

Chapter 1

Introduction

I. LIFE CYCLES AND DIVERSITY OF VASCULAR PLANTS

The subjects of this thesis are the pteridophytes and seed plants that are conventionally classified as the vascular plants or tracheophytes. Vascular plants were traditionally defined by the possession of specialized conducting tissues, called phloem and xylem. Mosses are now believed to have inherited their conducting tissues from a common ancestor with the tracheophytes (Mishler & Churchill 1984) but are not considered in this thesis.

Vascular plants can be divided into four groups with respect to life cycle. These groups are homosporous pteridophytes, heterosporous pteridophytes, gymnosperms and angiosperms. This is not intended to be a phylogenetic classification.

There are about a quarter of a million species of vascular plant alive today. The vast majority are angiosperms and most of the remainder are homosporous pteridophytes. Heterosporous pteridophytes and gymnosperms contribute only a small number of species (Table 1.1).

TABLE 1.1

Estimated number of extant species in each of the major groups of vascular plants (data from Parker 1982).	
Homosporous pteridophytes	12,000 species
Heterosporous pteridophytes	<1,000 species
Gymnosperms	<1,000 species
Angiosperms	>200,000 species

A. Homosporous pteridophytes

Homosporous pteridophytes produce a single type of spore. Spores are dispersed and develop into photosynthetic or mycoparasitic gametophytes. A gametophyte's gender is indeterminate at the time of spore dispersal, and a single gametophyte may produce eggs and/or sperm. Sperm are motile, and require free water to fertilize eggs. The young sporophyte is nourished by the maternal gametophyte during early development but later becomes

nutritionally independent.

The sporophytes of most homosporous pteridophytes are long-lived perennials but a few species are annuals (Cousens 1988). Annual species include the aquatic ferns of the genus *Ceratopteris* (Lloyd 1974a, 1974b; Hickok, Warne & Slocum 1987). *Anogramma leptophylla* produces annual sporophytes that are borne on a perennial gametophyte (Mehra & Sahnun 1976). Vegetative propagation by rhizomes is common (Tiffney & Niklas 1985). Sporophytes are generally small in stature, relative to gymnosperms. With the exception of some members of the Ophioglossaceae (Kato 1988), modern pteridophytes lack a lateral meristem which would allow increases in stem thickness by secondary growth.

The gametophyte stage is poorly known. Subterranean mycoparasitic gametophytes occur in the Psilotaceae, Lycopodiaceae, Ophioglossaceae and Schizaeaceae (Wagner, Wagner & Beitel 1985). In the Vittariaceae, gametophytes reproduce vegetatively by gemmae (Sheffield & Farrar 1988). Some populations of *Vittaria* never produce sporophytes (Farrar 1978).

B. Heterosporous pteridophytes

Heterosporous pteridophytes produce two types of spores. The smaller microspores develop into male gametophytes and the larger megaspores develop into female gametophytes. Gametophytes develop within the spore wall (endosporic development) and are almost totally dependent on the food reserves of the spore. Fertilization usually occurs after spores are dispersed from the parental sporophyte. The young sporophyte is nourished from the food reserves of the female gametophyte.

Among modern pteridophytes, heterospory occurs in the Selaginellales, Isoetales, Marsileales and Salviniiales. Most members of the last three orders are aquatic or semi-aquatic. In the Filicales, the monotypic genus *Platyzoma* is also heterosporous but differs from other heterosporous groups in having gametophytes that are photosynthetic and develop outside the spore wall (exosporic development).

C. Seed plants

Gymnosperms and angiosperms together comprise the seed plants or spermatophytes. Seed plants are heterosporous but, unlike heterosporous pteridophytes, megaspores develop into female gametophytes while attached to the maternal sporophyte and surrounded by sporophyte tissues. The female gametophyte and surrounding structures are referred to as an ovule. Female gametophytes encounter male gametophytes while the ovule is still attached to the maternal sporophyte. After fertilization, the ovule is called a seed. Seeds are dispersed from the sporophyte. There are thus two dispersal stages in the life cycle: the dispersal of male gametophytes, and the dispersal of seeds. The microspores and male gametophytes of seed plants are called pollen grains.

The nutritional relationships between generations vary among seed plants. In *Ginkgo* and cycads, the maternal sporophyte fully provisions the female gametophyte before fertilization. The embryo then digests the female gametophyte. In other gymnosperms, the maternal sporophyte supplies most resources to the female gametophyte after fertilization, at the same time as the embryo is digesting the gametophyte. In angiosperms, the female gametophyte is greatly reduced. Its nutritional role is taken by the endosperm, a distinctive angiosperm tissue formed by a second fertilization of two nuclei from the female gametophyte. The maternal sporophyte supplies resources to the endosperm or directly to the embryo.

Most modern gymnosperms are woody trees or shrubs. Several members of *Gnetum* are lianes. The cycad *Zamia pseudoparasitica* is an epiphyte (Norstog 1987) and the podocarp *Podocarpus ustus* is a parasite on other conifers (de Laubenfels 1959). Angiosperms have a greater range of growth forms than all other vascular plants. Angiosperms may be trees, shrubs or herbs; perennial or annual; terrestrial, aquatic or marine; lianes, epiphytes or parasites; to name just some of the variety. With the exception of conifer forests and the moss-lichen tundra, angiosperms dominate all major terrestrial vegetation zones (Friis, Chaloner & Crane 1987).

D. The paleobotanical perspective

Vascular plants are first recorded in rocks of mid-Silurian age, approximately 420 Myr ago (Edwards & Feehan 1980; but see Taylor 1988; absolute dates throughout thesis are based on Harland et al. 1982; 1 Myr is a megayear). These plants were homosporous. The earliest record of heterospory comes from the middle Devonian, approximately 380 Myr ago (Andrews, Gensel & Forbes 1974), closely followed by the earliest known seeds (Famennian c. 365 Myr ago: see Gillespie, Rothwell & Scheckler 1981). The first undisputed angiosperms appear in the early Cretaceous (Barremian c. 120 Myr ago: Doyle 1978). There is general consensus that heterosporous pteridophytes have evolved from homosporous pteridophytes; that gymnosperms had heterosporous ancestors; and that angiosperms are derived from gymnosperms. Heterosporous pteridophytes have had several independent origins. There is still no consensus as to whether the seed habit and angiospermy evolved only once or on a number of occasions. Gymnosperms are believed to have evolved from one or more groups of progymnosperms (Devonian pteridophytes with secondary growth). The immediate ancestors of angiosperms are still disputed.

Homosporous, heterosporous, gymnospermous and angiospermous life cycles make their first appearances in that order in the fossil record. This does not imply, however, that modern pteridophytes are more primitive than seed plants, or that modern gymnosperms are adaptively inferior to angiosperms. All extant lineages have undergone an equal period of evolution since their divergence from a common ancestor. The major radiation of the modern "polypodiaceous" ferns occurs near the Cretaceous/Paleocene boundary, after the rise to dominance of angiosperms. Lovis (1977) has referred to this radiation of a homosporous lineage as "the most recent major innovation in the evolution of the world's flora".

1. *Changes in species diversity*

I will discuss three studies that have attempted to describe changes in species diversity from compilations of fossil floras (Niklas, Tiffney & Knoll 1983; Knoll 1986; Lidgard & Crane 1988). Such studies are subject to many uncertainties and biases (Koch

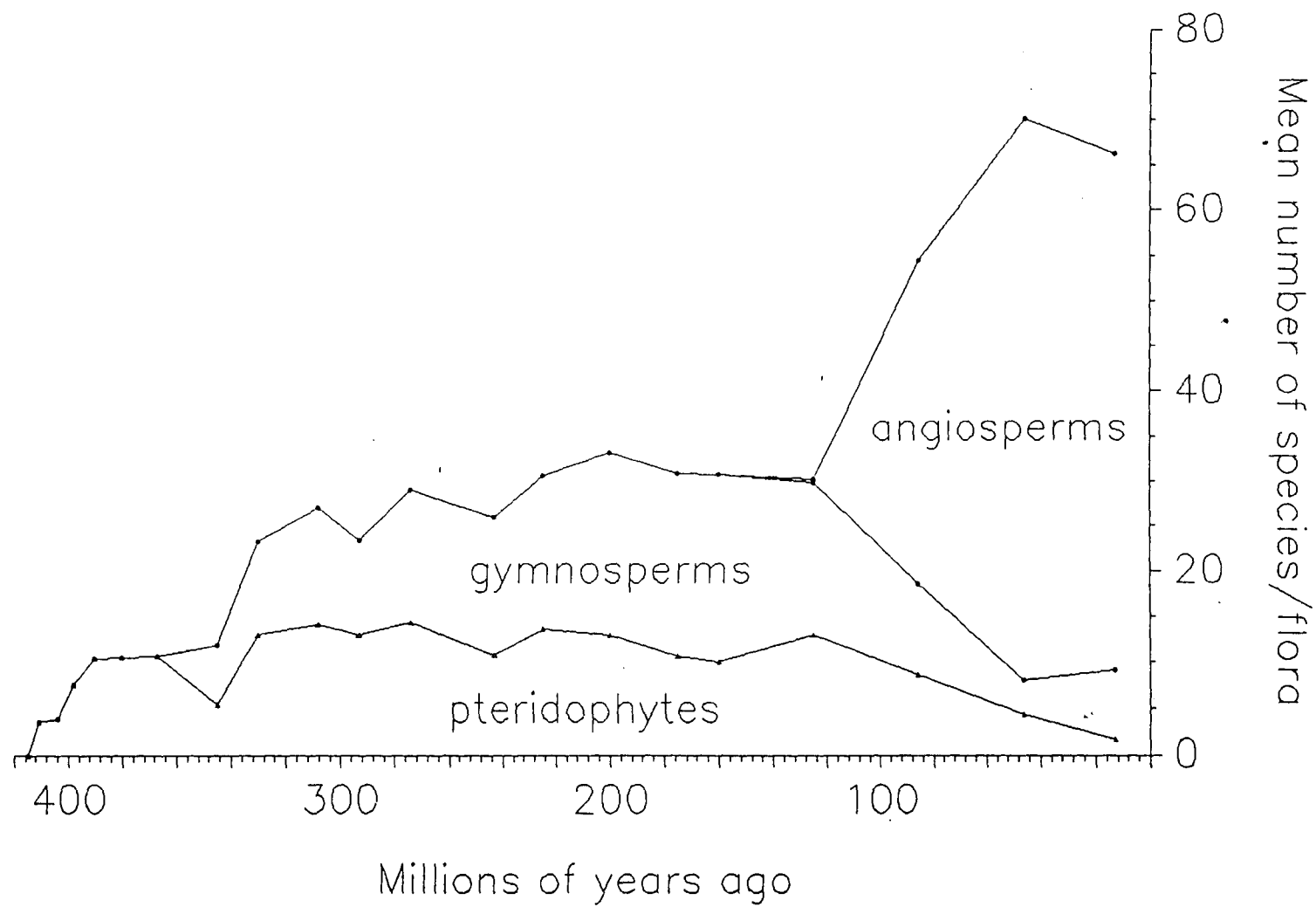
1978; Niklas et al. 1980; Ferguson 1985) but are the most accessible sources of information.

Niklas et al. (1983) compiled data from approximately 18,000 fossil plant species citations. From their Figure 1, the history of vascular plants can be divided into four phases: (1) a Silurian-Devonian flora of low diversity dominated by morphologically simple plants; (2) a Carboniferous-Permian flora dominated by pteridophytes but with a significant minority of gymnosperms (c. 360-250 Myr ago); (3) a gymnosperm-dominated flora during the Triassic, Jurassic and early Cretaceous (c. 250-130 Myr ago); and (4) the appearance and rise of flowering plants during the Cretaceous and Tertiary. Niklas et al. detected a four-fold increase in observed species diversity between (1) and (2), and a further three-fold increase in observed diversity between (3) and (4). The transition from (2) to (3) coincided with climatic changes at the end of the Permian and was not associated with major changes in observed diversity.

Knoll (1986) examined the published records of 391 fossil floras ranging in age from the latest Silurian (c. 410 Myr ago) to Pliocene (ending 1.6 Myr ago). The floras consisted of compression macrofossils that had accumulated in lowland deltaic or floodplain environments. The choice of flood-plain floras excluded distinctive communities such as the coal swamp floras of the Carboniferous and Permian. For each flora, Knoll calculated the minimum number of species that could account for the macrofossil remains. This number was used as an index of diversity within floras and used to detect changes in diversity over evolutionary time. I have replotted Knoll's data to show changes in diversity of pteridophytes, gymnosperms and angiosperms (Figure 1.1).

Knoll recognized three relatively brief periods of rapid increase in diversity, separated by much longer intervals of more or less constant diversity. The first period of rapid increase occurred during early Devonian times (c. 400 Myr ago) and corresponded to a radiation of homosporous plants. The second period of rapid increase occurred during the Mississippian (c. 350 Myr ago) and was associated with the initial radiation of gymnosperms. From Mississippian times until the mid Cretaceous

Figure 1.1. Mean number of species per paleoflora (after Knoll 1986).



(c. 120 Myr ago) within-floral diversity stayed relatively constant despite major changes in the taxonomic composition of the flora. During this long interval, floodplain communities were dominated by a gymnosperm overstorey and a pteridophyte understorey. The third period of rapid increase corresponded to the radiation of angiosperms into both overstorey and understorey communities (see Knoll 1986; Knoll & Niklas 1987).

Knoll's study estimated the mean number of species per flora rather than the global species diversity. This approach should be less sensitive than Niklas et al.'s study to biases caused by different geological periods having different areas of exposed sediments. Knoll defined a paleoflora as "the total assemblage of fossil plants recovered from a reasonably homogeneous package of rocks". The floras used in his compilation consisted of plants "from a mixture of similar and contiguous habitats occurring over an area a few to several tens of kilometers in linear dimension" and represent "accumulation over a period of several thousand, and perhaps as much as a few million years". These floras do not directly correspond to a modern flora.

Both Niklas et al. (1983) and Knoll (1986) recognized a major increase in diversity during the Mississippian (= early Carboniferous) followed by little change in diversity until the Cretaceous rise of angiosperms. Niklas et al. divided this interval into an early pteridophyte-dominated period and a later gymnosperm-dominated period. However, Knoll observed little change in the relative proportions of gymnosperms and pteridophytes over the same period. The difference is probably explained by Knoll's data excluding the pteridophyte-dominated floras of the coal swamps. Interestingly, several pteridophytes from the coal swamps evolved gymnosperm-like characters (see Chapter 5).

Lidgard and Crane (1988) compiled data on fossil leaves from 197 floras of Late Jurassic to Paleocene age (160-60 Myr ago). Absolute species diversity showed only a moderate increase in this period, despite a rapid rise in the number of angiosperm species. Of particular interest is their analysis of changes in the proportion of different groups in individual floras. From 115 to 90 million years ago, the percentage of angiosperm species

rose from near zero to an average of about 70%. This increase in relative importance was primarily at the expense of cycadophytes and pteridophytes. Conifers showed little or no decline in relative importance over the same period. All Lidgard and Crane's floras appear to come from the Northern Hemisphere. The rapid increase in percentage of angiosperms could be explained either by a local adaptive radiation or by invasion from elsewhere.

Crane (1987) gives an excellent discussion of regional differences in floras during the rise of angiosperms, but there is again a paucity of information from the Southern Hemisphere. From his reconstructions, angiosperms first became established as weedy herbs or shrubs in relatively open habitats. Large angiosperm trees may not have become widespread until the latest Cretaceous or early Tertiary. The major radiation of modern angiosperm herbs did not take place until later in the Tertiary. These groups of herbs account for a large part of current angiosperm diversity.

2. *The abominable mystery*

Darwin described the sudden appearance of angiosperms in the fossil record as an "abominable mystery". This quote is almost obligatory in any discussion of the origin of angiosperms. The riddle of angiosperm origins has three parts. First, which groups of modern and fossil plants are the closest relatives of angiosperms? Second, where and when did the angiosperms originate? Third, what characters were responsible for the adaptive success of angiosperms? The third question is discussed in Chapter 8. The first two questions are briefly considered below.

Recent cladistic analyses (Crane 1985; Doyle & Donoghue 1986) place the angiosperms in an "anthophyte" clade with the modern Gnetales (*Gnetum*, *Welwitschia*, *Ephedra*) and two Mesozoic taxa, *Pentoxylon* and the Bennettitales (cycadeoids). An association between the Gnetales (particularly *Gnetum*) and angiosperms was frequently suggested before cladistics, but equally frequently discounted. The morphological similarities between *Gnetum* and some angiosperms are so striking that "a botanist, not familiar with *Gnetum*, would guess the plant to be a

dicotyl" (Chamberlain 1935, p. 409). However, these similarities were often dismissed as examples of convergent evolution. It is an interesting psychological question why so many botanists seem to have had a strong emotional commitment to the angiosperms remaining a mysterious group without obvious affinities (e.g. Swamy 1974).

One piece of evidence that suggested an association between the Gnetales and angiosperms was the fact that the Gnetales, unlike other gymnosperms, produce wood with vessels. Thompson (1918) studied vessel ontogeny in *Gnetum* and concluded that vessels were formed by a different means than occurs in angiosperms and that the shared possession of vessels was only a superficial resemblance. Thompson's conclusions were almost universally accepted and cited, despite conflicting reports (published in the same journal) that vessel formation was essentially similar in *Gnetum* and some angiosperms (Bliss 1921; MacDuffie 1921). Muhammad & Sattler (1982) reinvestigated the vessels of *Gnetum* and concluded that there were no fundamental differences in vessel formation between the two groups.

Gnetum and angiosperms share common features of wood chemistry (Melvin & Stewart 1969; Michell, Ingle & Stewart 1969; Shio & Higuchi 1978) and have similar development of the photosynthetic apparatus (Jeske & Senger 1978, 1981). These chemical characters are particularly convincing evidence for a relationship because they provide independent support for a hypothesis originally based on morphological resemblance. Crane (1985) and Doyle & Donoghue (1986) did not use these characters in their cladistic analyses. Their results suggested that the Gnetales form a monophyletic sister group to the angiosperms (ignoring fossil taxa for the moment). Thus, *Gnetum*, *Welwitschia* and *Ephedra* are claimed to be each other's closest living relatives. Jeske & Senger's (1981) data on pigment synthesis suggest that *Gnetum* and *Welwitschia* may be more closely related to angiosperms than they are to *Ephedra*, in which case the Gnetales would be a paraphyletic group.

I will now address the timing of angiosperm origins. Hickey & Doyle (1977; also Doyle 1978) argued that the fossil record suggests that the "primary adaptive radiation of the flowering

plants took place during the Early Cretaceous" and that there is no need to invoke a long pre-Cretaceous history for the flowering plants. This conclusion was based on a detailed analysis of pollen and macrofossils from the Early Cretaceous Potomac series of North America. The earliest strata with angiosperm fossils show a low diversity of "primitive" morphologies. More recent strata show a gradual increase in diversity and morphological "advancement". This pattern was interpreted as "consistent with the orderly diversification of a monophyletic group . . . , but not with either markedly polyphyletic origin or random immigration of already advanced and differentiated types from some other area".

On the other hand, several lines of evidence suggest an earlier origin of the angiosperm lineage. First, quite dissimilar lineages are already in existence by the first appearance of undisputed angiosperm fossils in the Barremian and these fossils have characters that are more "advanced" than characters present in some modern angiosperms (Walker & Walker 1986). Even Hickey & Doyle (1977) believed that "the lines leading to monocots and dicots had already diverged by the earliest appearance of angiosperms in the fossil record". Second, *Sarmiguelia* from the Late Triassic (c. 230 Myr ago) is claimed to be an angiosperm (Cornet 1989). These fossils are about 100 million years older than the Early Cretaceous angiosperms. Third, "molecular clocks" place the divergence of monocots and dicots much earlier than the Cretaceous. Martin, Gierl & Saedler (1989) placed the monocot-dicot divergence in the Carboniferous (c. 300 Myr ago). Wolfe et al. (1989) placed the divergence in the Jurassic-Triassic, about 200 Myr ago. Moreover, several lines of evidence suggest that the monocot-dicot split is not the most ancient divergence in the angiosperm lineage (Kubitzki & Gottlieb 1984a). Fourth, biogeographical evidence is difficult to accommodate within an Early Cretaceous radiation, particularly the concentration of "primitive" angiosperm families in the Malay archipelago and the existence of pairs of families with one member represented on either side of Wallace's Line (Whitmore 1988).

