UC Davis
UC Davis Previously Published Works

Title
Allele frequency dynamics in a pedigreed natural population

Permalink
https://escholarship.org/uc/item/9zs6376k

Journal
Proceedings of the National Academy of Sciences of the United States of America, 116(6)

ISSN
0027-8424

Authors
Chen, Nancy
Juric, Ivan
Cosgrove, Elissa J
et al.

Publication Date
2019-02-05

DOI
10.1073/pnas.1813852116

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Allele frequency dynamics in a pedigreed natural population

Nancy Chen[a,b,c,1], Ivan Juric[d], Elissa J. Cosgrove[e], Reed Bowman[f], John W. Fitzpatrick[g], Stephan J. Schoech[h], Andrew G. Clark[e], and Graham Coop[a,b]

[a]Center for Population Biology, University of California, Davis, CA 95616; [b]Department of Evolution & Ecology, University of California, Davis, CA 95616; [c]Department of Biology, University of Rochester, Rochester, NY 14627; [d]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44195; [e]Department of Molecular Biology & Genetics, Cornell University, Ithaca, NY 14850; [f]Avian Ecology Program, Archbold Biological Station, Venus, FL 33960; [g]Cornell Lab of Ornithology, Cornell University, Ithaca, NY 14850; and [h]Department of Biological Sciences, University of Memphis, Memphis, TN 38152

Edited by Nils C. Stenseth, University of Oslo, Oslo, Norway, and approved November 19, 2018 (received for review August 10, 2018)

A central goal of population genetics is to understand how genetic drift, natural selection, and gene flow shape allele frequencies through time. However, the actual processes underlying these changes—variation in individual survival, reproductive success, and movement—are often difficult to quantify. Fully understanding these processes requires the population pedigree, the set of relationships among all individuals in the population through time. Here, we use extensive pedigree and genomic information from a long-studied natural population of Florida Scrub-Jays (*Aphelocoma coerulescens*) to directly characterize the relative roles of different evolutionary processes in shaping patterns of genetic variation through time. We performed gene dropping simulations to estimate individual genetic contributions to the population and model drift on the known pedigree. We found that observed allele frequency changes are generally well predicted by accounting for the different genetic contributions of founders. Our results show that the genetic contribution of recent immigrants is substantial, with some large allele frequency shifts that otherwise may have been attributed to selection actually due to gene flow. We identified a few SNPs under directional short-term selection after appropriately accounting for gene flow. Using models that account for changes in population size, we partitioned the proportion of variance in allele frequency change through time. Observed allele frequency changes are primarily due to variation in survival and reproductive success, with gene flow making a smaller contribution. This study provides one of the most complete descriptions of short-term evolutionary change in allele frequencies in a natural population to date.

population genetics | pedigrees | fitness | gene flow | genetic drift

An evolving natural population is essentially a vast pedigree, with genetic material transmitted down this pedigree following the laws of Mendelian inheritance (except in rare cases of meiotic drive). We often cannot directly observe the actual processes underlying genetic change. Instead, population genetic studies typically rely on current day patterns of genetic variation—or, if temporal samples are available, the variation in allele frequencies through time—to make inferences about the effects of genetic drift, natural selection, and gene flow in driving evolutionary change. However, these evolutionary mechanisms can be precisely understood in terms of the differential genetic contributions of individuals to the population pedigree over time, combined with the stochasticity of Mendelian segregation.

Knowledge of the population pedigree allows us to trace expected individual genetic contributions, i.e., the expected number of copies of a neutral allele contributed by a given individual, to the population in future generations. Individual genetic contributions can be estimated analytically (1–3) or via gene dropping simulations, i.e., simulations of Mendelian transmission of alleles down their pedigree of descendants (4). The long-term expected genetic contribution of an individual is an individual's reproductive value, a general measure of individual fitness (5–7). Indeed, the reproductive value of an individual influences many aspects of the survival of an individual's genotype, from the probability of loss of a new, weakly beneficial mutation to the complex distribution of genomic blocks passed on to future generations (8).

Analyses of known pedigrees have been used to estimate individual genetic contributions to assess founder effects in human populations (1–3, 9–11) and to predict the probability of gene loss in captive breeding populations (4, 12). Also, empirical pedigree calculations have long been used to understand genetic models of human diseases (13) and are increasingly used in natural populations to understand the genetic basis of quantitative trait variation, fitness consequences of inbreeding, and much more (14). To date, empirical pedigrees and gene dropping approaches have been rarely used to study the temporal spread and loss of individual alleles (15–18).

Here, we combine genomic data with a known population pedigree to describe and predict allele frequency change at many loci in an exhaustively sampled free-living population of Florida Scrub-Jays (*Aphelocoma coerulescens*) at Archbold Biological Station. Intensive study since 1969 has resulted in lifetime fitness measures for thousands of individuals on an extensive pedigree. Recently,

Significance

Evolution is change in the genetic composition of populations. In nature, individuals reproduce, die, and move among populations, leading to changes in the population frequency of the alleles they carry. Here we study an extensive family tree (pedigree) for an exhaustively sampled natural population of Florida Scrub-Jays to directly characterize the mechanisms underlying how genetic material is transmitted to future generations. We link individual fitness with long-term genetic contributions and quantify the relative roles of evolutionary processes governing allele frequency change. This study clearly illustrates how short-term evolutionary change occurs within a natural population.

Author contributions: N.C., A.G.C., and G.C. designed research; N.C., E.J.C., R.B., J.W.F., and S.J.S. performed research; N.C. and I.J. contributed new reagents/analytic tools; N.C., E.J.C., R.B., J.W.F., and S.J.S. performed data collection; N.C. and G.C. analyzed data; and N.C. and G.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

Data deposition: All data and code used in this study can be found at Figshare, 10.6084/m9.figshare.7044368.

See Commentary on page 1834.

[1]To whom correspondence should be addressed. Email: nancy.chen@rochester.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1813852116/-/DCSupplemental.

Published online December 31, 2018.

