SIMCOAL: A General Coalescent Program for the Simulation of Molecular Data in Interconnected Populations With Arbitrary Demography

L. Excoffier, J. Novembre, and S. Schneider

SIMCOAL (version 1.0) is a computer program for the simulation of molecular genetic diversity in an arbitrary number of haploid populations examined for a set of fully linked loci. It is based on the retrospective coalescent approach initially described by Kingman (1982a,b), and clearly exposed in a series of other articles (Donnelly and Tavaré 1995; Ewens 1990; Hudson 1990). The coalescent backward approach does not simulate the genetic history of the whole population, like in conventional forward simulations, but rather reconstructs the gene genealogy (coalescent history) of samples of genes drawn from different demes in a population. For neutral genes, this coalescent process essentially depends on the history and on the demography of the population, and is independent from the mutational process. In our program, we simulate mutations starting from the most recent common ancestor (MRCA) of all genes in the sample, and add them independently on all branches of the genealogy assuming a uniform and constant Poisson process. Using this two-step (coalescent-mutation) approach, many replicates of haploid samples of DNA sequences, RFLP, or microsatellite data can be simulated very quickly. The analysis of a large number of simulated samples allows one to obtain the empirical distribution of practically any statistic that can be derived from genetic data, including statistics for which no analytical derivation is available (Hudson 1993). Typical applications of our program include the study of the effect of complex demographies on the pattern of genetic diversity within and between populations, like in the case of bottlenecks, complex cases of admixture, or metapopulation systems. While our program generates haploid samples of genes or haplotypes, diploid data can be generated under the hypothesis of Hardy–Weinberg equilibrium by taking random pairs of haplotypes to form diploid genotypes. The program runs on PC-compatible computers under Windows 95/98/NT, but we provide the C++ source code that should compile under other operating systems, provided the compiler follows the latest C++ ISO committee specifications. It is freely available on the website at http://anthropologie.unige.ch/~laurent/simcoal.

Principle of the Simulation Algorithm

The coalescent algorithm we have implemented is inspired from that described in Hudson (1990), but it is based on a generation-by-generation approach instead of a continuous time approximation. We start from a given sample of genes that are found in one or more demes connected by an arbitrary pattern of migrations.

We first simulate the gene genealogy of the sample, independently from the mutational process, going backward one generation at a time. At each generation we first look for the occurrence of what we call an "historical event." Historical events are fully defined in the input file (see below) and allow one to implement, at any generation, a population fission, massive migration, rescaling of deme size, a change of population growth pattern, and/or modification to the pattern of migrations between populations. Second, we implement a phase of migration according to a predefined migration matrix between populations. Third, we check for potential coalescent events within each deme. In each deme, the probability of a coalescent event is given by $i(t)[i(t) - 1]/[2N(t)]$ (Kingman 1982a), where $i(t)$ is the number of gene lineages remaining in a deme at time t, and $N(t)$ is the total number of gene copies in a deme at time t. N can either remain stationary, vary exponentially from generation to generation, or drastically change after an historical event. Only one coalescent event is allowed per deme per generation, which is a very good approximation if N is much larger than i. If a coalescent event happens, two lineages of the deme are chosen at random to coalesce. We record the coalescence time and which lineages have coalesced. Finally, we resize the deme size according to the predefined pattern of demographic growth. The coalescence and deme size rescaling are performed independently in each deme before going to the previous generation, and restarting the four steps defined above. The process is carried out backward in time until all gene lineages have coalesced into a single lineage. The realized gene genealogy can be seen as a binary tree, with terminal nodes being the initial sampled genes, internal nodes being the result of the coalescence of lineages, and a root node being the MRCA of the sample.

Once a genealogy is obtained, we implement the mutational process. We recursively explore the binary tree, starting from the root node, and randomly assign mutations to each branch of the tree according to a Poisson distribution with parameter $\mu\, t_j$, where μ is the mutation rate per generation and per locus, and t_j is the length (in generations) of the jth branch. The impact of these mutations will depend on the mutation model and the type of data to simulate, which can either be DNA sequences, RFLP haplotypes, or a series of fully linked microsatellite loci. Before adding the simulated mutations on the tree, we assign arbitrary initial states to the an-

cestral gene at the root node. For each branch of the tree we then recursively assign mutations to particular sites or loci. Each site can therefore be hit several times (finite sites model) in the history of the sample. We describe below in more details the particularities of the mutation models we have implemented for each data type.

