J Mol Evol (1996) 42:432–442

JOURNAL OF MOLECULAR EVOLUTION

Molecular Evolution of Nitrate Reductase Genes

Jizhong Zhou,[1,*] Andris Kleinhofs[1,2]

[1] Department of Genetics and Cell Biology, Washington State University, Pullman, WA 99164-6420, USA
[2] Department of Crop and Soil Sciences, Washington State University, Pullman, WA 99164-6420, USA

Received: 10 April 1995 / Accepted: 10 September 1995

Abstract. To understand the evolutionary mechanisms and relationships of nitrate reductases (NRs), the nucleotide sequences encoding 19 nitrate reductase (NR) genes from 16 species of fungi, algae, and higher plants were analyzed. The NR genes examined show substantial sequence similarity, particularly within functional domains, and large variations in GC content at the third codon position and intron number. The intron positions were different between the fungi and plants, but conserved within these groups. The overall and nonsynonymous substitution rates among fungi, algae, and higher plants were estimated to be 4.33×10^{-10} and 3.29×10^{-10} substitutions per site per year. The three functional domains of NR genes evolved at about one-third of the rate of the N-terminal and the two hinge regions connecting the functional domains. Relative rate tests suggested that the nonsynonymous substitution rates were constant among different lineages, while the overall nucleotide substitution rates varied between some lineages. The phylogenetic trees based on NR genes correspond well with the phylogeny of the organisms determined from systematics and other molecular studies. Based on the nonsynonymous substitution rate, the divergence time of monocots and dicots was estimated to be about 340 Myr when the fungi–plant or algae–higher plant divergence times were used as reference points and 191 Myr when the rice–barley divergence time was used as a reference point. These two estimates are consistent with other estimates of divergence times based on these reference points. The lack of consistency between these two values appears to be due to the uncertainty of the reference times.

Key words: Nitrate reductases — Base composition — Intron number and locations — Molecular clock — Relative rate test — Angiosperms — Monocots — Dicots

Introduction

Eukaryotic nitrate reductases (NRs) catalyze the reduction of nitrate to nitrite and are classified into three groups based on the electron donors: NADH-specific (EC 1.6.6.1), NAD(P)H-bispecific (EC 1.6.6.2), and NADPH-specific (EC 1.6.6.3) (Guerrero et al. 1981). The NADH-specific NR is present in most higher plant and some algae species, while the NADPH-specific NR is found in fungi (Horner 1983; Cooley and Tomsett 1985; Renosto et al. 1982). The NAD(P)H-bispecific NR has been found in soybean (Streit et al. 1985), some grass species including rice (Shen et al. 1976), maize (Redinbaugh and Campbell 1981), and barley (Dailey et al. 1982), and some algal species (Losada 1976; Guerrero et al. 1981).

The NADH NR gene copy number varies among different species. *Hordeum vulgare* (barley) has a single NADH NR gene, while *Oryza sativa* (rice), *Triticum aestivum* (wheat), and many other grasses have at least two NADH NR genes per haploid genome (Hamat et al. 1989; Kleinhofs et al. 1988; Zhou 1993). Two NADH NR genes occur in *Arabidopsis* (Cheng et al. 1988).

* *Present address:* Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
Correspondence to: A. Kleinhofs

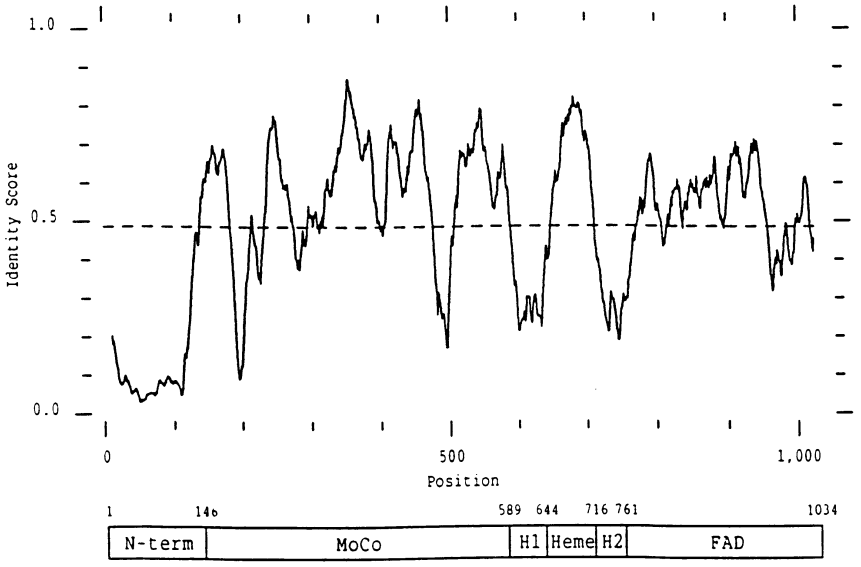

Fig. 1. Amino acid sequence homology of 17 NR genes. The plot was generated using the program PLOTSIMILARITY in the GCG package based on the optimal amino acid sequence alignment with window size of 21 amino acids. The window of comparison is moved along all sequences one position at a time. The average identity over the entire window is plotted at the middle position of the window. Two valleys observed in the MoCo region are due to the large insertions in the MoCo domain of fungal NR genes. The number for domain divisions corresponds to the consensus sequence generated from the optimal multiple amino acid sequence alignment.

All eukaryotic NRs have three highly conserved functional domains binding FAD, heme, and molybdenum cofactor (MoCo) (Kleinhofs et al. 1989). The three functional domains are separated by two short hinge regions (Kubo et al. 1988). They are encoded in a linear order with the MoCo region at the N-terminal, the heme region in the middle, and the FAD region at the C-terminal (Fig. 1). The amino acid sequence of the three functional domains is more highly conserved among species than the N-terminal and the hinge regions. The MoCo domains of higher plant NRs share about 40% amino acid homology with chicken liver sulfite oxidase (Campbell and Kinghorn 1990; Hoff et al. 1992; Warner and Kleinhofs 1992). The heme domains have 32–48% amino acid homology with the superfamily of cytochrome b_5 proteins such as cytochrome b_2, b_5, flavocytochrome b_2, and sulfite oxidase. The FAD domains share 47% identity with the FAD domain of cytochrome b_5.

NR genetics, biochemistry, and molecular biology have been investigated intensively (see Kleinhofs and Warner 1990 for a detailed review), but little is known about NR gene evolution. To understand the evolutionary mechanisms and relationships of NR genes, 19 NR DNA sequences from fungi, algae, and higher plants were analyzed. The results showed that there are large variations in GC content at the third codon position and intron number. The intron positions were different between the fungi and plants but conserved within these groups. The evolutionary rates were different for different regions of NR genes and the NR gene has apparently evolved at a constant rate at nonsynonymous sites.

