

# Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations

Ian R. Bradbury\*, Paul Bentzen

Marine Gene Probe Laboratory, Department of Biology, Dalhousie University, 1355 Oxford St., Halifax, Nova Scotia B34 4H6, Canada

**ABSTRACT:** Indirect genetic approaches such as those based on the association between genetic and geographic distance (isolation by distance, IBD) may provide one of the best means of estimating dispersal in marine systems. We evaluated the scale-dependency and the ecological covariates of the IBD relationship through a combined modeling and meta-analytical approach. Using Wright's fixation index  $F_{ST}$ , simulations were used to generate IBD relationships,  $F_{ST}/(1-F_{ST})$  and geographic distance, for various combinations of average dispersal distance and number of generations. IBD linearity increased with time since colonization and dispersal distance, although the effect of dispersal distance predominated after initial colonization. Simulations suggest that declines in the IBD slope were associated with increases in the spatial scale of observation, suggesting that the IBD pattern is non-linear at very short and long average dispersal distances. We hypothesized that non-linear IBD would be common in marine populations, and apparent through a biologically significant non-zero intercept and decreasing slope with increasing geographic scale. Predictions of common non-linearity were examined through IBD relationships (i.e. slope, intercept,  $R^2$  of the regression) and life history parameters from the published literature for 18 species of anadromous/philopatric fishes. As predicted, IBD parameters (intercept and slope) were consistently correlated with many life history traits (e.g. fecundity, egg size). However, the statistical removal of adult size eliminated most significant life history–IBD correlations. Increases in IBD slope were associated with decreasing gene flow ( $p = 0.014$ ,  $R^2 = 0.33$ ), and decreasing migration distance ( $p = 0.039$ ,  $R^2 = 0.23$ ). Non-linearity was further supported by consistent declines in the IBD slope with increased geographic scale in anadromous fish as well as several marine species, suggesting the ubiquity of this phenomenon. We conclude that isolation by distance patterns may reflect dispersal phenotype and are biologically significant. Nonetheless, non-linearity in IBD pattern is probably the norm in aquatic organisms, resulting from large ranges and limited dispersal. Accordingly, approximations of demographic parameters based on the IBD must be made cautiously, taking into account possible non-linearity, scale dependencies, and assumptions of genetic drift–dispersal equilibrium.

**KEY WORDS:** Dispersal · Isolation by distance · Anadromous · Life history · Marine fish

*Resale or republication not permitted without written consent of the publisher*

## INTRODUCTION

Estimates of dispersal have been notoriously difficult to obtain in marine environments because of the wide geographic ranges and high dispersal potential of many organisms. Indirect methods, such as genetic approaches, may yield the best estimates of dispersal

at the scales needed for conservation efforts (Taylor et al. 2000, Grivet et al. 2005). Genetic methods meet the sampling requirements needed to quantify realized long-range dispersal, permitting insights otherwise impossible to obtain (e.g. Neigel 1997). Genetic studies integrate over generations, focus on successful dispersers, and hence generate a picture of long-

\*Email: [ibradbur@dal.ca](mailto:ibradbur@dal.ca)

term patterns in connectivity and dispersal (Palumbi 2003).

Under restricted dispersal ( $d <$  species range, where  $d$  is the species average dispersal distance per generation) a decrease in genetic correlation with increasing geographic distance is commonly observed, termed isolation by distance (IBD). Although first described by Wright (1943), IBD has received increased attention following Slatkin's (1993) study, and is currently the most frequently employed hypothesis in studies of spatial genetic pattern. Positive relationships between geographic separation and genetic differentiation appear to be common, and hold the potential for insights into current and historical patterns of gene flow (Slatkin 1993, Palumbi 2003).

Simulation and modeling studies suggest that the IBD slope may allow estimation of demographic parameters such as  $d$  and neighborhood size (Rousset 1997, Palumbi 2003). Rousset (1997) used Malecot's (1955) lattice model to suggest that the product of effective organism density and dispersal distance may be estimated using the slope ( $b$ ) of the regression of Wright's  $F_{ST}/(1 - F_{ST})$  against geographic distance. Specifically  $1/b = 4D\sigma^2$ , where  $D$  is the effective adult density (no.  $\text{km}^{-1}$ ), and  $\sigma^2$  is the variance of parent-offspring axial distance. Rousset (2000) extended this approach to one based on individual differences rather than the population level differences used in his previous work (Rousset 1997). However, given the difficulty of sampling a large portion of the whole population, the earlier approach of Rousset (1997) has been more frequently applied to marine systems. In a related simulation-based approach, Palumbi (2003) used a stepping-stone simulation to estimate the relationship between  $d$  and IBD regression slope for several marine invertebrates. Increasingly, both simulation (e.g. Palumbi 2003) and modeling approaches (e.g. Rousset 1997) are being used to estimate marine dispersal (e.g. Kinlan & Gaines 2003, Buonaccorsi et al. 2004, Gomez-Uchida & Banks 2005). Nonetheless, comparisons with direct measures necessary to evaluate modeled or simulated estimates for aquatic organisms are lacking, and the applicability and validity of these approaches requires examination.

The success of IBD-based approaches ultimately depends on the nature of the IBD relationship. Following colonization, when IBD is absent (Hutchison & Templeton 1999), the slope of the IBD relationship should increase and eventually stabilize as equilibrium is approached (Slatkin 1993). IBD should develop most quickly in 1-dimensional systems and over short distances (e.g. Castric & Bernatchez 2003), but increase in spatial scale with time. At equilibrium, no variation in slope of the IBD should be observed across the species range (Hutchison & Templeton 1999). Slatkin (1993)

suggested that when IBD exists, equilibrium is established at the scale in question, gene flow is sufficient to prevent isolated demes, and long-distance dispersal is rare enough to prevent homogenization.

Scale dependence of the IBD slope has been noted (Rousset 1997, Leblois et al. 2003, Buonaccorsi et al. 2004), but rarely considered in the literature (but see Castric & Bernatchez 2003, Leblois et al. 2003). Yet, scale-dependency is critical to the usage of IBD in dispersal estimation as it could directly alter dispersal estimates and subsequent spatial management decisions. The spatial scale over which linear isolation is present in 1 dimension should increase with the diffusion approximation,  $\sqrt{2N_{em}n}$  (Slatkin 1993) where  $N_{em}$  is the effective number of migrants and  $n$  is the no. of generations since colonization. If the spatial scale is small relative to the diffusion approximation, it is likely that equilibrium would be established; however, in newly colonized systems, or in highly philopatric species, the probability that equilibrium will be attained over long distances may be low. Rousset (1997, 2000) suggested that independent estimates of  $d$  be used to delineate the scale over which linearity is observed. Linearity should hold at distances greater than  $\sigma$  and less than some value (e.g.  $0.56\sigma / \sqrt{2u}$ ,  $u$  is the mutation rate), and the ability to detect IBD will depend on the range of distances evaluated (Rousset 1997). This scenario is problematic, as independent estimates of  $\sigma$  or  $u$  are absent or rare in many situations (i.e. marine systems), making *a priori* predictions of the scale of linearity impossible. Moreover, the potential range of dispersal distances in marine species may vary by 3 to 4 orders of magnitude, making predictions difficult (Bradbury & Snelgrove 2001).

In the present study, we examined the biological significance of the IBD relationship, and the influence of potential biases resulting from the spatial scale-dependency of the IBD through a combined modeling and meta-analytical approach. We first developed a 1-dimensional stepping-stone simulation to test the assumption of linearity in the IBD relationship over various values of  $d$  and  $n$ . The simulated IBD relationships were then used to evaluate estimates of dispersal resulting from non-linearity, using both simulation (Palumbi 2003) and modeled (Rousset 1997) approaches. We tested the hypothesis of common IBD non-linearity using a database of IBD relationships for anadromous/philopatric fishes. Non-linearity was examined using iterative least-squares regression at increasing bin sizes to explore the effect of increasing study scale on IBD slope. Furthermore, the biological significance of the IBD was examined through the covariation of IBD parameters and life history traits expected to co-evolve with dispersal phenotype (Hutchings & Morris 1985).

