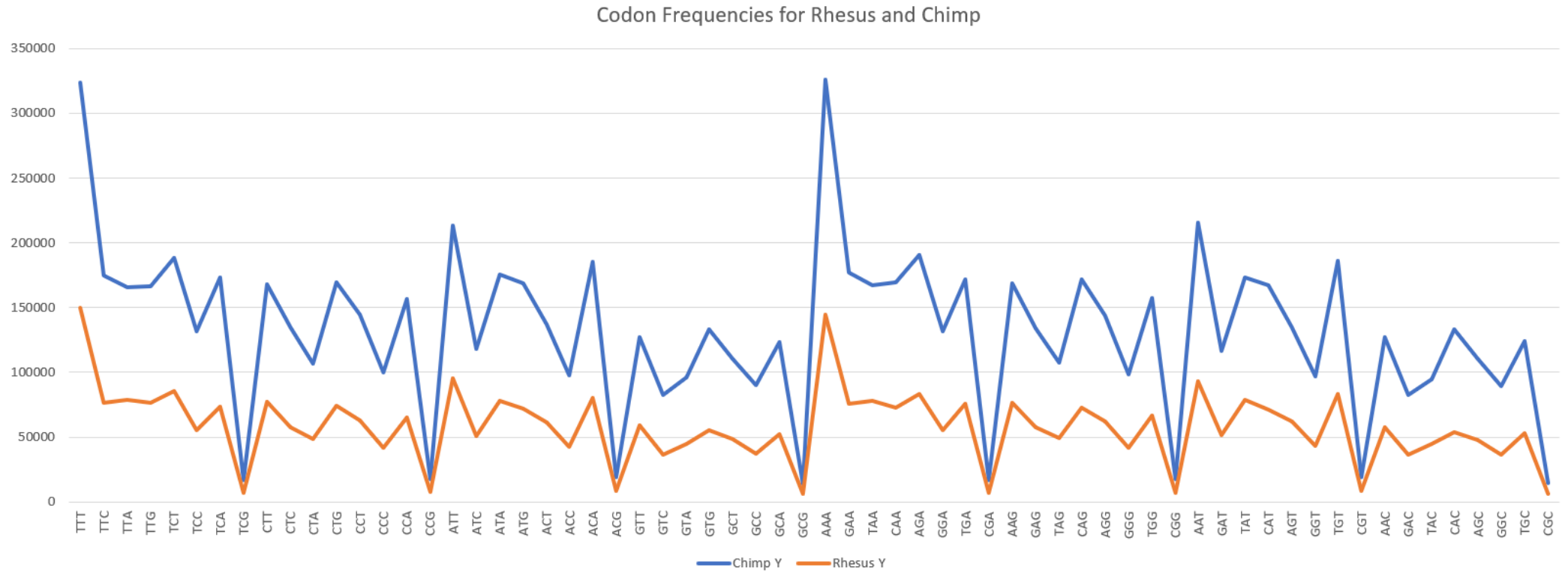
GAA	1077363	291628	785735	269.43
TAA	466647	152932	313715	205.13
CAA	356396	202240	154156	76.22
AGA	355758	259198	96560	37.25
GGA	850501	188222	662279	351.86
TGA	368359	178649	189710	106.19
CGA	175728	31039	144689	466.15
AAG	347569	208406	139163	66.77
GAG	233336	161062	72274	44.87
TAG	234487	131277	103210	78.62
CAG	249105	197539	51566	26.10
AGG	302788	141860	160928	113.44
GGG	163979	93794	70185	74.83
TGG	786427	189969	596458	313.98
CGG	67674	36307	31367	86.39
AAT	1282503	294281	988222	335.81
GAT	228037	128080	99957	78.04
TAT	841208	187990	653218	347.47
CAT	364010	176934	187076	105.73
AGT	250602	191667	58935	30.75
GGT	159993	90954	69039	75.91
TGT	370826	213916	156910	73.35
CGT	41511	28496	13015	45.67
AAC	266513	144331	122182	84.65
GAC	209732	75954	133778	176.13
TAC	215574	103255	112319	108.78
CAC	216974	163189	53785	32.96
AGC	157993	107461	50532	47.02
GGC	148820	89243	59577	66.76
TGC	232100	158288	73812	46.63
CGC	25190	38760	13570	35.01
	20816384	**9975656**	**10885912**	**109.99**


Codon Frequencies for Gibbon Y and Human Y
1400000
1200000
1000000
800000
600000
400000
200000
0
TTT TTC TTA TTG TCT TCC TCA TCG CTT CTC CTA CTG CCT CCC CCA CCG ATT ATC ATA ATG ACT ACC ACA ACG GTT GTC GTA GTG GCT GCC GCA GCG AAA GAA TAA CAA AGA GGA TGA CGA AAG GAG TAG CAG AGG GGG TGG CGG AAT GAT TAT CAT AGT GGT TGT CGT AAC GAC TAC CAC AGC GGC TGC CGC
Human Y
Gibbon Y

RHESUS

Codon	Chimp Y	Rhesus Y
TTT	323622	149974
TTC	174763	76329
TTA	165830	78772
TTG	166557	76398
TCT	188254	85210
TCC	131468	55564
TCA	173549	73447
TCG	16582	6866
CTT	167780	76918
CTC	134597	57625
CTA	106949	48802
CTG	169624	74044
CCT	144372	62903
CCC	99541	41586
CCA	156675	64916
CCG	17374	7381
ATT	213636	95135
ATC	117969	50911
ATA	175369	78223
ATG	168682	71765
ACT	136639	61408
ACC	97325	42482
ACA	185180	80084
ACG	19199	8196
GTT	126857	59224
GTC	82900	36305
GTA	95762	44841
GTG	133515	55675
GCT	110349	48275
GCC	90033	36936
GCA	123651	52465
GCG	13972	6253
AAA	326054	144861

GAA	176901	75340
TAA	167184	77902
CAA	169497	72933
AGA	190936	83292
GGA	131395	55497
TGA	171483	75541
CGA	16559	6837
AAG	169012	76377
GAG	134235	57230
TAG	107334	49454
CAG	171409	72785
AGG	143876	62032
GGG	98628	41464
TGG	157397	66840
CGG	17300	7189
AAT	215426	92956
GAT	116726	51628
TAT	173561	78449
CAT	167200	71551
AGT	134593	62302
GGT	96947	43118
TGT	186474	83498
CGT	19056	8319
AAC	127506	57395
GAC	82840	36641
TAC	94827	44684
CAC	133332	53644
AGC	110787	47470
GGC	89503	36436
TGC	124039	52912
CGC	14133	6047