Output Files

The final state of each gene is recorded and output to several files according to two data formats. The first format is compatible with the Arlequin program (Schneider et al. 2000). The data generated after each simulation are output to a separate Arlequin project file. An Arlequin batch file is also created, listing all simulated files, and allowing one to compute different statistics on the whole set of simulated files with Arlequin, and thus to obtain their distribution after extraction from the Arlequin output files. We provide a complete example of such analysis on our website (http://anthropologie.unige.ch/~laurent/simcoal). Two other sets of output files are compatible with the NEXUS file format. The firs set includes two files with the "*.trees" extension that list all the simulated trees, with branch lengths expressed either in units of generations scaled by the population size (N), and therefore representing the true coalescent history of the sample of genes, or in units of average number of substitutions per site, and therefore representing the realized mutational tree. These two sets of trees can, for instance, be visualized with the software TREEVIEW (Page 1996). Second, the other set includes a single file with the "*.paup" extension. It lists all the simulated genes together with their true genealogic structure. This file can be analyzed with David Swofford's PAUP* software (1999).

Mutation Models

RFLP Data

Only a pure two-allele model is implemented. Several fully linked RFLP loci can be simulated, assuming a homogeneous mutational process over all loci. A finite-sites model is used, and mutations can hit the same site several times, switching the RFLP site on and off. We thus assume that there is the same probability for a site loss or for a site gain.

Microsatellite Data

We have implemented a pure stepwise mutation model (SMM) with or without constraints on the total size of the microsatellite. Several fully linked microsatellite loci can be simulated under the same mutation model constraints.

DNA Sequence Data

We have implemented here several simple finite-sites mutational models. The user can specify the percentage of substitutions that are transitions (the transition bias), the amount of heterogeneity in mutation rates along a DNA sequence according to either a discrete or continuous gamma distribution. We can therefore simulate DNA sequences under a Jukes and Cantor (1969) model or under a Kimura (1980) two-parameter model, with or without gamma correction for heterogeneity of mutation rates (Jin and Nei 1990). Other mutation models that depend on the nucleotide composition of the sequence were not considered here because of their complexity and because they require specifying many additional parameters, like the mutation transition matrix and the equilibrium nucleotide composition.

Demographic Models

There is no limitation on sample sizes, deme sizes, or number of simulated demes, other than the available memory in the computer and the computational time. Deme sizes should nevertheless always be much larger than the sample sizes for the single coalescent event per generation assumption to hold. Each sample of gene is assumed to be drawn from a different deme that can exchange an arbitrary number of migrants with other demes at any generation. As the migration rates are specified by a migration matrix of size equal to the number of simulated demes, either an island model, a one-dimensional, or a two dimensional stepping-stone model can be simulated, with demes arranged on a flat surface, a cylinder, or a torus. Several migration matrixes can be specified in the input file, so that any matrix can be used at any time in the simulation process. The history of several completely separated demes can also be simulated, as in the case of a series of population fissions or admixtures. Historical events can be defined to specify which parental populations are used to produce daughter populations after a fission event, as well as its timing. Note that the program can also be used to simulate the genetic divergence of different species, with or without bottlenecks. The size of each deme can either remain stationary over time or can change exponentially, separately in each population. Positive or negative exponential growths can be modeled. Historical events can be used to instantaneously modify the size of a deme or to alter its exponential growth pattern at any generation in the past.

Input File

A single input file is needed for the simulation. It should specify the number of gene samples to simulate, the sample sizes, the deme sizes, the exponential growth rate of each deme, the number of linked loci to simulate, and the overall mutation rate for all simulated loci. One can also include a series of historical events that can be used to resize a deme; implement an episode of migration from a source to a sink deme (for instance to simulate a population splitting into two subpopulations); reset the exponential growth rate of a given deme; or switch from a given migration matrix to another one. Note that several historical events can happen in the same generation. The data type (RFLP, microsatellite, or DNA) must also be specified in the input file. For DNA data, the proportion of substitutions being transitions must be specified to set a given transition bias, as well as the amount of rate heterogeneity among sites, which can be done by setting the alpha parameter of a gamma distribution. For microsatellite data, the potential size range constraint must also be provided. Additional information on the syntax of the input files can be found on the website mentioned above.

