Creation Research Society Quarterly 2020 57:123–132

The Distribution of Genetic Variants in the Human Genome Reflects Created Diversity

Marshall Jordan*

Abstract

Those who believe the record of Genesis account for genetic differences by two mechanisms: created diversity and mutations. For the Y chromosome, mutations appear to be sufficient to explain worldwide Y chromosome diversity. Also, for rare autosomal variants, mutations appear to explain their origins within a few thousand years. However, for common autosomal variants, creationists have proposed that common genetic variants represent created diversity which God put into Adam's genome. Evolutionists oppose this view, insisting that all genetic variation is due to mutations accumulated over millions of years. To test these opposing interpretations of the origin of genetic variation, the pattern of distribution of genetic variants across the human genome was determined to see whether the variants on Y chromosome and the autosomes have different mechanisms of origin. This investigation finds that the concentration of common genetic variants found on the Y chromosomes in the 1000 Genome database is 183 per million bases versus a mean of 2,958 per million bases found on the autosomes, indicating that the Y chromosomes lack created variants. The common variants on the X chromosome fall in an intermediate position of 1,890 per million bases, an indication that some created variants are carried on the X chromosome. Consistent with a mutational origin for all of the Y chromosome common variants and a created origin for most of the autosomal common variants, these findings provide an independent line of evidence supporting recent created diversity as the explanation for most human genetic variation.

Introduction

In the 1000 Genome Project database (1000 Genomes Project Consortium, 2015), the typical human genome differs from the reference genome at 4.1 to 5 million positions. The vast majority of these are common genetic variants (CGVs), which have minor allele frequencies (AF) of 5% to 50%. Rare variants (AF < 0.5%) are 40,000 to 200,000 per genome. Have these genetic variants all come from mutations accumulated over millions of years, or are they explained by created diversity and recent mutation accumulation? Analysis of the 1000 Genome Project database

* Marshall C Jordan, MD, FACS, Albuquerque, NM, marshalljordanmd@gmail.com
Accepted for publication July 15, 2020

was performed in an attempt to answer this question.

Creationists have proposed that most human traits, and the CGVs which cause them, are due to created diversity (Sanford et al., 2018), whereas rare variants are mutations that are responsible for many genetic diseases (Venkataraman and Rivas, 2019; Horton and Lucassen, 2019; Wright et al., 2018). The Created Heterozygosity and Natural Processes Model of human variation explains current human genetic variation consistent with the historical record of Genesis (Jeanson and Lisle, 2016). The model assigns most CGVs (AF > 5%) to the category of created diversity and most rare variants (AF < 0.5%) to the category of acquired diversity, or mutation. Thus, according to this model, most of the 4.1 to 5 million CGVs found in the typical human genome are the result of God's design, and mutations have produced only a small fraction of the variants in a typical human genome.

God may have created about 1% of the nucleotide pairs in Adam's genomes as heterozygous (Carter and Lightner, 2016). This would mean that about 30 million variants were designed, not only in Adam, but also in Eve, if she were a clone of Adam with respect to her autosomes. These variants would have been passed down to their descendants as CGVs distributed on their chromosomes. Computer models by Carter and Powell (2016) show that most of these CGVs would have survived the Flood in the 8 people on the Ark. Mutations have added a few rare variants to these large numbers of created CGVs to produce the genetic variants found on human genomes today.

To understand the implications of the created diversity argument, one should consider the chromosomes of those who survived the Flood. Noah's family contained 10 unique copies of each autosomal chromosome, if the wives of Noah's three sons (which account for six of the 10) were not closely related. The autosomes of Noah's three sons are combinations of those found in Noah and his wife (the other four copies), and so are not unique. There would also have been nine unique copies of the X chromosome, if Noah and his wife were also not close relatives, two in each daughter-in-law, two in Noah's wife, and one in Noah (which would not get passed to his sons). But there would only have been one Y chromosome, which Noah passed down to his three sons and from which all Y chromosomes on Earth today have descended. Noah's one Y chromosome, by definition, cannot contain any diversity because genetic diversity is due to different bases residing on paired chromosomes. This means that all of the Y chromosome variants in the world today have resulted from the mutations which have accumulated onto the Y chromosomes as they have descended from Noah. As such, there should be few common variants on the individual Y chromosomes of the world, when compared to the autosomes. The total number of variants on a typical man's Y chromosome should be few and should reflect the mutation rate and the years since Noah.

The origin of variants on the X chromosome requires special consideration given the unique history of this chromosome. Adam had only one X chromosome. If Eve's two X chromosomes were clones of Adam's one X, then there should be no created diversity on the X chromosomes today, and only the few mutations accumulated since Creation. In that case, the pattern and concentration of variants on the X chromosomes in the 1000 Genome database should closely resemble the pattern and concentration of variants on the Y chromosomes in the database, both having accumulated all their variants by mutation in a few thousand years. On the other hand, if there is created variation between the two copies of Eve's X chromosome, this created diversity should be found today as many CGVs on the X chromosomes in the 1000 Genome database.