The question of the time of angiosperm origin is intimately linked to the question of the place of origin. There are two major contenders for this honor. One view is that angiosperms

originated in western Gondwana, in what are now parts of Africa and South America (Raven & Axelrod 1974; Doyle 1984). This scenario is compatible with an Early Cretaceous origin (Doyle 1984). More recently, Audley-Charles (1987) and Takhtajan (1987) have strongly argued for an origin in eastern Gondwana, possibly on a shard split from the north of Australia during the Late Jurassic. This shard now forms parts of Burma, Thailand and Malaya. Fossil evidence is lacking from this region during the critical Jurassic-Early Cretaceous period.

The major argument against an origin in eastern Gondwana is that *Clavatipollenites* (the earliest recognizable angiospermous pollen in Australia) does not appear in Australia until the Early Albian, up to 10 Myr after its first appearance in Europe and Africa (Truswell, Kershaw & Sluiter 1987). However, the earliest recognized angiosperm pollen types, including *Clavatipollenites*, are more structurally advanced than the pollen of some extant families of primitive dicots. These more primitive pollen types would be difficult to distinguish from the pollen of some gymnosperms (Walker & Walker 1986). Therefore, *Clavatipollenites* may have originated in western Gondwana, but this does not preclude earlier angiosperms being present in eastern Gondwana.

3. Overview

In summary, the first land plants were homosporous and, ever since, plants with homosporous life cycles have made a significant contribution to terrestrial floras, particularly as small herbaceous perennials. Heterosporous pteridophytes first appeared during the Devonian. Heterosporous, arborescent lycopods and sphenopsids were important members of Carboniferous coal swamp communities. Gymnospermous seed plants also made their first appearance during the Devonian. Gymnosperms quickly became the dominant arborescent forms, except in the coal swamps, and maintained this dominance until the rise of angiosperms. The radiation of angiosperms during the Cretaceous and Tertiary resulted in a great increase in total species diversity but a decrease in gymnosperm and pteridophyte diversity.

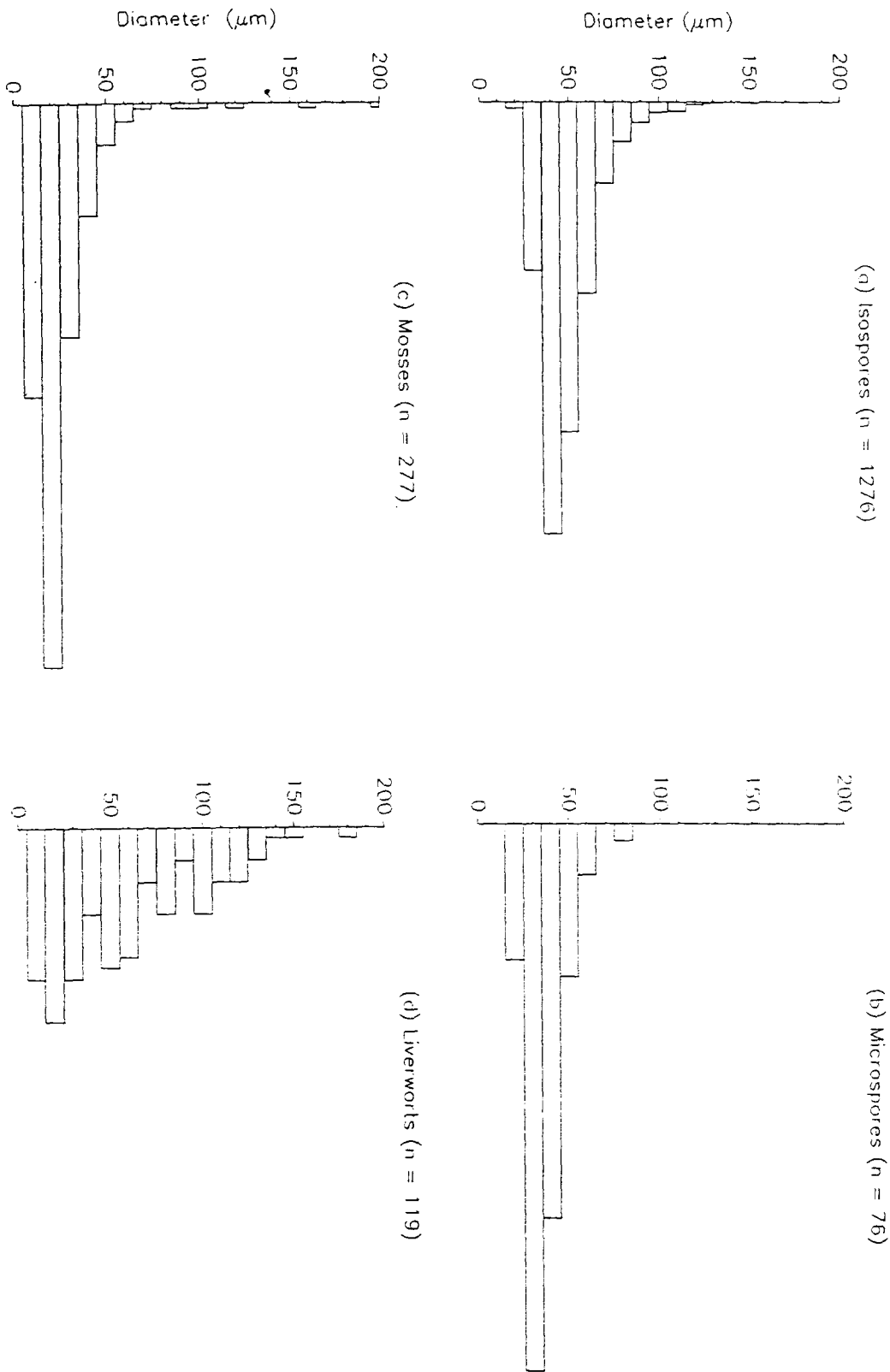
about 0.06 μg and a single spore of *C. richardii* weighs about 0.8 μg . This last figure is important because it is probably close to the upper limit of isospore weights.

Figure 1.2 also gives the distribution of diameters for 277 moss species and 119 liverwort species compiled from Erdtman (1965). The ranges of spore diameters for mosses and liverworts are similar -- from less than 10 μm to about 200 μm -- but the shapes of the distributions are very different. Most mosses have spores less than 25 μm . Liverworts have a comparatively even distribution of spore sizes in the range 10-60 μm . There is little overlap between the distributions of moss and pteridophyte spore sizes. Most moss spores are smaller than most pteridophyte spores. I will not comment further on the size distributions of bryophyte spores, but I believe these patterns probably reflect major adaptive differences between mosses, liverworts and pteridophytes.

Figure 1.2. Spore size distributions for (a) isospores of homosporous pteridophytes; (b) microspores of heterosporous pteridophytes; (c) moss spores; and (d) liverwort spores. The horizontal scale has been adjusted so that the area under the histogram is the same for all four groups.

Spore sizes were taken from Erdtman (1965) for mosses and liverworts, and from Erdtman & Sorsa (1971) for pteridophytes. If more than one set of measurements was given for a species, I used only the first set of measurements. I used the largest quoted diameter as my measure of spore size. If only one diameter was quoted, this was assumed to be the largest diameter. My tabulation of diameters was uncritical and I did not take account of whether the quoted diameter included a perine or spore processes. The thickness of the perine was usually negligible when this information was provided. Some diameters may be of immature spores. Despite the limitations of the data set, I believe the figure reliably identifies real differences in the size distributions of different spore types.

Figure 1.2. Spore size distributions



B. Spore size in modern heterosporous pteridophytes

Erdtman & Sorsa (1971) give approximate diameters for megaspores of 18 heterosporous species. The lowest value (c. 67 μm : *Selaginella firmula*) is much smaller than other megaspores and appears to be in error because Erdtman & Sorsa list all *Selaginella* megaspores as greater than 100 μm (their Table 1). If the remaining 17 species are considered, the smallest megaspores are possessed by *Regnellidium diphyllum* (c. 180 μm) and *S. d'albertisii* (c. 190 μm) and the largest megaspores by *Isoetes japonica* (560-660 μm) and *I. stellenbosensis* (c. 630 μm). A sample of 17 species probably does not give a good idea of the range of megaspore sizes. Megaspores of *S. exaltata* are "reputed" to reach 1500 μm diameter (Erdtman & Sorsa 1971). Tryon (1986) gives a range of c. 200-1000 μm for *Selaginella* megaspores and c. 100-800 μm for *Isoetes* megaspores. The lower limits of megaspore size can be compared to the upper limit of isospore size (150 μm ; see above). There appears to be little overlap between the upper limit of the isospore distribution and the lower limit of the megaspore distribution.

Microspore diameters for 76 species range from about 20 μm to 80 μm (Figure 1.2b; data from Erdtman & Sorsa 1971). Thus, microspores are on average slightly smaller than isospores. There was no significant correlation between megaspore and microspore size for the 17 species in which both measures were available ($r = 0.12$).

Platyzoma microphyllum is a species of great theoretical interest because it produces two sizes of spore but retains many features of homosporous life cycles such as exosporic development and potentially hermaphrodite gametophytes (from the large spores only). For these reasons *Platyzoma* is often considered to represent a less evolved form of heterospory. Spore sizes support this conjecture. The mean size of large spores is 175 μm and the mean size of small spores is 91 μm (Tryon 1964). Thus, the large spores are at the lower limit of the megaspore range and the small spores are at the upper limit of the microspore range.

C. Spore size in the fossil record

Paleobotanists frequently work with dispersed spores and have no sure way of telling whether the parental sporophyte was homosporous or heterosporous. By convention, spores larger than 200 μm are usually classified as megaspores. Spores smaller than 200 μm are sometimes called miospores, a collective term that encompasses isospores and microspores.

Dispersed spores from the Silurian (ends 408 Myr ago) belong to only three form genera and all are less than 100 μm in diameter. There is a gradual increase in spore diversity to about 50 recognized genera by the end of the Devonian (360 Myr ago). The increase in diversity is accompanied by an increase in maximum spore size. Maximum spore size first approaches and then exceeds 200 μm . During the Devonian, spores greater than 200 μm were the upper tail of a continuous distribution and several spore species have been described with average diameters in the range 200–300 μm . By Upper Carboniferous times (c. 300 Myr ago), however, there is a distinct gap in the size distribution with miospores smaller than 200 μm and megaspores larger than 300 μm (Chaloner 1967). Megaspores reached their greatest size during the Carboniferous. Megaspores of the arborescent lycopod *Lepidocarpon* have been found that are 11 mm in length and about 5 mm in width (Phillips 1979).

Devonian spores larger than 200 μm are now known to include large isospores of homosporous species as well as the megaspores (and even microspores) of early heterosporous species. Turnau & Karczewska (1987) have shown that the distribution of spore sizes for three Middle Devonian spore species is bimodal. The spores of the smaller size class (presumed microspores) have average diameters of 97 μm , 100 μm and 207 μm . Another spore species has a unimodal distribution with a range from 167 μm to 307 μm and is presumably homosporous.

Among modern seed plants, larger seeds are associated with establishment in shaded rather than open habitats (Salisbury 1942; Harper et al. 1970; Foster 1986) and, to a lesser extent, with arid conditions (Baker 1972). The increase in maximum isospore size during the Devonian may have been associated with the evolution of arborescent growth forms (Niklas et al. 1980).

Larger spores may also have allowed the colonisation of drier, previously unoccupied habitats.

D. Seed size in the modern flora

Seed sizes tend to be measured by weight, unlike spore sizes which tend to be measured by diameter. Harper, Lovell & Moore (1970) reviewed the range of seed sizes in the modern world flora. The largest seed is possessed by the double coconut palm *Lodoicea maldivica* (18-27 kg). The smallest seeds are possessed by mycotrophic and parasitic species that rely on an external food source for seedling establishment. Some examples are *Goodyera repens* (2 μ g), *Monotropa hypopitys* and *Orobancha picridis* (3 μ g), and *Pyrola uniflora* (4 μ g; seed weights from Salisbury 1942). Orchid seedlings are mycotrophic and orchid seeds range in weight from 0.3 μ g to 14 μ g (Arditti 1979). These weights can be compared with 0.8 μ g for the large isospores of *Ceratopteris richardii* (see above). Apart from a few heterotrophic angiosperms, there appears to be no overlap between the distributions of isospore and seed weights.

Angiosperm seed weights, therefore, span about twelve orders of magnitude. I have found fewer data on gymnosperm seed weights (Fig. 1.3). Harper et al. (1970) give the weight of *Cycas circinalis* and *Araucaria bidwilli* seeds as about 50 g and 10 g respectively. In contrast, an average seed of *Pinus sylvestris* or *Picea excelsa* weighs about 6 mg (Salisbury 1942). The seeds of *Sequoia sempervirens* are about 1.6 mm in length (Dallimore & Jackson 1948) which probably corresponds to a weight of about 2 mg.

E. Ecological correlates of seed size

Studies of the modern flora have identified correlations between seed size and particular growth forms or habitats: (1) on average, herbs have smaller seeds than shrubs which have smaller seeds than trees (Salisbury 1942; Baker 1972; Levin 1974; Rockwood 1985; Hodgson & Mackey 1986; Mazer 1989); (2) species that become established in the shade usually have larger seeds than species that become established in open habitats (Salisbury 1942; Levin 1974; Foster & Janson 1985; Foster 1986; Hodgson &

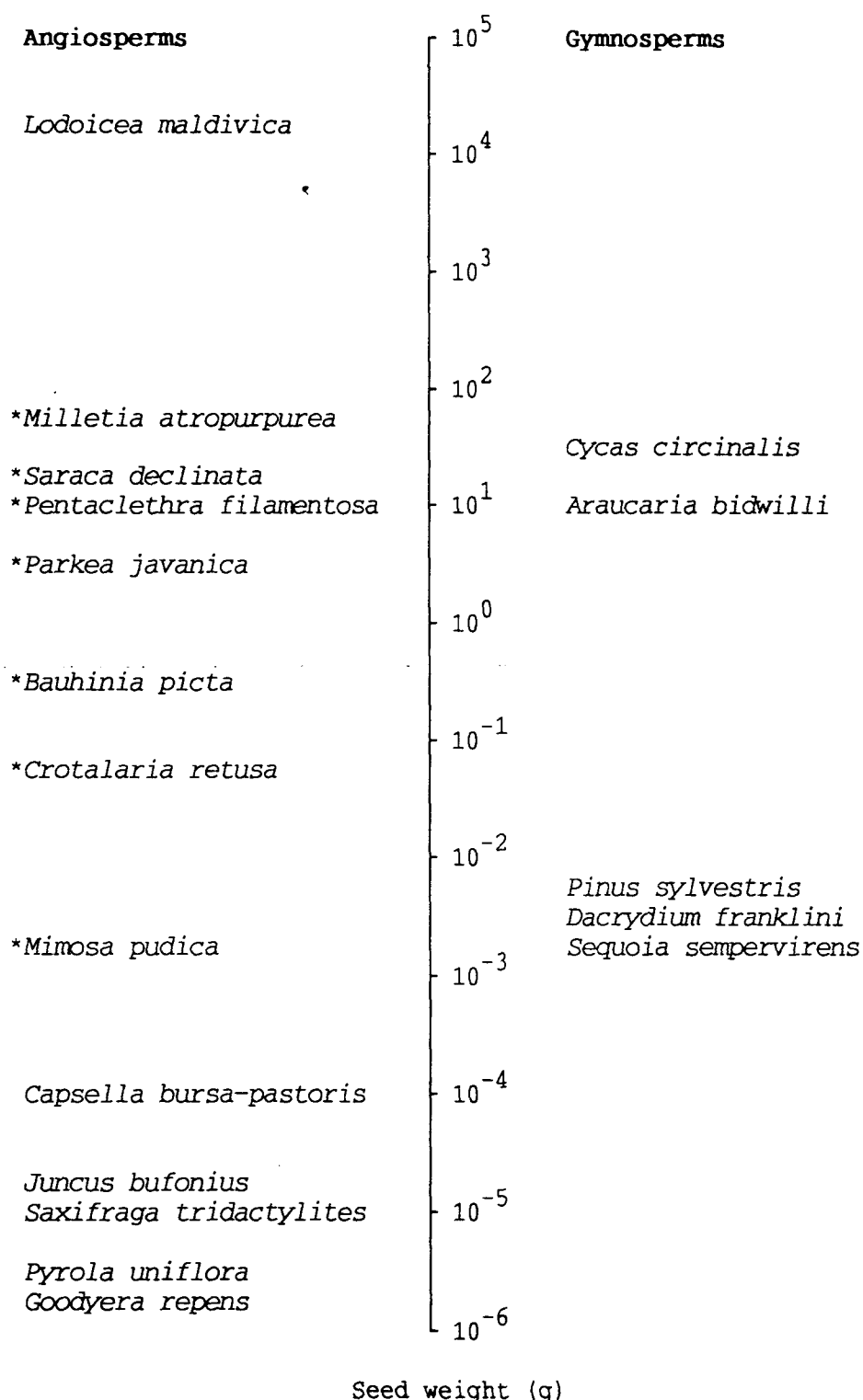


Figure 1.3. Seed weights of selected angiosperms and gymnosperms. Species marked with an asterisk are members of the Leguminosae (sensu lato). Sources: *L. maldivica*, *C. circinalis*, *A. bidwilli* (Harper et al. 1970); *J. bufonius*, *C. bursa-pastoris*, *S. tridactylites* (Grime et al. 1981); *P. uniflora*, *G. repens*, *P. sylvestris* (Salisbury 1942); *D. franklini*, *S. sempervirens* (Dallimore & Jackson 1948); Leguminosae (Corner 1951). Some seed weights were estimated from the linear dimensions of the seed.

Mackey 1986; Mazer 1989); (3) average seed size increases with increasing aridity (Baker 1972); and (4) annual herbs tend to have smaller seeds than perennial herbs (Baker 1972; Hodgson & Mackey 1986). Correlations detected in some studies were not necessarily confirmed in other studies. For example, Mazer (1989) found that trees had larger seeds than other life forms but did not find significant differences in seed size between shrubs and herbs.

Hodgson & Mackey (1986) have cautioned against purely adaptive interpretations of seed size. They argued that variation in size is constrained by anatomical and developmental features such as type of placentation, number of ovules per carpel, presence or absence of endosperm in the ripe seed, and the pattern of embryogenesis. These features are taxonomically linked. In consequence, different families have characteristic ranges of seed size and may be limited to particular types of ecological specialization. Mazer (1989) also found significant differences in seed size among families. These studies have general implications for formal statistics, because seed sizes of related species are not truly independent data points and the appropriate degrees of freedom for significance tests are unclear.

The nature of the data bases on which these correlations are based should be mentioned. Salisbury used data from about 300 species but made no pretence of collecting a "random" sample of the English flora. Baker's data base included almost 2500 representatives of the Californian flora covering a much wider range of habitats than Salisbury. Hodgson & Mackey's data come from an almost exhaustive survey of the much more restricted Sheffield flora. Foster & Janson used seed weights from the woody plants of a moist tropical forest in Peru. Levin's data on seed weight and oil content come from a non-systematic sample of the world flora with an emphasis on the Northern Hemisphere. Mazer's data comprises 50% of species from the Indiana Dunes region, randomly chosen from the published flora. A further point should be emphasized: though some of the correlations appear highly significant the differences in mean seed weight between categories are usually small relative to the range of seed weight

within categories.

Seed weight is only an indirect measure of maternal investment because, among other reasons, seeds differ in the composition of their food reserves. Seeds with a high oil content are probably more costly than an equivalent sized seed with predominantly carbohydrate food reserves. Oil content tends to increase with plant woodiness and shade tolerance, variables that are also associated with increases in seed weight (Levin 1974). The obvious explanation is that these factors favor greater food reserves and that increased seed weight or increased oil content are alternative means by which food reserves can be increased.

F. Seed size in the fossil flora

Chaloner & Sheerin (1981) plotted the sizes of Devonian and Carboniferous propagules. An interesting feature of their data is that the minimum size of gymnosperm seeds is considerably greater than the minimum size of megaspores. The smallest megaspores were about 200 μm in diameter. (This is actually based on a comparison to the modern flora because fossil spores greater than 200 μm are classified as megaspores by default.) By comparison, the smallest seeds were about 2 mm in length.

