Journal Title: International journal of plant sciences.

Vol: 164 **Iss:** 3

Month/Year: 2003

Pages: S43-S54

Article Author:

Article Title: Monson,; Gene duplication, neofunctionalization, and the evolution of C-4 photosynthesis

Imprint: Chicago, IL ; University of Chicago Pres

If there are any transmission problems with this document please provide the details below and send this sheet back to us as soon as possible.

FAX 360-650-7397 or ARIEL 140.160.178.79

Date Received:

Pages are: Missing Cut Off Other
(Please circle the appropriate description) top bott left right

Page Numbers:

Additional Information:

Int. J. Plant Sci. 164(3 Suppl.):S43–S54. 2003.

1058-5893/2003/16403S-0004$15.00

GENE DUPLICATION, NEOFUNCTIONALIZATION, AND THE EVOLUTION OF C_4 PHOTOSYNTHESIS

Russell K. Monson[1]

Department of Environmental, Population, and Organismic Biology, University of Colorado, Boulder, Colorado 80309, U.S.A.

The evolution of C_4 photosynthesis provides one of the most interesting examples of evolutionary novelty in plants. As an adaptation that enhances plant carbon gain in warm climates with high light and relatively low atmospheric CO_2 concentration, the complex interactions between C_4 anatomy and biochemistry appear to have evolved over thirty times independently within the angiosperms. Past theories have explained the multiple appearances of C_4 photosynthesis solely on the basis of global decreases in atmospheric CO_2 concentration during the past 50 million years. The premise of such theories is that the C_4 pathway provides selective advantages in terms of plant carbon gain in an atmosphere of low CO_2 concentration. These "carbon balance" theories, however, are limited in their ability to explain why or how C_4 photosynthesis evolved so many times independently and why certain patterns in the taxonomic distribution of C_4 photosynthesis exist; e.g., the absence of C_4 photosynthesis in canopy-forming forest tree species and the paucity of C_4 species within eudicots compared to monocots. In this review, I present the case that one of the most often overlooked aspects of C_4 evolution is the potential for genetic limitation, specifically that associated with gene duplication and subsequent modification, which is crucial to the evolution of C_4 biochemistry. I describe the research to date that provides insight into the origins of C_4 genes, and I derive the conclusion that the evolution of C_4 photosynthesis is largely a story of gene duplication while plants are still in the ancestral, C_3 state. Once a reservoir of key, duplicated, and preserved C_3 genes is present, a small amount of subsequent modification within gene promoter regions is all that is necessary to transform certain C_3 patterns of gene expression to C_4 patterns. Quantitative theory predicts that the most likely factors to be associated with the accumulation of a reservoir of duplicated C_3 genes are large population size, short generation time, and frequent recruitment of sexually produced individuals. When combined with the selective pressures of reduced atmospheric CO_2 concentration, consideration of population and life history factors, and the genetic constraints that they impose, could help explain certain patterns of C_4 distribution.

Keywords: monocots, eudicots, nonfunctionalization, adaptive, PEP carboxylase, pyruvate orthophosphate dikinase, NADP-malic enzyme, carbonic anhydrase, Rubisco.

Introduction

As a morphological and biochemical novelty, the evolution of C_4 photosynthesis has received much attention in recent years, including its relevance to global climate change (Ehleringer et al. 1991; Cerling et al. 1993, 1997) and changes in the atmospheric CO_2 concentration (Ehleringer and Monson 1993; Ehleringer et al. 1997; Cowling 2001; Sage 2001). Patterns in C_4 evolution are generally inferred with the assumption that past reductions in the atmospheric CO_2 concentration have suppressed photosynthesis and stimulated photorespiration to such an extent that the C_4 CO_2-concentrating mechanism was selectively favored (Sage 1999). According to this assumption, favorable plant carbon balance is the principal driver of diversification within C_4 taxa. Several more specific aspects of C_4 evolution have been explained by the carbon balance perspective, including the high frequency with which C_4 photosynthesis has evolved, its restricted diversification within eudicots, and its absence from forest tree species (Monson 1989; Ehleringer et al. 1997; Monson 1999). Broader perspectives have not been generally considered in explaining these patterns.

With the recent publication of gene sequences for the genomes of several organisms, it is has become possible to gain new insight into the origins of novel genes and their role in the evolution of developmental, morphological, and biochemical traits. One of the principal mechanisms known to influence the evolution of novel genes is gene duplication followed by advantageous mutations in one of the duplicates to provide a novel role (referred to as neofunctionalization). The rate of gene duplication and, more importantly, the frequency of neofunctionalization following duplication represent potential limitations to the rate at which novel traits appear in populations (Lynch and Conery 2000). In this article, I explore the potential for these factors to affect the frequency and pattern of C_4 evolution. I make the argument that, when combined with perspectives on plant carbon balance and historical

[1] Fax 303-492-8699; e-mail russell.monson@colorado.edu.

Manuscript received October 2002; revised manuscript received December 2002.

changes in the earth's atmospheric CO_2 concentration, consideration of genetic constraints might provide the basis for a broader understanding of C_4 evolutionary patterns.

Historical Patterns in the Evolution of C_4 Photosynthesis

C_4 photosynthesis is found in only 3% of the ca. 250,000 vascular plant species (Sage et al. 1999). Despite its restricted taxonomic distribution, C_4 photosynthesis contributes to 30% of global primary productivity (Gillon and Yakir 2001), reflecting the widespread distribution of C_4-dominated grasslands at tropical and subtropical latitudes. Ca. 75% of all C_4 species are in the monocots, despite the fact that monocot species only represent 25% of the angiosperms; the two monocot families with C_4 species are the Poaceae and Cyperaceae. The diversity of C_4 species is much lower within the eudicots, with only two of the 15 C_4 eudicot families containing more than 10 C_4 species (Sage et al. 1999). The eudicot family that contains the most C_4 species, the Amaranthaceae, is very large (ca. 170 genera and 2400 species) and contains 50% of all eudicot C_4 species. Most C_4 species are herbaceous; there are some woody C_4 shrub species and some C_4 species that are arborescent. However, there are no known C_4 species represented among canopy-forming forest tree species (Sage 2001).

In terms of geological time, the appearance of C_4 photosynthesis is a relatively recent phenomenon. The earliest, definitive evidence of C_4-specific Kranz anatomy in fossil leaves is from specimens that date to 12.5 Ma (Nambudiri et al. 1978; Tidwell and Nambudiri 1989). Estimation of C_4 age from phylogenetic analyses utilizing molecular clock techniques indicate that in the grasses C_4 first appeared 20–30 Ma, in the late Oligocene or early Miocene (Kellogg 1999). Analysis of carbon stable isotopes from pedogenic soil paleosols indicate global expansion of C_4 grasslands 7–5 Ma (Cerling 1999). In several analyses, Ehleringer and coworkers have argued that the widespread appearance of C_4 photosynthesis is correlated with global reductions in the atmospheric CO_2 concentration from greater than 500 ppmv during the Miocene and continuing into the Pleistocene to less than 300 ppmv during the Pleistocene (Ehleringer et al. 1991, 1997; Ehleringer and Monson 1993; also Kuypers et al. 1999).

The cause of C_4 expansion during periods of reduced atmospheric CO_2 concentration is presumed to be the C_4 advantage of reduced photorespiration rates and concomitantly higher photosynthetic quantum yield (moles of CO_2 assimilated per moles of photons absorbed) at low CO_2 concentrations and high temperatures (Ehleringer and Björkman 1977; Ehleringer 1978; Ehleringer and Pearcy 1983). Similar arguments involving differences in the photosynthetic quantum yield between C_4 monocots and C_4 eudicots were used to explain the lower levels of taxonomic diversity observed among C_4 eudicots (Ehleringer et al. 1997). According to a theory presented in the latter article, at high temperatures and in areas with summer monsoon activity, the photosynthetic quantum yield of C_4 monocots would have exceeded that of C_3 monocots at higher CO_2 concentrations, such as those characteristic of the late Miocene (7 Ma). However, the photosynthetic quantum yield of C_4 eudicots would not have exceeded that of C_3 eudicots until CO_2 concentrations reached much lower values, such as those characteristic of glacial periods during the Pleistocene (50–18 ka). With less time available for the C_4 eudicot advantage, it is assumed that C_4 eudicots have had fewer opportunities for taxonomic diversification. (In contrast to this argument, however, a recent molecular analysis of the Amaranthaceae provides evidence that C_4 photosynthesis in this family has existed since at least the late Miocene [ca. 10 Ma], with the potential for C_4 diversification since that time [H. Freitag, personal communication].)

Monson (1989, 1999) used similar quantum yield arguments to provide a possible reason for the absence of C_4 photosynthesis in forest trees. Several woody species have been identified with C_4 photosynthesis (particularly in the Amaranthaceae and Chamaesyce [=Euphorbia] [Pearcy and Troughton 1975; Winter 1981]), although dominant forest tree species are virtually all C_3 plants. In studies of C_3-C_4 intermediate species, it was clear that "adaptive troughs," represented by biochemical inefficiencies and low photosynthetic quantum yields, are possible along the evolutionary path from C_3 to C_4 photosynthesis (Monson et al. 1986; Monson 1989). These adaptive troughs could impose a barrier to the evolution of fully expressed C_4 photosynthesis in species native to shaded, forested environments.