Materials and Methods

DNA Sequences and Sequence Alignment. Nineteen NR sequences from fungi, algae, and higher plants were analyzed (Table 1). DNA sequences were taken from GenBank. The sequences of *Arabidopsis thaliana Nia1* and *Chlamydomonas reinhardtii* NR genes were kindly provided by Drs. N. Crawford (University of California at San Diego) and P. Lefebvre (University of Minnesota at Minneapolis), respectively.

The multiple NR amino acid sequences were aligned using the GCG (Genetics Computer Group, Devereaux et al. 1984) program PILEUP based on the alternative peptide symbol comparison table StructGappep.cmp. This alignment was then edited to obtain the optimal multiple amino acid sequence alignment (available upon request) using the multiple alignment editor program MALIGNED (Stephen Clark, Division of Cellular and Molecular Biology, Ontario Cancer Institute, Toronto, Canada). The multiple nucleotide sequences were first aligned using PILEUP, and then gaps were deleted or inserted according to the optimal amino acid sequence alignment using the GCG multiple alignment editor program LINEUP.

Relative Rate Test, Phylogenetic Analysis, and Base Composition. The number of nucleotide substitutions was calculated with Nei and Gojobori's (1986) unweighted pathway method using the computer program kindly provided by Dr. A. Hughes (Pennsylvania State University at University Park). The relative rate tests (Sarich and Wilson 1967) were performed as described by Li and Tanimura (1987). Since the synonymous substitution values for most of the pairwise comparisons were too great to be reliably estimated, only the nonsynonymous and overall nucleotide substitutions were considered. The classification of species and the choice of outgroup for each test relied on the established taxonomic evidence (Cronquist 1981). In order to simplify the computational task, not all of the possible pairwise relative rate tests were performed. Since the variation in the estimated number of nucleotide substitutions among the different pairs of species in the comparison of two lineages was considerably less than the sampling variances (Nei

Table 1. NR gene sequences used in this study

Species[a]	Abbreviation	Electron donor	Subclass	Amino acids
Fungi				
Aspergillus nidulans	And	NADPH	Euascomycetidae	873
Aspergillus niger	Ang	NADPH	Euascomycetidae	867
Neurospora crassa	Ncr	NADPH	Euascomycetidae	982
Algae				
Chlamydomonas reinhardtii	Cre	NAD(P)H	Chlorophyceae	882
Chlorella vulgaris	Cvu	NADH	Chlorophyceae	318[b]
Volvox carteri	Vca	NADH	Chlorophyceae	864
Dicots				
Arabidopsis thaliana, Nia1	At1	NADH	Dilleniidae	917
Arabidopsis thaliana, Nia2	At2	NADH	Dilleniidae	917
Betula pendula	Bpe	NAD(P)H	Hamamelidae	898
Cucurbita maxima	Cma	NADH	Dilleniidae	918
Lycopersicon esculentum	Les	NADH	Asteridae	911
Nicotiana tabacum, Nia1	Nt1	NADH	Asteridae	904
Nicotiana tabacum, Nia2	Nt2	NADH	Asteridae	904
Phaseolus vulgaris	Pvu	NADH	Rosidae	881
Spinacia oleracea	Sol	NADH	Caryophyllidae	926
Monocots				
Hordeum vulgare, Nar1	Hv1	NADH	Liliidae	915
Hordeum vulgare, Nar7	Hv2	NAD(P)H	Liliidae	891
Oryza sativa	Osa	NADH	Liliidae	916
Zea mays	Zma	NADH	Liliidae	502[c]

[a] Accession numbers or references for sequences are as follows: And, Johnstone et al. (1990); Ang, M77022; Ncr, X61303; Cre, Dr. P. Lefebvre (personal communication); Cvu, X56771; Vca, X64136; At1, Dr. N. Crawford (personal communication); At2, J03240; Bpe, X54079; Cma, M33154; Les, X14060; Nt1, X14058; Nt2, X14059; Pvu, X53603; Sol, M32600; Hv1, X57854; Hv2, X60173; Osa, X15819, X15820; Zma, M27821

[b] Partial sequence for heme binding domain

[c] Partial sequence of C-terminal

and Gojobori 1986) for the individual pairwise estimates, the pairwise estimates of the number of nucleotide substitutions in the comparison of two lineages were averaged, the standard error of this mean was recalculated, and this average and standard error were used in the relative rate test. Relative rate tests were performed on sets of species representing comparisons between (1) kingdoms or classes; (2) subclasses within a kingdom or class; (3) families within a subclass; and (4) genes within a genome.

The phylogenetic trees were constructed by the UPGMA (unweighted pair-group method using arithmetic means) and bootstrap parsimony method using the programs in the PHYLIP phylogeny inference package (Felsenstein 1989).

The base composition was estimated using the program DI_NUC, which was written by Susan Johns at the VADMS center (Visualization Analysis and Design in Molecular Sciences), Washington State University.

Results

Sequence Homology

The length of NR genes varied from 864 to 982 amino acids among fungi, algae, and higher plants (Table 1). The *Neurospora crassa* NR gene is the largest, while the *Volvox carteri* NR gene is the smallest. The length differences among these NR genes are mainly due to deletions or insertions in the N-terminal and hinge regions. Also, two large insertions (14–22 amino acids) were observed in the MoCo domains of fungal NR genes. No gap was observed in the heme domain of any NR gene, which may reflect the high functional constraint on this region.

Based on the alternative peptide symbol comparison table StructGappep.cmp in the GCG package, the overall amino acid sequence identity among these NR genes is greater than 39%. The amino acid sequences of the three functional domains are highly conserved, while those of the N-terminal and the two hinge regions are more diverged (Fig. 1). A region containing 11 amino acids, identical for all known NR genes, is located in the MoCo domain. Perhaps this region is involved in binding the molybdenum cofactor or nitrate.

