doi: 10.1046/j.1420-9101.2004.00658.x

Molecular evolution of plant haemoglobin: two haemoglobin genes in nymphaeaceae *Euryale ferox*

E. GULDNER, E. DESMARAIS, N. GALTIER & B. GODELLE

CNRS UMR 5000 – 'Génome, Populations, Interactions', Université Montpellier, Montpellier, France

Keywords:

duplication;
nonsynonymous substitution rate;
Nymphaeaceae;
plant haemoglobin;
symbiosis.

Abstract

We isolated and sequenced two haemoglobin genes from the early-branching angiosperm *Euryale ferox* (Nymphaeaceae). The two genes belong to the two known classes of plant haemoglobin. Their existence in Nymphaeaceae supports the theory that class 1 haemoglobin was ancestrally present in all angiosperms, and is evidence for class 2 haemoglobin being widely distributed. These sequences allowed us to unambiguously root the angiosperm haemoglobin phylogeny, and to corroborate the hypothesis that the class 1/class 2 duplication event occurred before the divergence between monocots and eudicots. We addressed the molecular evolution of plant haemoglobin by comparing the synonymous and nonsynonymous substitution rates in various groups of genes. Class 2 haemoglobin genes of legumes (functionally involved in a symbiosis with nitrogen-fixing bacteria) show a higher nonsynonymous substitution rate than class 1 (nonsymbiotic) haemoglobin genes. This suggests that a change in the selective forces applying to plant haemoglobins has occurred during the evolutionary history of this gene family, potentially in relation with the evolution of symbiosis.

Introduction

Two types of haemoglobin, called class 1 and class 2 haemoglobin, occur in vascular plants. They are distinct from each other through their function and expression pattern (Jacobsen-Lyon *et al.*, 1995; Trevaskis *et al.*, 1997; Hunt *et al.*, 2001). Class 1 haemoglobins have been found in both monocots and eudicots and show a conserved expression pattern (Hunt *et al.*, 2001). They are expressed in germinating seeds [barley (Duff *et al.*, 1998)], in roots of mature plants (*Parasponia, Trema* and *Casuarina* GLB1 (Bogusz *et al.*, 1988; Bogusz *et al.*, 1990; Jacobsen-Lyon *et al.*, 1995; Andersson *et al.*, 1997; Franche *et al.*, 1998) and rosette leaves subjected to hypoxic conditions, and in response to sucrose [*Arabidopsis* (Trevaskis *et al.*, 1997)]. Class 2 haemoglobins were first described in plants achieving symbiosis with nitrogen-fixing bacteria (symbiotic haemoglobins). In most legume species (symbiosis with *Rhizobium* and *Bradyrhizobium*) and in several species from Fagales, Rosales and Cucurbitales (symbiosis with *Frankia*), class 2 haemoglobins are expressed at high concentration in nodules and facilitate oxygen diffusion to the nitrogen-fixing symbionts (Appleby, 1992). Class 2 haemoglobins were also found in several nonsymbiotic eudicots (nonsymbiotic haemoglobins), where their function remains unclear. In these species, they are expressed in several tissues (Szabados *et al.*, 1990; Strozycki *et al.*, 2000), and might participate in embryogenesis and seed maturation [*Cichorium* (Hendriks *et al.*, 1998), *Arabidopsis* (Hunt *et al.*, 2001)]. At this moment no class 2 haemoglobin gene has been found in any monocot.

Expression patterns, therefore, suggest that the evolution of haemoglobin may have played a role in the evolution of symbiosis between angiosperms and nitrogen-fixing bacteria, stimulating research about the molecular phylogeny and evolution of this gene family. Hypotheses about plant haemoglobin evolution have gradually progressed thanks to the characterization of new sequences (Andersson *et al.*, 1996; Trevaskis *et al.*, 1997; Arredondo-Peter *et al.*, 1998; Strozycki *et al.*, 2000). We are a long way from the first postulate of a horizontal transfer of the haemoglobin gene from the