One criticism that can be applied to these past theories is that they are founded on the assumption that the adaptive advantages that are obvious in the final evolutionary product were the same advantages that caused the initial appearance of the trait. In C_4 photosynthesis, there are reasons to suspect that this may not be the case. Using C_3-C_4 intermediate species as models to understand the sequential process of C_4 evolution, Monson (1989, 1999), Rawsthorne (1992), and Monson and Rawsthorne (2000) have proposed schemes in which the CO_2-concentrating function of the C_4 pathway and its advantages in terms of improved quantum yield do not appear until late in the evolutionary sequence. It is possible that a different type of CO_2-concentrating mechanism, one that depends on the decarboxylation of glycine rather than malate or aspartate, was present during the initial stages of C_4 evolution. The glycine-decarboxylation mechanism may indeed provide a slight photosynthetic advantage through a CO_2-concentrating mechanism (von Caemmerer 1989; Schuster and Monson 1990); however, its effectiveness is considerably reduced compared with fully expressed C_4 species, and it is significantly less effective in enhancing the quantum yield for CO_2 assimilation (Monson et al. 1986). Thus, there is reason to question the significance of differences in the quantum yield between fully expressed C_3 and C_4 plants as the initial force driving the evolution of C_4 photosynthesis. Given these reservations to the current interpretation of C_4 evolution, I argue that there are other possible causes and constraints that merit evaluation. In this article, I introduce gene duplication and neofunctionalization, which have had an important role in C_4 evolution but have not been considered in past discussions of C_4 diversification.

Gene Duplication as an Evolutionary Force and Limitation

One of the most frequently discussed issues in evolutionary biology concerns the genetic origins of morphological and biochemical novelties (Lynch and Conery 2000; Schubert et al.

2000; Shubin and Marshall 2000; Bruce et al. 2001; Van de Peer et al. 2001; Averof 2002; Mazet and Shimeld 2002). Where do the genes come from that produce novel traits in organisms, and why do selection and genetic drift not silence novel genes in a population before gene frequencies reach a stable value? Much of this discussion has focused on gene duplication as the most likely source for novel traits. Gene duplication refers to the production of two duplicate loci on the same chromosome, where previously there had been only one through unequal crossing over, tandem duplication of linked genes, or other illegitimate chromosomal rearrangements. In plants, duplicate genes appear to be mostly unlinked, although this is based on observations with only two species to date (McGrath et al. 1993; Thomas et al. 1993). Gene duplication was originally discussed in the work of Muller (1936), but its full importance in the evolutionary process was not realized until Ohno's seminal book over 30 yr later (Ohno 1970). Since the theoretical work of Ohno, and subsequently several others (Nei and Roychoudhury 1973; Kimura and King 1979; Ohta 1987, 1988, 2000), empirical evidence for the evolutionary importance of gene duplication has been observed in the prevalence of multigene families, which are ubiquitous among species (Huynen and van Nimwegen 1998; Friedman and Hughes 2001; Harrison and Gerstein 2002).

Past theoretical models that balance the rate of deleterious mutation against the spread of duplicated alleles in a randomly mating population have predicted that it is much more rare for a duplicated gene to reach a stable frequency in a population than to be silenced through mutation (made into a nonfunctional "pseudogene") (Haldane 1933; Ohno 1970; Ohta 2000). In order to promote the evolution of a novel trait, the duplicated gene must reach high enough frequencies to diverge and take on a new function, neofunctionalization, before being silenced. Without neofunctionalization, silencing by genetic drift is inevitable (Nei and Roychoudhury 1973; Kimura and King 1979; Maruyama and Takahata 1981; Watterson 1983; Ohta 1988). Some insight into the extent of gene loss after duplication can be gained by examining ancient polyploids in which the entire genome has been duplicated. In one study, it was concluded that 92% of the duplicated genes in yeast polyploids were lost within 100 Ma after genome duplication (Seoighe and Wolfe 1998).

Rates of gene duplication and nonfunctionalization have been recently estimated in an analysis of whole-genome sequence data from three eukaryotes (*Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans*) (Lynch and Conery 2000). The rate of duplication ranges across a factor of 10 depending on species and gene (from 0.2 to 2 duplications per gene per 100 million years). This rate of duplication is within the same range as the rate of point mutation, indicating that as an evolutionary force and source of genetic novelty, gene duplication may be as important as mutation. The half lives of duplicated genes in the species that have been examined range from 3 to 7 million years, and it is estimated that 90% of the duplicates are silenced within 50 million years after duplication (Lynch and Conery 2000). Finally, the rate of nonsynonymous mutation (i.e., mutations that affect the amino acid sequence of translated proteins) decreases over time relative to the rate of synonymous mutation (i.e., mutations without effect on amino acid sequence); this indicates that selection to retain specific gene sequences is relaxed immediately after a duplication event but increases over time. In this regard, it has been hypothesized that one of the advantages conferred by gene duplication is an immediate relaxation in selection pressure on the gene duplicate, which increases its potential to evolve through random drift (Ohno 1970).

Theoretical estimates of the frequency with which gene duplicates are retained through neofunctionalization tend to be greater than those derived from empirical data (Hughes and Hughes 1993). It was generally assumed that mutations in a duplicated gene resulted in only one of two possibilities—nonfunctionalization (to form a pseudogene) or neofunctionalization (to form a novel gene). It is now clear that there is an additional possibility, subfunctionalization, which, if prevalent, could enhance the retention frequency of gene duplicates (fig. 1). In subfunctionalization, it is recognized that many genes control more than one function (i.e., subfunctions), either through differentiation of *cis*-regulatory elements that control tissue-specific patterns of expression or differentiation within the coded region producing alternative functional domains of a protein. If gene duplication results in the production of duplicate subfunctionalized genes, subsequent modification through mutation may result in the reciprocal loss of subfunctions; thus, the two genes may diverge to independently partition control over gene expression. In such a scenario, the two genes would not only share control over expression patterns, but they would also share in the selective advantages they contribute to the organism (Arntz and Delph 2001). This places a selective premium on both genes simultaneously, thus resisting the tendency for the duplicated gene to be lost through nonfunctionalization. Divergence of subfunctionalized genes does not exclude future neofunctionalization, and in fact subfunctionalization may increase the chances of neofunctionalization since duplicated genes are retained longer and this provides more time for potentially advantageous mutations to accumulate. It has proven difficult to identify examples of subfunctionalization, although a strong case has been made with regard to developmental genes in zebra fish (DeMartino et al. 2000; Bruce et al. 2001). It is possible that many subfunctionalizations are masked by subsequent neofunctionalization (Mazet and Shimeld 2002).

The Evidence for Gene Duplication and Neofunctionalization in C_4 Evolution

The evolution of C_4 photosynthesis involves numerous biochemical and anatomical changes, most likely involving hundreds to thousands of key mutational events, in up to several tens of genes. A molecular survey study utilizing cDNAs as probes for transcript abundance in the mesophyll and bundle-sheath cells of the C_4 monocot, sorghum, identified 25 different mesophyll-specific gene sequences and eight different bundle-sheath-specific gene sequences (Wyrich et al. 1998). Compared to the ancestral C_3 state, each of these sequences would have had to evolve unique regulatory properties to provide differential expression between the cell types. This analysis probably underestimates the true number of gene changes required to evolve C_4 photosynthesis, since a significant number of developmental genes must also have emerged to produce the unique anatomical and ultrastructural features of C_4 plants. Given

