Dynamic Evolution of Nitric Oxide Detoxifying Flavohemoglobins, a Family of Single-Protein Metabolic Modules in Bacteria and Eukaryotes

Jennifer H. Wisecaver,[1] William G. Alexander,[2] Sean B. King,[1] Chris Todd Hittinger,[2] and Antonis Rokas*,[1]

[1]Department of Biological Sciences, Vanderbilt University

[2]Laboratory of Genetics, Genome Center of Wisconsin, DOE Great Lakes Bioenergy Research Center, Wisconsin Energy Institute, J. F. Crow Institute for the Study of Evolution, University of Wisconsin–Madison, Madison

***Corresponding author:** E-mail: antonis.rokas@vanderbilt.edu.

Associate editor: Miriam Barlow

Abstract

Due to their functional independence, proteins that comprise standalone metabolic units, which we name single-protein metabolic modules, may be particularly prone to gene duplication (GD) and horizontal gene transfer (HGT). Flavohemoglobins (flavoHbs) are prime examples of single-protein metabolic modules, detoxifying nitric oxide (NO), a ubiquitous toxin whose antimicrobial properties many life forms exploit, to nitrate, a common source of nitrogen for organisms. FlavoHbs appear widespread in bacteria and have been identified in a handful of microbial eukaryotes, but how the distribution of this ecologically and biomedically important protein family evolved remains unknown. Reconstruction of the evolutionary history of 3,318 flavoHb protein sequences covering the family's known diversity showed evidence of recurrent HGT at multiple evolutionary scales including intrabacterial HGT, as well as HGT from bacteria to eukaryotes. One of the most striking examples of HGT is the acquisition of a flavoHb by the dandruff- and eczema-causing fungus *Malassezia* from *Corynebacterium* Actinobacteria, a transfer that growth experiments show is capable of mediating NO resistance in fungi. Other flavoHbs arose via GD; for example, many filamentous fungi possess two flavoHbs that are differentially targeted to the cytosol and mitochondria, likely conferring protection against external and internal sources of NO, respectively. Because single-protein metabolic modules such as flavoHb function independently, readily undergo GD and HGT, and are frequently involved in organismal defense and competition, we suggest that they represent "plug-and-play" proteins for ecological arms races.

Key words: **phylogenetics, gene tree-species phylogeny reconciliation, gene innovation, horizontal gene transfer, gene duplication, fungi, *Malassezia*.**

Introduction

An organism's total metabolic network can be subdivided into modules, which correspond to functional units composed of enzymatic reactions and protein complexes (Muto et al. 2013; Kanehisa et al. 2014). Interestingly, the evolutionary modes of proteins that make up these metabolic modules are often correlated with module complexity. For example, proteins that participate in small metabolic modules are typically more variable in their gene copy numbers (Wapinski et al. 2007) as well as more likely to undergo gene duplication (GD) (Li et al. 2006; Prachumwat and Li 2006) and horizontal gene transfer (HGT) (Cohen et al. 2011) than proteins that interact and function as parts of large multiprotein modules.

Taken one step further, the correlation between evolutionary mode and module complexity predicts that the proteins comprising the smallest metabolic modules—those encoded by a single gene whose product catalyzes one self-contained metabolic reaction, which we name single-protein metabolic modules—may be particularly amenable to GD and HGT. Intriguingly, recent studies on two such protein families, namely the cytolytic toxins aerolysins and the antibacterial lysozymes, show that both exhibit complex evolutionary histories marked by GD and recurrent HGT (Danchin et al. 2010; Moran et al. 2012; Metcalf et al. 2014).

Proteins in the flavohemoglobin family (flavoHbs) are prime examples of single-protein metabolic modules. FlavoHbs catalyze the nitric oxide (NO) dioxygenase reaction that converts NO, a ubiquitous but toxic nitrogen compound, to innocuous nitrate, a common source of nitrogen for many organisms. FlavoHbs are widespread in bacteria (Forrester and Foster 2012), and functional characterization of these proteins in Proteobacteria, Firmicutes, and Actinobacteria suggests that they confer protection from both endogenous NO, a natural byproduct of denitrification metabolism, as well as extracellular sources of NO that is released as an antimicrobial by plant and animal hosts and by competing microorganisms (Cramm et al. 1994; Crawford and Goldberg 1998; Hu et al. 1999; Stevanin et al. 2002; Frey and Kallio 2003).

