Inferring genetic connectivity in real populations, exemplified by coastal and oceanic Atlantic cod

Ingrid Spies\textsuperscript{a,b,c,1}, Lorenz Hauser\textsuperscript{d}, Per Erik Jorde\textsuperscript{e,c}, Halvor Knutsen\textsuperscript{c,e,f}, André E. Punt\textsuperscript{d}, Lauren A. Rogers\textsuperscript{a,c}, and Nils Chr. Stenseth\textsuperscript{c,1}

\textsuperscript{a}National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Seattle, WA 98115; \textsuperscript{b}Quantitative Ecology and Resource Management, University of Washington, Seattle, WA 98195; \textsuperscript{c}Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, N-0316 Oslo, Norway; \textsuperscript{d}School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195-5020; \textsuperscript{e}Institute of Marine Research, N-4817 His, Norway; and \textsuperscript{f}Centre for Coastal Research, University of Agder, 4604 Kristiansand, Norway

Contributed by Nils Chr. Stenseth, March 19, 2018 (sent for review January 5, 2018; reviewed by Howell Tong and Nigel G. Yoccoz)

Genetic data are commonly used to estimate connectivity between putative populations, but translating them to demographic dispersal rates is complicated. Theoretical equations that infer a migration rate based on the genetic estimator $F_{ST}$, such as Wright’s equation, $F_{ST} \approx 1/(4N_e m + 1)$, make assumptions that do not apply to most real populations. How complexities inherent to real populations affect migration was exemplified by Atlantic cod in the North Sea and Skagerrak and was examined within an age-structured model that incorporated genetic markers. Migration was determined under various scenarios by varying the number of simulated migrants until the mean simulated level of genetic differentiation matched a fixed level of genetic differentiation equal to empirical estimates. Parameters that decreased the $N_e / N_t$ ratio (where $N_e$ is the effective and $N_t$ is the total population size), such as high fishing mortality and high fishing gear selectivity, increased the number of migrants required to achieve empirical levels of genetic differentiation. Higher maturity-at-age and lower selectivity increased $N_e / N_t$ and decreased migration when genetic differentiation was fixed. Changes in natural mortality, fishing gear selectivity, and maturity-at-age within expected limits had a moderate effect on migration when genetic differentiation was held constant. Changes in population size had the greatest effect on the number of migrants to achieve fixed levels of $F_{ST}$, particularly when genetic differentiation was low, $F_{ST} \approx 10^{-3}$. Highly variable migration patterns, compared with constant migration, resulted in higher variance in genetic differentiation and higher extreme values. Results are compared with and provide insight into the use of theoretical equations to estimate migration among real populations.

Wright’s equation | population dynamics | genetic connectivity | fisheries management | population genetic theory

Dispersal is fundamental to the dynamics of ecological systems. Individuals may disperse and interact in many complex ways, resulting in population structures that can range from complete panmixia to distinct populations, metapopulations, and isolation-by-distance (1–3). Knowledge of dispersal rates is important for defining units for management. Patterns of dispersal are likely to be stochastic, and factors affecting population-level dispersal rates are complicated and not fully understood (4, 5). Dispersal can be affected by population size, recruitment success, and life history parameters to oceanographic and annual climatic variation (5, 6).

Genetic data are commonly used to estimate connectivity between putative populations, but translating them to demographic dispersal rates is extremely difficult (7). Interpretation of genetic data can be particularly problematic when genetic differentiation is not significantly different from zero, because it is difficult to distinguish between moderate dispersal rates and panmixia (8). Lack of genetic differentiation may be due to high dispersal rates between populations, a lack of statistical power, or insufficient time for differences to accrue among separate groups. There are several indirect formulae that can be used to estimate dispersal (e.g., refs. 9 and 10); one of the most common is $F_{ST} \approx 1/(4N_e m + 1)$, hereafter referred to as Wright’s equation (11), where $N_e$ is the effective population size of each population under the island model in migration–drift equilibrium and $m$ is the migration rate, and the effective number of migrants is the product of $N_e$ and $m$, $M_e = N_e m$, implying that the reproductive success of immigrants is equal to residents.