Chen et al. (19) generated genome-wide single-nucleotide polymorphism (SNP) data for nearly every individual in the population over the past two decades to demonstrate how decreased immigration into the population led to increased levels of inbreeding and decreased mean fitness over time. In this study, we link individual lifetime reproductive success with long-term genetic contributions and allele frequency change. We show how the population pedigree presents a powerful opportunity to directly elucidate the relative roles of drift, gene flow, and selection in governing allele frequency dynamics over time.

Results

Individual Fitness and Genetic Contributions. First, we consider a series of inferences that can be made purely with the pedigree, ignoring the SNP genotypes for the moment. We estimated fitness for all 926 individuals who bred in our study population in 1990–2013 and were born before 2002 (the age cohorts who are all dead by the end of 2014). Lifetime reproductive success was highly variable in our study population: The total number of nestlings produced over an individual's lifetime ranged from 0 to 43, with 197 individuals (21%) producing no nestlings despite having at least one breeding attempt (*SI Appendix*, Fig. S1). Only 43% produced any grand-offspring (range 0 to 189), and 33% produced great-grand-offspring (range 0 to 210). As might be expected in these monogamous birds in which the sexes experience equal annual mortality (20, 21), we found no significant differences in individual fitness between males and females (Wilcoxon rank sum test, $p > 0.65$ for all three measures of fitness).

Using the detailed population pedigree, we calculated both the genealogical and expected genetic contribution of each individual to the study population from 1990 to 2013. Fig. 1 *A* and *B* shows results for two illustrative males, both of whom first bred in 1994. Male A lived until 2006 and had 41 offspring, whereas Male B only lived until 2000 and had 7 offspring. We define an individual's genealogical contribution to a given year as the proportion of nestlings in the birth cohort who are genealogically descended from the focal individual, while an individual's expected genetic contribution is the expected proportion of alleles at a locus in the nestling cohort that comes from the focal individual. Beyond a few generations, few genealogical descendants are expected to inherit any genetic material, so the number of genealogical descendants should quickly outnumber the number of genetic descendants. Fig. 1 *A*–*C* nicely demonstrates this pattern in our data, providing empirical illustration for a substantial body of theory on the relationship between genetic and genealogical ancestry (8, 22, 23). An individual's genealogical contribution in 2013 is correlated with its expected genetic contribution in 2013 (Spearman's $\rho = 0.99$, $p < 2 \times 10^{-16}$), but its genealogical contribution is significantly larger (paired Wilcoxon test, $p < 2 \times 10^{-16}$; Fig. 1*C*).

Individual fitness is a central concept in evolutionary biology but is notoriously difficult to measure (24). Here, we tested for a relationship between various proxies for fitness and the expected genetic contribution to the population. All three measures of fitness (number of offspring, grand-offspring, and great-grand-offspring) are significantly correlated with both the total