FlavoHbs have also been functionally characterized in several fungal species (Zhu and Riggs 1992; Takaya et al. 1997; Ullmann et al. 2004; Zhou et al. 2011). For example, the

Article

baker's yeast *Saccharomyces cerevisiae* possesses a single gene encoding a flavoHb protein (Yhb1). Yhb1 is targeted to the cytosol under aerobic conditions where it mitigates stress due to exogenous NO, whereas under anaerobic conditions the protein is targeted to the mitochondria where it likely protects from endogenous mitochondrial NO production (Cassanova et al. 2005). In contrast, filamentous fungi in the genus *Aspergillus* possess two flavoHb isoforms that localize to the cytosol and mitochondria and protect these cells from exogenous and endogenous NO, respectively (Zhou et al. 2011). More recently, flavoHbs have been identified in a handful of nonfungal microbial eukaryotes (Iijima et al. 2000; Mastronicola et al. 2010; Rafferty et al. 2010; Runckel et al. 2014). Although, the full taxonomic distribution and evolutionary history of flavoHbs is not yet known, all hypotheses for the origin of eukaryotic flavoHbs include some form of HGT from bacteria, either through a single transfer event from the alpha-proteobacterial ancestor of the mitochondria to an ancestral eukaryote or through multiple independent events from diverse bacterial donors into different eukaryotic lineages (Vinogradov et al. 2005, 2013; Hoogewijs et al. 2012).

To better understand the evolutionary diversification of this ecologically and biomedically significant single-protein metabolic module, we performed a comprehensive phylogenetic analysis of flavoHbs and investigated how evolutionary processes, such as GD and HGT, shaped its history and function. The phylogeny of 3,318 flavoHbs shows evidence for recurrent HGT across multiple evolutionary scales, including between bacterial divisions and between the bacterial and eukaryotic domains of life. All of these horizontally transferred flavoHbs appear functional and likely to mediate NO resistance; indeed, our experiments here validate that a fungal flavoHb recently acquired via HGT from Actinobacteria does function in alleviating NO stress. Lastly, examination of the distribution of N-terminal targeting peptides across the flavoHb phylogeny suggests that some eukaryotic flavoHbs arose via GD and are specialized to handle endogenous NO in the mitochondria. These results, together with the results from earlier studies of aerolysins and lysozymes, raise the hypothesis that single-protein metabolic modules, perhaps due to their functional independence and their propensity to undergo GD and HGT, are frequently involved in organismal defense and competition.

Results

Distribution of FlavoHbs across the Tree of Life

Our search identified 3,318 flavoHbs from 2,363 bacteria and 204 eukaryotes (table 1 and fig. 1, supplementary table S1, Supplementary Material online); flavoHbs were absent from Archaea, as previously reported (Forrester and Foster 2012). The 3,010 bacterial flavoHbs were distributed across ten bacterial divisions, with the majority found in Proteobacteria (1,749 flavoHbs), Actinobacteria (649 flavoHbs), and Firmicutes (562 flavoHbs). Notably, flavoHbs were not ubiquitously distributed across the bacterial phylogeny and were uncommon or absent in some well-represented divisions such as Bacteriodetes (having only 30 flavoHbs present in 509 examined genomes; supplementary table S2, Supplementary Material online) and Cyanobacteria (0 flavoHbs identified in 109 examined genomes; supplementary table S2, Supplementary Material online).

The majority of the 308 flavoHbs identified in eukaryotes were from Fungi (280 flavoHbs). Other eukaryotes possessing flavoHbs included dictyostelid cellular slime molds (Amoebozoa; 7 flavoHbs), filamentous water molds (Heterokonta; 12 flavoHbs), the intestinal parasite *Giardia* (Fornicata; 6 flavoHbs), and two trypanosomatid parasites of insects *Crithidia mellificae* and *Angomonas deanei* (Euglenozoa; 3 flavoHbs). Six flavoHb sequences from animals were discarded due to suspected genome assembly contamination; in all these cases, flavoHbs were encoded on extremely small and unplaced assembly scaffolds. Additionally, all but one of these sequences showed high percent identity to bacterial sequences (76.7–100% protein sequence identity), the exception being a sequence from *Hydra vulgaris* (Cnidaria), whose top BLAST hit (41.9% protein sequence identity) was the distantly related oomycete *Phytophthora infestans* (Heterokonta) (supplementary table S1, Supplementary Material online).