For the purposes of this study, we define dispersal as the geographic displacement of an organism from its natal area, ideally measured from fertilization to successful reproduction of an individual. Based on this definition, seasonal migrations may not necessarily increase dispersal unless straying rates correlate with migration distance. Similarly, we define IBD as the presence of a positive association between genetic (i.e.  $F_{ST}$ ) and geographic distance. The IBD relationship is characterized by the slope, intercept, and  $R^2$  of the least-squares regression, which give a measure of the rate of genetic change with distance (slope) as well as the fit ( $R^2$ ) of the linear approximation (e.g. Crispo & Hendry 2005). We chose to focus on anadromous fishes because  $N_{em}$  (i.e. effective number of migrants) and  $n$  (number of generations since colonization) are expected to be relatively small, and the potential for resolving non-linear structure is therefore greatest. Anadromous fishes are well known for their homing tendencies: estimates average between 90 to 100% homing in many species (Hendry et al. 2004). Moreover the temperate and high latitude occurrence of anadromous fish and recent de-glaciation in many regions suggests colonization times of 10 000 to 20 000 yr (Castric & Bernatchez 2003) and hence relatively low  $n$ . Finally, the nature of coastal habitat conforms to the requirements of a 1-dimensional stepping-stone (Rousset 1997), and IBD should therefore be more apparent than the 2-dimensional situation.

## METHODS

**Stepping-stone simulations.** The simulations were a 1-dimensional, circular, stepping-stone based on Palumbi (2003). This situation is applicable to many coastal species whose populations are distributed along a linear coastline, with gene flow occurring in 1 dimension through neighboring demes. Demes were equally spaced (1 km), and possessed a constant and uniform population size ( $N = 1000$ ). Although multiple values ranging from 100 to 10 000 were examined, an  $N$  of 1000 was chosen as it seemed representative of species of interest and provided realistic computational time (see 'Discussion'). Each of 1000 demes was started with equal allele frequencies (0.5) in a single-locus, 2-allele system. In each generation, genetic drift was set as per Kimura (1980), and dispersal between demes followed. The probability of dispersal from Demes  $x$  to  $y$  was approximated as  $p(x,y) = \alpha/2 \exp(-\alpha |x-y|)$ , where  $1/\alpha$  is the mean parent offspring dispersal (i.e.  $d$ ). This dispersal probability distribution is characteristic of marine species, and has been used in models of marine reserve dynamics (Botsford et al.

2001). In each successive generation, drift, and dispersal occurred, and deme-specific allele frequencies were estimated.  $F_{ST}$  was estimated based on 10 populations of equal spacing, beginning at a random deme along the stepping-stone, based on total allele frequencies. Simulations were run 5 times and  $F_{ST}$  values averaged, simulating 5 independent loci and allowing for estimates of variance. IBD was examined by altering the sampled deme spacing and examining changes in  $F_{ST}/(1-F_{ST})$  with geographic distance (Rousset 1997). The number of generations and  $d$  were varied to examine their effect on the shape of the IBD relationship. The linearity in each IBD was examined by iterative regression at increasing distances where least-squares regressions were fitted to data sets of increasing spatial scales (i.e. 0–100, 0–200, 0–300 km, etc.) and the slope or intercept plotted against maximum distance in each bin (e.g. Castric & Bernatchez 2003).

To examine the relationship between the IBD slope and  $d$ , IBD's were generated for 20 values of  $d$  ranging from 1 to 100 km with a study scale of 100 km (i.e. 1 to 100 percent study scale). To simulate field sampling, these IBD relationships were randomly sampled without replacement ( $n = 10$ ) and ordinary least-squares regressions were used to estimate the slopes. Values of  $d$  were estimated from IBD slopes using 2 approaches. First, following Palumbi (2003) and Kinlan & Gaines (2003) we utilized our full IBD slope – known  $d$  relationship to estimate  $d$  of the subsampled slope, simulating the estimation in wild populations (e.g. Kinlan & Gaines 2003). Our estimate was compared with the relationship from Kinlan & Gaines (2003), based on the equation  $d = 0.0016(\text{IBD slope})^{-1.0001}$ . Second,  $d$  was estimated using the approach outlined in Rousset (1997) and as used in Gomez-Uchida & Banks (2005) (see 'Introduction'). Assuming symmetrical exponential dispersal, the IBD slope approximated by  $4D\sigma^2$  was converted to  $d$  using  $1/\alpha = 1/\sqrt{2/\sigma^2}$  (Buonaccorsi et al. 2004) and the effective deme size ( $N_e$ ) of 1000 individuals.

**Literature survey and data collection.** We examined the prevalence of IBD in the literature through a comprehensive search utilizing both the Web of Science and CSA Fisheries and Aquatic Sciences. Annual publication frequencies were compiled for the period 1980 to 2004 and used to generate IBD data (see Table 1) for anadromous species. In cases where  $F_{ST}$  values were reported, they were used directly; however, in most cases it was necessary to convert reported allele frequencies to  $F_{ST}$  using Arlequin 2.0 (Schneider et al. 2000). Distances between sampling sites were measured as the shortest within-water distance (no dispersal over land) between 2 sampling sites, from charts or using Microsoft Streets and Trips (Version 9.00). Linear IBD equations were estimated using ordinary least-

squares regression of geographic distance against  $F_{ST} / (1 - F_{ST})$  (Rousset 1997). For inclusion we required that studies provided 3 or more estimates of  $F_{ST}$  at multiple spatial scales and demonstrated a clear increase in  $F_{ST}$  at larger spatial scales. We excluded cases where no increase was apparent, because such cases might reflect factors other than uniform dispersal (such as dispersal barriers) that would confound the analysis.

Data on various life history parameters were collected for each species following Hutchings & Morris (1985). However, because the distance of freshwater movement was of interest, migration distance was added to the list of characters. It is questionable whether or not estuarine/freshwater (= upstream) migration distance is representative of marine migration distances; however, estimates of these 2 distances were correlated in most cases where data were available. The exception was Atlantic sturgeon *Acipenser oxyrinchus*, in which the upstream distance is relatively short (~50 km) compared with marine migration (~1400 km), and hence not representative of the migration potential of this species. Therefore, the value of marine migration was used in this instance. Mean values were used for all analyses, necessitating 2 assumptions: first, that intraspecific variation in these characters was not important, and second, that the study populations were broadly representative of the species in question. Although we acknowledge that these assumptions may contribute noise to the data, we suggest that interspecific variation in life history attributes is much larger than intraspecific variation, given the wide taxonomic diversity represented in this study.

**Data analysis.** Relationships between IBD parameters,  $d$ , and life history were explored graphically and using linear regression. Co-variation between life history traits and migration variables were explored statistically and graphically using Pearson correlation coefficients and linear regression. We used Stearns' (1992) method of correcting for size through the use of allometric residuals to investigate the co-variation between size and dispersal phenotype (i.e. IBD regression parameters). Life history characters were regressed against maximum length and we used the regression residuals in further analysis. Correlation analyses were conducted before and after the statistical removal of the effects of size to examine the consequences of size on co-variation. Finally, we examined co-variation of all measured variables to explore the co-evolution of these traits. Principal components analysis (PCA) was conducted with the effect of size removed to evaluate co-variation of IBD parameters with life history traits (e.g. Hutchings & Morris 1985).