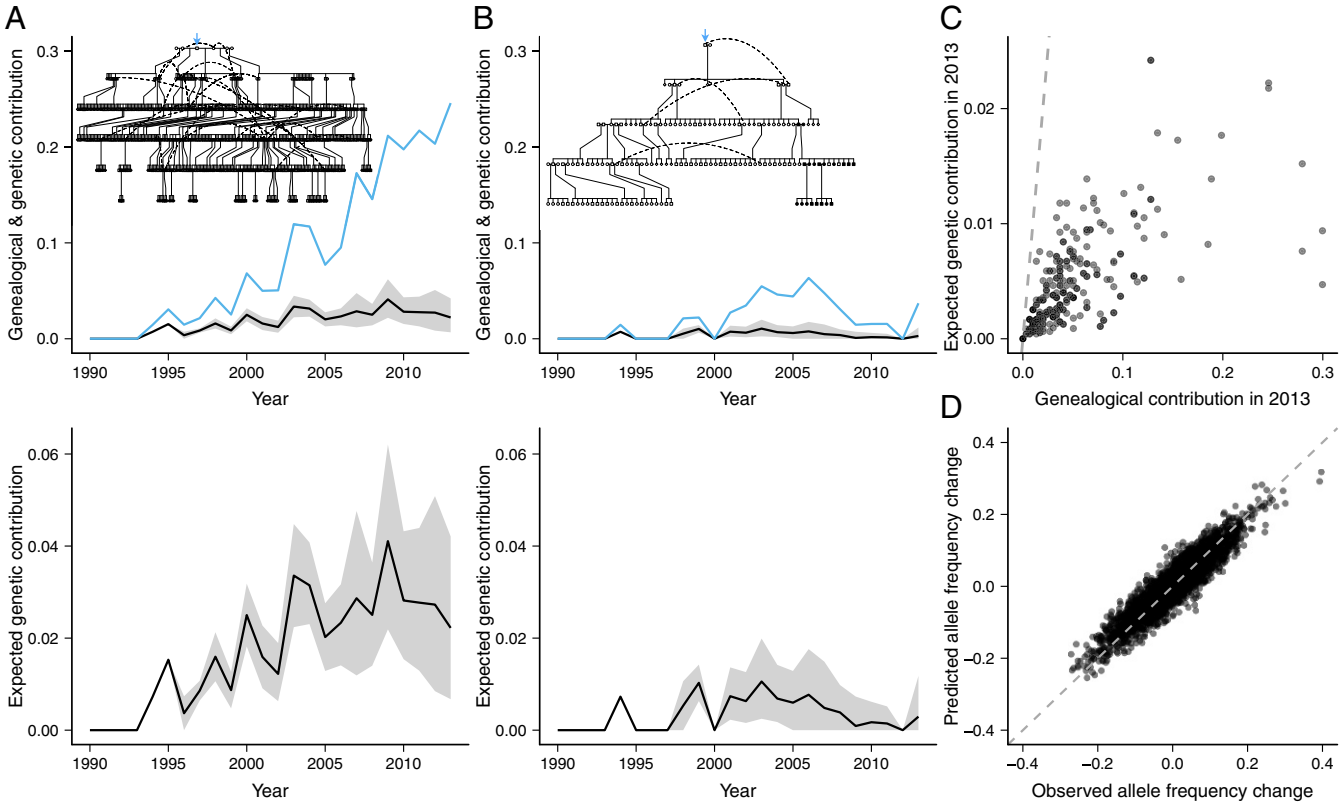

Fig. 1. Genealogical (*Top*) and expected genetic contributions (*Bottom*) to the study population over time for two males who first bred in 1994 with total lifetime reproductive success of (*A*) 41 and (*B*) 7. Blue lines indicate the proportion of nestlings each year who are genealogical descendants. Black lines indicate mean expected genetic contribution for each year, and gray shading is the 95% confidence interval for their contribution at a neutral locus. The pedigree of all descendants of each individual in the study population is shown, with an arrow indicating the focal individual, and solid symbols denoting individuals still alive in 2013. (*C*) Genealogical contributions and expected genetic contributions to the population in 2013 for all breeders born before 2002 who first bred in 1990 or later (926 individuals). The dotted line indicates a one-to-one relationship. (*D*) Predicted versus observed change in allele frequencies from 1999 to 2013.

expected genetic contribution from 1990 to 2013 (Spearman's $\rho = 0.92, 0.85, 0.78$, respectively; $p < 2 \times 10^{-16}$ for each comparison) and the expected genetic contribution to the 2013 nestling cohort (Spearman's $\rho = 0.57, 0.83, 0.87$, respectively; $p < 2 \times 10^{-16}$ for each comparison; *SI Appendix*, Fig. S2). The correlation between individual fitness and expected genetic contribution in 2013 increases with the number of generations considered in the measure of fitness.

Allele Frequency Predictions. In previous work, we genotyped >80% of all adults and nearly every nestling born in 1989–1991, 1995, and 1999–2013 (3,404 individuals total) at 10,731 autosomal SNPs (19). Here, we investigate allele frequency dynamics in the birth cohort from 1999 to 2013. In theory, we should be able to predict the allele frequency of a particular SNP in a given year simply by summing the individual genetic contributions of each founder to the population that year weighted by the founder's genotype at that SNP. Note that immigrants are considered founders, so this approach incorporates gene flow. We generated allele frequency predictions for each autosomal SNP in 1999–2013. We can nearly perfectly predict the allele frequency for each SNP in any given year ($\beta = 0.99$). More importantly, we can predict the overall net change in allele frequencies from 1999 to 2013 ($\beta = 0.87$; Fig. 1*D*).

Effect of Gene Flow. Previous work showed high levels of immigration into our study population, with immigrants comprising 32 to 55% of all breeding adults in a given year (19). We estimated the cumulative expected genetic contribution of new immigrants appearing in our study population from 1991 onward (Fig. 2*A*). Total expected genetic contributions of individual immigrant cohorts in 2013 range from 0.003 to 0.083 and are significantly correlated with the number of individuals in that cohort (Spearman's $\rho = 0.52$, $p = 0.01$). Immigrants arriving since 1990 are, in aggregate, expected to contribute 75% of the alleles present in the 2013 nestling cohort. We fitted a model to project the contributions of immigrants into the future (*SI Appendix*, Fig. S3). We predict that it takes, on average, 32 y for 95% of neutral alleles to be replaced by immigration.

With the high expected genetic contribution of immigrants, we predicted that gene flow could play an important role in governing allele frequency trajectories over time. While the majority of SNPs show small frequency changes, we do observe a few large allele frequency shifts over this 15-y time period: The difference in allele frequencies between 1999 and 2013 is >0.15 for 129 SNPs and >0.2 for 11 SNPs. We used gene dropping simulations to model the expected allele frequency distributions at each SNP in the nestling cohorts from 1999 to 2013. Unlike our previous pedigree-based simulations to generate individual genetic contributions, here we began simulations with the observed founder genotypes for each SNP. The mean allele frequency of these gene dropping simulations is equal to the allele frequency predictions generated above.

Indeed, we found that gene flow alone can cause large allele frequency shifts (one example is shown in Fig. 2*B*). This allele increased in frequency by 0.26 between 1999 and 2013, yet the observed allele frequency trajectory lies well within expectations from our gene dropping simulations. For this SNP, the allele frequency in incoming immigrants significantly increased over time (Mann–Kendall test, $p = 0.002$), from 0.51 in the 1990 founders to 0.71 in immigrants appearing in 2013, likely causing the population allele frequency to increase as well. As gene dropping begins with founder genotypes, any change in allele frequency due to incoming immigration is reflected in the simulation results. In the absence of data on the pedigree and the genotypes of immigrants, such trajectories could resemble selection, but our gene-dropping approach shows that these large changes in allele frequencies are actually likely the result of gene flow.

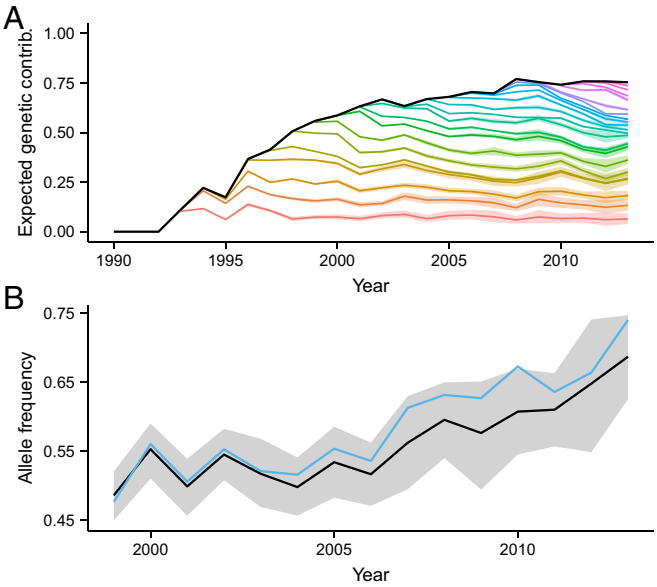

Fig. 2. (*A*) The expected genetic contribution of different cohorts of recent immigrants (based on the year they were first observed in our population). The black line shows the total expected genetic contribution of immigrants appearing in the population after 1990. Each colored line shows the mean added contribution of successive cohorts of immigrants, with shading to show the 95% confidence intervals. (*B*) Observed (blue) and simulated (black) allele frequencies over time for an SNP with significantly increasing immigrant allele frequencies.