General Considerations

The simulations are usually very fast, allowing one to perform thousands of simulations in a reasonable amount of time. For instance, generating 1000 samples of 20 sequences of 300 bp in 20 demes under an island model ($Nm = 0.5$; $\theta_{\text{intrademe}} = 2N\mu = 1$, where m is the migration rate between demes, μ is the mutation rate for 300 bp, and $N = 1000$ genes per deme) takes about 39 min 30 s on a 400 MHz PC. Generating the same samples with a larger migration rate ($Nm = 10$) takes 34 min 15 s. Finally, generating 1000 samples of 400 sequences of 300 bp takes only 2 min in a single population of size 20,000, and 2 min 12 s in a single population of size 40,000. The execution time thus depends primar-

ily on the size of the demes and the overall migration pattern, which will affect the total number of generations to simulate before reaching the MRCA of all sampled genes. In the absence of migration (as in a single population) the simulations are much faster for each generation because we do not need to check each remaining lineage to decide if it migrates or not. Other factors, like an exponential growth rate, the mutation model, the mutation rate, or the heterogeneity of mutation rates for DNA sequences have less effect on the total execution time. Note, however, that a program based on a continuous time approximation, as described in Hudson (1990) would be much faster, but it would be very difficult to make it as general as the present one. In most instances, our experience tells us that it takes far more time to analyze the simulated data than to generate them.

We have checked for simple cases that our simulations lead to expected results regarding coalescence times. We give here just two simple examples. We first simulated a sample of two genes drawn from a haploid population of size 10,000. Over 50,000 replicates, the mean coalescence time and its standard error were found to be 10,007 and 9983, respectively, in very close agreement with the expected value of 10,000 generations for both parameters. We also simulated two samples of five genes drawn from two haploid populations of size 1000 that exchange migrants at a rate of 0.0001. Over 50,000 replicates, the average coalescent times of genes within and between populations were found equal to 1992 and 6999, in close agreement with their expected values of 2000 and 7000, respectively (see Slatkin 1991:170). The accuracy of the simulation of the DNA mutation models can be seen in Table 3 of Excoffier and Yang (1999), where we have simulated a Kimura two-parameter model assuming a gamma distribution of mutation rates ($\alpha = 0.26$ and $\alpha = 0.40$) and 10 times more transitions than transversions ($\kappa = 20$). These parameters were correctly estimated from simulated data with maximum likelihood methods when the overall divergence of the simulated sequences was large. Note that the estimated mutation parameters were biased for short divergence times between sequences, which was due to the inefficiency of maximum likelihood methods in that case (Excoffier and Yang 1999).

Even though our simulation program is designed to simulate haploid data, it could be used to simulate diploid data under the hypothesis of Hardy–Weinberg equilibrium, by grouping pairs of haplotypes to form diploid genotypes. However, because we simulate haploid migrants, it would not be safe to use our program to simulate high levels of gene flow with diploid migrants, where one could expect comigration of lineages during the simulated period. Our program has been used extensively over the last few years for a variety of problems, like checking the validity of molecular estimators of admixture (Bertorelle and Excoffier 1998), checking our ability to estimate heterogeneity of mutation rates for human and chimpanzee mtDNA (Excoffier and Yang 1999), studying population divergence with bottlenecks (Gaggiotti and Excoffier 2000), or studying the effect of genetic barriers between geographically structured groups of demes (Dupanloup de Ceunink et al. 2000). We have not yet explored all the potential of the program, but we think its versatility will be valuable to other researchers for getting expectations and distribution of different statistics in complex demographic models, or to check by fast simulation the behavior of new parameter estimators. The SIMCOAL program is freely available at http://anthropologie.unige.ch/~laurent/simcoal.