Correspondence: Guldner Emilie, CNRS UMR 5000 – 'Génome, Populations, Interactions', Université Montpellier 2 – CC 063, Place E. Bataillon, 34095 Montpellier, France. Fax: +33 467 14 45 54; e-mail: guldner@univ-montp2.fr

animal kingdom to a progenitor of the dicotyledonous angiosperms. Molecular phylogeny analyses support the grouping of class 1 vs. class 2 haemoglobins (Andersson *et al.*, 1996; Trevaskis *et al.*, 1997; Arredondo-Peter *et al.*, 1998; Strozycki *et al.*, 2000), the two clusters being separated by a long internal branch. The recent identification of haemoglobin genes in nonvascular plants (liverwort *Marchantia polymorpha* (Hunt *et al.*, 2001) and mosses *Physcomitrella patens* (Arredondo-Peter *et al.*, 2000) and *Ceratodon purpureus* (accession number AF309562) provides a way to root the angiosperm haemoglobin phylogeny, potentially casting light on the evolutionary history of this gene family. According to Hunt *et al.* (2001), the root of the tree takes place between the class 1 and class 2 clusters (Fig. 1a), suggesting that class 1 and class 2 haemoglobins arose through an ancient duplication event that occurred before the divergence between monocots and eudicots. This hypothesis requires that class 2 genes have been lost in the monocot lineage.

Nonvascular plants, however, are quite divergent from angiosperms (Soltis *et al.*, 1999), casting doubts on the location of the root of the haemoglobin tree. The so-called 'long branch attraction' artefact (Felsenstein, 1978) may be responsible for the moss sequences being connected to the long branch that separates class 1 from class 2 haemoglobins, rather than to any other branch in the un-rooted tree. An alternative scenario involving a eudicot-specific duplication event, and paraphyletic class 1 haemoglobins, would be more parsimonious with respect to gene gain/loss (Fig. 1b).

Also, the published haemoglobin trees suggest that the rate of amino-acid substitution varies between lineages (Strozycki & Legocki, 1995; Strozycki *et al.*, 2000). The divergence between class 2 haemoglobins from various legume species, for example, appears comparable with the divergence between monocot and eudicot class 1 genes despite a more restricted taxonomic range, suggesting a possible acceleration of evolutionary rate in the class 2 lineage. This acceleration, if confirmed, would be an interesting evolutionary feature of the class 2 symbiotic genes. It might, however, contribute to possible phylogenetic artefact.

These arguments emphasize the need for a solid rooting of the haemoglobin tree. This would help in clarifying the phylogenetic history of the gene family in symbiotic and nonsymbiotic plants. Furthermore, the measurement of rate variation between lineages being highly dependent on the assumed phylogeny and rooting, this would also allow assessing the extent of evolutionary rate variation between lineages, possibly connecting features of the molecular evolutionary process in haemoglobin to the evolution of symbiosis. In this study, we made use of Nymphaeaceae *Euryale ferox,* a deeply-branching lineage of the angiosperm tree, to root the haemoglobin tree, and reconsider the molecular evolution of plant haemoglobins.

Material and methods

Young leaves from the Nymphaeaceae *E. ferox* were collected in the botanical garden of Montpellier (France). DNA extraction was performed with the DNeasy Plant Kit (QIAGEN, Courtaboeuf, France).

Degenerate primers for polymerase chain reaction (PCR) were defined from aligned plant haemoglobin sequences available in GenBank, and used to amplify each haemoglobin gene in two overlapping regions. Class 1 primers are: TTCAGYGARGAGCARGARGC (exon 1 forward)/GTGGWCTCCCTCACSGTRAC (exon3 reverse), and GACTCCGACCTGCCACTTGAAC (exon 2 forward)/GCRGCBACCARYTGRTCRTARGC (exon 4 reverse). Class 2 primers are: TTCAGYGARGAGCARGARGC (exon 1 forward)/GTRGCRCCYAGYCTYTTCAG (exon 3 reverse), and AACCCSMARCTCARGYCYCAT (exon 2 forward)/GCMRCCAACTKRTCRTARGC (exon 4 reverse). The PCR reactions included *Taq* DNA polymerase buffer (Promega)/1.5 mM MgCl2/200 μM each dNTP/

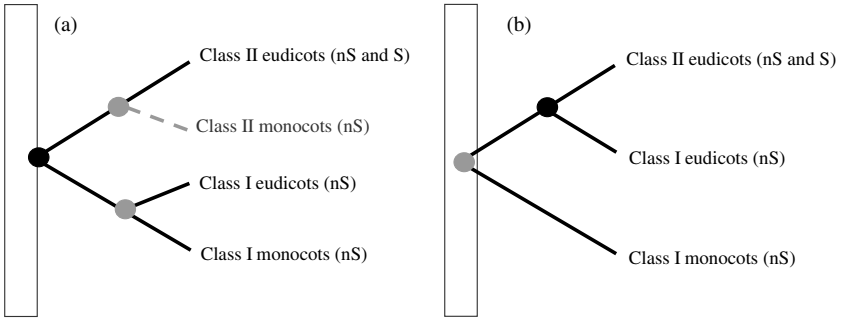

Fig. 1 Evolutionary history of plant haemoglobin genes: competing hypothesis. (a) A gene duplication gave rise to class 1 and class 2 haemoglobins before the divergence between monocots and eudicots. Class 2 haemoglobins have been lost (or are still undetected) in cereals. (b) A gene duplication occurred in the eudicot lineage – black dot: gene duplication; gray dot: monocots/eudicots divergence; S: symbiotic gene; nS: nonsymbiotic gene.