Short-Term Selection. Given that our gene dropping simulations accurately account for the effects of both gene flow and drift, we then tested for significant net allele frequency changes from 1999 to 2013 as well as between all adjacent years during this time period. We compared observed allele frequency shifts to the expected distribution of allele frequency shifts generated from the gene dropping simulations (Fig. 3*A*). At a false discovery rate (FDR) of 0.25, 18 SNPs showed significant changes in allele frequency between 1999 and 2013 (*SI Appendix*, Table S1 and Fig. 3). For allele frequency shifts between adjacent years, we find some hits if we treat each year as an independent test (*SI Appendix*, Table S2 and Fig. S4); no SNPs survived multiple testing correction across years. The gene dropping simulations provide a good fit to observed data (*SI Appendix*, Fig. S5), except perhaps in 2001–2002. The largest fire in our study area during the 50-y study occurred in February 2001, at the height of a severe drought (25), and this unusually intense fire resulted in elevated competition in unburned territories (26) and may have led to stronger selection in 2001–2002. We observed slightly elevated adult mortality (33% compared with mean 26% for this time period) as well as lower productivity and later breeding in 2001 (27). However, overall, our results suggest that allele frequency change in our population from 1999 to 2013 is largely consistent with a neutral model.

Variance in Allele Frequencies Through Time. Finally, to quantify the relative roles of different evolutionary processes in shaping patterns of genetic variation genome-wide, we constructed a model for the variance in allele frequency change in 1999–2013. We assume that allele frequencies change due to just three processes: differential survival of individuals, immigration, and reproduction. We partitioned the proportion of allele frequency

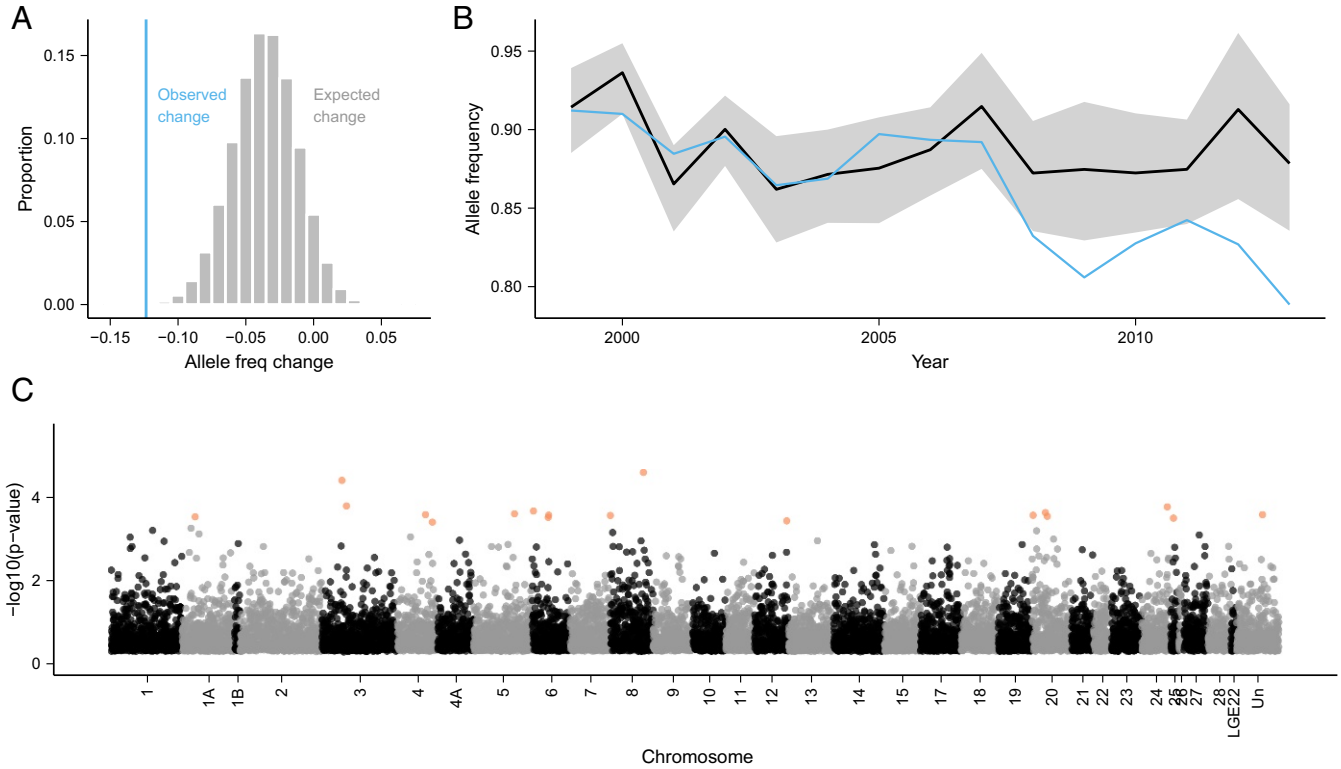

Fig. 3. (*A*) Distribution of expected allele frequency shifts between 1999 and 2013 for the SNP shown in *B* (gray histogram). The blue line indicates the observed allele frequency change. (*B*) Observed (blue) and simulated (black) allele frequency trajectories for one of the significant SNPs in 1999–2013. Gray bars indicate 95% confidence intervals for the gene dropping simulations. (*C*) Manhattan plot for allele frequency shifts in 1999–2013. Significant SNPs (FDR < 0.25) are highlighted in orange.

