Unfurling Fern Biology in the Genomics Age

MICHAEL S. BARKER AND PAUL G. WOLF

Twenty-first century technology is addressing many of the questions posed by 20th-century biology. Although the new approaches, especially those involving genomic data and bioinformatic tools, were first applied to model organisms, they are now stretching across the tree of life. Here, we review some recent revelations in the ferns. We first examine how DNA sequence data have contributed to our understanding of fern phylogeny. We then address evolution of the fern plastid genome, including reports of high levels of RNA editing. Recent studies are also shedding light on the evolution of fern nuclear genomes. Initial analyses of genomic data suggest that despite their very high chromosome numbers homosporous ferns may have experienced relatively few rounds of genome duplication. Genomic data are enabling researchers to examine speciation rates and the mechanisms underlying the formation of new fern species. We also describe genetic tools that have been used to study gene function and development in ferns. Recent findings in fern biology are providing insights that are not only pertinent to this major component of the land flora but can also help to provide an improved evolutionary context for research on flowering plants.

Keywords: expressed sequence tag, polyploidy, speciation, hybridization, development

Genomics and related tools and concepts were initially developed using model organisms, yet their applications are now shifting to the rest of the tree of life—the unexpected results of which will ultimately influence our broader understanding of biology. Here we illustrate such developments by focusing on a plant group that has an extensive fossil history and remains a conspicuous component of the land flora: the ferns. Ferns have several characteristics that distinguish them from the more familiar seed plants, making ferns an ideal system for addressing previously intractable questions. But ferns are also a major clade of land plants, and knowledge of their basic biology and evolutionary history is essential if we are to make appropriate inferences about the seed plants, including the economically important flowering plants. We first describe some chief characteristics of ferns, centering on their life cycles, and place them in an evolutionary context. We then address a series of general themes, many of which are broadly applicable to all organisms, especially plants. Within each theme we present some unanswered questions, both old and new; we review some genomic tools and explain how they have enabled us to go further than ever before in addressing those questions; and we describe some of the current limitations (of both the tools and the information hidden in genomes), the need for new (and more-balanced) data, and the areas where we believe more research is needed. Although our own research is on what some might consider a rather esoteric group of plants, we seek to illustrate that the research implications extend beyond ferns to the evolution of genomes in general, and places research on economically important plants into better evolutionary context.

Which plants are considered ferns? As our knowledge of evolutionary relationships expands, we must adjust how names are applied and introduce new names. Although this can be frustrating for those not familiar with the group in question, it is essential for conveying information accurately. Figure 1 depicts our current understanding of the relationships among the major groups of vascular plants. This includes all the land plants (embryophytes) except for mosses, liverworts, and hornworts. The tree is based on an accumulation of data from many sources and research groups, more details of which will be presented as we focus on relationships within ferns. Here we need to put the ferns in context. Evidence from DNA sequences of several genes, as well as information on genome structure (Raubeson and Jansen 1992), indicates that a major split occurred, probably about 400 million years ago (MYA). This split gave rise to the extant lycophytes and a clade containing the remaining vascular plants (Pryer et al. 2004). The lycophytes include the extant club mosses (Lycopodiaceae) and spike mosses (Selaginellaceae), as well as several extinct lineages. The remaining vascular plant lineage underwent a later split into "monilophytes" and seed plants; the latter comprise the gymnosperms (among them the conifers) and the angiosperms (the species-rich and economically important flowering plants). The monilophytes comprise four extant lineages: leptosporangiate ferns (about 11,000 species), marattioid ferns (including the large king fern),

BioScience 60: 177–185. ISSN 0006-3568, electronic ISSN 1525-3244. doi:10.1525/bio.2010.60.3.4