variable quantities of each primer, depending on their degeneracy/genomic DNA/5 units *Taq* DNA polymerase (Promega) in a total volume of 50 μL. The reaction consists of a first step at 94 °C for 2 min, followed by 30 cycles with a denaturation step at 91 °C for 30 s, an annealing step at variable temperature (from 55 °C to 62 °C) for 20 s and an extension step at 72 °C for 30 s.

The amplified fragments, purified with the GenElute PCR DNA purification kit (Sigma-Aldrich, Saint Quentin Fallavier, France), were cloned with the pGEM-T Easy Vector System (Promega). For each amplified fragment, four to six clones were sequenced using the ThermoSequenase sequencing kit with dye primer (Amersham Pharmacia Biosciences, Orsay, France) following the provider's instructions. The sequences were analysed with the ALFwin Sequence Analyser (Amersham Pharmacia Biosciences).

Phylogenetic analyses were performed on 52 plant haemoglobin amino-acid sequences (143 sites, 132 sites with gaps removing, 121 variable sites, 110 informative sites). We used methods of maximum likelihood [using the JTT model (Jones *et al.*, 1992)] with the program NJML+ (Ota & Li, 2001), and Neighbour-Joining (NJ, Saitou & Nei, 1987) with the program PHYLO_WIN (Galtier *et al.*, 1996), with 1000 bootstrap replicates for each analysis. Phylogenetic analysis based on Bayesian inference was also performed with *MrBayes* program (Huelsenbeck & Ronquist, 2001). We used the JTT model assuming a discretized gamma distribution with four classes. 120 000 Monte Carlo Markov chain (MCMC) steps were performed using four Markov chains running at different temperatures. The 20 000 first steps were discarded (burnin). Trees were sampled every 100 steps, and a consensus tree (out of 1000 sampled trees) was constructed. Clade posterior probabilities allow evaluation of the robustness of each node.

Pairwise distances between protein sequences were calculated using the PAM correction for multiple substitutions (Dayhoff *et al.* 1978), as implemented in program PROTDIST from package PHYLIP (Felsenstein 1989). These distances, used to build the NJ tree, were also used to examine sequence similarities between the newly isolated sequences of *E. ferox* and others plant haemoglobin genes.

The heterogeneity of evolution rate between lineages was measured using a generalization of the relative-rate test (Wu & Li, 1985) to more than three species (Robinson-Rechavi *et al.*, 1998; Robinson-Rechavi & Huchon, 2000). The synonymous (*Ks*) and nonsynonymous (*Ka*) evolutionary distances were estimated for each sequence pair within the group of eudicot class 1 genes (11 genes from eight different species, 55 pairs), the group of monocot class 1 genes (four genes from three different species, six pairs), and the group of class 2 (symbiotic) genes of legumes (27 genes from 11 different species, 351 pairs). We used the method of Yang & Nielsen (2000) as implemented in program yn00 of PAML (Yang, 1997).

Results and discussion

Two haemoglobin sequences in Nymphaeaceae *E. ferox*

All the fragments produced by degenerate amplifications on the *E. ferox* genomic DNA were cloned and sequenced. Two sequences (called *Hbn1* and *Hbn2*) showed significant similarity to known haemoglobin sequences. *Hbn2* shows three introns and four exons in positions identical to those of other plant haemoglobins (Jensen *et al.*, 1981; Landsmann *et al.*, 1986). Concerning *Hbn1*, amino-acid sequence similarity to known haemoglobin sequences was recovered when a nonstandard intron–exon map was used: exon 3 appears 42 nucleotides longer (and exon 4, 42 nucleotides shorter) than in other plant haemoglobin genes. PAM distances on amino acid sequences were estimated between each of these two new genes and all others known sequences. *Hbn1* is much closer to the class 1 haemoglobins (average PAM distance: 0.41 replacements per site) than to class 2 nonsymbiotic (0.86 rep. per site) and symbiotic (1.19 rep. per site) haemoglobins. In contrast, *Hbn2* is more similar to class 2 nonsymbiotic genes (0.76 rep. per site) than to class 1 (0.90 rep. per site) and symbiotic (1.05 rep. per site) haemoglobins. These results show that at least two haemoglobin genes actually exist in *E. ferox*, presumably belonging to class 1 (*Hbn1*) and class 2 (*Hbn2*), respectively.

