



# Maruyama's allelic age revised by whole-genome GEMA simulations



Shuhao Qiu, Alexei Fedorov\*

Program in Bioinformatics and Proteomics/Genomics, University of Toledo, Health Science Campus, Toledo, OH 43614, USA  
Department of Medicine, University of Toledo, Health Science Campus, Toledo, OH 43614, USA

## ARTICLE INFO

### Article history:

Received 31 July 2014

Accepted 16 February 2015

Available online 21 February 2015

### Keywords:

Genomics

Computational biology

Allelic age

SNP

Haplogroups

## ABSTRACT

In 1974, Takeo Maruyama deduced that neutral mutations should, on average, be older than deleterious or beneficial ones. This theory is based on the diffusion approximation for a branching process, which considers mutations independently of one another and not as multiple groups of interconnected mutations with strong linkage disequilibrium (haplotypes). However, mammalian genomes contain thousands of haplotypes, in which beneficial, neutral, and deleterious mutations are tightly linked to each other. This complex haplotype organization should not be ignored for estimation of allelic ages. We employed our GEMA computer simulation program for genome evolution to re-evaluate Maruyama's phenomenon in modeled populations that include haplotypes approximating real genomes. We determined that only under specific conditions (high recombination rates and abundance of neutral mutations), the deleterious and beneficial mutations are younger than neutral ones as predicted by Maruyama. Under other conditions, the ages of negative, neutral, and beneficial mutations were almost the same.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

Investigations of “allelic age” began in 1970s. This term was defined as the number of generations a mutant allele has persisted in the population since its first occurrence [7,10–12]. Initially, the prediction of allelic age relied upon mathematical modeling—a diffusion approximation for a branching process. In 1973, Kimura and Ohta [7] inferred that the “average ages of neutral alleles, even if their frequencies are relatively low, are quite old.” Specifically, they demonstrated that a neutral mutation whose current frequency is 10% has an expected age (measured in generations) roughly equal to the effective population size  $N_e$ . This result complicates experimental verification of allelic age predictions. Thus, allelic age estimates currently come from either mathematical modeling or indirect experimental hints about the distribution patterns of mutations with various population frequencies. In 1974, Takeo Maruyama [11] modeled semidominant mutations and made a principal prediction that neutral mutations, on average, are significantly older than both deleterious and beneficial alleles. This prediction has been widely accepted and became an important landmark in this field. A year after Maruyama's paper, Wen-Hsiung Li [10] inferred the age of deleterious mutations having various degrees of dominance. He demonstrated that the mean age decreases with increasing selection coefficients against heterozygotes. Allelic age has been nicely reviewed in the late 1990s [5] and early 2000s [18]. The allelic age has been indirectly estimated in several independent experimental studies that statistically

examined the distributions of multiple mutant alleles. Slatkin and Rannala [17] estimated the allelic age by use of intra-allelic variability. Further, Rannala and Reeve applied high-resolution multipoint linkage-disequilibrium mapping [14], while Genin and colleagues analyzed shared haplotypes of rare disease mutations [4]. Last year, Kiezun and co-authors [6], concluded from analysis of large-scale population sequencing studies and computer simulations, that deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency. However, the allelic ages for neutral, deleterious and beneficial mutations are still unclear because the direct measurement of the age is impossible.

Recent whole-genome sequencing of numerous individuals revealed that each human individual bears millions of mutations [1]. These millions of mutations form intricate patterns of haplotypes, where neutral, beneficial, and deleterious mutations are tightly linked with each other and strongly influence the ages of their neighbors. A haplotype structure for a gene strongly depends on the local recombination rate, which may vary thousands of times from one chromosomal location to another [2].

In order to examine the role of haplotypes on the allelic age, we applied whole-genome computer simulations of SNP dynamics using our GEMA program package. A “naturally occurring” intense influx of 40 novel mutations per person has been applied in this computer modeling. Such intense mutation influx generated thousands of SNPs in each modeled individual. The time of the arrival for each mutation has been recorded and used for the calculation of its age. These simulations allow the direct measurement of the average age of mutations with high accuracy. In these computational experiments, we changed various parameters such as recombination rate, degrees of dominance, and distributions of mutations by their selection coefficients. These various

\* Corresponding author at: Department of Medicine, University of Toledo, Health Science Campus, Toledo, OH 43614, USA. Fax: +1 419 383 3102.  
E-mail address: [Alexei.fedorov@utoledo.edu](mailto:Alexei.fedorov@utoledo.edu) (A. Fedorov).

conditions drastically altered the patterns of haplotype ensembles in the modeled genome. We demonstrated that Maruyama's effect appears only for specific sets of parameter ranges and quantitatively described its variation under different conditions.