Based on these considerations, the concentrations of variants on all of the chromosomes in 1000 Genome database was determined to test the hypothesis that human genetic diversity is mostly due to created variation with a few added mutations.

Methods

The following will be described: the 1000 Genome database, the Python scripts written to analyze the database, the determination of the adjusted sequence length of the chromosomes, the computation of variant concentrations, and the determination of the number of individual Y chromosome variants per male.

The variant call files (VCF) of the 1000 Genome database for each chromosome were down loaded from the International Genome Sample Resource (IGSR) website (www.internationalgenome.org/Data/Variant Calls/EBI FTP site). These files contain the variants from the genomes of 2,504 people from 26 populations from around the world. Each individual contributed variants from two copies each of the 22 autosomes, each female contributed variants from two copies of the X chromosome, and each male contributed variants from one copy of the X and one copy of the Y chromosome.

Custom Python scripts (available at https://github.com/marshalljordanmd/Big-Data) were written to count the variants on the chromosome files of the 1000 Genome database and to classify them according to their allele frequencies (AF) into three groups, the rare variants (AF less than 0.5%), the uncommon variants (AF between 0.5 and 5%), and the common variants (AF greater than 5%). Each variant with a listed AF greater than 50% was converted to the minor allele frequency by subtracting from 100%. A Python script ("big_data.py")

extracted the AF values from the INFO column for each listed variant of each chromosome file.

To compare the Y chromosome variants to the autosomal variants derived from men only, another Python script ("male_var.py") was written to compute the AF for each male variant by counting the individual genotypes listed in the columns for the male samples of each chromosome file. A similar script ("genotype.py") was written to count all of the genotypes, both male and female, and compute the AF for each variant listed in the 1000 Genome files for the autosomes 18–22 , the X chromosome, and the Y chromosome. Due to the complexity of the X chromosome, three custom Python scripts were required to count the X chromosome genotypes for the whole sample, male-only samples, and female-only samples. The major difficulty with genotype counting on the X was due to male variants being recorded as haploid, except on the "pseudoautosomal" regions on the tips of the X chromosome, where male variants are recorded as diploid variants.

To compute the concentrations of variants found in each chromosome file, in order to normalize the variant counts to length of bases, a determination was needed of the length of DNA that was sequenced for each chromosome. The chromosome lengths were listed on the 1000 Genome Browser on the website of the National Center for Biotechnology Information (www.ncbi.nlm.gov). Inspection of the chromosomes on this browser showed that there was a large variation in the amount of DNA actually sequenced compared to the listed length due to the various amounts of heterochromatin and the size of unsequenced centromeres. For the five "acrocentric" chromosomes (13, 14, 15, 21, 22), variant data only from the long chromosome arm was given, the short arm being entirely heterochromatin. On other chromosomes, the centromere gap was much larger than the average of about 3 million bases (MB) of unsequenced DNA found in most of the chromosomes . For these reasons, and to have a uniform criteria for determination of the length of DNA sequenced for each autosome and the X chromosome, a Python script ("std_dev.py") was written to measure the distance between each variant and compute the mean distance and standard deviation. This script was used to determine an adjusted sequence length by excluding lengths of heterochromatin, the centromeres, and gaps between variants larger than 1MB. This adjusted sequence length was used to normalize the variant counts for each autosome file and the X chromosome file to give the concentration of variants in units of variants per MB.

The Y chromosome presented unique difficulties in arriving at an adjusted sequence length. The length of DNA sequenced was considerably less than that of the other chromosomes, being about one third of its length, as determined by the above inspection routine used for the other chromosomes. The complexity of the Y chromosome and the difficulty of sequencing in regions of high X chromosome homology, found in the "ampliconic" and "X-transposed" regions, rendered half of the sequenced length to be unreliable for variant calls, according to Poznik et al (2013). Therefore, the 1000 Genome Project 10.4 MB inclusion mask, suggested for the Y chromosome by Poznik et al (2016), was used when computing the variant concentrations. The segments of the Y chromosome making up this 10.4 MB mask were downloaded from the technical page of the IGSR (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/chrY). The Python scripts were altered to count only the variants in this 10.4 MB region, and 10.4 MB was used as the adjusted sequence length.

To determine the average number of Y chromosome variants for each individual male, an allele count (AC) dictionary was made with 616 keys, representing the minor AC bins, and with the number of variants for each AC as values. Then the key was multiplied by the value for each AC bin, and these were summed and divided by 1,233 to get the average total number of Y chromosome variants per individual male.

Results

Based on the adjusted sequence lengths, the mean distance between variant positions ranged between 29 bases (SD 124) in the chromosome 16 file and 35 bases (SD 237) in the chromosome 1 file. For the X chromosome file, the mean distance between variant positions was 43 and the standard deviation was 221. For the Y chromosome file, using the 10.4 MB mask as the adjusted sequence length, the mean distance between variant positions was 168 and the standard deviation was 182 (see Table I).