## RESULTS

### Simulations

IBD was apparent in all simulations, but with varying departures from linearity. The degree to which linearity was violated decreased with increasing  $d$  and number of generations, as expected (Fig. 1). After 100 generations, IBDs were characterized by an initial rapid increase in  $F_{ST}$  with increasing geographic distance; however, at distances of 20 to 30 km genetic differentiation achieved its maximum value. This maximum decreased with  $d$  ranging from  $F_{ST} = 0.003$  at  $d = 1$ , to 0.00014 at  $d = 10$ . In contrast, after 10 000 generations the increase in  $F_{ST}$  with distance was much more gradual, with IBD slopes of 0.001 at  $d = 1$ , to 0.000012 at  $d = 10$ . The standard error of the genetic differentiation increased with increasing distance in all examples (Fig. 1). Iterative regression analysis with increased geographic range indicated that the IBD slope stabilized as  $n$  and  $d$  increased (Fig. 1). Nonetheless, after 10 000 generations a maximum  $F_{ST}$  was still obtained at a distance of 20 km for a dispersal distance of 1 km. Iterative regression analysis revealed a decrease in slope with increasing spatial scale in 75% of simulations. In the remaining simulations equilibrium seemed to be established across the study scale, and slopes were relatively constant across all distances. For simulations at 100 generations, maximum  $F_{ST}$  was attained by <40 km in each simulation. However, for the simulations run for 1000 and 10 000 generations, the effect of  $d$  on the geographic scale at which maximum  $F_{ST}$  was achieved, was 3 to 4× that of generation time.

Estimates of dispersal were based on our simulated  $d$ -slope relationship (as in Palumbi 2003) as well as the model-based approach (Rousset 1997) and both seemed to approximate dispersal reasonably well when  $d$  was 2/3 the scale of study. As  $d$  increased, the IBD regression slope decreased and the  $R^2$  increased (Fig. 2a). However, the variance in estimates increased with increasing  $d$  (Fig. 2b). Over short distances (<40 km) both approaches underestimated  $d$  by 10 to 20%. At moderate distances (40 to 80 km), the simulation approach tended to overestimate  $d$  by 20 to 30% (Fig. 2b). As  $d$  approached the spatial scale of the study, both approaches were characterized by increased variability and inaccuracy as local panmixia was achieved. Interestingly, estimates of  $d$  based on the equation from Kinlan & Gaines (2003, see 'Methods') varied widely from known estimates by as much as 38% at short  $d$  to 380% at longer  $d$  (not shown).

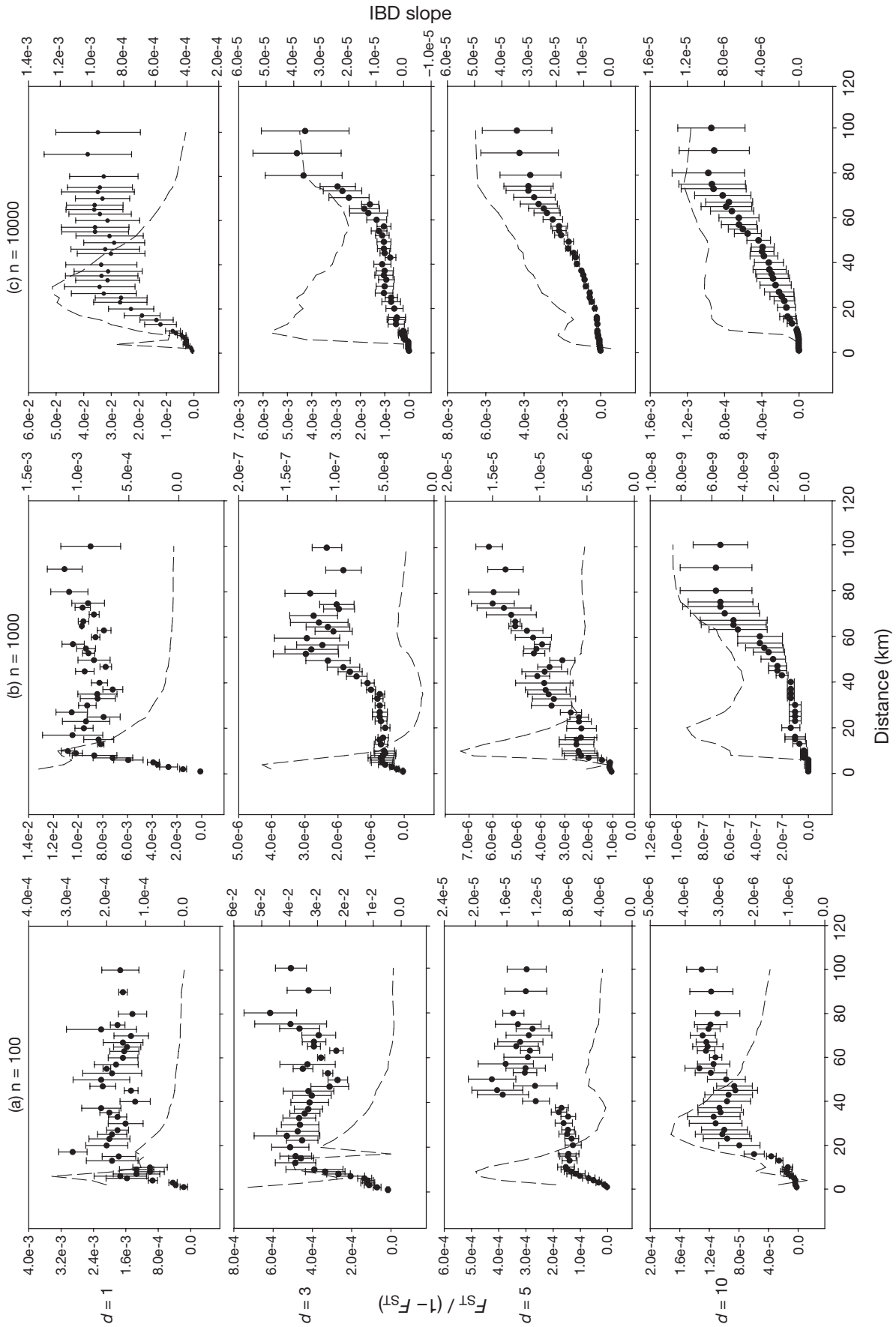


Fig. 1. Isolation by distance (IBD) relationships based on 1-dimensional stepping-stone simulations. Simulations were based on a 2-allele model, averaged over 5 loci with 1000 demes, and run for 100, 1000, or 10 000 generations, with mean dispersal distances (d) of 1, 3, 5 and 10 km and the dispersal relationship of Botstford et al. (2001). Dashed lines: iterative regression slopes based on regression utilizing increasing bin sizes (e.g. Castric & Bernatchez 2003)

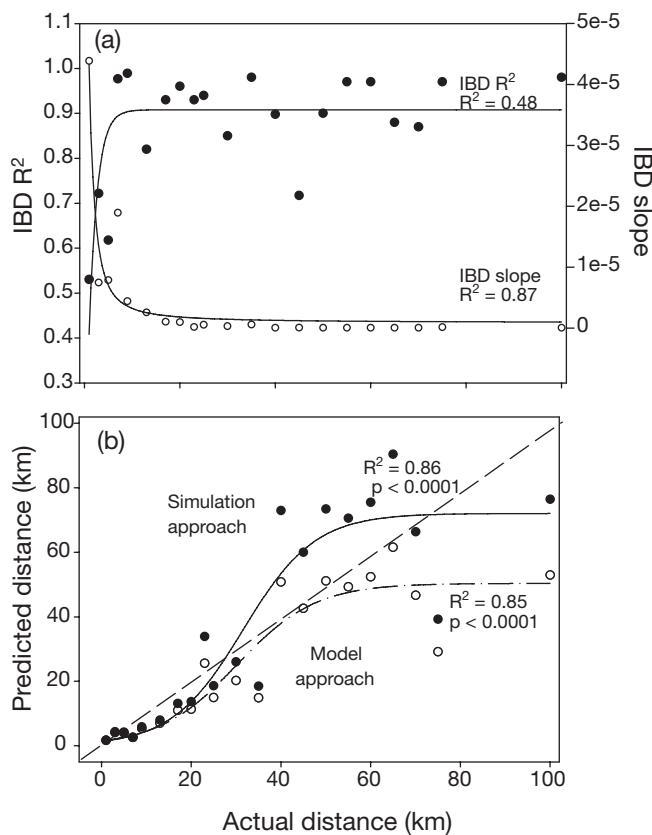


Fig. 2. (a) Slope and fit ( $R^2$ ) of linear isolation by distance (IBD) relationships generated from stepping-stone simulation at 20 dispersal distances between 1 and 100 km. All simulations ran for 1000 generations with 1000 demes and 1000 individuals per deme using methodologies as in Fig. 1 (see 'Methods'). (b) Comparison of average dispersal estimates based on approaches outlined in model-based (i.e. Rousset 1997) and simulation-based (i.e. Palumbi 2003) for IBDs generated in (a) (see 'Methods'). Complete IBDs were sub-sampled randomly to simulate field sampling, and slopes of sub-sampled sets were used to estimate average dispersal distances using relationships from Rousset (1997) and Buonacorsi et al. (2004) for model-based and from (a) for simulation-based estimates