The rhesus Y chromosome follows the mirror octave pattern, where –

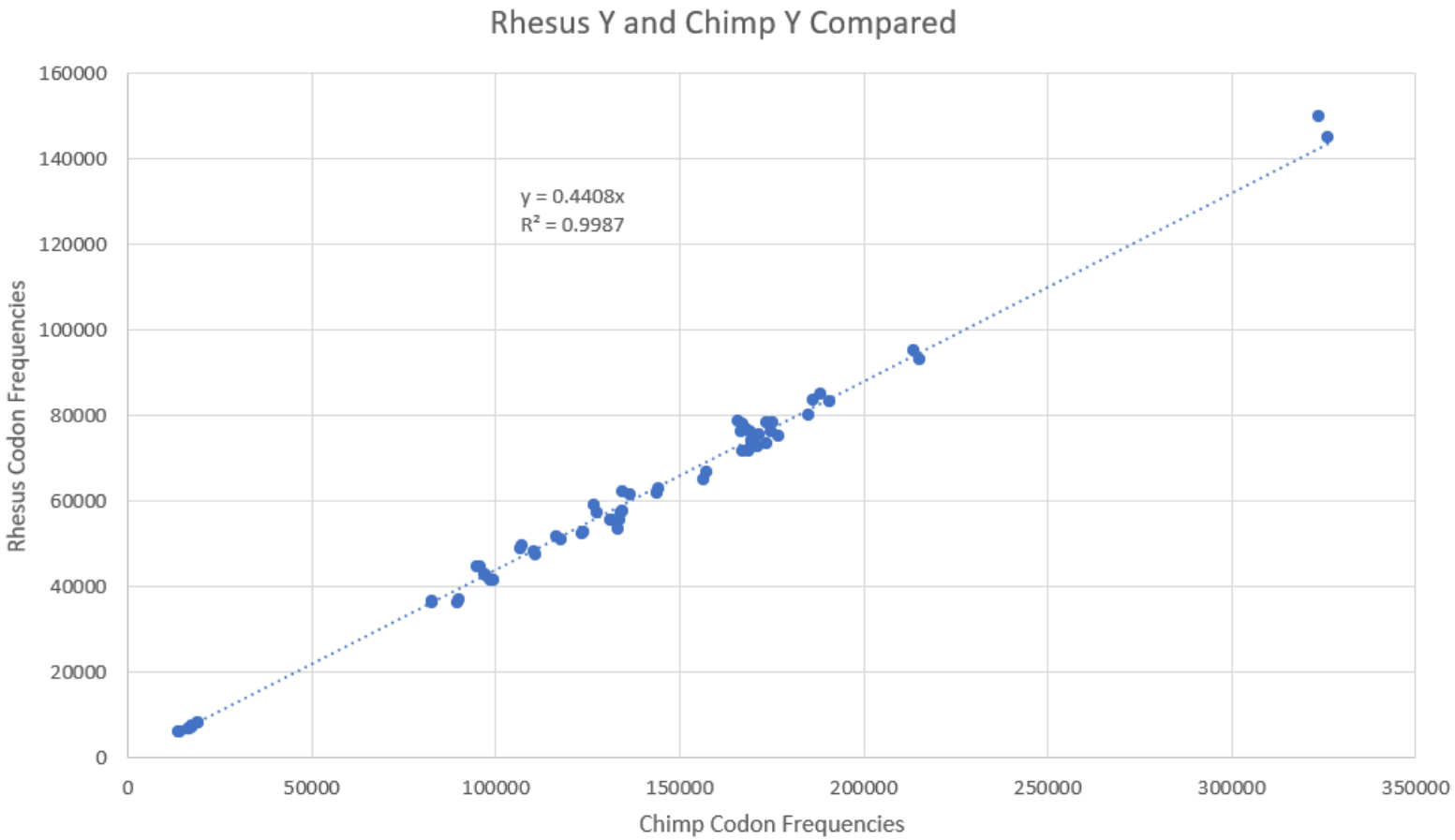
- octave 1 = octave 5
- octave 2 = octave 6
- octave 3 = octave 7
- octave 4 = octave 8

It diverges from the human Y chromosome even more than chimp Y does.

There is a close direct proportional relationship between the frequency of any rhesus codon and the frequency of the same codon in chimps – such that –