For each of the chromosome files, the adjusted sequence lengths and the variant counts are listed on Table II with the counts divided into the three groups—rare, uncommon, and common. The variant concentrations in each frequency group, computed by dividing the variant counts by the adjusted sequence lengths, are listed on Table III. The autosome common variant concentrations per MB ranged from 2,676 for chromosome 1 to 3,374 for chromosome 19. The X chromosome common variant concentration was 1,890 and the Y chromosome common variant concentration was 183. The mean of the autosome common variant concentrations was 2,958 with a standard deviation of the mean of 172.58. The concentration of common variants on the Y was 6.19% of the mean for the autosomes. The variant concentrations for the three frequency groups for both the X and the Y chromosomes were significantly below the mean of the concentrations on the autosomes at $p < 0.0001$. The results for the common

Table I. ***The distances between variants in bases*****, expressed as mean and standard deviation,** ***with total variant counts and variants per bp for the 24 chromosome files of the 1000 Genome database.*** **Mean and SD for autosomes and the X chromosome were obtained using the Python script "std_dev.py" which had been adjusted to exclude heterochromatin and gaps larger than 1 MB between variants. Mean and SD were obtained for the Y chromosome from the Python script "y_variant_dist.py," using the 10.4 MB mask. "CHR" is the chromosome, "LEN" is the adjusted sequence length, "DIST" is the mean of the distances between variants, "SD" is the standard deviation of the mean, "VAR" is the total number of variants in the chromosome file, and "VAR / BP" is the variants divided by the adjusted sequence length.**

CHR	LEN	DIST	SD	VAR	VAR / BP
1	228.2	35.28	237.36	6467979	0.0283
2	239.1	33.77	146.13	7081503	0.0296
3	194.9	33.42	56.39	5832220	0.0299
4	187.5	32.80	86.48	5736878	0.0306
5	177.9	33.78	87.92	5265703	0.0296
6	167.9	33.42	103.86	5024045	0.0299
7	156.1	33.99	114.99	4716660	0.0302
8	143.3	31.17	110.69	4597039	0.0321
9	123.0	34.54	340.04	3560643	0.0290
10	132.3	33.13	164.75	3992142	0.0302
11	131.8	32.57	183.41	4045566	0.0307
12	130.8	33.81	85.30	3868362	0.0296
13	96.1	33.62	155.46	2857873	0.0297
14	88.3	33.25	39.80	2655038	0.0301
15	82.5	34.03	339.30	2424665	0.0294
16	79.1	29.33	124.23	2697907	0.0341
17	78.2	33.57	126.17	2329260	0.0298
18	74.9	33.04	115.97	2267139	0.0303
19	56.0	30.54	80.05	1832467	0.0327
20	59.8	32.99	135.11	1812814	0.0303
21	35.6	32.16	196.13	1105518	0.0311
22	34.3	34.48	290.32	1073411	0.0313
X	152.1	43.86	221.60	3467421	0.0228
Y	10.45	168.33	182.37	61956	0.0059

variant concentrations, the CGVs, are displayed on Figure 1.

The results of genotype counting of the whole sample (males and females) versus males-only for autosomes 18–22 and X and Y are listed in Table IV. By comparison of the males-only to the whole sample counts, it can be seen that the variant counts in the common and uncommon groups are not much changed, whereas the rare variant counts are reduced when counting males-only. This is easily explained by the fact that, when counting singleton variants (those with only one person in the whole sample carrying the variant), if half the sample is female, then half of these singleton variants will not be counted in the male-only counts. Also, for doubletons, one expects that, on average, 25% of them would also not be counted since they would belong to only two females in the whole sample. Conversely, the common variant count is not much changed by counting of male-only genotypes, since, while the AC may be cut in half by only counting males, the AF remains unchanged since the sample size, the denominator, is also cut in half by only counting males. Therefore, counting male-only genotypes did not alter the finding that common variant counts are reduced on the Y chromosomes and the X chromosomes.

The average number of variants per male on the Y chromosomes in the 1000 Genome database was obtained by multiplying the AC bin number by the number of variants in the bin and adding them together. A total count of 732,415 was obtained. Dividing this total count by the sample size of 1,233 males, one finds that the average male in the database carries on his Y chromosome 594 variants: 68 rare, 178 uncommon, and 348 common. Assuming 200 generations since Noah, these mutations would have accumulated at a rate of 2.97 per generation. A result consistent with the Y chromosome mutation rate reported by Jeanson (2019) and Jeanson and Holland (2019).

