Novel Populations from Simulated Admixed Populations

Shah Yunn Naing[1,2], Eugene WJ Thia[1,2], Daeun Roh[1,2], Cedric Chew[1,2], Sandi Kyaw Tun[1,2], Mee Khin Wai[1,2] and Maurice HT Ling[1,2,3,4]*

[1]School of Life Sciences, Management Development Institute of Singapore, Singapore

[2]School of Applied Sciences, Northumbria University, United Kingdom

[3]School of Data Sciences, Perdana University, Malaysia

[4]HOHY PTE LTD, Singapore

*Corresponding Author: Maurice HT Ling, School of Life Sciences, Management Development Institute of Singapore, Singapore; School of Applied Sciences, Northumbria University, United Kingdom; School of Data Sciences, Perdana University, Malaysia; HOHY PTE LTD, Singapore.

Abstract

Admixtures of two relatively distinct populations; as a result of clashing, mixing, and merging; can drastically affect its population genetics. Studies have suggested that admixed populations are instrumental in establishing novel populations. Computer simulations is a common method to study population admixtures. Although it is plausible to study the emergence of novel populations from simulated admixed populations, studies in this area have been sparse. Here, we attempt to demonstrate the emergence of novel populations from admixed populations using simulation. Our results show that all admixed populations have the potential to result in the emergence of novel populations despite large majority (up to 90%) of one of the two source populations. The null hypothesis of no significant allelic changes can be rejected with a p-value of 5.3E-05. Therefore, our simulation study supports current studies suggesting that admixed populations are instrumental in establishing novel populations.

Keywords: Population genetics; Forward simulation; Island; Admixture populations; Island

Introduction

Mixing of populations resulting in genetic admixtures is an important process in evolution [1] as it is one of the fastest evolutionary processes to affect the genetic structure of a population [2] and have been observed in many species; including, human [3], fish [4], cat [5], cattle [6], and moth [7]. This can lead to a rapid genetic diversification, which may result in rapid reduction of detrimental mutation load [8], improved health [9, 10], and disease resistance [11]. There are many recent studied examples of human population admixtures [12-14]; including, Denisovans and Neanderthals leading to Asian population [15], hunter-gatherers and Neolithic farmers in Iron Age Europe [16], northern and southern China [17], Latin America [18], and the entire Asia at large [19]. There are also studies suggesting that admixed populations are instrumental in establishing novel populations [20-22].

One of the recent methods to study population admixtures is using simulations and comparing simulation results to laboratory or empirical findings [15, 23-25]. The two main approaches of simulation [26] are forward time simulation, where the evolution of initial population is tracked under various genetic models over multiple generations; and backward time (also known as coalescent simulation), where the simulation begins with the current time and work backwards into history. Hence, it is plausible to study the emergence of novel populations from simulated admixed populations but studies in this area have been sparse as recent simulation studies mainly focus on the admixture models [27-31] or determining population history [29, 31-34] and adaptations to novel environments [35].

Therefore, direct evidence suggesting that emergence of novel populations can originate from admixed populations is lacking. In this study, we examine the genetic structure and possible novel population emergence of using various admixtures from two simulated populations. Our results show that all admixtures have the potential to result in the emergence of novel populations, which is consistent with current studies [20-22].

Methods

Ancestral tribe definition. Mixtures of two ancestral populations, Tribe A and Tribe B, we simulated for 100 generations using Island [36], which had been used in previous studies [37, 38]. Thirty marker genes (M1 to M30) were defined for both tribes and 10 alleles (A1 to A10) were defined for each marker gene. Ancestral Tribe A consisted of organisms with high proportion of A1 to A5 (19.8% of each allele) with low proportion of A6 to A10 (0.2% of each allele) for all 50 marker genes. Inversely, ancestral Tribe B consisted of organisms with low proportion of A1 to A5 (0.2% of each allele) with high proportion of A6 to A10 (19.8% of each allele) for all 50 marker genes. Each ancestral tribe consisted of 10 thousand individuals. Specification of allelic compositions and commands for generating for ancestral Tribes A and B, as well as the generated ancestral Tribes A and B, are provided in supplementary materials.