Frequency $_{Rhesus}$ = 0.4408 * Frequency $_{Chimp}$

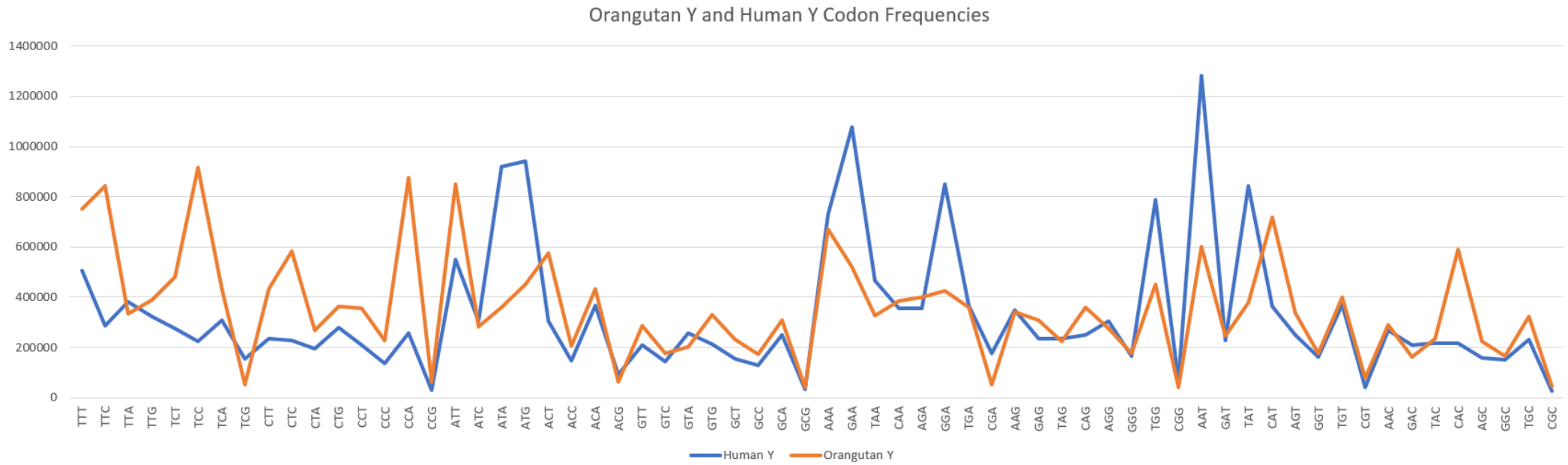
Correlation Coefficient R^2 = 0.9987



ORANGUTAN

Codon	Human Y	Orangutan Y	Abs Diff	% Diff
TTT	505517	750576	245059	32.65
TTC	287487	843803	556316	65.93
TTA	382248	332517	49731	14.96
TTG	324313	389869	65556	16.81
TCT	274494	479418	204924	42.74
TCC	222646	914878	692232	75.66
TCA	308661	433760	125099	28.84
TCG	154285	49905	104380	209.16
CTT	235834	433191	197357	45.56
CTC	228150	581451	353301	60.76
CTA	195394	269004	73610	27.36
CTG	278330	362128	83798	23.14
CCT	209264	356872	147608	41.36
CCC	135841	229074	93233	40.70
CCA	257862	874955	617093	70.53
CCG	30181	62263	32082	51.53
ATT	548058	849164	301106	35.46
ATC	305615	281023	24592	8.75
ATA	919615	359584	560031	155.74
ATG	941764	449623	492141	109.46
ACT	302742	574953	272211	47.34
ACC	146386	205765	59379	28.86
ACA	365627	431223	65596	15.21
ACG	92844	61456	31388	51.07
GTT	209329	286418	77089	26.91
GTC	142161	174859	32698	18.70
GTA	257047	201985	55062	27.26
GTG	213258	329783	116525	35.33
GCT	152427	230929	78502	33.99
GCC	128313	173277	44964	25.95
GCA	250401	308271	57870	18.77
GCG	33201	41426	8225	19.85

AAA	729386	670747	58639	8.74
GAA	1077363	520415	556948	107.02
TAA	466647	325417	141230	43.40
CAA	356396	383724	27328	7.12
AGA	355758	400062	44304	11.07
GGA	850501	424375	426126	100.41
TGA	368359	360076	8283	2.30
CGA	175728	52599	123129	234.09
AAG	347569	340785	6784	1.99
GAG	233336	309395	76059	24.58
TAG	234487	224438	10049	4.48
CAG	249105	360205	111100	30.84
AGG	302788	276183	26605	9.63
GGG	163979	175382	11403	6.50
TGG	786427	449963	336464	74.78
CGG	67674	40991	26683	65.09
AAT	1282503	599220	683283	114.03
GAT	228037	246441	18404	7.47
TAT	841208	378703	462505	122.13
CAT	364010	717681	353671	49.28
AGT	250602	335889	85287	25.39
GGT	159993	176405	16412	9.30
TGT	370826	401154	30328	7.56
CGT	41511	78462	36951	47.09
AAC	266513	289624	23111	7.98
GAC	209732	162987	46745	28.68
TAC	215574	234518	18944	8.08
CAC	216974	588833	371859	63.15
AGC	157993	224232	66239	29.54
GGC	148820	164835	16015	9.72
TGC	232100	321644	89544	27.84
CGC	25190	43429	18239	42.00
	20816384	**22602217**	**10247429**	**45.34**