change from year to year due to survival/reproduction and gene flow using a model that accounts for variation in population sizes over time and overlapping generations (Fig. 4). The change in allele frequency due to births is a result of both variation in family size and Mendelian segregation of alleles in heterozygotes. We further divided the variance in allele frequency change due to births into these two components and found that the noise due to Mendelian segregation comprises 24 to 48% of the variance due to births, and 12 to 23% of the overall variance. Our model results reflect patterns we observed in the field. For instance, the number of nestlings born in 2012 was unusually low (*SI Appendix*, Fig. S6), leading the survivors to have a disproportionate impact on allele frequency variation in 2011–2012. Overall, we found that 90% of the variance in allele frequencies is driven by variation in survival and reproductive success (fitness) among individuals. If variation in fitness is heritable, then the effects of drift can be compounded over the generations, even at unlinked loci (28–30). Simulations of allele frequency change on pedigrees in which we randomized family sizes over breeding individuals showed that heritable variation in reproductive success has no detectable effect on the variance in allele frequency change in 1999–2013: The mean difference between randomized and observed pedigrees was −0.8%, with a 95% confidence interval of (−11.3%, 8.7%). These results suggest that drift is the predominant force driving allele frequency change over time, which is consistent with our small population size.

Discussion

We capitalized on a long-term demographic study of Florida Scrub-Jays with extensive pedigree and genomic data to demonstrate how short-term evolutionary processes operate in a natural population. We estimated genealogical and expected genetic contributions for hundreds of individuals, and linked genetic contributions to both individual fitness measures and allele frequency change over time. In our population of Florida Scrub-Jays, we observed huge variation in individual fitness: 75% of the 445 individuals who first bred in our population before 1997 have no living descendants by 2013, but 6 of these individuals are each genealogical ancestors to >25% of the birth cohort in 2013. However, many of these genealogical descendants receive little genetic material from a particular ancestor, thanks to the vagaries of Mendelian segregation and recombination during meiosis (8, 22, 23). Here, we empirically show how genealogical contributions outstrip expected genetic contributions after just a few generations.

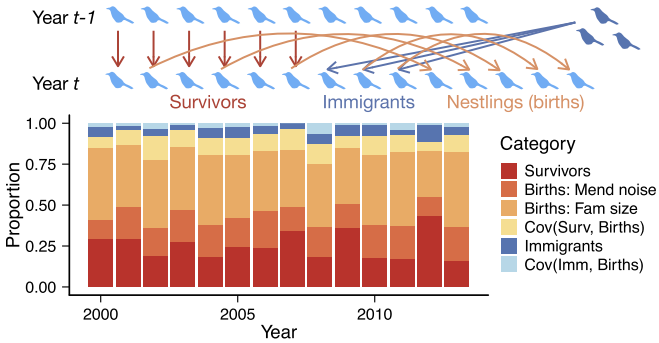

Fig. 4. Schematic and results for our model of the variance in autosomal allele frequency change from year to year due to survival (Surv)/reproduction (red/orange/yellow) or gene flow (Imm; blue). The variance in allele frequency change due to births is further partitioned into the variance due to variation in family size (Fam size) and additional noise due to Mendelian segregation of heterozygotes (Mend noise).

Individual fitness is defined as an individual's genetic contribution to future generations but is typically measured using single-generation proxies such as lifetime reproductive success. Similar to ref. 24, we found that lifetime reproductive success is correlated with an individual's expected genetic contribution to the population in the future. Florida Scrub-Jays rarely move once they become an established breeder on a territory, giving us confidence in our measures of total lifetime reproductive success. Our estimates of the total number of grand-offspring or great-grand-offspring, however, may be an underestimate, because a few of the individuals in our sample still have surviving offspring, and any descendants of emigrants are not counted. We believe the latter is a minor issue, because we know that emigration rates are extremely low from annual surveys of the surrounding areas. The correlation between the number of descendants and expected genetic contribution in 2013 is higher for fitness proxies that include more generations. Longer-term fitness proxies can be more accurate, in part, because they include variation in offspring quality (24), an idea we could explore by estimating the genetic correlation of the number of offspring and the number of grand-offspring (31).

The high expected genetic contribution of immigrants is consistent with previous results showing that immigrants play an important role in maintaining levels of genetic variation in the population (19). Genome-wide, allele frequency changes are primarily driven by variation in individual survival and reproduction. The contribution of new immigrants to allele frequency changes from year to year (Fig. 4) is much smaller than the cumulative expected genetic contribution of immigrants compounded over generations (Fig. 2*A*). This discrepancy occurs because, in our model, immigrants are included in allele frequency change only in the year they appear, while their genetic contributions to future years is folded into variation in survival and births. The change in allele frequencies we see due to variation in survival and births, except for the deviation due to Mendelian segregation of heterozygotes, includes the contribution of natural selection. Disentangling drift from selection would require testing for associations between reproductive success and individual phenotypes. Thus, these proportions should be viewed as including the contributions of both drift and selection to allele frequency change.

We used gene dropping to predict allele frequency changes over time for individual SNPs across the genome and showed that SNP trajectories can sometimes be strongly driven by gene flow. Our results emphasize the importance of knowing the underlying demography of population, as large allele frequency shifts that ordinarily may be attributed to selection could be due to processes such as drift and gene flow. Although we did detect signatures of selection changing allele frequencies in a few adjacent years, overall, we found little evidence of strong directional selection on single alleles on this short timescale.

One of the reasons why we detect so few selected loci is the accuracy with which we can predict allele frequency change from individual genetic contributions and observed founder genotypes. By conditioning on the population pedigree and founder genotypes, our gene dropping simulations appropriately accounted for variation in population sizes over time and relatedness within the birth cohort, as well as the effects of gene flow. One could argue that using gene dropping to test for selection is conservative, as the pedigree itself encodes information about variation in fitness. However, variation in offspring number is a natural part of genetic drift (32), while heritable variation in fitness at unlinked loci can act to compound genetic drift over the generations (28–30). Therefore, gene dropping simulations on the population pedigree provide the correct null model for heritable fitness variation for neutral alleles are that are unlinked to selected alleles. We further explored the performance of gene dropping for alleles linked to fitness by permuting families within each year and found that heritable variation in reproductive success does not significantly decrease variance in allele frequencies.