An Applied Example: Human-Chimpanzee Divergence

SIMCOAL was used to get empirical confidence intervals around estimated parameters of human-chimpanzee divergence in a recent article (Gaggiotti and Excoffier 2000). A new method was developed to estimate divergence times between populations of unequal sizes from the mean number of pairwise differences within and between populations [more details can be found in Gaggiotti and Excoffier (2000)]. The parameters we estimated from the molecular diversity found among 2996 human and 222 chimpanzee (*Pan troglodytes*) mitochondrial DNA sequences of 277 bp were the following: (1) divergence time $T = 2\mu t = 48.47$; (2) size of the ancestral population $\theta_0 = 2N_0\mu = 32.04$; (3) relative sizes of the chimpanzee and human populations (compared to the ancestral population) being $k = 0.80$ and $1 - k = 0.20$, respectively. From these parameters, we estimated the mutation rate for the whole sequence assuming an effective population size of approximately 5000 females for humans (e.g., Vigilant et al. 1989) as $\mu = \theta_0/(2N_0) = 48.47/50{,}000 = 6.4 \times 10^{-4}$. The time of divergence between humans and chimps was then obtained as $t = T/(2\mu) = 37{,}867$ generations. We therefore took those parameters to simulate the genetic diversity of 2996 human and that of 222 chimpanzee sequences in 1000 replicates from which new estimates of θ_0, T, and k were each time obtained. The 2.5 and 97.5 percentile values of the three empirical distributions were then used as approximate limits of 95% confidence intervals around the estimated parameters.

From the Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, 12 rue G. Revilliod, 1227 Geneva, Switzerland (Excoffier, Novembre, and Schneider) and Colorado College, Colorado Springs, Colorado (Novembre). We are grateful to Ziheng Yang and Alan Rogers for providing us with source code for generating discrete and continuous gamma variates, respectively. This work was made possible by Swiss National Science Foundation grant no. 31-056755.99. J.N. was funded by the Howard Hughes Medical Institute Grant to the Colorado College. Address correspondence to Laurent Excoffier at the address above or e-mail: Laurent.Excoffier@anthro.unige.ch.

References

Bertorelle G and Excoffier L, 1998. Inferring admixture proportions from molecular data. Mol Biol Evol 15: 1298–311.

Donnelly P and Tavaré S, 1995. Coalescents and genealogical structure under neutrality. Annu Rev Genet 29: 401–421.

Dupanloup de Ceunink I, Schneider S, Langaney A, and Excoffier L, 2000. Inferring the impact of linguistic barriers on population differentiation: application to the Afro-Asiatic-Indo-European case. Eur J Hum Genet (in press).

Ewens WJ, 1990. Population genetics theory—the past and the future. In: Mathematical and statistical developments of evolutionary theory (Lessar S, eds). Dordrecht: Kluwer Academic; 177–227.

Excoffier L and Yang Z, 1999. Substitution rate variation among sites in the mitochondrial hypervariable region I of humans and chimpanzees. Mol Biol Evol 16:1357–1368.

Gaggiotti O and Excoffier L, 2000. A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. Proc R Soc Lond B 267:81–87.

Hudson RR, 1990. Gene genealogies and the coalescent process. In: Oxford surveys in evolutionary biology (Futuyma DJ and Antonovics JD, eds). New York: Oxford University Press; 1–44.

Hudson RR, 1993. The how and why of generating gene genealogies. In: Mechanisms of molecular evolution (Takahata N and Clark AG, eds). Sunderland, MA: Sinauer; 23–36.

Jin L and Nei M, 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7:82–102.

Jukes T and Cantor C, 1969. Evolution of protein molecules. In: Mammalian protein metabolism (Munro HN, eds). New York: Academic Press; 21–132.

Kimura M, 1980. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.

Kingman JFC, 1982a. The coalescent. Stoch Proc Appl 13:235–248.

Kingman JFC, 1982b. On the genealogy of large populations. J Appl Prob 19A:27–43.

Page RDM, 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput Appl Biosci 12:357–358.

Schneider S, Roessli D, and Excoffier L, 2000. Arlequin: a software for population genetics data analysis. User manual, ver 2.000. Geneva: University of Geneva.

Slatkin M, 1991. Inbreeding coefficients and coalescence times. Genet Res Camb 58:167–175.

Swofford DL, 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods), ver. 4.0. Sunderland, MA: Sinauer.

Vigilant L, Pennington R, Harpending H, Kocher H, and Wilson AC, 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci USA 86:9350–9354.