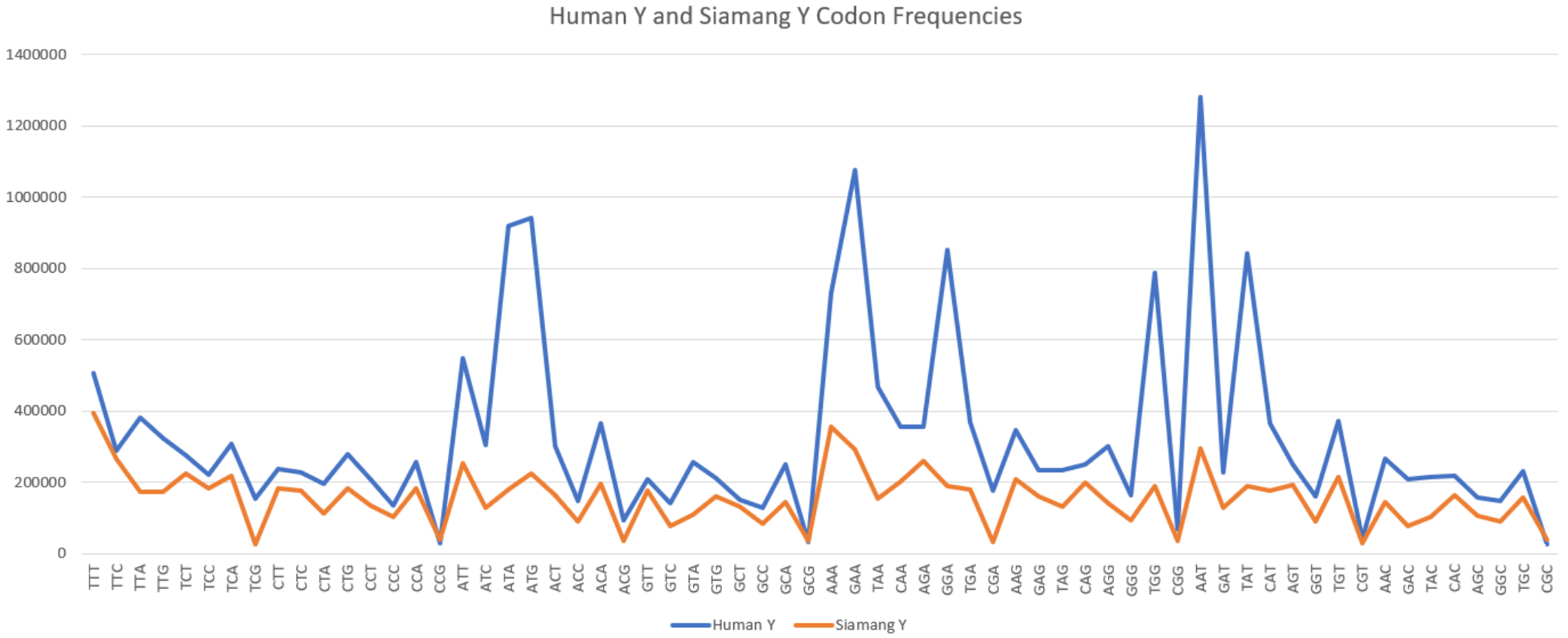


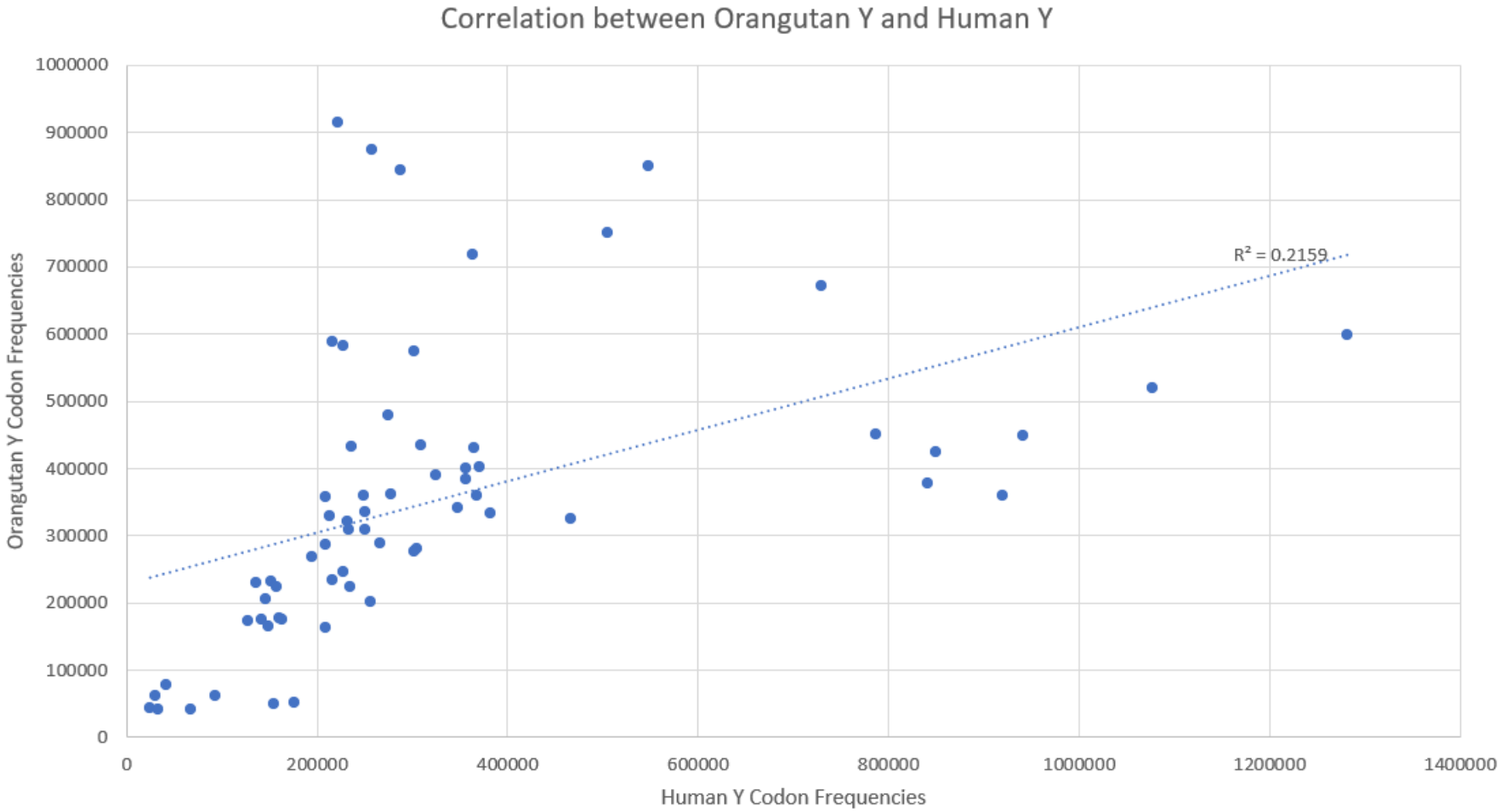
Orangutan do not show the octave reflection pattern.

