Sex-linked Markers Facilitate Genetic Parentage Analyses in Knobbed Whelk Broods

D. WALKER, A. J. POWER, AND J. C. AVISE

From the Department of Genetics, University of Georgia, Athens, GA 30602 (Walker and Avise); Marine Extension Service, 20 Ocean Science Circle, University of Georgia, Savannah, GA 31411–1011 (Power).

Address correspondence to DeEtte Walker, Georgia Institute of Technology, 315 Ferst Drive, Atlanta, GA 30332–0363, or e-mail: lw158@mail.gatech.edu.

Abstract

To explore the potential of sex-linked polymorphisms for genetic parentage analyses in natural populations, we have employed a recently discovered “X-linked” microsatellite marker (in conjunction with polymorphic autosomal loci) to deduce biological paternity and maternity for large numbers of encapsulated embryos within individual broods of the knobbed whelk (*Busycon carica*). Empirical findings illustrate how such sex-linked genetic tags can in special instances find at least three novel utilities in genetic dissections of large-clutch species: clarification of paternity assignments that had remained ambiguous from di-locus autosomal data alone; elucidation of linkage relationships among pairs of autosomal loci; and illumination of maternity (and thereby paternity also) in broods for which neither biological parent was known from independent evidence.

By virtue of their high polymorphism and simple Mendelian inheritance (Goldstein and Schlötterer 1999), autosomal microsatellite loci often provide superb markers for assessing genetic paternity and maternity (Jones and Ardren 2003). Most such analyses on natural or captive animals have been conducted on small-clutch species such as birds (Birkhead and Moller 1992) and mammals (e.g., Coltman et al. 1999). For organisms with much larger clutches, such as many fishes (Avise 2001) and invertebrates (e.g., Brockman et al. 2000), genetic parentage analyses pose special challenges in data collection and statistical analysis (e.g., Fiumera et al. 2001; Neff 2001), but they also offer exceptional opportunities to characterize rates and patterns of de novo mutations (Jones et al. 1999) and to assess details of sexual selection and population demographics (review in Avise et al. 2002).

A favorable biological setting for paternity or maternity analysis occurs when one parent of a clutch is known or suspected from independent evidence (such as pregnancy or nest-guarding). By comparing genotypes of the known parent and its offspring, alleles contributed by the brood’s other-sex parent(s) can be deduced and sometimes used to quantify those parents’ relative genetic contributions to a clutch. Jones and Avise (1997; see also DeWoody et al. 2000a, b) introduced a quantitative graphical method for deducing genetic parentage in large half-sib clutches based on di-locus summaries. The approach in principle identifies full-sib cohorts within a multi-sire (or multi-dam) brood, based on patterns of allelic association at pairs of polymorphic autosomal loci. The method works well when the biological parents of a brood do not share alleles, but it has limitations when polymorphism is weak or when the number of parents contributing to a brood is high relative to the number of offspring sampled.

One way to offset these limitations is to “brute-force” the parentage analyses by assaying large numbers of variable loci (thereby increasing the probabilities of parental exclusion). However, suitable microsatellite loci are not easy to identify in all species, and the assays are also rather expensive and labor intensive. Furthermore, as the number of loci scored increases, so too (but at a much faster rate) does the number of multi-locus genotypic combinations among the embryos of multiply sired broods, meaning that even if many offspring are assayed, few will fall into particular multi-locus “cells.” Because numerical counts of embryos tend to be much higher in pairwise or di-locus treatments, such data analyses can remain valuable for specifying particular sire genotypes. Thus, novel or elegant analyses that can extract clear parentage information from even small numbers of microsatellite loci and modest numbers of progeny are of both academic and utilitarian interest. Here, using empirical...
In the knobbed whelk *Busycon carica* (Mollusca, Gastropoda, Melongenidae), a breeding female lays a string of egg capsules collectively containing hundreds or thousands of tiny embryos (Power et al. 2002). We collected adults and egg-case strings from oyster bars and intertidal mud flats near Wassaw Island, Georgia. Some females were captured in the act of depositing an egg string, so these were “known” mothers of their respective broods. Other egg strings were collected not in association with any particular female. Embryos in these strings represent “neither-parent-known” situations.

