There has been a recent burst of studies of the function of genitalia, many of which share several important shortcomings. Given that further studies on this topic are likely (there are probably millions of species showing rapid genital divergence), I discuss the studies critically to promote clear formulation of hypotheses and interpretation of results in the future. I also emphasize some possibly important but neglected variables, including female stimulation, phylogenetic contexts, and the behavior of male genitalia, and outline simple techniques that could improve future studies.

Sexual selection and genitalia

One of the most widespread and well-documented phenomena in morphological evolution is the tendency for male genitalia to diverge relatively rapidly compared with other body parts [1–5]. Although this trend was previously thought to result from selection favoring mating barriers between species, post-copulatory sexual selection (i.e. sexual selection after male and female genitalia have come into contact) is currently thought to be largely responsible for rapid divergence in genital evolution [3,4,6,7]. However, the relative importance of three different types of sexual selection (sperm competition, cryptic female choice and sexually antagonistic coevolution) is debated [4,7,8].

Recently, there has been a welcome burst of experimental studies designed to test these hypotheses by checking the effects of male genital morphology and behavior on male reproductive success [9–28]. Given that many of these studies share important shortcomings, and because there is likely to be further work on this topic (after all, there are probably millions of species showing rapid genital divergence), it is useful to discuss them critically to help researchers formulate hypotheses and interpret results more precisely. Here, I emphasize some possibly important but neglected variables, including female stimulation, phylogenetic contexts and the behavior of genitalia, and outline simple techniques that could improve future studies.

Variables commonly measured

The variables most commonly measured in recent experimental studies of genital function have been the relationships between male genital morphology and either sperm transfer or sperm precedence when females mate with multiple males. Often these variables have been measured when females mated with males that differ in genital morphology or behavior [10–12,15,16,19–22,25]. Although sperm transfer and paternity are relevant, they provide an incomplete picture of possible genital functions. For instance, there are more than 20 female processes that could result in cryptic female choice [29] (Box 1). Both theoretical and empirical considerations provide strong reasons to expect that other effects that have seldom been tested (e.g. induction of oviposition and, or of female resistance to further mating) are often important. Under natural selection, females are likely to evolve to trigger oviposition and increased resistance to the sexual advances of other males on the basis of stimuli from copulation. In this way, females set up the conditions under which sexual selection on males could favor elaboration of male signaling or manipulative abilities [8,29]. In addition, studies of more than 60 species have shown that induction of oviposition and female refractory behavior are the two most widespread responses to male seminal products in insects [30]. Sexual selection on the communicative and manipulative properties of male seminal products is likely to have often been similar to that on male genitalia [8,29]. Finally, female re-mating and ovulation were both affected by male genital morphology in one group where these female responses were checked [26,27].

In sum, a failure to find an effect on sperm precedence and sperm transfer use does not justify rejection of post-copulatory sexual selection on genitalia [22] (nor does a similar finding with respect to copulatory courtship rule out post-copulatory sexual selection [10]).

The dauntingly long list of possible cryptic female choice mechanisms (Box 1) means that tests of post-copulatory selection are asymmetrical, in that confirmation is technically much easier than rejection [29]. Finding an effect on one female response is evidence that post-copulatory sexual selection is operating, but failing to find an effect on one, or even on several female responses, does not necessarily justify the conclusion that no such selection occurs. Only after testing a series of possibilities (the list will vary in different species) can the hypothesis be rejected. In other words, one of the strengths of the sexual selection hypothesis (sexual selection could arise in the context of many different female responses) also makes it especially difficult to test exhaustively. This is of course frustrating for researchers looking for clear tests of ideas and simple interpretations of experimental results. But nature is not always as uncomplicated as one might wish. I believe that the most powerful tests of the cryptic female choice hypothesis per se (tests that are most likely to result in confident rejection) are unlikely to come from working through lists such as that in Box 1, but from testing...
Box 1. The difficulty of ruling out cryptic female choice.

Many female-influenced processes (Table I) can affect the chances of any one male siring her offspring, and can be triggered by stimuli from either copulation or the genitalia of the male; therefore, such processes are capable of resulting in cryptic female choice on male genitalia (after [29]; see references therein for methods used to quantify female responses). Not all processes in Table I apply to all species (e.g. if there is no spermatophore or copulatory plug, cryptic female choice by refraining from removing these structures or by facilitating spermatophore discharge is not possible), and there are probably others. Multiple processes will be feasible in any given species, and tests of genital function that consider only a subset of the functions that are feasible in a species cannot provide convincing tests of the possibility of cryptic female choice in that species.

Table I. Female processes whose selective performance can result in cryptic female choice.