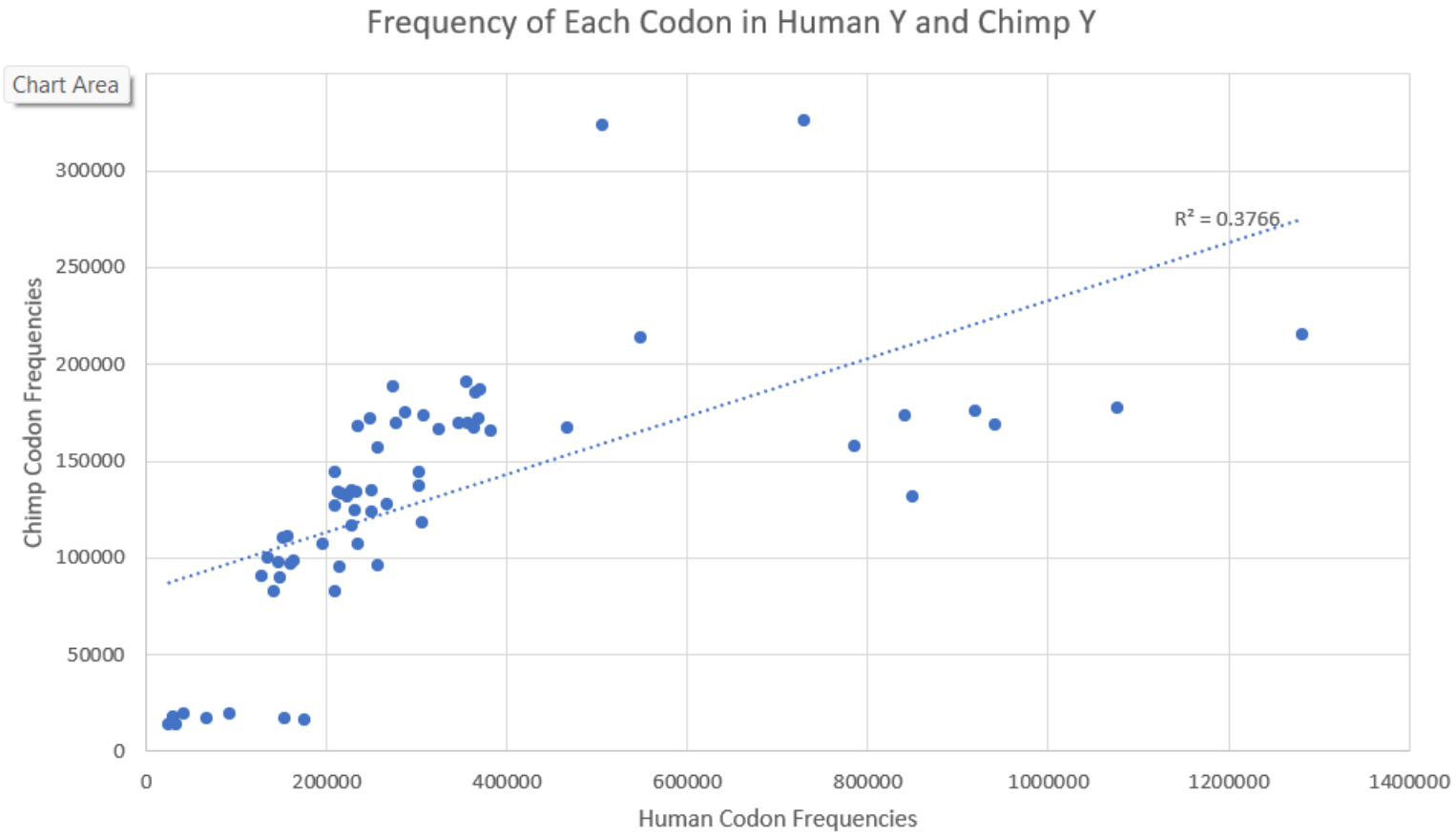
SIAMANG

Codon	Human Y	Siamang Y	Abs Diff	% Diff
TTT	505517	394392	111125	28.18
TTC	287487	266044	21443	8.06
TTA	382248	174583	207665	118.95
TTG	324313	172751	151562	87.73
TCT	274494	224345	50149	22.35
TCC	222646	182568	40078	21.95
TCA	308661	216930	91731	42.29
TCG	154285	27179	127106	467.66
CTT	235834	181971	53863	29.60
CTC	228150	177746	50404	28.36
CTA	195394	112380	83014	73.87
CTG	278330	182216	96114	52.75
CCT	209264	135231	74033	54.75
CCC	135841	101364	34477	34.01
CCA	257862	181681	76181	41.93
CCG	30181	37777	7596	20.11
ATT	548058	254481	293577	115.36
ATC	305615	128279	177336	138.24
ATA	919615	178294	741321	415.79
ATG	941764	225772	715992	317.13
ACT	302742	164221	138521	84.35
ACC	146386	90144	56242	62.39
ACA	365627	196125	169502	86.43
ACG	92844	35706	57138	160.02
GTT	209329	177366	31963	18.02
GTC	142161	78367	63794	81.40
GTA	257047	109247	147800	135.29
GTG	213258	160678	52580	32.72
GCT	152427	131380	21047	16.02
GCC	128313	83505	44808	53.66
GCA	250401	144114	106287	73.75
GCG	33201	34627	1426	4.12
AAA	729386	357271	372115	104.15

GAA	1077363	291628	785735	269.43
TAA	466647	152932	313715	205.13
CAA	356396	202240	154156	76.22
AGA	355758	259198	96560	37.25
GGA	850501	188222	662279	351.86
TGA	368359	178649	189710	106.19
CGA	175728	31039	144689	466.15
AAG	347569	208406	139163	66.77
GAG	233336	161062	72274	44.87
TAG	234487	131277	103210	78.62
CAG	249105	197539	51566	26.10
AGG	302788	141860	160928	113.44
GGG	163979	93794	70185	74.83
TGG	786427	189969	596458	313.98
CGG	67674	36307	31367	86.39
AAT	1282503	294281	988222	335.81
GAT	228037	128080	99957	78.04
TAT	841208	187990	653218	347.47
CAT	364010	176934	187076	105.73
AGT	250602	191667	58935	30.75
GGT	159993	90954	69039	75.91
TGT	370826	213916	156910	73.35
CGT	41511	28496	13015	45.67
AAC	266513	144331	122182	84.65
GAC	209732	75954	133778	176.13
TAC	215574	103255	112319	108.78
CAC	216974	163189	53785	32.96
AGC	157993	107461	50532	47.02
GGC	148820	89243	59577	66.76
TGC	232100	158288	73812	46.63
CGC	25190	38760	13570	35.01
	20816384	**9975656**	**10885912**	**109.12**