Here we have traced only single alleles down the pedigree. The incorporation of linkage and haplotype information would allow the quantification of realized, actual genetic contributions for each individual instead of just expected genetic contributions. By tracing the inheritance of genomic blocks down the pedigree, we could explore the relationship between reproductive value and the distribution of surviving genetic material, quantify the actual genetic contribution of recent immigrants across the genome, and pinpoint specific haplotypes linked to fitness. However, even single SNP analyses on a population pedigree provide substantial insights to the evolutionary forces governing allele frequency dynamics over time. As genomic resources for pedigreed populations expand, our ability to directly observe the causes and consequences of short-term evolution will increase dramatically.

Materials and Methods

Study System and Dataset. The Florida Scrub-Jay is a nonmigratory, cooperatively breeding bird restricted to oak scrub in Florida (33). A population of Florida Scrub-Jays has been intensively monitored at Archbold Biological Station (Venus, FL) for decades. Woolfenden, Fitzpatrick, Bowman, and colleagues began monitoring the northern half in 1969 (19, 33), and Mumme, Schoech, and colleagues began monitoring the southern half in 1989 (34, 35). All individuals in the entire population are uniquely banded, allowing identification of immigrant individuals each year. The entire population is censused every few months, and all nests of all family groups are closely monitored, providing documentation of survival and reproductive success for all individuals in the population. All fieldwork was approved by the Institutional Animal Care and Use Committees at Cornell University (IACUC 2010-0015), the University of Memphis (0667), and Archbold Biological Station (AUP-006-R) and permitted by the US Fish and Wildlife Service (TE824723-8, TE-117769), the US Geological Survey (banding permits 07732, 23098), and the Florida Fish and Wildlife Conservation Commission (LSSC-10-00205).

Because of the very low rate of extra-pair paternity and limited natal dispersal distances in this population (19–21, 36), we have a detailed and accurate population pedigree. To avoid any artifacts caused by study tract expansion before 1990, we began all our analyses in 1990 and truncated the pedigree accordingly. Our final pedigree consists of 6,936 individuals. We used the pedigree to estimate individual fitness for all adults who first bred in 1990 or later and were born before 2002 (926 individuals), by counting the total number of offspring, grand-offspring, or great-grand-offspring produced by a given individual over its lifetime. We restricted our sample to age cohorts of breeders who all died before the end of 2014 to ensure an accurate and unbiased survey of lifetime reproductive success. Of these individuals, 5% had offspring who were still alive at the end of 2014 and may produce additional grand-offspring, and 13% had grand-offspring who were still alive at the end of 2014 and therefore may produce additional great-grand-offspring. Here, we define offspring as 11-d-old nestlings (the age at which they are first banded).

For our genomic analyses, we focused on a core set of ∼68 territories in a geographic area that has been consistently monitored starting in 1990. In a previous study, we genotyped 3,984 individuals at 15,416 genome-wide SNPs, resulting in near-complete sampling of all nestlings and breeders in these core territories in 1989–1991, 1995, and 1999–2013 (19). Information on SNP discovery, genotyping, and pedigree verification can be found in ref. 19. Here, we removed SNPs with minor allele frequency <0.05. Our final dataset consists of 10,731 autosomal SNPs in 3,404 individuals. All data used in this study can be found at Figshare 10.6084/m9.figshare.7044368.

Expected Genetic Contributions. We quantify individual genetic contribution as the expected proportion of alleles in the nestling cohort that comes from the focal individual. The expected genetic contribution of an individual to a given year can be calculated as

$$G = \frac{1}{n}\sum_{m}\sum_{p}\left(\frac{1}{2}\right)^{g}, \quad [1]$$

where n is the total number of nestlings born that year, m is the number of nestlings related to the focal individual, p is the number of paths in the pedigree linking the focal individual and the nestling, and g is the number of generations separating the focal individual from the nestling in that path (1–3). We used pedigree-based simulations to estimate expected individual genetic contributions instead. Our simulation results match theoretical expectations but also provide estimates of the variance around the expected values.

We used gene dropping simulations to obtain expected genetic contributions of individual breeders and of different immigrant cohorts to our population over time. A founder is by definition any individual in the pedigree whose parents are unknown. Thus, all immigrants are founders. We assigned genotypes to all founders as follows: For individual simulations, we assigned the genotype "22" to the focal individual and "11" to all other founders. To assess the expected genetic contributions of different immigrant cohorts, we assigned immigrants appearing in different years different alleles. We then simulated Mendelian transmission of alleles down the pedigree 1,000,000 times using custom C code. The distribution of allele counts in the nestling cohort each year gives the distribution of expected genetic contributions over time.

Immigrant Contribution Projection. The proportion of resident alleles in the birth cohort over time [$r(t)$] can be written as

$$r(t) = (1 - i)^t, \quad [2]$$

where i is the per-year replacement rate by immigrant alleles, and t is the number of years following 1992. We began in 1992 because no parents in 1990–1992 were recent immigrants. We fitted this model using nonlinear least squares to estimate i, then used Eq. **2** to calculate the expected time until neutral alleles were 95% replaced by immigrant alleles.

Allele Frequency Predictions. In the absence of selection, the allele frequency of an autosomal SNP in any given year can be written as a function of the individual genetic contributions of each founder and the founder allele frequencies. Let F be the number of founders, $G_{i,y}$ be the expected genetic contribution of founder i to the population in year y, and p_i be the allele frequency of founder i. We can predict the expected allele frequency in year y as follows:

$$\hat{p}_y = \sum_{i=1}^{F} G_{i,y} p_i. \quad [3]$$

Here we iteratively trimmed the population pedigree until all founders were genotyped, and estimated individual genetic contributions using simulations on the trimmed pedigree. We evaluated prediction accuracy by fitting linear regressions.