### Meta-analysis

Significant isolation by distance was detected in the majority of species examined, and IBD parameters showed wide variation among species (Table 1), with IBD slope varying by 3 orders of magnitude. The majority of IBD parameters and life history traits displayed covariation with size (i.e. 'maximum length', Table 2). Further associations between IBD and life history variables with and without size effects were explored with pairwise correlation analysis (Table 3). Life history traits covaried significantly with IBD parameters (IBD slope and intercept). The IBD intercept was significantly positively related to size-dependent traits (i.e. maximum length, age at maturity, length at maturity, Table 3). In contrast,

IBD slope displayed significant negative relationships with size-related traits as well as egg size. However, the statistical removal of size effects (see 'Methods') eliminated all significant correlations between IBD and life history (Table 1). IBD slope and intercept were significantly correlated with average study  $F_{ST}$  as well as migration distance. The slope increased with increasing average  $F_{ST}$  ( $p = 0.014$ ; Fig. 3a) but decreased with increasing migration distance ( $p = 0.039$ ; Fig. 3b), whereas the intercept increased with increasing migration distance ( $p = 0.001$ ; Fig. 3c). PCA with the size effect removed (Fig. 3d) suggested 3 clusters of co-varying characters. Fecundity, and all other egg-related traits clustered, as did length at maturity, migration distance, IBD intercept and, finally, IBD slope, average  $F_{ST}$ , age at maturity, and the fit of the regression ( $R^2$ ).

Iterative regression analysis with increasing geographic distance revealed consistent declines in slope with increasing distance (Fig. 4) while, conversely, the intercept increased with increasing distance. Similarly, analysis of several marine species with contrasting life histories (Fig. 5) indicated clear evidence of diminishing IBD slope with increasing geographic distance.

## DISCUSSION

Dispersal in wild populations is a fundamental component of most life histories. When dispersal is much less than the species' geographic range, an association between genetic and geographic distance may develop, and IBD is currently the most commonly applied hypothesis for spatial genetic pattern. In the present study, we examined scale-dependency and ecological covariates of the genetic IBD relationship through a combined modeling and meta-analytical approach. We suggest that IBD relationships in anadromous fishes are rarely linear and that assumptions of drift–dispersal equilibrium should be evaluated carefully in anadromous as well as marine species. Approaches to the estimation of dispersal ( $d$ ) that utilize linear approximations of the IBD may be biased by this assumption. Nonetheless, accurate estimation of  $d$  was possible in a subset of simulations, and the IBD slope was significantly related to life history traits associated with dispersal, suggesting their utility in characterizing dispersal phenotype, especially in species for which direct measures may be lacking.

### Simulations and dispersal

Simulations exploring various values of dispersal distance and number of generations suggest that non-linear patterns may be quite common, especially fol-

Table 1. Summary of average  $F_{ST}$  values and isolation by distance (IBD) parameters for various anadromous fish species

Species	Mean $F_{ST}$	IBD slope	$R^2$	Intercept	Source
<i>Oncorhynchus kisutch</i>	0.081	2.72e-04	0.74	-0.0494	Hendry et al. (2004)
<i>Oncorhynchus tshawytscha</i>	0.065	1.68e-05	0.06	0.0579	Hendry et al. (2004)
<i>Oncorhynchus keta</i>	0.023	4.857e-06	0.46	0.01724	Hendry et al. (2004)
<i>Oncorhynchus gorbusha</i>	0.013	3.259e-06	0.23	0.0061525	Hendry et al. (2004)
<i>Oncorhynchus nerka</i>	0.053	4.879e-05	0.54	0.025	Hendry et al. (2004)
<i>Salmo clarki</i>	0.074	8.478e-05	0.87	0.03238	Hendry et al. (2004)
<i>Oncorhynchus mykiss</i>	0.034	3.388e-05	0.35	0.0164	Hendry et al. (2004)
<i>Salvelinus malma</i>	0.146	5.986e-05	0.71	0.02262	Everett et al. 1997
<i>Salmo salar</i>	0.059	3.067e-05	0.5	0.02984	Hendry et al. (2004)
<i>Salvelinus fontinalis</i>	0.098	5.319e-04	0.17	-0.01056	M. W. Jones (unpubl. data)
<i>Salvelinus alpinus</i>	0.057	7.087e-05	0.67	0.02814	Bernatchez et al. (1998)
<i>Coregonus lavaretus</i>	0.089	1.675e-04	0.81	0.0031982	Hansen et al. (1999)
<i>Acipenser oxyrinchus</i>	0.0910	3.578e-05	0.06	0.0768	Wirgin et al. (2000)
<i>Acipenser brevirostrum</i>	0.0839	8.273e-05	0.99	0.0106	Walsh et al. (2001)
<i>Retropinna retropinna</i>	0.050	2.9307e-03	0.79	-0.0606	Mitchell et al. (1993)
<i>Thaleichthys pacificus</i>	0.0061	5.51e-06	0.46	-0.0023504	McLean et al. (1999)
<i>Alosa sapidissima</i>	0.00951	4.429e-06	0.09	0.0045817	Epifanio et al. (1995)
<i>Morone saxatilis</i>	0.0487	3.025e-05	0.29	-0.0006407	Wirgin et al. (1989, 1993)

Table 2. Regression of life history and dispersal parameters on maximum length. Isolation by distance (IBD) parameters were calculated from linear regression of  $(F_{ST}/1-F_{ST})$  and distance (see Table 1)

Parameter	p-value	$R^2$
IBD intercept	0.001	0.49
IBD slope	<0.000	0.63
IBD $r^2$	0.169	0.12
Mean $F_{ST}$	0.21	0.10
Hatch time	0.699	0.01
Fecundity	0.01	0.34
Egg size	0.014	0.32
Hatch size	0.07	0.06
Age at maturity	<0.00	0.61
Length at maturity	<0.00	0.87
Migration distance	<0.00	0.52

lowing recent colonization events, and in species that display limited dispersal relative to their geographic range. In simulations that represented both scenarios, maximum  $F_{ST}$  values were attained at short distances, resulting in a declining slope within the study scale. Moreover, as expected, the IBD relationship weakened as the  $d$  approached the maximum scale studied. We suggest, as have other authors (Slatkin 1993, Rousset 1997), that the assumption of linearity may be valid at scales larger than the  $d$ , and scales smaller than those at which maximum genetic distance is achieved. The suggestion of an upper boundary  $(0.56\sigma/\sqrt{2u})$  to the scale of linearity may be useful under equilibrium conditions, but is unlikely to be useful under non-equilibrium conditions (Fig. 1). The scale at which linearity will occur will depend not only on  $d$ , but also on the number of generations, making prediction difficult. We observed that for  $d = 10$  the scale of maximum differen-

tiation varied with increasing  $n$  up to as much as 60 km (Fig. 1).

With the exception of the very recent colonizations (100 generation simulations), it appears that dispersal distance was the major factor in determining the distance at which maximum genetic differentiation was achieved (Fig.1). Crispo & Hendry (2005) employed a meta-analytical approach to evaluate whether time since colonization had a significant effect on the strength of the IBD relationship measured using the p-value and the  $R^2$ . Their conclusions were similar to ours in that the effect of time since colonization on the strength of the IBD was weak in the context of other factors (i.e. dispersal distance). Another key outcome of these simulations was the observation that variance in  $F_{ST}$  increased with increasing geographic distance, suggesting that increased sampling at these scales may be necessary to adequately define the IBD pattern. This finding is consistent with the model presented by Hutchison & Templeton (1999).