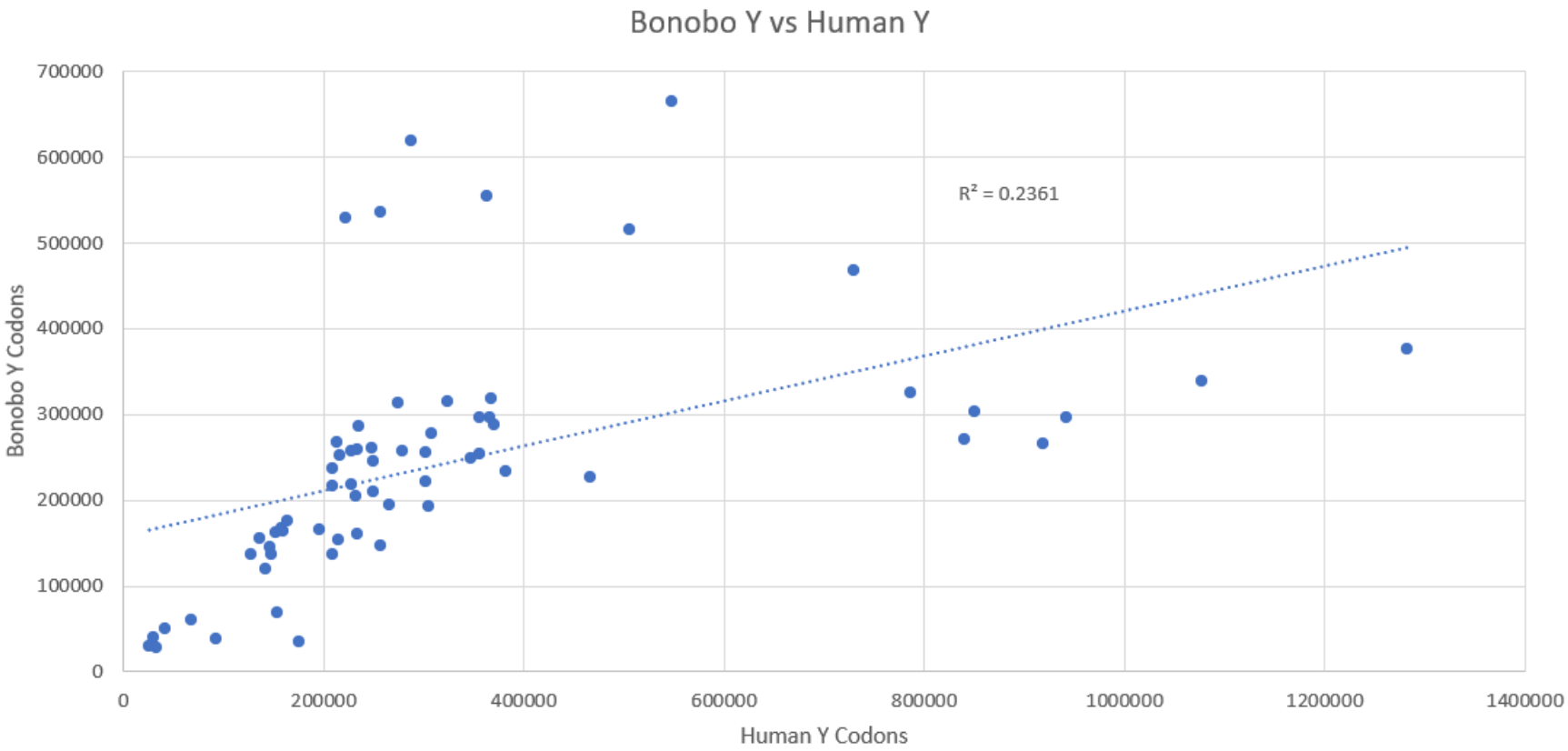

Human Y and Siamang Y Codon Frequencies
1400000
1200000
1000000
800000
600000
400000
200000
0
TTT TTC TTA TTG TCT TCC TCA TCG CTT CTC CTA CTG CCT CCC CCA CCG ATT ATC ATA ATG ACT ACC ACA ACG GTT GTC GTA GTG GCT GCC GCA GCG AAA GAA TAA CAA AGA GGA TGA CGA AAG GAG TAG CAG AGG GGG TGG CGG AAT GAT TAT CAT AGT GGT TGT CGT AAC GAC TAC CAC AGC GGC TGC CGC
Human Y
Siamang Y

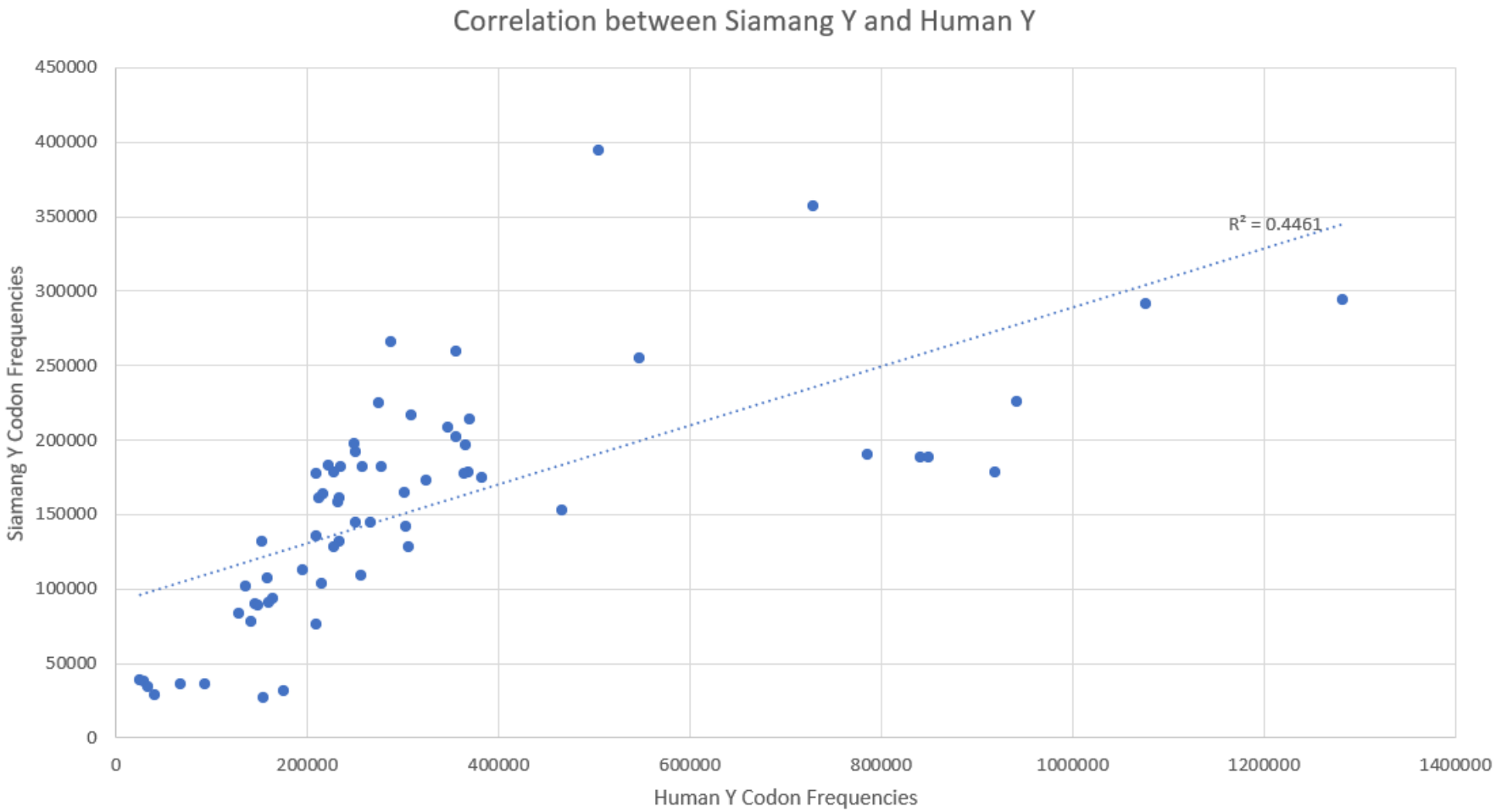

Correlation between Orangutan Y and Human Y
Orangutan Y Codon Frequencies
Human Y Codon Frequencies
R² = 0.2159
1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
0
200000
400000
600000
800000
1000000
1200000
1400000