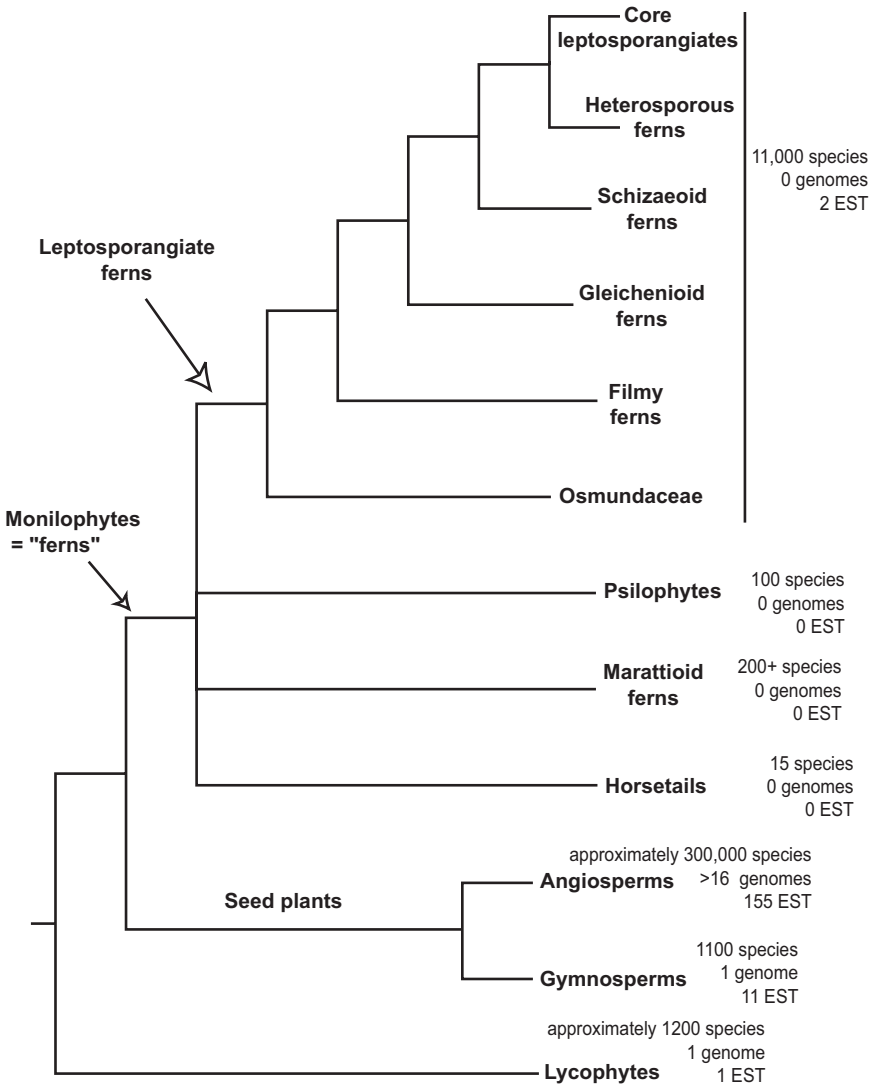

Figure 1. Working phylogenetic framework, based primarily on Pryer and colleagues (2004). The core leptosporangiate ferns represent about 90% of fern species. For each major clade we list the approximate number of species, the number of complete genomes sequenced and available, and the number of species for which expressed sequence tag (EST) data are available. Genome and EST information was obtained from the Plant Genome Database (www.plantgdb.org/). Note that for the seed plants there are 17 genomes and 166 EST collections, whereas for the sister group, the ferns, there are no genomes and only two EST collections.

psilophytes (adder's tongue ferns, moonworts, and whisk ferns), and the horsetails (see figure 1). For the purpose of this review we will use the general term "fern" for all monilophytes. The important concept here is that the ferns are the sister group to all seed plants, thereby providing information needed for comparative studies. For example, any differences between gymnosperms and angiosperms could be a result of changes that evolved in either group. It is only by comparison with the outgroup, ferns, that the direction of evolutionary changes can be inferred and thus studied in an appropriate context.

How do ferns differ from seed plants? All plants have an alternation of generations: Gametes (egg and sperm) are produced by mitosis during the haploid gametophyte stage of the life cycle; fertilization restores diploidy in the zygote; and then the zygote divides mitotically to become the sporophyte, in which meiosis results in haploid spores that germinate and divide to become the next generation of gametophytes. Thus, the plant life cycle differs fundamentally from that of animals, in which gametes are produced by meiosis rather than mitosis. For seed plants, most people are familiar with the sporophyte, the large plant body that we see with the naked eye. Gametophytes of seed plants are nutritionally dependent on the sporophytes (the male gametophytes are the pollen grains). However, in ferns, the spores germinate to become independently growing gametophytes, often large enough to see (if you know where to look) without a microscope. We will discuss the genomic and evolutionary consequences of this type of life cycle later. A second feature that differentiates ferns from seed plants is that most ferns are homosporous. Seed plants and other heterosporous plants produce two kinds of spores: (1) large megaspores that develop into the larger female gametophytes in which eggs or egg cells form, and (2) small microspores that develop into microgametophytes in which sperm cells form. In contrast, homosporous plants produce a single type of spore and gametophyte, although fern gametophytes in nature are usually unisexual because of a pheromonal sex-determination system (Schneller et al. 1990, Hamilton and Lloyd 1991). Heterospory has evolved independently several times, including in the relatively small clade of aquatic ferns within the (otherwise homosporous) leptosporangiate fern lineage (Pryer et al. 2004).

Fern phylogeny

The evolutionary position of ferns relative to other vascular plants is described above. Until fairly recently, several conflicting phylogenetic hypotheses (within ferns) appeared equally likely, but these were made on the basis of intuitive interpretations of morphological change, rather than on

hard data. Part of the problem was the lack of phylogenetically useful characters. Ferns lack complex structures such as seeds and flowers that provide such a wealth of information about angiosperms. In fact, some of the earlier work on fern evolution was done at the "genome" scale by examining chromosome number and meiotic pairing behavior, a method pioneered by the work of Irene Manton (1958). During the 1980s, studies that used variation for restriction sites in chloroplast DNA provided considerable resolution of relationships among closely related species, usually within genera (Gastony et al. 1992, Conant et al. 1994). However, broader-scale relationships were not resolved until a further technical advance in molecular biology. First developed in 1977, DNA sequencing entered the realm of plant systematics by 1990 (Doebley et al. 1990). Initially, the focus was a single gene (*rbcL*) for which primers had been developed, and later, more genes were added to the repertoire. The effect on plant systematics was profound. A series of papers, sequentially adding more taxa and more genes (from the chloroplast and nuclear genomes), resulted in a well-resolved framework of relationships among most major groups of leptosporangiate ferns (Hasebe et al. 1995, Pryer et al. 2004, Schuettpelz and Pryer 2007). Although the fine details of a large phylogenetic tree might be of interest only to fern specialists, the tree itself can have many applications. Below we present several themes concerning genome evolution, gene expression, and development. Such studies should be done in an evolutionary framework, so that comparisons are appropriate. Thus, a robust phylogenetic hypothesis provides the information necessary for choosing taxa for comparative study (Pryer et al. 2002).

Plastid genomes

Within plant cells, various plastids are found; the most important in green plants is the chloroplast, where photosynthesis occurs. Plastids contain their own DNA, although most of the proteins expressed in plastids are nuclear encoded. When variation at the DNA level was first used to infer plant phylogeny, most studies focused on the plastid genome. Unlike nuclear genomes, plastid genomes are relatively conserved in structure and sequence such that comparisons across green plants are feasible, yet there is sufficient variation for evolutionary analysis (Palmer 1987). Early work employed restriction-site variation, and later studies examined nucleotide sequences of plastid genes. Furthermore, variation also exists for overall plastid genome organization. Most embryophyte plastid genomes include a large (15 to 25 kilobase [kb] pairs) inverted repeat. Studies that used probes from tobacco showed that the gene order in the plastid genome of the fern *Adiantum capillus-veneris* was different from that of seed plants, especially in the region of the inverted repeat, where the order in the fern was reversed (Hasebe and Iwatsuki 1992). Details of these differences were revealed by the complete nucleotide sequence of the *A. capillus-veneris* plastid genome (Wolf et al. 2003). It appears that a series of overlapping inversions, each about 20 kb, resulted in the *Adiantum* gene order (Wolf and Roper 2008). Previous attempts to infer these inversion events (Stein et al. 1992) failed because no phylogenetic framework was available for choosing appropriate taxa. Another unusual aspect of the fern chloroplast genome is that many of the protein-coding genes appear to contain stop codons (Wolf et al. 2003). However, sequencing of cDNAs (complementary DNA, derived from the messenger RNA [mRNA]) revealed that the RNA is edited at a minimum of 350 sites across the *Adiantum* plastid genome (Wolf et al. 2004), a rate 10 times higher than any other vascular plant. The extent of RNA editing in ferns, and the function and evolution of this unusual molecular phenomenon, remain a mystery. The phenomenon is poorly studied in many plant groups and not well characterized for nuclear genes. Although plastid genomes are relatively small (about 150 kb) and simple in structure, they have provided a wealth of information in plant biology. Furthermore, developing analytical tools at this scale is a useful stepping stone toward examining the much larger and more complex nuclear genomes.