Tiffney (1984, 1986) has collected published data on seed size from the paleontological literature. Sample sizes are small for all periods prior to the Cretaceous (Table 1.2), but some interesting observations can be made on minimum seed sizes. Before the Cretaceous and the first recognized angiosperm fossils, the smallest recorded seed had a volume of approximately 0.2 mm^3 ($n = 96$). Seeds of Cretaceous gymnosperms ranged between 18 mm^3 and 2180 mm^3 ($n = 12$) whereas seeds of Cretaceous angiosperms ranged from 0.02 mm^3 to 55 mm^3 with a mean seed volume of only 1.6 mm^3 ($n = 203$). Not only were angiosperm seeds smaller on average than the seeds of contemporary gymnosperms, but the smallest angiosperm seeds were about an order of magnitude smaller than the smallest seeds from all earlier periods. Large angiosperm diaspores first appeared in the early Tertiary.

At face value, the fossil evidence suggests that the earliest angiosperms were small-seeded and later radiated into

TABLE 1.2

Minimum and maximum seed sizes (in mm^3) by geological period. Cretaceous seeds divided into gymnosperm and angiosperm seeds. Paleogene and Neogene floras (Tertiary) are dominated by angiosperms (data from Tiffney 1984, 1986).

	Dev	Miss	Pennsyl	Triass	Jurass
n	4	9	7	42	28
max	13	500	318,000	19,000	16,000
min	2.0	0.2	7.0	0.6	1.0

	Cretaceous		Tertiary	
	Gymno	Angio	Paleo	Neo
n	12	203	485	578
max	2,200	55	61,000	25,000
min	18	0.02	0.07	0.04

Tiffney reported diaspore volumes rather than seed volumes. In most cases, the diaspore is probably a seed. He defines a diaspore as "the reproductive unit that is dispersed or sown" and diaspore volume includes fruit tissue in berries and drupes. His data for angiosperms may include some multi-seeded diaspores. However, multi-seeded diaspores cannot be responsible for the generally smaller size of angiosperm seeds in the Cretaceous. Tiffney estimated diaspore size by the volume of a rectangular prism ($L \times B \times D$) that could just surround the diaspore. Tiffney compared volume and weight for diaspores of 52 modern species. This comparison suggested that every 1 mm^3 would have corresponded to a weight of about 1 mg. Direct measures of seed density give slightly higher values ($1.1\text{--}1.5 \text{ mg/mm}^3$; Hodgson & Mackey 1986) but this is to be expected because a rectangular prism overestimates diaspore volume. Maximum and minimum seed sizes are sensitive to new data. The early Cretaceous pteridosperm *Ktalenia* has seeds approximately 2.8 mm long and up to 1.0 mm wide (Taylor & Archangelsky 1985). These specimens would reduce the minimum seed size for Cretaceous gymnosperms to about 2 mm^3 ($L \times B \times D$).

large-seeded forms. There is another possible interpretation. The fossil record of early angiosperms is dominated by Northern Hemisphere sites. If angiosperms originated in eastern Gondwana (Audley-Charles 1987), the early dominance of small-seeded angiosperms in the Northern Hemisphere may reflect more rapid dispersal of small-seeded forms from a distant center of origin. The same caveat applies to the conclusion that the earliest angiosperms were "weedy" species of disturbed sites (Hickey & Doyle 1977).

Angiosperms probably evolved from gymnosperms with relatively small seeds. Tiffney (1986) gives seed volumes for three Jurassic taxa that are putatively related to the angiosperms (*Pentoxylon*, 8 mm^3 ; *Caytonia*, 1.6 mm^3 ; cycadeoids, 5.2 mm^3 ; see Crane 1985, and Doyle & Donoghue 1986 for the presumed relationship of these groups to angiosperms).

Tiffney's data also supports the hypothesis, based on Chaloner & Sheerin's measurements of Carboniferous and Devonian propagules, that the smallest gymnosperm seeds were considerably larger than the smallest megaspores. Tiffney's estimates of seed volume were calculated as the product of length by width by breadth. Thus, the quoted seed volumes overestimate actual seed volumes but are useful for comparisons of relative size. A small megaspore, $200 \text{ }\mu\text{m}$ in diameter, has a volume of 0.008 mm^3 when calculated by the same method. This is over an order of magnitude smaller than the smallest pre-Cretaceous (gymnosperm) seeds.

III. SUMMARY

This brief survey of the fossil and modern flora has revealed a number of patterns. These patterns suggest some interesting questions which will be addressed in this thesis.

(1) Most modern vascular plants are heterosporous (including seed plants) but there are also substantial numbers of homosporous species. What accounts for the evolutionary persistence of homosporous life cycles? (Chapter 3)

(2) Almost all homosporous species have smaller propagules than almost all heterosporous species. What explains this association?

(Chapter 4)

(3) Dimorphic spores (heterospory) are strongly associated with sexual differentiation of gametophytes and endosporic development. What explains these associations? (Chapter 4)

(4) Seed plants have been much more successful than other heterosporous lineages. What features of their life cycle account for this success? (Chapter 5)

(5) Why are the smallest gymnosperm seeds much larger than the smallest megaspores? (Chapter 7)

(6) The rise to dominance of angiosperms is associated with a great increase in the total numbers of species within floras. Why are angiosperms more speciose than other groups? (Chapter 7)

(7) Seeds of the earliest angiosperms were generally smaller than those of contemporary gymnosperms and many were smaller than the smallest gymnosperm seeds in any period. Is this difference significant for understanding the success of angiosperms? (Chapter 7)

(8) The vast majority of modern vascular plants are angiosperms. What features explain this success? (Chapter 8)

(9) There is a long interval between the first appearance of seed plants and the first appearance of angiosperms. Were angiosperm-like plants absent prior to the Cretaceous because conditions were unsuitable or would angiosperm-like plants have been successful whenever they arose? In other words was the appearance of angiosperms waiting for the right conditions or waiting for the right adaptation? (Chapter 8)

Chapter 2

Models of parental allocation

A major theme of my thesis is the evolutionary significance of changes in propagule size. This chapter discusses two classes of model that describe parental allocation among offspring. Multiple strategy models (Section II of this chapter) are concerned with somatic polymorphism, the production of two or more types of propagule where the differences between morphs are not due to genetic segregation. Such models are relevant to my theme because somatic polymorphisms can involve size differences among propagules. Size-versus-number models (Section III of this chapter) address the evolutionary trade-off between producing many small propagules or a few large propagules. Both classes of model have wider applications than propagule size. In an Appendix to this chapter, I describe the optimal pattern of parental allocation when offspring differ in quality. These models are useful for describing microevolutionary change and optimal allocation strategies within species. They are less applicable to understanding why different lineages have vastly different propagule sizes.

I. MODELS AND STRATEGIES

The number and type of propagules produced by a sporophyte should be subject to natural selection. This section discusses how a sporophyte's allocation of resources to propagule production should be divided among propagules. Two classes of models can be identified: "allocation models" and "size-versus-number models" (Lloyd 1985, 1987, 1988). Some models combine properties of both classes.

Allocation models describe the optimal allocation among different types of propagules. The properties of the propagules, including their cost, are usually assumptions of the model. Such models consider how many propagules of each type should be produced for a given total allocation. Lloyd (1984) has presented a useful classification of possible strategies. In a *uniform*

strategy, all sporophytes produce a single type of propagule. In a *labile strategy*, propagules do not fall into discrete classes but show continuous adaptive variation. Uniform and labile strategies are both characterized by a unimodal distribution of propagule traits. The theoretical distinction is that variation around the mean is non-adaptive in uniform strategies but is adaptive in labile strategies. Uniform strategies should be subject to natural selection to reduce the variance around the mean. In a *conditional strategy*, individual sporophytes produce a single type of propagule but different sporophytes produce different types of propagule. Unlike a genetic polymorphism, the choice of propagule type is environmentally determined. In a *multiple strategy*, individual sporophytes produce two or more distinct types of propagule.

Size-versus-number models usually assume a single class of propagules and describe the optimal investment in individual propagules. A sporophyte has a finite amount of resources available for investment in propagules. The more resources supplied to a propagule, the greater should be the propagule's chances of survival and future reproduction, but the fewer resources should remain for the sporophyte to produce other propagules. Thus, a sporophyte could invest a small amount in each of a large number of inexpensive propagules or a large amount in each of a smaller number of propagules. The sporophyte's options lie on a continuum of increasing propagule "size" and decreasing propagule number.

I have been referring to a *sporophyte's* optimal investment in *propagules* but this is only a particular case of the more general problem of a *parent's* optimal investment in *offspring*. The more inclusive terminology is adopted below. Section II considers models of allocation among two or more types of offspring. These models are principally concerned with defining the conditions that favor a multiple strategy. Section III considers size-versus-number models. Such models usually assume a single class of offspring.

II. MULTIPLE STRATEGY MODELS

The idea that parents should produce a single type of offspring has an intuitive appeal. One type of offspring will give the maximum return per unit investment, and a parent's fitness will be maximized by producing only this type. Some plants, however, produce two or more distinct types of seeds. A single brood may contain seeds with and without obvious dispersal mechanisms, and/or a mixture of dormant and non-dormant seeds. Models that predict a multiple strategy rather than a uniform strategy have one of two features: "allocation-dependence" or "bet-hedging". Allocation-dependence is a general feature of models that predict spatial diversification within broods (dispersal), whereas bet-hedging is a general feature of models that predict temporal diversification within broods (variable dormancy). Ellner's (1986) model is an exception that predicts variable dormancy because of allocation-dependence.

Allocation-dependence arises when the increment in parental fitness from an additional offspring depends on the number and/or type of other offspring produced by the parent. Haig & Westoby (1988c) referred to such cases as frequency-dependence within broods. Lloyd (1988) introduced the term "allocation-dependence" to clarify the distinction between situations where the marginal return for a given allocation depends on a parent's previous allocation (allocation-dependence) and situations where the marginal return is independent of the parent's previous allocation but dependent on the allocations of other parents in the population (frequency-dependence). The distinction is useful and I adopt Lloyd's terminology. An optimization model is appropriate if there is allocation-dependence but no frequency-dependence because the strategies employed by other parents are not considered. Optimization models are generally considered to be inappropriate if there is frequency-dependence.

A number of models have shown an advantage of producing a mixture of dispersed and non-dispersed offspring (Hamilton & May 1977; Comins, Hamilton & May 1980; Comins 1982; Motro 1982a, 1982b; Schoen & Lloyd 1984). These models include effects of non-independence among parents (frequency-dependence), but it is sib-competition within broods (allocation-dependence) that is

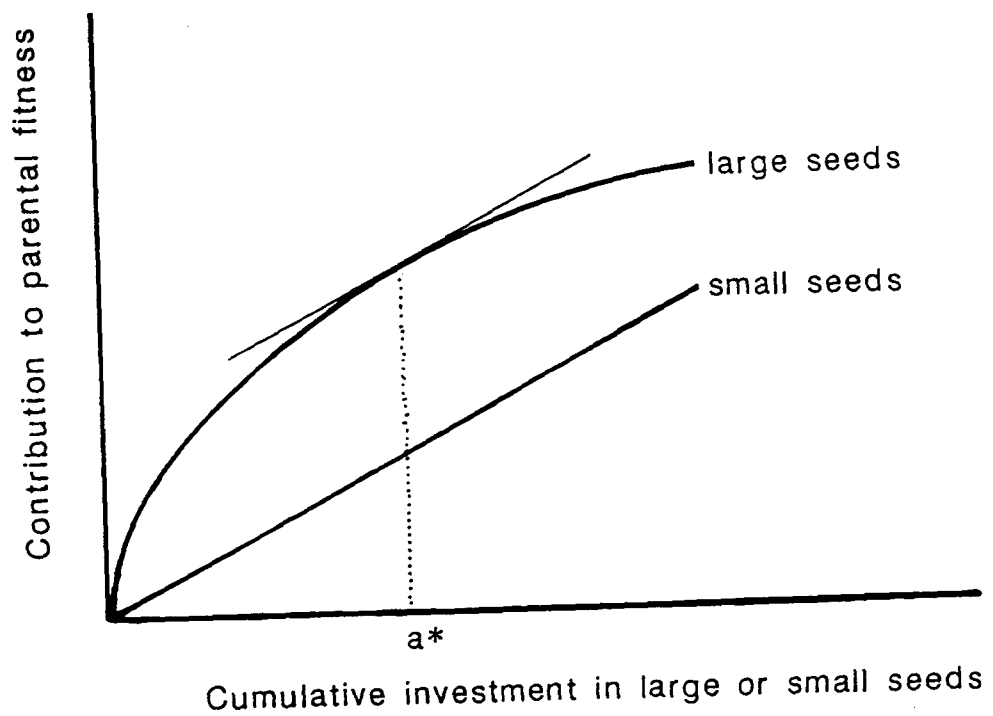
responsible for separately diminishing returns.

A simplified version (without frequency-dependence) of Schoen & Lloyd's (1984) "near and far dispersal model" will illustrate the means by which allocation-dependence can favor a multiple strategy. Consider a parent that produces poorly-dispersed large seeds and well-dispersed small seeds. Assume (1) that the parent's return per unit investment from large seeds declines as their number increases because of local resource competition; (2) that parental fitness increases linearly with number of small seeds; and (3) that the initial return per unit investment is lower for small seeds than for large seeds. The parent's optimal strategy is to produce large seeds until the marginal gain from producing additional large seeds equals the return per unit investment from small seeds. Thereafter, the parent should invest in small seeds (Figure 2.1).

My second assumption (linear returns from small seeds) was made for graphical convenience, but a multiple strategy is also predicted if there are separately diminishing returns on both large and small seeds. In general, a multiple strategy may be superior to either uniform strategy if there are separately diminishing returns on investment in the two types. The parent benefits by exploiting the more profitable "early returns" from both types of offspring (Lloyd 1984). At the parental optimum, the marginal returns from each type are the same but the return per unit investment may differ for the two types. The proportion of investment in each type will depend on the total amount of parental investment. For small total investments, a greater proportion will go to the type giving the highest initial returns. If the marginal return from this type never falls below the maximal return from the other type, the parent should only produce the higher-yielding type.

Bet-hedging models all have the property that the success of a strategy is unpredictable for a single reproductive attempt. A multiple strategy can be favored even though there exists a uniform strategy with a higher arithmetic mean fitness. This is possible because the geometric mean is usually a better estimate of long-term fitness than the arithmetic mean. A strategy with a small variance can have a higher geometric mean than an

Figure 2.1. An example of allocation-dependence (see text). A parent should invest in large seeds until a^* and then switch to investment in small seeds. At a^* the marginal return from large seeds equals the return from investment in small seeds.



alternative strategy with a higher arithmetic mean but a large variance (Gillespie 1977; Real 1980; Rubenstein 1982; Seger & Brockmann 1987). A multiple strategy can reduce the variance of parental fitness in one of two ways (Venable 1985). In *high-risk-high-risk heteromorphisms* offspring fitnesses are negatively correlated. Thus, when one type does badly, the other does well, and vice versa. In *high-risk-low-risk heteromorphisms*, one type gives the parent a higher average yield but with a high variance. The other type gives a lower average yield but with a low variance. The low-risk morph is generally able to exploit a wider range of conditions than the high-risk morph. "Bet-hedging" models have been widely used to explain the evolution of variable seed dormancy (Cohen 1966, 1967, 1968; Bulmer 1984; Levin, Cohen & Hastings 1984; Ellner 1985; Brown & Venable 1986; Silvertown 1988).

In summary, bet-hedging and allocation-dependent models can predict different types of offspring within a brood. Sib-competition and environmental uncertainty are common factors, so one might predict that multiple strategies would also be common. However, models that predict multiple strategies usually only consider two predetermined types of offspring and show that a parent producing both types has superior fitness to a parent producing only one type. It is always possible that there exists an even better uniform strategy in which the parent produces a third type of offspring.

III. SIZE-VERSUS-NUMBER MODELS

A. The Smith-Fretwell model

Smith & Fretwell (1974) developed a simple graphical model of the evolutionary trade-off between the amount invested in individual offspring and the number of offspring produced. They assumed that parental investment in an offspring could be represented by a single quantity and that an offspring's fitness could be expressed as a function of this investment (Figure 2.2). Parental fitness, for a given allocation, was defined as the return in offspring fitness per unit investment. Thus, parental fitness corresponded to the slope of a line from the origin to the appropriate point on the function. The line with the steepest

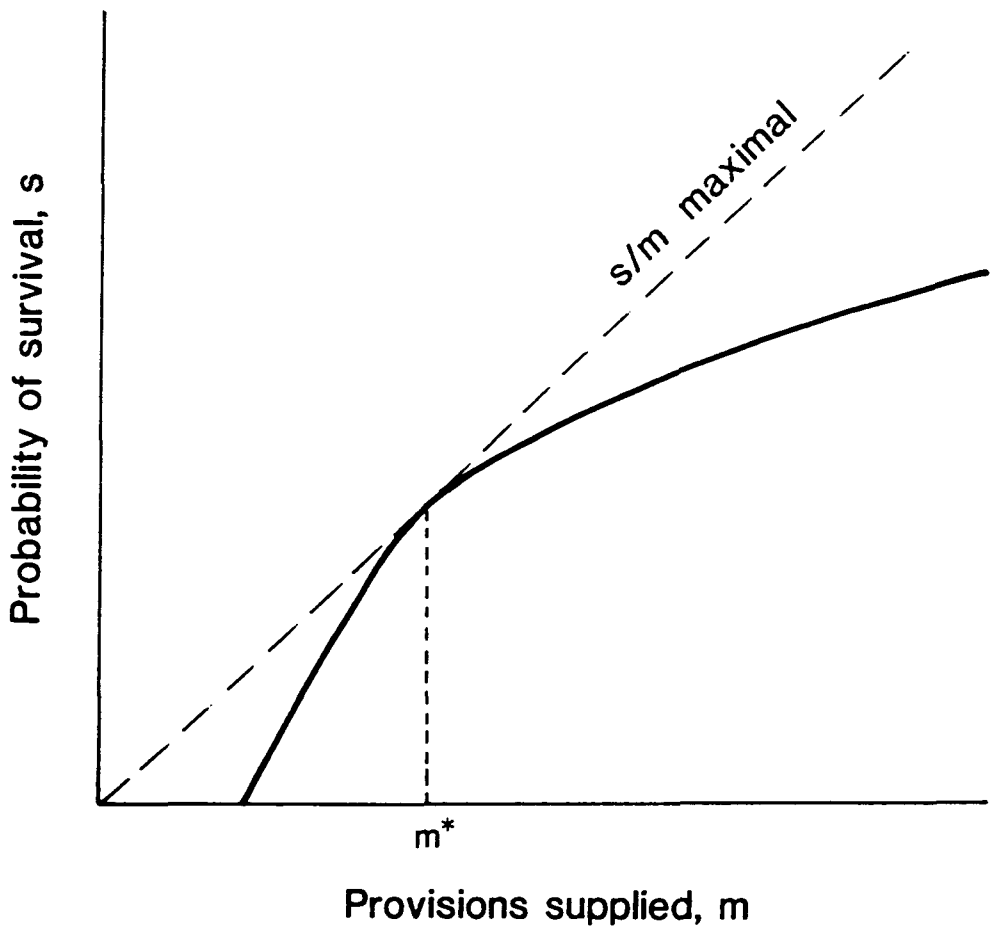
gradient gave the allocation to individual offspring that maximized parental fitness. Smith & Fretwell predicted that a parent should invest an equal amount in each offspring because only one value maximized return per unit investment.

The optimal parental allocation depended on the precise form of the function relating offspring fitness to the amount of parental investment per offspring (henceforth, the Smith-Fretwell function). Smith & Fretwell (1974) argued that offspring should have no chance of survival below some minimum investment and that offspring fitness should increase with investment above this minimum. Maximum parental fitness would be given by a tangent to the curve through the origin (Figure 2.2).

The form of the function proposed by Smith & Fretwell appears biologically reasonable. A minimum effective investment must exist, if only because nucleic acids need to be replicated (Smith & Fretwell 1974). The assumption that offspring fitness increases with investment is supported by ample evidence that larger food reserves enhance seedling fitness (McGinley, Temme & Geber 1987; Haig & Westoby 1988c). The additional assumption of a "sufficiently convex function" presupposes diminishing returns at higher levels of investment. McGinley et al. (1987) have reviewed the evidence for a convex relationship between seed size and seedling fitness.