Discussion

The concentration of CGVs is significantly reduced on the Y chromosomes as compared to the autosomes of the 1000 Genome database, consistent with

the Created Diversity Model. This result overturns the notion that all variants on the autosomes are solely due to accumulated mutations. If the source of genetic variation on both the Y chromosomes and the autosomes were due only to mutations, one would expect them to have similar concentrations of CGVs. The mechanism of genetic variation must differ between the autosomes and the Y chromosomes. Interpretation of these results based on the historical record of Genesis offers an explanation for these remarkable findings, namely that the autosomes are replete with created diversity, while the Y chromosomes lack created diversity and only have the few mutations that have accumulated since Noah .

The variant concentrations for the X chromosomes are intermediate to those of the autosomes and the Y chromosomes. This suggests that the X chromosomes differ from the Y chromosomes by having some created diversity, but not as much as the autosomes. The implications of these findings are that Eve's two X chromosomes were seeded with created variant positions, and are not an identical clone of Adam's.

The concentration of rare variants is also reduced on the Y chromosomes. This can be explained by the fact that only 1,233 men contributed one copy each of the Y chromosome to the database, whereas these 1,233 men contributed two copies each of the autosomes to the database. The whole sample of 2,504 people thus contribute 5,008 chromosomes to each of the autosomal files. So the Y chromosome rare variants were found on 1,233 chromosomes, whereas the whole sample autosomal rare variants were found on 5,008 chromosomes. The fewer chromosomes sequenced will result in fewer rare variants recorded for the Y chromosomes due to the effect of sample size on the probability of discovery of low-frequency variants in the population.

Table II. ***Chromosome lengths and variant counts.*** **"CHR" is the chromosome name. "SIZE" is the chromosome length in millions of bases (MB) on the NCBI 1000 Genome Browser. "SEQ" is the sequenced length covered in the file, minus gaps over 1MB (aka "the adjusted sequence length"). "<0.5%," "0.5–5%," ">5%" are variants counts for the rare, uncommon, and common allele frequency groups. "TOTAL" is the sum of the three groups, which is all the variants in the chromosome file.**

CHR	SIZE (MB)	SEQ.	<0.5%	0.5–5%	>5%	TOTAL
1	249.3	228.2	5111842	745512	610625	6467979
2	243.2	239.1	5634777	797577	649149	7081503
3	198.0	194.9	4607628	662904	561688	5832220
4	191.2	187.5	4484322	673473	579083	5736878
5	180.9	177.9	4169319	604297	492087	5265703
6	171.1	167.9	3900585	593941	529519	5204045
7	159.1	156.1	3700430	552416	463814	4716660
8	146.4	143.3	3644484	523788	428767	4597039
9	141.2	123.0	2812347	405938	342358	3560643
10	135.5	132.3	3130071	461441	400630	3992142
11	135.0	131.8	3193629	466650	385287	4045566
12	133.9	130.8	3037436	452868	378058	3868362
13	115.2	96.1	2236056	334631	287186	2857873
14	107.3	88.3	2090083	307965	256990	2655038
15	102.5	82.5	1914145	279164	231356	2424665
16	90.4	79.1	2156250	296673	244984	2697907
17	81.2	78.2	1852509	258608	218143	2329260
18	78.1	74.9	1779090	264493	223556	2267139
19	59.1	56.0	1430530	212992	188945	1832467
20	63.0	59.8	1434105	207606	171103	1812814
21	48.1	35.6	863351	127201	114966	1105518
22	51.3	34.3	836746	127937	108728	1073411
X	155.3	152.1	2781668	398336	287417	3467421
Y	59.4	10.4	50842	9216	1898	61956

The Effect of Sample Size on Variant Discovery

To see how sample size affects variant discovery, consider a variant of AF = 0.1 in a large population. The probability that a randomly selected genome from that population does not contain the variant is 1 – 0.1 = 0.9. The probability, P, that a series of n randomly selected samples from the population does NOT contain

Table III. Chromosome variant concentrations, in variants per MB, obtained by dividing the variant count of each frequency group by the adjusted sequence length in MB. "CHR" is the chromosome name. "SEQ" is the adjusted sequence length in millions of bases (MB), "RARE," "UNC," "COMMON" are the concentrations of variants per MB for the three groups of rare, uncommon, and common variants, corresponding to Allele Frequencies (AF) <0.5%, 0.5–5%, and >5%, respectively. The concentration means and standard deviations (SD) are given for the autosomes, and the concentrations for the X and Y chromosomes are compared to them for statistical significance. The * indicates statistical significance at p < 0.0001 comparing X and Y concentrations to the means of the autosomal concentrations.