Received October 10, 1999
Accepted June 15, 2000

Corresponding Editor: Stephen J. O'Brien

SPAM (Version 3.2): Statistics Program for Analyzing Mixtures

E. M. Debevec, R. B. Gates, M. Masuda, J. Pella, J. Reynolds, and L. W. Seeb

A number of recent studies have addressed individual identification and population-specific discrimination of individuals (e.g., Banks and Eichert 2000; McParland et al. 1999; Olsen et al. 2000; Shriver et al. 1997). However, often the primary interest is the relative contributions of distinct population segments in an admixture of organisms. This coarser focus is particularly relevant in complex mixtures with multiple contributors and applications where differentiation among populations is not sufficient for individual identification. Such mixed population analysis has been applied extensively in the Pacific Northwest and Alaska, where local salmon populations may spawn in separate freshwater streams but are harvested together in salt- or freshwater fisheries (e.g., Scribner et al. 1998; Seeb and Crane 1999). The statistical theory developed to estimate the contributions of discrete populations or stocks in a mixture is commonly called mixed stock analysis (MSA) or genetic stock identification (GSI) in the fisheries literature (Fournier et al. 1984; Millar 1987; Pella and Milner 1987). This approach has also been utilized as an introgression index to calculate the percentage of genes from source or parental populations (e.g., Planes and Doherty 1997).

The first step in performing mixture analyses is to assemble a baseline of genotypic and/or phenotypic characteristics for all populations present in the mixture. The mixture is then randomly sampled and the same suite of characteristics is measured for each individual in the sample. The maximum likelihood contribution estimates for the baseline populations are the unique mixture proportions that lead to the greatest likelihood of obtaining a mixture sample with the observed set of characters.

The Statistics Program for Analyzing Mixtures (SPAM) is a software tool for computing the maximum likelihood estimates for a mixed population or stock analysis. SPAM has two basic modes of operation. Estimation mode uses baseline and mixture information to estimate stock contributions and their associated measures of precision. Simulation mode uses baseline information to generate hypothetical mixtures with known contributions to assess baseline potential. SPAM provides a common framework for implementing all aspects of a mixed stock analysis within the Windows 95/98 operating system. The use of ASCII files for all input and output is facilitated by the SPAM session environment which monitors analysis progress and allows for editing and/or viewing any of its associated files.

Although originally developed for loci such as allozymes displaying relatively low variability, SPAM can be used with highly variable loci such as microsatellites. Because SPAM calculates genotype probabilities across multiple loci, probabilities can become exceedingly small with many highly variable loci. Smouse and Chevillon (1998) reviewed the analytical aspects of population-specific discrimination for individuals and concluded a modest number of codominant loci, each with a small number of alleles and each allele in moderate frequency, provide the highest potential discriminatory power.

SPAM requires two or more ASCII input files for analyses. Detailed formatting instructions for each input file is provided in the SPAM documentation. Baseline files describe the genetic and/or phenotypic makeup of each population potentially in the mixture. Three character types can be used: PHENOTYPE or MTDNA, with one response per individual; LOCUS, with two responses per individual; or ISOLOCUS, with four responses per individual. Isoloci, resulting from a duplication event, share alleles with identical electrophoretic mobility (Allendorf and Thorgaard 1984), and are expected to include four gene doses per individual. No assumptions concerning inheritance are made. Isolocus data are treated as equal frequencies at two disomic loci, equivalent to equal frequencies at a single tetrasomic locus. Baseline populations can be described with either relative or absolute frequencies for each character.

Estimation mode requires a mixture file providing character data of the individuals in the mixture sample. The mixture sample is assumed to be representative of the true mixture. SPAM can select a subset of all characters described in the baseline and mixture files for use in an analysis, eliminating the need to create multiple input files for a reduced set of characters. A control file is used in every SPAM analysis to define input and output options, as well as to set parameters used in the maximum likelihood search for contribution estimates.

SPAM searches for maximum likelihood estimates of population proportions using three numerical algorithms: conjugate gradient (CG), iteratively reweighted least squares (IRLS), and expectation-maximization (EM). A detailed description of these algorithms can be found in Masuda et al. (1991) and Pella et al. (1996). Standard error estimates are available from either the infinitesimal jackknife or bootstrap resampling. Confidence intervals can be calculated using bootstrap resampling of just the mixture, just the baseline, or both. Estimates, standard errors, and confidence intervals can be requested for user-defined population aggregates to improve the accuracy of estimates by pooling populations that are genetically indistinguishable (Wood et al. 1987).

Primary results from SPAM are output in either an estimation file or a simulation file. Other output files can be requested, including bootstrap summaries, baseline and mixture summaries, conditional genotype and population probabilities, and population contribution estimates for each resample.