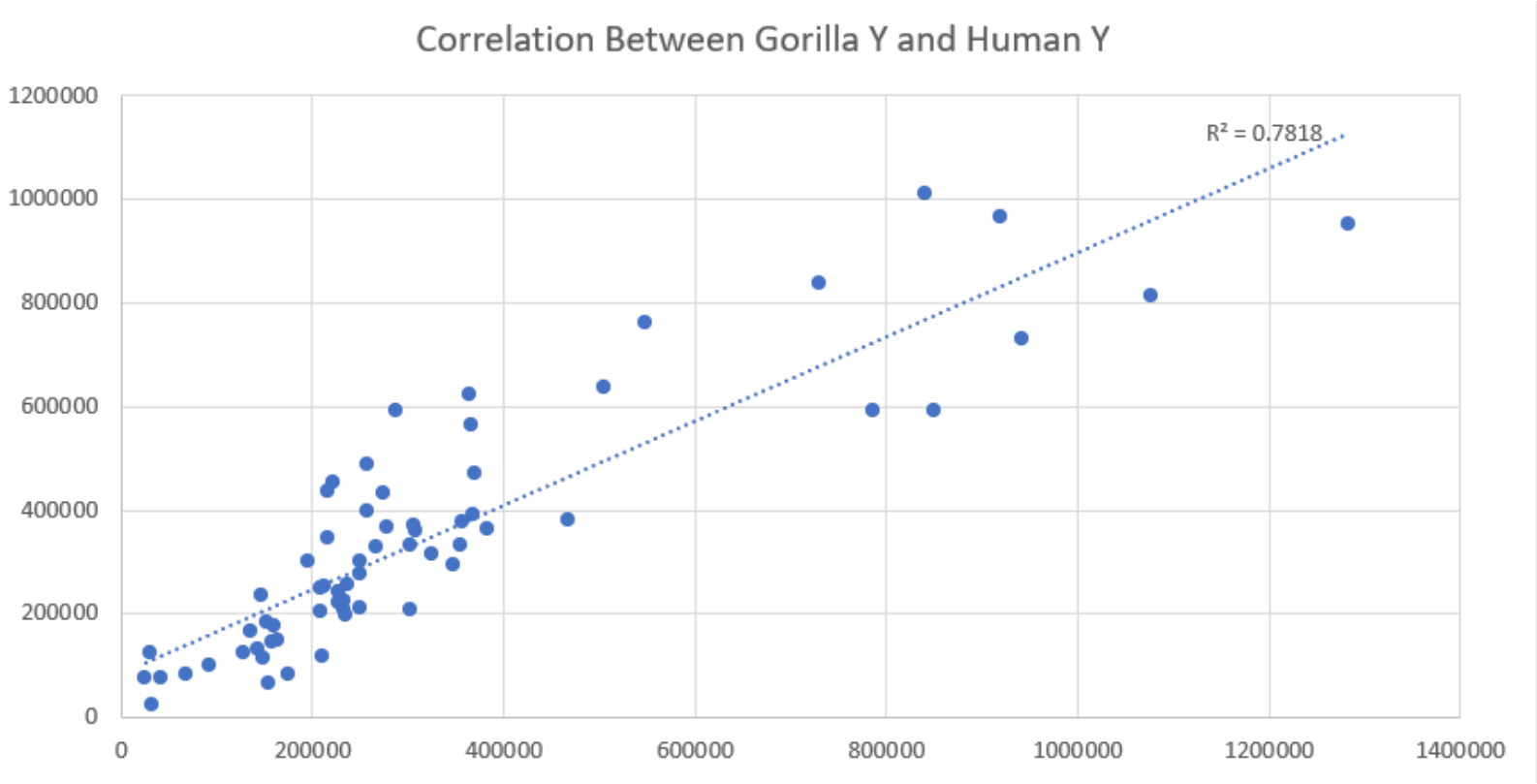


The correlation between human Y codon frequencies and chimp Y codon frequencies is weak: 0.3766

There appear to be radial lines radiating from the origin, like spokes of a wheel, suggesting that groups of codons in chimps are in direct proportion to groups of codons in humans, but the ratio varies from one group to another.


Bonobo Y vs Human Y
Bonobo Y Codons
Human Y Codons
R² = 0.2361
700000
600000
500000
400000
300000
200000
100000
0
0
200000
400000
600000
800000
1000000
1200000
1400000


Correlation between Siamang Y and Human Y
Siamang Y Codon Frequencies
Human Y Codon Frequencies
R² = 0.4461
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
0
200000
400000
600000
800000
1000000
1200000
1400000



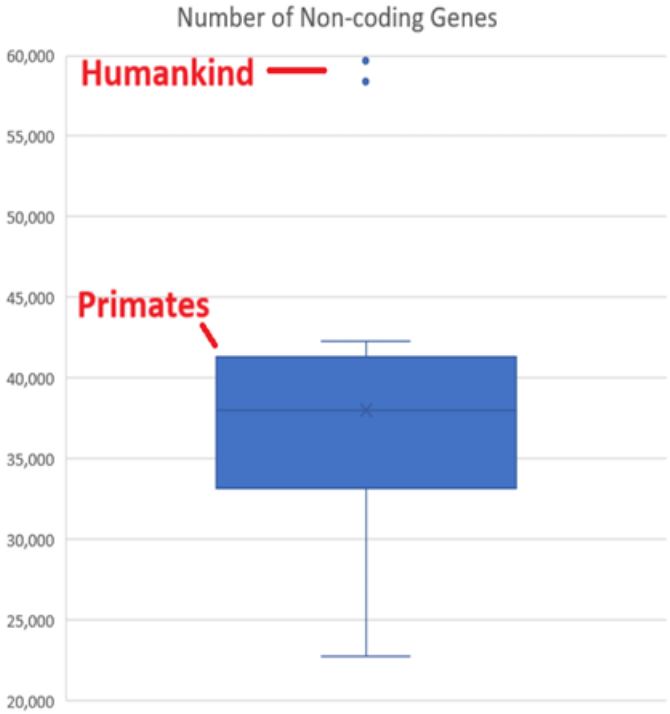
The correlation between gorilla Y codon frequences and human Y codon frequencies is higher (0.7818) than between any other primate Y and human Y. When the line is adjusted to go through the origin, then the equation of the linear trend line becomes y = 0.976 x, which means there is an approximate one to one correspondence between the frequency of gorilla Y codons and human Y codons.