Nuclear genomes: High chromosome numbers and paleopolyploidy

The roots of plant genomics extend back to the early 20th century when cytologists began studying chromosomes. By squashing and staining actively dividing cells, cytologists were able to observe, under a microscope, a variety of chromosome features, including numbers, sizes, and pairing behavior. These data proved invaluable for making inferences about the nature of plant species, and provoked numerous questions about genome evolution, some of which endure today. Among these long-standing questions is how the high chromosome numbers of homosporous ferns originated and are maintained. By the 1950s it was clear that fern nuclear genomes, particularly those of homosporous species, possessed exceptionally high chromosome numbers relative to other plants (Manton 1950). Homosporous fern genomes contain an average of $n = 57.05$ chromosomes, over threefold more than the flowering plant average of $n = 15.99$ (Klekowski and Baker 1966). However, the heterosporous fern species were found to possess an average of $n = 13.62$ chromosomes, very close to the average of flowering plants—another heterosporous lineage. This striking difference between homosporous and heterosporous plants spawned a number of hypotheses. An influential early hypothesis argued that most homosporous ferns were polyploids (Klekowski and Baker 1966), that is, species with more than two complete sets of chromosomes in their somatic cells. According to this hypothesis, additional non-Mendelian genetic variation could be generated by abnormal pairing during meiosis of different versions of chromosomes, or homoeologs, rather than the normal homologous pairing (Haufler 2002). It was proposed that homosporous ferns evolved this extra source of genetic variation to compensate for what was believed to be their primary mode of reproduction, intragametophytic self-fertilization, an extreme form of inbreeding that results in

100% homozygosity in a single generation (Klekowski 1973). If genetic variation were produced by homoeologous pairing it would not be affected by self-fertilization, and would thus provide a selectively advantageous buffer against the loss of genetic diversity. Although a few early studies supported this idea (Hickok and Klekowski 1974, Hickok 1978, Chapman et al. 1979), subsequent genetic investigations using analyses of isozymes did not support the homoeologous pairing hypothesis. Isozyme studies revealed that homosporous ferns with the lowest chromosome numbers for their genus were genetically diploid with Mendelian inheritance (Gastony and Gottlieb 1985, Haufler and Soltis 1986, Haufler 1987, Wolf et al. 1987). Further, these studies also revealed that many homosporous ferns were predominantly outcrossing, and therefore did not suffer from extreme inbreeding (Haufler and Soltis 1984, Gastony and Gottlieb 1985, Holsinger 1987). Thus, the fundamental rationale for the homoeologous pairing hypothesis was rejected.

To explain the paradoxical combination of high chromosome numbers and diploid gene expression, Haufler (1987) suggested that ferns experienced multiple rounds of ancient polyploid speciation followed by gene silencing, but not chromosome loss. Ancient polyploidy, or paleopolyploidy, is now recognized as a major force in the evolution of flowering plants (Blanc and Wolfe 2004, Cui et al. 2006, Barker et al. 2008, Tang et al. 2008), but its full role in fern genome evolution remains unresolved. Over the last 20 years, a few studies have supported a paleopolyploid origin for the high chromosome numbers of homosporous ferns. Consistent with paleopolyploidy in the history of homosporous ferns, silenced copies of multiple nuclear genes have been identified in genetically diploid homosporous fern genomes (Pichersky et al. 1990, McGrath et al. 1994, McGrath and Hickok 1999). Furthermore, the active process of gene silencing without chromosome loss in a polyploid genome has been demonstrated in a fern (Gastony 1991). However, the first genetic linkage map for a fern, the diploid homosporous fern *Ceratopteris richardii* ($n = 39$), failed to identify remnants of duplicated chromosomes (Nakazato et al. 2006), although most loci were duplicated, and a faint signal of synteny (similar gene order) was detected among duplicated chromosomal segments.

To gain further insight into the origin of the high chromosome numbers of homosporous ferns, researchers have recently begun applying modern genomic tools. One genomic tool, expressed sequence tags (ESTs), has been particularly useful in ferns. In large and putatively complex genomes, such as those of ferns, whole-genome sequencing so far has not been economically feasible. An alternative to sequencing a whole nuclear genome is to sequence a large portion of the transcriptome, that part of the genome that is transcribed to RNA. By extracting mRNA from a plant and reverse transcribing it into cDNA, we can sequence many of the expressed genes. Such sequences have been referred to as ESTs when the cDNAs are sequenced from one end to provide a sequence "tag" for a particular expressed gene. However, the term "EST" is evolving, and now often refers to any sequenced cDNA data, regardless of how, or why, they were collected and assembled. Bioinformatic tools are then used to cluster overlapping EST sequences into a condensed set of contiguous sequences—contigs—and to identify unique, nonoverlapping sequences or singletons (figure 2). These two sets of sequences, contigs and singletons, are pooled into a collection referred to as unigenes, or "unique genes." This EST sequencing approach is particularly well suited for studying the genomics of nonmodel organisms because ESTs