CHR	SEQ	RARE	UNC	COMMON
1	228.2	22401	3267	2676
2	239.1	23567	3336	2715
3	194.9	23641	3401	2882
4	187.5	23916	3592	3088
5	177.9	23436	3397	2766
6	167.9	23232	3537	3154
7	156.1	23706	3539	2971
8	143.3	25433	3655	2992
9	123.0	22865	3300	2783
10	132.3	23659	3488	3028
11	131.8	24231	3541	2923
12	130.8	23222	3462	2890
13	96.1	23268	3482	2988
14	88.3	23670	3488	2910
15	82.5	23202	3384	2804
16	79.1	27260	3751	3097
17	78.2	23689	3307	2790
18	74.9	23753	3531	2985
19	56.0	25545	3803	3374
20	59.8	23982	3472	2861
21	35.6	24251	3573	3229
22	34.3	24395	3730	3170
Mean (SD) of autosomes		23680 (1012.13)	3502 (142.40)	2958 (172.58)
X	152.1	18288*	2619*	1890*
Y	10.4	4887*	886*	183*

* indicates p < 0.0001 compared to the autosomal means for each group

the variant with allele frequency, AF, is given in Equation 1.*

$$P = (1- AF)\hat{}n \qquad \text{Equation 1}$$

Equation 1 shows that, for a given sample size n, increasing AF reduces the probability that the variant will be missed. Figure 2 is a plot of P, the probability of missing the variant, versus n, the sample size, for variants of several different AF. This plot has the sample size scale ranging up to 1000, a sample size approaching that of the 1000 Genome database for the Y chromosomes, which are from 1,233 males. The three AFs on Figure 2 are 5%, the value separating common from uncommon variants, 0.5%, the value separating uncommon from rare variants, and 0.02%, the value for the "singletons" of the database, which are variants found in only one in 5,000. On Figure 2, note that 80% of variants of AF 0.02% in the population will be absent in a sample of 1,000. This difficulty in finding rare variants is acknowledged for Phase 3 of the 1000 Genome Project, where the power to detect variants of AF greater than 1% is estimated to be > 99%, but the power to detect variants of AF of 0.1% is only 75%. (The 1000 Genomes Project Consortium, 2015).

From Equation 1 and Figure 2, while only 20% of singleton variants will be detected with a sample size of 1,000, all of the CGVs will be detected. A sample size exceeding 1,000 will lead to a greater detection rate of singleton variants, but will not change the detection rate of CGV. If any CGVs are missed, this would be due to sequencing errors and low coverage, not to low sample size.

Based on this statistical argument, the deficit in rare variants on the Y chromosomes compared to the autosomes can be explained by the smaller sample size for the Y chromosomes. But the deficit in CGVs is not due to this difference in sample size, but is due to the absence of created variants on the Y chromosomes.

Data Quality Considerations

Data quality in the 1000 Genome database could impact these results. Low coverage of the Y chromosomes in the 1000 Genome database (the lack of repeat sequencing of variant positions to eliminate errors) could explain some of the deficit in Y chromosome variants. The error rate for variant discovery is thought to be greater for the rare variants. Increasing depth of coverage should reduce both the false discovery rate and the false negative rate.

The problem of low coverage affecting data quality is addressed by Ponzik, et al (2016) concerning the sequence reads of 1000 Genome Project used in his study. He states, "We applied stringent quality control to meet the Project's requirement of false discovery rate (FDR) < 5% for SNVs (single nucleotide variants)... In our validation analysis with independent datasets, genotype concordance was greater than 99% for SNVs..." Table I in Ponzik's paper shows that, in the 60,555 SNVs discovered in the Y chromosome 1000 Genome samples he studied, there was a false discovery rate of 3.9% and a concordance rate with the 1000 Genome database of 99.6%. Poznik et al (2016) then addressed the false negative rate of variant discovery by comparing the variable sites called on 143 high-coverage (80x) Complete Genomics (CG) Y chromosome samples to those found in the low-coverage data (4x) from the 1000 Genome database. They found 3,834 of 17,194 called sites on the high-coverage CG analysis were not called in the low-coverage 1000 Genome Y chromosome analysis, for a false negative rate of 22%. Most of the missed calls (87%) were for rare variants (Poznik et al., 2016, Supplemental Notes 1.3.1 and 1.3.2). High coverage of the CG samples found 22% more variants on the Y chromosome than the low coverage of the 1000 Genome Project. Most of the missed variants were rare variants.