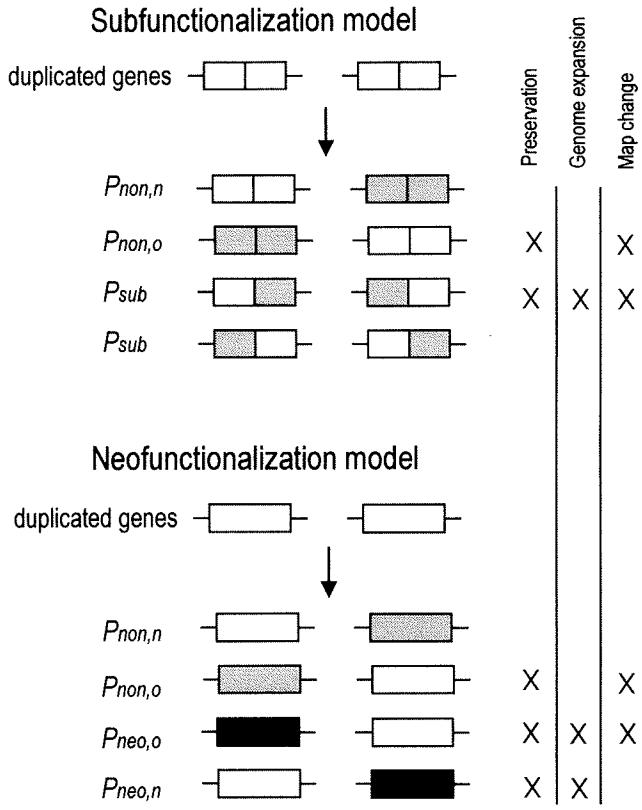

Fig. 1 A schematic showing the possible paths for nonfunctionalization, subfunctionalization, or neofunctionalization following gene duplication. The statistical probabilities for each event can be defined with $P_{non,n}$ and $P_{non,o}$ representing the probabilities of nonfunctionalization for the newly arisen gene and original gene, respectively; P_{sub} representing the probability of subfunctionalization; $P_{neo,n}$ and $P_{neo,o}$ representing the probabilities of neofunctionalization for the newly arisen gene or original gene, respectively. In the upper part of the figure, a duplicate gene pair is shown entering a series of probabilities in which one member of the duplicated gene is nonfunctionalized ($P_{non,n}$, $P_{non,o}$) by deleterious mutations (shown by gray) or subfunctionalized (P_{sub}) by deleterious mutations in complementary parts of both gene copies. In the lower part of the figure, a duplicate gene pair is shown entering a series of probabilities in which one member of a duplicated gene is nonfunctionalized ($P_{non,n}$, $P_{non,o}$) by deleterious mutations or neofunctionalized ($P_{neo,n}$, $P_{neo,o}$) by beneficial mutations in one member of a duplicated gene. The outcome of each event in terms of preservation of the newly arisen gene, expansion of the genome, or a genetic map change is represented on the right, with an "X" indicating a positive outcome. Redrawn from Lynch et al. 2001.

such a large diversity of molecular evolutionary events that must have occurred, and given that C_4 photosynthesis has evolved over thirty times independently (Kellogg 1999), the C_4 evolutionary story represents an unprecedented example of gene duplication with subsequent diversification through neofunctionalization. When it is further recognized that many of these evolutionary events likely occurred within the relatively short time span of 5–7 million years (the time of principal C_4 expansion; Cerling 1999), it is clear that an analysis of C_4-specific genes and gene families will provide substantial insight into the genetic basis for a unique, relatively rapid type of metabolic and anatomical evolution. In the next few sections, I provide a review of some of the most important features of C_4-specific genes, and I attempt to interpret these features within the context of the processes involving gene duplication and neofunctionalization. Additional reviews of the evolutionary aspects of C_4 genes are by Ku et al. (1996) and Monson (1999).

PEP Carboxylase

The enzyme phosphoenolpyruvate (PEP) carboxylase (PEPc) catalyzes the irreversible β-carboxylation of PEP to form oxaloacetic acid (OAA) and, in the process, releases inorganic phosphate (Pi). PEPc genes are found in eubacteria, cyanobacteria, green algae, and vascular plants, but to date, they have not been isolated from fungi and animals. It is likely that the principal and most ancient role of PEPc is as an anaplerotic enzyme, catalyzing the production of OAA for use as a respiratory substrate, although a recent theory has been forwarded to indicate a primary role for PEPc in balancing reductant levels between chloroplasts and the cytosol (Scheibe 1990; Backhausen et al. 1994). In plants, PEPc exists as multigene families, with different isoforms taking on specific roles, including the regulation of ion balance in root cortical cells and stomatal guard cells, the production of amino-group acceptor molecules in symbiotic nitrogen fixation, and the initial fixation of inorganic C (as HCO_3^-) in C_4 photosynthesis and Crassulacean acid metabolism (O'Leary 1982; Kluge 1983; Chollet et al. 1996; Gehring et al. 1998). In C_4 plants, ancestral C_3 isoforms of PEPc coexist with the more advanced C_4 isoforms (Backhausen et al. 1994).

In many ways, PEPc epitomizes the evolutionary diversification that is possible through gene duplication and subsequent neofunctionalization; multigene families are present, with enzyme isoforms specialized to not only represent the original anaplerotic function but also several novel functions. Insight into the pattern of gene duplication and neofunctionalization can be gained by studying phylogenetic affinities among the different PEPc isoforms (fig. 2). Most of the divergence between the C_3 and C_4 roles for PEPc has occurred

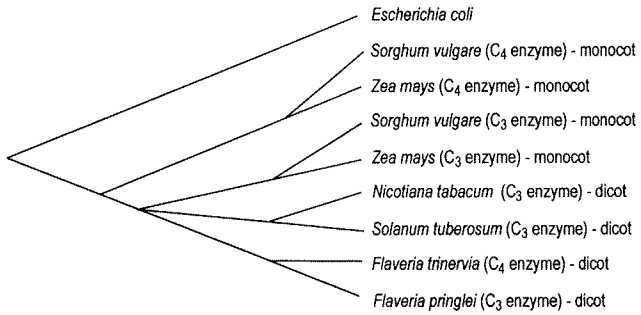

Fig. 2 Phylogenetic representation of various PEP carboxylase isoforms that have been characterized by amino acid sequence analysis. It appears that the C_4 PEPc genes in the monocots *Zea mays* and *Sorghum vulgare* share a common C_3 ancestor gene, and the evolution of C_4 PEPc has occurred independently in monocots and eudicots. Redrawn and synthesized from past studies, including Toh et al. (1994) and Gehring et al. (1998).

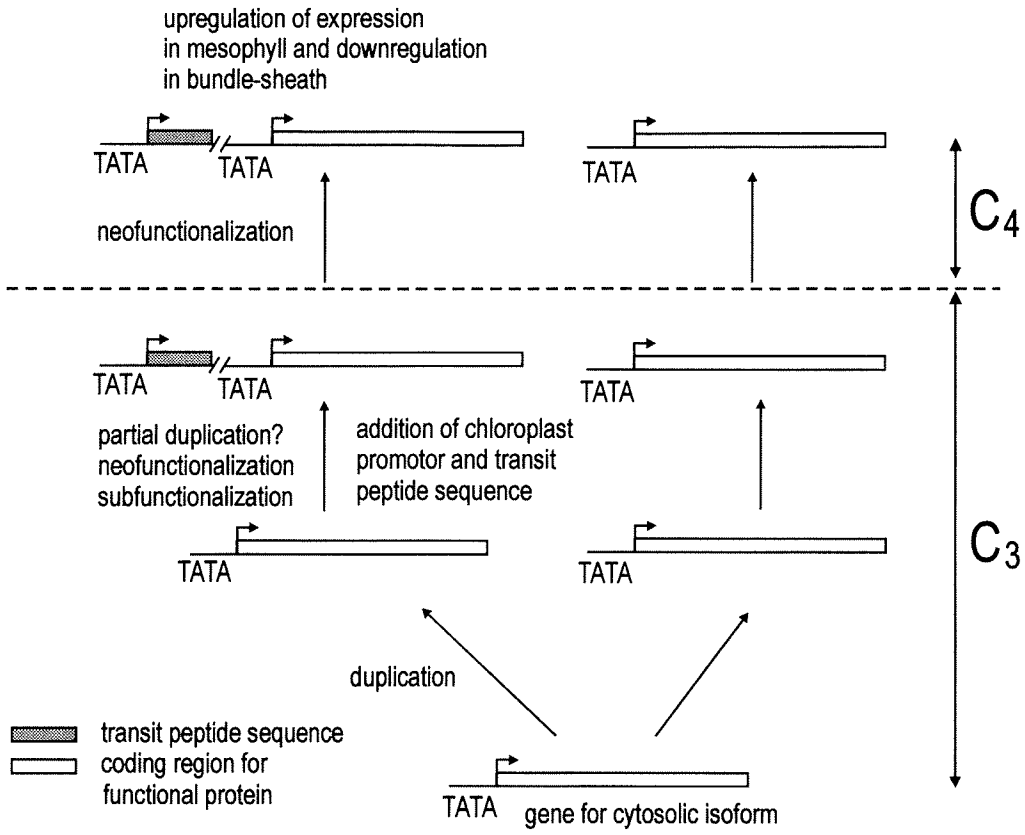

Fig. 3 Evolutionary scheme showing the duplication, subfunctionalization, and neofunctionalization for the pyruvate, Pi dikinase (PPdk) genes in the C_4 monocot, *Zea mays*. Note that the duplication and a significant portion of the neofunctionalization occurs while still in the C_3 state. Adapted from Sheen (1991).

after major taxonomic splits, as evidenced by greater intrageneric similarity than intergeneric similarity. In the C_4 monocots that have been examined (maize and sorghum), there is a single C_4 PEPc gene (Kawamura et al. 1992), whereas in the C_4 eudicot genus *Flaveria*, a small group of C_4 PEPc genes has appeared (Poetsch et al. 1991). It is not currently clear whether the *Flaveria* C_4 PEPc subgroup is monophyletic or polyphyletic with regard to C_3 ancestor genes.