Base Composition

Base composition at the third codon position varies greatly among different NR genes, while at the first and second codon positions it is less variable. The average GC content was 68.4% (±17.8%) at the third codon position, 41.5% (±2.8%) at the second codon position, and 56.4% (±5.1%) at the first codon position. Monocot and algal NR genes had very high GC contents (87.8% ± 9.4%) at the third codon position, while fungal and dicot

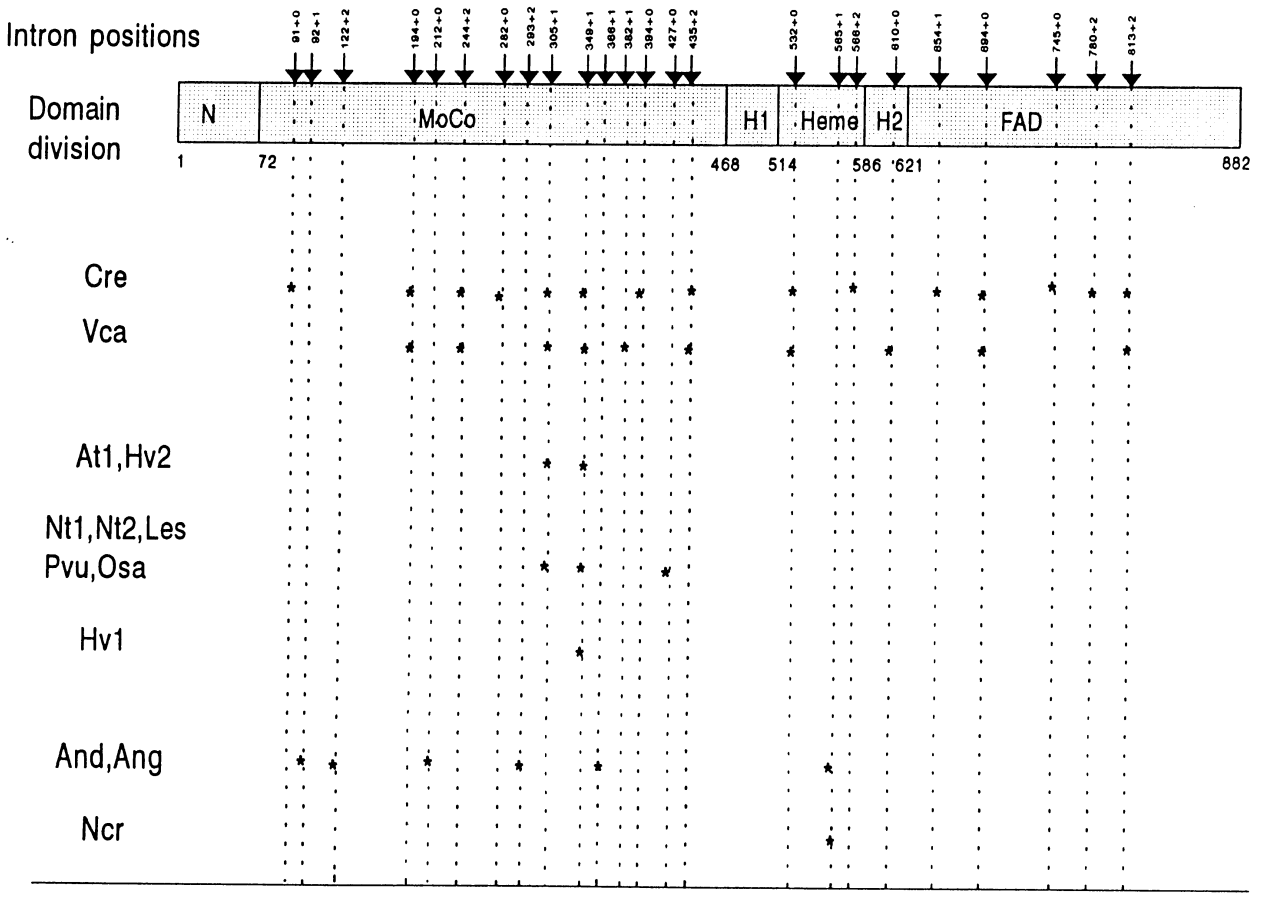

Fig. 2. Intron number and locations of NR genes. The *abbreviations* for individual genes are in Table 1. The codon positions and domain divisions are with reference to the *Chlamydomonas reinhardtii* NR amino acid sequence. The intron positions are described by both phase and codon. A codon followed by +1 or +2 is the next codon split by an intron after the first or second base, respectively. A codon followed by +0 has an intron between it and the next codon. The presence of an intron is shown by *.

NR genes had lower GC contents at this position (57.0% ± 9.3%). Considerable differences in GC content at the third codon position were also observed within dicot and algal NR genes.

NR Gene Intron Numbers and Locations

Extensive variation in intron number was observed among eucaryotic NR genes (Fig. 2). *Chlamydomonas reinhardtii* and *Volvox carteri* NR genes have 15 and 9 introns, respectively, which are located in the MoCo, heme, and FAD regions, while the barley NADH and *Neurospora crassa* NADPH NR genes have only one intron. Rice and most of the dicot NR genes have three introns in the MoCo region, while *Aspergillus* NR genes have six introns residing in the MoCo and heme regions.

The intron locations are conserved among NR genes within fungal, algal, and higher plant lineages (Fig. 2). The intron location in *Neurospora* NR gene is conserved from that in *Aspergillus* NR genes. Eight of the ten intron positions in *Volvox* are identical to those in *Chlamydomonas reinhardtii.* The three introns in NR genes from *Oryza sativa* (rice), *Phaseolus vulgaris* (bean), *Lycopersicon esculentum* (tomato), and *Nicotiana tabacum* (tobacco) are located precisely at the same positions. The barley NADH and NAD(P)H NR gene intron positions are also identical to those in the rice NADH NR gene.

The intron locations in higher plant NR genes appear to be conserved from algal NR genes. Two of the three NR gene introns in higher plants have identical positions to the introns in *Chlamydomonas* and *Volvox.* The intron positions between higher plant, algal, and fungal NR genes, however, are completely different (Fig. 2).

Introns found in the known eucaryotic NR genes do not divide the coding sequence into exons which represent the functional domains of the enzyme (Fig. 2). Although the exact boundaries for different functional domains are not known, most of the introns appear to be within and not between the functional domains.