## 2. Materials and methods

Computer simulations were performed using a new v3 release of our Perl program GEMA\_v3.pl, named Genome Evolution with Matrix Algorithms (GEMA). The previous release (GEMA\_v2.pl) has been described in detail [13]. Both v2 and v3 versions are freely available from our web site: <http://bpg.utoledo.edu/~afedorov/lab/GEMA.html>. The V3 release of GEMA has only a small addition compared to v2, which, upon creation of a new mutation, records the time of its arrival (measured in generations, as \$g variable inside a multidimensional array @matrix). Finally, the age of every SNP is periodically recorded into a new fifth column of the GEMA backup file.

In the described simulations with GEMA\_v3.pl, we always used the following parameters: (1) unsaturated mode; (2) duration: 10,000 generations; (3) population size ( $N = 100$ ); (4) number of offspring per mating pair ( $\alpha = 5$ ); (5) mutation rate per gamete ( $u = 20$ ); (6) recombination rate ( $r = 1$  or  $r = 48$ ); (7) dominance coefficient ( $h = 0$ ,  $h = 0.5$ , or  $h = 1$ ); (8) MatingScheme: permanent random male–female pairs; and (9) upon generation of a random mutation, a random number generator imbedded into GEMA program assigned a selection coefficient to it either according to the “experiment B” or “experiment C” distributions demonstrated in Fig. 1. Experiments B and C were first described in our paper [13], and we kept their original names in this paper for clarity. Those two experiments were chosen for the ease of interpretation of the results. The effects of all deleterious mutations in these experiments are equal to each other since their selection coefficients ( $s$ ) always equal to  $-1$ . Consequently, all beneficial mutations are also equal to each other ( $s = +1$  for all beneficial mutations).

Our GEMA modeling approximates natural conditions in a way in which we consider thousands of genes in genome of virtual individuals and the real influx of novel mutations (which is about 40 new mutations per individual). As we demonstrated in Qiu et al. [13], several hundreds of genes in the modeling genome have approximately the same effect on SNP dynamics as 25,000 genes observed in humans. In addition, the length of modeling genes does not significantly influence the SNP dynamics. Due to these reasons and for the speed of computations, we used a 0.6 Mb long DNA segment with a random nucleotide sequence as the genome for modeled individuals. Thousands of nucleotide-long segments of this sequence were used to model 600 genes. The simplification of our modeling, compared to real conditions, is that all genes in our simulations have the same properties. This includes the same recombination rate, same frequencies of deleterious, beneficial, and

neutral mutations and the same dominance coefficient. In real human genes, these parameters vary significantly from gene to gene. However, these simplifications allow us to evaluate the influence of each parameter on the dynamics of SNP in the population.

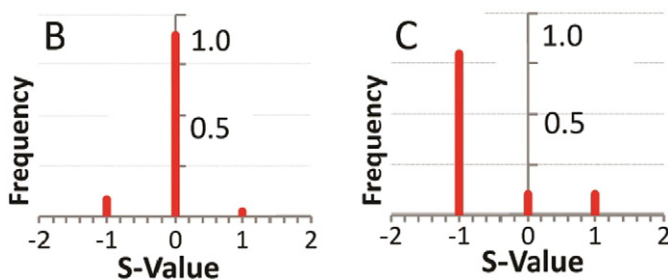
The snapshot of all SNPs in all modeled individuals was recorded after every 1000 generations as backup files. These backup files contain the following information on each SNP: position; selection coefficient; mutant nucleotide; modeled individuals bearing this SNP including location on a maternal or paternal DNA; and the time of SNP arrival (in generations). Backup files was processed with our Perl scripts AllelicAge\_10bin.pl and AllelicAge\_csv.pl, that calculate the frequency of each SNP, its selection coefficient and the time of its arrival, and present this information in an output table in Excel format (Supplementary Materials, Tables S1 and S2). These tables were used to calculate the distribution of SNPs by their population frequency, the number of SNPs with particular selection coefficient within a designated range of population frequencies (from 10% to 30% range or in 40%–60% range), and the distribution of SNPs within a particular range of population frequency by their age. The SNP frequency stands for the frequency of the mutant alleles in the entire modeled population.