The goal of this study was to examine how characteristics of natural populations and varying patterns of migration affect the relationship between genetic differentiation and demographic connectivity among populations that are exchanging migrants. We compared the relationship between genetic differentiation and migration in simulated age-structured populations with Wright–Fisher populations to explore how well Wright’s equation estimated migration in real populations. Simulations explored the demographic relationships within a particularly well-studied system, Atlantic cod (Gadus morhua) in the North Sea, and two smaller populations on the Norwegian coast in the Skagerrak: a coastal population at Risør, referred to as the outer coast, and a fjord population, Søndeled, referred to as the inner fjord (Fig. 1). Simulated populations were at migration–drift equilibrium, and simulations incorporated 13 microsatellite markers, similar to a study on population structure in this system (12). Connectivity between these two regions has been examined with genetics and tagging (12–15), and research suggests that adults are generally

Significance

Estimates of migration are important for understanding the dynamics of natural populations. A statistic known as $F_{ST}$ is often used to measure levels of genetic differentiation among natural populations. Equations that translate $F_{ST}$ into estimates of migration are based on “ideal” populations, which are subject to many simplifying assumptions compared with real populations. Therefore, theoretical estimates of migration might not be realistic. We modeled populations of Atlantic cod in the North Sea and the adjacent Skagerrak region to compare how migration is related to the complexities of real populations, and how actual migration compares with predictions based on theory. Results are intended to help apply population genetic theory to practical situations.

Author contributions: I.S., P.E.J., H.K., L.A.R., and N.C.S. designed research; I.S. performed research; I.S. and P.E.J. analyzed data; and I.S., L.H., P.E.J., H.K., A.E.P., L.A.R., and N.C.S. wrote the paper.

Reviewers: H.T., London School of Economics and Political Science; and N.G.Y., UiT The Arctic University of Norway.

The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

To whom correspondence may be addressed. Email: ingrid.spies@noaa.gov or n.c.stenseth@bbv.uio.no.

This article contains supporting information online at www.pnas.orglookup/suppl/doi:10.1073/pnas.180096115/-/DCSupplemental.

Published online April 19, 2018.

www.pnas.org/cgi/doi/10.1073/pnas.180096115

PNAS | May 8, 2018 | vol. 115 | no. 19 | 4945–4950
Table 1. Empirical levels of genetic differentiation using 13 microsatellite loci, measured by \( F_{ST} \), between Atlantic cod samples from two different years of samples (year shown) from the North Sea and the inner fjord and the outer coast (12)

| Location/year | Inner fjord (2005) | Outer coast (2000) |
|---------------|--------------------|--------------------|
| N. Sea 2000/2001 | 0.0039 (P = 0.0013) | 0.0003 (P = 0.7022) |
| N. Sea 2002 | 0.0051 (P = 0.0001) | 0.0001 (P = 0.3414) |

North Sea cod were sampled from the German Bight in 2002 and off Hirtshals in 2000, 2001.
the estimated number of migrants when levels of genetic differentiation were held constant (Fig. 3). This difference reflects the property that, when $F_{ST}$ is very low, small errors in $F_{ST}$ increase errors in estimating $N_m$ (e.g., ref. 22): the outer coast and the North Sea were characterized by lower levels of genetic differentiation ($F_{ST} = 0.0001$ to $0.0003$) than the inner fjord and the North Sea ($F_{ST} = 0.0039$ to $0.0051$). Table 1). Simulations used base case parameterization, with double and half the inner fjord and outer coast populations. The targeted empirical levels of genetic differentiation are shown as dotted horizontal lines, and zero is shown as a solid horizontal line in $F_{ST}$ distribution of migration used to simulate stochasticity intended to mimic natural processes, with the median level lower than the mean and occasional years of extremely high migration (13). Dispersal is a stochastic process; in fact, stochasticity in dispersal rates may be key to the persistence of metapopulations (24). Simulations with stochastic migration showed that levels of genetic differentiation among populations connected by migration varied around a central tendency, similar to previous studies (25). These complement observations of temporal differences commonly observed in studies of genetic differentiation (26), and may support other theories explaining temporal changes in allele frequencies, such as sweepstakes reproductive success (27). Levels of uncertainty documented here with 100 simulations for each case show that temporal samples provide varying estimates of genetic differentiation whether migration is static (Fig. 5) or stochastic (Fig. 4) when populations are in migration–drift equilibrium, although stochastic migration is more likely in natural systems. In addition, variance typically increased with stochastic migration (Fig. 4). These results support the use of multiple temporal samples to provide a more accurate mean estimate of genetic differentiation among populations than a single estimate (8).