<table>
<thead>
<tr>
<th>Female process</th>
<th>Male response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allow penetration deep enough to enable sperm deposition at the optimum site for storage or fertilization</td>
<td>Add more nutrients to eggs before laying them</td>
</tr>
<tr>
<td>Refrain from terminating copulation prematurely</td>
<td>Oviposit all available mature eggs</td>
</tr>
<tr>
<td>Transport sperm to storage and fertilization sites</td>
<td>Prepare uterus for implantation</td>
</tr>
<tr>
<td>Modify internal conditions (e.g. pH) inside reproductive tract to reduce the defenses against microbial invasion that can kill sperm</td>
<td>Refrain from removing copulatory plug produced by male</td>
</tr>
<tr>
<td>Nourish or otherwise maintain sperm alive in storage site</td>
<td>Aid male in the formation of copulatory plug (add necessary components, or promote breakage of male genitalia to leave plugs)</td>
</tr>
<tr>
<td>Discard sperm from previous males</td>
<td>Modify genital morphology following first copulation to make subsequent insemination more difficult</td>
</tr>
<tr>
<td>Move sperm from previous males to a site where the current male can remove them</td>
<td>Refrain from removing spermatophore before all sperm are transferred</td>
</tr>
<tr>
<td>Accede to male manipulations that result in discharge of his spermatophore</td>
<td>Abort previously formed zygotes</td>
</tr>
<tr>
<td>Accelerate vitellogenesis (grow more immature eggs to maturity more quickly)</td>
<td>Refrain from aborting zygotes from current sperm</td>
</tr>
<tr>
<td>Ovulate</td>
<td>Delay re-mating for longer</td>
</tr>
<tr>
<td></td>
<td>Refrain from mating with additional males</td>
</tr>
<tr>
<td></td>
<td>Invest more in caring for offspring</td>
</tr>
</tbody>
</table>

Predictions of the consequences of female monandry and polyandry [3,6] (for data on the bumblebee Bombus that do not fit current theories, see [4]).

Interpretations based on incomplete sets of alternate hypotheses

A second problem shared by some of the current crop of studies [21,22,28] is that the interpretations of results failed to discuss the possibility that male genital structures are under sexual selection to stimulate the female. Given that the first thorough discussion of sexual selection and genital evolution emphasized this hypothesis [3], these omissions are surprising. It can sometimes be technically difficult to test the stimulation hypothesis, but this is not sufficient reason to ignore it. In fact, as explained in Box 2, experimental testing of the role of female stimulation is potentially easier for genitalia and other contact courtship devices than for other potential male signals to females.

Box 2. Experimental ‘blinding’ of the female: the relative ease of experimental tests of female preference regarding male genitalia

Experimental tests of female preferences for male morphological traits are often technically difficult and give only incomplete answers; these difficulties might seem especially great with genitalia. However, it is in fact often probably easier to perform strong experimental tests of possible female preferences regarding male genitalia and other contact courtship structures than it is to test preferences regarding many other male traits. Most experimental tests of female preferences for morphological traits have involved altering the male structure, and new techniques offer great promise in studies of genital function [22]. However, even the best known of male manipulation studies (e.g. [50,51]) have an infrequently noted logical weakness: they do not control for the possible correlated effect of experimentally altering the morphology of the male on his subsequent behavior. Does a male widowbird alter his behavior as a result of having had his tail length altered, and does this behavior cause the female to change her response to him? Controlling for this problem is easier for male structures, such as genitalia and contact courtship structures [3,29,44], that are perceived by the female via her tactile sense, because the sensitivity of the female can be altered while the male is left untouched. Thus, one can easily change the perception of the male by the female by altering the properties of a subset of her entire array of tactile sense organs, modifying only those that will be contacted by the male structure. In effect, one can selectively ‘blind’ a female to the genital charms of the male. Possible techniques include local extirpation of sensitivity by shaving off sensory hairs, touching the sense organs with a very hot or a very cold object [26], exposing sense organs to a microlaser [22], or locally masking the sense organs by applying a thin patch of adhesive [26,27,52]. A further promising, but as yet unexplored technique would be to apply local anesthetics to the female (this technique could be especially useful to test internal stimulation of the female). Even if the behavior of the male proves to be changed by such manipulations (e.g. by the failure of the female to respond to his genital morphology or behavior), the original link in the chain of cause-and-effect is likely to have been changes in the perception of male stimuli by the female.

A possible limitation to the power of such an experiment would be if the male can sense the alteration itself (e.g. in the tsetse fly study mentioned earlier [26], a male might sense that the small zone of the abdomen of the female was subjected to high temperature the day before), and alters his behavior as a result. This limitation would make techniques for blinding the female that are less easily sensed by the male preferable over others that are more easily sensed by the male.

The most extensive application of this female blinding strategy to date showed that three different cryptic female choice processes were altered by experimentally blinding females to male genital stimuli [26,27]. In addition, a previously unappreciated pre-copulatory level involving a less-discriminating female response to stimulation from the male genitalia was discovered [26].
This is because it is possible to ‘blind’ the female selectively to the charms of the male (by anesthetizing or otherwise inactivating the sense organs in the area of her body that is contacted by the male structures) while leaving her otherwise intact and unaltered. This technique deserves more use in the future.

**The importance of pattern in studies of process**

Another surprising omission in most of the new studies is any attempt to place the experimental results in a phylogenetic context by taking advantage of comparisons with closely related species. Such comparisons are often feasible, because male genitalia are described in the taxonomic literature for most of the groups involved in these studies. Comparison with closely related species is not simply a decoration for a paper on genital function; it is crucial (especially when a phylogeny is available), because it can reveal which aspects of the male genitalia have diverged most, and are thus most likely to be under selection (whether sperm competition, female choice, male female conflict, or lock and key). Some studies seem to assume that ‘bigger is better’ in male genital structures under sexual selection [9,16,17,28], but this is only reasonable if phylogeny shows larger structures have been derived from smaller ones [16].