## 3. Results

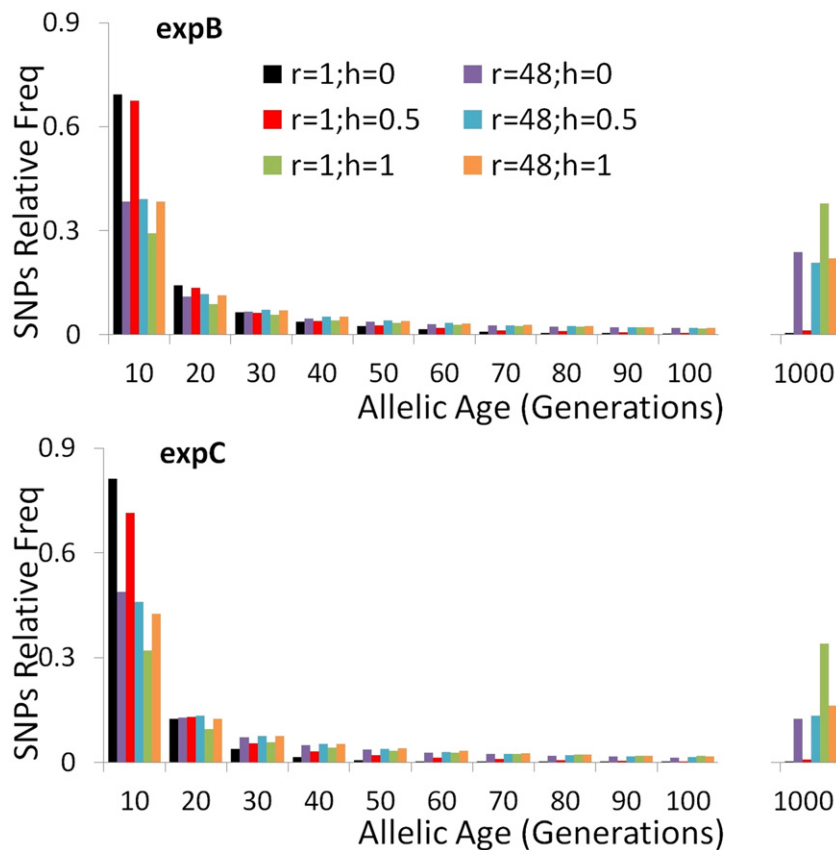
Computer simulations of whole-genome SNP dynamics were performed using the program GEMA\_v3.pl. In these computations, the following three parameters were always the same for every experiment: (1) population size was 100 modeled individuals ( $N = 100$ ); (2) every modeled individual had 40 novel mutations ( $\mu = 20$  mutations per gamete); and (3) the mating scheme was a default GEMA choice—permanent random male–female pairs (MatingScheme = 1) with 5 offspring per mating pair ( $\alpha = 5$ ). Also, genomes of modeled individuals always consisted of 600 genes each 1000 nucleotide long. [As we discussed previously, the exact number of genes above a certain threshold ( $\sim 200$ ) does not significantly influence SNP dynamics [13]]. Variable parameters for each computational experiment were the following: (1) number of recombination events per gamete ( $r$ ) was either  $r = 1$  or  $r = 48$ ; (2) gene dominance coefficient ( $h$ ) for every gene was either  $h = 0$  (dominant genes),  $h = 0.5$  (co-dominant genes), or  $h = 1$  (recessive genes); and (3) distribution of mutations by their selection coefficients corresponded to the “Experiment B” or “Experiment C” shown in Fig. 1. We specifically used  $r = 48$ , because it represents the average number of pieces of paternal and maternal genomes in a human gamete [13]. The alternative  $r = 1$  settings model the regions with low recombination rate frequency, which are abundant in the human genome.

The distribution of SNPs by their age for different modeled parameters is shown in Fig. 2. This distribution has been combined for 12 independent experiments. The total number of all SNPs in specific experiment varied from 152,582, for simulations with  $r = 1$ ,  $h = 0$ , and “experiment C”, to the 505,970 SNPs for  $r = 1$ ,  $h = 1$ , and “experiment B” simulations. Since the number of SNPs varies from one experiment to another, we performed their normalization by division by the total number of SNPs in each experiment. Hence, the results in Fig. 2 are presented as relative SNP frequencies counted within 10-generation bins. The details for every SNP from these data are provided in the supplementary Table S1. In all experiments the youngest SNPs were the most numerous ones, as expected from population genetics. We observed that, when the recombination rate was high ( $r = 48$ ), the older SNPs were more abundant than when the recombination rate was low ( $r = 1$ ). A special case that does not follow this rule is provided by the combination of low recombination rate ( $r = 1$ ) with recessive dominance coefficient ( $h = 1$ ). As we explained previously [13], these specific conditions may result in an un-stable number of SNPs in the population, periodically producing gigantic peaks of SNP numbers.

The calculated mean age of SNPs, for which population frequencies belong to a particular range (10%–30% or 40%–60%) is shown in Fig. 3.