Nucleotide Substitution Rates

In order to estimate NR gene evolutionary rates among different lineages, divergence times of 1,000 Myr for fungi and plants (Wolfe et al. 1989; Martin et al. 1989) and 750 Myr for green algae and higher plants (Amati et

Table 2. Estimated nonsynonymous nucleotide substitution rates for the entire and different regions of NR genes ($\times 10^{-10}$/site/year)[a]

	I[b]		*II*[c]		Average
Regions	n_1	Rate ± SE	n_2	Rate ± SE	Rate ± SE
Entire gene	42	3.24 ± 0.14	24	3.39 ± 0.15	3.29 ± 0.14
MoCo-Heme-FAD	42	2.86 ± 0.13	24	2.93 ± 0.14	2.89 ± 0.13
MoCo	42	2.59 ± 0.17	24	2.54 ± 0.17	2.57 ± 0.17
Heme	48	2.47 ± 0.36	39	2.92 ± 0.44	2.67 ± 0.40
FAD	45	3.47 ± 0.26	26	3.45 ± 0.26	3.46 ± 0.26
N-terminal	42	7.20 ± 1.42	24	8.16 ± 1.45	7.55 ± 1.43
Hinge 1	45	5.34 ± 1.34	26	7.39 ± 1.72	6.09 ± 1.48
Hinge 2	35	10.3 ± 2.22	26	8.50 ± 2.19	9.56 ± 2.20

[a] The nucleotide substitution rate was calculated by $K/2T$, in which K is the number of nucleotide substitutions per site and T is the divergence time
[b] *I* is the calculated substitution rate and its standard error based on the mean number of the nonsynonymous substitutions and the average variance over the n_1 pairwise comparisons between fungi and plants using fungi–plant divergence time, 1,000 Myr, as a reference point
[c] *II* is the calculated substitution rate and its standard error based on the mean number of the nonsynonymous substitutions and the average variance over the n_2 pairwise comparisons between algae and higher plants using algae–higher plant divergence time, 750 Myr, as a reference point

al. 1988; Meagher et al. 1989) were assumed. Based on these divergence times, the nucleotide substitution rates were calculated for the entire NR gene and for each of the individual regions of the sequence. The estimates of the nonsynonymous and overall nucleotide substitution rates were very similar for the two reference divergence times (Tables 2, 3). The comparisons of both nonsynonymous and overall substitution rates indicated that different regions of the NR genes evolved at different rates. The three functional domains evolved at about one-third of the rate of the other three regions (Tables 2, 3). The FAD domain evolved somewhat faster than the MoCo and heme domains (Tables 2, 3). The hinge 2 region evolved faster than the N-terminal, which in turn evolved more rapidly than the hinge 1 region (Tables 2, 3).

Relative Rate Test

The molecular clock hypothesis (Zuckerkandl and Pauling 1965) predicts that homologous genes in different species should evolve at similar rates. To test this hypothesis for NR genes, the relative rate test was applied. No significant deviation from the expectation that ''species'' 1 and 2 show equal divergence from the reference ''species'' was observed for nonsynonymous substitutions (Table 4). These data suggest that NR genes evolved at a constant rate at the nonsynonymous sites and can be used as a molecular clock.

Some species have two or more NR genes. To determine if these genes evolved at similar rates, the relative rate test was performed between different NR genes in barley, *Arabidopsis*, and *Nicotiana tabacum* (tobacco). No significant differences in nonsynonymous substitution rates were observed between NR genes within these species (Table 4).

Although the relative rate tests for the entire NR sequences are consistent with rate constancy at the nonsynonymous sites, this observation does not guarantee that each region of the NR gene evolved at a constant rate. Heterogeneity in the evolutionary rate in different regions may compensate for one another and lead to relative homogeneity in the rate for the entire gene. To examine this possibility, the relative rate tests were performed at nonsynonymous sites for each of the six regions: MoCo, heme, FAD, N-terminal, hinge 1, and hinge 2. No significant differences (at the 5% level) were detected, suggesting that each of these regions also evolved at a constant rate at nonsynonymous sites (data not shown).

Significant deviations from the expectation of molecular clock behavior for *overall* nucleotide substitutions rates were observed for 10 of the 26 tests (Table 4). The overall nucleotide substitution rate in monocots was significantly lower than that in dicots. The overall nucleotide substitution rates were not constant between different dicot subclasses.

Molecular Phylogeny

The UPGMA phylogenetic tree based on the nonsynonymous substitutions calculated from the entire NR gene sequences and the bootstrap parsimony tree based on the entire NR genes revealed four distinct groups separating fungi from plants, algae from higher plants, and monocots from dicots (Figs. 3, 4). Within the monocot and dicot groups, rice appeared to be more closely related to maize than to barley, and *Spinacia oleracea* (spinach) was closer to tobacco than to *Phaseolus vulgaris* (bean).

Dating of Monocot–Dicot Split

The apparent constancy of the NR gene nonsynonymous substitution rate encouraged us to use it to estimate species divergence times. Since the N-terminal and the two hinge regions (about 15% of the entire sequence) have

Table 3. Estimated overall nucleotide substitution rates for the entire and different regions of NR genes ($\times 10^{-10}$/site/year)[a]

	I		*II*		Average
Regions	n_1	Rate ± SE	n_2	Rate ± SE	Rate ± SE
Entire gene	42	4.20 ± 0.16	24	4.57 ± 0.17	4.33 ± 0.16
MoCo-Heme-FAD	42	3.83 ± 0.15	24	3.78 ± 0.15	3.81 ± 0.15
MoCo	42	3.60 ± 0.19	24	3.74 ± 0.20	3.65 ± 0.19
Heme	48	3.32 ± 0.41	39	3.92 ± 0.49	3.59 ± 0.45
FAD	45	4.36 ± 0.29	26	4.68 ± 0.31	4.48 ± 0.30
N-terminal	42	7.89 ± 1.39	24	9.15 ± 1.56	8.35 ± 1.45
Hinge 1	45	6.74 ± 2.16	26	9.30 ± 2.32	7.68 ± 2.18
Hinge 2	30	11.4 ± 2.90	26	9.16 ± 2.50	10.4 ± 2.71

[a] See Table 2 footnotes

Table 4. Relative rate test of the molecular clock hypothesis for the entire NR gene sequenced[a]