Simulation and analysis. Ten admixed source populations were generated by randomly selecting individuals from the ancestral populations in the following proportions: (a) 10% Tribe A 90% Tribe B, (b) 20% Tribe A 80% Tribe B, (c) 30% Tribe A 70% Tribe B, (d) 40% Tribe A 60% Tribe B, (e) 50% Tribe A 50% Tribe B, (f) 60% Tribe A 40% Tribe B, (g) 70% Tribe A 30% Tribe B, (h) 80% Tribe A 20% Tribe B, and (i) 90% Tribe A 10% Tribe B. Each admixed population consisted of 1000 individuals. Ten replicates of one hundred generations were simulated for each admixed source population and the population size was kept constant at 1000 individuals during simulation. Normalized Chi-Square statistic [37], which is the quotient of Chi-square statistic and number of alleles, was used to analyze the simulated populations based on the expected allelic frequencies of ancestral Tribe A and Tribe B separately. The commands for simulation and analysis are provided in supplementary materials.

Results and Discussion

Population genetics changed over time in 1:1 admixed population suggesting potential emergence of novel populations. We hypothesized that the population genetics of 1:1 admixed population should be relatively stable. However, regression analysis on 110 data points (11 generations of 10 replicates) yields the following regression equation: Normalized Chi-Square Statistic = 22.116 Generation + 4924.3 when using the allelic frequencies of Tribe A as expected frequencies (Figure 1A) and Normalized Chi-Square Statistic = 22.572 Generation + 4828.3 when using the allelic frequencies of Tribe B as expected frequencies (Figure 1B). Both gradients are significantly larger than zero (p-value ≤ 12E-13), which suggests that the population structure is deviating with increasing generations from the expected population genetics of either Tribe A or B. This implies that the null hypothesis expecting that the population genetics of 1:1 admixed population should be stable with respect to a control population is rejected and suggests the potential of novel population emergence with 1:1 admixed population.

It may be plausible that the pseudorandom number generator used in Island [36] may have an impact on the simulation. An analysis of the codes of Island [36] shows that the pseudorandom number generator is from random module in Python Standard Library, which implements Mersenne Twister [39], which has a period of 2**19937-1. Mersenne Twister has been shown by several studies to be suit-

able for most Monte Carlo simulation studies [40-42]. The same random module was used in other tools; such as, DOSE [43, 44]; where null hypotheses have been reported [45, 46]. Although there has been reported cases where Mersenne Twister [42, 47-49] were found not to be suitable, this is unlikely the case as a recent study using Island [36] has also demonstrated instances of null hypothesis [37].
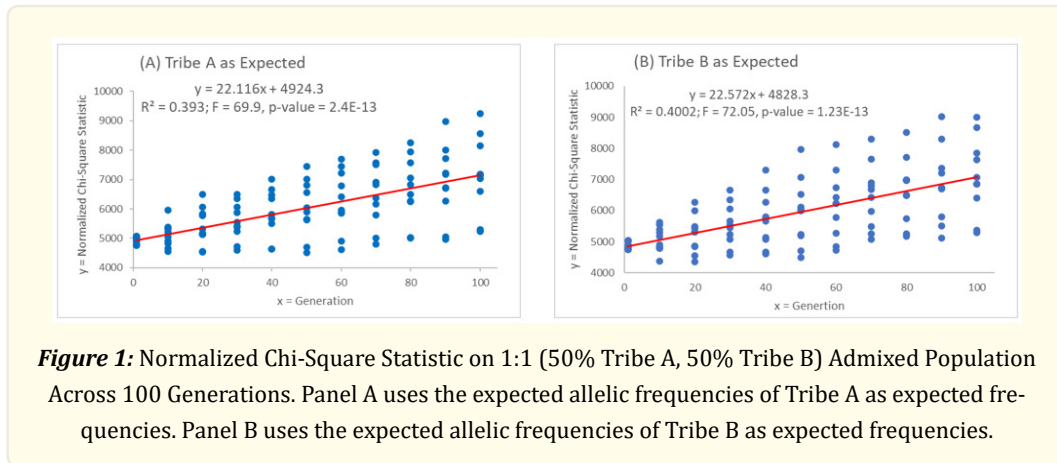

Figure 1: Normalized Chi-Square Statistic on 1:1 (50% Tribe A, 50% Tribe B) Admixed Population Across 100 Generations. Panel A uses the expected allelic frequencies of Tribe A as expected frequencies. Panel B uses the expected allelic frequencies of Tribe B as expected frequencies.