Age-structured populations were characterized by higher differentiation than non-age-structured populations, which follows from lower $N_e/N_m$ ratio in age-structured populations if census size is constant (Fig. 6 A and B vs. Fig. 6 C and D). $F_{ST}$ predicted by Wright’s equation was within the range of 100 simulations in all cases of age-structured populations when $N_m$ was known (Fig. 5); mean simulated $F_{ST}$ values were higher than theoretical estimates when migration was low (Fig. 5A) and lower than theoretical estimates when migration was high (Fig. 5B). Wright’s equation is compared with two-way migration and equal
population size scenarios in Fig. 5 because these most closely match the assumptions associated with that equation. One-way migration increased genetic differentiation over two-way (bidirectional) migration regardless of relative population size or migration rate, by decreasing the exchange of genes (Fig. 5).

Mutation, which is typically negligible over short evolutionary time periods, decreased levels of genetic differentiation when mutation rates were similar to migration rate (refs. 28 and 29, Fig. 6B vs. Fig. 6E, and Fig. 6D vs. Fig. 6F), and the finite-island model with mutation and migration (30) predicted \( F_{ST} \) accurately (Fig. 6D). Therefore, information about relative population sizes and patterns of migration can be used to provide additional insight into the amount of migration between two populations.

Research on genetic population structure in natural populations is often designed to inform management decisions and conservation efforts. Knowledge of genetic divergence (e.g., \( F_{ST} \)) is insufficient for assessing demographic connectivity (7, 22) for two reasons: (i) demographic connectivity depends on the migration rate (\( m \)) rather than the number of migrants (\( N_m \)), and (ii) linking \( F_{ST} \) to levels of migration (i.e., Wright’s equation) rests on a number of assumptions that may be unrealistic (22). Here, we solved the first problem by including an external estimate of population size \( N \) and \( N_e \) from simulated populations, as described in SI Appendix, which allowed us to extract a separate estimate of \( m \) from the composite estimate of \( N_m \). We solved the second problem by examining robustness to assumptions through computer simulations.

Briefly, our simulations demonstrated high levels of robustness to several key assumptions, including the number of populations (two rather than infinite), their relative sizes (unequal rather than equal), and the direction of migration between them (one-way rather than two-way). We found that \( F_{ST} \), with and without stochastic migration, rapidly approached its pseudo drift–migration equilibrium value, as expected when the number of migrants is not extremely low. These characteristics are expected to be fairly common in nature, especially for marine organisms, and facilitate the adoption of Wright’s equation in real populations.

Some factors particular to our model, including maturity-at-age, fishing mortality, and fishing gear selectivity, did have an effect on the number of migrants required to reach the observed \( F_{ST} \). However, the effect was driven by how these factors influenced the effective to census population size ratio (\( N_e/N \)). We demonstrated that reasonable variation in these factors typically resulted in small changes in the estimated migration. These findings suggest a general procedure for dealing with violations of the assumptions associated with Wright’s equation by evaluating their impact on the \( N_e/N \) ratio and adjusting the estimated number of migrants accordingly. In conclusion, we find that Wright’s equation provides a useful starting point for estimating dispersal from genetic data (\( F_{ST} \)) in real-world organisms that depart strongly from assumptions of Wright’s equation, provided we have prior information on population sizes and basic biological knowledge of the population system, such as the age or stage of dispersal.

Materials and Methods

The model consisted of a large donor population and two recipient populations with different levels of genetic differentiation (\( F_{ST} \)) with the donor, and exemplified a fairly common phenomenon in the marine environment: source sink dynamics involving a dispersive larval phase (e.g., refs. 13, 31, and 32). The recipient populations differed in size and genetic differentiation with the donor population; therefore, model runs that explored the relationship between the donor population and the first recipient population were performed separately from runs between the donor population and the second recipient population. The model incorporated age-structured population dynamics, one-way larval migration from the larger to the smaller population, and stochastic mutation at expected rates during the simulation period. In each year, the genetic composition of each new recruit was generated by randomly selecting (with replacement) one male and one female from the group of spawning fish. Migrants were subject to the same level of fishery selection, fishing mortality, and natural mortality as fish in the recipient population. Mutation was simulated by random draws from a Beta(6,1) distribution that were scaled down by 0.01. Migration was determined by varying the number of simulated migrants (\( M_0 \)) until the mean simulated level of genetic differentiation, \( F_{ST} \), measured by theta, \( \theta \) (33), fell within observed levels of genetic differentiation (Table 1). The model provided estimates of the level of migration between the North Sea and the Skagerrak and was known within the framework of the operating model; therefore, migration is generally referred to as a known value rather than an estimate.