			Nonsynonymous substitution	Overall substitution
''Species'' 1	''Species'' 2	''Species 3'' reference	$K_{13} - K_{23}$ ± SE	$K_{13} - K_{23}$ ± SE
Algae	Higher plant	Fungi	0.033 ± 0.029	−0.005 ± 0.033
Algae	Monocots	Fungi	0.024 ± 0.029	0.027 ± 0.033
Algae	Dicots	Fungi	0.039 ± 0.029	−0.015 ± 0.032
Monocots	Dicots	Fungi	0.015 ± 0.022	−0.042 ± 0.033
Monocots	Dicots	Algae	−0.032 ± 0.019	−0.168 ± 0.022***
Within fungi				
And	Ang	Ncr	−0.005 ± 0.014	0.000 ± 0.034
Within algae				
Cre	Vca	Ncr	0.012 ± 0.017	0.026 ± 0.034
Within monocots				
Hv1	Osa	Les	0.008 ± 0.009	0.010 ± 0.018
Hv1	Hv2	Les	−0.013 ± 0.013	−0.023 ± 0.019
Osa	Hv2	Les	−0.021 ± 0.013	−0.030 ± 0.017
Within dicots				
Dilleniidae	Caryophyllidae	Monocots	0.001 ± 0.013	−0.033 ± 0.018
Dilleniidae	Asteridae	Monocots	0.020 ± 0.012	−0.042 ± 0.018*
Dilleniidae	Hamamelidae	Monocots	0.022 ± 0.012	0.045 ± 0.017**
Dilleniidae	Rosidae	Monocots	0.016 ± 0.012	0.050 ± 0.017**
Asteridae	Caryophyllidae	Monocots	0.019 ± 0.012	−0.009 ± 0.018
Hamamelidae	Caryophyllidae	Monocots	0.021 ± 0.012	0.078 ± 0.017***
Rosidae	Caryophyllidae	Monocots	0.015 ± 0.012	0.083 ± 0.017***
Asteridae	Rosidae	Monocots	0.004 ± 0.012	0.092 ± 0.017***
Asteridae	Hamamelidae	Monocots	0.002 ± 0.012	0.087 ± 0.018***
Rosidae	Hamamelidae	Monocots	0.006 ± 0.012	−0.005 ± 0.016
Within Dilleniidae				
At1	At2	Les	−0.014 ± 0.013	0.016 ± 0.015
At1	Cma	Hv2	−0.009 ± 0.013	0.070 ± 0.018***
At2	Cma	Hv2	0.017 ± 0.013	−0.070 ± 0.017***
Within Asteridae				
Nt1	Nt2	Les	−0.002 ± 0.004	−0.002 ± 0.005
Nt1	Les	Pvu	0.009 ± 0.011	−0.007 ± 0.016
Nt2	Les	Pvu	0.008 ± 0.011	−0.006 ± 0.016

[a] K_{13} or K_{23} is the number of nucleotide substitutions per site between ''species'' 1 or ''species'' 2 and the reference ''species'' 3. SE is the standard error of ($K_{13} - K_{23}$). The null hypothesis is $K_{13} - K_{23} = 0$. *Significant at 5% level. **Significant at 1% level. ***Significant at 0.1% level

evolved rapidly, they were excluded in estimating the divergence time for the monocot–dicot split. The divergence time for the monocot–dicot split was estimated based on three reference time points: the divergence time between fungi and plants (1,000 Myr), between algae and higher plants (750 Myr), and between rice and barley (70 Myr, Stebbins 1981; Wolfe et al. 1989). The divergence time for the monocot–dicot split was estimated to be about 340 Myr when the fungi–plant and algae–higher plant divergence times were used as reference points. A

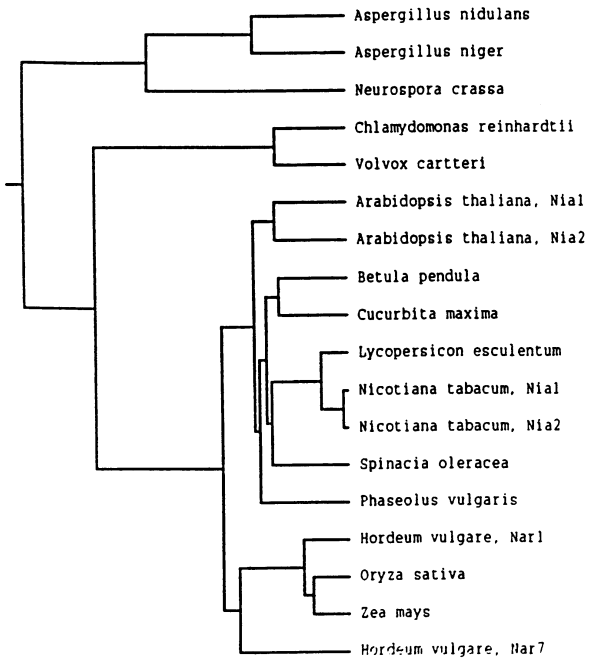

Fig. 3. Phylogenetic tree constructed with UPGMA based on the nonsynonymous substitutions of the entire NR gene sequences.

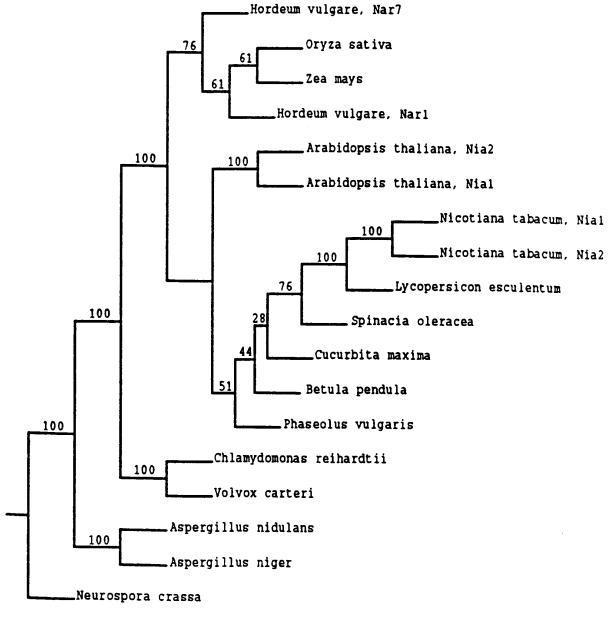

Fig. 4. Phylogenetic tree constructed with the bootstrap parsimony method based on the entire NR gene sequences. The *numbers* above the horizontal lines are bootstrap confidence estimates on the branches in 100 replicates.

value of 191 Myr was obtained when the rice–barley divergence time was used as a reference point (Table 5).

In addition, the divergence time for the barley NADH and NAD(P)H NR genes was estimated as 160 ± 12 Myr using the barley–rice divergence time as a reference point (data not shown). This result suggests that the barley ancestral NR gene duplicated about the time or shortly after the divergence of the monocots from the dicots.

Discussion

Intron loss and gain have been proposed as explanations for their origin. The intron loss hypothesis suggests that the progenitor genes shared by procaryotes and eucaryotes were interrupted by introns, which were then lost in procaryotes during evolution (Gilbert 1978; Darnell and Doolittle 1986; Gilbert et al. 1986). The intron gain hypothesis, on the other hand, suggests that introns have more recent origins and were inserted into the eukaryotic genes after the divergence of the prokaryote and eukaryote lineages (Cavalier-Smith 1985).