Potential emergence of novel populations from all admixtures. Given that novel populations may emerge from 1:1 admixture, it is plausible to conceive potential emergence of novel populations from all non-1:1 admixed population. Using the allelic frequency of Tribe A as the standard, our results show a significant decrease (r-square = 0.8148, t = -5.549, p-value = 8.61E-4) in the gradient the mean normalized Chi-Square statistic across 100 Generations with increasing proportion of Tribe A (Figure 3A). The same trend is observed when the allelic frequencies of simulated populations were matched with the expected allelic frequencies of ancestral Tribe B (r-square = 0.9846, t = -21.179, p-value = 1.32E-07; Figure 3B). This suggests that all admixed populations deviate from their respective ancestral populations with increasing generations. Furthermore, this deviation from ancestral population is observed even at large majority - 90% of Tribe A or 90% of Tribe B.

An analysis of the allelic frequencies (Table 1) shows that average of 6.78% of the alleles show significant changes (p-value < 0.05) between the 1st and 100th generation, with a standard deviation of 1.749%. Hence, the null hypothesis of no significant allelic changes can be rejected with a p-value of 5.3E-05. In addition, the highest percentages of significant allelic changes are observed at 90% of Tribe A or 90% of Tribe B, which further supports that all admixed populations deviate from their respective ancestral populations with increasing generations with potential emergence of novel populations. This is consistent with studies suggesting that admixed populations are instrumental in establishing novel populations [20-22].
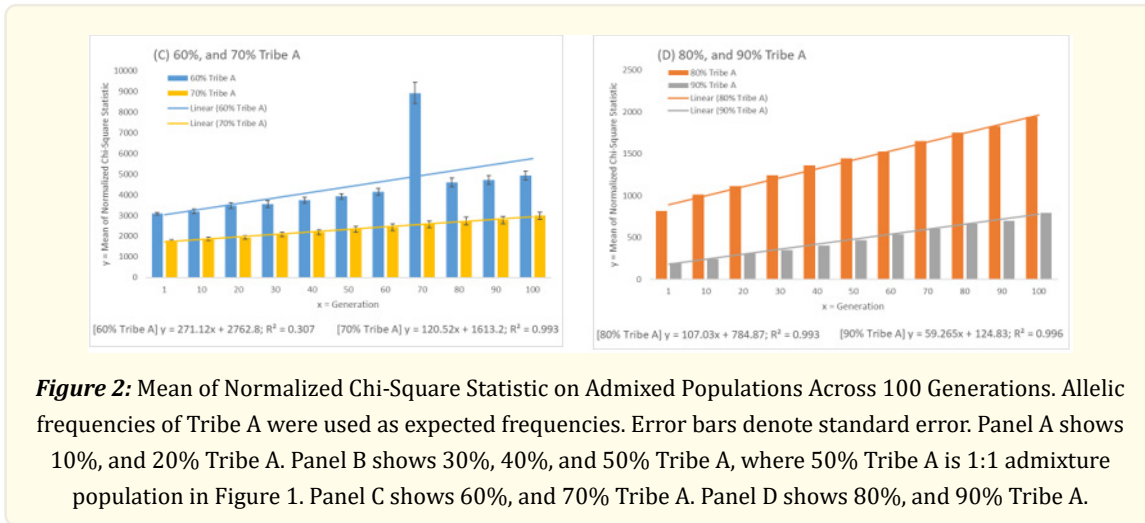
Figure 2: Mean of Normalized Chi-Square Statistic on Admixed Populations Across 100 Generations. Allelic frequencies of Tribe A were used as expected frequencies. Error bars denote standard error. Panel A shows 10%, and 20% Tribe A. Panel B shows 30%, 40%, and 50% Tribe A, where 50% Tribe A is 1:1 admixture population in Figure 1. Panel C shows 60%, and 70% Tribe A. Panel D shows 80%, and 90% Tribe A.
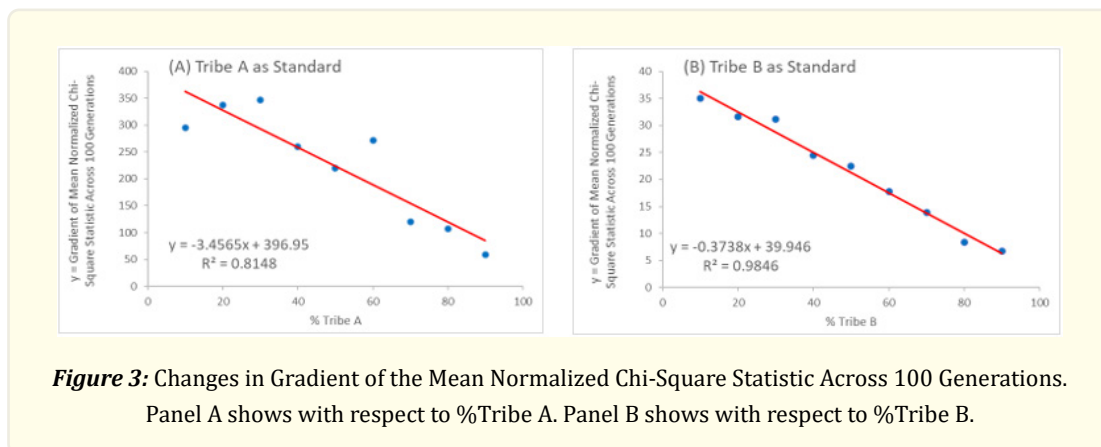