Our simulations suggest that estimates based on either approach provided reasonably accurate estimates of  $d$ . Patterns in estimated  $d$  suggest that at short distances (<40 km) both approaches underestimated  $d$  by 10 to 20%. At moderate distances (40 to 80 km), the simulation approach tended to overestimate  $d$  by 20 to 30% (Fig. 2b) but, as expected, both approaches failed as the IBD relationship weakened as  $d$  approached the scale of study and local panmixia was achieved. Moreover, estimates based on the simulation (i.e. Palumbi 2003) approach appear highly dependent on simulation design and parameters. Our slope-dispersal distance relationship (Fig. 2a) and simulated field IBDs were parameterized similarly and as such yielded relatively accurate estimates, yet this is difficult to dupli-

Table 3. Correlation of dispersal phenotype (isolation by distance, IBD, parameters) and life history characters. Entries above diagonal represent correlations with size included; those below the diagonal represent correlations with effect of size removed (Stearns 1992). Only correlation coefficients that were significant at  $p < 0.05$  are reported

	IBD slope	IBD $R^2$	IBD intercept	Mean $F_{ST}$	Hatch time	Fecundity	Egg size	Hatch size	Age at maturity	Length at maturity	Maximum length	Migration distance
IBD slope	—	+	0.623	+	—	—	0.532	—	-0.578	-0.762	-0.795	-0.613
IBD $R^2$	—	—	—	+	+	—	—	—	—	—	—	-0.582
IBD intercept	—	—	—	+	+	+	+	+	0.714	0.727	0.699	0.733
Mean $F_{ST}$	+	0.489	—	—	+	—	+	+	0.587	+	+	+
Hatch time	—	+	+	+	—	-0.722	0.461	0.732	+	—	—	+
Fecundity	+	—	—	—	-0.825	—	-0.584	0.465	0.486	0.587	+	+
Egg size	—	+	—	—	0.630	-0.816	—	0.795	+	0.655	0.567	0.590
Hatch size	—	+	+	+	0.781	-0.919	0.825	—	+	+	+	+
Age at maturity	+	+	+	0.581	+	+	—	+	—	0.691	0.779	0.530
Length at maturity	—	—	+	0.497	+	—	+	+	—	—	0.935	0.828
Migration distance	—	—	+	—	+	—	+	+	—	0.630	0.719	—

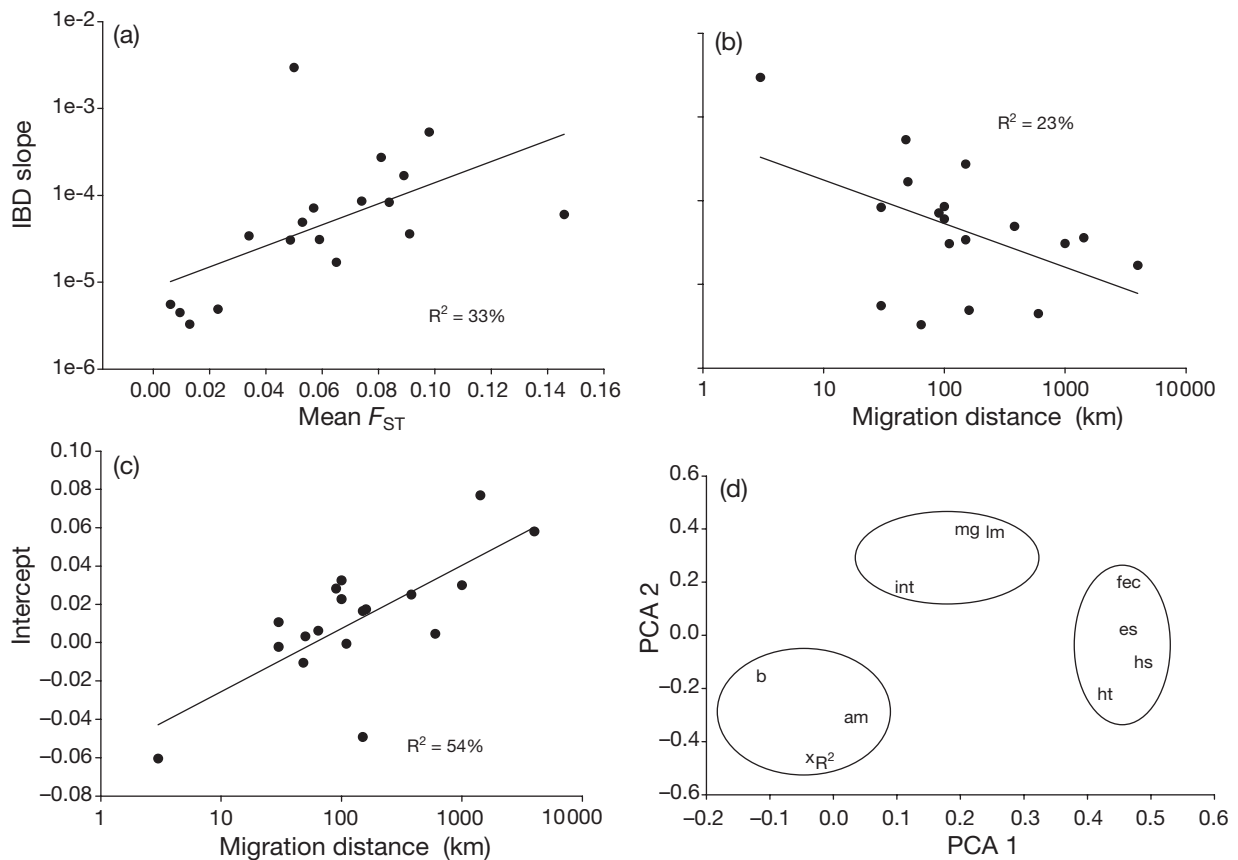


Fig. 3. Relationships between isolation by distance (IBD) parameters and life history attributes and mean  $F_{ST}$ . (a)–(c) Mean  $F_{ST}$  and migration distance in relation to slope and intercept of IBD relationships. Data from sources in Table 1. (d) Principle components analysis scores for all life history and migratory components (data from Tables 1 & 3) with effect of size removed (Stearns 1992); fec: fecundity; x: mean study  $F_{ST}$ ; int: intercept of IBD equation; b: slope of IBD regression;  $R^2$ : fit of IBD regression; am: age at maturity; lm: length at maturity; mg: migration distance; es: egg size; hs: hatch size; ht: hatch time