The conservation of two intron positions between higher plant and algal NR genes indicates that they probably had the same origin. The NR genes in the ancestor of algae and higher plants may have had all of the introns and subsequently most introns may have been lost during higher-plant evolution. The precise loss of an intron could occur through an exact deletion or through recombination with a DNA copy of a spliced gene transcript (Gilbert et al. 1986). The loss of introns has also been proposed in some other plant genes, such as alcohol dehydrogenase (ADH) (Gilbert et al. 1986). The *Arabidopsis* ADH gene has only six of the nine introns found in the maize ADH gene (Chang and Meyerowitz 1986).

Both intron loss and gain could account for the observation that fungi and plants do not have NR gene introns at identical positions. Introns might have been inserted into the NR genes after the divergence of fungi and plants, so the intron positions would not be conserved. The NR genes in the ancestor of fungi and plants, however, may have had all the introns present in current NR genes plus additional ones. Many introns might have been lost independently after the divergence of fungi and plants. Intron 1 in *Aspergillus* and *Chlamydomonas* and intron 6 in *Aspergillus* and intron 10 in *Chlamydomonas* are located only 4 bp apart (Fig. 2). These two intron locations may be identical and the positions may have ‘‘slipped’’ during evolution (Gilbert et al. 1986; Liaud et al. 1990, 1992; Martinez et al. 1989). Current NR gene intron data do not differentiate between the intron loss and gain hypotheses.

Gilbert (1978) suggested that novel protein structure and functions could evolve into new combinations by exchange or shuffling of exons. The exon shuffling hypothesis further argues that introns might be vestigial and serve as sites for recombination among exons. According to the exon shuffling theory, exon borders should delimit structural or functional regions within proteins, and some exons should be common to more than one gene family (Blake 1978). The homology of the three functional do-

Table 5. Estimations of monocit–dicot divergence time based on the nonsynonymous substitutions of the MoCo-heme-FAD functional domains of NR genes[a]

Reference divergence	Reference divergence time (Myr)	K_{AC}	K_{BC}	K_{AB}	Monocot–dicot divergence time (Myr)
Fungi–plants	1,000	0.576	0.562	0.192 ± 0.012	346 ± 21
Algae–higher plants	750	0.428	0.443	0.192 ± 0.012	339 ± 21
Rice–barley	70	0.203	0.195	0.073 ± 0.007	191 ± 12

[a] When using the divergence times of fungi–plants (including algae) and algae–higher plants as reference points, K_{AC}, K_{BC}, and K_{AB} are the nonsynonymous substitutions of the MoCo-heme-FAD functional domains among monocots (*A*), dicots (*B*), and fungi or algae (*C*). When using the divergence time for a rice–barley split as a reference point, K_{AC}, K_{BC}, and K_{AB} are the nonsynonymous substitutions of the MoCo-heme-FAD functional domains among rice (*A*), barley (*B*), and dicots (*C*). The divergence time for monocot–dicot split was estimated using the method of Li and Graur (1991, pp. 117–118). The standard errors for the monocot–dicot divergence times were obtained by dividing the standard errors of K_{AB} with twice the nonsynonymous substitution rate for MoCo-heme-FAD regions (2.89×10^{-10}) (Table 2)

mains of NR genes with other eukaryotic genes (Crawford et al. 1988; Campbell and Kinghorn 1990; Hoff et al. 1992) suggests the possibility that exon shuffling was involved in the evolution of NR genes. This hypothesis, however, is not supported by the observations with NR genes where the recognized functional domains are not separated by the known introns. Thus, in order to reconcile the exon shuffling hypothesis with the intron position information, one has to postulate the existence of many more introns in the progenitor NR gene. Alternatively, the currently recognized NR functional domains may be quite different in size from those ''shuffled'' during the early evolution of NR. The different intron positions between the fungal and plant NR genes suggest that extensive rearrangements of the NR gene structure have taken place since its early evolution.

The average overall nucleotide substitution rate of the entire NR gene (4.33×10^{-10}) is slightly lower than the rates estimated for *Adh1* (alcohol dehydrogenase 1) (5.18×10^{-10}) (Gaut and Clegg 1991) and for *Rbc*L (ribulose-1,5-bisphosphate carboxylase large subunit) ($5.0–6.5 \times 10^{-10}$) (Doebley et al. 1990). The nonsynonymous substitution rate (3.29×10^{-10}) is slightly higher than the rate for *Adh1* (2.5×10^{-10}) (Gaut and Clegg 1991) and much higher than that for GAPDH (glyceraldehyde-3-phosphate dehydrogenase) (1.41×10^{-10}) (Martin et al. 1989). The apparent difference in the nucleotide substitution rates between NR and *Adh1* or *Rbc*L genes may be due to the difference in the choice of the reference time for calculating the evolutionary rate. The divergence times for fungi–plants and for algae–higher plants were used to estimate the NR gene nucleotide substitution rates, while the rice–barley divergence time was used for calculating the *Adh1* and *Rbc*L gene nucleotide substitution rates (Gaut and Clegg 1991; Doebley et al. 1990).

The relative rate tests suggest that NR genes evolved at a constant rate at the nonsynonymous sites among fungi, algae, and higher plants, whereas the overall nucleotide substitution rate was not constant. Although the constancy of the synonymous substitution rates in NR genes could not be tested, the inconstancy of the overall nucleotide substitution rates and the constancy of the nonsynonymous substitution rates imply that the synonymous substitution rates would not be constant.

One possible explanation for the near constancy of the nonsynonymous substitution rates and the lineage-dependent synonymous substitution rates is stringent functional constraint at the protein level. Although no direct test of such a hypothesis is feasible, Sueoka's (1962, 1988, 1992) directional mutation pressure theory suggests that the regression of GC content at the first and second codon positions with GC content at the third codon position can be used as an index of the severity of selective constraint. For NR, the regression coefficients were found to be 0.0804 between the second and third codon positions and 0.2559 between the first and third codon positions. According to Sueoka's theory, this indicates that the constraint at the second codon position is very high, and somewhat less at the first codon position. This theory would account for the observed pattern of NR rate variation as an effect of powerful selective constraints at nonsynonymous sites. The very good clock-like behavior that results is similar to the situation in GS (glutamine synthetase) genes (Pesole et al. 1991).

Relative rate tests revealed that the NR genes within barley and *Arabidopsis* have evolved at constant rates even though their expression patterns are different (Warner et al. 1987; Wilkinson and Crawford, personal communication). These data suggest that the functional difference of these NR genes in plant growth may not have affected their evolutionary rates.