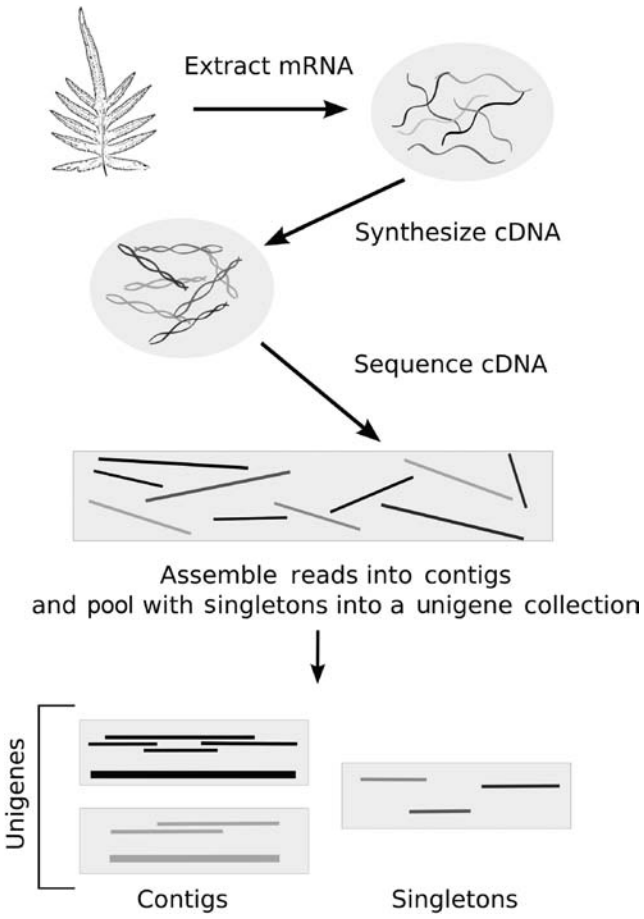

Figure 2. Expressed sequence tag libraries provide a sampling of an organism's transcriptome, or the expressed fraction of the genome. Libraries are constructed by extracting mRNA from a focal organism and synthesizing cDNA. Before sequencing, the cDNA is frequently normalized to reduce the frequency of highly expressed transcripts so that a greater diversity of genes, especially copies that have relatively low expression, are sequenced. Further, the cDNA is often randomly fragmented to facilitate shotgun sequencing of the transcriptome. Bioinformatic tools are used to assemble the fragmented cDNA reads into contiguous sequences (contigs) that represent a particular transcript. These contigs, plus all of the remaining singleton sequences that did not assemble, are pooled together into a collection of unique genes, or unigenes.

provide a broad sampling of an organism's transcriptome regardless of genome complexity. Further, ESTs are relatively low cost, and with new ("next-generation") sequencing technologies, the cost of de novo EST sequencing has steadily declined. This reduction in the cost of DNA sequencing is reflected in the growing number of ESTs available on GenBank, which currently stands at more than 64 million entries (January 2010). Of these 64 million sequences, ferns are represented by slightly more than 15,000 ESTs on GenBank. Considering the low cost and analytic flexibility of ESTs, they will most likely play a significant role in furthering our understanding of fern genomes.

How can ESTs be used to evaluate the paleopolyploid hypothesis and the origin of high chromosome numbers in ferns? First, gene family phylogenies can be constructed (based on sequence similarity) from an EST collection for a species, and gene duplication events identified. By plotting the ages of all gene duplications, typically in terms of their number of synonymous substitutions (K_s or dS), ancient genome duplications may be inferred as peaks in the histogram (figure 3). These peaks reflect the very large number of duplications of similar age that one expects to result from ancient polyploidy, against a backdrop of small-scale duplications occurring continually. By using relative rate corrections to account for variation in substitution rates across lineages, it is possible to discern whether paleopolyploidizations observed in two or more taxa are shared (Barker et al. 2008). Using this approach to analyze EST collections from two polypod ferns (*C. richardii* and *A. capillus-veneris*) has revealed that each lineage shares only a single detectable genome duplication with a median peak at approximately 1.6 K_s (figure 3; Barker 2009). By combining fossil dates (Schneider et al. 2004) with nuclear gene phylogenies from EST data, this duplication event is thought to have occurred nearly 180 MYA. This places the paleopolyploidization somewhere along the branch leading to all or most of the polypod ferns, the largest extant clade of homosporous ferns.

Although we have evidence for one ancient genome duplication in the ancestry of most extant ferns, it is not clear if such a low frequency of paleopolyploidy is sufficient to create and maintain the extraordinary chromosome numbers of homosporous ferns. Differences in the rate of chromosomal change and loss may also play a significant role. Following duplication, the chromosomes of a polyploid genome may form complexes of three or more chromosomes (multivalents) during meiosis. A distinguishing trait of diploidy is the formation of pairs (*di*-ploidy) of chromosomes (bivalents) during meioses. Although many plants have experienced at least one round of ancient genome duplication, these species all behave as genetic diploids rather than polyploids. So how do most plant genomes, which have experienced rounds of ancient genome duplications, return to this diploid genetic state? Subsequent to genome duplication, plant nuclear genomes undergo a series of changes that restore the diploid genetic system, a process known as diploidization. Mechanisms such as chromosomal fusion, illegitimate recombination, and transposition cause rearrangements and often a net loss of genetic material in plant genomes at varying rates. Combined with silencing of duplicated genes through mutation or outright loss, the actions of this suite of genomic changes lead to diploidization (Doyle et al. 2008). Significant variation in the rate of these genomic changes is well known from the numerous whole-genome sequences available for the angiosperms. At one extreme is *Arabidopsis*, whose small nuclear genome contains only five chromosomes but has experienced at least three rounds of whole-genome duplication in less than 200 MY (Bowers et al. 2003, Tang et al. 2008; Barker et al. 2009). In contrast, *Vitis* appears to have only one duplication (the oldest one in *Arabidopsis*), yet it has a larger genome, with 17 chromosomes (Jaillon et al. 2007). Thus, genome structure (and chromosome number) is a product of variation in diploidization processes as well as genome duplication. In the EST gene duplication plots (see figure 3), genic diploidization is apparent as a