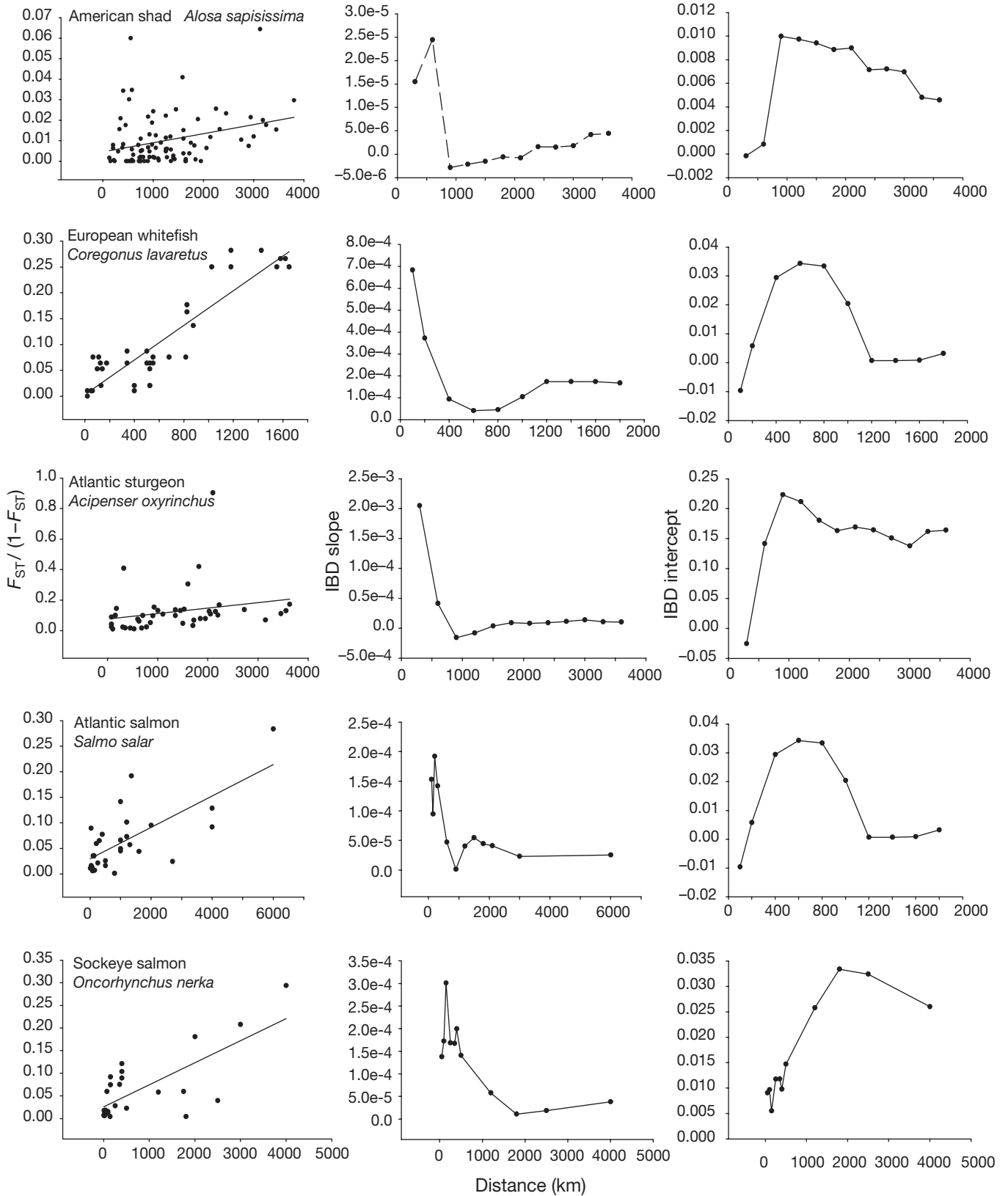


Fig. 4. Isolation by distance (IBD) relationships, slope and intercept of iterative linear regression analysis with an increasing scale of geographic distances for 5 species of anadromous fish (see Table 1)

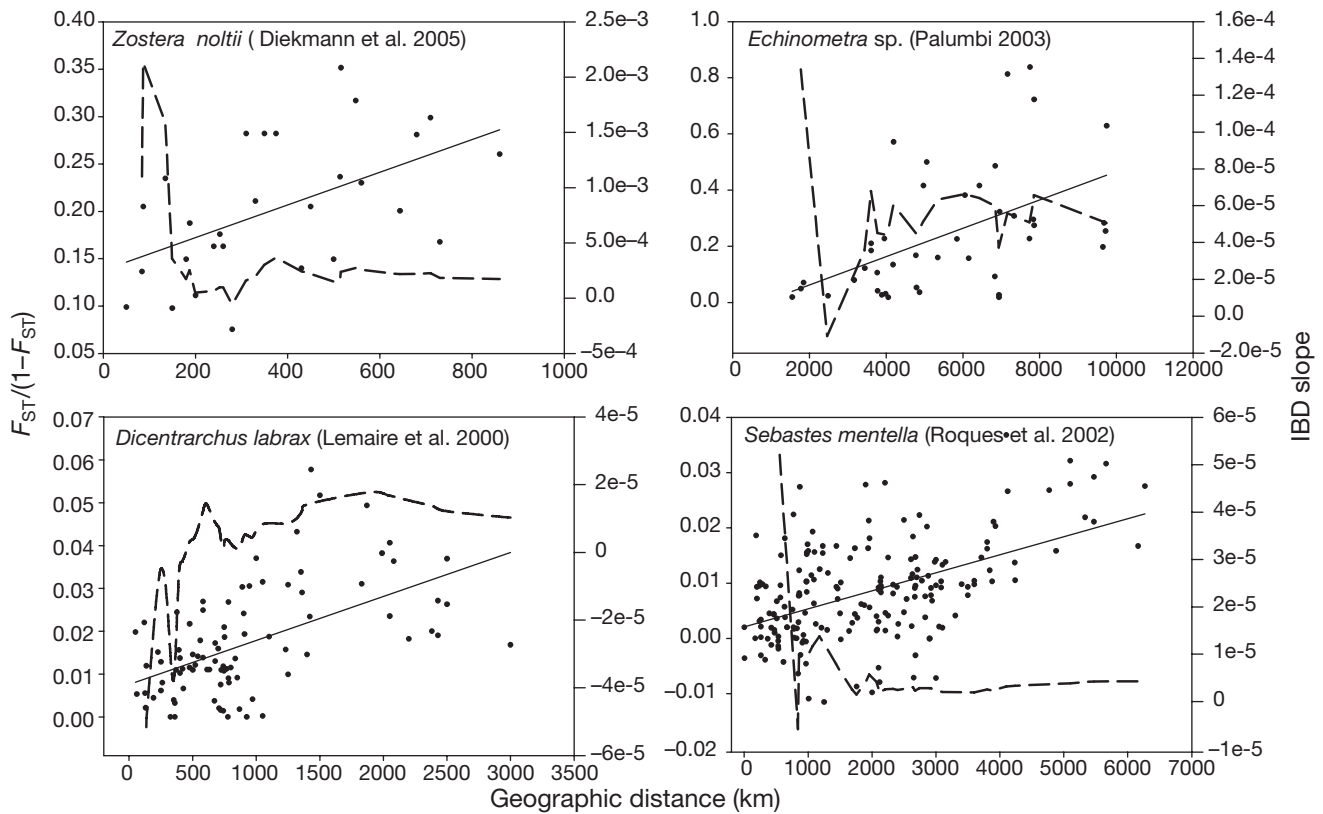


Fig. 5. Isolation by distance (IBD) relationships for 4 marine species. Dashed line: changes in slope with increases in geographic range based on iterative regression analysis (see 'Methods')

cate in a true field setting. Application of the Kinlan & Gaines (2003) slope–dispersal distance relationship (see 'Methods') produced estimates that varied by as much as 38 to 380% from known dispersal distances, suggesting errors in the range of 2 orders of magnitude, depending on simulation. As such, our results suggest that a simulation-based approach may be inappropriate in marine dispersal estimation unless the simulation can be better tailored to simulate field conditions.

Accordingly, we suggest that simulation design and conditions (i.e. number of generations, study scale, etc.) may have a large impact on real dispersal estimates, especially at larger scales. This relationship may explain discrepancies between the approaches noted by other authors (e.g. Gomez-Uchida & Banks 2005). As such, the simulation approach therefore seems unlikely to yield reliable results in natural populations. Moreover, the discrepancies noted with the simulation approach are also likely to be due to fact that the assumption of linearity is violated multiple times (i.e. once during the actual IBD regression estimation, and again during the estimation of each simulated IBD in the slope–dispersal distance relationship).