Neutral Allele Dynamics. To generate expected allele frequency distributions over time, we used gene dropping simulations on a trimmed pedigree. For each SNP, we iteratively trimmed the pedigree until all founders in the final trimmed pedigree had a known genotype. Briefly, we removed all ungenotyped founders and set all offspring of these individuals as founders, then repeated these two steps until all remaining founders had observed genotypes. Note that the trimmed pedigree can differ across SNPs because of variable missing data across individuals; however, missing data rates are low (<5%), so these differences are slight. Using the observed genotype for each founder, we simulated Mendelian transmission of alleles down the pedigree a million times and estimated allele frequencies each year in genotyped nestlings from a core set of 54 to 76 territories.

We used Mann–Kendall tests from the R package Kendall (37) to test for trends in the allele frequencies of incoming immigrant cohorts through time. We tested for net directional selection between 1999 and 2013, as well as between all adjacent years during that time period, by comparing observed allele frequency shifts to the distribution of expected allele frequency shifts generated from the gene dropping simulations. For each test, we calculated p values by counting the number of simulations in which the simulated value is more different from the median value of all of the simulations compared with the observed value. We used an FDR threshold of 0.25 for significance.

Variance in Allele Frequencies Model. To quantify the proportion of variance in the change in allele frequencies due to gene flow and variation in individual survival and reproductive success, we modeled the population as follows: Adults who survive or immigrate into the population then produce offspring. From our detailed census and other population monitoring records, we generated a list of individuals present in our population each year in 1990–2013 and categorized them as survivors, immigrants, or nestlings (new births; *SI Appendix*, Fig. S6*A*). We only included an individual in a given year if it was observed in at least two months during March through June. We conservatively considered individuals who left our study population but later returned as survivors during the intervening time period to minimize inflating the variance in allele frequencies.

Let N_t be the total number of individuals in the population in year t, N_s be the number of individuals who survived from year $t-1$ to t, N_i be the number of new immigrants into the population in year t, and N_b be the number of individuals born in year t. Thus, the population size in year t is $N_t = N_s + N_i + N_b$. If we denote the allele frequencies in each category as p_j, then we can write the change in allele frequencies between years $t-1$ and t for a given SNP as

$$\Delta p = \frac{N_s}{N_t}(p_s - p_{t-1}) + \frac{N_i}{N_t}(p_i - p_{t-1}) + \frac{N_b}{N_t}(p_b - p_{t-1}). \quad [4]$$

The variance in allele frequency change over time is then

$$\begin{aligned}\mathrm{Var}(\Delta p) = &\left(\frac{N_s}{N_t}\right)^2 \mathrm{Var}(p_s - p_{t-1}) + \left(\frac{N_i}{N_t}\right)^2 \mathrm{Var}(p_i - p_{t-1}) \\ &+ \left(\frac{N_b}{N_t}\right)^2 \mathrm{Var}(p_b - p_{t-1}) + 2\frac{N_s N_b}{N_t^2}\mathrm{Cov}(p_s - p_{t-1}, p_b - p_{t-1}) \\ &+ 2\frac{N_i N_b}{N_t^2}\mathrm{Cov}(p_i - p_{t-1}, p_b - p_{t-1}). \quad [5]\end{aligned}$$

Note that we assume that survivors and immigrants in a given year are unrelated and accordingly set $\mathrm{Cov}(p_s - p_{t-1}, p_i - p_{t-1}) = 0$.

We further partitioned the change in allele frequency due to the birth cohort $\mathrm{Var}(p_b - p_{t-1})$ into the change due to variation in family size and the deviation due to Mendelian segregation of alleles from heterozygotes ($\Delta p_{b,\mathrm{mend}}$) (29). If p_m and p_f are the allele frequencies of the parents weighted by the number of offspring they produced in year t, then

$$p_b - p_{t-1} = \left(\frac{1}{2}(p_m + p_f) - p_{t-1}\right) + \Delta p_{b,\mathrm{mend}}, \quad [6]$$

where the first term denotes the expected change in allele frequencies due to the variation in family size, and the second term denotes the additional independent noise due to Mendelian transmission. We can then estimate the variance due to Mendelian noise as

$$\mathrm{Var}(\Delta p_{b,\mathrm{mend}}) = \mathrm{Var}\left(p_b - \frac{1}{2}(p_m + p_f)\right), \quad [7]$$

with the alternate term for the variance due to family size variation following from Eq. **6**.

We estimated each of the terms on the left and right sides of Eq. **5** averaged across all autosomal SNPs. We then divided each of the terms on the right by the total to quantify the proportion of allele frequency change due to which individuals survive to the focal year, appear as new immigrants, or are born, as well as the contribution of survivors and immigrants to the birth cohort and Mendelian segregation of heterozygotes. We verified our model using simulations.

Although we have genomic data from nearly every individual present in the population from 1999 to 2013, we still have a small number of ungenotyped individuals in each year (*SI Appendix*, Fig. S6*B*). To account for missing genotypes, we corrected each term in Eq. **5** for sampling. Normally, the error in allele frequency estimation due to sampling can be statistically modeled, but relatedness among individuals and nonrandom sampling make error estimation more complicated in this case. Therefore, we empirically calculated the error in allele frequency estimation using simulations.

To assess the effect of heritable variation in reproductive success on allele frequency change, we performed simulations that permuted parental assignments within each birth cohort, effectively breaking any inheritance of reproductive success while keeping variation in family sizes consistent. We then simulated genotypes for 10,000 loci on the randomized pedigree and the observed pedigree and compared the variance in net allele frequency change in 1999–2013. See *SI Appendix* for the full derivation of the model and more details on our simulations. All statistical analyses were done in R (38). All code is available from Figshare 10.6084/m9.figshare.7044368.

ACKNOWLEDGMENTS. We thank the many students and staff who collected the field data at Archbold Biological Station over the past half century.

Thanks go to Doc Edge for statistical help and Vince Buffalo for helpful discussions on pedigree randomization. Thanks go to the Coopons, Arvid Ågren, Jeff Jensen, and an anonymous reviewer for comments. This work was supported by National Science Foundation (NSF) Grants DEB0855879 and DEB1257628 and by the Cornell Lab of Ornithology Athena Fund. N.C. was supported by NSF Postdoctoral Fellowship in Biology 1523665. N.C. and G.C. acknowledge additional support from National Institutes of Health Grant R01-GM108779 and NSF Grants 1262327 and 1353380.