**Fig. 1.** Distribution of computer-generated mutations by their selection coefficients ( $s$ -values). B—“Experiment B” models a discrete distribution of mutations characterized predominantly by neutral mutations, occurring at a frequency of 90% within the population, while the remaining 10% is characterized by deleterious and beneficial mutations occurring in a ratio of 9:1. C—“Experiment C,” the ratio of deleterious to beneficial mutations occurs again in the ratio of 9:1. However, this model is characterized by a preponderance of mutations with deleterious effects (81%). Neutral mutations in this case comprise 10% and beneficial - 9% of overall nucleotide changes occurring within the population.



**Fig. 2.** Distribution of relative frequencies of all SNPs in population by their age (measured in generations since SNP arrival). The extreme right block of columns marked as “1000” on the horizontal axis shows the relative frequencies of all accumulated SNPs at 110–1000 generations.

These SNPs were grouped by their selection coefficients ( $s$ ) into being deleterious ( $s = -1$ ), neutral ( $s = 0$ ), or beneficial ( $s = +1$ ). Fig. 3 illustrates that recombination rate ( $r$ ), dominance coefficient ( $h$ ), and the distribution of mutation by selection coefficients (experiments B or C) may significantly influence on the mean allelic age. In 75% of experiments, no difference in the allelic ages for neutral, beneficial, or deleterious mutations was detected. Only with high recombination ( $r = 48$ ) and “Experiment B” settings (90% neutral SNPs) was the mean age of neutral mutations 1.4–2.6 times higher than for deleterious or beneficial ones, in accordance to Maruyama’s predictions [11].

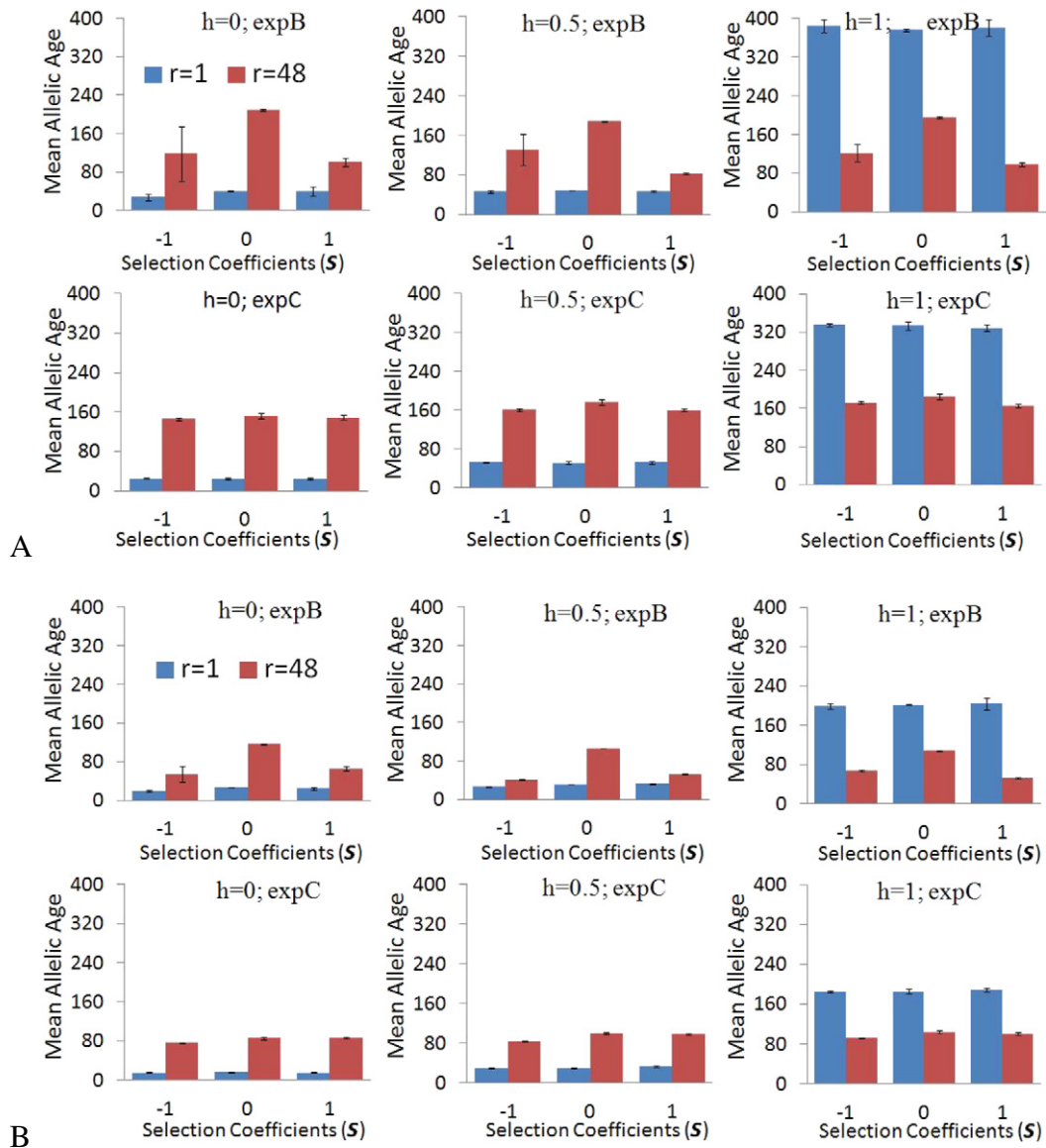
Finally, the distribution of SNPs by their ages is demonstrated in Fig. 4. For proper comparison of different experiments with various parameters, the number of SNPs has been normalized by division by the total number of SNPs in the experiment. Thus, Fig. 4 represents SNP frequency density and provides an overall view of the age distribution of all SNPs. Statistical information about these distributions including beneficial, deleterious, and neutral groups of SNPs is presented in Table 1 (detailed information on the age of each SNP is presented in supplementary Tables S1 and S2).

