Communicative & Integrative Biology 6:6, e27496; November/December; © 2013 Landes Bioscience

Variability of non-symbiotic and truncated hemoglobin genes from the genome of cultivated monocots

Gustavo Rodríguez-Alonso and Raúl Arredondo-Peter*

Laboratorio de Biofísica y Biología Molecular; Departamento de Bioquímica y Biología Molecular; Facultad de Ciencias; Universidad Autónoma del Estado de Morelos; Cuernavaca, Morelos, México

Abbreviations: Hb, hemoglobin; MCL, maximum composite likelihood; nsHb, non-symbiotic hemoglobin; tHb, truncated hemoglobin

Keywords: GC content, cereals, globin, neutral selection

Non-symbiotic (nsHb) and truncated (tHb) hemoglobins (Hbs) have been detected in a variety of land plants. The evolution of land plant nsHbs and tHbs at the protein level is well documented; however, little is known about the evolution of genes coding for these proteins. For example, the variability of the land plant *nshb* and *thb* genes is not known. Here, we report the variability of the *nshb* and *thb* genes from the genome of the cultivated monocots *Brachypodium distachyon*, *Hordeum vulgare* (barley), *Oryza glaberrima* (rice), *O. rufipogon* (rice), *O. sativa* (rice) var indica, *O. sativa* (rice) var japonica, *Panicum virgatum* (switchgrass), *Setaria italica* (foxtail millet), *Sorghum bicolor* (sorghum), *Triticum aestivum* (wheat), and *Zea mays* ssp. *mays* (maize) using sequence comparison and computational methods. Our results revealed that in cultivated monocots variability is higher in *nshb*s than in *thb*s, and suggest that major substitution events that occurred during the evolution of the cultivated monocot *hb*s were A→G and T→C transitions and that these genes evolved under the effect of neutral selection.

Non-symbiotic (nsHb) and truncated (tHb) hemoglobins (Hbs) are O_2-binding proteins that have been detected in a variety of land plants, ranging from primitive bryophytes to evolved angiosperms.[1-5] Phylogenetic analysis revealed that land plant *nshb* and *thb* genes apparently evolved from different ancestors. Also, sequence analysis showed that 2 classes of nsHbs exist in higher plants: class 1 and class 2 nsHbs (nsHb-1 and nsHb-2, respectively).[6,7] However, recent analysis revealed that apparently only nsHbs-1 exist in cultivated monocots, which diverged into clade I and clade II nsHbs (nsHb-I and nsHb-II, respectively) from a nsHb-1 ancestor.[8]

The evolution of land plant nsHbs and tHbs at the protein level is well documented.[3,4,8-10] However, little is known about the evolution of genes coding for these proteins. For example, the variability of the land plant *nshb* and *thb* genes is not known. Monocots are useful models for gene analysis because of the availability of several (fully or partially) sequenced genomes from these plants. Here, we report the variability of the *nshb* and *thb* genes from the cultivated monocots *Brachypodium distachyon*, *Hordeum vulgare* (barley), *Oryza glaberrima* (rice), *O. rufipogon* (rice), *O. sativa* (rice) var indica, *O. sativa* (rice) var japonica, *Panicum virgatum* (switchgrass), *Setaria italica* (foxtail millet), *Sorghum bicolor* (sorghum), *Triticum aestivum* (wheat), and *Zea mays* ssp. *mays* (maize). Our results revealed that in cultivated monocots, variability is higher in *nshb*s than in *thb*s and that these genes evolved under the effect of neutral selection.

Protein and gene sequences for nsHbs and tHbs from the above cultivated monocots were obtained from the GenBank (www.ncbi.nlm.nih.gov/genbank) and Phytozome (www.phytozome.org) databases as described by Rodríguez-Alonso and Arredondo-Peter[8] (**Tables S1** and **S2**) using the sequence of *O. sativa* nsHbs 1 to 5 and tHb (GenBank accession number AAK72229.1, AAK72228.1, AAK72230.1, AAK72231.1, ABN45744.1, and EEC80902.1, respectively) as probes. Pairwise sequence alignment was performed using the Needle program (http://www.ebi.ac.uk/Tools/psa/emboss_needle/). Differences between the aligned sequences were quantitated using a Phyton's script developed by one of the authors (Rodríguez-Alonso G). Values were normalized based on the number of aligned sequences. The maximum composite likelihood (MCL) matrix and GC content quantitation were obtained using the MEGA 5.0 program.[11] Testing of the neutral mutation hypothesis was performed by calculating the Tajima's