Both UPGMA and bootstrap parsimony trees based on NR genes suggest that the bambusoids (rice) are more closely related to the panicoids (corn) than to the pooids (barley). These results are consistent with the *Rbc*L-based and *Adh1*-based analyses (Doebley et al. 1990; Gaut and Clegg 1991). However, they do not agree with the phylogeny based on ribosomal RNA (rRNA) sequence data, which suggests that the bambusoids are closer to the pooids than to the panicoids (Hamby and Zimmer 1987). The UPGMA and bootstrap parsimony trees based on NR genes also suggest that Caryophylli-

dae is closer to Asteridae than to Rosidae. This agrees with Ritland and Clegg's (1987) topology for these three subclasses from two chloroplast gene sequences using UPGMA and with Wolfe et al.'s (1987) topology from 15 chloroplast gene sequences using the neighbor-joining method. One exception to classical taxonomy is the position of *Cucurbita maxima* (squash), which is usually classified with *Arabidopsis* in the subclass Dilleniidae (Cronquist 1981). Based on NR genes, it is not clustered with *Arabidopsis.*

The estimation of the monocot–dicot divergence time based on NR gene sequences is dependent on the chosen reference times (Table 5). Using fungi–plant and algae–higher plant divergence times to calibrate the molecular clock, the monocot–dicot split was estimated at about 340 Myr. This result is consistent with the estimations based on the GAPDH genes using the fungi–plant split reference time (Martin et al. 1989), and on the GAPDH and *Rbc*L genes using the brophyte–spermatophyte split reference time (450 Myr) (Martin et al. 1993). Using the divergence time of rice and barley to calibrate the clock, the monocot–dicot split was estimated at approximately 200 Myr. This estimate is consistent with that based on chloroplast DNA sequences using the rice–barley split reference time (Wolfe et al. 1989). These results point to the reference times as the major source of error in estimating the monocot–dicot divergence time. The estimated divergence times for the fungi–plant split and for algae–higher plant split may involve greater uncertainty (Clegg 1990), but the calculation based on the estimated rice–barley split time may be more sensitive to errors. Other possible sources of error are that fungi and algae may be too distantly related to higher plants to serve as proper outgroups, the sequenced rice and barley NADH NR genes may not be orthologous, and the sequenced cereal NR genes may not be representative of monocots. Additional NR gene sequence data may help resolve this controversy.

Several molecules have been used to study the evolution of plant species, e.g., rRNA (Hamby and Zimmer 1992), GAPDH (Martin et al. 1989, 1993), GS (Pesole et al. 1991), cytochrome *c* (Scogin 1981; Boulter 1972), *Rbc*S (ribulose-1,5-bisphosphate carboxylase small subunit) (Martin et al. 1983; Meagher et al. 1989), and *Rbc*L and other chloroplast genes (Wolfe et al. 1989; Clegg and Zurawski 1992). Not all molecules have equal value in analyzing evolutionary events and phylogenetic relationships. Woese (1987) argued that a useful chronometer has to meet certain specifications: (1) clock-like behavior, (2) phylogenetic range, and (3) size and accuracy. The molecule has to evolve at a constant rate among different lineages. The rates of change of the molecule have to be commensurate with the spectrum of evolutionary distances being measured. The molecule has to be large enough and consist of a fairly large number of loosely coupled ''domains'' (functional units) which are somewhat independent of one another in an evolutionary sense. Based on these specifications, NR genes may be useful for studying plant molecular evolution. NR genes have evolved at a constant rate at the nonsynonymous sites and consist of three distinct highly conserved functional domains and three other more rapidly diverged regions. These different regions have evolved at variable rates, so they may be useful in resolving the evolutionary relationships among both distantly and closely related organisms. The three highly conserved functional domains consist of about 85% of the gene, so it is easier to obtain optimal sequence alignment for NR genes than for rRNA genes in which the best alignment is often difficult to achieve (Rothschild et al. 1986; Pesole et al. 1991). NR genes are also large, about 900 amino acids, so they provide more informative sites (bp) for phylogenetic studies than other genes. Finally, NR genes have extensive variation in intron numbers and the positions of many introns are conserved. The NR intron data and the conservation of the NR gene functional domains with other eukaryotic genes (Crawford et al. 1988; Campbell and Kinghorn 1990; Hoff et al. 1992) may be very useful for examination of the intron loss or gain and exon shuffling hypotheses.

Acknowledgments. We are grateful to Drs. Robert L. Warner, Michael Kahn, Michael Moody, and Andrzej Kilian for their suggestions and discussions on this study. This work was supported by NSF grant DCB-8904551, by the Washington Technology Center and by the Washington Barley Commission. This publication is Scientific Paper Number 96 02-03, College of Agriculture and Home Economics Research Center, Washington State University, Pullman, WA, USA. Project No. 0745 and 0223.

References

Amati BB, Goldschmidt-Clermont M, Wallace CJA, Rochaix J-D (1988) cDNA and deduced amino acid sequences of cytochrome c from *Chlamydomonas reinhardtii*. Unexpected functional and phylogenetic implications. J Mol Evol 28:151–160

Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? Nature 273:267–268

Boulter D (1972) Amino acid sequences of cytochrome c and plastocyanins in phylogenetic studies of higher plants. Syst Zool 22:549–553

Campbell H, Kinghorn JR (1990) Functional domains of assimilatory nitrate reductases and nitrite reductases. TIBS 15:315–319

Cavalier-Smith T (1985) Selfish DNA and the origin of introns. Nature 315:283–284

Chang C, Meyerowitz EM (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412

Cheng CL, Dewdney J, Nam HG, den Boer BGW, Goodman HM (1988) A new locus (*NIA1*) in *Arabidopsis thaliana* encoding nitrate reductases. EMBO J 7:3309–3314

Clegg MT, Zurawski G (1992) Chloroplast DNA and the study of plant phylogeny: present status and future aspects. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 50–91

Clegg M (1990) Dating the monocot-dicot divergence. Trend Ecol Evol 5:1–2

Cooley RN, Tomsett AB (1985) Determination of the subunit size of NADPH-nitrate reductases from *Aspergillus nidulans*. Biochim Biophys Acta 831:89–93

Crawford NM, Smith M, Bellissimo D, Davis RW (1988) Sequence and nitrate regulation of the *Arabidopsis thaliana* mRNA encoding nitrate reductases, a metalloflavoprotein with three functional domains. Proc Natl Acad Sci USA 85:5006–5010

Cronquist A (1981) An integrated system of classification of flowering plants. Columbia University Press, New York