#### 4. Discussion

Maruyama made a non-obvious and intriguing theoretical prediction about the average age of deleterious, beneficial, and neutral mutations. Experimental verification of SNP age encounters two major problems. First is ascertaining the real age of mutations that occurred many generations ago. Second is assessing the deleterious, beneficial, or neutral effects for mutations, which is unknown for the vast majority of human SNPs. The unexpectedly young age has been deduced only for several hundred mutations located in about 50 different loci that are associated with recent strong positive selection in the human genome [15,16,19]. Among the 22 strongest candidate loci for positive selection

in humans presented by Sabeti et al. [16] in Table 1, the authors characterized 41 possibly functional SNPs and additional closely located 439 SNPs, which are in strong linkage disequilibrium and propagate with these beneficial mutations by genetic hitchhiking. Because the set of characterized beneficial mutations is tiny compared to all known human SNPs, and because the set of hitchhiking SNPs is many fold larger than the set of known beneficial SNPs, it is impossible to evaluate the Maruyama effect from these principal public data sets. One of the most comprehensive experimental evaluations of the Maruyama effect has been reported by Keizun and co-authors [6]. Using whole-genome computation analysis the authors examined thousands of putatively deleterious missense SNPs inside protein coding sequences and compared them with synonymous mutations. The authors concluded that deleterious alleles are, on average, younger than neutral ones. However, the analysis was qualitative and did not provide a precise quantitative estimation of the Maruyama effect. Also, the influence of important genomic parameters on the mutation age (e.g., local recombination rates, coefficient of dominance for genes under analysis) has not been examined.

Presently, even with the availability of about 3,000 completed human genomes in public databases, there is a limitation of genomic data on families that includes sequences from members of several generations. Yet this kind of information is required for evaluating allelic age. However, in a few years, the technology race to develop a fast sequencing device with “\$100 per genome” capacity should be accomplished. With such technology, whole-genome sequencing analysis of large pedigrees will become routine. In addition, there are several long-running selection projects with laboratory animals (like mice and rats), where frozen materials from animals across numerous generations have been preserved [20]. The availability of cheap sequencing in the near future will provide unprecedented genomic data on extra-long pedigrees, across multiple generations of humans and other species.



**Fig. 3.** Mean allelic age of SNPs with different selection coefficients calculated for different experimental conditions. Allelic age was measured in generations passed after the SNP arrival. All analyzed SNPs had current population frequencies in the range from 40% to 60% (Panel A) or in the range from 10% to 30% (Panel B). Each error bar reflects 95% confidence interval (CI) for the respective experiment. The modeled selection coefficient ( $s$ ) for each SNP had three possible values: either +1 (for beneficial mutations), 0 (for neutral mutations), or -1 (for deleterious mutations). All computations were performed for a population size of  $N = 100$ ; twenty novel mutations per gamete ( $\mu = 20$ ) and five offspring per individual ( $\alpha = 5$ ). In each individual experiment, the modeled number of recombination events per gamete was either  $r = 1$  or  $r = 48$ ; the dominant coefficient for each gene was  $h = 0$  (dominant genes),  $h = 0.5$  (co-dominant genes), or  $h = 1$  (recessive genes). The distribution of mutations by selection coefficients was either from “Experiment B” (90% of starting SNPs neutral) or “Experiment C” (10% of starting SNPs neutral), as described in Fig. 1. Exact allelic age for each SNP in these experiments is provided in Supplementary Table S2.