The History of FlavoHbs Is Incongruent With the Species Phylogeny

Of the 17 higher taxa represented in the maximum likelihood (ML) phylogeny of flavoHbs, three taxa were represented by only a single flavoHb, five were represented by a monophyletic clade of flavoHbs, and nine were represented by multiple flavoHb clades (table 1 and fig. 1). For example, flavoHbs in Actinobacteria formed two distantly related clades: Clade 1 (340 flavoHbs including the *Mycobacterium* spp. type I isoform; Thakur et al. 2014) and Clade 2 (308 flavoHbs including the *Mycobacterium* spp. type II isoform; Gupta et al. 2011). Similarly, fungal flavoHbs formed five distinct clades: Clade 1 (24 flavoHbs from budding yeasts in the Saccharomycetales including *S. cerevisiae*), Clade 2 (3 flavoHbs from *Schizosaccharomyces* spp. and 40 flavoHbs from other Saccharomycotina outside the Saccharomycetaceae, including the opportunistic human pathogen *Candida albicans*), Clade 3 (175 flavoHbs from species of Basidiomycota and Pezizomycotina), Clade 4 (36 flavoHbs from Pezizomycotina), and Clade 5 (2 flavoHbs from the dandruff-causing fungus *Malassezia globosa* and from the eczema-causing fungus *Malassezia sympodialis*).

To evaluate the statistical phylogenetic support for the nine higher taxa whose flavoHb phylogenetic distributions were not monophyletic, we performed approximately unbiased (AU) comparative topology tests (Shimodaira 2002). For three taxa (Amoebozoa, Bacteroidetes, and Heterokonta), the constrained ML phylogeny, in which all flavoHbs from a given lineage were forced to be monophyletic, was not statistically significantly worse than the unconstrained ML phylogeny depicted in figure 1 (table 1). In contrast, monophyly was significantly rejected for the remaining six taxa (Bacteria, Actinobacteria, Firmicutes, Proteobacteria, Eukaryota, and Fungi; P value ≤ 0.05; table 1).