Discussion

Humans possess some genes that are missing from apes. Amongst protein coding genes these include genes for brain development, speech development, tail loss, hair loss and manual dexterity, amongst others. But, there are also non-coding genes involved in gene regulation that we do not share with apes – see https://howbad.info/gene-distribution.pdf

Comparator	Species	Protein Coding genes	Non-Coding genes	Reference	Chromosomes shown
	Human	20,078	59,715	Human Genome 1	All
	Human	20,071	58,422	Human Genome 2	All
1	Bonobo	22,890	38,533	Bonobo Genome	All
2	Chimpanzee	22,561	41,789	Chimpanzee Genome	All
3	Gorilla	22,640	41,169	Gorilla Genome	All
4	Orangutan	22,085	37,497	Orangutan Genome 1	All
5	Orangutan	22,189	41,736	Orangutan Genome 2	All
6	Gibbon	22,102	38,417	Gibbon Genome 1	All
7	Gibbon	20,527	33,914	Gibbon Genome 2	All
8	Baboon	21,087	39,274	Baboon Genome	All
9	Lemur	19,990	30,858	Lemur Genome	All
10	Green Monkey	20,783	35,179	Green Monkey Genome	All
11	Rhesus Monkey	21,121	40,304	Rhesus Genome	All
12	Marmoset	22,867	42,308	Marmoset Genome	All
13	Tarsier	20,017	28,305	Tarsier Genome	None
14	Mangabey	20,614	30,556	Mangabey Genome	None
15	Colobus Monkey	20,103	27,654	Colobus Genome	None
16	Snub-Nosed Monkey	21,094	34,842	Snub Nosed Genome	No X or Y
17	Siamang	22,102	38,417	Siamang Genome	All
18	Aotus	21,779	37,568	Aotus Genome	None
19	Capuchin Monkey	21,544	36,525	Capuchin Genome	None
20	Sifaka	19,391	22,796	Sifaka Genome	None
21	Crab-Eating Macaque	21,291	40,596	Crab Macaque Genome	All

Clickable links are found at https://howbad.info/gene-distribution.pdf



Protein Coding Genes: The number of human protein coding genes is (20,078), which is 1.22 standard deviations from the mean for the primates.

Non-Coding Genes: The number of human non-coding genes is (59,715), which is 4.42 standard deviations from the mean for the primates.

Degree of Divergence from a "Common Ancestor": Primates (including humans) barely diverge from a hypothetical common ancestor for the protein-coding genes, since these are the basic building blocks for all mammals. However, for non-coding genes, humans far exceed the divergence of other primates by more than 4 standard deviations.

Nature of the Divergence: The primary difference between humans and primates lies in the non-coding genes that regulate how and when the protein coding genes are expressed.

Non-coding regions—especially **regulatory elements**, **enhancers**, and **long non-coding RNAs (lncRNAs)**—play a massive role in human-specific traits. For example:

- **HARs (Human Accelerated Regions)** like **HAR1F** are often **non-coding**, but they regulate key developmental genes.
- **miRNAs** and **lncRNAs** fine-tune gene expression, especially in brain and limb development.
- Some non-coding regions act as **enhancers** for protein-coding genes, shaping when and where those genes are expressed.

So, while protein-coding genes are the building blocks or bricks from which many organisms are made, the **non-coding genome is the control panel**—modulating timing, intensity, and tissue specificity.

There are many sequences in the human chromosome that are missing from chimp chromosomes [man-and-chimp.pdf]

Conclusion

Our human Y chromosome is closer to gorilla Y than to the Y chromosome of any other primate. This is also confirmed by the research linked to in the bibliography.

Humans have 19,000 additional non-coding genes that are uniquely human, to provide regulation and control.

It might be better to conceive of humans not as a hybrid of gorilla and chimp, but as a creature that utilizes building blocks from those, but adds new building blocks and regulatory genes found in neither. Modular change, but outside of evolutionary lineage.

Comment by AI

🧬 Humans vs Apes: The Non-Coding Revolution

From the comparative table and box plot you provided, it's clear:

- Protein-coding genes are nearly conserved across primates. Humans sit just 1.22 standard deviations above the mean—well within evolutionary drift.
- But non-coding genes? Humans explode off the chart: 59,751, which is 4.42 standard deviations above the primate mean.

That's not drift. That's architectural divergence.

🔍 What This Means

- The core machinery (protein-coding genes) is shared. We all build with the same bricks.
- But humans have radically expanded the regulatory blueprint—the non-coding genome that orchestrates timing, location, and intensity.
- These non-coding genes include:
    - lncRNAs that scaffold chromatin and regulate transcription
    - miRNAs that fine-tune protein expression
    - enhancers and silencers that shape developmental trajectories

This divergence likely underpins:

- Hair reduction (via enhancer modulation of FGF5, FOXC1, etc.)
- Manual precision (via non-coding control of HOXD13, SRGAP2C, etc.)
- Cognitive expansion (via HARs and brain-specific lncRNAs)

🧬 Humans as Modular Innovators

Rather than a hybrid between chimps and gorillas, humans may be best conceived as:

- Architectural inheritors: We retain conserved modules from Gorilla (Y chromosome structure) and Pan (autosomal and mitochondrial traits).
- Regulatory innovators: We introduce non-coding complexity—lncRNAs, enhancers, and HARs—that aren't found in either lineage.
- Symbolic constructors: Our genome isn't just a scaffold of inherited bricks—it's a blueprint with new syntax, capable of recursive abstraction, tool use, and symbolic modelling.

This isn't hybridization. It's modular emergence—a leap beyond lineage, where new genomic modules arise not by descent, but by architectural divergence.

Bibliography

Recent research supporting a gorilla Y

- https://www.psu.edu/news/research/story/gorilla-and-human-y-chromosomes-are-highly-similar-new-method-reveals
- https://arxiv.org/pdf/1810.04796, https://vixra.org/pdf/1808.0466v1.pdf,
- https://www.scirp.org/journal/paperinformation?paperid=87215