Figure 2 presents the phylogenetic tree of 52 plant haemoglobin genes, constructed from 132 amino-acid sites with phylogenetic analysis based on Bayesian inference. Both groups of class 1 and class 2 genes are supported by high clade posterior probabilities (0.97 and 1.00 respectively). Phylogenetic reconstructions obtained using the NJ and maximum of likelihood methods are congruent with the Bayesian tree in that they present two strongly supported clades composed of class 1 genes (including *Hbn1*), and class 2 genes (including *Hbn2*) respectively. Note that both the class 1 and class 2 genes of *E. ferox* branch out as the earliest diverging lineages, in agreement with recent angiosperm molecular phylogenies (Soltis *et al.*, 1999; Soltis *et al.*, 2000). These results unambiguously confirm the previous rooting of the haemoglobin tree (Hunt *et al.*, 2001), i.e. the gene duplication that gave rise to class 1 and class 2 haemoglobins is older than the split between monocots and eudicots. Thus, class 2 haemoglobins must have been lost (or are still undetected) in cereals, and more research is needed to find out whether the lack of class 2 haemoglobin is restricted to cereals, or general to all monocots.

The unambiguous rooting of the tree gives an opportunity to compare molecular evolutionary rates between haemoglobin lineages.

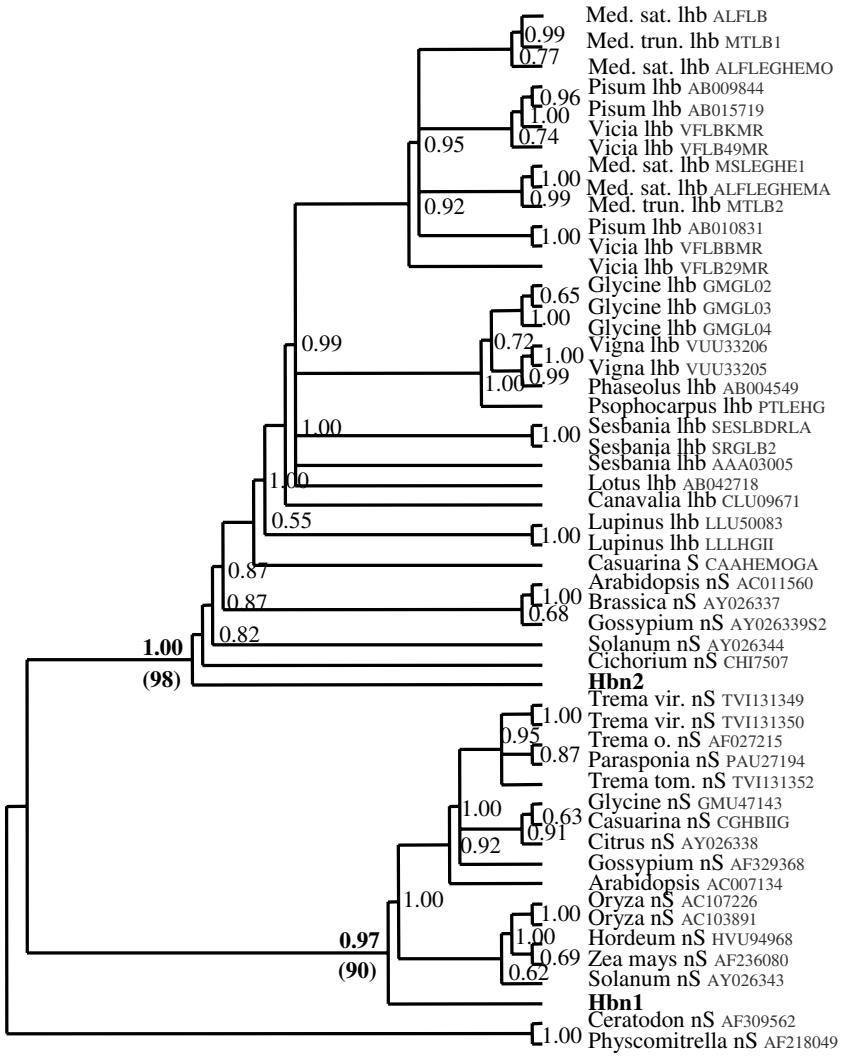

Fig. 2 Bayesian tree: 52 species, 132 amino acid sites. Clade posterior probabilities are given for each node. Neighbour-Joining bootstrap supports (1000 replicates) are indicated within parenthesis. lhb: symbiotic haemoglobin of legumes; S: nonlegume symbiotic gene; nS: nonsymbiotic gene.