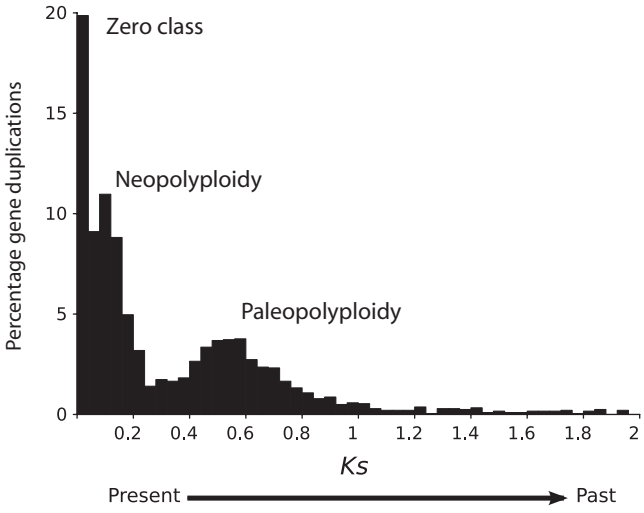

Figure 3. An example age distribution of gene duplications for a plant genome. The x-axis is the number of synonymous substitutions per site among duplicate genes, whereas the y-axis is the percentage of gene duplications in a particular K_s bin. As genes duplicate and diverge, substitutions accumulate over time and can be used to place events in phylogenetic order and time. Because of ongoing gene duplications, such as tandem duplications, most duplications in the genome are young and placed in the zero class. Most of these small-scale duplicates are quickly silenced, yielding an exponential distribution. However, many duplicates are retained from whole-genome duplications and these create peaks in the age distribution. Recent neopolyploidizations are represented by taller and tighter peaks than older paleopolyploidizations because of gene loss and error in estimating the K_s increases over time for a particular gene duplication. These processes cause whole-genome duplications to appear as shorter and broader peaks as the age of the duplication increases.

shrinking of the older peaks as duplicated genes are lost over time. Analyses of angiosperm genomes suggest that many angiosperm lineages have experienced two to three rounds of paleopolyploidy in a time frame of $K_s < 2$ (Cui et al. 2006, Barker et al. 2008). Considering that only a single duplication event has been observed in current analyses of fern genomes, they may experience a lower rate of whole-genome duplication than angiosperms.

So, why do homosporous ferns have so many more chromosomes than angiosperms? One possibility is that homosporous fern genomes may be less dynamic than angiosperm genomes, and may experience chromosomal loss at a much slower rate. Consistent with this perspective is the exceptionally low density of genes observed in a fern genome (Rabinowicz et al. 2005) and the observation that genome size and chromosome number are strongly correlated in ferns (Nakazato et al. 2008). These observations imply that although fern genomes appear to lose duplicate genes through gene silencing, physical genetic material may be lost at a rate that is slow compared with many angiosperms. It is not entirely clear why this occurs, but additional information—including EST data from more fern lineages, as well as fern whole-genome sequences—is needed to determine how, and ultimately why, fern nuclear genomes are different from those of seed plants, and whether such differences relate to homospory and heterospory or to life-history differences.

Neopolyploidy

Because recent genomic analyses suggest that high chromosome numbers in ferns are not indicative of multiple rounds of ancient polyploidy, what does this imply for estimates of recent polyploidy? Because of their large chromosome numbers, ferns were thought to be the most highly polyploid lineage of plants. Recently formed polyploid species that have not undergone diploidization, and that therefore still exhibit polyploid genetics, are termed neopolyploids. Past estimates of the proportion of fern species that are neopolyploid were often based on chromosome count cutoffs, and the exceptionally high chromosome numbers of homosporous ferns were taken as evidence of rampant neopolyploidy. For example, Grant (1981) estimated that more than 95% of fern species were of polyploid origin by assuming that all species with more than $n = 14$ chromosomes were polyploids. Vida (1976) made a more realistic estimate of 43.5% by considering the base numbers in each fern genus as diploid. Molecular genetic tools, however, can use aspects of gene expression as an independent portrait of neopolyploidy rather than relying on chromosome number per se. By examining patterns of gene expression from isozymes, it became apparent that species with the lowest chromosome numbers for their genus, regardless of how high this number may be, are actually diploids (Haufler and Soltis 1986). Species with polyploid patterns of gene expression were nearly always multiples of this base number. Analyses of duplicate genes from plant EST data also support the observation that species with base numbers for their genus do not show evidence of recent genome duplications. Importantly, EST analyses of known neopolyploids demonstrate extra peaks consistent with recent duplication (Barker and Rieseberg 2008).

Recent estimates of polyploidy have been calculated for ferns by considering a phylogenetic pattern of chromosome number evolution. An analysis of the well-studied North American fern flora indicates that 30.9% are neopolyploids, and this number rises to 38.9% if odd ploidy levels such as triploids and pentaploids are included (Flora of North America Editorial Committee 1993). These results agree with a global analysis that identified 32.86% of leptosporangiate and 24.0% of eusporangiate fern species as neopolyploids (Wood et al. 2009). Compared with most estimates of polyploidy before the advent of molecular tools, these numbers are quite low but similar to estimates of the incidence of polyploid species among flowering plants (Wood et al. 2009). However, the percentage of speciation events due to polyploidy is twice as large in ferns (31.37%) as it is in angiosperms (15.00%; Wood et al. 2009). Considering that ferns and angiosperms have similar levels of polyploid incidence, but with half as many new species initiated by polyploidy in angiosperms, it seems that, once established, angiosperm polyploids are approximately twice as successful at producing new species as fern polyploids—an inference that departs from historical views of neopolyploidy in ferns.