*Correspondence to: Raúl Arredondo-Peter; Email: ra@uaem.mx
Submitted: 11/28/2013; Accepted: 12/12/2013
Citation: Rodríguez-Alonso G, Arredondo-Peter R. Variability of non-symbiotic and truncated hemoglobin genes from the genome of cultivated monocots. Communicative & Integrative Biology 2013; 6:e27496; http://dx.doi.org/10.4161/cib.27496

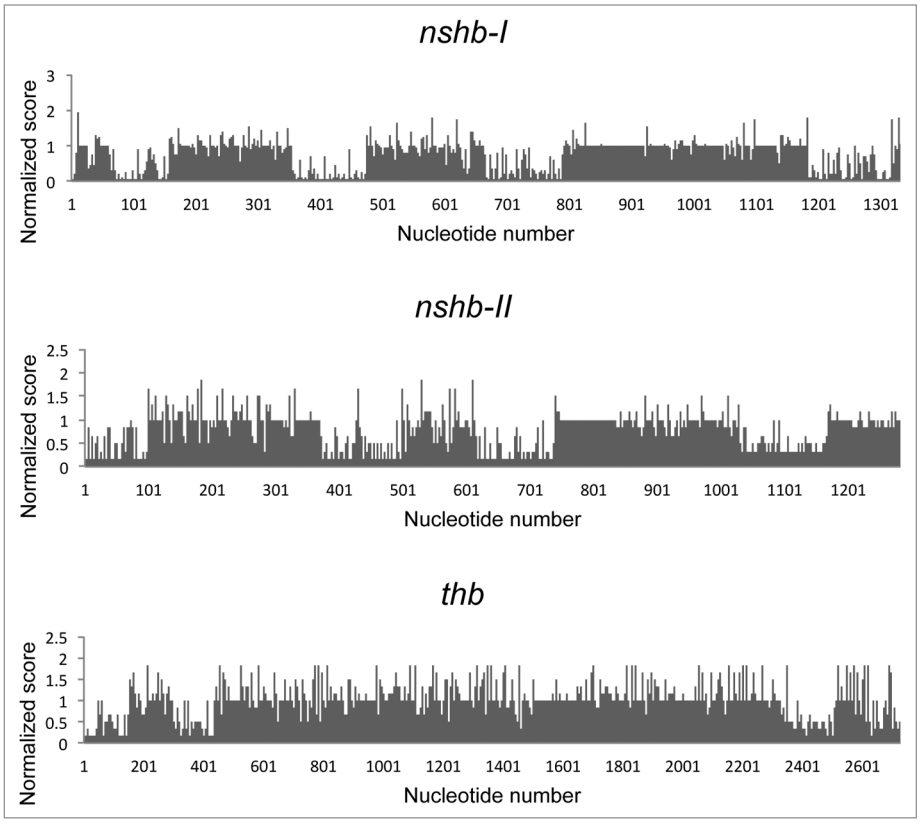


Figure 1. Variability of the *nshb-I*, *nshb-II* and *thb* genes from cultivated monocots. High and low score regions approximately correspond to introns and exons into the *nshb-I* and *nshb-II* genes, respectively.

		A	T	C	G	
nshb-I	A	-	2.69	4.57	30.72	R= 2.144
	T	3.4	-	10.97	5.22	
	C	3.4	6.45	-	5.22	
	G	20.6	2.69	4.57	-	
nshb-II	A	-	2.74	5.49	17.44	R= 1.463
	T	3.8	-	23.36	7.02	
	C	3.8	11.64	-	7.02	
	G	9.45	2.74	5.49	-	
thb	A	-	4.28	6.63	11.27	R= 1.077
	T	6.01	-	19.07	6.71	
	C	6.01	12.31	-	6.71	
	G	10.1	4.28	6.63	-	

Figure 2. Maximum composite likelihood (MCL) matrix for *nshb-I*, *nshb-II*, and *thb* genes from cultivated monocots. The *R* value indicates the transition/transversion ratio.

D value[12] using the DNAsp program (http://www.ub.edu/dnasp/).

Figure 1 shows that in cultivated monocot *nshb-I* and *nshb-II* genes the 3 introns interrupting land plant *nshb*s[1,4,13] are differentiated from exons because of the existence of high and low variability scores, respectively. This indicates that in cultivated monocot *nshb-I* and *nshb-II* genes, the variability of introns is higher than that of exons. In contrast, in cultivated monocot *thb* genes the variability score of exons and introns is similar. Quantitation of similarity between pairwise sequence alignments of cultivated monocot Hb proteins and exons from the *hb* genes showed that the average variability values (obtained from the similarity values reported in **Figures S1** and **S2**) are 30.36% for nsHbs, 25.52% for *nshb*s, 11.46% for tHbs, and 11.23% for *thb*s. This result indicates that in cultivated monocots, the average variability of nsHbs is higher than that of tHbs.