The Smith-Fretwell model uses the device of a function that relates an offspring's fitness to the amount of parental investment it receives. An offspring's fitness is implicitly assumed to be independent of the amount of resources received by other individuals in the population. However, competition among propagules is likely to affect offspring fitness, and success in competition should be influenced by the amount of resources received by competitors. Therefore, a parent's fitness from a given allocation may depend on the allocation patterns of other parents in the population (i.e. fitnesses are frequency-dependent). Optimization models are generally considered to be inappropriate when the success of a strategy depends on the strategies employed by other individuals. The favored alternative is usually an explicit genetic model or an ESS analysis. An interesting theoretical question is to what extent the relevant

Figure 2.2. Probability s that an offspring will survive to reproduce as a function of m , the provisions supplied by its parent. The parent's fitness is greatest when m^* is supplied to each of T/m^* offspring, where T is the total parental investment.



population of strategists includes parents belonging to different species.

I believe that the model can still be useful, despite frequency-dependence, if a slightly circular definition of the Smith-Fretwell function is adopted. Suppose that the function is defined for the offspring of a rare parent in a population producing offspring of a given size or size distribution. If the population is at an ESS, the optimal size predicted by the model will correspond to the established size in the population. The conventional analysis would, therefore, describe the selective consequences of variant allocation patterns in a population near an ESS, but the model would not adequately describe dynamics away from the ESS. I use this approach in discussing the origin of heterospory (Chapter 4).

B. Variations on Smith and Fretwell

In a metaphorical sense, a parent can be said to possess information about its expected reproductive success from alternative allocation strategies. The source of this information is natural selection, the differential reproductive success of parents employing different strategies in the past. The information derived from natural selection is stored in the parent's genotype. A parent may adopt alternative allocation patterns in response to environmental cues. These cues could be interpreted as sources of information, but they have only acquired information content through the process of natural selection. (I am assuming plants lack any process analogous to learning.) If environments change, the information stored in the genome may become outdated.

The precision of adaptation is related to the number of different cues recognized by the parent in determining allocation. At one extreme, all reproductive attempts would be the same. The relevant Smith-Fretwell function would be the relationship between investment and offspring fitness averaged over all environments in the past. Parents should allocate the same amount to each offspring at every reproductive attempt. The allocation pattern observed would be adaptive on average but would be inferior to alternative strategies at some reproductive

attempts. At the other extreme, each reproductive attempt would be different and the allocation strategy would take account of the precise circumstances of the individual parent on the individual occasion.

There must be limits to the quality of information provided by past natural selection. The strategies of most real parents probably lie somewhere between these extremes. A simple extension to the model would be to assume that parents can recognize a limited number of situations with different Smith-Fretwell functions. Under the modified model, a parent should still invest an equal amount in each offspring within a brood, but this amount could vary among broods. Models of the Smith-Fretwell type thus predict either a uniform or a conditional strategy depending on the information available to the parent.

A parent might disperse its offspring into two or more habitats where offspring have different Smith-Fretwell functions. If the distribution of available habitats is the same for all parents at all reproductive attempts, the optimal strategy is to invest the same amount in all offspring (McGinley et al. 1987). The optimal investment is given by the tangent from the origin to a function that is the weighted average of the Smith-Fretwell functions for the different habitats (each habitat weighted by the proportion of offspring it receives). Thus, a parent's ability to direct offspring to specific habitats alters the weighting of habitats but does not change the prediction that there should be a single optimum investment.

The model of the previous paragraph assumed that there were no differences in the distribution of habitat types at different reproductive attempts or, more precisely, the model assumed that parents have no information about such differences. The appropriate Smith-Fretwell function was, therefore, based on the long-term average distribution of habitats. If a parent can distinguish between different types of reproductive event, allocation should be determined by the Smith-Fretwell function appropriate to each event (Haig & Westoby 1988c).

Temme (1986) investigated the optimal allocation to individual offspring if parents are able to distinguish between two (or more) classes of offspring with different Smith-Fretwell

functions. Temme argued that a parent's optimal strategy was to invest different amounts in different individuals (see Appendix for extensions to Temme's model).

If offspring number is determined before the size of the resource pool is known, a parent may have insufficient resources to provision all offspring with the amount that maximizes return per unit investment. Lloyd (1987) has examined this problem. A parent could either reduce the size of the brood and invest the optimal amount in the remainder, or provision all offspring with a suboptimal amount. In general, brood reduction is favored when a relatively small proportion of parental care has been invested, but reduced investment in all offspring is favored when offspring have received substantial investment. Brood reduction may be adaptive if reduction occurs before major resource commitment. The initiation of more offspring than the parent is able to provision could allow offspring number to be adjusted to changes in the resource pool or level of predation (Lloyd 1980, 1987). It could also allow offspring of inferior quality to be selectively aborted and allow resources to be redirected to offspring of higher quality (Westoby & Rice 1982; see Appendix below).

Parker & Begon (1986) examined the relationship between clutch size and egg size in insects, but their models have more general applicability. If there is competition within clutches, an offspring's fitness depends on the number and size of its siblings. Egg size should be greater in clutches with a larger total investment. The effect is due to allocation-dependence. Larger offspring size is also favored by competition between offspring of different parents. One way to look at this is that competitors (intraspecific or interspecific) are part of the environment. Thus, competition alters the Smith-Fretwell function and influences optimal offspring size. Parker & Begon's models predict constant egg size within clutches but possible variation among clutches.

All models so far have assumed that a parent's reproductive effort can be measured as a single quantity, but parents actually invest several different resources in offspring. McGinley & Charnov (1988) presented a model in which two or more resources could be considered simultaneously. In their model, the ratio of

resources invested in offspring was fixed by the relative pool-sizes. For a given ratio, investment could be represented as a single variable and the optimal allocation found as in the Smith-Fretwell model. The optimal allocation of a resource depended on its relative pool-size. The greater the supply of the resource relative to the supply of other resources, the greater the optimal allocation to individual offspring. Their model predicted that all offspring within a brood should receive the same level of investment but that this level could vary among broods. However, variation would be predicted within broods if different parts of the parent had autonomous resource pools (e.g. Watson & Casper 1984).

In summary, models of the Smith-Fretwell type usually predict that a parent should invest the same amount in all members of a brood (uniform or conditional strategy). This prediction is robust to spatial variability in the environment (McGinley et al. 1987), but not to detectable variation in offspring quality (Temme 1986). If offspring number is determined before the size of the resource pool, a parent may reduce the size of the brood or adjust the amount invested in each offspring (Lloyd 1987). The optimal investment of a single resource will depend on the relative pool-sizes of other resources (McGinley & Charnov 1988).

IV. OVERVIEW

Harper, Lovell & Moore (1970) reviewed contemporary knowledge on the shapes and sizes of seeds and concluded that single plants and populations usually have a unimodal and continuous distribution of seed sizes. The mean seed size of a species was claimed to be remarkably constant over a wide range of planting densities within an experiment, though it could vary considerably from site to site and from year to year. Seed size was more variable in species with determinate growth, presumably because growth form placed stronger constraints on seed number. In addition, Harper et al. reported considerable variation in the weight of individual seeds from the same plant. For example, the heaviest and lightest viable seeds on a single plant of *Trifolium subterraneum* varied in the ratio 17:1. Somatic polymorphisms,

with two or more sharply defined types of seeds, were common in four families, the Compositae, Chenopodiaceae, Gramineae and Cruciferae (Harper et al. 1970). Such polymorphisms are largely restricted to short-lived, fugitive species (Harper 1977, p. 71).

I am only aware of one study of an entire flora that examines the relative frequency of species with unimodal size distributions and species with seed polymorphisms. Lloyd (1984) surveyed the indigenous angiosperm flora of New Zealand (c. 350 genera). The vast majority of species produced a single type of fruit or seed. Only 14 genera contained species that produced two distinct types of structure. In nine genera this was due to a genetic polymorphism and in one genus to a conditional strategy. Multiple strategies were restricted to members of four genera. Lloyd's survey was based on morphology and would not have detected multiple strategies that were not associated with differences in seed form. A unimodal size distribution appears to be the predominant strategy in the New Zealand flora, though it is not clear to what extent New Zealand is typical of the rest of the world.

Harper et al. (1970) emphasized that seed size was less plastic than other components of reproductive yield. During the last decade, the emphasis of ecological studies has shifted towards the variation in seed size within species and seed size has been viewed as a plastic character (Silvertown 1989). Seed size is never absolutely constant within a species or within a brood of a single individual. For example, Michaels et al. (1988) measured seed size variation in 39 species from the Illinois flora. They obtained coefficients of variation greater than 10% in all species but none greater than 100%. Rather than waste time with the semantic question as to whether this means seed size is highly variable or relatively constant we should ask how much of the observed variation in seed size is due to adaptive responses by parents and how much is due to uncontrolled variation in development (McGinley et al. 1987; Haig & Westoby 1988c). The models discussed above identify factors that should favor adaptive variation in seed size and factors that should favor a uniform seed size.

A crucial question has not been addressed by the models.

Seed size variation between species is much greater than seed size variation within species. What explains this variation? Environmental factors are undoubtedly important but there is still substantial variation in seed size among species within habitats. Size-versus-number trade-offs can only be defined for a particular kind of unit. If I have a billion dollars to invest in a chain of businesses, my optimal investment in each business depends on whether the businesses are international hotels or hot dog stands. Similarly, the optimal size of a propagule will depend on whether the propagule is an isospore establishing a fern prothallus or an avocado seed establishing a large-leaved seedling.

Developmental constraints are often seen as a reason why characters should be non-optimal. I prefer the alternative view (outlined above) that a character can only be optimized given a particular set of constraints. That is, a phenotype is only ever described as optimal with respect to an implicit or explicit set of alternative phenotypes. The constraints define the set of alternatives, and the optimal phenotype may change if the constraints change. In optimality models, some characters are allowed to vary whereas other characters (the constraints) are treated as invariants. The constraints of one model may be variables in another model. The distinction between constraints and variables is based on the plausible assumption that, in evolutionary time, some characters vary less readily than others. Thus, whether a character is defined as a variable or a constraint is partly a matter of time-scale.

This view of optimum seed size and developmental constraints raises two kinds of question. First, in what ways do propagules differ in their development, structure or physiology so as to have different optimal sizes? For example, families with a single seed per carpel tend to have larger seeds than families with many seeds per carpel (Hodgson & Mackey 1986). Does ovule number influence the size-versus-number trade-off, or is ovule number correlated with some other character that does? Second, how are species that produce different types of propagule able to coexist? Much has been written on ecological coexistence but I will not review the various models here (for an introduction to

lottery models see Fagerström 1988; for a recent development of competition models see Tilman 1988).

Appendix

**Brood reduction and optimal parental investment
when offspring differ in quality**

Smith & Fretwell (1974) predicted that an optimal parent should invest the same amount in each offspring. However, their model assumed that the function relating offspring fitness to parental investment was the same for all offspring or, more precisely, the model assumed that a parent could not detect differences in offspring quality before resources were committed. Temme (1986) argued that equal investment did not maximize parental fitness if offspring differed in quality. Rather, the parent should invest different amounts in different offspring such that the marginal return from additional investment would be the same for all offspring.

In this appendix, I extend Temme's argument by deriving the optimal parental investment for different relative frequencies of offspring types. I also consider when it is in a parent's interests to abort rather than provision offspring of lower quality. My model addresses the basic question: how should a parent allocate resources among offspring when the offspring have different expectations of fitness given the same parental investment? The answer is simple. Parental resources are optimally distributed when (1) the marginal return from each provisioned offspring equals the average return from all offspring, and (2) offspring whose quality falls below some threshold are aborted and yield no return on their cost.

THE MODEL

The model presented below assumes that the number of offspring produced by a parent is determined by the parent's allocation to individual offspring and the size of the resource pool. That is, the number of offspring is constrained by the allocation to individual offspring and the size of the resource pool. However, the allocation to individual offspring is assumed to be unconstrained by total allocation. This implies that total

allocation is large relative to the individual allocation and that a change in a parent's allocation strategy or a change in total allocation results in an appropriate change in the number of offspring. Changes in allocation strategy are assumed to be without direct cost to the parent, but different strategies may give different returns on investment. Three assumptions are made about the function relating offspring fitness to parental investment: (1) there is some minimum investment below which an offspring has zero fitness, (2) offspring fitness is an increasing function of parental investment, and (3) the function is convex (i.e., there are diminishing marginal returns). All these assumptions are either implicit or explicit in the models of Smith & Fretwell (1974) and Temme (1986). In my model, a parent cannot vary the proportion of offspring of a given quality but can adjust the amount invested in individual offspring of different quality.

Suppose that offspring belong to two types, A and B, with relative frequencies p and $(1 - p)$. A parent invests an amount a in each A offspring and b in each B offspring. Each offspring makes an independent contribution to parental fitness, and the offspring's contribution is a function of the resources it receives. Differences in quality between A and B offspring are expressed as different functions $f_A(a)$ and $f_B(b)$. Parental fitness, w , can be defined as the average return per unit investment.

$$w = \frac{pf_A(a) + (1 - p)f_B(b)}{pa + (1 - p)b} \quad (2.1)$$

By assumption of the model, a and b are subject to diminishing returns. If the marginal returns from A and B offspring are unequal, the parent could increase its average return (fitness) by allocating a little more to each offspring with the higher marginal return and a little less to each offspring with the lower marginal return. If the marginal returns are equal but less than the average return, the parent could increase its fitness by reducing the total number of offspring and allocating a little more to each offspring. Therefore, the optimal set of allocations

$(\hat{a}, \hat{b})$ occurs when the marginal returns from investment in each type is equal to the average return per unit investment. That is,

$$\hat{w} = f'_A(\hat{a}) = f'_B(\hat{b}) \quad (2.2a)$$

$$\hat{w} = \frac{pf_A(\hat{a}) + (1-p)f_B(\hat{b})}{p\hat{a} + (1-p)\hat{b}} \quad (2.2b)$$

The values of $\hat{a}$ and $\hat{b}$ depend on p . The model's behavior can be examined by considering the extreme conditions when $p = 1$ (all type A) and $p = 0$ (all type B).

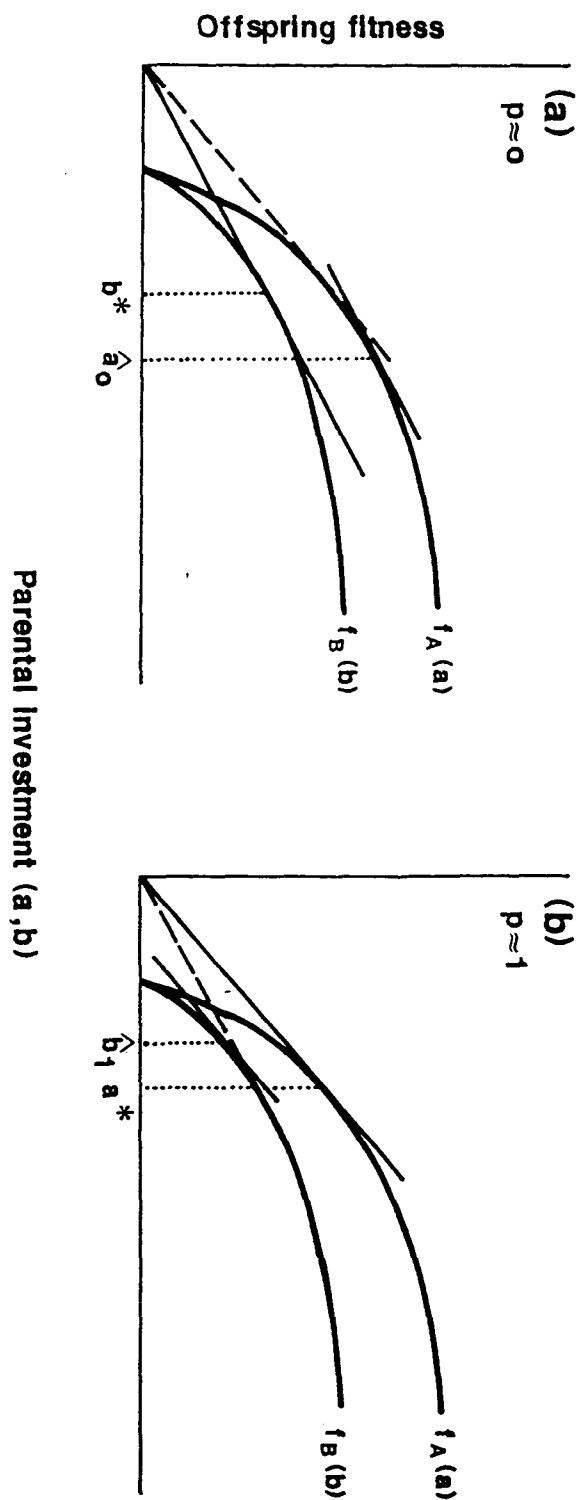
$$p = 1, \quad f'_A(a^*) = f_A(a^*)/a^* \quad (2.3a)$$

$$p = 0, \quad f'_B(b^*) = f_B(b^*)/b^* \quad (2.3b)$$

The allocations a^* and b^* are the Smith-Fretwell allocations for A and B offspring respectively. Suppose that, for an equivalent investment, A offspring have superior fitness to B offspring. If a parent produces predominantly B offspring ($p \approx 0$), each B offspring should receive b^* and the rare A offspring should receive $\hat{a}_0$ such that $f'_A(\hat{a}_0) = f_B(b^*)/b^*$ (Figure 2.3a). The optimal allocation to each A offspring will be greater than a^* . As the proportion of A offspring increases, $\hat{a}$ approaches a^* and $\hat{b}$ decreases below b^* . If the parent produces predominantly A offspring ($p \approx 1$), each A offspring should receive a^* and the rare B offspring should receive $\hat{b}_1$ such that $f'_B(\hat{b}_1) = f_A(a^*)/a^*$ (Figure 2.3b). The greater the proportion of A offspring, the greater is parental fitness at the optimal strategy and the smaller is the investment by the parent in individual offspring of either type.

Temme (1986) similarly assumed equal marginal returns from investment in both types, but did not consider variation in the proportion of offspring belonging to each type. The dependence of optimal allocation on p is easily explained. If A offspring are more common, the parent can achieve a higher average return. Therefore, marginal returns from individual offspring of both types will be greater (because marginal returns equal average

Figure 2.3. The optimal parental investment in type A and type B offspring is given by the intercept of the solid tangent with the appropriate curve: (a) offspring predominantly type B; (b) offspring predominantly type A.



returns for the optimal strategy) and the allocation to individual offspring of both types will be less (because of diminishing marginal returns).

The model can be extended to consider selective brood reduction. Suppose that the parent aborts all B offspring at some cost c per aborted offspring, then

$$w = \frac{pf_A(\tilde{a})}{p\tilde{a} + (1-p)c} \quad (2.4a)$$

The optimal investment in A offspring when B offspring are aborted, $\tilde{a}$, occurs when marginal returns equal the average return.

$$\tilde{w} = f'_A(\tilde{a}) = \frac{f_A(\tilde{a})}{\tilde{a} + (1-p)c/p} \quad (2.4b)$$

The optimal strategy is represented graphically in Figure 2.4. The line AA' is the tangent to $f_A(a)$ drawn from the point $(1-p)c/p$ to the left of the origin. The point where this tangent touches the fitness function defines $\tilde{a}$. The result is the same as the Smith-Fretwell model for A offspring except that there is an additional cost $(1-p)c/p$ associated with each A offspring. The factor $(1-p)/p$ is the number of aborted B offspring per provisioned A offspring.