The existence of a multigene C_4 PEPc subgroup in *Flaveria*, combined with the likelihood that eudicot C_4 species are relatively recent in their evolutionary origin, indicates that gene duplication and neofunctionalization has occurred relatively quickly in this group, at least compared with the oft-quoted scale of millions of years for the time between duplication and pseudogene formation in a variety of organisms (Lynch and Conery 2000). Rapid neofunctionalization in *Flaveria* is likely from the focused effect of a few key mutations in regulatory elements of the genes rather than numerous mutations in the coding region. There are only five C_4-specific differences in the amino acid sequence of the PEPc's of *F. trinervia* (C_4) and *F. pringlei* (C_3) (Hermans and Westhoff 1992). In contrast, mutations in the promoter region appear to have caused numerous changes in C_4 PEPc gene expression, including differential expression between bundle-sheath and mesophyll cells, higher levels of expression in mesophyll cells, and light inducibility during leaf development (Gilmartin et al. 1990; Hermans and Westhoff 1992).

Pyruvate, Orthophosphate Dikinase

Pyruvate Pi dikinase (PPdk) catalyzes the production of PEP from pyruvate and ATP. The ancestral role of the enzyme appears to be in the anaplerotic production of PEP for amino acid biosynthesis (Aoyagi and Bassham 1984). In C_4 species, PPdk takes on an additional, unique C_4 role in producing the large amounts of PEP required by PEPc in mesophyll cells. The enzyme exists as cytosolic and chloroplast forms in both C_3 and C_4 species (Glackin and Grula 1990; Matsuoka 1995). Three different PPdk enzymes can be found in the C_4 monocot, maize, two of which are cytosolic and carry out the ancestral role, and one of which is chloroplastic and carries out the unique C_4 role (Sheen 1991). In the C_4 eudicot, *F. trinervia*, two PPdk enzymes are found, the chloroplastic C_4 enzyme and a single cytosolic C_3 enzyme (Roesche et al. 1995; Roesche and Westhoff 1995).

In both C_3 and C_4 plants, including monocots and eudicots, the chloroplastic gene and at least one of the cytosolic genes utilize the same coding region but have overlapped promoters (fig. 3). In maize, the chloroplastic promoter overlaps one of the cytosolic promoters, but one of the cytosolic genes exists

as a nonoverlapped single gene. These unique arrangements are best explained by the restricted duplication of promoter elements and perhaps a short segment of the coding region (which subsequently mutated into the transit peptide sequence) rather than the entire gene (fig. 3). Subsequent neofunctionalization would have occurred through mutations in the duplicated promoter, leading to chloroplast-specific expression. Given that the overlapped form of the promoters occurs in both C_3 and C_4 species, and monocots and eudicots alike, the initial duplication and subsequent neofunctionalization must have occurred in the C_3 state, well before the appearance of C_4 photosynthesis. Subsequent evolution of the C_4 chloroplastic PPdk would have occurred through changes in the chloroplastic promoter, without prior duplication, leading to differential upregulation in the C_4 mesophyll cell chloroplast.

As in C_4 PEPc, neofunctionalization to provide a specific C_4 role for PPdk is capable of occurring with a relatively few number of mutational events; neofunctionalization may appear more rapid than appears justified because the initial steps following promoter duplication occurred while still in the C_3 state. Subsequent C_4 evolution required changes only in the differential regulation of expression between mesophyll and bundle-sheath cells and upregulation of expression in mesophyll cells. In maize, the region of the C_4 chloroplast promoter involved in the upregulation of expression in mesophyll cells is only 19 nucleotides long (Matsuoka and Numazawa 1991), indicating that few mutations are required to evolve a C_4-specific pattern of expression.

Evolutionary patterns in PPdk genes may provide novel insight into the origins of a subfunctionalized gene. The overlapped state of the chloroplastic and cytosolic promoters in the PPdk gene is a clear example of subfunctionalization; i.e., a single gene is controlling expression of a protein in two functional roles. It is likely that the subfunctionalized form of the gene appeared after duplication and neofunctionalization of a nonoverlapped cytosolic PPdk gene, which still exists in maize (fig. 3). The combined advantages of chloroplastic and cytosolic expression of PPdk from the new, overlapped gene may have allowed the ancestral, nonoverlapped gene to be silenced in some genera (i.e., in *Flaveria* where the nonoverlapped gene is absent). In evolving the novel C_4-specific function for the chloroplastic PPdk, the existence of the subfunctionalized gene appears to have been a key factor since it allowed upregulation of the chloroplastic function, while retaining normal levels of expression for the cytosolic function.

NADP-Malic Enzyme

The oxidative breakdown of malate to form CO_2 and pyruvate, and to provide NADPH reductant, is catalyzed by NADP-malic enzyme (NADPme). In its ancestral non-C_4, "housekeeping" role, NADPme fulfills a variety of functions (Drincovich et al. 2001), including the provision of carbon skeletons for ammonia assimilation (Chopra et al. 2002) and reductant for wound-induced production of lignin and flavonoids (Casati et al. 1999; Maurino et al. 2001). In its derived C_4 role, NADPme is isolated in bundle-sheath cell chloroplasts where it catalyzes the decarboxylation of malate, thus feeding CO_2 to the reductive pentose phosphate pathway. In taking on this C_4-specific role, modifications had to occur to the NADPme gene to provide for differential expression in a specific cell type and upregulation of gene expression to provide relatively high enzyme activities.

Within C_3, C_4, and C_3-C_4 intermediate *Flaveria* species, three isoforms of NADPme have been identified, one of which is cytosolic and two of which are chloroplastic (Drincovich et al. 1998; Lai et al. 2002). Gradients in expression can be found with respect to photosynthetic pathway type, with C_3 species predominantly expressing the cytosolic (72-kD) form, C_3-C_4 intermediate species predominantly expressing both of the chloroplastic (64-kD and 62-kD) forms, and C_4 species predominantly expressing only one of the chloroplastic (62-kD) forms (Drincovich et al. 1998). The fact that genes for both chloroplastic forms are found in C_3 species, and that the gene sequences for the transit peptides of the C_3 and C_4 chloroplastic enzymes are nearly identical, supports the hypothesis that duplication of the ancestral gene for the cytosolic 72-kD isoform was followed by neofunctionalization to a novel gene for one of the chloroplastic forms and still another duplication and modification to form the other chloroplastic form, all in the C_3 state. The evolution of C_4 photosynthesis, then, required modification of one or both of the chloroplastic forms to provide the C_4-specific pattern of expression. The fact that both chloroplastic forms are upregulated in C_3-C_4 intermediate species, but only one form (the 62-kD form) is upregulated in C_4 species, indicates that an initial evolutionary step occurred through the enhanced expression of both available chloroplastic genes, followed by eventual domination in the expression of one gene over the other (Lai et al. 2002).

The modification in gene sequence required for neofunctionalization of the two chloroplastic forms for C_4-specific roles is mostly restricted to regulatory, untranscribed regions at both the 3′ and 5′ extremes. The cytosolic and chloroplastic NADPme genes are 75% identical in the transcribed region but only 29%–36% identical in the 3′ untranscribed region; even the two chloroplastic forms exhibit low identity (39%) in the 3′ untranscribed region (Lai et al. 2002). Observations with chimeric constructs of the 3′ and 5′ ends of the chloroplastic genes support the hypothesis that elements in the 5′ untranscribed region control cell-specific expression patterns and elements in the 3′ untranscribed region act as expression enhancers to upregulate NADPme activity (Marshall et al. 1997; Taylor et al. 1997). Clearly, the evolution of unique NADPme expression patterns in C_4 taxa was facilitated by the fact that most of the gene duplication and neofunctionalization occurred in C_3 ancestors, with more restricted modification occurring in regulatory elements to provide unique C_4 capabilities.

Carbonic Anhydrase

Carbonic anhydrase (CA) is crucial to facilitating rapid equilibrium between the CO_2 and HCO_3^- forms of inorganic carbon and thus for relaxing the potential for diffusional limitations in the supply of CO_2 as a substrate for Rubisco or HCO_3^- as a substrate for PEPc. In C_3 chloroplasts, where Rubisco catalyzes the primary assimilation of atmospheric CO_2, CA is present at high activities. In C_4 mesophyll cells, the expression of CA activity is downregulated in chloroplasts since Rubisco is absent but upregulated in the cytosol, where

PEPc catalyzes the assimilation of HCO_3^-. There are two separate genes for CA, one that contains the code for a functional transit peptide and produces the chloroplastic form and one that lacks a functional transit peptide and produces the cytosolic form. Both genes are present in C_3, C_3-C_4 intermediate, and C_4 species of *Flaveria* (Ludwig and Burnell 1995). The length of cDNAs from the two genes in *Flaveria* species is similar, leading to the hypothesis that the chloroplastic gene is ancestral and the cytosolic gene appeared following duplication and neofunctionalization in which mutations rendered the transit peptide sequence ineffectual (fig. 4). These events presumably occurred while still in the C_3 state. During C_4 evolution, further neofunctionalization without further duplication took place through mutation in the regulatory regions of both genes. This resulted in the downregulation of expression for the chloroplast gene and the upregulation of expression for the cytosolic gene.