Speciation

With approximately 30% of fern speciation events resulting from genome duplication, the remaining 70% of speciation events most likely result from divergence among species at the same ploidy level. Most of this divergence will occur at the diploid level. Although molecular tools have significantly increased our ability to recognize cryptic species and discriminate species boundaries (Haufler 2008), we are just now beginning to use genomic approaches to understand the types of intrinsic genetic mechanisms that initiate or maintain species discreteness, a critical component of speciation. Two potential mechanisms of genetic incompatibility that can cause reproductive isolation between species are chromosomal rearrangements and Bateson-Dobzhansky-Mueller (BDM) incompatibilities. Chromosomal rearrangements, such as inversions or translocations, may cause reproductive isolation because meiotic recombination in hybrid individuals will most likely yield gametes missing critical chromosomal sections or genes. Alternatively, BDM incompatibilities arise when new alleles evolve in geographically separated populations. These new alleles, having evolved in isolation, may be genetically incompatible when they are present in the same genome, and thus contribute to reproductive isolation between species. An example of a BDM incompatibility is seen in a necrotic *Arabidopsis* hybrid containing two species-specific alleles at a disease resistance gene. This allelic combination has been shown to be necessary and sufficient to cause the incompatibility (Bomblies et al. 2007). So what types

of genetic incompatibilities contribute to fern speciation? The observation that most diploid hybrids are sterile, but recover their fertility when their genomes double, indicates that chromosomal rearrangements are widespread in ferns. However, it is not clear whether these initiate speciation or these rearrangements are built up after initial divergence. A genetic mapping study (Nakazato et al. 2007) of divergent populations of the diploid homosporous fern *C. richardii* provides some insight into this question. Using two geographically separated populations of *C. richardii* that yield only partially fertile F1s, Nakazato and colleagues (2007) examined changes in spore viability and distortions in the inheritance patterns of loci from the two parents in two genetically distinct F2 populations. The F2 populations showed substantial increases in spore viability relative to the F1 generation, an indication of significant chromosomal rearrangements between the parental populations. However, Nakazato and colleagues (2007) also observed evidence for numerous BDM incompatibilities between various nuclear loci of the parents, and nuclear–cytoplasmic incompatibilities. Most mammals and many angiosperm species studied appear to be isolated primarily by BDM incompatibilities. In contrast, both chromosomal rearrangements and BDM incompatibilities were roughly equally responsible for reproductive isolation between the divergent *Ceratopteris* populations. Additional linkage mapping and genomic analyses of other fern species will very likely be fruitful in elucidating the primary mechanisms of fern speciation.

Evolution of development

Research on the evolution of plant development has focused on seed plants. However, many of the genes involved in development of the flower, for example, have homologues in nonflowering clades (Hasebe 1999). This again illustrates the importance of examining the basic biology of taxa other than model organisms. If genes involved in developmental pathways evolved for functions different from those seen in model organisms, then we may be missing key aspects in our understanding of the evolution and function of these systems (Cronk 2001). Fortunately, there is an effort to study development in a wider range of plants (Pryer et al. 2002, Nishiyama 2007), and several techniques can be applied to species other than the usual model organisms (Floyd and Bowman 2007).

An indispensable tool for determining gene function is the ability to "knock out" a particular gene and examine the effect on the phenotype. Not only can this approach be used to study the function of a single gene but it can also be used to reveal biochemical and developmental pathways and the sequential expression of genes involved. Such techniques are well developed in most model organisms, but until recently they were not possible in ferns. The first approach to be used in ferns was RNA interference (RNAi), which was applied successfully to knock out targeted genes involved in cytoskeleton formation in the fern *Marsilea vestita* (Klink and Wolniak 2000). RNAi uses antisense or double-stranded RNA that corresponds to a gene targeted for silencing. The technique mimics a suite of naturally occurring systems involved in gene regulation in eukaryotes and defense mechanisms in prokaryotes (Shabalina and Koonin 2008). In eukaryotes, genomic regions that encode naturally interfering RNAs are closely coupled with the genes they are regulating. However, the addition of synthetic RNA that is complementary to the mRNA of a target gene can in many cases completely block translation of that gene. Stout and colleagues (2003) reported RNAi in *C. richardii* and demonstrated RNAi silencing of genes selected from a *C. richardii* EST library. Blocking a gene's action and observing the phenotype is one important approach in the study of gene function. More recently, Kawai-Toyooka and colleagues (2004) developed a DNA interference (DNAi) approach for targeted gene silencing in the fern *A. capillus-veneris*. DNAi uses synthetic fragments of promoterless double-stranded DNA, easily generated by PCR (the polymerase chain reaction), that are then delivered directly to living cells. As for RNAi, the complementarity of the introduced nucleic acids is directed at a target gene. However, DNAi has advantages over RNAi in that it is easier to generate, is more stable, and appears to target the nuclear genes rather than the transcripts. DNAi interferes with transcription so it provides a more permanent gene silencing than RNAi. Also, the DNAi system can be used to target nontranscribed regions of the genome, perhaps enabling the future study of regulatory elements. Both RNAi and DNAi will likely be profitable for augmenting our knowledge of the roles of nuclear genes in ferns by facilitating reverse genetic approaches: manipulating gene targets and examining the effect on phenotype in new model organisms.

One important way in which ferns can provide unique research opportunities is in the study of genes that are associated with only one of the two life-cycle stages. Gametophyte EST libraries have been developed for seed plants (Honys and Twell 2003, Lee and Lee 2003), but the gametophytes are not actually independent in these plants, so it is difficult to factor out effects from the sporophyte. Ferns, however, have truly independent sporophyte and gametophyte stages, so it should be possible to examine genes expressed only in one stage. Are the same patterns seen across ferns, and which stage genes share homologies with reproductive genes in seed plants? The need for EST collections in ferns is paramount.