For each brood, individual embryos (and their mother, if known) were genotyped at each of two autosomal microsatellite loci (*bc3.4* and *bc3.12*) and one X-linked microsatellite locus (*bc2.2*). Other adult males and females sampled from the same locality also were genotyped. DNA extraction methods, PCR primers, and PCR assay conditions are detailed in Avise et al. (2004). PCR products were electrophoresed through 4.2% acrylamide gels with 2.1 μl deionized formamide, 0.5 μl loading buffer, and 0.3 μl ROX 500 Size Standard (Applied Biosystems). GENESCAN 3.1 and GENOTYPER 2.5 software were used to analyze data collected from runs on an ABI377 automated sequencer.

The program GENEPOL (Raymond and Rousset 1995) was used to estimate allele frequencies in the adult population, to test genotypic frequencies for Hardy-Weinberg proportions (Guo and Thompson 1992), and to assess linkage disequilibrium at the population level. Genetic exclusion probabilities were calculated under either of two assumptions (Selvin 1980): that one parent (the mother) was known and that neither parent was known. Other analysis methods and their underlying logic, detailed below, relate explicitly to parentage analyses facilitated by the sex-linked polymorphism.

### Results and Discussion

Totals of 12 and 36 different alleles were detected at autosomal loci *bc3.4* and *bc3.12*, respectively, in our population sample of 139 adult whelks, and expected heterozygosities were 0.82 and 0.96 (Avise et al. 2004). No significant deviations from Hardy-Weinberg equilibrium (HWE) were detected, nor were these two loci demonstrably in linkage disequilibrium in the adult sample (all $P > 0.05$). Under a model in which it is assumed that one parent is known, the combined genetic exclusion probability across these two autosomal loci was 0.97; under the “neither-parent-known” model, the combined exclusion probability was 0.93.

At the sex-linked microsatellite locus (*bc2.2*), 15 distinct alleles were detected in adult whelks (Figure 1), and expected heterozygosity was 0.89. As shown previously (Avise et al. 2004), alleles at *bc2.2* behave analogously to X-linked alleles in mammals: each male is hemizygous (having received his single allele from his mother), and each female (the heterogametic sex) carries two allelic copies (one from her sire and one from her dam).

Several formal models are available for investigating various aspects of genetic parentage under a variety of genetic assumptions and biological scenarios (review in Jones and Ardren 2003). Here we use a straightforward empirical approach to introduce the notion that in addition to sex-ratio estimation (Avise et al. 2004), there are at least three other ways that a highly polymorphic sex-linked locus, by virtue of its special mode of genetic transmission, can contribute to genetic parentage analyses in large-clutch species: in clarifying paternity assignments that had remained uncertain from di-locus autosomal data alone; by facilitating linkage assessments among pairs of autosomal loci; and in improving maternity analyses (and thereby paternity analyses also) in broods for which neither biological parent was known at the outset from independent evidence. Other researchers have identified and utilized sex-linked microsatellites (or other sex-linked molecular markers) to assist with kinship assessments in nature (e.g., Buchholz et al. 1998; Komdeur et al. 1997; Rabenold et al. 1991); but to our knowledge, this is the first study to explore special applications for such markers in parentage analyses of large-clutch broods (but see also Imhof et al. 1998). To exemplify these benefits, we will employ whelk broods 564F and U4. Although these broods represent best-case scenarios, they serve to illustrate the special kinds of parentage information that sex-linked loci can provide in one-parent-known and neither-parent-known settings, respectively.

#### Paternity Analysis

From brood 564F, we first genotyped several hundred embryos and their known dam at each of two autosomal

**Figure 1.** Histogram of allele frequencies at sex-linked locus *bc2.2* in our sample of 139 adult knobbed whelks from Wassaw Island.
microsatellite loci. At each locus considered singly, each embryo’s paternally derived allele is apparent by “subtracting” that embryo’s maternally derived allele from its diploid genotype. An accumulation of such data for many embryos in a brood yields numerical counts of paternal alleles, and, following Jones and Avise (1997) and DeWoody et al. (2000b), these can be grouped into di-locus blocks in an effort to identify paternal “gametotypes” at both loci jointly (Table 1). If the sires share no marker alleles with one another or with the known dam, such tables should reveal discrete cohorts of offspring (often in recognizable blocks of four; i.e., when each sire is heterozygous at both loci). In such cases, di-locus genotypes of all sires of a brood can be specified, and the relative contributions of those sires can thereby be tallied. However, Table 1 indicates unresolved ambiguities that severely obscured the identification of di-locus paternal gametotypes in brood 564F. These ambiguities presumably arose because some of the sires must have shared alleles at the autosomal marker loci employed.