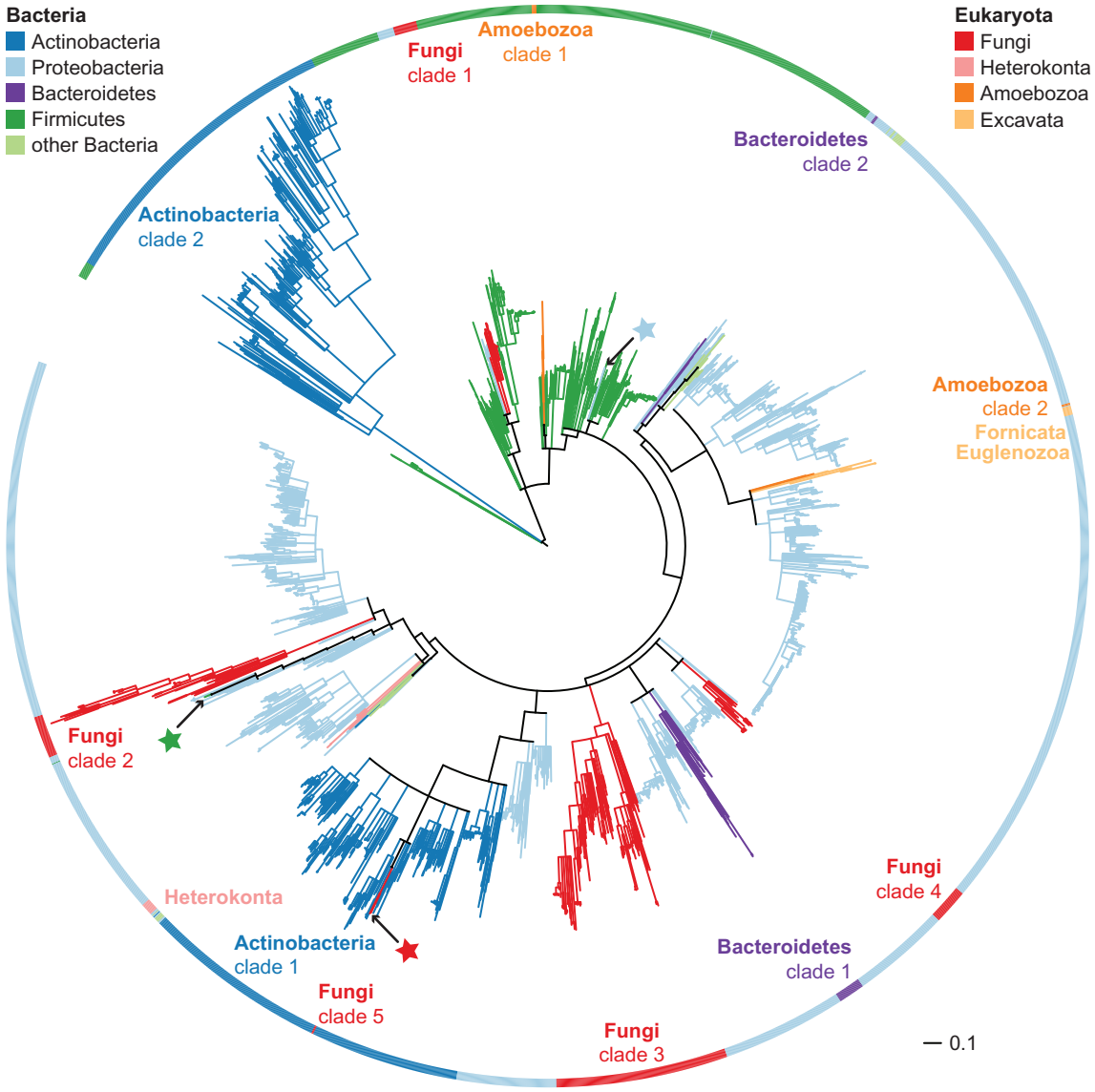

Fig. 1. ML phylogeny of flavoHbs from across the tree of life. The phylogeny was visualized using ITOL version 3.0 (Letunic and Bork 2011). Branches with SH-like local support less than 0.90 were collapsed using TREECOLLAPSECL version 4.0 (http://emmahodcroft.com/TreeCollapseCL.html, last accessed April 19, 2016). Colored stars indicate horizontally transferred flavoHbs in *Malassezia* (red) and *Chondromyces apiculatus* (blue), as well as a contamination-derived flavoHb in *Virgibacillus halodenitrificans* (green).

Similar to the patterns present across the full flavoHb phylogeny, the phylogenetic placement of flavoHbs within fungal Clade 3—the largest clade of fungal flavoHbs—is incongruent with the species phylogeny. Of the eight taxonomic classes present in fungal Clade 3, 1 was represented by only a single flavoHb, two were monophyletic, and five were not monophyletic (supplementary fig. S3, Supplementary Material online). Monophyly was significantly rejected for three classes (Agaricomycetes, Dothideomycetes, and Leotiomycetes; AU test, P value ≤ 0.05; supplementary table S3, Supplementary Material online). Interestingly, many abundant free-living or opportunistic plant and animal pathogens (e.g., species of *Neurospora*, *Aspergillus*, and *Fusarium*) contain multiple Clade 3 copies of flavoHb. In contrast, no table biotrophic plant pathogens such as species of *Puccinia*, *Ustilago*, and *Blumeria* appear to contain no flavoHbs.

The occurrence of multiple gene family members in a genome and their distribution across multiple distinct clades of the gene family phylogeny are telltale signs of gene innovation processes, such as GD and HGT (Gogarten and Townsend 2005). To reconstruct patterns of HGT, GD, and gene loss (GL) of flavoHb across the tree of life, we used a gene tree-species phylogeny reconciliation approach. For tractability (see Materials and Methods), we performed this analysis on a reduced set of taxa whose species phylogeny had limited polytomies and, critically, whose flavoHb gene tree was still representative of the entire data set (supplementary figs. S1 and S2, Supplementary Material online).

Duplicate FlavoHbs Can Have Specialized Functions

Regardless of the relative costs assigned to GD, GL, and HGT, hundreds of evolutionary events were required to reconcile

Table 1. Distribution of flavoHbs in Bacteria and Eukaryotes.

Lineage	No. FlavoHbs	No. Taxa	No. Genera	Monophyletic	AU test of Monophyly	
					Diff -lnL	*P* value
Bacteria	3,010	2,380	307	No	611.7	1.00E-45
Proteobacteria	1,749	1,447	182	No	986.6	3.00E-07
Actinobacteria	649	405	67	No	235.9	0.014
Firmicutes	562	481	39	No	473.1	0.001
Bacteroidetes	30	29	12	No	88.4	0.204
Planctomycetes	11	9	2	Yes (1.0)	—	—
Deinococcus-Thermus	4	4	1	Yes (0.611)	—	—
Acidobacteria	1	1	1	n/a	—	—
Chlamydiae	2	2	1	Yes (1.0)	—	—
Lentisphaerae	1	1	1	n/a	—	—
Verrucomicrobia	1	1	1	n/a	—	—
Eukaryota	308	204	92	No	573.7	3.00E-12
Fungi	280	186	85	No	367.6	4.00E-04
Heterokonta	12	7	2	No	28.2	0.421
Amoebozoa	7	5	2	No	120.4	0.12
Euglenozoa	3	2	2	Yes (0.886)	—	—
Fornicata	6	4	1	Yes (1.0)	–	—
All lineages	3,318	2,584	399	n/a	—	—