Evolutionary rates of plant haemoglobins

As long as it is assumed that the right topology has been inferred, the estimation of the branch lengths of a phylogeny may provide useful information about molecular evolutionary processes. In the case of the plant haemoglobin phylogeny, several nodes remain unresolved within each group of genes (Fig. 2). However, at this time we have sufficiently reliable knowledge of angiosperm phylogeny (Soltis *et al.*, 1999; Soltis *et al.*, 2000), which can help to resolve some of these ambiguities. We thus modified the haemoglobin gene phylogeny to make it in agreement with the canonical angiosperm phylogeny, grouping together genes belonging to the same family or order. No change was made within the legume class 2 group, in which many recent duplication events have occurred, confusing species phylogeny. Branch lengths of the haemoglobin gene tree were then estimated by fitting PAM distances to the modified topology (least square fit). This 'true' haemoglobin phylogeny (Fig. 3) shows longer branches for the class 2 sequences, and especially for symbiotic genes. The level of significance of this trend was assessed using a generalization of the relative-rate test to more than three species (Robinson-Rechavi *et al.*, 1998). The difference of amino-acid substitution rates between class 1 and class 2 genes, using moss sequences as out-group, was found to be highly significant ($P < 0.001$).

The acceleration observed for class 2 genes might have been caused by an increase of mutation rate (e.g. if located in a rapidly evolving genomic region). Alternatively, this pattern might reflect a change in the selective forces applying to the haemoglobin gene. To distinguish between these two hypotheses, we contrasted the synonymous (*Ks*) and nonsynonymous (*Ka*) evolutionary rates for each sequence pair within two groups of class 1 genes (eudicots and monocots respectively) on one hand, and the group of symbiotic class 2 genes of legumes on

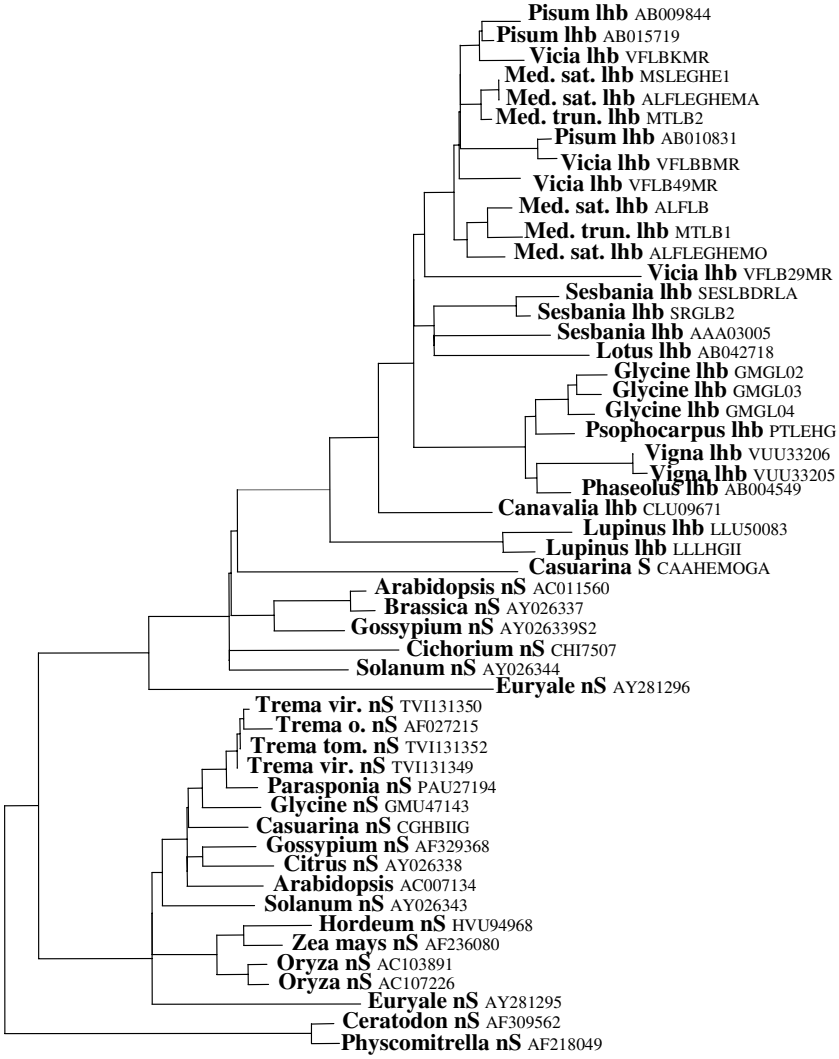

Fig. 3 Haemoglobin gene phylogeny: 52 species, 132 amino acid sites. Branch lengths were estimated by fitting PAM distances between amino-acid sequences to a modified haemoglobin tree topology (see text).

