Simulating the effects of migration rates on Neolithic range expansion

Neha J. Angal and Christopher R. Tillquist

Anthropology, University of Louisville

UL of

Abstract
In this simulation study, we investigated how migration rates, fecundity, and carrying capacity may have influenced the development of clines during a range expansion. Using NEMO, an open source simulation environment for population genetics, we simulated range expansions, which typically generate a diversity cline. Migration occurred as a 2D–stepping–stone model with migration rates of 0.2 and 0.3. We modeled ten biallelic loci, tracked average heterozygosity for 1024 demes, and sampled from the first generation where the lattice filled and the final generation. Sampling at the time when all demes are first-filled, revealed independent effects of migration rates, fecundity, and carrying capacity on lattice-wide heterozygosity.

Background
Continental colonization events can result in differential spatial distributions of diversity due to serial founder effects and allele surfing during range expansion events. Establishing which factors mostly strongly impact diversity during an expansion may provide insights into colonization processes. These processes and resulting distributions have been investigated for several decades in Europe using complex models and simulation studies, with a particular emphasis placed upon determining if Europe was peopled in the late Paleolithic or the Neolithic (Arenas et al. 2012, Arenas et al. 2013, Klopfstein et al. 2006, Rendine et al. 1986, Barbunjani et al. 1995, Wegman et al. 2006). However, many of these studies relied on small lattices that are less representative of continental space, and utilized programs that are not publically available. Hence results that are not reproducible, and relied upon parameters that may not be realistic for modeling human populations.

Objectives and Hypotheses
The goal of this study was to determine what distributions of diversity develop in the context of a range expansion, which parameters drive those distributions, and how long signals can persist. We expected that migration rate would have the strongest apparent effect upon distributions of diversity.

Materials & Methods
Forward simulations were performed using NEMO (Guillaume et al. 2006). Result files were processed with gawk scripts, PGDSpider (Lischer and Excoffier 2012), and analyzed using Arlequin 3.5. Files were post-processed with gawk scripts and Vim. Heterozygosities were calculated using the dplyr package (Wickham and Francois 2015) and plotted in R using the heatmaps.2 module of gplots (Warnes et al., 2016). Processing and plotting of Arlequin results files were computed using R version 3.2.4 (Very Secure Dishes; R Core Development Team).

Experimental Design
Two sets of simulations were run for experimental and control scripts: a set of one stage and a set of two stage models. Experimental scripts modeled range expansions across a mostly empty lattice, and control scripts modeled instantaneously filled lattices. Both lattices were 32 by 32 for a total of 1024 demes. Ten biallelic loci were modeled for each simulation, and all ten were maximally polymorphic at the start, such that allele frequencies were set at 0.5. One stage models varied migration rates (either 0.2 or 0.3), mean fecundity (either 3 or 4), and carrying capacity (either 200 or 500), and ran for 500 generations. Two stage models varied the same parameters and corresponding values, with the first stage lasting for 250 generations and the second stage lasting for 500 generations, for a total of 750 generations. Twenty replicate runs were completed for each simulation. For each experimental simulation, results files for the first generation at which the lattice was filled (determined by each patch containing at least 25 individuals) and the final generation were obtained; corresponding files for control simulations at the same generation were also obtained. Parameter choices were selected based on previous simulations in the literature.
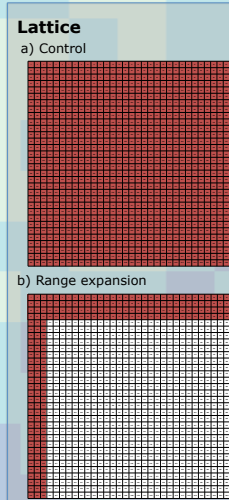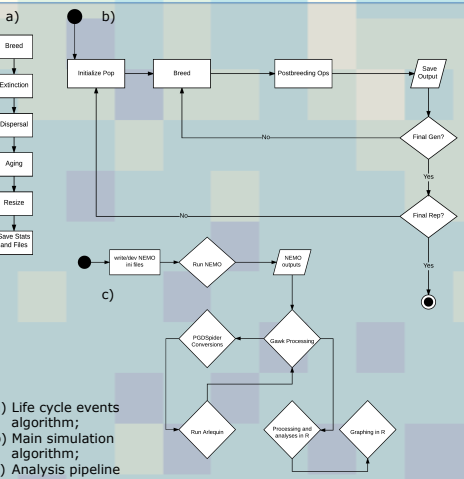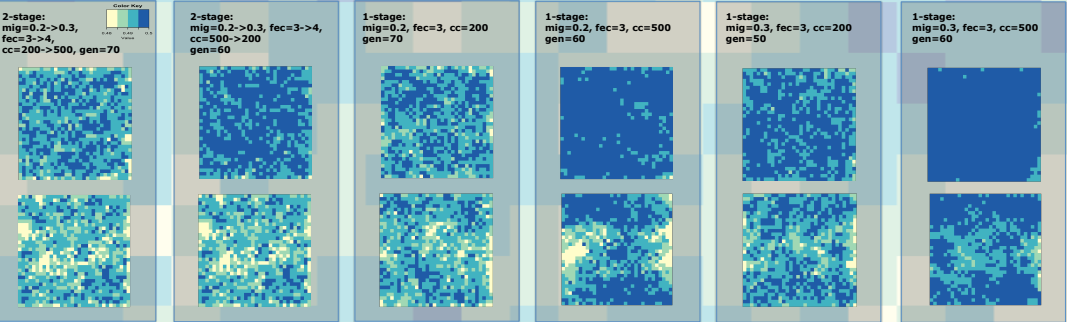
Results
Across all comparisons of controls versus simulated range expansions, only heatmaps of the first-generation-to-fill differed. While at the final generations there were no differences between control and experimental simulations, there was variability in the degree of diversity with experimental simulation showing lower diversity. With regard to the one stage models, all differed between control and experimental simulations except for the simulations where migration was 0.3, mean fecundity was 4, and carrying capacity was 500.