Suppose that a parent provisions each A offspring with $\tilde{a}$. The parent can achieve equivalent fitness by aborting B offspring at cost c or by provisioning B offspring with amount b if

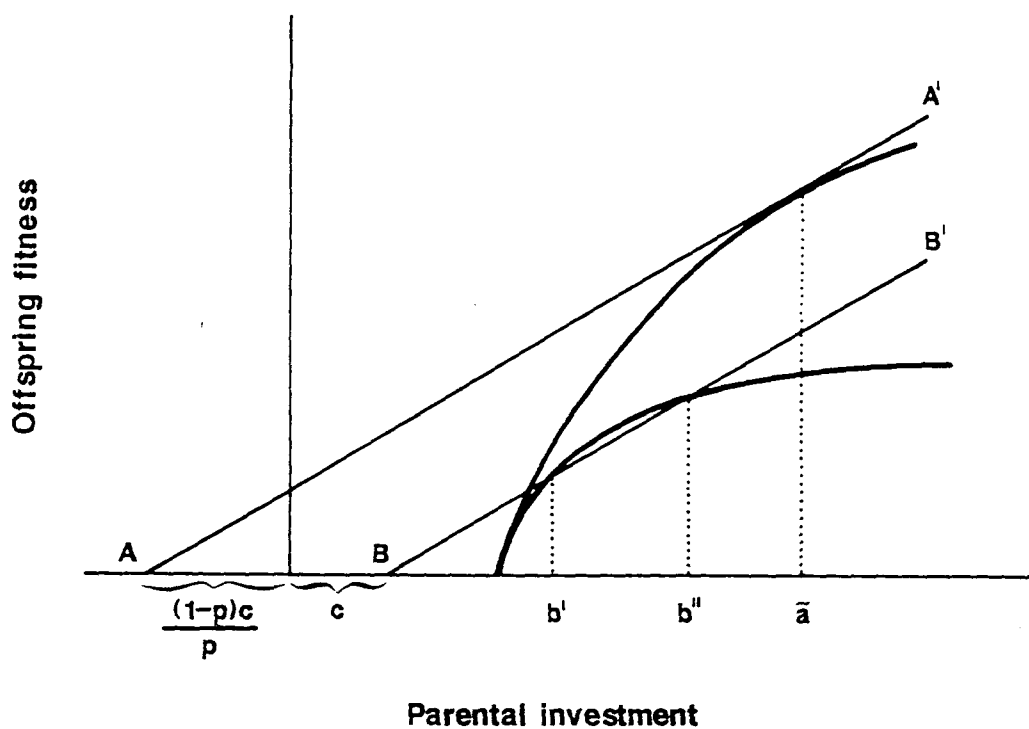
$$\frac{pf_A(\tilde{a})}{p\tilde{a} + (1-p)c} = \frac{pf_A(\tilde{a}) + (1-p)f_B(b)}{p\tilde{a} + (1-p)b} \quad (2.5a)$$

which simplifies to

$$\frac{pf_A(\tilde{a})}{p\tilde{a} + (1-p)c} = \frac{f_B(b)}{(b-c)} \quad (2.5b)$$

This result has the intuitive interpretation shown in Figure 2.4.

Figure 2.4. The optimal parental investment in A offspring when B offspring are aborted is given by the intercept of the tangent AA' with $f_A(a)$. The proportion of A offspring is p and the cost of aborting a B offspring is c . A strategy in which B offspring are provisioned is superior to the abortion strategy if the line BB' (parallel to AA') intersects $f_B(b)$ (see text).



The left-hand side of (2.5b) is the average return per unit investment for the best abortion strategy. This is represented by the slope of the line AA' . The right-hand side of (2.5b) is the average return on that portion of investment in B that is in excess of the fixed cost c . BB' is the line, parallel to AA' , that cuts the horizontal axis at the point c . The right-hand side of (2.5b) equals the left-hand side at b' and b'' where BB' cuts $f_B(b)$. Since this function is convex (by the assumption of diminishing returns), the parent could achieve higher fitness by investing b between b' and b'' . Therefore, there exist allocations $\hat{a}$ and b that give greater average returns than the abortion strategy. If BB' did not intersect $f_B(b)$, then the abortion strategy would be superior to any strategy in which B offspring were provisioned. The distance between the horizontal intercepts of AA' and BB' is c/p ($= [1 - p]c/p + c$). The smaller the value of c/p , the smaller the difference in quality between A and B offspring that would favor the abortion of the lower quality type.

The model's overall dynamics can be summarized. As p (the proportion of higher-quality A offspring) increases, the optimal allocations, $\hat{a}$ and b , decrease. Above some threshold value of p , the parent's optimal strategy is to abort B offspring at cost c . If B offspring are aborted, increases in p continue to decrease $\hat{a}$ because this reduces the cost of abortion per A offspring (i.e. reduces $[1 - p]c/p$). At the limit (when $p = 1$), $\hat{a}$ equals the Smith-Fretwell allocation a^* . For given $f_A(a)$ and $f_B(b)$, the smaller the value of c , the smaller is the threshold value of p .

The model presented here is a specific case of the economic principle that a profit-maximizing firm tends to equalize the ratio of marginal product over cost for all activities (see Bloom, Chapin & Mooney 1985; Lloyd 1988; or any economics text). This principle defines both the optimal parental allocation among offspring that differ in quality and the optimal time spent foraging in patches of different quality. In both cases, an individual attempts to maximize the return (fitness or energy) from an investment (resources or time) in units (offspring or patches) that differ in quality. There is a strict analogy between Charnov's (1976) conclusion that an optimal predator

should abandon a patch once the marginal capture rate drops to the average capture rate for the habitat and my conclusion that an optimal parent should terminate investment in an offspring once the marginal return falls to the average return from all offspring.

Because of the close analogy to optimal foraging theory, theorems about foraging can be translated into theorems about parental allocation. Arditi and Dacorogna (1988) have developed an elegant model of optimal foraging in a one-dimensional environment with an arbitrary food distribution. In their model, a forager moves through the environment, adjusting its speed to the local availability of food. The slower the forager's movement, the greater the investment of time in foraging. The optimal strategy is to move at the maximal rate (without foraging) through areas with food availability below some threshold and to slow down in areas above the threshold so as to reduce local food availability to the threshold. Their model can be translated into a statement about parental allocation for an arbitrary distribution of offspring quality. Successive offspring of a parent can be represented as contiguous patches encountered by a forager. Aborted offspring are patches below the threshold, and the cost of abortion is the time spent in such patches. A parent's optimal strategy is to abort all offspring whose quality falls below some threshold and to provision each remaining offspring until the marginal return on investment equals the average return.

DISCUSSION

Offspring differ in quality to the extent that they have different expected fitnesses given the same parental investment. Temme (1986) and I have shown that equal investment does not maximize parental fitness if a parent can detect differences in offspring quality. Rather, a parent should equalize the expected marginal returns from all provisioned offspring. As a result, some offspring should receive more resources than others. The "expected fitnesses" and "expected marginal returns" are the expectations given the information available to the parent. A parent might only recognize broad categories of offspring (e.g.

offspring with gross developmental abnormalities versus normal offspring) or might be able to make finer distinctions.

Variation in quality is not necessarily genotypic. Thus, seeds produced late in the season are often smaller than seeds produced early in the season (Cavers & Steel 1984). A possible explanation is that late seeds have lower expected fitness than early seeds given the same parental investment. Other explanations are possible (McGinley & Charnov 1988) but the hypothesis does make testable predictions about the relative success of seeds sown at different times.

Brood reduction is an extreme case of differential investment. Lloyd (1987) demonstrated that brood reduction may be adaptive if total reproductive allocation is uncertain at the time when offspring number is determined. This note demonstrates that brood reduction may also be adaptive if offspring differ in quality. My model identifies when it is in a parent's interests to abort rather than provision offspring of lower quality. This formalizes the hypothesis that selective brood reduction can enhance parental fitness if abortion occurs before substantial resources are committed to offspring (see Queller 1987; Stearns 1987; for recent discussions of this hypothesis).

Chapter 3

Sex expression in homosporous pteridophytes

Homosporous reproduction is the ancestral condition in vascular plants. After dispersal, a spore germinates to produce a photosynthetic gametophyte, which may attempt to reproduce as a male, as a female, or as a hermaphrodite. Gametophytes are required to be photosynthetic (a few species are mycotrophic) because spores contain insufficient resources for a spore's gametophyte to reproduce successfully as a female, without first accumulating additional resources.

Mating occurs within small local populations determined by the range of sperm movement. Therefore, nearby gametophytes are potential mates, but also potential competitors for fertilizations. The sporophyte that develops from a fertilized egg becomes established at the site of fertilization, and sporophytes produced by neighbouring gametophytes are potential competitors for resources. Whether a gametophyte will have higher expected fitness by producing eggs or sperm depends on the sex expression and relative condition of other gametophytes within the local population. Gametophytes have labile sex expression because the amount of resources they control relative to nearby gametophytes will depend on microsite differences and relative time of germination.

Patterns of sex expression in homosporous ferns are consistent with adaptive sex determination by gametophytes in response to local environmental conditions which determine growth rates, and in response to information about the presence, sex expression and relative condition of nearby gametophytes. The minimum amount of resources necessary for male reproduction is less than the minimum amount necessary for female reproduction. Therefore, smaller gametophytes tend to reproduce as males, particularly when in the vicinity of larger female gametophytes.

I. INTRODUCTION

Sex allocation theory is concerned with the optimal allocation of reproductive resources between male and female functions. The theory has been extensively developed for seed plants (Charnov 1982; Goldman & Willson 1986), but other groups of plants have been comparatively neglected. Models developed for seed plants are not directly applicable to homosporous species because of important differences in the life cycles of the two groups. This chapter presents a theoretical account of optimal sex expression in plants with homosporous life cycles. The discussion will concentrate on homosporous ferns but should also be relevant to homosporous psilophytes, lycopods and sphenopsids. Willson (1981) has already considered sex expression in homosporous ferns from an evolutionary perspective, and this chapter builds on her work by formulating a more explicit theoretical framework. The ideas developed in this chapter are used in Chapter 4 to formulate a model for the origin of heterosporous life cycles.

Sex allocation in seed plants is usually considered to be a property of sporophytes, even though eggs or sperm are produced by gametophytes. This is possible because the sex expression of gametophytes is determined at spore formation. Moreover, almost all the resources for reproduction by gametophytes are supplied by their parent sporophyte. In contrast, models of sex expression in homosporous ferns should be gametophyte-centered because their spores are indeterminate with regard to sex and because a substantial proportion of reproductive resources are acquired by the gametophyte's own activities rather than supplied by its parent sporophyte.

Homosporous ferns also differ from seed plants in the relative timing of fertilization and propagule dispersal. The dispersal stage of homosporous ferns is the spore and fertilization occurs after dispersal whereas the dispersal stage of seed plants is the seed and fertilization occurs before dispersal. Therefore, models of sex allocation in homosporous ferns should differ from equivalent models for seed plants in two important respects. Firstly, the local mating population will be determined by the dispersal distance of sperm rather than the dispersal distance of pollen. Secondly, fern sporophytes develop

at the site of fertilization whereas seed plant sporophytes are dispersed and may develop at a location distant from the site of fertilization.

In the model which follows, I will define a gametophyte's male reproductive effort as all costs involved in the production and successful function of sperm. A gametophyte's female reproductive effort will be defined as the costs of producing eggs and of supporting the young sporophyte. The model seeks that sex expression of a gametophyte which maximizes the gametophyte's expectation of reproductive success. To make the model more explicit the chapter is presented in the form of premises, followed by inferences which follow from the premises, followed by discussion of the evidence in relation to inferences and premises. The chapter should not be interpreted as testing the model by strong inference, however, because the predictions were made in awareness of the evidence. Rather the aim is to spell out a coherent structure of theory within which the available evidence can be organized. An explicit model can be of use in understanding evidence that conflicts with the model, as well as evidence that conforms to the model. Provided that the inferences of a model are logical consequences of the model's premises, discrepancies between evidence and inferences indicate the need for modified premises.

II. THE MODEL

A. Premises

(A1) Resources committed to male reproduction are unavailable for female reproduction. Similarly, resources committed to reproduction are unavailable for vegetative growth. Growth may allow greater reproductive effort in the future, but a delay in reproduction runs the risk of losses due to death, injury or missed opportunity.

(A2) Sexual reproduction occurs within small local populations of gametophytes that are defined by the limited mobility of sperm. A gametophyte's optimal sex allocation depends on the sex expression, number and condition of other gametophytes within the local population but a spore cannot predict these factors at the

time of dispersal.

(A3) To succeed as a female parent a gametophyte needs sufficient resources to support a young sporophyte. The more resources supplied to the sporophyte the better should be its chances of survival, though this investment is probably subject to diminishing returns.

(A4) A young sporophyte will compete for light and other resources with nearby sporophytes. Small early growth advantages may be very important for ultimate success.

(A5) Success as a male parent depends on fertilizing those eggs which go on to become successful sporophytes. Expected fitness of a gametophyte should increase with the number of sperm produced but with diminishing returns because there can only be a finite number of eggs within the range of sperm movement.

(A6) Sperm-producing gametophytes within a local population compete for fertilizations of the available eggs.

(A7) The cost of producing a sperm is much less than the cost of supporting a young sporophyte. Therefore, a growing gametophyte will reach a sufficient size for male reproduction before it reaches a sufficient size for female reproduction. Similarly, a small gametophyte may have some chance of reproductive success as a male, even if it has no chance of success as a female.

(A8) The costs of male reproduction are experienced before fertilization, but the major costs of female reproduction are experienced after fertilization. A large gametophyte may produce archegonia but most of its resources are not committed to female reproduction until they are supplied to a young sporophyte.

(A9) Sporophytes produced by intergametophytic matings have, on average, superior fitness to the completely homozygous products of intragametophytic selfing. This advantage is sufficient for intergametophytic mating to have been favored by natural

selection.

(A10) Archegoniate gametophytes release substances which can be detected by other gametophytes. (These substances will be referred to as "signal molecules" in the next section.)

B. Inferences

(B1) A gametophyte's sex expression should vary in response to different 'social' environments. Labile sex expression is predicted because the composition of local gametophyte populations should be subject to large stochastic variation (A2). The predicted flexibility will be limited by the information available to the gametophyte. Signal molecules (A10) indicate the presence of archegoniate gametophytes, and the concentration of signal molecules may give additional information about their number, distance or condition. However, a given concentration is likely to have several possible interpretations. The first evidence of nearby antheridiate gametophytes may be the fertilization of an egg.

(B2) A gametophyte that detects signal molecules before it has reached a sufficient size for female reproduction should produce antheridia. This is because the gametophyte is growing near an archegoniate neighbour (A10) and has the opportunity to fertilize this neighbour's eggs. Female reproduction would be more costly (A7) and the neighbour has a head-start in competition as a female parent (A4). Antheridiate gametophytes could subsequently become hermaphroditic because of diminishing returns on male investment (A5). There may be some chance of successful female reproduction, even if the gametophyte is in direct competition with a more advanced archegoniate neighbour.

(B3) A gametophyte that attains a size sufficient for female reproduction without detecting signal molecules should produce archegonia. An archegonia-first strategy is favored over an antheridia-first strategy because there is no point in producing antheridia if there are no eggs to fertilize. Moreover, unlike male reproduction, substantial resources are not committed to

female reproduction until after fertilization (A8). If non-archegoniate gametophytes are present, they will produce antheridia in response to signal molecules released by the archegoniate gametophyte (A10 & B2). If the gametophyte occurs alone, its eggs will remain unfertilized. Such gametophytes should become hermaphroditic if intragametophytic selfing is a better option than waiting. Archegoniate gametophytes that subsequently become hermaphroditic could also be explained by diminishing returns on female investment (A3).

(B4) An archegoniate gametophyte should provision only one sporophyte because two sporophytes growing side by side are unlikely to increase maternal fitness. The division of resources between siblings may result in both being outcompeted by the sporophyte of another gametophyte (A3 & A4).

(B5) Gametophytes growing under poor conditions should tend to reproduce as males because of the lower minimum costs of male reproduction (A7). This prediction applies within species but not between species.

C. Evolution of signal molecules

A central premise~~of~~ the model (A10) was that archegoniate gametophytes release substances detectable by other gametophytes. Such substances are known, and are called antheridiogens because non-archegoniate gametophytes respond to them by forming antheridia (see below). I referred to these substances as "signal molecules" because I wished to show that the induction of antheridia is adaptive and predictable *a priori*. Antheridiogen is a substance released by one gametophyte which elicits a response by another gametophyte. I can suggest two models for the origin of such a system.

(1) Archegoniate gametophytes released a substance with some function unrelated to its future role as an antheridiogen. Non-archegoniate gametophytes were selected to form antheridia in response to the substance because this increased their chance of paternity and avoided competition as a female with a more advanced neighbour.

(2) The future antheridiogen was a hormone which initiated antheridium formation. Archegoniate gametophytes were selected to release this substance as a pheromone because this caused neighbors to redirect resources from growth to antheridia, thus reducing competition and facilitating out-crossing.

No matter what their origin, I believe antheridiogen systems are maintained because they benefit both the signalling and the responding gametophyte. Willson (1981) recognized this possibility but emphasized conflicts of interest. She suggested that antheridiogens could be (1) a method by which parental sporophytes could manipulate gametophyte sex expression to ensure adequate fertilization or (2) a form of allelopathy by which female gametophytes could stunt the growth of their competitors. I recognize that parental sporophytes and female gametophytes may benefit from antheridium induction but doubt that induction is against the interests of the induced gametophyte. Antheridial development is under the control of genes expressed in this gametophyte and there should be strong selection against non-adaptive responses. In their discussion of animal signals, Krebs & Dawkins (1984) concluded that cooperative signals should evolve towards a compromise between economy and detectability whereas non-cooperative signals should be more expensive and exaggerated. Antheridiogens probably belong to the first category.

III. THE EVIDENCE

A. Antheridiogen systems in ferns

The model assumed that gametophytes have been selected to favor out-crossing over self-fertilization (A9). Electrophoretic studies have shown that intergametophytic matings predominate in most species of homosporous ferns, but that a few species show high frequencies of intragametophytic selfing (Haufler 1987; Soltis & Soltis 1987). It is not surprising that ferns, like angiosperms, show variation in the degree of inbreeding. As in angiosperms, ecological factors are likely to be important in explaining the variation (see Lloyd 1974b). Past mating history will also influence the relative merits of intergametophytic and intragametophytic fertilizations. Past outcrossing allows the accumulation of deleterious recessives, and thus increases the

relative merits of current out-crossing (see Appendix below; Hedrick 1987). My model only applies to species with intergametophytic mating systems.

In the remainder of this section I discuss sex expression in homosporous ferns with particular emphasis on antheridiogen systems. *Pteridium aquilinum* will be used as an example. *Pteridium* gametophytes begin to release antheridiogen shortly before they produce archegonia. The antheridiogen induces vegetative gametophytes to form antheridia. The archegoniate gametophytes are insensitive to this effect and only form antheridia after a prolonged period without fertilization. In multispore culture, the largest gametophytes are the first to release antheridiogen. Therefore, these gametophytes develop archegonia and all others develop antheridia. Antheridiate gametophytes are of two types. The smallest lack a notch meristem and produce antheridia only. The others differentiate a notch meristem and later develop archegonia. In single spore culture all gametophytes grow to a size at which they release antheridiogen and form archegonia. However, if the medium contains antheridiogen, all gametophytes precociously develop antheridia (summarized from Näf 1979).

Näf (1979) listed 10 species with demonstrated antheridiogens and a further 24 species which responded to *Pteridium* antheridiogen. Since then several other species have been shown to respond to antheridiogens (e.g. Haufler & Ranker 1985; Gemmrich 1986b). As far as is known, sex expression in these species resembles that in *Pteridium*, though there is variation in details. For example, archegoniate gametophytes of *Ceratopteris thalictroides* produce antheridia after a brief unisexual stage (Klekowski 1970a) but *Bommeria* gametophytes remain unisexual unless they are fragmented (Haufler & Gastony 1978). This variation clearly affects the probability of self-fertilization and may be the result of selection for different mating systems. Among homosporous ferns, antheridiogens have been looked for but not found in *Polypodium crassifolium* (Schraudolf 1967) and *Acrostichum danaeifolium* (Lloyd & Gregg 1975). At present there is insufficient evidence to determine the frequency of species with and without antheridiogens but antheridiogens

would be expected to be less common in apomicts and species which predominantly self-fertilize (Lloyd 1974b).