NADP Malate Dehydrogenase

NADP malate dehydrogenase (NADPmdh) uses reducing power from NADPH in the chloroplasts of C_3 and C_4 mesophyll cells to convert oxaloacetate (OAA) to malate. In C_3 mesophyll cells, this process helps maintain the reductant balance between chloroplasts and the cytosol, and it may act as a "safety valve" to use excess NADPH during periods when the photosynthetic light reactions and reductive pentose phosphate pathway are out of balance (Scheibe 1990; Backhausen et al. 1994). In C_4 mesophyll cells, NADPmdh has taken on the role of converting the OAA produced by PEPc, following the initial fixation of atmospheric CO_2 to malate, just prior to the transport of malate to bundle-sheath cells where it is decarboxylated. In this C_4 role, the activity of NADPmdh is upregulated in mesophyll cell chloroplasts and downregulated in bundle-sheath cell chloroplasts. Two different genes for NADPmdh have been isolated from the C_4 monocots, sorghum and maize, one of which is expressed at low levels and likely represents the gene controlling the C_3-specific role and one which is expressed at higher levels and presumably represents the C_4-specific role (Luchetta et al. 1991). Thus, it is likely that gene duplication followed by neofunctionalization occurred to produce the C_4-specific gene. It is not clear if both of these genes are present in the C_3 ancestors to maize and sorghum, so it cannot be determined at this time whether neofunctionalization resulted in the production of a C_3 gene or occurred specifically during the evolution of C_4 photosynthesis. In the C_4 eudicot, *Flaveria trinervia*, a single gene has been isolated for NADPmdh (McGonigle and Nelson 1995), raising the possibility of neofunctionalization without prior duplica-

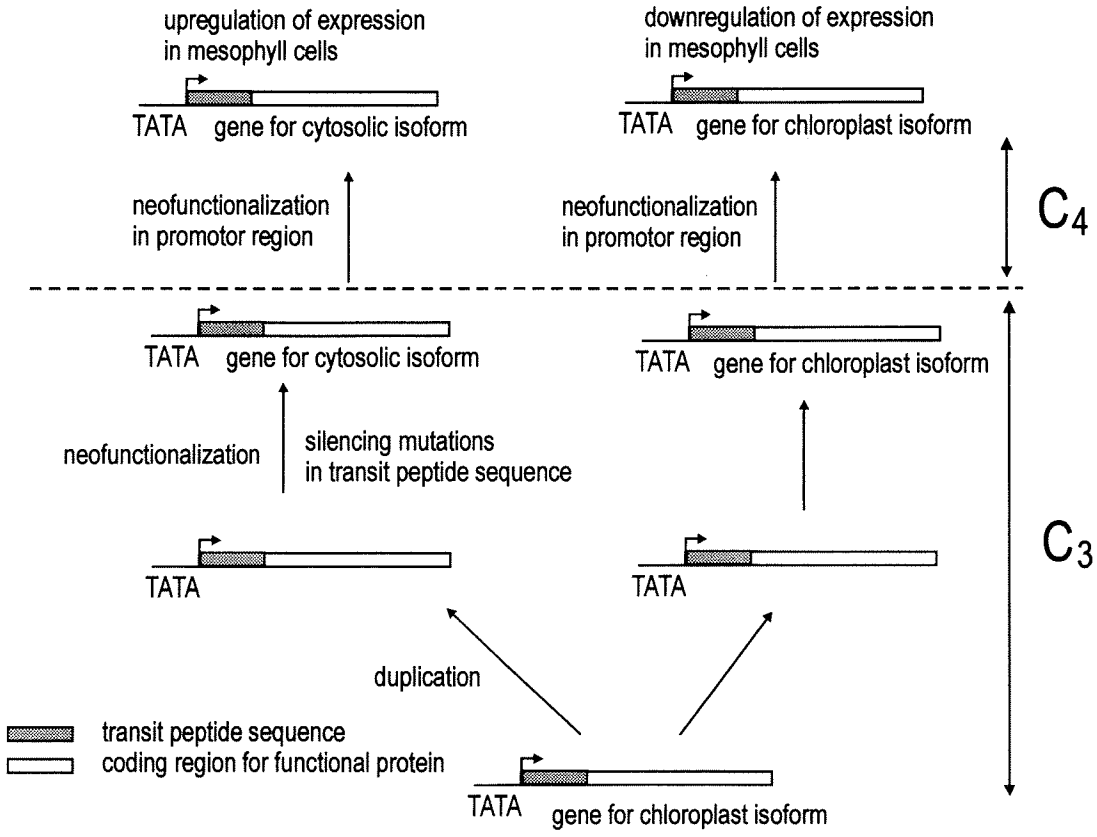

Fig. 4 Evolutionary scheme showing the duplication and neofunctionalization of carbonic anhydrase (CA) genes in *Flaveria* sp. Following duplication, neofunctionalization occurs while still in the C_3 state through silencing mutations in the transit peptide sequence of one duplicate, ultimately leading to the cytosolic isoform. Neofunctionalization for the C_4 role occurs through mutations in the promoter region and the evolution of differential expression patterns. Adapted from Ludwig and Burnell (1995).

tion or neofunctionalization with prior duplication followed by silencing of one member of the duplicates. In either example, it might be that the upregulation of NADPmdh for its C_4 role produced enough activity in C_4 mesophyll chloroplasts to accommodate its original C_3 role as well, eliminating the need for two loci.

Ribulose 1,5-Bisphosphate Carboxylase/ Oxygenase (Rubisco)

Rubisco catalyzes the assimilation of CO_2 in the mesophyll cells of C_3 plants and the bundle-sheath cells of C_4 plants. Rubisco is a multimeric enzyme composed of large, chloroplast-encoded and small, nuclear-encoded subunits. In C_4 leaves, expression of Rubisco is downregulated in mesophyll cells and upregulated in bundle-sheath cells. Downregulation of Rubisco large subunits in C_4 mesophyll cells may result from posttranscriptional processes (Schäffner and Sheen 1991; Boinski et al. 1993). Differential expression of small subunits results from transcriptional control of the nuclear gene involving interactions between a 3′ suppressor sequence and a 5′ promoter sequence (Viret et al. 1994), although posttranscriptional control may also be involved in differential expression of the small subunit (Schäffner and Sheen 1991). The *cis*-acting elements that control suppression or nonsuppression of the Rubisco small subunit gene appear to be influenced by a mesophyll cell-specific, regulatory protein (i.e., some type of *trans*-acting element) (Nomura et al. 2000). Small subunit expression results from a multigene family in some species (Gutteridge and Gatenby 1995). It is currently not known whether small subunit, multigene families are common in C_4 species. If C_4 ancestor species possessed small subunit, multigene families, then the evolutionary involvement of a *trans*-acting protein may make sense, since it would allow for effective, simultaneous downregulation of multiple gene copies. In any case, the current state of knowledge does not permit conclusions concerning whether gene duplication and subsequent neofunctionalization had an important role in the evolution of the differential expression of Rubisco in C_4 species.

Gene Duplication and Neofunctionalization as Constraints to C_4 Evolutionary Patterns

Gene Duplication and Neofunctionalization for C_3 Roles

It is clear that gene duplication preceded the evolution of many of the novel biochemical steps involved in C_4 photosynthesis. Many of the duplicated genes destined for C_4 roles, however, were shaped by neofunctionalization and selection to produce novel C_3 roles, prior to being co-opted for C_4 photosynthesis. This conclusion is supported by several genes that eventually take on C_4 roles that are found in both C_3 and C_4 congeners (e.g., chloroplastic PPdk, NADPme, mesophyll CA). In most cases, it appears that further duplication was not required when novel C_3 genes were finally co-opted for C_4 roles. For example, chloroplastic PPdk and NADPme appear to have been upregulated for C_4 photosynthesis through direct modification of promoter regions in C_3 genes, without additional duplication. Apparently, the C_4-dependent upregulation of activity resulted in sufficient activity to handle both the new C_4 role and the more primitive C_3 role.

The original duplication and neofunctionalization of genes for divergent C_3 roles, and retention for later neofunctionalization into C_4 roles, are probably key processes that have allowed C_4 photosynthesis to appear so quickly in response to reduced atmospheric CO_2 concentration and to have evolved so many times independently. The requirement to build up a reservoir of duplicated and neofunctionalized C_3 genes, in support of C_4 evolution, might also explain why C_4 photosynthesis appeared only 20–30 Ma, despite decreases in atmospheric CO_2 concentration which started at least 100 Ma. The time lag between the initial decreases in CO_2 concentration and the appearance of C_4 photosynthesis may be partly related to genetic limitations.