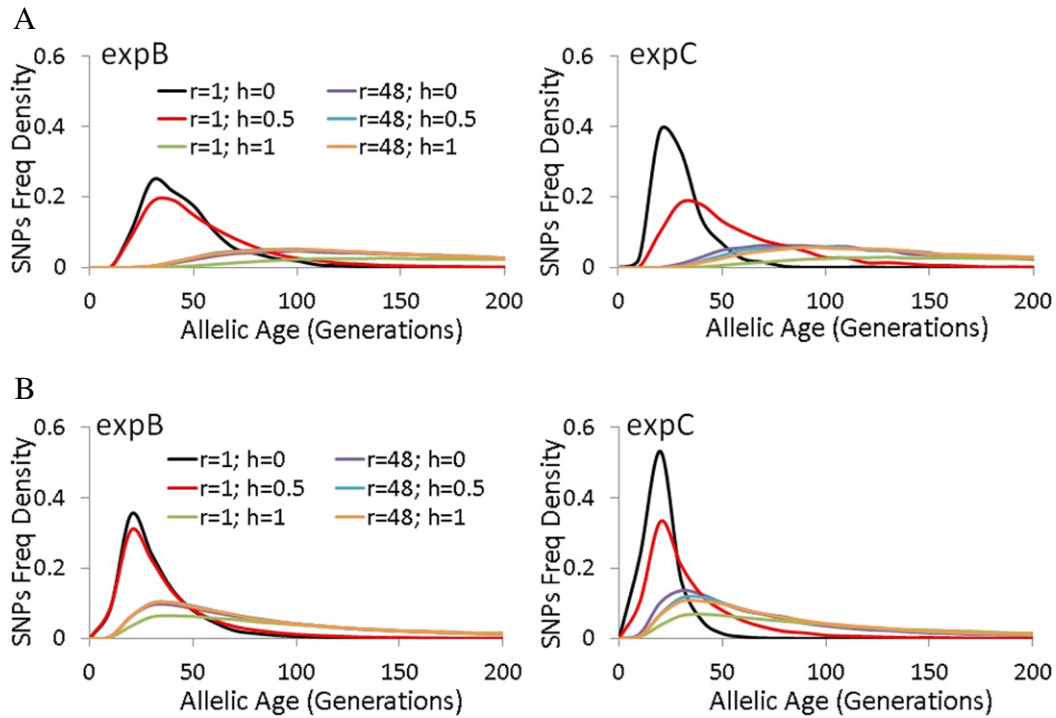
Such data open the possibility of a direct investigation of the fate and the age of many mutations. Hence, a precise estimation of the Maruyama predictions will soon be possible.

In this respect, mathematical modeling provides an important insight into this problem. However, existing mathematical approaches for inferring allelic age consider only one mutation at a time, while possible interactions of SNPs with one another have been ignored [7,10,12]. The “1000 Genomes” project recently revealed 38 million SNPs within the pool of sequenced genomes, and demonstrated that two non-related humans from the same population have over three million SNP differences between them [1]. Each human gene bears hundreds of SNPs, arranged in several major haplogroups having strong linkage disequilibrium between SNPs from the same haplotype. Since mutations never exist alone, to understand their dynamics they should be modeled/analyzed in the context of haplotypes. Keeping this in mind, we implemented whole-genome computational simulations to

investigate how different haplogroup structures influence the average age of SNPs. In our GEMA simulations, we used the lowest estimated value of the influx of novel mutations observed in humans (20 novel mutations per gamete) [3,8,9]. Such an influx, even in a very small population of 100 modeled individuals, generates thousands of SNPs randomly distributed among 600 genes. Closely located mutations are linked together and form haplotypes. The length of the haplotypes depends on the recombination rate ( $r$ ).

These haplotypes compete with one another via natural selection. Each non-neutral mutation contributes to the total fitness of the model individual, which is calculated by taking into account all beneficial and deleterious mutations and the dominance coefficients ( $h$ ) of the genes in the modeled genomes. In our simulations we applied the ultimate selection mode, in which only the fittest offspring survive and form the next generation. Our computations demonstrated that the recombination rate, dominance coefficient, and overall distribution





**Fig. 4.** Distribution of SNPs by their age. Panel A—shows all SNPs having frequencies from 40% to 60% in the population. Panel B—all SNPs having frequencies of 10% to 30% in the population. Each curve represents an experiment with specific *h* and *r* parameters. The exact numbers of SNPs in the experiments are provided in the Table 1 and Supplementary Table S2.

**Table 1**  
Distributions of beneficial, deleterious, and neutral SNPs by their ages. The parameters provided in the table are the following: *s*—the selection coefficients of SNPs; *r*—recombination rate; *h*—dominance coefficient; <age> ± CI—mean allelic age and its 95% confidence interval; STDEV—standard deviation for the distribution of SNPs by their ages; #SNPs—number of SNPs; and “F Range”—the SNP population Frequency range for the analyzed SNPs (“high” = 40%–60%; “low” = 10%–30%).