The antheridiogen of one species may cross-react with the antheridiogens of related species. For example, antheridiogens of *Pteridium aquilinum* (A_{Pt}) and *Pteris vittata* (A_{Ps}) induce antheridia on each other's gametophytes and on the gametophytes of several related species. However, some species that react to A_{Pt} do not react to A_{Ps} and vice versa (Gemmrich 1986b). Antheridiogens in the Schizaeaceae are chemically related to gibberellic acid (Nakanishi et al. 1971; Yamane et al. 1979). The antheridiogen of *Anemia phyllitidis* has recently been synthesized and given the name antheridic acid (Corey & Myers 1986). Antheridic acid is also an antheridiogen in *A. rotundifolia* and *A. flexuosa* (Yamane et al. 1987), but a slightly different molecule is the antheridiogen in *A. mexicana* (Nester, Veysey & Coolbaugh 1987). Species that respond to schizaeaceous antheridiogens do not respond to A_{Pt} and vice versa (Näf 1979).

Different strains within the same species sometimes show different responses to antheridiogen, as indicated by different sex ratios in multispore culture. Differences in antheridiogen response between two strains of *Ceratopteris richardii* are the result of alleles at one or two nuclear loci (Scott & Hickok 1987). Artificial mutagenesis in the same species has produced strains that are insensitive to conspecific antheridiogen (Warne, Hickok & Scott 1988).

Fern spores usually require light to germinate (Miller 1968). In at least some species, spores will germinate in the dark if antheridiogen is present in the medium (Näf 1966; Gemmrich 1986b; Yamane et al. 1987). Schraudolf (1985a) and Schneller (1988) have discussed the biological function of antheridiogen-induced dark germination. The light-requirement for germination makes adaptive sense because gametophytes that develop in the dark will be unable to photosynthesize. Enforced dormancy may allow subsequent germination under more favorable conditions. The response to antheridiogen is probably also adaptive because, under natural conditions, antheridiogen indicates the presence of a nearby archegoniate gametophyte (in the light). Food reserves in the spore may be sufficient to

produce an antheridium and fertilize an egg of the nearby gametophyte, even though food reserves might be insufficient for reproduction under other circumstances. Gametophytes of *Polypodium crassifolium* (Schraudolf 1967), *Onoclea sensibilis* (Rubin, Robson & Paolillo 1984) and *Pteris vittata* (Gemmrich 1986a) form antheridia when grown entirely in the dark. Thus, spore food reserves are sufficient for antheridium formation in these species.

Light also affects the growth form of developing gametophytes. A germinating spore forms a protonema which elongates by tip growth. If exposed to low light intensities, a protonema will continue to elongate until its spore's nutrient reserves are exhausted. If the protonema encounters high light intensities, tip growth is replaced by planar growth and a prothallus is formed (Cooke & Racusen 1988). Tip growth can be reinitiated by transferring a prothallus to low light (Näf, Sullivan & Cummins 1974). Cooke & Racusen (1988) expressed surprise when a geometric analysis showed that a planar prothallus was no more efficient at photoreception than a protonema of equivalent volume. However, their analysis assumed an homogenous environment. In a dappled environment, tip growth would be more efficient at reaching regions of high light intensity and planar growth would be more efficient at exploiting such regions.

The model predicted that poor growing conditions should favor maleness (B5). Gametophytes of *Acrostichum danaeifolium* have delayed maturity and a higher proportion of males when grown in soil culture rather than on agar (Lloyd & Gregg 1975). The same is true for *Onoclea sensibilis* (Rubin et al. 1985). Field populations of *Matteuccia struthiopteris* have slower growth rates and more males than laboratory populations (von Aderkas 1983). Growth inhibition by auxins increases the proportion of males in cultures of *Ceratopteris thalictroides* (Hickok & Kiriluk 1984). *Polypodium crassifolium* and *Onoclea sensibilis* gametophytes develop as males when grown in darkness or near darkness, but develop archegonia under equivalent conditions in the light (Schraudolf 1967; Näf et al. 1974; Rubin et al. 1984). Many other studies show an association between slow growth and unisexual

males but these are equivocal because slow growth could be a consequence of antheridium formation.

Some observations are not directly predicted by the model. Occasional males have been observed in single spore culture of *Matteuccia* (von Aderkas 1983) and *Onoclea* (Rubin et al. 1985). The fact that males sometimes develop in the absence of antheridiogen suggests that sperm mobility may sometimes exceed the effective range of antheridiogens. This suggestion is a corollary of the hypothesis that gametophytes should only produce antheridia when they have some chance of male reproductive success. Auxin-induced males of *Ceratopteris thalictroides* release antheridiogen (Hickok & Kiriluk 1984). This observation is difficult to accomodate within the model.

The model I have presented depends on interactions within small local populations but "small" was never defined. The size of the local population will be determined by the density of gametophytes and the distance over which gametophytes interact. Three such distances can be identified: the range of sperm movement; the effective range of antheridiogens; and the distance over which competition between young sporophytes is significant. I do not know if these distances are similar or greatly dissimilar, but such information is important for the model. Moreover, I do not know how often species with cross-reacting antheridiogens are found in mixed populations. Schraudolf (1985b) reported that single, female gametophytes of *Anemia phyllitidis* grown on natural substrates were surrounded by a hemisphere of high antheridiogen activity with a 10 cm radius.

There have been few studies of gametophyte populations in the field. Cousens (1981) reported mixed species populations of gametophytes at densities up to 4.5 cm^{-2} , but also found isolated gametophytes bearing sporophytes. Schneller (1988) described high densities of dormant spores in the soil under established sporophyte populations, but much lower densities at nearby sites away from established sporophytes.

B. Sex expression in *Equisetum*

Sphenopsids of the genus *Equisetum* are homosporous, but apparently lack an antheridiogen system. Gametophytes are initially male or female, and both sexes occur at significant frequencies in single-spore culture (Duckett 1970, 1972, 1979). Moreover, reused medium does not alter the sex ratio in multispore culture (Hauke 1977). Sex expression appears to be determined by environmental factors rather than by intrinsic differences among spores (Duckett 1977).

In culture, most male gametophytes remain male throughout life, but many archegoniate gametophytes later convert to producing antheridia (Duckett 1970, 1972). This change to male sexual expression can be delayed by transfer to fresh medium, so the effect is thought to be due to nutrient depletion or a build up of metabolites (Duckett 1977). Such effects are likely to be less important in the field and bisexual gametophytes are, in fact, less common in wild populations (Duckett & Duckett 1980). Poor growing conditions appear to favor male sexual expression. The proportion of female gametophytes is higher on media allowing faster growth rates (Duckett 1972; Hauke 1977) and unfavorable conditions result in a preponderance of males in wild populations (Duckett & Duckett 1980).

The major difference between *Equisetum* and *Pteridium* with respect to sex determination is that *Equisetum* gametophytes lack information about the presence and condition of conspecific neighbors. Provided that acquiring information has minimal costs, additional information should allow a more adaptive response. Why then has *Equisetum* not evolved an antheridiogen system? Two kinds of explanation suggest themselves: (1) Antheridiogens would be adaptive but they have not evolved because of lack of mutations, lack of appropriate precursors etc, or because gametophytes grow in environments with too much flowing water for communication by chemical to be effective. (2) Antheridiogens could have evolved but the information provided would not be useful. This could happen if successful gametophytes almost always occur in large gametophyte populations, so that the presence of male and female conspecifics is a near certainty, and an antheridiogen would provide little additional information. Alternatively the range of

sperm movement could be much greater than the effective range of any possible antheridiogen, so that a gametophyte's reproductive success would often be determined by other gametophytes about which the gametophyte could not obtain information.

Bilderback et al. (1973) measured *Equisetum* sperm moving at $300 \mu\text{ms}^{-1}$ upon release from their antheridium. Sperm could keep swimming for 2 hrs in distilled water. If the initial speed was maintained for this period in a straight line, sperm could swim over 2 m. Little is known about the effective ranges of antheridiogens under field conditions, but it seems conceivable they would be much smaller than this. Duckett & Duckett (1980) studied large natural populations of *Equisetum* gametophytes on previously inundated mud along the margins of lakes and reservoirs. Gametophytes were recorded at densities ranging from 500 to $200,000 \text{ m}^{-2}$ (when present at all). Suppose, for the sake of argument, that local mating populations cover half a square meter, then these densities would be consistent with populations always being large enough for potential mates of both sexes to be available. Mud flat environments may also be subject to mass flow of water which might reduce the effectiveness of antheridiogens and increase the range of sperm movement. Thus, there is at present insufficient evidence to decide whether the lack of antheridiogens is more likely to be due to primitive absence or to features of *Equisetum*'s habitat.

The natural history of pteridophyte gametophytes is largely unknown. Further field information on gametophyte densities and ranges travelled by antheridiogens and sperm will be of great interest, particularly information that will allow comparisons between species with and without antheridiogens.

IV. SIMPLE POLYEMBRYONY

Females were predicted to bear only one sporophyte (B4) and this prediction is, in general, supported by observation. However, females usually produce several archegonia and this poses two problems:- (1) Why are gametophytes with several sporophytes rare? and (2) If gametophytes only provision one sporophyte, why produce more than one archegonium? Miller (1968) suggested that "polyembryony is prevented only by the failure of spermatozoids

to reach several archegonia." This is a possible answer to the first question but is unsatisfactory for the second. Sperm are chemically attracted to archegonia (Miller 1968) and, if fertilizations are limiting, one archegonium should be as good as several. Moreover, Miller's hypothesis is contradicted by evidence that multiple fertilizations are common. Buchholz (1922) collected many reports of gametophytes with more than one zygote or embryo (simple polyembryony) and inferred that most of the extra embryos died during very early development. He proposed that simple polyembryony was a mechanism of "developmental selection" whereby gametophytes only commit resources to embryos of superior vigor.

Buchholz's long-neglected hypothesis suggests a possible reinterpretation of recent studies designed to measure genetic load in fern populations. Gametophytes of homosporous ferns are often hermaphroditic. A gametophyte that fertilizes one of its own eggs ("intragametophytic selfing") will produce a sporophyte that is homozygous at all loci, because all gametes produced by a gametophyte are genetically identical (Klekowski 1979). Several studies have used this property of homosporous life cycles to estimate the genetic load of fern populations. In such studies, gametophytes are isolated and allowed to self-fertilize. Those gametophytes which fail to produce a viable sporophyte are assumed to carry one or more recessive lethals (for references see Lloyd 1974b; Hedrick 1987). I will call this the "genetic load" interpretation of embryo death. The alternative "developmental selection" interpretation is more inclusive because recessive lethals are still a possible cause of death, but the gametophyte can also play a role in an embryo's death.

My purpose in the remainder of this section is to argue that current ideas on genetic load in homosporous ferns are too simple, and that some phenomena are better explained by Buchholz's (1922) concept of developmental selection. (I will also introduce the term "gibs" to describe gametophytes with the same parent. This is by analogy to "sibs", which are sporophytes that share a parent.)

Three sorts of evidence seem to contradict the genetic load interpretation.

(a) Spores collected from a single sporophyte often produce self-sterile gametophytes at frequencies significantly less than 50% (Klekowski 1970b; Ganders 1972; Masuyama 1986). If a sporophyte is heterozygous for a single lethal, half of its spores should produce self-sterile gametophytes. If the sporophyte is heterozygous for lethals at k unlinked loci, $[1 - (1/2)^k]$ of its spores should be self-sterile. Given linkage, any proportion between 50% and 100% is possible. However, no combination of single-locus lethals can give fewer than 50% self-sterile gametophytes.

Two hypotheses have attempted to explain how some families of gametophytes can contain fewer than 50% self-sterile gametophytes. Klekowski (1970b, 1976) proposed that lethal alleles might be inherited at polyploid loci. Electrophoretic evidence now makes a polyploid explanation unlikely (Haufler 1987; Soltis & Soltis 1987). Ganders (1972) proposed that embryo death might be explained by homozygosity at two or more diploid loci, which were individually non-lethal. Put another way, variation in background genotype might result in incomplete penetrance of lethal alleles.

(b) Isolated gametophytes sometimes produce a viable sporophyte after previous self-fertilizations have resulted in embryo death ("leaky lethality" e.g. Klekowski 1970b; 1972; Ganders 1972; Masuyama 1986).

"Leaky lethality" can also be ascribed to incomplete penetrance, but its cause cannot be variation in background genotype because the genotypes of successful and unsuccessful zygotes are identical. Klekowski (1970b) speculated that zygotes might vary in their egg cytoplasm or position on the gametophyte thallus.

(c) Hedrick (1987) obtained two estimates of the mean number of lethals per zygote for each of five species of ferns. If X is the proportion of self-sterile gametophytes, and Y is the proportion of sterile matings between gib pairs, Hedrick

proposed that the mean number of lethals per zygote x could be estimated by

$$x = -2 \ln(1 - X) \quad (3.1)$$

and by

$$x = -4 \ln(1 - Y) \quad (3.2)$$

For two of the five comparisons there was an almost three-fold difference between estimates. Hedrick (1987) offered no explanation for these differences. (Hedrick's use of the Poisson distribution assumes gametic equilibrium, but this can only be an approximation because selection produces disequilibrium among lethal alleles. However, I doubt that this explains the discrepancy.)

Developmental selection is capable of explaining (a,b,c) above, and is evolutionarily plausible. For a gametophyte, success as a female parent depends on its ability to support a sporophyte to nutritional independence, and on the sporophyte's probability of reproduction after independence. The sporophyte's survival will often depend on success in competition with neighbouring plants. For this reason, a gametophyte should only provision one sporophyte because two sporophytes growing side-by-side would be in direct competition for resources from the gametophyte and from the environment. If a gametophyte always provisions the first viable sporophyte, success as a female parent depends on the quality of this sporophyte. However, this is not the case if a gametophyte can actively abort embryos. The fewer resources committed to an aborted embryo relative to the total cost of supporting a sporophyte, the greater should be the advantages of selective abortion.

Four factors determine whether a gametophyte can achieve higher expected fitness by aborting or by provisioning an embryo. They are: (1) the expected fitness of the embryo if provisioned; (2) the probability that a future embryo will have higher expected fitness; (3) the costs of abortion; and (4) the risks of delaying reproduction. Suppose that gametophytes can assess some measure of embryo vigor and that this measure is correlated with the embryo's expected fitness if provisioned. Then a

gametophyte's fitness will be maximized if embryos of low vigor are aborted but embryos of high vigor are provisioned. Such a relationship is illustrated in Figure 3.1. In the figure, the fate of embryos of intermediate vigor is probabilistic, but this is not an essential feature of the model. A step-function with a threshold above which all embryos are provisioned and below which all embryos are aborted is also possible.

A heterozygous sporophyte could segregate alleles at several loci affecting embryo vigor, and the proportion of self-sterile gametophytes would depend on the frequency of homozygous embryos that fall below some critical vigor. Clearly, such a model is compatible with any proportion of self-sterile gametophytes among the progeny of a single sporophyte [cf (a)]. Similarly, there should be no simple relationship between the proportion of self-sterile gametophytes and the proportion of sterile gib-matings [cf (c)].

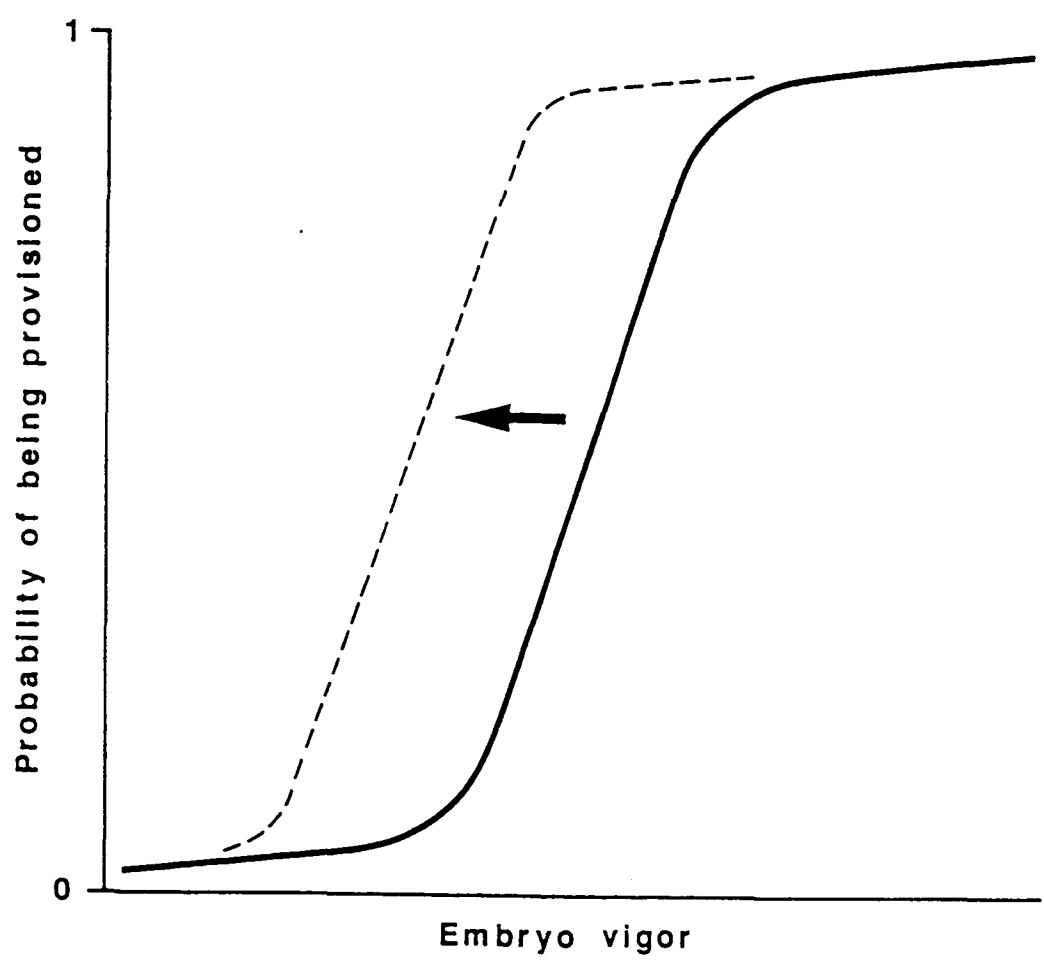
The developmental selection hypothesis can explain "leaky lethals" [cf (b)] in one of two ways. (1) The relationship between vigor and abortion could be probabilistic. (2) The relationship could be deterministic, but with the abortion threshold changing between fertilizations (Figure 3.1). A changing threshold might be adaptive. As time passes, gametophytes gain information about local sources of sperm. Once one sperm is received, the receipt of identical sperm from the same source is likely. Therefore, the abortion of an embryo is not an irrevocable rejection of that embryo's genotype. Gametophytes might be expected to become less selective as the number of embryos that fail to reach a given standard increases.

In summary, the developmental selection hypothesis proposes that gametophytes are sometimes self-sterile because selfed embryos have low vigor and are consequently aborted. This mechanism is similar to that proposed by Seavey & Bawa (1986) for post-zygotic self-incompatibility in angiosperms.

V. CONCLUSIONS

The microsite where a spore lands has a profound influence on whether the resulting gametophyte will have a higher expected fitness as a male or as a female. Such circumstances favor

Figure 3.1. Proposed relationship between embryo vigor and whether the embryo is provisioned (solid curve). Embryos of low vigor are aborted. "Leaky lethality" (see text) could be explained by the indeterminate fate of embryos with intermediate vigor, or because the position of the function shifts for later fertilizations (broken curve).



individuals that determine sex in response to information about the environment (Charnov & Bull 1977). The nature of sex expression should depend on the information available to the gametophyte. *Equisetum* gametophytes can assess environmental quality from their own growth rate but have little information about conspecific neighbors. Slow-growing gametophytes tend to develop as males but sex expression is not responsive to the presence or sex expression of neighboring gametophytes. Ferns that produce antheridiogens possess additional information. Poor environments tend to favor maleness, but this response can be modified by the gametophyte's 'social' environment.

The mechanism of sex determination appears to be very different in homosporous and heterosporous life cycles. In homosporous species, gametophytes determine sex in response to environmental cues. In heterosporous species, sporophytes determine a gametophyte's sex before spore dispersal. However, the underlying mechanisms may be similar, in the following sense. Within any given species, smaller spores should tend to produce smaller gametophytes. If gametophytes with few reproductive resources tend to develop as males, a gametophyte that develops from a smaller-than-average spore should be more likely to develop as a male than should a gametophyte from a larger-than-average spore. In this manner, a sporophyte can influence a gametophyte's sex expression by determining spore size. If homosporous sporophytes suddenly produced spores of two sizes, the spores would immediately differ in sex expression because of sex determining mechanisms already present in gametophytes. In the next chapter I consider when it is in a sporophyte's interests to produce spores of two distinct size classes.