There would appear to be two possible paths by which a complex metabolic pathway such as C_4 photosynthesis could have been constructed through the evolutionary process. In the first, progenitor C_4 genes would have been duplicated and neofunctionalized in sequence, with one neofunctionalization triggering selective pressures to drive subsequent neofunctionalization. In the second, genetic recruitment may have occurred from a reservoir of previously duplicated and neofunctionalized genes that had no connections in the ancestral C_3 state, in other words, a reservoir of genes that had been neofunctionalized for independent functions. In either case, a clear prerequisite would have been a preexisting population of several coexisting, but different, duplicated genes. The existence of such a prerequisite places a potential genetic constraint on the evolution of C_4 photosynthesis; those groups capable of fostering the emergence of C_4 photosynthesis would have to be capable of accumulating a reservoir of duplicated, progenitor C_4 genes. One issue worth further exploration concerns the population and life history characteristics that might be required to establish a reservoir of duplicated genes capable of being co-opted for C_4 function.

Population and Life History Considerations

Ultimately, the rate by which duplicated genes appear in a population will depend on the sexual recruitment rate. Thus, life history traits related to reproduction and survival rates will have an important effect on the frequency at which novel traits can evolve through gene duplication. One life history trait related to the rate of sexual recruitment, and likely to affect the rate of gene duplication and subsequent neofunctionalization, is generation time. Processes such as gene duplication are likely to scale with generation time rather than chronological time (Nei 1975). Thus, populations of annuals or short-lived perennials would be expected to foster more frequent appearance of duplicate genes than populations of long-lived perennials. At this time, it is not clear whether mutation rates after duplication, which affect neofunctionalization, are most dependent on generation time (i.e., tied to the meiotic process) or chronological time (i.e., independent of the meiotic process). However, the frequency with which new alleles produced through nondeleterious mutations are exposed to selection should be directly dependent on generation time; thus, a population of annual plants will cause new mutations and mutation combinations to emerge and face selection each year, whereas long-lived perennial plants can hide mutations for long periods of time. From such reasoning, it is clear that

species with short generation times may foster the more frequent appearance of gene duplicates and tighter coupling of neofunctionalized alleles to selection. This combination of traits may have favored the evolution of pathways that depend so heavily on gene duplication, such as C_4 photosynthesis, in plants with short generation times. Having come to this conclusion, it should be noted that nonfunctionalization through the accumulation of deleterious mutations probably also scales with generation time (Nei 1975). Thus, while short-lived plants may have higher frequencies of gene duplication, they probably also have higher rates of nonfunctionalization of those duplicated genes. The exact manner in which the competing processes of neofunctionalization and nonfunctionalization are resolved with respect to generation time is not clear at this time.

Large populations are expected to increase the probability of neofunctionalization by (1) providing a larger reservoir of preexisting heterozygous alleles at the original locus, which provides an immediate, positive selective advantage to the newly arisen gene duplicate, and (2) providing more time for neofunctionalization, since silencing mutations must work on a larger set of duplicated alleles to cause their ultimate extinction from the population. Past modeling efforts have revealed that the mean number of generations required for nonfunctionalization of a duplicate gene is dependent on population size (fig. 5), supporting the hypothesis that more time will be available for neofunctionalization in large populations. Other factors that are likely to affect rates of neofunctionalization are (1) the intrinsic properties of the duplicated gene, including any selective advantages it brings with it, (2) the rate by which new mutations appear in the gene, and (3) the probability by which those mutations cause a beneficial change in the gene's function (i.e., impart some positive selective value to the gene), as opposed to causing a neutral or detrimental change.

From this initial analysis, it would appear that taxa with relatively large populations, frequent rates of sexual recruitment, and short generation times would have an advantage in terms of accumulating a reservoir of duplicated and neofunctionalized genes. All three of these expectations can be used to propose that C_4 photosynthesis should most frequently appear in ruderal, probably herbaceous, species with short generation times and within widespread species that include large numbers of interbreeding individuals.

Two of the most widely discussed enigmas of C_4 evolution concern its infrequent appearance in eudicots, compared with monocots, and its lack of representation in trees from closed forest-canopy ecosystems. Past explanations for these phenomena have relied on a foundation of plant carbon balance (Ehleringer et al. 1997; Monson 1999); both, however, might be explained by the absence of adequate conditions to promote the retention of duplicated genes. At this time, a comprehensive set of comparative data on the population and life history traits of C_3 and C_4 species is not available; however, anecdotal evidence does exist, most of which supports the theory. The absence of C_4 photosynthesis from forest trees is consistent with general aspects of the theory in that forest trees tend to be long lived and, therefore, there are fewer opportunities for recruitment and the exposure of mutations to new selection regimes. Herbaceous species, especially those exhibiting the ruderal growth habit, would generally exhibit more frequent

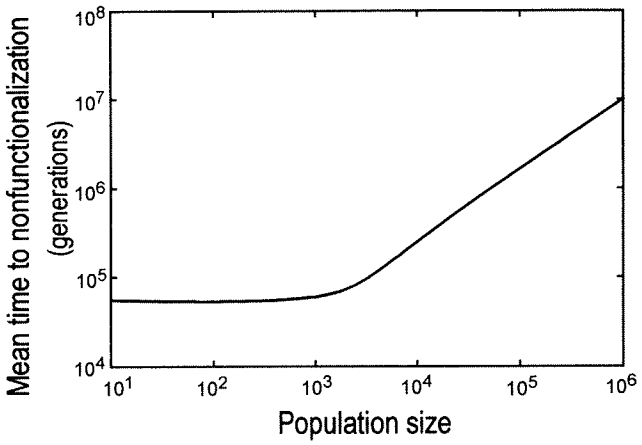

Fig. 5 Graph showing the mean time to nonfunctionalization following gene duplication as a function of population size. The line was generated using the diffusion-theory model of Watterson (1983), as presented by Lynch and Force (2000).

recruitment and shorter generation times. These aspects of tree versus herb comparisons are often reflected in the *K*- versus *r*-selection theories that have been proposed as organizing principles in understanding the evolution of plant life history traits; forest trees are often offered as the exemplar of *K*-selected species, whereas ruderal herbs are often offered as the typical *r*-selected species. Among the herbs, there is general support that the grasses, which harbor the largest diversity of C_4 species, occur in relatively large populations compared with most eudicot groups; the grasses tend to occur in widespread graminoid-dominated ecosystems, and they reproduce quite successfully with wind pollination, a trait that must be accompanied by large population size and relatively high plant density. The generally smaller population sizes of eudicot species may expose them to significant influences by random genetic drift, and this may have limited the capacity for eudicots to develop the reservoirs of duplicated genes required for the evolution of C_4 photosynthesis; i.e., eudicot species with small population sizes may have experienced "bottlenecks" with regard to the accumulation of duplicated genes and the evolution of C_4 photosynthesis. In this regard, it is interesting that the eudicot family with the largest diversity of C_4 species, the Amaranthaceae, is geographically widespread with large numbers of ruderal (*r*-selected) genera.

One trait in perennial grasses that does not appear to support the theory is their clonal growth habit and persistence for decades or centuries with minimal apparent sexual reproduction. Opportunities for the introduction of genetic diversification should be rather limited in this situation. In many regards, the frequent appearance of C_4 photosynthesis in long-lived perennial grasses and its absence in long-lived forest tree species is enigmatic and could be used to argue against any role for longevity in the evolution of C_4 photosynthesis. Past studies, however, have revealed levels of genetic diversity within the grasses that are as high as those in aclonal eudicot species (Ellstrand and Roose 1987). It may be that opportunities for sexual reproduction are more frequent than expected in the grasses. Additionally, the rapid expansion of global

grasslands during the Miocene (Pascual and Jaureguizar 1990; Retallack 1992; Kennet 1995) may have provided unique opportunities for sexual recruitment and gene duplication during the exact period when atmospheric CO_2 concentration decreased below the 500 ppmv threshold, which appears to be so crucial for grass carbon balance (Ehleringer et al. 1997). The combination of rapid population expansion, frequent episodes of sexual recruitment, and low atmospheric CO_2 concentration may underlie the frequent appearance of C_4 photosynthesis in perennial grasses, and the absence of this combination may explain the absence of C_4 photosynthesis in forest trees.

Conclusions

Most past theories attempting to explain the distribution of C_4 photosynthesis focus on interactions between CO_2 assimilation and environmental gradients in temperature and precipitation. Such theories are founded on the assumption that the primary factor controlling C_4 distribution is the capacity to maintain a favorable carbon balance in the face of various environmental constraints. In this article, I present the view that one perspective that has been ignored in the development of C_4 evolutionary theory is the possible constraint by genetic processes that must have underlain the appearance of novel C_4 genes. It is clear that gene duplication and subsequent neofunctionalization are critical processes to the appearance of C_4 photosynthesis. Species that are characterized by certain population attributes, such as large population size, short generation time, and frequent recruitment of sexually produced individuals, are expected to exhibit the most frequent appearance of duplicated genes, and thus they have the highest probability of producing neofunctionalized traits. When this perspective is included in the analysis of C_4 diversity and taxonomic distribution, it may begin to explain certain patterns that have eluded past theories, such as why C_4 photosynthesis is less frequent in eudicot taxa and why C_4 photosynthesis is absent from forest trees.