the other hand. Under the hypothesis of a variable mutation rate but a constant selective regime, an equal *Ks*/*Ka* ratio would be expected in all groups. Figure 4 leads to rejection of this hypothesis: the *Ks*/*Ka* ratio is much lower in the class 2 symbiotic group (average: 4.6) than in class 1 groups (average: 9.3 for eudicots and 12.5 for monocots). This indicates that the selective forces applying (or having applied) to symbiotic class 2 haemoglobins are distinct from those applying to class 1 haemoglobins.

These within-group comparisons do not reflect the evolutionary process that accompanied the functional divergence of the two genes after the duplication, but rather what occurred after their current function was acquired. Again, two main hypotheses might be invoked to explain the observed difference in *Ks*/*Ka* ratio between the two lineages. First, a lower *Ks*/*Ka* in class 2 genes might result from the relaxation of functional constraints, i.e. a decrease of the strength of purifying selection. Amino acids might be freer to vary in class 2 than in class 1 genes, for a yet undetermined reason. This is the simplest explanation, consistent with the neutralist theory, and should be considered as the null hypothesis. Alternatively, recurrent adaptation might also explain an increase of nonsynonymous substitution rate in class 2 symbiotic genes. Advantageous (nonsynonymous) mutations have a higher fixation probability than neutral mutations, and contribute to decreasing the *Ks*/*Ka* ratio.

The reason why the adaptive hypothesis could be considered as a reasonable alternative to the neutralist explanation in this case is that the legume haemoglobins are functionally involved in the symbiosis between legumes and Rhizobium. Interactions between species potentially generate the need for a recurrent adaptation (Red-Queen-like evolution), one species having to adapt to the innovations found by the other one. Nitrogen

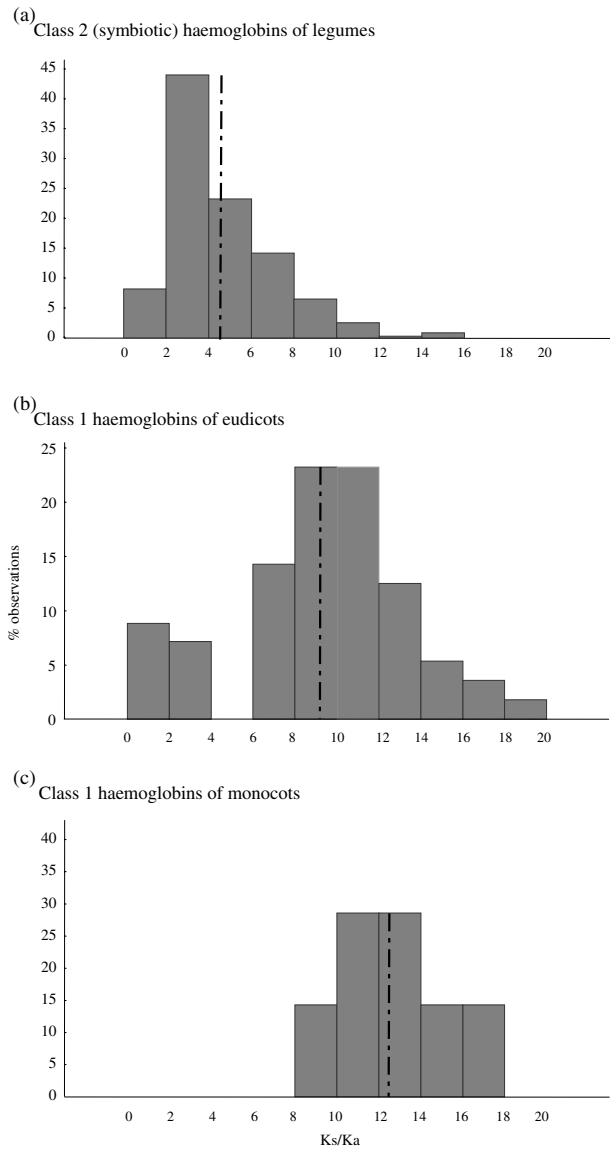

Fig. 4 Distribution of pairwise *Ks/Ka* ratios for legume symbiotic genes (a), eudicots class 1 genes (b) and monocots class 1 genes (c). Within each group, synonymous and nonsynonymous distances were estimated for all sequence pairs. The vertical lines indicate the average values.