Admittedly, simulation design parameters such as dimensionality,  $N_e$ , and locus characteristics may contribute to IBD pattern and  $d$  estimation. Given the linear nature of the habitat of many coastal species and anadromous species, the assumption of a 1-dimensional habitat seems appropriate (e.g. Gomez-Uchida & Banks 2003). Genetic drift will depend on effective population sizes, and reasonable values are critical to interpretation. Reported  $N_e$  estimates for anadromous fish fall between 100 and 1000 individuals (e.g. Hansen et al. 2002).  $N_e$  estimates for marine fish are rare, yet seem to be 1000 to 10 000 and higher (e.g. Hauser et al. 2002). Our chosen value of 1000 individuals seems realistic in these circumstances. Our choice of a 2-allele, multi-locus system has been used frequently in simulation studies (e.g. Hendry et al. 2004). The simulation of multiple loci through repetitive simulations with identical parameters produces independent estimates, allowing variance estimates for distance measures. Our simulations assume negligible mutation over time scales of interest ( $\sim 1000$  yr). Although mutation rate has been shown to affect observations of spatial structure (e.g. Epperson 2005), mutation rate does seem to effect dispersal estimates made using the IBD (Leblois et al. 2003).

### Life history and dispersal

The expected co-evolution of dispersal and life history attributes motivated us to evaluate the biological significance of the IBD relationship in natural populations. Dispersal may be a direct cause of life history evolution, acting as a source of selection for larger size, better swimming ability, and enhanced sensory abilities (Hutchings & Morris 1985, Roff 1988). This interpretation is consistent with observations that energetic constraints (Bernatchez & Dodson 1987) and habitat may favor dispersal of larger individuals within a given population (Knouft & Page 2003). Furthermore, this interpretation is consistent with our observation that the dispersal phenotype represented in the IBD parameters is directly related to organism size. We observed significant associations between life history and the IBD relationship, as would be predicted if the IBD pattern was correlated with dispersal phenotype. As size increased across species, the IBD slope decreased significantly and the intercept increased, confirming the prediction that gene flow and  $d$  increase with increasing size. The elimination of associations between IBD parameters and life history attributes when the effect of size was removed further supports this linkage. This is consistent with Roff (1988), who found significant differences in the body size of migratory and resident (non-migratory) marine fishes.

### Equilibrium and IBD

The utility of the IBD relationship in demographic estimation is fully dependent on the equilibrium status of the species in question. Several authors have suggested that a diagnostic for populations in drift-dispersal equilibrium is the presence of a positive monotonic relationship across all geographic distances (Hutchison & Templeton 1999). This relationship was rarely observed in any of our simulations or in real IBD patterns from the literature. Admittedly, our focus on anadromous fishes may bias conclusions towards instability because of homing and recent colonization times; however, the patterns observed within anadromous fishes were common to several of the marine species we examined (Fig. 5). Indeed, several authors have supported the suggestion that potentially highly dispersive taxa (which include many marine species) may seldom attain equilibrium (e.g. Benzie 1999). Grosberg & Cunningham (2000) suggested that there are few examples of marine species that fully satisfy the expectations of equilibrium across their geographic ranges.

Several explanations may account for the failure to achieve equilibrium by marine species. First, the fre-

quency of disturbance may occur at a similar time scale as stabilization, so that equilibrium is seldom attained. Castric & Bernatchez (2003) suggested that deglaciation 14 000 to 20 000 yr ago may not have been sufficient to allow equilibrium to develop in anadromous brook char. Many marine species are expected to have extremely large effective population sizes, which would slow the attainment of genetic equilibrium. Second, we suggest that, in conjunction with disturbance frequency, restricted dispersal relative to the species range may significantly delay the attainment of equilibrium, particularly in coastal and anadromous species. Several recent studies using non-genetic approaches suggest that pelagic dispersal in marine systems may be on the order of 10s to 100s of kilometers (e.g. Thorrold et al. 2001). It appears that despite large dispersal potential (Bradbury & Snelgrove 2001), behavior, and mortality of long-distance dispersers in concert with complex circulation may reduce effective dispersal associated with a pelagic larval period.

### Dispersal estimation and non-linearity

Dispersal estimates are central to decisions regarding spatial management practices (e.g. Marine Protected Area [MPA] design), and the potential effect that non-linearity in IBD may have on the estimation of demographic parameters must therefore be evaluated before these estimates can be implemented in management decisions. Evaluation of genetic dispersal estimates by direct measures are absent from the marine literature, but a few cases are found in terrestrial studies (e.g. Rousset 1997, Spong & Creel 2001, Sumner et al. 2001) and tend to differ by less than a factor of 2. While no such comparisons are available for marine species, Gomez-Uchida & Banks (2005) calculated  $d$  for darkblotched rockfish *Sebastes crameri* comparing the approach of Rousset (1997) with that of Palumbi (2003), and found that estimates of  $d$  based on the 2 approaches differed by 2 orders of magnitude (1 and 111 km respectively). Consistent with this discrepancy, our estimates of  $d$  based on the equation of Kinlan & Gaines (2003) similarly differed from known values by as much as 2 orders of magnitude, supporting the hypothesis that their equation significantly overestimates at moderate to high  $d$ .

We suggest that the discrepancies observed between actual and estimated dispersal, and between estimation approaches are probably linked to the non-linearity of the IBD. This non-linearity has been observed through the reduction of the IBD slope at larger distances in a few field studies (e.g. Hellberg 1995, Castric & Bernatchez 2003). This reduction is typically attributed to the fact that either the  $F_{ST}$  has reached its

upper boundary at large distances, or a non-equilibrium state is present (Castric & Bernatchez 2003, Leblois et al. 2003). Either condition may be plausible in marine species, given that the range of dispersal is relatively unknown, and equilibrium status is therefore impossible to evaluate. Simulations suggest a working range for which  $d$  may be predicted. At large scales of study relative to dispersal distance (big scale, short  $d$ ) simulations suggest that sampling within the range of maximum  $F_{ST}$  becomes a possibility. This condition does not seem to have a large effect on estimates of dispersal because both approaches performed well in our study. Nonetheless, as  $d$  approaches the study scale the IBD signal begins to deteriorate and predictions of dispersal distance deviate from the expected value. Ideally, estimates of  $d$  would be based on the section of the IBD relationship that is less than the maximum  $F_{ST}$ . For anadromous fishes, the scale of linearity seems to be on a scale of 200 to 500 km (Fig. 4). In marine organisms the wide variety of life histories and potential dispersal scales may prevent such generalizations. Alternatively, iterative regression analysis or autocorrelation may be used to examine the scales of linearity on a group-by-group basis.

The leap from genetic pattern to demographic process inevitably involves simplifying assumptions inherent to the specific models and simulations. In general, the presence of IBD across the geographic scale under consideration is a prerequisite of dispersal estimation. Palumbi (2004) suggested that IBD should be required for dispersal estimates, but this stipulation may limit estimates to species with restricted dispersal. Species with broad-scale dispersal, which do not show IBD, would be excluded from further analysis. As noted earlier, the lack of IBD may be a result either of historical processes, or sampling at scales both too small and too large to resolve an increase in the IBD slope. Therefore, the nature of this bias is not entirely clear and is probably case-dependent. More specifically, both Rousset's (1997, 2000) approach, which is based on Malecot's lattice model, and Palumbi's simulation approach assume an even distribution of populations across the lattice, a condition rarely observed in natural populations.

## CONCLUSION

Measures of dispersal based on genetic pattern are essential to conservation and management of many species, as dispersal directly determines the species response to selection and their capacity to respond to disturbance in a changing environment. However, genetic structure reflects both contemporary and historical factors, and most populations are probably not

in drift–dispersal equilibrium across their range. As such, the simplifying assumptions involved in the estimation of demographic data from patterns of genetic differentiation may be invalid in certain systems and each species must be evaluated on a case-by-case basis. Nonetheless, these approaches may provide one of the most effective tools for the examination of dispersal in many marine species.

*Acknowledgments.* The authors thank L. Barrett, P. V. R. Snelgrove, J. Hutchings, and B. Laurel, who assisted and advised on data analysis and interpretation. I.R.B. was supported by a National Sciences Engineering Research Council Postgraduate Scholarship as well as a Killam Postgraduate Scholarship. P. Bentzen is supported by an NSERC discovery grant.