SPAM provides a convenient Windows environment from which the user can edit input files, perform analyses, and view output files. Every analysis produces a child window that monitors progress and reports the creation of output files. When an analysis is complete, the user can select an output file for viewing from the Results menu. The selected file is opened in an external text editor chosen by the user,

or in the default Notepad text editor. SPAM analyses can be run one at a time or in batch mode by selecting multiple control files for submission. Complete details are provided in the full SPAM documentation. Associated input ASCII files can be created in any text editor or generated using other software.

SPAM can be obtained free of charge from the Gene Conservation Laboratory, Alaska Department of Fish and Game. A zipped file containing the executable program, sample files, and complete documentation can be downloaded from the website at http://www.cf.adfg.state.ak.us/geninfo/research/genetics/Software/SpamPage.htm. Alternatively, requests can be made by e-mail to Joel_Reynolds@fishgame.state.ak.us.

From the Gene Conservation Laboratory, Division of Commercial Fisheries, Alaska Department of Fish and Game, 333 Raspberry Rd., Anchorage, AK 99518 (Debevec, Gates, Reynolds, and Seeb), Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, Alaska (Debevec), MathSoft, DAPD, Seattle, Washington (Gates), and Alaska Fisheries Science Center, Auke Bay Laboratory, Juneau, Alaska (Masuda and Pella). This article is contribution PP-194 of the Alaska Department of Fish and Game, Commercial Fisheries Division, Juneau, Alaska. Address correspondence to Lisa W. Seeb at the address above.

References

Allendorf FW and Thorgaard GH, 1984. Tetraploidy and the evolution of salmonid fishes. In: Evolutionary genetics of fishes (Turner BJ, ed). New York: Plenum; 1–53.

Banks MA and Eichert W, 2000. WHICHRUN (version 3.2): a computer program for population assignment of individuals based on multilocus genotype data. J Hered 91:87–89.

Fournier DA, Beacham TD, Riddell BE, and Busack CA, 1984. Estimating stock composition in mixed stock fisheries using morphometric, meristic and electrophoretic characteristics. Can J Fish Aquat Sci 41:400–408.

Masuda M, Nelson S, and Pella J, 1991. User's manual for GIRLSEM and GIRLSYM: the computer programs for computing conditional maximum likelihood estimates of stock composition from discrete characters. U.S. Department of Commerce, NOAA-NMFS, Auke Bay Laboratory, Auke Bay, Alaska.

McParland TL, Ferguson MM, and Liskauskas AP, 1999. Genetic population structure and mixed-stock analysis of walleyes in the Lake Erie–Lake Huron corridor using allozyme and mitochondrial DNA markers. Trans Am Fish Soc 128:1055–1067.

Millar RB, 1987. Maximum likelihood estimation of mixed stock fishery composition. Can J Fish Aquat Sci 44:583–590.

Olsen JB, Bentzen P, Banks MA, Shaklee JB, and Young S, 2000. Microsatellites reveal population identity of individual pink salmon to allow supportive breeding of a population at risk of extinction. Trans Am Fish Soc 129: 232–242.

Pella JJ, Masuda M, and Nelson S, 1996. Search algorithms for computing stock composition of a mixture from traits of individuals by maximum likelihood. NOAA technical memo NMFS-AFSC-61. Washington, DC: U.S. Department of Commerce.

Pella JJ and Milner GB, 1987. Use of genetic marks in stock composition analysis. In: Population genetics and fishery management (Ryman N, Utter F, eds). Seattle: Washington Sea Grant, University of Washington Press; 247–276.

Planes S and Doherty PJ, 1997. Genetic and color interactions at a contact zone of *Acanthochromis polyacanthus:* a marine fish lacking pelagic larvae. Evolution 51: 1232–1243.

Scribner KT, Crane PA, Spearman WJ, and Seeb LW, 1998. DNA and allozyme markers provide concordant estimates of population differentiation: analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon. Can J Fish Aquat Sci 55:1748–1758.

Seeb LW and Crane PA, 1999. Allozymes and mtDNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the South Alaska peninsula. Trans Am Fish Soc 128:88–103.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, and Ferrell RE, 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 60:957–964.

Smouse PE and Chevillon C, 1998. Analytical aspects of population-specific DNA fingerprinting for individuals. J Hered 89:143–150.

Wood CC, McKinnell S, Mulligan TJ, and Fournier DA, 1987. Stock identification with the maximum-likelihood mixture model: sensitivity analysis and application to complex problems. Can J Fish Aquat Sci 44:866–881.

Received December 7, 1999
Accepted August 30, 2000

Corresponding Editor: Robert Angus