Appendix

**Mutation/selection balance and gametophytic selfing
in homosporous ferns**

Over the past decade there has been increasing interest in the mating systems of homosporous ferns. All gametes produced by a gametophyte are genetically identical. Therefore, if a gametophyte fertilizes one of its own eggs ("gametophytic selfing"), the resulting sporophyte will be homozygous at all loci. This unique property of homosporous life cycles is commonly believed to allow the efficient detection of recessive deleterious alleles (genetic load). Gametophytes are isolated and allowed to self-fertilize. Those gametophytes which fail to produce a sporophyte are assumed to carry a recessive sporophytic lethal. Thus, the frequency of lethal-free sporophytes imposes an upper limit on the incidence of gametophytic selfing in the field because a sporophyte that carries a recessive lethal cannot be homozygous at all loci (for references to genetic load studies see Lloyd 1974b, Klekowski 1984).

A second significant feature of fern mating systems is simple polyembryony: the fertilization of more than one egg on the same gametophyte. Though several eggs may be fertilized, usually only one embryo is matured. Klekowski (1982) considered the effect of simple polyembryony on the equilibrium frequency of recessive lethal alleles in a randomly mating fern population. My aim in this note is to extend Klekowski's analysis to a fern population with gametophytic selfing. After I completed an earlier version of this note, Hedrick (1987) published a paper that addressed the equilibrium frequency of recessive deleterious alleles given a mating system with some gametophytic selfing. Hedrick did not consider simple polyembryony nor sporophytic selfing (see below). In my model, I consider these factors but only for the restricted case of recessive lethal alleles.

MUTATION SELECTION BALANCE AT A SINGLE LOCUS

Let p be the frequency of the wildtype allele A_1 and $(1 - p) = q$ be the frequency of the recessive lethal A_2 . The recessive lethal can only be present in heterozygous sporophytes, and its frequency must be half the frequency of heterozygotes. Therefore, two sporophyte genotypes are possible, A_1A_1 and A_1A_2 , with frequencies $(1 - 2q)$ and $2q$ respectively. Matings are assumed to be of three types: (i) outcrossing (mating with an unrelated gametophyte); (ii) sporophytic selfing (mating with a gametophyte derived from the same parent sporophyte); and (iii) gametophytic selfing. The probability of a given type of mating is assumed to be the same for A_1 and A_2 gametophytes. Allele frequencies after mating and selection will be represented by p' and q' .

Outcrossing

Fern mating systems are complicated by simple polyembryony: the fertilization of several archegonia on the one gametophyte. Despite multiple fertilizations, usually only one embryo completes development. Klekowski (1982) studied the effects of simple polyembryony on the frequency of recessive lethals in a randomly mating population. He assumed that all matings with A_1 gametophytes are successful but A_2 "female" gametophytes remain in the breeding pool until they are fertilized by an A_1 sperm.

	A_1A_1	A_1A_2	A_2A_2
Frequency of progeny genotypes	p^2	$2pq + q^2$	-

For out-crossing, the new gene frequency after selection is

$$q' = (2pq + q^2)/2$$

and the change in gene frequency due to selection is

$$\delta q = q' - q = -q^2/2$$

(after Klekowski 1982).

These calculations assume that there is no cost of unsuccessful A_2A_2 matings and that an A_2 gametophyte can always be fertilized by an A_1 sperm. Thus, the only natural selection is the reduced success of A_2 sperm. I will make these assumptions though they may underestimate selection against A_2 .

Sporophytic Selfing

If sporophytic selfing involves gametophytes from an A_1A_1 parent, all offspring will be A_1A_1 . However, if the gametophytes have an A_1A_2 parent, one quarter of offspring will be A_1A_1 and the remainder A_1A_2 (because of simple polyembryony). The calculations for a heterozygous parent use Klekowski's assumption that there is no cost of unsuccessful A_2A_2 matings and that an A_2 gametophyte can always be fertilized by an A_1 sperm.

Genotype of parent	Genotype frequencies among progeny		
	A_1A_1	A_1A_2	A_2A_2
A_1A_1	$1 - 2q$	—	—
A_1A_2	$q/2$	$3q/2$	—

For sporophytic selfing, the new gene frequency after selection is

$$q' = 3q/4$$

and the change in gene frequency due to selection is

$$\delta q = -q/4.$$

Gametophytic Selfing

Under gametophytic selfing, A_1 gametophytes produce viable sporophytes, but A_2 gametophytes produce sporophytes that fail to survive.

Genotype frequencies among progeny		
A_1A_1	A_1A_2	A_2A_2
p	$-$	q (die)

For gametophytic selfing, the new gene frequency after selection is

$$q' = 0$$

and the change in gene frequency due to selection is

$$\delta q = -q.$$

Mixed-mating Population

If the probabilities of the three types of mating are r , s , t , then after mating and natural selection:

$$q' = [r(2q - q^2)/2 + 3sq/4]/(1 - tq)$$

And:

$$\begin{aligned}\delta q_s &= [-rq^2/2 - sq/4 - tq + tq^2]/(1 - tq) \\ &= [-rq^2/2 - sq/4 - tpq]/(1 - tq)\end{aligned}$$

If $q \ll 1$,

$$\delta q_s \approx -rq^2 - sq/4 - tq.$$

The three terms give the contributions to selection against A_2 of the three kinds of mating. Unless s and t are both small (of similar magnitude to q) the contribution of outcrossing will be insignificant, and

$$\delta q_s \approx -sq/4 - tq.$$

Mutation/Selection Balance

If u is the mutation rate from $A_1 \rightarrow A_2$ and v is the reverse mutation rate from $A_2 \rightarrow A_1$, then the change in gene frequency due to mutation (after selection) is:

$$\delta q_m = up' - vq'$$

$$\approx up', \quad \text{because } up' \gg vq'.$$

$$\approx u, \quad \text{because } p' \approx 1.$$

At equilibrium δq_s will be equal and opposite to δq_m .

$$u \approx s\hat{q}/4 + t\hat{q}$$

$$\hat{q} \approx 4u/(s + 4t)$$

where $\hat{q}$ is the equilibrium frequency of the lethal allele. The approximation of $\hat{q}$ is invalid for s and t both very small.

The equilibrium frequency of a lethal allele under random mating is $\hat{q} = \sqrt{u}$ without simple polyembryony (Crow & Kimura 1970) and $\hat{q} = \sqrt{2u}$ with simple polyembryony (Klekowski 1982). Single-locus mutation rates are generally believed to be of the order of 10^{-5} or 10^{-6} per generation (Falconer 1981). Because u is small, $4u/(s + 4t)$ will be much smaller than $\sqrt{u}$ or $\sqrt{2u}$ for all but negligible rates of selfing.

Inbreeding reduces the equilibrium frequency of deleterious recessives because more alleles are exposed to natural selection in homozygotes. If gametophytic selfing is at all common, the frequency of recessive lethals should be much lower than in a randomly mating population. The frequency at a single locus should be of similar magnitude to the mutation rate at the locus. These results are not easily extended to the equilibrium frequency of self-sterile gametophytes because such gametophytes could carry a recessive lethal at any of an unknown number of loci. However, as stated previously, the incidence of successful gametophytic selfing cannot exceed the frequency of lethal-free sporophytes.

Hedrick (1987) and I obtained similar results for the frequency of recessive lethal alleles under gametophytic selfing. His equation (5b) gives $\hat{q} \approx u/t$ for a recessive lethal with no sporophytic selfing. This is directly equivalent to $\hat{q} \approx 4u/(s + 4t)$ when $s = 0$. Neither of our models allows mutation rates to be

estimated from the proportion of self-sterile gametophytes because the number of potentially lethal loci is unknown and "genetic load" is not a satisfactory explanation of all embryo deaths (see Section IV, this chapter).

Chapter 4

The origin of heterospory

The life cycles of heterosporous and homosporous pteridophytes differ in three main respects: (1) heterosporous pteridophytes produce spores of two distinct size classes, homosporous pteridophytes produce spores of a single size class; (2) gametophytes of heterosporous species are either male or female, depending on spore size, but gametophytes of homosporous species are potentially hermaphroditic; (3) gametophyte development is endosporic in heterosporous pteridophytes but exosporic in homosporous pteridophytes.

This chapter presents a model that explains the association of these three characters. In the model, an initially homosporous population is subject to natural selection for increased spore size. Larger spores benefit female reproduction to a greater extent than male reproduction because the minimum costs of male reproduction are less than the minimum costs of female reproduction. Above some critical spore size, the population is invaded by sporophytes producing smaller spores which reproduce predominantly as males. Under the model, the evolution of heterospory would have had three phases: (1) a gradual increase of spore size in a homosporous population; (2) the sudden introduction of smaller microspores; and (3) the subsequent divergence in size and specialization of the two spore types. The model explains haploid dioecy as a consequence of pre-existing mechanisms of sex determination, and endosporic development as a consequence of an increased dependence on spore food reserves for reproduction. An important corollary of the model is that homosporous life cycles are adaptively superior to heterosporous life cycles when propagule size lies below some threshold but adaptively inferior when propagule size lies above this threshold.

I. INTRODUCTION

Pteridophytes can be classified as homosporous or heterosporous. Homosporous species produce spores with a unimodal size distribution whereas heterosporous species produce spores of two distinct size classes. Most extant pteridophytes are homosporous. Heterospory is restricted to the orders Isoetales, Selaginellales, Marsileales and Salviniiales and to the monotypic genus *Platyzoma*. Other heterosporous groups are known only from the fossil record. Heterosporous forms are believed to have evolved from homosporous ancestors and this must have occurred on several independent occasions.

In all species for which I have evidence, heterospory is associated with differences in the sex expression of the two classes of spores. The smaller microspores develop into male gametophytes. Female reproduction is restricted to the larger megaspores. This contrasts with homosporous species in which spores of the same size may develop as male, female or bisexual gametophytes (G. Smith 1955; Sussex 1966; Sporne 1975).

This chapter presents a model for the evolution of heterospory from a homosporous ancestor, based on the reasonable assumption that gamete formation is controlled by genes expressed in gametophytes but spore size is controlled by genes expressed in sporophytes. Therefore, sex expression should evolve so as to maximize the expected fitness of individual gametophytes but spore size should evolve to maximize the expected fitness of individual sporophytes. A model of the origin of heterospory must explain why *gametophytes* from spores of different sizes should have different optimal sex expressions and under which circumstances *sporophytes* benefit from producing spores of two sizes.

For the purposes of the model, fitness will be defined as the number of sporophytes of the next generation to which an individual contributes a genome. The "individual" may be a gametophyte or sporophyte, thus gametophyte and sporophyte fitnesses are defined separately. Female and male fitnesses refer to genomes contributed through eggs and sperm respectively. The female/male distinction can apply to both gametophyte and sporophyte fitness. The distinction between natural selection

which increases gametophyte fitness and that which increases sporophyte fitness is important because recent theoretical studies have shown that the different genotypes of gametophyte and sporophyte define different, sometimes conflicting, evolutionary interests (Westoby & Rice 1982; Queller 1983, 1984; Law & Cannings 1984; Haig 1986).

The evolution of anisogamy has received considerable theoretical treatment (Parker, Baker & Smith 1972; Bell 1978; Maynard Smith 1978; Cox & Sethian 1985). The model presented here is not a simple restatement of these models with spores substituted for gametes. The major difference is that, in anisogamy models, zygote fitness is a function of the combined mass of the fusing gametes whereas, in my model, sporophyte fitness is a function of maternal gametophyte size alone. The paternal contribution to a sporophyte is a single sperm, regardless of the size of the sperm-producing gametophyte.

II. A MODEL FOR THE ORIGIN OF HETEROSPORY

A. Optimal Sex Expression of Homosporous Gametophytes

A homosporous gametophyte can allocate resources to male reproduction, female reproduction or both. The mechanisms which determine this allocation should have evolved so as to maximize the gametophyte's expected fitness given the information available to the gametophyte about itself, its environment, and nearby gametophytes. Optimal sex expression will depend, in part, on the amount of resources available to the gametophyte. This relationship is discussed in the previous chapter and is summarized below.

The minimum effective male investment and the minimum effective female investment are probably determined by the respective minimum costs of producing an antheridium and of supporting a sporophyte to nutritional independence. These costs may change independently during the course of evolution. For example, a deterioration in the conditions for sporophyte establishment may result in an increase in the minimum female investment without changing minimum male costs.

Above the minimum investment for each sex, the more resources committed to male or female reproduction the greater

should be a gametophyte's chances of reproductive success. However, both forms of expenditure are subject to diminishing returns. Male reproductive success is limited by the range of sperm movement. The number of eggs within this range must be finite as must be the number of sporophytes the area is able to support. This places limits on the amount of effective male investment. Female reproductive success is usually limited to a single sporophyte. Beyond some point, additional resources supplied by the gametophyte should have little effect on the sporophyte's chances of survival.

I will make two assumptions about the relative returns from male and female investment: (1) The minimum male investment which has some chance of reproductive success is less than the minimum female investment which has some chance of success; (2) The point of diminishing returns is lower for male than for female investment. As a consequence of these assumptions gametophytes which can make only a small total commitment to reproduction should tend to reproduce as males. The amount of resources stored in a spore can influence total reproductive commitment, both as 'capital' and through the 'compound interest' of growth.

B. Optimal Spore Size of Homosporous Sporophytes

A sporophyte's fitness is the sum of the fitnesses of all gametophytes which develop from its spores. Spore size has two major effects on sporophyte fitness: (1) The amount of resources invested in a spore may affect the gametophyte's chances of reproductive success, and (2) the more resources committed to each spore, the fewer spores can be produced. Therefore, there is a trade-off between spore size and number. A sporophyte should invest that amount in each spore which maximizes the return in gametophyte fitness per unit investment.

The role of spore food reserves may vary among homosporous species. At one extreme, spore reserves could be important for germination and early growth of the gametophyte but most of the resources used in sexual reproduction would come from activities of the gametophyte such as photosynthesis. At the other extreme, reproduction could be dependent on resources supplied by the parental sporophyte and the contribution of the gametophyte's own

activities could be negligible.

A number of factors could result in selection for increased spore size. If conditions are unfavorable for gametophyte growth, larger reserves may be necessary for successful establishment. If favorable conditions are short-lived, larger spores may allow more rapid completion of the gametophyte stage. If competition among gametophytes is intense, small early growth advantages may give large increases in expected fitness.

In this section, I develop a model of optimal spore size in homosporous pteridophytes, which is based on Smith & Fretwell's (1974) treatment of the optimal parental investment in individual offspring (discussed in Chapter 2). Expected gametophyte fitness (W_g) is considered to be a function of spore size (s).

$$W_g = g(s) \quad (4.1)$$

W_g is defined as the expected number of gametes contributed to sporophytes in the next generation by a gametophyte that develops from a spore of size s . Most importantly, this is the expectation that holds *before dispersal of the spore*. Before dispersal, the expected number of (future) successful gametes will be the average of the expectations for a gametophyte in all the environments to which the spore could be dispersed. After dispersal, the spore/gametophyte will find itself in one or other of these environments and its expected fitness will change accordingly. Because it is defined before dispersal, W_g is considered to be a function of spore size (s), but not a function of the specific circumstances in which the individual gametophyte will find itself. "Spore size" is used to describe s but the intended meaning is the cost of the spore to the sporophyte rather than the actual dimensions of the spore. The function $g(s)$ is defined for "mutant" spores of size s in a population producing spores of size s^* . This definition allows W_g to be defined for spore sizes not present in the population and avoids the problem of frequency dependent fitness because the "mutant" spores are assumed to be rare. Finally, note that W_g is a measure of the absolute contribution to the next generation rather than a measure of relative fitness.

Expected sporophyte fitness (W_s) is the sum of the expected gametophyte fitnesses of all spores produced by a sporophyte. If a sporophyte's total reproductive effort is k , then the sporophyte can produce k/s spores of size s .

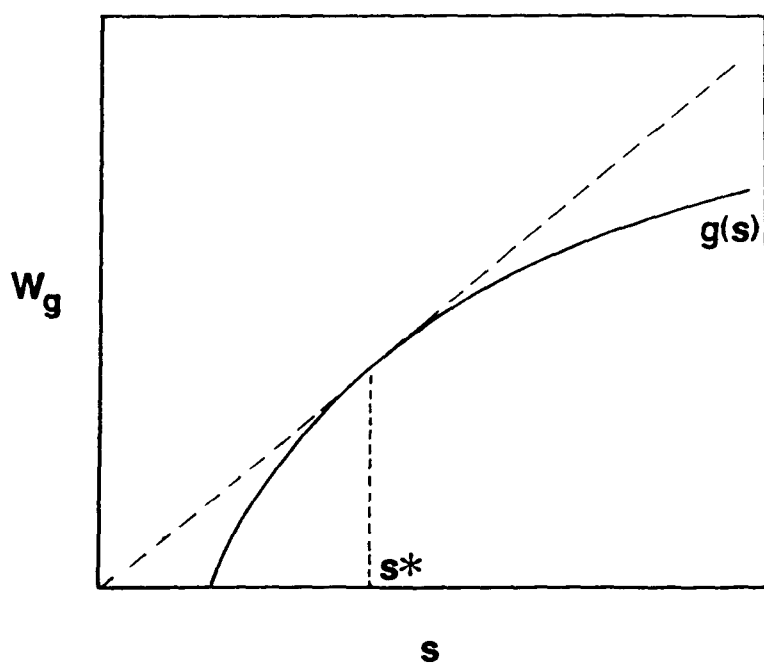
$$W_s = k[g(s)/s] \quad (4.2)$$

On a graph of expected gametophyte fitness against spore size, W_g is given by the height of the curve at the spore size being considered and W_s is proportional to the slope of the line from the origin to this point on the curve. W_s is maximized at a spore size at which the tangent to the curve passes through the origin (Fig. 4.1). The typical spore size s^* should correspond to this optimum and a sporophyte should produce k/s^* spores of size s^* .

Implicit in this model is the assumption that gametophyte fitnesses are independent within broods. This allows optimal spore size to be determined for a single spore considered in isolation and sporophyte fitness to be calculated as the product of spore number and individual spore fitness. In the following discussion, I will use the term "gib" to refer to the relationship between gametophytes produced by the same sporophyte. If spores are widely dispersed and gametophytes rarely interact with gibs, then independence is probably an acceptable assumption. However, if gibs often interact within local populations, the assumption of independence may be invalid. Gibs may interact as mates or as competitors for matings and resources. A 'mutant' spore need not be rare within its local population. The assumption of independence will be retained to simplify the development of my model but the consequences of relaxing this assumption will be discussed at the end of the next section.

At this point I depart from earlier treatments in two ways. First, I develop Smith & Fretwell's (1974) treatment by considering expected gametophyte fitness to be the sum of expected male fitness and expected female fitness even though the spore/gametophyte is not yet committed (at the time of dispersal) to reproduce as a male or as a female. Second, I depart from conventional sex-ratio theory (e.g. Charnov 1982) by not

Figure 4.1. Model of optimal spore size. Expected gametophyte fitness, W_g , is a function of spore size s (i.e. $W_g = g(s)$). The optimal spore size occurs at s^* where $g(s)/s$ is a maximum (after Smith & Fretwell 1974).



separating the probability that an individual will become male or female from the expected number of male or female gametes it will produce. Rather, I use the compound quantities $m(s)$ and $f(s)$, defined as the number of successful male gametes and the number of successful female gametes expected to be produced by a spore of size s , via the gametophyte produced by the spore. It is important to note that the expectations are assessed before spore dispersal, and therefore before sex has been determined and indeed before the gametophyte resulting from the spore has been exposed to any of the various circumstances which will influence its sex expression, with the exception only of spore size.

Given these definitions, expected fitness as a gametophyte is the sum of the expected male and female contributions

$$W_g = m(s) + f(s) \quad (4.3)$$

Expected sporophyte fitness is an aggregate across all spores produced, and can also be expressed as a sum of components due to male and female reproduction

$$W_s = k[m(s)/s + f(s)/s] \quad (4.4)$$

The functions $m(s)$ and $f(s)$ combine (1) the expectation that a spore of size s reproduces as a male or female with (2) the expected number of successful gametes given reproduction as a particular sex. These two expectations are conventionally treated as separate functions; I combine them for two reasons. First, $m(s)$ and $f(s)$ are the biologically relevant functions for assessing how selection should act on a sporophyte's determination of spore size before dispersal. Second, useful assumptions can be made about $m(s)$ and $f(s)$ (see below), whereas it is difficult to assume anything about sex expression, or about expected reproduction as a given sex, independent of the other component. This is because sex expression is determined by gametophytes *after* dispersal, in response to environmental factors which are correlated with expected reproduction as a given sex. These factors could include the sex expression of other gametophytes in the local mating population (Willson 1981).