The results of varied combinations of parameters differed, and in all cases lower diversity was observed in the experimental relative to the control simulations. The one exception was the set of simulations were migration was 0.2, mean fecundity was 3, and carrying capacity was 200. In these, both the control and experimental simulations had low diversity at generation 500 and appeared very similar to one another. As a whole, the one stage models documented the influence of each parameter separately on diversity. Holding other variables constant: higher carrying capacity resulted in higher diversity; higher migration resulted in higher diversity; and higher fecundity resulted in higher diversity, with the most marked difference observed when both carrying capacity and fecundity were high.

Summary
Simulations of range expansion models using variable migration rates, fecundity, and carrying capacities all resulted in loss of diversity. Parameter choice played a strong role in the overall results, such that a lower migration rate (m=0.2), lower fecundity (f=3), and lower carrying capacity (cc=200), resulted in lower diversity in demes outside of the origin. Generally, these patterns followed the theoretical expectations of a model of serial founder effects. Results from our simulations are in accord with empirical data, supporting a strong role of drift in generating human population structure within the context of continental colonization events.

Future planned simulations will work to tease out the interactions between the three parameters in the context of selection.

2-stage: mig=0.2->0.3, fec=3->4, cc=200->500, gen=70

2-stage: mig=0.2->0.3, fec=3->4, cc=500->200 gen=60

1-stage: mig=0.2, fec=3, cc=200 gen=70

1-stage: mig=0.2, fec=3, cc=500 gen=60

1-stage: mig=0.3, fec=3, cc=200 gen=50

1-stage: mig=0.3, fec=3, cc=500 gen=60

Lattice
a) Control

b) Range expansion

Model parameters

Lattice Model	Model	MR 1	MR 2	CC 1	CC 2	G 1	G 2	Gen Fill	Gen Done
2 stage models									
Range Expansion	2 stage	0.2	0.3	200	500	3	4	70	750
RE	2 stage	0.2	0.3	500	200	3	4	60	750
RE	2 stage	0.3	0.2	200	500	3	4	60	750
RE	2 stage	0.3	0.2	500	200	3	4	60	750
Control	2 stage	0.2	0.3	200	500	3	4	70	750
CN	2 stage	0.2	0.3	500	200	3	4	60	750
CN	2 stage	0.3	0.2	200	500	3	4	60	750
CN	2 stage	0.3	0.2	500	200	3	4	60	750
1 stage models									
Range Expansion	1 stage	0.2		200		3		70	500
RE	1 stage	0.2		200		4		20	500
RE	1 stage	0.2		500		3		60	500
RE	1 stage	0.2		500		4		20	500
RE	1 stage	0.3		200		3		50	500
RE	1 stage	0.3		200		4		20	500
RE	1 stage	0.3		500		3		60	500
RE	1 stage	0.3		500		4		20	500
Control	1 stage	0.2		200		3		70	500
CN	1 stage	0.2		200		4		20	500
CN	1 stage	0.2		500		3		60	500
CN	1 stage	0.2		500		4		20	500
CN	1 stage	0.3		200		3		50	500
CN	1 stage	0.3		200		4		20	500
CN	1 stage	0.3		500		3		60	500
CN	1 stage	0.3		500		4		20	500