Figure 3: Changes in Gradient of the Mean Normalized Chi-Square Statistic Across 100 Generations. Panel A shows with respect to %Tribe A. Panel B shows with respect to %Tribe B.

% Tribe A	% Tribe B	Number (Percentage)		
		Significant Allelic Changes	*Significant Increases*	*Significant Decreases*
10	90	26 (8.67%)	4 (1.33%)	22 (7.33%)
20	80	17 (5.67%)	3 (1.00%)	14 (4.67%)
30	70	18 (6.00%)	4 (1.33%)	14 (4.67%)
40	60	21 (7.00%)	0 (0.00%)	21 (7.00%)
50	50	24 (8.00%)	6 (2.00%)	18 (6.00%)
60	40	13 (4.33%)	3 (1.00%)	10 (3.33%)
70	30	20 (6.67%)	6 (2.00%)	14 (4.67%)
80	20	15 (5.00%)	3 (1.00%)	12 (4.00%)
90	10	29 (9.67%)	4 (1.33%)	25 (8.33%)

Table 1: Number of Significant Allelic Changes from 1st to 100th Generation.

Conclusion

Our simulation results support current studies [20-22] suggesting that admixed populations are instrumental in establishing novel populations.

Supplementary Materials

Data files for this study can be downloaded at http://bit.ly/AdmixturePopulations.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Rius M and Darling JA. "How Important is Intraspecific Genetic Admixture to the Success of Colonising Populations?". Trends in Ecology & Evolution 29.4 (2014): 233-242.

2. Korunes KL and Goldberg A. "Human Genetic Admixture". PLOS Genetics 17.3 (2021): e1009374.

3. Yang MA and Fu Q. "Insights into Modern Human Prehistory Using Ancient Genomes". Trends in genetics: TIG 34.3 (2018): 184-196.

4. Jaisuk C and Senanan W. "Effects of Landscape Features on Population Genetic Variation of a Tropical Stream Fish, Stone Lapping Minnow, Garra cambodgiensis, in the Upper Nan River Drainage Basin, Northern Thailand". PeerJ 6 (2018): e4487.

5. Howard-McCombe J., et al. "On the Use of Genome-Wide Data to Model and Date the Time of Anthropogenic Hybridisation: An Example from the Scottish Wildcat". Molecular Ecology 30.15 (2021): 3688-3702.

6. Manomohan V., et al. "Legacy of Draught Cattle Breeds of South India: Insights into Population Structure, Genetic Admixture and Maternal Origin". PloS One 16.5 (2021): e0246497.