It is important to note that the theory presented here is not meant to replace the assumption that the dominant factor driving the evolution of C_4 photosynthesis over the past 20–30 million years has been the pressure to maintain favorable carbon balance in an atmosphere of progressively lower CO_2 concentrations. The current theory is, instead, meant to complement past theories by pointing out that genetic constraints may have come into play in determining which species can respond to the reduced CO_2 concentration by evolving the novel anatomical and biochemical aspects of C_4 photosynthesis. Clearly, the adaptive physiological features of C_4 photosynthesis as a syndrome must be put forward as the dominant evolutionary drivers; the genetic constraints proposed here might merely be the filters that determine whether the genetic materials were available to permit biochemical adaptation to a progressively changing atmosphere.

Acknowledgments

I am grateful for the very insightful discussions with Drs. Jeffry Mitton, Michael Grant, Herman Heilmeier, Ned Friedman, Helmut Freitag, and Deborah Goldberg. My synthesis draws heavily on theory developed by Dr. Michael Lynch (Indiana University) and his colleagues; I must acknowledge their innovative work as the true foundation for the synthesis. Ultimately, the opinions and speculations of this article, however, are mine, and I take full responsibility for current and future errors, as well as all overstatements and oversights.

Literature Cited

Aoyagi K, JA Bassham 1984 Pyruvate orthophosphate dikinase mRNA organ specificity in wheat and maize. Plant Physiol 76: 278–280.

Arntz AM, LF Delph 2001 Pattern and process: evidence for the evolution of photosynthetic traits in natural populations. Oecologia 127:455–467.

Averof M 2002 Arthropod hox genes: insights on the evolutionary forces that shape gene functions. Curr Opinion Gen Dev 12: 386–392.

Backhausen JE, C Kitzmann, R Scheibe 1994 Competition between electron acceptors in photosynthesis: regulation of the malate valve during CO_2 fixation and nitrate reduction. Photosynth Res 42: 75–86.

Boinski JJ, JL Wang, P Xu, T Hotchkiss, JO Berry 1993 Post-transcriptional control of cell type–specific gene expression in bundle sheath and mesophyll chloroplasts of *Amaranthus hypochondriacus*. Plant Mol Biol 22:397–410.

Bruce AE, AC Oates, VE Prince, RK Ho 2001 Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate *hoxB5* genes. Evol Dev 3:127–144.

Casati P, MF Drincovich, GE Edwards, CS Andreo 1999 Malate metabolism by NADP-malic enzyme in plant defense. Photosynth Res 61:99–105.

Cerling TE 1999 Paleorecords of C_4 plants and ecosystems. Pages 445–469 *in* RF Sage, RK Monson, eds. C_4 plant biology. Academic Press, San Diego, Calif.

Cerling TE, JM Harris, BJ MacFadden, MG Leakey, J Quade, V Eisenmann 1997 Global vegetation change through the Miocene-Pliocene boundary. Nature 389:153–158.

Cerling TE, Y Wang, J Quade 1993 Expansion of C_4 ecosystems as an indicator of global ecological change in the late-Miocene. Nature 361:344–345.

Chollet R, J Vidal, MH O'Leary 1996 Phosphoenolpyruvate carboxylase: a ubiquitous, highly-regulated enzyme in plants. Annu Rev Plant Physiol Plant Mol Biol 47:273–298.

Chopra J, N Kaur, AK Gupta 2002 A comparative developmental pattern of enzymes of carbon metabolism and pentose phosphate pathway in mung bean and lentil nodules. Acta Physiol Plant 24: 67–72.

Cowling SA 2001 Plant carbon balance, evolutionary innovation and extinction in land plants. Global Change Biol 7:231–239.

DeMartino S, Y-L Yan, T Jowett, JH Postlewaite, ZM Varga, A Ashworth, CA Austin 2000 Expression of *sox11* gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. Dev Dyn 217:2471–2484.

Drincovich MF, P Casati, CS Andreo 2001 NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. FEBS Lett 490:1–6.

Drincovich MF, P Casati, CS Andreo, SJ Chessin, VR Franceschi, GE

Edwards, MSB Ku 1998 Evolution of C_4 photosynthesis in *Flaveria* species: isoforms of NADP-malic enzyme. Plant Physiol 117: 733–744.

Ehleringer JR 1978 Implications of quantum yield differences to the distributions of C_3 and C_4 grasses. Oecologia 31:255–267.

Ehleringer JR, O Björkman 1977 Quantum yields for CO_2 uptake in C_3 and C_4 plants: dependence on temperature, CO_2 and O_2 concentrations. Plant Physiol 59:86–90.

Ehleringer JR, TE Cerling, BR Helliker 1997 C_4 photosynthesis, atmospheric CO_2 and climate. Oecologia 112:285–299.

Ehleringer JR, RK Monson 1993 Evolutionary and ecological aspects of photosynthetic pathway variation. Annu Rev Ecol Syst 24: 411–439.

Ehleringer JR, RW Pearcy 1983 Variation in quantum yields for CO_2 uptake among C_3 and C_4 plants. Plant Physiol 73:555–559.

Ehleringer JR, RF Sage, LB Flanagan, RW Pearcy 1991 Climate change and the evolution of C_4 photosynthesis. Trends Ecol Evol 6: 95–99.

Ellstrand NC, ML Roose 1987 Patterns of genotypic diversity in clonal plant species. Am J Bot 74:123–131.

Friedman R, AL Hughes 2001 Gene duplication and the structure of eukaryotic genomes. Genome 11:373–381.

Gehring HH, V Heute, M Kluge 1998 Toward a better knowledge of the molecular evolution of phosphoenolpyruvate carboxylase by comparison of partial cDNA sequences. J Mol Evol 46:107–114.

Gillon J, D Yakir 2001 Influence of carbonic anhydrase activity in terrestrial vegetation on the O^{18} content of the atmospheric CO_2. Science 291:2584–2587.

Gilmartin PM, L Sarokin, J Memelink, N-H Chua 1990 Molecular light switches for plant genes. Plant Cell 2:369–378.

Glackin CA, JW Grula 1990 Organ-specific transcripts of different size and abundance derive from the same pyruvate, orthophosphate dikinase gene in maize. Proc Natl Acad Sci USA 87:3004–3008.

Gutteridge S, AA Gatenby 1995 Rubisco synthesis, assembly and regulation. Plant Cell 7:809–819.

Haldane JBS 1933 The part played by recurrent mutation in evolution. Am Nat 67:5–9.

Harrison PM, M Gerstein 2002 Studying genomes through the aeons: protein families, pseudogenes and proteome evolution. J Mol Biol 318:1155–1174.

Hermans J, P Westhoff 1992 Homologous genes for the C_4 isoform of phosphoenolpyruvate carboxylase in a C_3 and C_4 *Flaveria* species. Mol Gen Genet 224:459–468.

Hughes MK, AL Hughes 1993 Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol Biol Evol 10:1360–1369.

Huynen MA, E van Nimwegen 1998 The frequency distribution of gene family sizes in complete genomes. Mol Biol Evol 15:583–589.

Kawamura T, K Shigesada, H Toh, S Okumura, S Yanagisawa, K Izui 1992 Molecular evolution of phosphoenolpyruvate carboxylase for C_4 photosynthesis in maize: comparison of its cDNA sequence with a newly isolated cDNA encoding an isozyme involved in the anaplerotic function. J Biochem 112:147–154.

Kellogg EA 1999 Phylogenetic aspects of the evolution of C_4 photosynthesis. Pages 411–444 *in* RF Sage, RK Monson, eds. C_4 plant biology. Academic Press, San Diego, Calif.

Kennet JP 1995 A review of polar climatic evolution during the Neogene, based on the marine sediment record. Pages 49–64 *in* ES Vrba, GH Denton, TC Partridge, LH Burckle, eds. Paleoclimate and evolution. Yale University Press, New Haven, Conn.

Kimura M, JL King 1979 Fixation of a deleterious allele at one of two duplicate loci by mutation pressure and random drift. Proc Natl Acad Sci USA 76:2858–2861.

Kluge M 1983 The role of PEPC in C_4-photosynthesis and Crassulacean acid metabolism. Physiol Veg 5:817–825.

Ku MSB, Y Kano-Murakami, M Matsuoka 1996 Evolution and expression of C_4 photosynthesis genes. Plant Physiol 111:949–957.

Kuypers MMM, RD Pancost, JS Sinninghe-Damaste 1999 A large and abrupt fall in atmospheric CO_2 concentration during Cretaceous times. Nature 399:342–345.

Lai LB, L Tausta, TM Nelson 2002 Differential regulation of transcripts encoding cytosolic NADP-malic enzyme in C_3 and C_4 *Flaveria* species. Plant Physiol 128:140–149.

Luchetta P, C Crétin, P Gadal 1991 Organization and expression of the two homologous genes encoding the NADP malate dehydrogenase in *Sorghum vulgare* leaves. Mol Gen Genet 228:473–481.

Ludwig M, JN Burnell 1995 Molecular comparison of carbonic anhydrase from *Flaveria* species demonstrating different photosynthetic pathways. Plant Mol Biol 29:353–365.

Lynch M, JS Conery 2000 The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155.

Lynch M, A Force 2000 The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459–473.