Direct quantitation of similarity from aligned sequences assumes that nucleotides exist approximately in the same proportion in the genomes of organisms and that point mutations occurred with the same frequency during the evolution of genomes. However, this is generally incorrect. Also, this method does not consider regressions. Thus, we generated MCL matrices to evaluate substitution rates for point mutations and regression frequencies into the cultivated monocot *hb* genes. **Figure 2** shows that the transition/transversion ratios (R) are 2.144 for *nshb-I*, 1.463 for *nsb-II*, and 1.077 for *thb* genes. This result suggests that major substitution events that occurred during the evolution of the cultivated monocot *hb*s were A→G and T→C transitions, and thus indicates that in these genes the GC content is high. Quantitation of the GC content in codons from exons of cultivated monocot *hb*s showed that GC content is slightly higher in *nshb*s than in *thb*s and that GC content in these genes is higher (~70–90%) in codon position 3 than in codon positions 1 and 2 (~55–60 and ~40%, respectively) (**Fig. 3**).

Testing of the neutral mutation hypothesis for the evolution of cultivated monocot *hb* genes was performed by estimating the Tajima's D value.[12] Results showed that the Tajima's D value for cultivated monocot *nshb* and *thb* genes is 0.39 and –0.272, respectively, and that $P > 0.1$ for both estimations. Thus, these estimations are not statistically different from values expected in a neutralist model. These results suggest that cultivated monocot *nshb* and *thb* genes evolved under the effect of neutral selection.

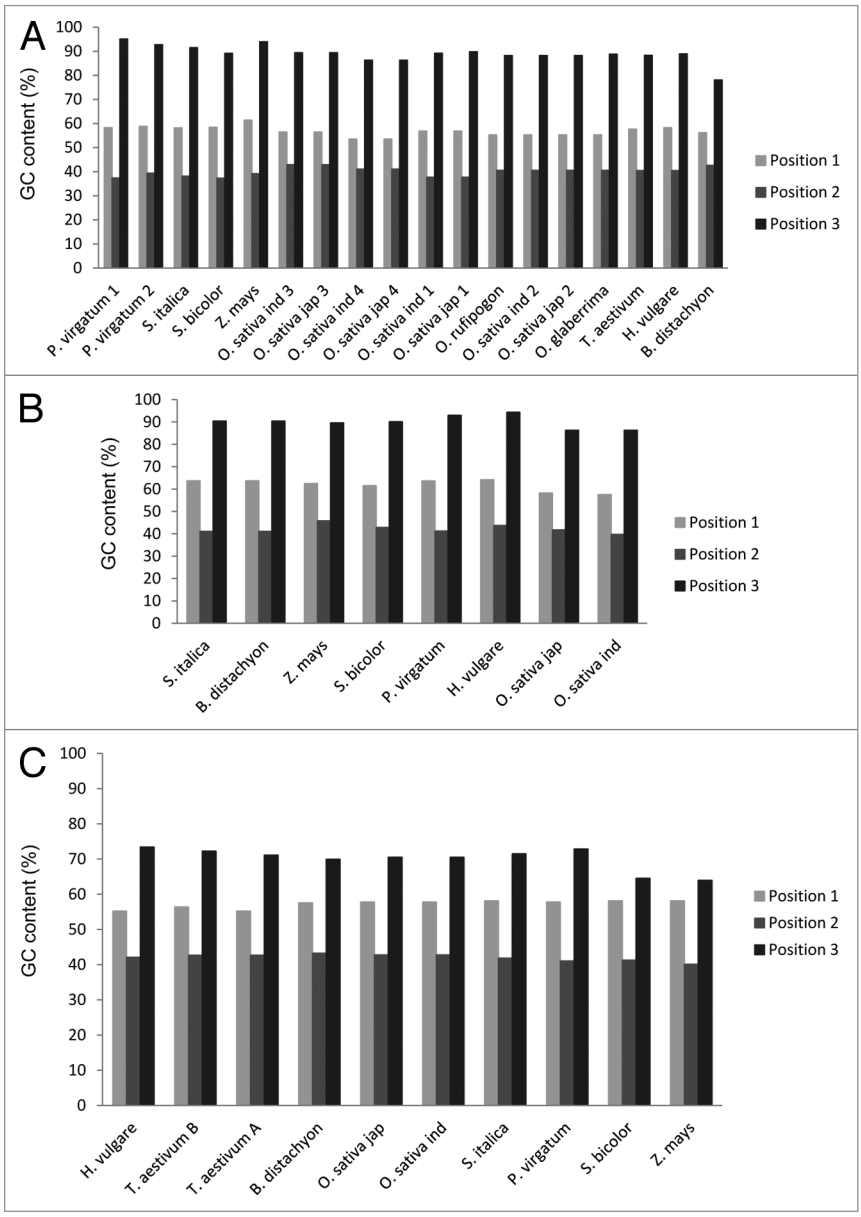
Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors wish to express their gratitude to Miss Gillian Klucas for English corrections. Work in the authors' laboratory has been funded by SEP-PROMEP (grant no.