The false negative rate of 22% does not alter the conclusion of this paper,

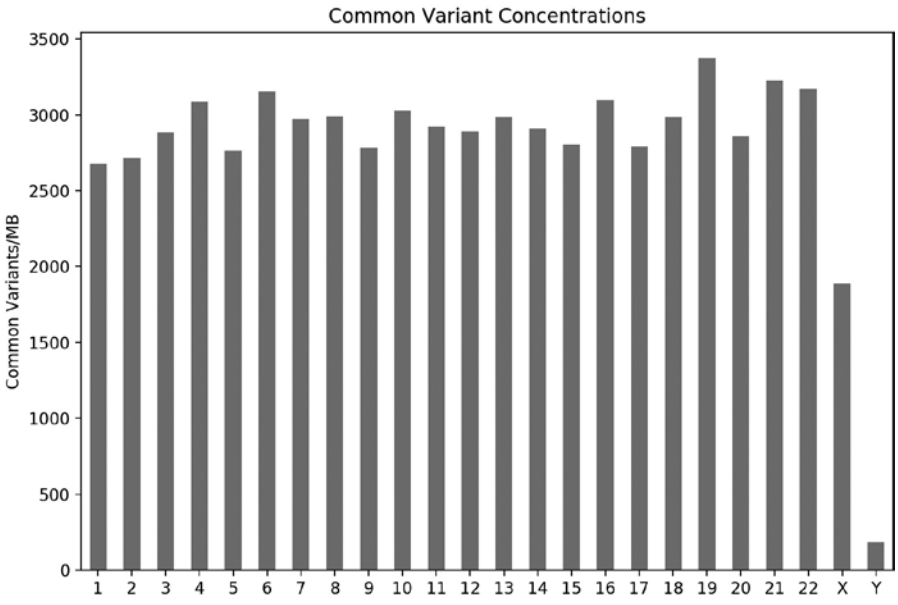

Figure 1. Common variant concentrations in variants per MB for the chromosome files from the 1000 Genome Project.

Table IV. ***Chromosome variant counts based on genotype counting from 1000 Genome database files obtained from International Genome Sequence Resources.*** **The "WHOLE SAMPLE" counts (males and females) are from "genotype.py" and the "MALE ONLY" counts are from "male_var.py." X chromosome variants based on genotypes were counted by "new_x.py" for the whole sample of males and females, and male-only variants based on genotypes were counted with "male_var_x.py." These Python scripts are available on GitHub as noted in the text of the paper. Counts are displayed as [Rare, Uncommon, Common].**

CHR	WHOLE SAMPLE	MALE ONLY
18	[1772063, 264455, 223481]	[1204873, 265219, 223103]
19	[1425486, 212934, 188905]	[972048, 212861, 188197]
20	[1428094, 207570, 171048]	[968455, 208136, 170923]
21	[857935, 127026, 114844]	[587414, 127333, 114526]
22	[853657, 130811, 111404]	[585829, 132065, 111228]
X	[2767440, 398236, 287277]	[1466272, 381575, 285758]
Y		[50580, 9241, 1858]

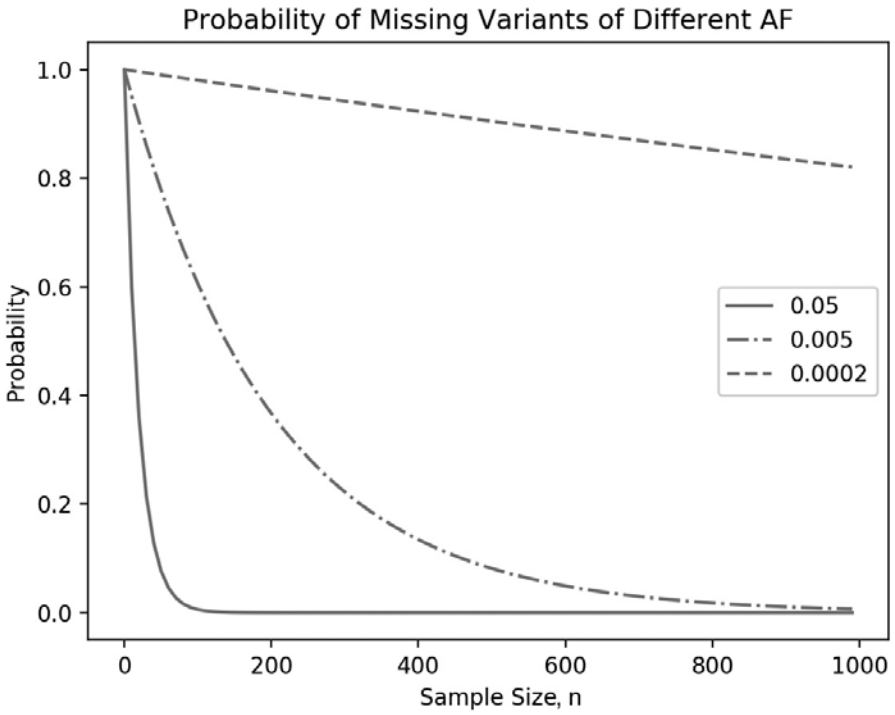

Figure 2. The probability of missing the variant after n samples selected, for variants of AF 0.05 (5%), 0.005 (0.5%), and 0.0002 (0.02%). P = (1 - AF) ^ n , where P is the probability the variant of allele frequency (AF) has not been found after n random selections from an infinitely large population; "n" is the sample size.

since increasing the common variant counts on the Y chromosomes by 22% (a large over-correction), raising the concentration from 183 to 223 per MB, would still result in a statistically significant deficit, at the $p < 0.01$ level, in CGVs on the Y chromosomes compared to the autosomes. The false negative rate of the 1000 Genome data due to low coverage does not explain the CGV deficit of the Y chromosomes. The lack of CGVs in the Y chromosomes of the 1000 Genome database is consistent with lack of created diversity on the world's Y chromosomes due to the descent of all men from Noah, as recorded in Genesis.