7. Elameen A., et al. "Large-Scale Genetic Admixture Suggests High Dispersal in an Insect Pest, the Apple Fruit Moth". PloS One 15.8 (2020): e0236509.

8. Simons YB and Sella G. "The Impact of Recent Population History on the Deleterious Mutation Load in Humans and Close Evolutionary Relatives". Current Opinion in Genetics & Development 41 (2016): 150-158.

9. Norris ET., et al. "Genetic ancestry, admixture and health determinants in Latin America". BMC Genomics 19.S8 (2018): 861.

10. Martínez-Magaña JJ., et al. "The Identification of Admixture Patterns Could Refine Pharmacogenetic Counseling: Analysis of a Population-Based Sample in Mexico". Frontiers in Pharmacology 11 (2020): 324.

11. Hamid I., et al. "Rapid Adaptation to Malaria Facilitated by Admixture in the Human Population of Cabo Verde". eLife 10 (2021): e63177.

12. Hellenthal G., et al. "A Genetic Atlas of Human Admixture History". Science 343.6172 (2014): 747-751.

13. Gopalan S., et al. "Human Genetic Admixture Through the Lens of Population Genomics". Philosophical Transactions of the Royal Society B: Biological Sciences 377.1852 (2022): 20200410.

14. Cuadros-Espinoza S., et al. "The Genomic Signatures of Natural Selection in Admixed Human Populations". The American Journal of Human Genetics 109.4 (2022): 710-726.

15. Browning SR., et al. "Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture". Cell 173.1 (2018): 53-61.e9.

16. Översti S., et al. "Human Mitochondrial DNA Lineages in Iron-Age Fennoscandia Suggest Incipient Admixture and Eastern Introduction of Farming-Related Maternal Ancestry". Scientific Reports 9.1 (2019): 16883.

17. Yang MA., et al. "Ancient DNA Indicates Human Population Shifts and Admixture in Northern and Southern China". Science 369.6501 (2020): 282-288.

18. Norris ET., et al. "Admixture-Enabled Selection for Rapid Adaptive Evolution in the Americas". Genome Biology 21.1 (2020): 29.

19. Xu S. "Human Population Admixture in Asia". Genomics & Informatics 10.3 (2012): 133-144.

20. Li H-S., et al. "Population Admixture Can Enhance Establishment Success of the Introduced Biological Control Agent Cryptolaemus montrouzieri". BMC Evolutionary Biology 18.1 (2018): 36.

21. Consuegra S., et al. "Winning the Invasion Roulette: Escapes from Fish Farms Increase Admixture and Facilitate Establishment of Non-Native Rainbow Trout". Evolutionary Applications 4.5 (2011): 660-671.

22. Duchen P., et al. "Demographic Inference Reveals African and European Admixture in the North American Drosophila melanogaster Population". Genetics 193.1 (2013): 291-301.

23. Yang X., et al. "AdmixSim: A Forward-Time Simulator for Various Complex Scenarios of Population Admixture". Frontiers in Genetics 11 (2020): 601439.

24. Schumer M, Powell DL and Corbett-Detig R. "Versatile Simulations of Admixture and Accurate Local Ancestry Inference with mixnmatch and ancestryinfer". Molecular Ecology Resources 20.4 (2020): 1141-1151.

25. Jin W., et al. "Exploring Population Admixture Dynamics via Empirical and Simulated Genome-wide Distribution of Ancestral Chromosomal Segments". The American Journal of Human Genetics 91.5 (2012): 849-862.

26. Yuan X., et al. "An Overview of Population Genetic Data Simulation". Journal of Computational Biology 19.1 (2012): 42-54.

27. Pfaffelhuber P and Rohde A. "A Central Limit Theorem Concerning Uncertainty in Estimates of Individual Admixture". Theoretical Population Biology 148 (2022): 28-39.

28. Swart Y., et al. "GWAS in the Southern African Context". PloS One 17.9 (2022): e0264657.

29. Fortes-Lima CA., et al. "Complex Genetic Admixture Histories Reconstructed with Approximate Bayesian Computation". Molecular Ecology Resources 21.4 (2021): 1098-1117.

30. Ravagni S, Sanchez-Donoso I and Vilà C. "Biased Assessment of Ongoing Admixture Using STRUCTURE in the Absence of Reference Samples". Molecular Ecology Resources 21.3 (2021): 677-689.