We then added the X-linked microsatellite locus (bc2.2) to the analysis of brood 564F, and this enabled us to fully resolve these paternity ambiguities using the following four-step procedural chain of logic:

(a) Group heterozygous daughters according to paternal allele at the X-linked gene

Because males are hemizygous and daughters inherit their father’s one-and-only sex-linked allele at bc2.2, and because the mother’s genotype is known, each heterozygous daughter (i.e., each individual that displays two bands on gels at the mother’s genotype is known, each heterozygous daughter bc2.2 Because males are hemizygous and daughters inherit their allele at the X-linked gene

(b) Generate a table of tri-locus paternal gametotypes for these heterozygous daughters

This table consists of paternally deduced alleles from the two autosomal loci. It follows the format in DeWoody et al. (2000b) but in this case applies strictly to heterozygous daughters that have been categorized for paternity according to the X-linked marker (step (a)). For brood 564F, the tri-locus summary for such daughters is presented in Table 2, from which it is evident that the brood must have had five sires (henceforth A–E) whose entire tri-locus genotypes now can be specified exactly (see footnote to Table 2). The additional father comes from the fact that sires C and D must have shared sex-linked allele 309 and, hence, were not distinguished by microsatellite locus bc2.2 alone. This analysis also shows that at locus bc3.4, allele 215 was shared by two sires (A and D), and alleles 211 and 213 were shared by three sires each (B, C, E and B, C, D, respectively).

(c) Repeat the analysis for homozygous daughters and hemizygous sons

The remaining embryos in the brood (those displaying a single band on gels at microsatellite locus bc2.2) could next be grouped into provisional sire categories based on the di-locus paternal genotypes identified for the heterozygous daughters in steps (a) and (b) above. Results for brood 564F are presented in Table 3, from which it is again clear that the same five sires likewise contributed to this portion of the brood.

(d) Sum the paternity counts for daughters and sons

By combining the paternity counts from Tables 2 and 3, the total absolute and relative contributions of different sires to the brood can be estimated. In the case of brood 564F, exactly five fathers (A–E) sired the following numbers (and proportions) of assayed offspring: 188 (47.6%), 97 (24.5%), 64 (16.2%), 18 (4.6%), and 28 (7.1%).
Linkage Analysis

An additional benefit of displaying paternal genotypes as in Tables 2 and 3 is that this facilitates evaluations of possible genetic linkage (DeWoody et al. 2000b). In the current case, autosomal loci bc3.4 and bc3.12 do indeed appear to reside on one chromosome, as evidenced by consistent inequalities between genotypic counts for parental and recombinant di-locus paternal genotypes in brood 564F. Thus, based in part on paternity information contributed by the X-linked marker, the parental and recombinant configurations of autosomal alleles at bc3.4 and bc3.12 could be distinguished for all five fathers of brood 564F, and recombination rates during paternal gametogenesis could be calculated separately for each sire’s germline (Figure 2). With our sample sizes of surveyed offspring, these recombination rates did not differ significantly across the five sires (independence G-test; G = 6.2, df = 4, P > .10). Based on total progeny counts across the five paternal half-sib cohorts in brood 564F, we estimate that autosomal loci bc3.4 and bc3.12 are separated by about 20.2 cM.