Lynch M, M O'Hely, B Walsh, A Force 2001 The probability of preservation of a newly arisen gene duplicate. Genetics 159:1789–1804.

Marshall JS, JD Stubbs, JA Chitty, B Surin, WC Taylor 1997 Expression of the C_4 Me1 gene from *Flaveria bidentis* requires an interaction between 5′ and 3′ sequences. Plant Cell 9:1515–1525.

Maruyama T, N Takahata 1981 Numerical studies of the frequency trajectories in the process of fixation of null genes at duplicated loci. Heredity 46:49–57.

Matsuoka M 1995 The gene for pyruvate, orthophosphate dikinase in C_4 plants: structure, regulation and evolution. Plant Cell Physiol 36:937–943.

Matsuoka M, T Numazawa 1991 Cis-acting elements in the pyruvate, orthophosphate dikinase gene in maize. Mol Gen Genet 228: 143–152.

Maurino VG, M Saigo, CS Andreo, MF Drincovich 2001 Non-photosynthetic malic enzyme from maize: a constitutively expressed enzyme that responds to plant defence inducers. Plant Mol Biol 45: 409–420.

Mazet F, SM Shimeld 2002 Gene duplication and divergence in the early evolution of vertebrates. Curr Opin Genet Dev 12:393–396.

McGonigle B, T Nelson 1995 C_4 isoform of NADP malate dehydrogenase: cDNA cloning and expression in leaves of C_4, C_3 and C_3-C_4 intermediate species of *Flaveria*. Plant Physiol 108: 1119–1126.

McGrath JM, MM Jancso, E Pichersky 1993 Duplicate sequences with a similarity to expressed genes in the genome of *Arabidopsis thaliana*. Theor Appl Genet 86:880–888.

Monson RK 1989 On the evolutionary pathways resulting in C_4 photosynthesis and Crassulacean acid metabolism. Adv Ecol Res 19: 57–110.

——— 1999 The origins of C_4 genes and evolutionary pattern in the C_4 metabolic phenotype. Pages 377–410 *in* RF Sage, RK Monson, eds. C_4 plant biology. Academic Press, San Diego, Calif.

Monson RK, BD Moore, MSB Ku, GE Edwards 1986 Co-function of C_3 and C_4 photosynthetic pathways in C_3, C_4 and C_3-C_4 intermediate *Flaveria* species. Planta 168:493–502.

Monson RK, S Rawsthorne 2000 CO_2 assimilation in C_3-C_4 intermediate plants. Pages 533–550 *in* RC Leegood, TD Sharkey, S von Caemmerer, eds. Photosynthesis: physiology and metabolism. Kluwer Academic, Dordrecht.

Muller HJ 1936 Bar duplication. Science 83:528–530.

Nambudiri EMV, WD Tidwell, BN Smith, NP Hebbert 1978 A C_4 plant from the Pliocene. Nature 276:816–817.

Nei M 1975 Molecular population genetics and evolution. North-Holland, Amsterdam. 288 pp.

Nei M, AK Roychoudhury 1973 Probability of fixation of nonfunctional genes at duplicate loci. Am Nat 107:362–372.

Nomura M, K Katayama, A Nishimura, Y Ishida, S Ohta, T Komari, M Miyao-Tokutomi, S Tajima, M Matsuoka 2000 The promoter of rbcS in a C_3 plant (rice) directs organ-specific, light-dependent

expression in a C_4 plant (maize), but does not confer bundle-sheath cell-specific expression. Plant Mol Biol 44:99–106.

Ohno S 1970 Evolution by gene duplication. Springer, Berlin.

Ohta T 1987 Simulating evolution by gene duplication. Genetics 115: 207–213.

——— 1988 Time for acquiring a new gene by duplication. Proc Natl Acad Sci USA 85:3509–3512.

——— 2000 Evolution of gene families. Gene 259:45–52.

O'Leary MH 1982 Phosphoenolpyruvate carboxylase: an enzymologist's view. Annu Rev Plant Physiol 33:297–315.

Pascual R, EO Juareguizar 1990 Evolving climates and mammal faunas in Cenozoic South America. J Hum Evol 19:23–60.

Pearcy RW, J Troughton 1975 C_4 photosynthesis in tree form *Euphorbia* species from Hawaiian rainforest sites. Plant Physiol 55: 1054–1056.

Poetsch W, J Hermans, P Westhoff 1991 Multiple cDNAs of phosphoenolpyruvate carboxylase in the C_4 eudicot *Flaveria trinervia*. FEBS Lett 292:133–136.

Rawsthorne S 1992 C_3-C_4 intermediate photosynthesis: linking physiology to gene expression. Plant J 2:267–274.

Rettallack GJ 1992 Middle Miocene fossil plants from Fort Ternan (Kenya) and evolution of African grasslands. Paleobiology 18: 383–400.

Roesche E, M Streubel, P Westhoff 1995 Primary structure of the photosynthetic pyruvate orthophosphate dikinase of the C_3 plant *Flaveria pringlei* and expression analysis of pyruvate orthophosphate dikinase sequences in C_3, C_3-C_4, and C_4 *Flaveria* species. Plant Mol Biol 26:763–769.

Roesche E, P Westhoff 1995 Genomic structure and expression of the pyruvate, orthophosphate dikinase of the eudicotyledonous C_4 plant *Flaveria trinervia* (Asteraceae). Plant Mol Biol 29:663–678.

Sage RF 1999 Why C_4 photosynthesis? Pages 1–9 *in* RF Sage, RK Monson, eds. C_4 plant biology. Academic Press, San Diego, Calif.

——— 2001 Environmental and evolutionary preconditions for the origin and diversification of the C_4 photosynthetic syndrome. Plant Biol 3:202–213.

Sage RF, M Li, RK Monson 1999 The taxonomic distribution of C_4 photosynthesis. Pages 551–584 *in* RF Sage, RK Monson, eds. C_4 plant biology. Academic Press, San Diego, Calif.

Schäffner AR, J Sheen 1991 Maize rbcS promoter activity depends on sequence elements not found in eudicot rbcS promoters. Plant Cell 3:997–1012.

Scheibe R 1990 Light-dark modulation regulation of chloroplast metabolism in a new light. Bot Acta 103:327–334.

Schubert M, LZ Holland, G Panopoulou, H Lehrach, ND Holland 2000 Characterization of amphioxus *Amph/Wnt8*: insights into the evolution of patterning of the embryonic dorsoventral axis. Evol Dev 2:85–92.

Schuster WS, RK Monson 1990 An examination of the advantages of C_3–C_4 intermediate photosynthesis in warm environments. Plant Cell Environ 13:903–912.

Seoighe C, KH Wolfe 1998 Extent of genomic rearrangement after genome duplication in yeast. Proc Natl Acad Sci USA 95: 4447–4452.

Sheen J 1991 Molecular mechanism underlying the differential expression of maize pyruvate, orthophosphate dikinase genes. Plant Cell 3:225–245.

Shubin NH, CR Marshall 2000 Fossils, genes and the origin of novelty. Paleobiology 26(suppl):324–340.

Taylor WC, E Rosche, JS Marshall, S Ali, CJ Chastain, JA Chitty 1997 Diverse mechanisms regulate the expression of genes coding for C_4 enzymes. Aust J Plant Physiol 24:437–442.

Thomas BR, VS Fords, E Pichersky, LD Gottlieb 1993 Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in *Clarkia* and comparison to the single gene in *Arabidopsis*. Genetics 135:895–905.

Tidwell WD, EMV Nambudiri 1989 *Tomlinsonia thomassonii*, gen. et sp. nov., a premineralized grass from the Upper Miocene Ricardo Formation, California. Rev Palaeobot Palynol 60:165–177.

Toh H, T Kawamura, K Izui 1994 Molecular evolution of phosphoenolpyruvate carboxylase. Plant Cell Environ 17:31–43.

Van de Peer Y, JS Taylor, I Braasch, A Meyer 2001 The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. J Mol Evol 53:436–446.

Viret J-F, Y Mabrouk, L Bogorad 1994 Transcriptional photoregulation of cell type–preferred expression of maize rbc-m3: 3′ and 5′ sequences are involved. Proc Natl Acad Sci USA 91:8577–8581.

von Caemmerer S 1989 A model of photosynthetic CO_2 assimilation and carbon isotope discrimination in leaves of certain C_3-C_4 intermediates. Planta 178:463–474.

Watterson GA 1983 On the time for gene silencing at duplicate loci. Genetics 105:745–766.

Winter K 1981 C_4 plants of high biomass in arid regions of Asia: occurrence of C_4 photosynthesis in Chenopodiaceae and Polygonaceae from the Middle-East and USSR. Oecologia 48:100–106.

Wyrich R, U Dressen, S Brockman, M Streubel, C Chang, D Qiang, AH Paterson, P Westhoff 1998 The molecular basis of C_4 photosynthesis in sorghum: isolation, characterization and RFLP mapping of mesophyll- and bundle-sheath-specific cDNAs as obtained by differential screening. Plant Mol Biol 37:319–335.