Ideally, one would start with a robust phylogeny, and study the genitalia of sister species as well as those of more distantly related species, thus distinguishing genital features that are unique to particular species from other features that are shared across species. A focus on recently derived traits can help solve an often difficult problem that must be faced when studying genitalia: which among the multitude of morphological details in a complex genital structure should one measure or modify experimentally? One would like to focus on the functionally most important aspects; but, at least some of the aspects that might be important depend on what is important to the female, and reading female minds is not an easy task (although see Box 2). Insight can be provided, however, by picking out the male traits that have evolved most rapidly in a particular group. By focusing studies in this way, the probability of successfully revealing possible associations between male reproductive success and genital traits might be increased over that which would be obtained if one simply blindly analyzed the entire genital form. This type of focus also seems preferable to combining data on different variables into composite variables using principal components techniques, which are powerful tools for answering some kinds of question, but whose results are often difficult to interpret in biological terms.

In short, studies of process can be strengthened by including data on patterns. It is ironic that the very literature that first revealed the phenomenon of rapid divergent genital evolution has been almost completely neglected in the recent experimental studies attempting to explain it.

**Other considerations**

Observations of the behavioral capacities of genitalia are also lacking in most studies, despite the fact that it is crucial to understand how sexually selected structures move if one hopes to understand how they function [31]. Although male genitalia are often more or less hidden inside the female during copulation, there are several partial solutions to this problem. The most effective technique is direct observation of male genital structures that contact only the outer surface of the female and that are thus not hidden inside her, or those that are periodically withdrawn from her [19,23,32–37]. Other techniques for studying male genital structures that remain inside the female include flash freezing at different stages of copulation [14,16,28,34–36]; observation through the semi-transparent body wall of the female [38]; observation of artificially induced behavior of intromittent genitalia via electric shocks [39]; partial anesthesia of the male [40]; stimulation of the genitalia after removing the head of the male [41]; and X-ray video recordings [41] (interpretation of the significance of genital form and behavior resulting from artificial inflation or erection outside the female needs care, however, because the genitalia of some groups assume quite different forms when they are constrained during copulation by contact with the female [42]). Although none of these techniques can give a complete understanding of genital behavior during copulation, they can sometimes help discriminate among possibilities. For example, observation of X-ray videos helped reject the otherwise seemingly reasonable possibility that a large hooked structure on the phallobase of the tsetse fly Glossina palpalis functions as a holdfast device, because the entire phallobase moves smoothly back and forth rhythmically within the female for extended periods during copulation, as in other species lacking the hook [41]. Inclusion of genital behavior patterns in phylogenies is an attractive future goal. The behavior associated with genitalia and other contact courtship devices can be even more diverse than the morphology itself [43,44].

Recent studies include both correlational and experimental approaches. Both are important, but each has important limitations (some of which can be partially compensated by the other). Demonstrations of correlations in unmodified animals have the advantage of using natural ranges of variation, and thus providing data concerning possible present-day selection on genital traits, but they are not equivalent to demonstration of cause-and-effect relations (although this seems to have been forgotten in some discussions). By contrast, experimental modifications can demonstrate cause-and-effect relations. However, at least those studies that have been performed to date (e.g. [22,26,27]) are limited in that they involved modifications that were beyond the range of naturally occurring variation. Therefore, they could not test the possibility of female discrimination regarding the kinds of more subtle morphological detail that often mark the differences between closely related species.

Finally, it has been common to ignore the possibility of ‘passive’ female effects; that is, the possibility that females have evolved to select among males or defend themselves from them by changing the rules of the game to which males must conform when attempting to win out in sperm competition [8,29]. Some explanations of differences in the numbers of sperm being transferred to the female as a
result of sperm competition [25,45] have ignored the possibility that female changes (e.g. in the volume of the transport sperm to storage and fertilization sites[47]) might have driven divergence in males. As noted by Wedell and Hosken [48] ‘females are the arenas in which post-copulatory sexual selection occurs, so at some fundamental level they are always involved passively or otherwise’. Female effects are especially likely in species in which the morphological fit between male and female during copulation rules out direct male control over sperm transport and storage [47]. These considerations emphasize the importance of including female morphology and sensitivity in studies of genital function [47,49]. Comparisons of these aspects of the female in closely related species have seldom been included, but such ‘pattern’ data could be helpful because they can point toward male traits that are particularly likely to be under selection.

There has been much recent progress in understanding the functions of animal genitalia, and new experimental studies will undoubtedly provide further advances. It would be a shame if, in the rush to gather new data, hard-earned lessons from previous studies are forgotten.

Acknowledgments
I thank D.J. Hosken and M.J. West-Eberhard for comments on previous versions of the article, and the Smithsonian Tropical Research Institute and the Universidad de Costa Rica for financial support.

References