Dailey FA, Warner RL, Somers DA, Kleinhofs A (1982) Characteristics of a nitrate reductases in a barley mutant deficient in NADH nitrate reductases. Plant Physiol 69:1200–1204

Darnell JE, Doolittle WF (1986) Speculations on the early course of evolution. Proc Natl Acad Sci USA 83:1271–1275

Devereaux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

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

Felsenstein J (1989) PHYLIP—Phylogeny inference package (version 3.2). Cladistics 5:164–166

Gaut BS, Clegg MT (1991) Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. Proc Natl Acad Sci USA 88:2060–2064

Gilbert W (1978) Why genes in pieces? Nature 271:501

Gilbert W, Marchionni M, McKnight G (1986) On the antiquity of introns. Cell 46:151–154

Guerrero MG, Vega JM, Losada M (1981) The assimilatory nitrate-reducing system and its regulation. Annu Rev Plant Physiol 32: 169–204

Hamat HB, Kleinhofs A, Warner RL (1989) Nitrate reductases induction and molecular characterization in rice (*Oryza sativa* L.). Mol Gen Genet 218:93–98

Hamby RK, Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 50–91

Hamby RK, Zimmer EA (1987) Ribosomal RNA sequences for inferring phylogeny within the grass family (*Poaceae*). Plant Syst Evol 160:29–37

Hoff T, Stummann BM, Henningsen KW (1992) Structure, function and regulation of nitrate reductases in higher plants. Physiol Plant 84:616–624

Horner RD (1983) Purification and comparison of *nit-1* and wild-type NADH: nitrate reductases of *Neurospora crassa.* Biochim Biophys Acta 744:7–15

Johnstone IL, McCabe PC, Greaves P, Gurr SJ, Cole GE, Brow MAD, Unkles SE, Clutterbuck AJ, Kinghorn JR, Innis MA (1990) Isolation and characterization of the *crn*A-*nii*A-*nia*D gene cluster for nitrate assimilation in *Aspergillus nidulans.* Gene 90:181–192

Kleinhofs A, Chao S, Sharp PJ (1988) Mapping of the nitrate reductases genes in barley and wheat. In: Miller TE, Koebner RMD (eds) Proceedings seventh international wheat genetics symposium, vol 1. Institute of Plant Science Research, Cambridge, pp 541–546

Kleinhofs A, Warner RL, Melzer (1989) Genetics and molecular biology of higher plant nitrate reductases. In: Poulton JE, Romeo JT, Conn EE (eds) Recent advances in phytochemistry, vol 23, Plant Nitrogen Metabolism. Plenum, New York, pp 117–155

Kleinhofs A, Warner RL (1990) Advances in nitrate assimilation. In: Miller BJ, Lea PJ (eds), Intermediary nitrogen metabolism, vol 16, Stumpf PK, Conn EE (eds), The biochemistry of plants. Academic Press, New York, pp 89–120

Kubo Y, Oguara N, Nakagawa H (1988) Limited proteolysis of the nitrate reductases from spinach leaves. J Biol Chem 263:19684–19689

Li WH, Graur D (1991) Fundamentals of molecular evolution. Sinauer, MA

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Liaud MF, Zhang DX, Cerff R (1990) Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate genes to the nucleus. Proc Natl Acad Sci USA 87:8918–8922

Liaud MF, Brinkmann H, Cerff R (1992) The β-tubulin gene family of pea: primary structures, genomic organization and intron-dependent evolution of genes. Plant Mol Biol 18:639–651

Losada M (1976) Metalloenzymes of the nitrate-reducing system. J Mol Catalysis 1:245–264

Martin W, Lydiate D, Brinkmann H, Forkmann G, Saedler H, Cerff R (1993) Molecular phylogenies in angiosperm evolution. Mol Biol Evol 10:140–162

Martin W, Gierl A, Saedler H (1989) Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339:46–48

Martin PG, Dowd JM, Stone SJL (1983) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. II The analysis of small subunit data to form phylogenetic trees. Aust J Bot 31:411–419

Martinez P, Martin W, Cerff R (1989) Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Biol 208:551–565

Meagher RB, Berry-lowe S, Rice K (1989) Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics 123:845–863

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:290–300

Pesole G, Bozzetti MP, Preparata G, Saccone C (1991) Glutamine synthetase gene evolution: a good molecular clock. Proc Natl Acad Sci USA 88:522–526

Redinbaugh MG, Campbell WH (1981) Purification and characterization of NAD(P)H:nitrate reductases and NADH:nitrate reductases from corn roots. Plant Physiol 68:115–120

Renosto F, Schmidt ND, Segel IH (1982) Nitrate reductases from *Penicillium chrysogenum:* the reduced flavin-adenine dinucleotide-dependent reaction. Arch Biochem Biophys 219:12–20

Ritland K, Clegg MT (1987) Evolutionary analysis of plant DNA sequences. Am Nat 130:S74–S100

Rothschild LJ, Ragan MA, Coleman AW, Heywood P, Gerbi SA (1986) Are rRNA sequence comparisons the Rosetta stone of phylogenetics? Cell 47:640

Sarich VM, Wilson AC (1967) Rates of albumin evolution in primates. Proc Natl Acad Sci USA 58:142–148

Scogin R (1981) Amino acid sequence studies and plant phylogeny. In: Yong DA, Sieger DS (eds) Phytochemistry and angiosperm phylogeny. Praeger, New York, pp 19–42

Shen TC, Funkhouser EA, Guerrero MG (1976) NADH- and NAD(P)H-nitrate reductases in rice seedlings. Plant Physiol 58: 292–294

Stebbins GL (1981) Coevolution of grasses and herbivores. Ann Mo Bot Gard 68:75–86

Streit L, Nelson RS, Harper JE (1985) Nitrate reductases from wild-type and *nr1*-mutant soybean (Glycine max[L]Merr) leaves. I Purification, kinetics, and physical properties. Plant Physiol 78:80–84

Sueoka N (1992) Directional mutation pressure, selective constraints, and genetic equilibria. J Mol Evol 34:95–114

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

Warner RL, Narayanan KR, Kleinhofs A (1987) Inheritance and ex-

pression of NAD(P)H nitrate reductase in barley. Theor Appl Genet 74:714–717

Warner RL, Kleinhofs A (1992) Genetics and molecular biology of nitrate metabolism in higher plants. Physiol Plant 85:245–252

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Wolfe KH, Gouy M, Yang YW, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

Zhou JZ (1993) Molecular evolution of nitrate reductases. PhD dissertation, Washington State University, Pullman, WA 99164

Zuckerkandl E, Pauling L (1965) Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel H (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166