The 22% false negative rate does not completely explain the rare variant deficit of the Y chromosomes. A 22% increase in the number of rare variants on the Y chromosome would raise the concentration from 4,887 per MB (Table III) to 5,962 per MB, which is 25% of the mean rare variant concentration found on the autosomes. The remaining rare variant deficit is explained by the lower sample size for the Y chromosomes.

Predictions of Coalescent Theory

An ancestral tree can be drawn from one couple, with branching at each generation, as offspring descend to the current population. When the branches are followed back in time to the original couple, a coalescence process is seen to occur with branch points representing individuals who have produced multiple children. Such trees have been built using genetic variant profiles, in which case the branch points represent common variants as the root of the tree is approached, going back in time. In other words, common variants are a marker of coalescence.

According to Coalescent Theory (Rosenberg and Norborg, 2002), the probability of coalescence of divergent sequences in the previous generation is inversely proportional to the number of samples (N) as seen in the following equation:

$$P(\text{coalescence}) = n(n-1) / 4N$$

Equation 2

Here, P is the probability of coalescence in the previous generation, n is the number of variants under consideration, and N is the effective population size. The reciprocal of Equation 2 is the time to coalescence, which is proportional to N. When applied to DNA sequences, the value of N for autosomes is four times as large as N for the Y chromosomes since there are four copies of each autosome for every Y chromosome. Therefore, autosomes should coalesce 4 times slower than Y chromosomes. If both the autosomes and the Y chromosome descend from the same ancestral couple, and if all variants are due to mutations accumulating, the autosomes should have at least 4 times fewer variants than the Y chromosomes. Yet, in the 1000 Genome database, the autosomes contain many more variants than the Y chromosomes. According to the Created Diversity Model, this is due to created variation on the autosomes.

Coalescent Theory can be used to estimate the time to the most recent common ancestor (TMRCA) from the DNA sequence data of a population (Rosenberg and Feldman, 2001). The theory assumes that genetic variants derive from mutations that are inherited by successive generations down to the present, a condition applicable to the Y

chromosome according to both creationists and evolutionists.

After constructing a phylogenetic tree from the 1000 Genome Y chromosome variant data, and applying a mutation rate of 0.76 X10^-9 mutations per base pair per year, Poznik (2016) estimated the TMRCA to be 190,000 years. This mutation rate, derived using evolutionary assumptions, is two orders of magnitude lower than that measured from pedigree-based high coverage Y chromosome data. When the more accurate mutation rates are used, Coalescent Theory estimates a TMRCA based on the Y chromosome variants of about 4,500 years (Jeanson and Holland, 2019).

Y Chromosome CGVs Are a Genetic Record of History

While all the variants on the Y chromosomes are due to mutations accumulated after Noah, the CGVs on the Y chromosome are due to mutations that occurred in the founders of the world's populations who spread out from Babel (Carter and Hardy, 2015; Sibley, 2017) and in men who fathered the large male lineages that persist in the world today. Historical events explain the survival of these lineages and the demise of others.

In contrast to the CGVs, the number of rare variants found on all the chromosomes in the database, is directly related to sample size. This is because mutations enter the population as rare variants, even as singletons. Every position in the genome is a potential variant site for a mutation. Among the 7.5 billion people in the world, variants are likely to exist today for every one of the 3.2 billion positions in the human genome, except for lethal variants. If all the genomes in the world were sequenced and variants recorded, the total variant concentrations in the database would approach 1 variant per base for all chromosomes as more and more rare variants were found. But the CGV deficit on the Y chromosomes compared to the autosomes would remain a permanent feature of human genetic variation due to lack of created variants on the Y. This prediction of the Created Diversity Model can be tested as the databases expand with more Y chromosomes sequences.

*[For a variant with AF = 0.1, the probability of randomly selecting a genome without this variant from a large population is 1 - 0.1 = 0.9. The probability of missing the variant on two consecutive selections is 0.9 X 0.9 = 0.81. The probability of missing the variant on three selections is 0.9^3 = 0.729. And so on. So the probability of missing the variant after n random selections is 0.9^n. Thus, P = (1 - AF)^n is the probability of missing the variant of frequency AF after n selections.]

Conclusion

The lack of common genetic variants found on the Y chromosomes versus the autosomes of the 1000 Genome database is consistent with most common variants on the autosomes being created diversity, which is absent from the Y chromosomes. The few variants found on the Y chromosome are due to mutations acquired with descent from Noah within the last 4,300 years. Human genetic variation is explained by two mechanisms: created diversity, which is the cause of most common variants, and mutations, which is the cause of most rare variants. Since common variants greatly outnumber rare variants in a typical human genome, created diversity, not mutations and natural selection, is the cause of most human genetic variation.