fixation by bacteria, for instance, involves the nitrogenase enzyme, which requires substantial amounts of energy (i.e. ATP) produced by bacteroid respiration. However, the oxygen necessary for respiration readily inhibits the activity of nitrogenase. Haemoglobin thus appears to be a key factor in the process of symbiotic exchange by facilitating oxygen diffusion to the nitrogen-fixing symbionts (Appleby, 1992). From an evolutionary perspective, increasing or decreasing the affinity of haemoglobin to oxygen might therefore be a way for legumes to control the bacterial growth and amount of symbiotic exchange or to adapt to a new strain of bacteria (Bever & Simms, 2000; Denison, 2000; West *et al.*, 2002).

Adaptation at the sequence level can definitely be invoked when *Ka* is higher than *Ks,* a pattern incompatible with neutral evolution or purifying selection. This is not what we observed for haemoglobin. Not observing *Ka/Ks,* however, does not imply that adaptation has not occurred. Indeed, this adaptive evolution might have involved only a (small) fraction of sites, and/or have occurred in specific branches of the tree. *Ks/Ka* measurements averaged over sites and over lineages make the detection of episodes of adaptive evolution difficult, and additional analyses involving models allowing for variable *Ks/Ka* ratio among sites/branches (Goldman & Yang, 1994; Yang, 1997; Yang, 2000) are required for distinguishing between the two hypotheses.

References

Andersson, C.R., Jensen, E.O., Llewellyn, D.J., Dennis, E.S. & Peacock, W.J. 1996. A new haemoglobin gene from soybean : a role for haemoglobin in all plants. *Proc. Natl. Acad. Sci. U. S. A.* **93**: 5682–5687.

Andersson, C.R., Llewellyn, D.J., Peacock, W.J. & Dennis, E.S. 1997. Cell-specific expression of the promoters of two nonlegume haemoglobin genes in a transgenic legume, *Lotus corniculatus*. *Plant Physiol.* **113**: 45–57.

Appleby, C. 1992. The origin and functions of haemoglobin in plants. *Sci. Progress* **76**: 365–398.

Arredondo-Peter, R., Hargrove, M.S., Moran, J.F., Sarath, G. & Klucas, R.V. 1998. Plant haemoglobins. *Plant Physiol.* **118** 1121–1125.

Arredondo-Peter, R., Ramirez, M., Sarath, G. & Klucas, R.V. 2000. Sequence analysis of an ancient haemoglobin cDNA isolated from the moss *Physcomitrella patens*. *Plant Physiol.* **122**: 1457.

Bever, J.D. & Simms, E.L. 2000. Evolution of nitrogen fixation in spatially structured populations of Rhizobium. *Heredity* **85**: 366–372.

Bogusz, D., Appleby, C., Landsmann, J., Dennis, E.S., Trinick, M.J. & Peacock, W.J. 1988. Functioning haemoglobin genes in non-nodulating plants. *Nature* **331**: 178–180.

Bogusz, D., Llewellyn, D.J., Craig, S., Dennis, E.S., Appleby, C. & Peacock, W.J. 1990. Nonlegume haemoglobin genes retain organ-specific expression in heterologous transgenic plants. *The Plant Cell* **2**: 633–641.

Dayhoff, M.O., Schwartz, R.M. & Orcutt, B.C. 1978. A model of evolutionary change in proteins; Matrices for detecting distant relationships. In: *Atlas of protein sequence and structure*, pp. 345–358. National Biomedical Research Foundation, Washington, DC.

Denison, R.F. 2000. Legume sanctions and the evolution of symbiotic cooperation by Rhizobia. *Am. Nat.* **156**: 567–576.

Duff, S.M.G., Guy, P.A., Nie, X., Durnin, D.C., Hill, R.D. & Nie, W.Z. 1998. Haemoglobin expression in germinating barley. *Seed Sci. Res.* **8**: 431–436.

Felsenstein, J. 1978. The number of evolutionary trees. *Syst. Zool.* **27**: 27–33.

Felsenstein, J. 1989. PHYLIP – Phylogeny Inference Package (version 3.2). *Cladistics* **5**: 164–166.

Franche, C., Diouf, D., Laplaze L, Auguy, F., Frutz, T., Rio, M., Duhoux, E. & Bogusz, D. 1998. Soybean (*lbc3*), *Parasponia*, and *Trema* haemoglobin gene promoters retain symbiotic specificity in transgenic Casuarinaceae : implications for haemoglobin gene evolution and root nodule symbioses. *Mol. Plant-Microbe Interact* **11**: 887–894.