Conclusions

Chromosomal studies have dominated much of fern evolutionary biology for the last half-century. Considering the uniqueness of fern genomes, this is likely to continue, as they provide a contrast to patterns of genome evolution that we observe in angiosperms. For example, ferns provide a unique opportunity for studying the process of genetic diploidization in genomes that may not experience

substantial chromosomal loss following polyploidy. In such an environment, are the forces driving sub- and neofunctionalization of duplicated genes the same as in the apparently much more dynamic genomes of angiosperms? And to what extent are the two independent gametophytic and sporophytic phases involved in the sub- and neofunctionalization of duplicate genes? With additional ESTs and whole-genome sequences for ferns, we will be better positioned to address these questions. Such data sets could also be used to begin identifying candidate genes involved in adaptation among fern species and discerning the roles of selection, hybridization, and neutral processes in different aspects of fern evolution. Further, the unique biology and phylogenetic position of ferns as sister to the seed plants demands that we also use new genomic tools to increase our knowledge of the evolution of plant development. The substantial morphological and reproductive differences between ferns and seed plants provide a distinct opportunity to improve our understanding of plant evolution. Ferns are an excellent test bed for our concepts of plant evolution and evolutionary theory, and new sequencing technologies promise to crack the black box of fern genomes. Although ESTs are starting to provide us with a picture of fern transcriptomes, whole nuclear-genome sequences will be needed to fully understand the complete architecture of fern nuclear genomes, including the noncoding regions, the amount of repetitive DNA, and how genes are organized on the chromosomes. Genome-sequencing projects started deliberately with study plants of economic importance, and small genomes had to be the first ones sequenced to minimize costs. However, ferns have very large genomes that are probably among the most complex (although we do not know this for sure). The fern genus *Ophioglossum* has the highest reported chromosome number ($2n = 1440$) of any organism (Khandelwal 1990). As new technologies emerge, it should be possible not only to sequence entire fern genomes, but the ability to do so might be a good test case for an emerging technology. If you can sequence a fern, you can sequence anything.

Acknowledgments

Thanks to National Science Foundation grant DEB-0228432 to P. G. W for funding. M. S. B is supported by the NSERC CREATE Training Program in Biodiversity Research. Carol Rowe, Mike Pfrender, Chris Haufler, Jerrry Gastony, and two anonymous reviewers made constructive comments on an earlier version of the manuscript.

References cited

Barker MS. 2009. Evolutionary genomic analyses of ferns reveal that high chromosome numbers are a product of high retention and fewer rounds of polyploidy relative to angiosperms. American Fern Journal 99: 136–141.

Barker MS, Rieseberg LH. 2008. Evolutionary genomics of hybridization: Detecting ancient hybridization and introgression by the inference of intrologs in plant genomes. Abstract presented at the Botany 2008 Conference; 26–30 July 2008, University of British Columbia.

Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, Knapp SJ, Rieseberg LH. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. Molecular Biology and Evolution 25: 2445–2455.

Barker MS, Vogel H, Schranz ME. 2009. Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. Genome Biology and Evolution 1: 391–399.

Blanc G, Wolfe, KH. 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell 16: 1679–1691.

Bomblies K, Lempe J, Epple P, Warthmann N, Lanz C, Dangl JL, Weige D. 2007. Autoimmune response as a mechanism for a Dobzhansky–Muller-type incompatibility syndrome in plants. PLoS Biology 5: 1962–1972.

Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature 422: 433–438.

Chapman RH, Klekowski EJ Jr., Selander RK. 1979. Homoeologous heterozygosity and recombination in the fern *Pteridium aquilinum*. Science 204: 1207–1209.

Conant DS, Stein DB, Valinski AEC, Sudarsanam P, Ahearn ME. 1994. Phylogenetic implications of chloroplast DNA variation in the Cyatheaceae, 1. Systematic Botany 19: 60–72.

Cronk QCB. 2001. Plant evolution and development in a post-genomic context. Nature Reviews Genetics 2: 607–619.

Cui LY, et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Research 16: 738–749.

Doebley J, Durbin M, Golenberg EM, Clegg MT, Ma DP. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbcL*) nucleotide sequence among the grasses (gramineae). Evolution 44: 1097–1108.

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF. 2008. Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42: 443–461.

Flora of North America Editorial Committee. 1993. Flora of North America, North of Mexico. Oxford University Press.

Floyd SK, Bowman JL. 2007. The ancestral developmental tool kit of land plants. International Journal of Plant Sciences 169: 1–35.

Gastony GJ. 1991. Gene silencing in a polyploid homosporous fern: Paleopolyploidy revisited. Proceedings of the National Academy of Sciences 88: 1602–1605.

Gastony GJ, Gottlieb LD. 1985. Genetic variation in the homosporous fern *Pellaea andromedifolia*. American Journal of Botany 72: 257–267.

Gastony GJ, Yatskievych G, Dixon CK. 1992. Chloroplast DNA restriction site variation in the fern genus *Pellaea*: Phylogenetic relationships of the *Pellaea glabella* complex. American Journal of Botany 79: 1072–1080.

Grant V. 1981. Plant Speciation. 2nd ed. Columbia University Press.

Hamilton RG, Lloyd RM. 1991. Antheridiogen in the wild: The development of fern gametophyte communities. Functional Ecology 5: 804–809.

Hasebe M. 1999. Evolution of reproductive organs in land plants. Journal of Plant Research 112: 463–474.

Hasebe M, Iwatsuki K. 1992. Gene localization on the chloroplast DNA of the maiden hair fern: *Adiantum capillis-veneris*. Botanical Magazine of Tokyo 105: 413–419.

Hasebe M, et al. 1995. Fern phylogeny based on *rbcL* nucleotide sequences. American Fern Journal 85: 134–181.

Haufler CH. 1987. Electrophoresis is modifying our concepts of evolution in homosporous pteridophytes. American Journal of Botany 74: 1987.

———. 2002. Homospory 2002: An odyssey of progress in pteridophyte genetics and evolutionary biology. BioScience 52: 1081–1093.

———. 2008. Species and speciation. Pages 303–333 in Ranker TA, Haufler CH, eds. Biology and Evolution of Ferns and Lycophytes. Cambridge University Press.

Haufler CH, Soltis DE. 1984. Obligate outcrossing in a homosporous fern: Field confirmation of a laboratory prediction. American Journal of Botany 71: 878–881.

———. 1986. Genetic evidence suggests that homosporous ferns with high chromosome numbers are diploid. Proceedings of the National Academy of Sciences 83: 4389–4393.

Hickok LG. 1978. Homoeologous chromosome pairing and restricted segregation in the fern *Ceratopteris*. American Journal of Botany 65: 516–521.

Hickok LG, Klekowski EJ Jr. 1974. Inchoate speciation in *Ceratopteris*: An analysis of the synthesized hybrid *C. richardii* x *C. pteridoides*. Evolution 28: 439–446.