31. Toyama KS, Crochet P-A and Leblois R. "Sampling Schemes and Drift Can Bias Admixture Proportions Inferred by Structure". Molecular Ecology Resources 20.6 (2020): 1769-1785.

32. Nakamichi R, Kitada S and Kishino H. "Exploratory Analysis of Multi-Trait Coadaptations in Light of Population History". Ecology and Evolution 12.3 (2022): e8755.

33. Kumari G., et al. "Genetic Diversity and Population Genetic Structure Analysis of an Extensive Collection of Wild and Cultivated Vigna Accessions". Molecular Genetics and Genomics 296.6 (2021): 1337-1353.

34. Allen A., et al. "Genetic Evidence Further Elucidates the History and Extent of Badger Introductions from Great Britain into Ireland". Royal Society Open Science 7.4 (2020): 200288.

35. Shang L., et al. "High Genetic Diversity with Weak Phylogeographic Structure of the Invasive Spartina alterniflora (Poaceae) in China". Frontiers in Plant Science 10 (2019): 1467.

36. Ling MH. "Island: A Simple Forward Simulation Tool for Population Genetics". Acta Scientific Computer Sciences 1.2 (2019): 20-22.

37. Johny A, Sumedha PR and Ling MH. "Simulation Suggests that One-Off Simple Supplementation from the Wild into Captive Population May Not Increase Captive Genetic Diversity". EC Veterinary Science 6.7 (2021): 107-111.

38. Kamarudin NJ., et al. "A Simulation Study on the Effects of Founding Population Size and Number of Alleles Per Locus on the Observed Population Genetic Profile: Implications to Broodstock Management". EC Veterinary Science 5.8 (2020): 176-180.

39. Matsumoto M and Nishimura T. "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudo-Random Number Generator". ACM Transactions on Modeling and Computer Simulation 8.1 (1998): 3-30.

40. Olteanu M and Paraschiv N. "The Influence of Random Numbers Generators upon Genetic Algorithms". 2013 IEEE INISTA (IEEE, Albena, Bulgaria) (2013).

41. Grauer JA. "A Comparison of Three Random Number Generators for Aircraft Dynamic Modeling Applications". (NASA Langley Research Center, Hampton, Virginia, USA), 20170005202 (TM-2017-219612) (2017).

42. Harase S. "On the F2-Linear Relations of Mersenne Twister Pseudorandom Number Generators". Mathematics and Computers

in Simulation 100 (2014): 103-113.

43. Castillo CF and Ling MH. "Digital Organism Simulation Environment (DOSE): A Library for Ecologically-Based in Silico Experimental Evolution". Advances in Computer Science: An International Journal 3.1 (2014): 44-50.

44. Castillo CF and Ling MH. "Digital Organism Simulation Environment (DOSE) Version 1.0.4". Current STEM 1 (2018): 1-106.

45. Castillo CFG and Ling MHT. "Resistant Traits in Digital Organisms Do Not Revert Preselection Status Despite Extended Deselection: Implications to Microbial Antibiotics Resistance". BioMed Research International (2014): 648389.

46. Castillo CF., et al. "Resistance Maintained in Digital Organisms Despite Guanine/Cytosine-Based Fitness Cost and Extended De-Selection: Implications to Microbial Antibiotics Resistance". MOJ Proteomics & Bioinformatics 2.2 (2015): 00039.

47. Shcherbyna Y, Kazakova N and Fraze-Frazenko O. "The Mersenne Twister Output Stream Postprocessing". Proceedings of the 3rd International Conference on Information Security and Information Technologies (ISecIT 2021) (Odesa, Ukraine) (2021): 9.

48. Gelencsér-Horváth A., et al. "Fast, Parallel Implementation of Particle Filtering on the GPU Architecture". EURASIP Journal on Advances in Signal Processing 1 (2013): 148.

49. Xin G, Wang P and Jiao Y. "Multiscale Quantum Harmonic Oscillator Optimization Algorithm with Multiple Quantum Perturbations for Numerical Optimization". Expert Systems with Applications 185 (2021): 115615.

Volume 4 Issue 1 January 2023