Galtier, N., Gouy, M. & Gautier, C. 1996. SEAVIEW and PHYLO_WIN : two graphic tools sequence alignment and molecular phylogeny. *CABIOS* **12**: 543–548.

Goldman, N. & Yang, Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Biol. Mol. Evol.* **11**: 725–736.

Hendriks, T., Scheer, I., Quillet, M.-C., Randoux, B., Delbreil, B., Vasseur, J. & Hilbert, J.-L. 1998. A nonsymbiotic haemoglobin gene is expressed during somatic embryogenesis in *Cichorium*. *Biochemica et Biophysica Acta* **1143**: 193–197.

Huelsenbeck, J. & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hunt, R.A., Watts, R.A., Trevaskis, B., Llewellyn, D.J., Burnell, J., Dennis, E.S. & Peacock, W.J. 2001. Expression and evolution of functionally distinct haemoglobin genes in plants. *Plant Mol. Biol.* **47**: 677–692.

Jacobsen-Lyon, K., Jensen, E.O., Jorgensen, J.-E., Marcker, K.A., Peacock, W.J. & Dennis, E.S. 1995. Symbiotic and nonsymbiotic haemoglobin genes of *Casuarina glauca*. *The Plant Cell* **7**: 213–223.

Jensen, E.O., Paludan, K., Hyldig-Nielsen, J.J., Jorgensen, P. & Marcker, K.A. 1981. The structure of a chromosomal leghaemoglobin gene from soybean. *Nature* **291**: 677–679.

Jones, D.T., Taylor, W.R. & Thornton, J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**: 275–282.

Landsmann, J., Dennis, E.S., Higgins, T.J.V., Appleby, C., Kortt, A. & Peacock, W.J. 1986. Common evolutionary origin of legume and non-legume plant haemoglobins. *Nature* **324**: 166–168.

Ota, S. & Li, W. 2001. NJML+ : an extension of the NJML method to handle protein sequence data and computer software implementation. *Mol. Biol. Evol.* **18**: 1983–1992.

Robinson-Rechavi, M. & Huchon, D. 2000. RRTree : relative-rate tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**: 296–297.

Robinson-Rechavi, M., Gouy, M., Gautier, C. & Mouchiroud, D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* **15**: 1091–1098.

Saitou, N. & Nei, M. 1987. The neighbor-joining method : a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.

Soltis, P.S., Soltis, D.E. & Chase, M.W. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–403.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W.H., Hoot, S.B., Fay, M.F., Axtel, M., Swensen, S.M., Prince, L., Kress, W.J., Nixon, K.C. & Farris, J. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atp*B sequences. *Bot. J. Linn. Soc.* **133**, 381–461.

Strozycki, P.M. & Legocki, A.B. 1995. Leghaemoglobins from an evolutionarily old legume, *Lupinus luteus*. *Plant Science* **110**: 83–93.

Strozycki, P.M., Karlowski, W.M., Dessaux, Y., Petit, A. & Legocki, A.B. 2000. Lupine *leghaemoglobin I* : expression in transgenic *Lotus* and tobacco tissues. *Mol. Gen. Genet.* **263**: 173–182.

Szabados, L., Ratet, P., Grunenberg, B. & Bruijn, F.G. 1990. Functional analysis of the *Sesbania rostrata* leghaemoglobin *glb3* gene 5′-upstream region in transgenic *Lotus corniculatus* and *Nicotiana tabacum* plants. *Plant Cell* **2**: 973–986.

Trevaskis, B., Watts, R.A., Andersson, C.R., Llewellyn, D.J., Hargrove, M.S., Olson, J.S., Dennis, E.S. & Peacock, W.J. 1997. Two haemoglobin genes in Arabidopsis thaliana : the evolutionary origins of leghaemoglobins. *Proc. Natl. Acad. Sci. U. S. A.* **94**: 12230–12234.

West, S.A., Kiers, E.T., Simms, E.L. & Denison, R.F. 2002. Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. Natl. Acad. Sci. U. S. A.* **269**: 685–694.

Wu, C. & Li, W.H. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U. S. A.* **82**: 1741–1745.

Yang, Z. 1997. PAML : a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**: 555–556.

Yang, Z. (2000). *Phylogenetic Analysis by Maximum Likelihood (PAML)*, Version 3.0. University College London, London, England (http://abacus.gene.ucl.ac.uk/software/paml.html)

Yang, Z. & Nielsen, R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**: 32–43.

Received 23 May 2003; revised 27 September 2003; accepted 27 September 2003