Figure 3. GC content in codon positions 1, 2, and 3 from the cultivated monocot *nshb-I* (**A**), *nshb-II* (**B**), and *thb* (**C**) genes.



UAEMor-PTC-01–01/PTC23) and Consejo Nacional de Ciencia y Tecnología (CoNaCyT grant nos. 25229N and 42873Q), México. Rodríguez-Alonso G was a graduate student financed by CoNaCyT (reg. no. 256280).

Supplemental Materials

Supplemental materials may be found here: www.landesbioscience.com/journals/CIB/article/27496

References

1. Garrocho-Villegas V, Gopalasubramaniam SK, Arredondo-Peter R. Plant hemoglobins: what we know six decades after their discovery. Gene 2007; 398:78-85; PMID:17540516; http://dx.doi.org/10.1016/j.gene.2007.01.035
2. Vázquez-Limón C, Castro-Bustos S, Arredondo-Peter R. Spectroscopic analysis of moss (*Ceratodon purpureus* and *Physcomitrella patens*) recombinant non-symbiotic hemoglobins. Commun Integr Biol 2012; 5:527-30; PMID:23336017; http://dx.doi.org/10.4161/cib.21473
3. Vázquez-Limón C, Hoogewijs D, Vinogradov SN, Arredondo-Peter R. The evolution of land plant hemoglobins. Plant Sci 2012; 191-192:71-81; PMID:22682566; http://dx.doi.org/10.1016/j.plantsci.2012.04.013
4. Vinogradov SN, Fernández I, Hoogewijs D, Arredondo-Peter R. Phylogenetic relationships of 3/3 and 2/2 hemoglobins in Archaeplastida genomes to bacterial and other eukaryote hemoglobins. Mol Plant 2011; 4:42-58; PMID:20952597; http://dx.doi.org/10.1093/mp/ssq040
5. Watts RA, Hunt PW, Hvitved AN, Hargrove MS, Peacock WJ, Dennis ES. A hemoglobin from plants homologous to truncated hemoglobins of microorganisms. Proc Natl Acad Sci U S A 2001; 98:10119-24; PMID:11526234; http://dx.doi.org/10.1073/pnas.191349198
6. Smagghe BJ, Hoy JA, Percifield R, Kundu S, Hargrove MS, Sarath G, Hilbert JL, Watts RA, Dennis ES, Peacock WJ, et al. Review: correlations between oxygen affinity and sequence classifications of plant hemoglobins. Biopolymers 2009; 91:1083-96; PMID:19441024; http://dx.doi.org/10.1002/bip.21256
7. Trevaskis B, Watts RA, Andersson CR, Llewellyn DJ, Hargrove MS, Olson JS, Dennis ES, Peacock WJ. Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. Proc Natl Acad Sci U S A 1997; 94:12230-4; PMID:9342391; http://dx.doi.org/10.1073/pnas.94.22.12230
8. Rodríguez-Alonso G, Arredondo-Peter R. Phylogenetic analysis reveals an apparent duplication of the non-symbiotic hemoglobin 1 gene early in the evolution of monocotyledonous plants. ScienceJet 2012; 1:27
9. Arredondo-Peter R. Evolutionary rates of land plant hemoglobins at the protein level. Global J Biochem 2011; 2:81-95
10. Vinogradov SN, Hoogewijs D, Arredondo-Peter R. What are the origins and phylogeny of plant hemoglobins? Commun Integr Biol 2011; 4:443-5; PMID:21966566
11. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011; 28:2731-9; PMID:21546353; http://dx.doi.org/10.1093/molbev/msr121
12. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 1989; 123:585-95; PMID:2513255
13. Garrocho-Villegas V, Arredondo-Peter R. Molecular cloning and characterization of a moss (*Ceratodon purpureus*) nonsymbiotic hemoglobin provides insight into the early evolution of plant nonsymbiotic hemoglobins. Mol Biol Evol 2008; 25:1482-7; PMID:18420592; http://dx.doi.org/10.1093/molbev/msn096