Maternity Analysis

A third utility for sex-linked polymorphisms arises in facilitating maternity analyses (and thereby paternity analyses also) in large broods for which neither parent is known from independent evidence. This can be illustrated by our genetic examination of 116 embryos from egg-case string U4, which was not collected in the presence of any adult female. Within that brood, at sex-linked locus bc2.2, all embryos displaying a single microsatellite band on the gels (i.e., presumptive hemizygous males or possibly homozygous females) carried either the 300 or the 306 allele, thus indicating that their dam must have been heterozygous 300/306. This was further evidenced by the presence of at least one copy of either allele 300 or 306 in all heterozygous embryos (known daughters) within that brood. In one or another of those heterozygous daughters, also displayed were paternal alleles 297, 303, 309, 315, 318, 330, and either 300 or 306. Thus (and barring de novo mutations), locus bc2.2 documents that at least seven sires clearly contributed to this brood. This estimate of seven sires, based solely on evidence from sex-linked locus bc2.2, is generally in line with comparable estimates from the two autosomal loci surveyed. At locus bc3.12, at least eight fathers were evidenced, because (following the logic of Kellogg et al. 1998) the embryos in brood U4 collectively displayed 15 different alleles of paternal origin (after subtracting two additional alleles likely to have been inherited from the dam). At autosomal locus bc3.4, genetic variation within the brood was less extensive, but the presence of seven alleles of paternal origin indicates genetic contributions from at least four sires.

Conclusion

With regard to paternity assessments in known-dam situations (e.g., when a female is caught in the act of...
egg-laying), our analysis of whelk brood 564F shows how a sex-linked marker can in some cases help to fully resolve sire genotypes that otherwise remained ambiguous due to allelic sharing between multiple fathers of a brood. With regard to assessments of genetic linkage in a natural population, further analyses of brood 564F show how the gametic phase of haplotypes at linked autosomal loci can be deduced with the assistance of a sex-linked marker. Traditionally, marker-based dissections of genomes (as in establishing linkage maps or analyzing quantitative trait loci) normally require use of controlled crosses or the availability of known pedigrees in large captive-bred populations (but see Moore and Kukuk 2002; Ritland 1996; Wu and Zeng 2001).

Despite the evidence (from offspring genotypes within brood 564F) that bc3.4 and bc3.12 are linked, these two autosomal loci did not appear to be in gametic-phase disequilibrium in the adult whelk population (admittedly, however, such disequilibrium would be hard to detect given the large numbers of alleles and genotypes at these microsatellite loci). This outcome is not surprising, because the estimated recombination rate ($r \geq 20\%$) between bc3.4 and bc3.12 implies that any initial disequilibrium should decay rapidly (as a function of $r$) such that a random-mating population is expected to approach gametic-phase equilibrium within a small number of generations (Hartl and Clark 1989; Hedrick 1985). This also implies that despite their physical linkage, loci bc3.4 and bc3.12 can provide high exclusionary power in parentage analyses (Chakraborty and Hedrick 1983). Indeed, given approximate gametic-phase equilibrium at the population level, linked loci can be interpreted as constituting a super-locus with more total potential alleles (and often a higher exclusionary power) than two otherwise comparable unlinked loci with independent assortment (see Jones et al. 1998).

Finally, with regard to parentage analyses in neither-parent-known broods, our data from whelk family U4 illustrate how a dam’s genotype can be deduced and the number of sires for a clutch can be closely estimated from a sex-linked marker, even alone. This is of interest because parentage assignments from autosomal loci are notoriously difficult to determine in natural settings where neither parent is known at the outset.

On the other hand, we also wish to emphasize that sex-linked markers do not offer a panacea for parentage analyses of large-clutch broods. Even a highly polymorphic sex-linked locus (as in the current situation involving whelks) cannot overcome all complications arising from shared alleles among parents, large numbers of sires (or dams) per brood, or limited sample sizes of offspring from a clutch. The two whelk families that we chose to highlight here were the best (i.e., most informative) examples among a dozen such families that we likewise have analyzed for genetic parentage (unpublished data). However, even in the other whelk broods for which $bc2.2’$ proved less optimal, this sex-linked locus nonetheless added significant information about genetic parentage, as we will document in a subsequent report detailing the temporal dynamics of sperm usage and multiple paternity across sequential egg cases within a clutch.

Acknowledgments

Work was supported by University of Georgia funds, by a Pew Foundation Fellowship in Marine Conservation to J.C.A., and by Marine Extension Funds to A.J.P. Field assistance was provided by Dodie Thompson, Rebecca Green, Mary Sweeney-Reeves, and Randal Walker.

References


Received March 23, 2004
Accepted August 3, 2004

Corresponding Editor: Robert Wayne