Experiment C							Experiment B						
<i>r</i>	<i>h</i>	<i>s</i>	<Age> ± CI	STDEV	No. SNPs	F range	<i>r</i>	<i>h</i>	<i>s</i>	<Age> ± CI	STDEV	No. SNPs	F range
1	0.5	-1	51.7 ± 1.4	34.5	2417	High	1	0.5	-1	46.3 ± 2.6	28.5	471	High
		0	51.1 ± 3.1	34.6	480				0	49.5 ± 0.3	31.0	34241	
		1	51.7 ± 2.4	32.7	732				1	47.6 ± 1.5	30.1	1476	
48	0.5	-1	160.5 ± 3.2	108.4	4338	High	48	0.5	-1	130.5 ± 32.1	83.4	26	High
		0	176.4 ± 5.5	121.6	1881				0	187.9 ± 0.8	132.1	92944	
		1	160 ± 3.4	108.5	3803				1	83 ± 1.6	48.8	3417	
1	0.5	-1	29.9 ± 0.6	26.5	8467	Low	1	0.5	-1	26.3 ± 0.8	21.5	2985	Low
		0	30.2 ± 1.3	26.0	1480				0	31.7 ± 0.1	26.0	128074	
		1	32.8 ± 1.2	28.7	2041				1	32.1 ± 0.8	25.8	4166	
48	0.5	-1	83.7 ± 1	80.1	26786	Low	48	0.5	-1	40.7 ± 1	30.3	3265	Low
		0	100.2 ± 2.5	98.5	6148				0	105.3 ± 0.3	106.4	462951	
		1	97.8 ± 2.1	93.8	7923				1	52.2 ± 0.9	42.4	9415	
1	0	-1	25 ± 0.9	12.9	878	High	1	0	-1	27.4 ± 6.2	10.0	10	High
		0	25.1 ± 2.3	14.0	145				0	40.7 ± 0.8	20.9	2800	
		1	24 ± 1.9	11.6	149				1	39.9 ± 9.4	24.0	25	
48	0	-1	145.7 ± 2.5	99.5	6175	High	48	0	-1	118.2 ± 56.4	70.5	6	High
		0	152 ± 4.7	104.3	1857				0	209.3 ± 2.2	152.1	17704	
		1	149.2 ± 4.1	102.7	2458				1	100.4 ± 8.5	57.2	173	
1	0	-1	15.6 ± 0.2	8.3	4893	Low	1	0	-1	20 ± 1.9	12.2	161	Low
		0	15.8 ± 0.6	8.5	832				0	27 ± 0.3	18.4	13366	
		1	15.6 ± 0.6	8.5	752				1	24.8 ± 2.7	16.6	144	
48	0	-1	76.4 ± 0.9	77.4	26512	Low	48	0	-1	54.6 ± 16.7	87.1	105	Low
		0	85.6 ± 2.4	87.9	5208				0	115.9 ± 1	122.2	61609	
		1	85.8 ± 2.2	85.9	5871				1	65.9 ± 5.1	63.4	599	
1	1	-1	335.6 ± 3.2	262.7	25814	High	1	1	-1	384.7 ± 13.5	314.3	2077	High
		0	333.2 ± 8.2	266.2	4076				0	376 ± 3	306.0	40146	
		1	329.1 ± 7.1	257.4	5111				1	380.4 ± 17.2	316.6	1295	
48	1	-1	171.4 ± 3	114.2	5655	High	48	1	-1	121.3 ± 18.4	74.6	63	High
		0	184.5 ± 5.3	121.9	2018				0	194.8 ± 2	138.3	18575	
		1	165.5 ± 3.5	110.5	3789				1	98.8 ± 3.9	62.9	1018	
1	1	-1	184.9 ± 1.6	222.2	76851	Low	1	1	-1	199.2 ± 5.7	257.5	7813	Low
		0	185.4 ± 4.2	220.3	10558				0	201.7 ± 1.6	258.1	99795	
		1	188.4 ± 4.2	227.7	11271				1	204.2 ± 11.7	265.1	1960	
48	1	-1	91.3 ± 0.9	87.8	35368	Low	48	1	-1	67 ± 2.1	53.2	2455	Low
		0	104.1 ± 2.5	102.5	6596				0	108.7 ± 0.9	110.2	58373	
		1	100.2 ± 2.2	98.4	7825				1	52.5 ± 2.5	44.6	1216	

of the entire pool of SNPs by their selection coefficients significantly affect the mean allelic age of SNPs. The Maruyama effect [11] was detected only when the recombination rate was high ( $r = 48$ ) and the neutral mutations were overabundant (90% of SNPs are neutral in Experiment B). Under these conditions, the average age of neutral mutations was 1.4 times higher than deleterious and 2.3 times higher than beneficial ones (Fig. 3 and Table 1). However, under the same conditions ( $r = 48$  and  $h = 0.5$ ) if the frequency of neutral mutations is decreased to 10% and the frequencies (but not ratio) of deleterious and beneficial mutations are increased (experiment C), the average ages of mutations with different selection coefficients were practically the same (no Maruyama effect).