1. O'Brien E, Jorde LB, Rönnlöf B, Fellman JO, Eriksson AW (1988) Founder effect and genetic disease in Sottunga, Finland. *Am J Phys Anthropol* 77:335–346.
2. Roberts DF (1968) Genetic effects of population size reduction. *Nature* 220:1084–1088.
3. Roberts DF, Bear JC (1980) Measures of genetic change in an evolving population. *Hum Biol* 52:773–786.
4. MacCluer JW, VandeBerg JL, Read B, Ryder OA (1986) Pedigree analysis by computer simulation. *Zoo Biol* 5:147–160.
5. Fisher R (1930) *The Genetical Theory of Natural Selection* (Oxford Univ Press, Oxford).
6. Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics* (Roberts Company, Greenwood Village, CO), Vol 42.
7. Grafen A (2006) A theory of Fisher's reproductive value. *J Math Biol* 53:15–60.
8. Barton NH, Etheridge AM (2011) The relation between reproductive value and genetic contribution. *Genetics* 188:953–973.
9. O'Brien E, Kerber R, Jorde L, Rogers A (1994) Founder effect: Assessment of variation in genetic contributions among founders. *Hum Biol* 66:185–204.
10. Heyer E, Tremblay M (1995) Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. *Am J Hum Genet* 56:970–978.
11. Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS (2005) The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* 69:288–295.
12. Lacy RC (1989) Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents. *Zoo Biol* 8:111–123.
13. Thompson EA (1986) *Pedigree Analysis in Human Genetics* (Johns Hopkins Univ Press, Baltimore).
14. Pemberton J (2008) Wild pedigrees: The way forward. *Proc R Soc B* 275:613–621.
15. Gratten J, et al. (2012) Selection and microevolution of coat pattern are cryptic in a wild population of sheep. *Mol Ecol* 21:2977–2990.
16. Gao Z, Waggoner D, Stephens M, Ober C, Przeworski M (2015) An estimate of the average number of recessive lethal mutations carried by humans. *Genetics* 199:1243–1254.
17. Milot E, et al. (2017) Mother's curse neutralizes natural selection against a human genetic disease over three centuries. *Nat Ecol Evol* 1:1400–1406.
18. Kukekova AV, et al. (2018) Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. *Nat Ecol Evol* 2:1479–1491.
19. Chen N, Cosgrove EJ, Bowman R, Fitzpatrick JW, Clark AG (2016) Genomic consequences of population decline in the endangered Florida scrub-jay. *Curr Biol* 26:2974–2979.
20. Quinn JS, Woolfenden GE, Fitzpatrick JW, White BN (1999) Multi-locus DNA fingerprinting supports genetic monogamy in Florida scrub-jays. *Behav Ecol Sociobiol* 45:1–10.
21. Townsend AK, Bowman R, Fitzpatrick JW, Dent M, Lovette IJ (2011) Genetic monogamy across variable demographic landscapes in cooperatively breeding Florida scrub-jays. *Behav Ecol* 22:464–470.
22. Matsen FA, Evans SN (2008) To what extent does genealogical ancestry imply genetic ancestry? *Theor Popul Biol* 74:182–190.
23. Gravel S, Steel M (2015) The existence and abundance of ghost ancestors in biparental populations. *Theor Popul Biol* 101:47–53.
24. Brommer JE, Gustafsson L, Pietiäinen H, Merilä J (2004) Single-generation estimates of individual fitness as proxies for long-term genetic contribution. *Am Nat* 163:505–517.
25. Abrahamson W, Abrahamson CR (2002) Persistent palmettos: Effects of the 2000-2001 drought on *Serenoa repens and Sabal etonia*. *Fla Sci* 65:281–292.
26. Valligny C (2003) *Effects of an Intense Winter Fire on Florida Scrub-Jay (Aphelocoma coerulescens) Habitat Use* (Villanova Univ, Villanova, PA).
27. Schoech SJ, Rensel MA, Bridge ES, Boughton RK, Wilcoxen TE (2009) Environment, glucocorticoids, and the timing of reproduction. *Gen Comp Endocrinol* 163:201–207.
28. Robertson A (1961) Inbreeding in artificial selection programmes. *Genet Res* 2:189–194.
29. Santiago E, Caballero A (1995) Effective size of populations under selection. *Genetics* 139:1013–1030.
30. Heyer E, Sibert A, Austerlitz F (2005) Cultural transmission of fitness: Genes take the fast lane. *Trends Genet* 21:234–239.
31. Zietsch BP, Kuja-Halkola R, Walum H, Verweij KJH (2014) Perfect genetic correlation between number of offspring and grandoffspring in an industrialized human population. *Proc Natl Acad Sci USA* 111:1032–1036.
32. Wright S (1938) Size of population and breeding structure in relation to evolution. *Science* 87:430–431.
33. Woolfenden GE, Fitzpatrick JW (1984) *The Florida Scrub Jay–Demography of a Cooperative-Breeding Bird* (Princeton Univ Press, Princeton).
34. Schoech SJ, Mumme RL, Moore MC (1991) Reproductive endocrinology and mechanisms of breeding inhibition in cooperatively breeding Florida scrub jays *(Aphelocoma c. coerulescens)*. *Condor* 93:354–364.
35. Mumme RL (1992) Do helpers increase reproductive success? *Behav Ecol Sociobiol* 31:319–328.
36. Aguillon SM, et al. (2017) Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. *PLoS Genet* 13:e1006911.
37. McLeod A (2011) *Kendall: Kendall Rank Correlation and Mann-Kendall Trend Test*. R Package Version 2.2. Available at https://CRAN.R-project.org/package=Kendall. Accessed December 1, 2015.
38. R Core Team (2013) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna). Available at www.R-project.org/. Accessed September 15, 2015.