Summary

The first two decades of the twenty-first century have witnessed a vast increase in knowledge in the science of genomics. Great controversy has attended the sequencing of the human genome as scientists with opposing worldviews give conflicting interpretations of the data. Evolutionists have assumed that all human genetic variation has developed by mutation and natural selection over millions of years, while those who believe the biblical record of recent creation maintain that most human genetic variation is created diversity placed by God into the genomes of our original parents.

The pattern of human genetic variation found in the 1000 Genome Project database provides a way to resolve this controversy. Created variants should be absent from the Y chromosomes in the world, as all have descended from Noah's single Y chromosome. The investigation reported here finds that the concentrations of common genetic variants in the Y chromosomes of the 1000 Genome database is dramatically less than the concentrations found on the autosomes, consistent with the Created Diversity Model. The notion that all human genetic variation is due to mutation is thus overturned. The history of Creation and the Flood recorded in Genesis provides the basis for understanding the origin of human genetic variation.

References

1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* 526:68–74.

Carr, S. 2013. *An Introduction to Population Genetics*. Chapter 3: "Coalescence Theory: Relating Theory to Data." Sinauer Associates, Inc. Publishers, Sunderland, MA.

Carter, R., and C. Hardy. 2015. Modeling biblical human population growth. *Journal of Creation* 29(1):72–79.

Carter, R., and J. Lightner. 2016. Human genetic data affirms biblical history on many levels and is an excellent resource for creation-based research. *Creation Research Society Quarterly* 52:249–255.

Carter, R., and M. Powell. 2016. The genetic effects of the population bottleneck associated with the Genesis Flood. *Journal of Creation* 30(2):102–111.

Elhaik, E., et al. 2014. The "extremely ancient" chromosome that isn't: a forensic bioinformatic investigation of Albert Perry's X-degenerate portion of the Y chromosome. *European Journal of Human Genetics* 22:1111–1116.

Elliot, M., and A. Mooers. 2007. www.sfu.ca./courses/basic869/869_lectures/MHP_Coalescent.

Horton, R., and A. Lucassen. 2019. Genomic testing in healthcare: A hybrid space where clinical practice and research need to co-exist. *Expert Review of Molecular Diagnostics* 19(11):963–967.

Iyengar, S.K., and R.C. Elston. 2007. The genetic basis of complex traits: Rare variants or "common gene, common disease"? *Methods in Molecular Biology* 376:71–84.

Jeanson, N. 2019. Testing the predictions of the young-Earth Y chromosome molecular clock: Population growth curves confirm the recent origin of human Y chromosome differences. *Answers Research Journal* 12:405–423.

Jeanson, N., and A. Holland. 2019. Evidence for a human Y chromosome molecular clock: Pedigree-based mutation rates suggest a 4,500-year history for human paternal inheritance. *Answers Research Journal* 12:393–404.

Jeanson, N., and J. Lisle. 2016. On the origin of eukaryotic genotypic and phenotypic diversity: Species' genetic clocks, population growth curves, and comparative nuclear genome analyses suggest created heterozygosity in combination with natural processes as a major mechanism. *Answers Research Journal* 9:81–122.

Poznik, G., et al. 2016. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics* 48(6):593–559.

Poznik, G., et al. 2013. Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science.* 341(6145):562–565.

Rosenberg, N.A., and M.W. Feldman. 2001. "The relationship between coalescence times and population divergence times." Chapter 9 in Slatkin, M., and M. Veuille (editors). 2002. *Modern Developments in Theoretical Population Genetics.* Oxford University Press, New York, NY.

Rosenberg, N.A., and M. Nordborg. 2002. Genealogical trees, coalescence theory and the analysis of genetic polymorphisms. *Nature Reviews: Genetics* 3(5):380–390.

Sanford, J., R. Carter, W. Brewer, J. Baumgardner, B. Potter, and J. Potter. 2018. Adam and Eve, designed diversity, and allele frequencies. In Whitmore, J.H. (editor). *The Proceedings of the International Conference on Creationism* 8:200–216. Creation Science Fellowship, Pittsburgh, PA.

Sibley, A. 2017. Dating the Tower of Babel events with reference to Peleg and Joktan. *Journal of Creation* 31(1):80–87.

Strachan, T., and A. Read. 2004. *Human Molecular Genetics* 3. Volume 3:256–295. Garland Science, New York, NY. ISBN 978-0-81534182-6.

Venkataraman, G.R., and M.A. Rivas. 2019. Rare and common variant discovery in complex disease: The IBD case study. *Human Molecular Genetics* 28(R2):R162-R169.

Wright, C.F., D.R. FitzPatrick, and H.V. Firth. 2018. Paediatric genomics: Diagnosing rare disease in children. *Nature Reviews: Genetics* 19:253–268.