Our results demonstrate the fruitfulness of the whole-genome computational simulation approach for population genetics, and its benefits over mathematical modeling. All in all, GEMA programs allow investigation of the integrative effects of thousands of mutations per individual, and evaluation of the effects of grouping of mutations into haplotypes.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ygeno.2015.02.005>.

### Acknowledgments

We are grateful to Dr. Robert Blumenthal, UT, for his valuable discussions and comments. We also appreciate the financial support from the Department of Medicine to conduct our research.

### References

- [1] G.R. Abecasis, D. Altshuler, A. Auton, L.D. Brooks, R.M. Durbin, et al., A map of human genome variation from population-scale sequencing, *Nature* 467 (2010) 1061–1073.
- [2] N. Arnheim, P. Calabrese, M. Nordborg, Hot and cold spots of recombination in the human genome: the reason we should find them and how this can be achieved, *Am. J. Hum. Genet.* 73 (2003) 5–16.
- [3] D.F. Conrad, J.E.M. Keebler, M.A. DePristo, S.J. Lindsay, Y.J. Zhang, et al., Variation in genome-wide mutation rates within and between human families, *Nat. Genet.* 43 (2011) 712–U137.
- [4] E. Genin, A. Tullio-Pelet, F. Begeot, S. Lyonnet, L. Abel, Estimating the age of rare disease mutations: the example of Triple-A syndrome, *J. Med. Genet.* 41 (2004) 445–449.
- [5] R.C. Griffiths, S. Tavaré, The ages of mutations in gene trees, *Ann. Appl. Probab.* 9 (1999) 567–590.
- [6] A. Kiezun, S.L. Pulit, L.C. Francioli, F. van Dijk, M. Swertz, et al., Deleterious alleles in the human genome are on average younger than neutral alleles of the same frequency, *PLoS Genet.* 9 (2013) e1003301.
- [7] M. Kimura, T. Ohta, The age of a neutral mutant persisting in a finite population, *Genetics* 75 (1973) 199–212.
- [8] A.S. Kondrashov, S.A. Shabalina, Classification of common conserved sequences in mammalian intergenic regions, *Hum. Mol. Genet.* 11 (2002) 669–674.
- [9] H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences, *Nature* 475 (2011) 493–496.
- [10] W.H. Li, The first arrival time and mean age of a deleterious mutant gene in a finite population, *Am. J. Hum. Genet.* 27 (1975) 274–286.
- [11] T. Maruyama, The age of a rare mutant gene in a large population, *Am. J. Hum. Genet.* 26 (1974) 669–673.
- [12] T. Maruyama, The age of an allele in a finite population, *Genet. Res.* 23 (1974) 137–143.
- [13] S. Qiu, A. McSweeney, S. Choulet, A. Saha-Mandal, L. Fedorova, et al., Genome evolution by matrix algorithms: cellular automata approach to population genetics, *Genome Biol. Evol.* 6 (2014) 988–999.
- [14] B. Rannala, J.P. Reeve, High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence, *Am. J. Hum. Genet.* 69 (2001) 159–178.
- [15] P.C. Sabeti, D.E. Reich, J.M. Higgins, H.Z. Levine, D.J. Richter, et al., Detecting recent positive selection in the human genome from haplotype structure, *Nature* 419 (2002) 832–837.
- [16] P.C. Sabeti, P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, et al., Genome-wide detection and characterization of positive selection in human populations, *Nature* 449 (2007) 913–918.
- [17] M. Slatkin, B. Rannala, Estimating the age of alleles by use of intraallelic variability, *Am. J. Hum. Genet.* 60 (1997) 447–458.
- [18] M. Slatkin, B. Rannala, Estimating allele age, *Annu. Rev. Genomics Hum. Genet.* 1 (2000) 225–249.
- [19] B.F. Voight, S. Kudaravalli, X. Wen, J.K. Pritchard, A map of recent positive selection in the human genome, *PLoS Biol.* 4 (2006) e72.
- [20] U. Wisloff, S.M. Najjar, O. Ellingsen, P.M. Haram, S. Swoap, et al., Cardiovascular risk factors emerge after artificial selection for low aerobic capacity, *Science* 307 (2005) 418–420.