Base case simulations were always run with two populations, one representing the North Sea and the other representing one of two recipient populations.

Table 2. Results of stochastic migration rates, using a lognormal distribution of migrants with a CV of 1.31

| Region      | \( M_0 (M_e) \) | \( F_{ST} \) (mean, 10% and 90% quantiles) |
|-------------|----------------|--------------------------------------------|
| Inner fjord | 55 (17)        | 0.0052 (0.0039, 0.0067)                     |
| Inner fjord | 80 (25)        | 0.0039 (0.0028, 0.0053)                     |
| Outer coast | 1,400 (433)    | 0.00034 (0.00013, 0.00053)                  |
| Outer coast | 2,925 (904)    | 0.00013 (0.00002, 0.00025)                  |

Mean \( F_{ST} \) (10% and 90% quantiles) and mean number of migrants were selected to match base case simulation results for constant migration (55 and 80 for inner fjord simulations and 1,400 and 2,925 for outer coast simulations). The number of migrants (\( M_0 \)) refers to the number of age-0-y migrants per year, and \( M_e \) represents the effective number of migrants per generation.
Simulated and theoretical levels of genetic differentiation between two age-structured populations without mutation over 100 simulations, for (A) 5 migrants per year, $m=0.15\%$, (B) 50 migrants per year, $m=1.5\%$, and (C) 500 migrants per year, $m=15\%$, where $m$ is calculated with respect to the recipient population. Within each panel, boxplots are (in this order) unequal population sizes ($N=5,400$ and $N=65,000$) with one-way migration from the large to the small population, unequal population sizes with two-way migration, equal population sizes (two populations size $N=5,500$) with one-way migration, and equal population sizes with two-way migration. Red diamonds represent theoretical predictions for theta ($33, 35$), and the red horizontal line is drawn at the theoretical prediction value. Boxplots show mean, 1 SD, and minimum and maximum values.
We define the migration rate, $m$, for a particular (recipient) population as $m = MN$, where $M$ is the number of migrants, and $N$ is the population size of the population of interest. This equation assumes that migration takes place randomly among all age classes, and that migration is one-directional. Migration of a single age class is defined relative to the size of that age class in the population, $m = M/N$. Most simulations were performed with one-way migration, and migration was presented relative to the Skagerrak, or smaller, population, unless specified otherwise. In the base case of unequal populations, migration for the smaller population (outer coast), $m = 1,400/9,492 = 0.11%$.

In the simple Wright island model, the effective number of migrants per generation can be represented as $M_e = Nm$. When generations overlap, the number of juveniles in the population per year is $N_0$, and the effective population size per generation ($N_e$) is then a complicated function of $N_0$ since it also involves the lifetime variance in reproductive success ($s$). However, under the assumption that survival and reproduction are the same for immigrant and native juveniles, this variance remains unchanged whether an individual is a juvenile migrant, $m_0$, or a member of the age-0-y class, $N_0$. The ratio $M_0/M$ must then equal $N_0/N$, so that $M_0 = M_0N_0/N_0$. As a numerical example (see Appendix, Table S1, first line): $M_0 = 55$ and $N_0/N_0 = (N_1/N_3)(N_2/N_0) = 0.195*5,444/438 = 0.309$, so that $M_0 = 55*0.309 = 17$, where * indicates multiplication. When $M_0 = 5$ and the population structure remains the same, $M_0 = 5*0.309 = 1.54$, and using this quantity in the theoretical formula for $F_{IS}$, we predict $F_{IS} = 1(1 + 4M_0(2(1 - s)))) = 0.075$, which is the value represented by the red diamond and dotted line in Fig. 5A.

ACKNOWLEDGMENTS. Fred Allendorf, Eric C. Anderson, Matt Jay, Kim Rand, Vladimir Minin, Tore Schweder, Asa Sourdiffe, Robin Waples, and Rebecca White provided technical support and useful feedback on the manuscript. Funding was provided by the Centre for Ecological and Evolutionary Synthesis, Oslo, Norway, and the National Oceanographic and Atmospheric Administration, US Department of Commerce.