## LITERATURE CITED

- Benzie JAH (1999) Genetic structure of coral reef organisms: ghosts of dispersal past. *Am Zool* 39:131–145
- Bernatchez L, Dodson JJ (1987) Relationship between biogenetics and behavior in anadromous fish migrations. *Can J Fish Aquat Sci* 44:399–407
- Bernatchez L, Dempson JB, Martin S (1998) Microsatellite gene diversity analysis in anadromous arctic char, *Salvelinus alpinus*, from Labrador, Canada. *Can J Fish Aquat Sci* 55:1264–1272
- Botsford LW, Hastings A, Gaines SD (2001) Dependence of sustainability on the configuration of marine reserves and larval dispersal distance. *Ecol Lett* 4:144–150
- Bradbury IR, Snelgrove PVR (2001) Contrasting larval transport in demersal fish and benthic invertebrates: the roles of behaviour and advective processes in determining spatial pattern. *Can J Fish Aquat Sci* 58:811–823
- Buonaccorsi VP, Westerman M, Stannard J, Kimbrel C, Lynn E, Vetter RD (2004) Molecular genetic structure suggests limited larval dispersal in grass rockfish. *Mar Biol* 145: 779–788
- Castric V, Bernatchez L (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis*). *Genetics* 163:983–996
- Crispo E, Hendry AP (2005) Does time since colonization influence isolation by distance? A meta-analysis. *Conserv Genetics* 6 665–682
- Diekmann OE, Coyer JA, Ferreira J, Olsen JL, Stam WT, Pearson GA, Serrão EA (2005) Population genetics of *Zostera noltii* along the west Iberian coast: consequences of small population size, habitat discontinuity and near-shore currents. *Mar Ecol Prog Ser* 290:89–96
- Epifanio JM, Smouse PE, Kobak CJ, Brown BL (1995) Mitochondrial DNA divergence among populations of American shad (*Alosa sapidissima*): How much variation is enough for mixed stock analysis. *Can J Fish Aquat Sci* 52: 1688–1702
- Epperson BK (2005) Mutation at high rates reduces spatial structure within populations. *Mol Ecol* 14:703–710
- Everett RJ, Wilmot RL, Krueger CC, Reynolds JB (1997) Population genetic structure of Dolly Varden from Beaufort Sea drainages of northern Alaska and Canada. *Am Fish Soc Symp* 19:240–249
- Gomez-Uchida D, Banks MA (2005) Microsatellite analysis of spatial genetic structure in darkblotched rockfish (*Se-*

- bastes crameri*): Is pooling samples safe? *Can J Fish Aquat Sci* 62:1874–1886
- Grivet D, Smouse PE, Sork VL (2005) A novel approach to an old problem: tracking dispersed seeds. *Mol Ecol* 14: 3585–3595
- Grosberg R, Cunningham CW (2000) Genetic structure in the sea: from populations to communities. In: Bertness MD, Gaines SD, Hay ME (eds) *Marine community ecology*. Sinauer Associates Sunderland, MA, p 61–84
- Hansen MM, Mensberg KLD, Berg S (1999) Postglacial recolonization patterns and genetic relationships among whitefish (*Coregonus* sp.) populations in Denmark inferred from mitochondrial DNA and microsatellite markers. *Mol Ecol* 8:239–252
- Hansen MM, Ruzzante DE, Nielsen E, Bekkevold D, Mensberg KLD (2002) Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. *Mol Ecol* 11:2523–2535
- Hauser L, Adcock GJ, Smith PJ, Bernal Ramirez JH, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proc Natl Acad Sci USA* 99:11742–11747
- Hellberg ME (1995) Stepping-stone gene flow in the solitary coral, *Balanophyllia elegans*: equilibrium and non-equilibrium at different spatial scales. *Mar Biol* 123:573–581
- Hendry AP, Castric V, Kinnison MT, Quinn TP (2004) The evolution of philopatry and dispersal: homing versus straying in salmonids. In: Hendry AP, Stearns SC (eds) *Evolution illuminated: salmon and their relatives*, Oxford University Press, p 91–52
- Hutchings JA, Morris DW (1985) The influence of phylogeny, size, and behavior on patterns of co-variation in salmonid life histories. *Oikos* 45:118–124
- Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53:1898–1914
- Kimura M (1980) Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical and pseudo-sampling methods. *Proc Natl Acad Sci USA* 77:522–526
- Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84:2007–2020
- Knouft JH, Page LM (2003) The evolution of body size in extant groups of North American freshwater fishes: speciation, size distributions, and Cope's rule. *Am Nat* 161: 413–421
- Leblois R, Estoup A, Rousset F (2003) Influence of mutational and sampling factors on the estimation of demographic parameters in a 'continuous' population under isolation by distance. *Mol Biol Evol* 20:491–502
- Lemaire C, Allegretti G, Naciri M, Bahri-Sfar L, Kara H, Bonhomme F (2000) Do discrepancies between microsatellite and allozyme variation reveal differential selection between sea and lagoon in the sea bass (*Dicentrarchus labrax*)? *Mol Ecol* 9:457–467
- Malecot G (1955) Remarks on the decrease of relationship with distance. *Cold Spring Harbor Symp Quant Biol* 20:52–53
- McLean JE, Hay DE, Taylor EB (1999) Marine population structure in an anadromous fish: Life history influences patterns of mitochondrial DNA variation in the eulachon, *Thaleichthys pacificus*. *Mol Ecol* 8:143–158
- Mitchell CP, Smith PJ, Northcote TG (1993) Genetic differentiation among populations of New Zealand common smelt, *Retropinna retropinna*, from Waikato basin. *NZ J Mar Freshw Res* 27:249–255
- Neigel JE (1997) A comparison of alternative strategies for estimating gene flow from genetic markers. *Annu Rev Ecol Syst* 28:105–128
- Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves *Ecol Appl* 13: 146–158
- Palumbi SR (2004) Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management. *Annu Rev Environ Res* 29:31–68
- Roff DA (1988) The evolution of migration and some life history parameters in marine fishes. *Environ Biol Fish* 22: 133–146
- Roques S, Sevigny JM, Bernatchez L (2002) Genetic structure of deep-water redbfish, *Sebastes mentella*, populations across the North Atlantic. *Mar Biol* 140:297–307
- Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228
- Rousset F (2000) Genetic differentiation between individuals. *J Evol Biol* 13:58–62
- Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.00: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–279
- Spong G, Creel S (2001) Deriving dispersal distances from genetic data. *Proc R Soc Lond B* 268:2571–2574
- Stearns SC (1992) *The evolution of life-histories*. Oxford University Press, Oxford
- Sumner J, Rousset F, Estoup A, Moritz C (2001) Neighbourhood size, dispersal, and density estimates in the prickly forest skink (*Gnypetoscincus queenslandiae*) using individual genetic and demographic methods. *Mol Ecol* 10: 1917–1927
- Taylor BL, Chivers SJ, Sexton S, Dizon AE (2000) Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation results. *Conserv Biol* 14: 1287–1297
- Thorrold SR, Latkoczy C, Swart PK, Jones CM (2001) Natal homing in a marine fish metapopulation. *Science* 291: 297–299
- Walsh MG, Bain MB, Squiers T, Waldman JR, Wirgin I (2001) Morphological and genetic variation among shortnose sturgeon *Acipenser brevirostrum* from adjacent and distant rivers. *Estuaries* 24:41–48
- Wirgin I, Proenca R, Grossfield J (1989) Mitochondrial DNA diversity among populations of striped bass in southeastern United States. *Can J Zool* 67:891–907
- Wirgin I, Ong T, Maceda L, Waldman JR, Moore D, Courtenay S (1993) Mitochondrial DNA variation in striped bass (*Morone saxatilis*) from Canadian rivers. *Can J Fish Aquat Sci* 50:80–87
- Wirgin I, Waldman JR, Rosko J, Gross R, Collins MR, Rodgers SG, Stabile J (2000) Genetic structure of Atlantic sturgeon populations based on mitochondrial DNA control region sequences. *Trans Am Fish Soc* 129:476–486
- Wright S (1943) Isolation by distance. *Genetics* 31:39–59