Holsinger KE. 1987. Gametophytic self-fertilization in homosporous plants: Development, evaluation, and application of a statistical method for evaluating its importance. American Journal of Botany 74: 1173–1183.

Honys D, Twell D. 2003. Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiology 132: 640–652.

Jaillon O, et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449: 463–467.

Kawai-Toyooka H, Kuramoto C, Orui K, Motoyama K, Kikuchi K, Kanegae T, Wada M. 2004. DNA interference: A simple and efficient gene-silencing system for high-throughput functional analysis in the fern *Adiantum*. Plant and Cell Physiology 45: 1648–1657.

Khandelwal S. 1990. Chromosome evolution in the genus *Ophioglossum*. Botanical Journal of the Linnean Society 102: 205–217.

Klekowski EJ Jr. 1973. Sexual and subsexual systems in homosporous pteridophytes: A new hypothesis. American Journal of Botany 60: 535–544.

Klekowski EJ Jr., Baker HG. 1966. Evolutionary significance of polyploidy in the Pteridophyta. Science 153: 305–307.

Klink VP, Wolniak SM. 2000. The efficacy of RNAi in the study of the plant cytoskeleton. Journal of Plant Growth Regulation 19: 371–384.

Lee JY, Lee DH. 2003. Use of serial analysis of gene expression technology to reveal changes in gene expression in *Arabidopsis* pollen undergoing cold stress. Plant Physiology 132: 517–529.

Manton I. 1958. Chromosomes and fern phylogeny with special reference to "Pteridaceae". Journal of the Linnean Society (Botany) 56: 73–92.

Manton I. 1950. Problems of Cytology and Evolution in the Pteridophyta. Cambridge University Press.

McGrath JM, Hickok LG. 1999. Multiple ribosomal RNA gene loci in the genome of the homosporous fern *Ceratopteris richardii*. Canadian Journal of Botany 77: 1199–1202.

McGrath JM, Hickok LG, Pichersky E. 1994. Assessment of gene copy number in the homosporous ferns *Ceratopteris thalictroides* and *C. richardii* (Parkeriaceae) by restriction fragment length polymorphisms. Plant Systematics and Evolution 189: 203–210.

Nakazato T, Jung MK, Housworth EA, Rieseberg LH, Gastony GJ. 2006. Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. Genetics 173: 1585–1597.

———. 2007. A genomewide study of reproductive barriers between allopatric populations of a homosporous fern, *Ceratopteris richardii*. Genetics 177: 1141–1150.

Nakazato T, Barker MS, Rieseberg LH, Gastony GJ. 2008. Evolution of the nuclear genome of ferns and lycophytes. Pages 175–198 in Ranker TA, Haufler CH, eds. Biology and Evolution of Ferns and Lycophytes. Cambridge University Press.

Nishiyama T. 2007. Evolutionary developmental biology of nonflowering land plants. International Journal of Plant Sciences 168: 37–47.

Palmer JD. 1987. Chloroplast DNA evolution and biosystematic uses of chloroplast DNA variation. The American Naturalist 130: 6–29.

Pichersky E, Soltis DE, Soltis PS. 1990. Defective chlorophyll a/b-binding protein genes in the genome of a homosporous fern. Proceedings of the National Academy of Sciences 87: 195–199.

Pryer KM, Schneider H, Zimmer EA, Banks JA. 2002. Deciding among green plants for whole genome studies. Trends in Plant Sciences 7: 550–554.

Pryer KM, Schuettpelz E, Wolf PG, Schneider H, Smith AR, Cranfill R. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. American Journal of Botany 91: 1582–1598.

Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA. 2005. Differential methylation of genes and repeats in land plants. Genome Research 15: 1431–1440.

Raubeson LA, Jansen RK. 1992. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255: 1697–2699.

Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallion S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. Nature 428: 553–557.

Schneller JJ, Haufler CH, Ranker TA. 1990. Antheridiogen and natural gametophyte populations. American Fern Journal 80: 138–147.

Schuettpelz E, Pryer KM. 2007. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56: 1037–1050.

Shabalina SA, Koonin EV. 2008. Origins and evolution of eukaryotic RNA interference. Trends in Ecology and Evolution 23: 578–587.

Stein DB, Conant DS, Ahearn ME, Jordan ET, Kirch SA, Hasebe M, Iwatsuki K, Tan MK, Thomason JA. 1992. Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proceedings of the National Academy of Sciences 89: 1856–1860.

Stout SC, Clark GB, Archer-Evans S, Roux SJ. 2003. Rapid and efficient suppression of gene expression in a single-cell model system, *Ceratopteris richardii*. Plant Physiology 131: 1165–1168.

Tang HB, Bowers JE, Wang XY, Ming R, Alam M, Paterson AH. 2008. Synteny and collinearity in plant genomes. Science 320: 486–488.

Vida G. 1976. The role of polyploidy in evolution. In Novak VJA, Pacltova B, eds. Evolutionary Biology. Czechoslovak Academy of Science.

Wolf PG, Roper JM. 2008. Structure and evolution of fern plastid genomes. Pages 159–174 in Ranker T, Haufler CH, eds. Biology and Evolution of Ferns and Lycophytes. Cambridge University Press.

Wolf PG, Haufler CH, Sheffield E. 1987. Electrophoretic evidence for genetic diploidy in the bracken fern (*Pteridium aquilinum*). Science 236: 947–949.

Wolf PG, Rowe CA, Sinclair RB, Hasebe M. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Research 10: 59–65.

Wolf PG, Rowe CA, Hasebe M. 2004. High levels of RNA editing in a vascular plant chloroplast genome: Analysis of transcripts from the fern *Adiantum capillus-veneris*. Gene 339: 89–97.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences 106: 13875–13879.

Michael S. Barker (msbarker@biodiversity.ubc.ca) is with the Biodiversity Research Centre and Department of Botany, at the University of British Columbia, Vancouver, Canada, and is also affiliated with the Department of Biology and Center for Genomics and Bioinformatics at Indiana University in Bloomington. Paul G. Wolf (wolf@biology.usu.edu) is with the Ecology Center and Department of Biology at Utah State University, in Logan.