Since I assume that the environmental sex determination of gametophytes is adaptive, the probability that a gametophyte will reproduce as a particular sex is confounded with its expectation of success as that sex. For example, a small spore that reproduces as a female need not have a small probability of success because small spores might only reproduce as females under circumstances where they have a high expectation of success. By working with the compound functions $m(s)$ and $f(s)$, the model avoids these difficulties and allows the issue of selection for heterospory to be approached.

The functions $m(s)$ and $f(s)$ combine factors intrinsic and extrinsic to the gametophyte. The intrinsic factors are the gametophyte-determined tendencies for spores of a given size to reproduce as a particular sex. These tendencies exist because smaller spores tend to give rise to smaller gametophytes and past natural selection has favored such gametophytes reproducing as males. The extrinsic factors determine the probability of reproductive success given a particular spore size and a particular sex expression.

The spore size/gametophyte fitness function ("fitness function" for short) of a homosporous pteridophyte must satisfy a number of biological constraints, which include: (1) The typical spore size s^* must be sufficient for male and female reproduction. (2) Sporophyte fitness should be at a local maximum at s^* . If not, natural selection should shift typical spore size towards the local maximum. (3) The expected male fitness of a typical spore should equal the expected female fitness, because every sporophyte has one parent of each sex and spores must contribute the same number of genomes through eggs as through sperm.

$$m(s^*) = f(s^*) = g(s^*)/2 \quad (4.5)$$

In the remainder of this section I will consider two possible forms of the fitness function which will illustrate some of the range of possible functions.

Consider a hypothetical species in which spore reserves are a small fraction of the resources necessary for successful

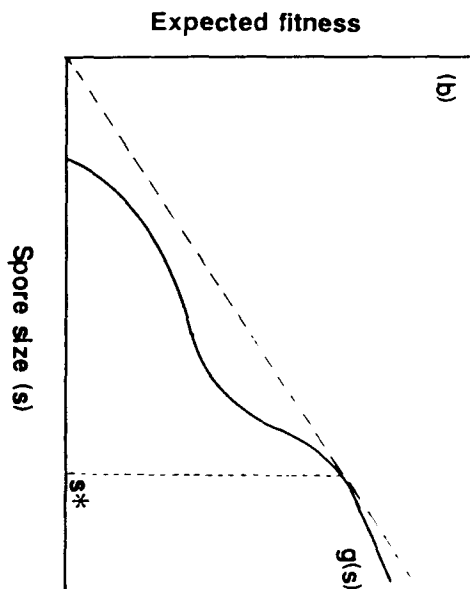
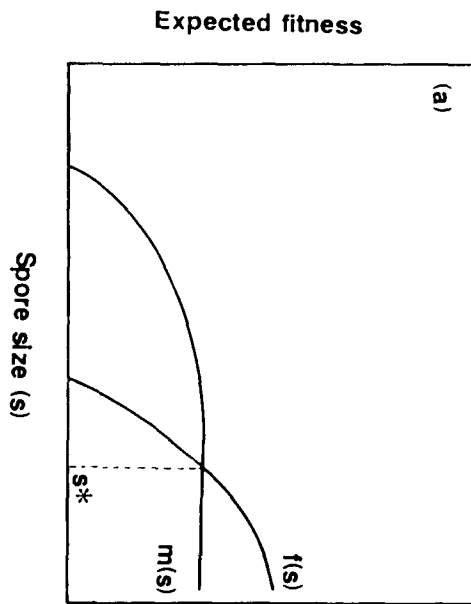
reproduction. Variation in spore size is assumed to influence the probability of a gametophyte becoming established but to have no other effect on reproductive success or sex expression. (By corollary, $m(s)$ and $f(s)$ are assumed to be equal.) Beyond a minimum viable spore size, increases in spore reserves should improve a gametophyte's probability of establishment, though with diminishing returns for spores larger than some optimum (Figure 4.1). This is the kind of relationship between offspring fitness and parental investment that was assumed by Smith & Fretwell (1974).

Now consider a second species in which spore reserves make a substantial contribution to gametophyte reproduction. Variation in spore size is likely to have different effects on expected male and female fitness because a gametophyte has different minimum costs for male and female reproduction. The minimum spore size at which a gametophyte has some chance of reproductive success is likely to be less for male than for female reproduction, and the spore size at which there are diminishing returns in expected male fitness is likely to be less than the equivalent spore size for expected female fitness. Possible forms of $m(s)$ and $f(s)$ are presented in Figure 4.2a and the sum of these functions is presented in Figure 4.2b. In these figures s^* is the spore size at which a sporophyte's overall fitness is a maximum, but the sporophyte would increase its expected male fitness by producing more smaller spores or increase its expected female fitness by producing fewer larger spores. The overall fitness function (Figure 4.2b) has a stepped appearance. This is due to the different minimum spore sizes for male and female reproduction. The first "step" corresponds to spores attaining a size sufficient to reproduce as a male. The second "step" occurs once spore size is sufficient for female as well as male reproduction.

C. The evolution of heterospory

I propose that heterospory arises in large-spored homosporous populations subject to natural selection for increased spore size. I assume that this selection is due to an increase in the minimum spore size necessary for female reproduction with little,

Figure 4.2. A possible relationship between spore size and expected gametophyte fitness. (a) Expected male fitness $m(s)$ and expected female fitness $f(s)$. (b) Expected gametophyte fitness $g(s)$, drawn as the sum of $m(s)$ and $f(s)$. Note that the predicted spore size s^* occurs at a maximum of $g(s)/s$ but at the intersection of $m(s)$ and $f(s)$.



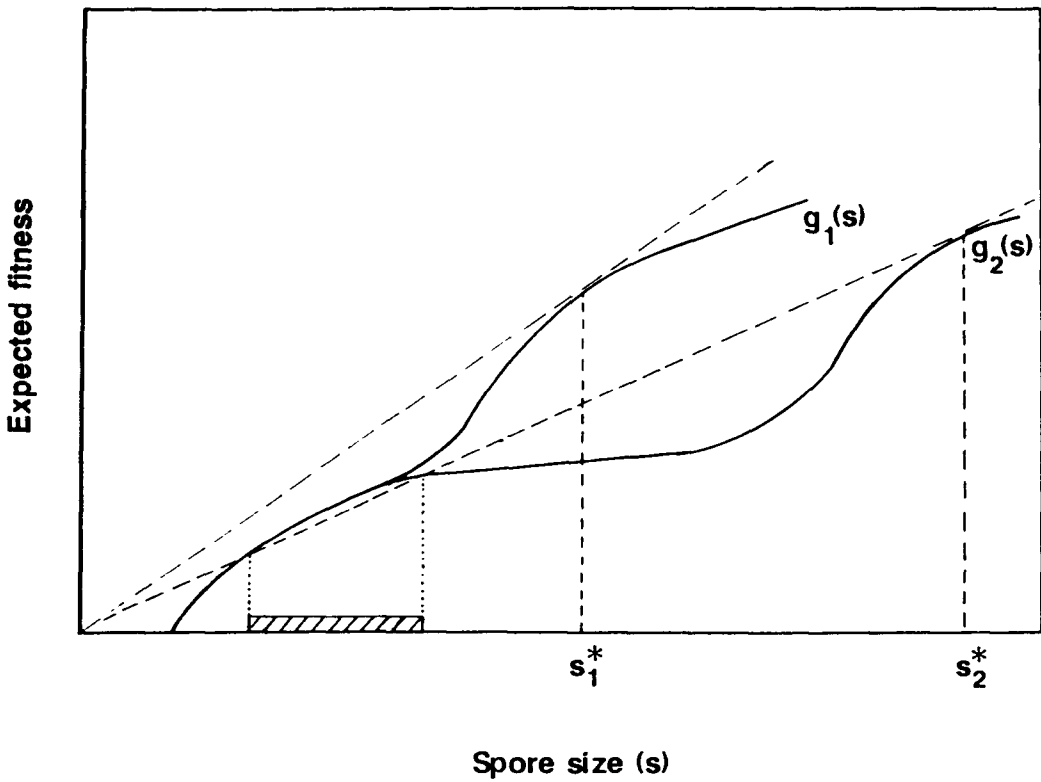
if any, change in the minimum spore size for male reproduction. As typical spore size increases, the 'steps' in the fitness function should move further apart until a point is reached at which a sporophyte can obtain a greater return in gametophyte fitness per unit investment by producing smaller spores (Figure 4.3). These spores would develop into gametophytes which had predominantly male sex expression because of the pre-existing labile sex determination of gametophytes. The larger spores would retain the full range of potential sex expression. The expected fitness of the two types of spores would be frequency-dependent because gametophytes from the smaller spores would have to fertilize eggs from the larger spores. Thus, the smaller spores could invade the population but could not totally replace the larger spores.

So far a gametophyte's expected fitness has been treated as a function of its own spore size but has been assumed to be independent of the size and number of other spores produced by its parent sporophyte ("allocation independence"; Lloyd 1988). At the equilibrium frequency, a sporophyte's return per unit investment should be the same for large and for small spores. The sporophyte would obtain the same return for a given effort by producing all large spores, all small spores or any combination of the two. This result is an artefact of frequency dependence of fitnesses in the global population but allocation independence within broods.

Strictly speaking, the model does not apply if allocation independence is violated. However, I believe the model is still useful for understanding the evolution of heterospory because the conditions which favor the origin of heterospory when allocation independence is assumed should also favor its evolution when this assumption is relaxed. The gametophytes produced by a single sporophyte (gibs) can influence each other's fitness by competing for the same resources, by competing for mates or by mating with each other.

Consider the extreme case of obligate gib-mating. The most efficient use of a sporophyte's resources would be to invest less in some spores which consequently specialize as males and to invest more in other spores which specialize as females. Such a

Figure 4.3. Model for the origin of heterospory. Fitness functions, $g_1(s)$ and $g_2(s)$, are shown for two homosporous populations. The two populations differ in the minimum spore size required for female reproduction. The predicted spore size for $g_1(s)$ is s_1^* . No other spore size gives an equivalent return per unit investment. For $g_2(s)$, the predicted isospore size is s_2^* but a sporophyte could obtain equivalent (or better) return per unit investment by producing smaller spores in the size range marked by the hatched bar.



sexual division of labour might also reduce competition for resources among giblings. The benefits to be gained from heterospory should increase as isospores become larger and the optimum spore sizes for male and female reproduction diverge.

D. Differentiation of microspores and megaspores

Once a class of small gametophytes has arisen, a number of evolutionary adjustments are expected. These gametophytes and the spores from which they develop should undergo specialization for male function because, as a consequence of their size, they have little chance of success as a female. The other class of gametophytes should also undergo change. The gametophytes of this class no longer have equal *a priori* expectations of male and female fitness because some fertilizations are achieved by gametophytes from the smaller spores. Therefore, natural selection for increased spore size would be expected because the existing spore size was an equal-weighted compromise between male and female function. This situation has potential for positive feed-back. As the size of large spores increases, their frequency relative to small spores should decrease and a greater proportion of male reproduction should be achieved by the small spores. This in turn could contribute to greater specialization of the large spores for female reproduction and further increases in spore size.

A large spore might be expected to retain the potential for male reproduction but megaspores of most heterosporous pteridophytes reproduce exclusively as females. This suggests that circumstances in which hermaphroditism is an advantage are sufficiently rare that the costs of retaining this potential outweigh the occasional advantage. However, this argument is the result of *a posteriori* reasoning.

Unlike homosporous species, most heterosporous pteridophytes have gametophytes which develop within the spore wall. Endosporic development is only possible when most of the resources necessary for reproduction are contained within the spore. Heterospory is proposed to arise under conditions where sporophyte development is heavily dependent on spore food reserves. Such conditions predispose gametophytes to endosporic development once the

conflict over spore size between male and female reproduction is resolved.

All modern heterosporous taxa are monoecious. That is, microspores and megaspores are produced on the same sporophyte. Monoecy should evolve when sporophytes can obtain greater fitness by producing spores of both types than by specializing on one or the other spore type (Charnov, Maynard Smith & Bull 1976). By definition, this means there is allocation dependence. Spore distributions around a parent sporophyte will be strongly leptokurtic. Therefore, some gametophytes are expected to have giblings as their only neighbours. Such gametophytes would be unable to mate if their giblings were all of the same sex, but matings would be possible if giblings were of both sexes. This is probably an important factor favoring monoecy.

E. Evidence

The model requires that the homosporous ancestors of heterosporous species had a sex-determining mechanism whereby gametophytes with substantial resources tended to reproduce as females and gametophytes with lesser resources tended to reproduce as males. The sex expression of many homosporous ferns is mediated by substances (antheridiogens) which are released by larger, archegoniate gametophytes and induce smaller gametophytes to develop antheridia (see Chapter 2). Such a mechanism is compatible with the model because it would predispose smaller spores to reproduce as males. However, the real significance of such systems is not whether antheridiogens were present in some ancestral group. Rather, the evolution of antheridiogen systems provides strong evidence that large size gives a gametophyte greater benefits in female than in male function (Willson 1981).

A variety of circumstantial evidence suggests that the minimum costs of male reproduction are less than the minimum costs of female reproduction. Some isospores are capable of male reproduction using spore reserves alone. For example, *Polypodium crassifolium* and *Anemia mexicana* can produce antheridia in total darkness (Schraudolf 1967; Nester & Schledlbauer 1982). Further evidence comes from endosporic gametophytes, which have limited photosynthetic activity. Endosporic microspores can be as small

as 24 μm (*Salvinia cucullata*) and endosporic megaspores as small as 180 μm (*Regnellidium diphyllum*: spore sizes from Erdtman & Sorsa 1971).

The model predicts that, above a critical spore size, a population of isospores is vulnerable to invasion by smaller male specialist spores. This prediction is consistent with the lack of overlap between the size ranges of isospores and megaspores in the modern spore flora. The model also predicts that heterospory evolved from homosporous species with large isospores. This prediction is supported by the Devonian spore record. Isospores larger than 200 μm are found in the same strata as the first megaspores but later disappear from the fossil record (see Chapter 1).

The remainder of this section considers spore size and sex expression in *Ceratopteris thalictroides* and *Platyzoma microphyllum*. These species have been chosen to illustrate hypothetical stages in the evolution of heterospory. No implication is intended about their future evolutionary history or their phylogenetic relationships, but there is some evidence to suggest that they may be related. The sporangium of *Platyzoma* has "a close resemblance" to the sporangium of *Ceratopteris* (Tryon 1964) and the two genera have been classified together on the basis of similar chromosome numbers (Lovis 1977).

Ceratopteris thalictroides is a homosporous fern that produces isospores about 100 μm in diameter (Lloyd 1974a). It will serve as a model of a large-spored homosporous species from which heterospory could have evolved. *Platyzoma microphyllum* is a heterosporous fern which produces small spores of mean diameter 91 μm and large spores of mean diameter 175 μm (Tryon 1964). The large spores are among the smallest known megaspores and the ratio of megaspore size to microspore size is less than in other heterosporous species. *Platyzoma* is the only heterosporous species known to have exosporic gametophytes and in several characters appears closer to the ancestral homosporous condition than do other heterosporous species (Tryon 1964; Duckett & Pang 1984). It will serve as a model of an early stage in the development of heterospory.

Ceratopteris spp. are aquatic annuals (Lloyd 1974a) and

their large spores are probably an adaptation for rapid gametophyte development. *C. thalictroides* has the shortest gametophyte generation documented for a homosporous fern. Antheridia may be found six days after spore germination and archegonia after 15 days (Klekowski 1970a). Gametophytes have two distinct morphologies. Cordate gametophytes are hermaphroditic but spathulate gametophytes produce antheridia only. Spathulate gametophytes develop in response to antheridiogens released by cordate gametophytes (Klekowski 1970a; Schledlbauer & Klekowski 1972). Spores have a unimodal size distribution and spore size is an accurate predictor of subsequent gametophyte morphology in multispore cultures. Smaller spores tend to produce spathulate (i.e. male) gametophytes (Schledlbauer 1976). *C. thalictroides*, thus, illustrates how the labile sex determination of homosporous pteridophytes can predispose spores of different sizes to different sex expressions.

The model predicted that microspores were the new spore type which initiated heterospory. Therefore, the first megaspores were predicted to have been essentially unmodified isospores with the full range of potential sex expression. Megaspores of most extant heterosporous pteridophytes reproduce exclusively as females but *Platyzoma microphyllum* provides evidence for this earlier stage. *Platyzoma* sporophytes produce large spores from 16-spored sporangia and small spores from 32-spored sporangia. Gametophytes which develop from its large spores are initially archegoniate but may later develop antheridia (Tryon 1964; Duckett & Pang 1984). The large spores thus resemble the isospores of those homosporous species with protogynous gametophytes (e.g. Klekowski 1969b). *Platyzoma*'s small spores produce gametophytes which reproduce exclusively as males and in this respect resemble the microspores of all other heterosporous species.

F. Discussion

Previous hypotheses for the origin of heterospory have usually been expressed as verbal models without an explicit statement of the selective pressure operating at each step. Sussex (1966) presented the "generally accepted interpretation of heterospory". During the early Devonian there was an evolutionary divergence of

spore size in some homosporous plants which resulted in two kinds of spore being produced in different sporangia on the same sporophyte. Megaspores increased in size while microspores remained approximately the same size as the ancestral isospores. My model differs in that microspores rather than megaspores are considered to be the primary innovation. Further, I envisage the *sudden* introduction of a markedly smaller spore rather than a *gradual* divergence of spore sizes.

Sussex (1966) believed that the evolution of heterospory involved a shift in the timing of the sex determining process from the gametophyte to the sporophyte generation. In my model the different sex expressions of the two spore types are a consequence of the pre-existing adaptive sex determination of homosporous gametophytes. A sporophyte can influence a gametophyte's gender by controlling the amount of nutrients stored in the gametophyte's spore without there being any fundamental change in the underlying mechanism of sex determination.

Sporne (1975) proposed that heterospory evolved in species which already possessed endosporic development. Separate male and female gametophytes evolved to avoid fertilization of a gametophyte's eggs by its own sperm. Natural selection then caused a reduction in the size of those spores destined to form male gametophytes because these gametophytes had no need of large spore reserves to support a sporophyte. My model has much in common with this scenario though I would emphasize that many homosporous species effectively avoid self-fertilization (e.g. Haufler & Soltis 1984).

Charlesworth (1988) has developed a formal genetic model based on the model presented in this chapter (i.e. Haig & Westoby 1988b) and coming to similar conclusions.

Recently, DiMichele, Davis & Olmstead (1989) have proposed that endosporic development preceded the evolution of heterospory. "By the precocious onset of sexuality, gametophytes could reach sexual maturity while still in the early endosporic phases of development". Gametophytic unisexuality was "a position effect of the metabolic microenvironment of developing spores" because endospory placed "developing gametophytes within the

heterogeneous metabolic sphere of the sporophytic parent". I interpret this primarily as a statement about proximal mechanisms rather than ultimate functions. In my model, endosporic development is a consequence of selection for rapid completion of the gametophyte stage, which entails a greater reliance on nutrients stored within the spore. Therefore, our two viewpoints are not incompatible.

My model proposes that heterospory has arisen in homosporous populations growing under conditions unfavorable for gametophyte growth, which therefore favor an increased reliance on spore reserves for reproduction. Large spore reserves would have evolved principally for the requirements of female reproduction and at some stage a point would have been reached at which a sporophyte could obtain as large a return per unit investment by producing smaller "microspores" as by producing the larger isospores. My model has two major strengths. First, spore dimorphism, unisexual gametophytes and endosporic development can all be explained by one simple model. Second, by defining the conditions which favor the evolution of heterospory, the model also defines the conditions under which a homosporous life cycle is adaptively superior. Modern homosporous species should be regarded as being well adapted in their own habitats rather than as evolutionary relicts which have unaccountably failed to evolve heterospory.