a) Life cycle events algorithm;
b) Main simulation algorithm;
c) Analysis pipeline

References: Arenas, Miguel, et al. 2013 Influence of Admixture and Paleolithic Range Contractions on Current European Diversity Gradients. Mol. Biol. Evol. 30(1):57-61. Arenas, Miguel 2012 Simulation of Molecular Data under Diverse Evolutionary Scenarios. PLoS Computational Biology 8(5):1-8. Barbujani, Guido et al. 1995 Indo-European origins: A computer- simulation test of five hypotheses. American Journal of Physical Anthropology 96(2):109-132. Bocquet-Appel, Jean-Pierre et al. 2005 Estimates of Upper Palaeolithic meta-population size in Europe from archaeological data. Journal of Archaeological Science 32(11): 1656-1668. Cavalli-Sforza, L.L., et al. 1994 The History and Geography of Human Genes. Princeton: Princeton University Press. Currat, Mathias, and Laurent Excoffier 2005 The effect of the Neolithic expansion on European molecular diversity. Proceedings of the Royal Society of London B: Biological Sciences 272:(1564): 679-688. Edmonds, Christopher A., Anita S. Lillie and L. Luca Cavalli-Sforza. 2004 Mutations arising in the wave front of an expanding population. PNAS 101(4):975-979. Excoffier, Laurent, et al. 2009 Genetic Consequences of Range Expansions. Annu. Rev. Ecol. Evol. Syst. 40:481-501. Excoffier, L. and H.E. L. Lischer. 2010 Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567. Fisher, R. A. 1937 The Wave of Advance of Advantageous Genes. Annals of Eugenics 355-369. Fix, Alan G. 1999 Migration and Colonization in Human Microevolution. Cambridge and New York: Cambridge University Press. Fix, Alan G. 1996 Gene Frequency Clines in Europe-Demic Diffusion or Natural Selection? The Journal of the Royal Anthropological Institute 2(4):625-643. Francois, Olivier et al. 2010 Principal Component Analysis under Population Genetic Models of Range Expansion and Admixture. Mol. Biol. Evol. 27(6):1257-1268. Gamble, Clive et al. 2005 The archaeological and genetic foundations of the European population during the Late Glacial: implications for 'agricultural thinking'. Cambridge Archaeological Journal 15(02):193-223. Guillaume, Frédéric and Jacques Rougemont. 2006 Nemo: an evolutionary and population genetics programming framework. Bioinformatics 22(20):2556-2557. Hoban, Sean et al. 2012 Computer simulations: tools for population and evolutionary genetics. Nature Reviews Genetics 13:110-122. Hofer, T. et al. 2009 Large Allele Frequency Differences between Human Continental Groups are more Likely to have Occured by Drift During Range Expansions than by Selection. Annals of Human Genetics 73:95-108. Klopfstein, Seraina et al. 2006 The Fate of Mutations Surfing on the Wave of a Range Expansion. Mol. Biol. Evol. 23(3):482-490. Lischer, Hel., and Excoffier L. 2012 PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298-299. R Core Team 2013 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. Ray, Nicolas et al. 2003 Intra-deme molecular diversity in spatially expanding populations. Mol. Biol. Evol. 20(1):76-86. Rendine, S., A. Piazza, and L. L. Cavalli-Sforza. 1986 Simulation and separation by principal components of multiple demic expansions in Europe. American Naturalist 1986: 681-706. Wegman, Daniel et al. 2006 Molecular Diversity After a Range Expansion in Heterogeneous Environments. Genetics 174:2009-2020.