NOTE.—For each taxonomic lineage (row): number of flavoHb sequences, number of unique species/strains containing one or more flavoHbs, and number of unique genera containing one or more flavoHbs. Genus and species/strain identifiers are from NCBI's taxonomy database (see supplementary table S1, Supplementary Material online). Euglenozoa and Fornicata, both members of the eukaryotic kingdom Excavata, were also monophyletic with each other (branch support = 0.997). Diff −lnL, difference in −log-likelihood between the best topology and the constrained topology that forced the lineage to be monophyletic. The −lnL of best topology was −225972.4 in all cases.

Table 2. Gene Tree-Species Tree Reconciliation Suggests Multiple GD and HGT Events in FlavoHb Evolution.

Transfer Cost	Candidate Solutions	Feasible Solutions	Solutions Analyzed	Event Score		Duplication		Loss		Transfer		Recovers HGT to		
				Avg	SD	Avg	SD	Avg	SD	Avg	SD	*Vhal* (%)	*Capi* (%)	*Mal* (%)
5	403,200	374,400	6,000	815.0	0.7	122.3	1.1	558.0	4.2	26.9	1.1	0	100	100
7	440	440	1,570	850.4	0.5	144.3	0.9	647.5	3.9	8.4	0.6	80	100	100
9	116	116	116	860.0	0.0	151.0	1.0	686.5	3.5	2.5	0.5	0	0	100
11	58	58	58	867.0	0.0	154.0	0.0	698.0	0.0	1.0	0.0	0	0	100

NOTE.—*Vhal, Virgibacillus halodenitrificans; Mal, Malassezia; Capi, Chondromyces apiculatus*. Duplication and loss costs were set to one. Transfer costs of 5, 7, 9, 11, 13, and 15 were all evaluated (results for transfer costs of 13 and 15 were identical to 11). Event score is the sum of costs for all duplications, losses, and transfers. Total event score as well as duplication, loss, and transfer counts are represented by the average (avg) and standard deviation (SD) of all analyzed solutions. At transfer cost = 5, incongruence surrounding the *Vhal flavoHb* was reconciled by two HGT events into surrounding Proteobacteria, rather than HGT into *Vhal*.

the flavoHb gene tree against the species phylogeny (table 2). Specifically, the reconciled flavoHb gene tree contained on average 122.3–154.0 GD events and 558.0–698.0 GL events, depending on the relative cost of HGT; that is, when the cost of HGT was high, more GD and GL events were required to reconcile the flavoHb gene tree to the species phylogeny (table 2).

In some cases, the occurrence of GD is coincident with the evolution of specialized or even novel flavoHb-related functions. For example, our reconciliation analysis suggests that flavoHb clades 1 and 2 of Actinobacteria likely arose via GD. Interestingly, protein sequences in Clade 1 are Type I flavoHbs, which are functional NO dioxygenases with similar sequence properties to the canonical flavoHb of *Escherichia coli* (Thakur et al. 2014). In contrast, proteins in Clade 2 are type II flavoHbs and possess unusual structural features and biochemical properties that may enable them to metabolize D-lactate produced as a byproduct of lipid peroxidation, rather than to detoxify NO (Gupta et al. 2011, 2012).

Although fungal Clades 3 and 4 group separately in figure 1, the notung analysis suggests that these flavoHb sequences also likely arose via GD. In support of this hypothesis, a constrained phylogeny in which Clades 3 and 4 were forced to be monophyletic was not statistically significantly worse than the unconstrained ML phylogeny (AU test, diff −lnL = 18.75, *P* value = 0.43). These flavoHb clades appear to have evolved specialized functions: specifically, analysis of N-terminal targeting peptides suggests that Clade 3 flavoHbs are primarily targeted to the cytosol, whereas Clade 4 flavoHbs are primarily targeted to mitochondria (supplementary table S1 and fig. S3, Supplementary Material online). These in silico results are supported by functional and subcellular localization experiments of the two flavoHbs, FHb1 and FHb2, in the fungus *Aspergillus oryzae*, which demonstrate that whereas FHb1, a member of Clade 3, is localized in the cytosol and likely protects cells from exogenous NO, the Clade 4 FHb2 is targeted to the mitochondria, protecting cells from internal NO stress (Zhou et al. 2011).

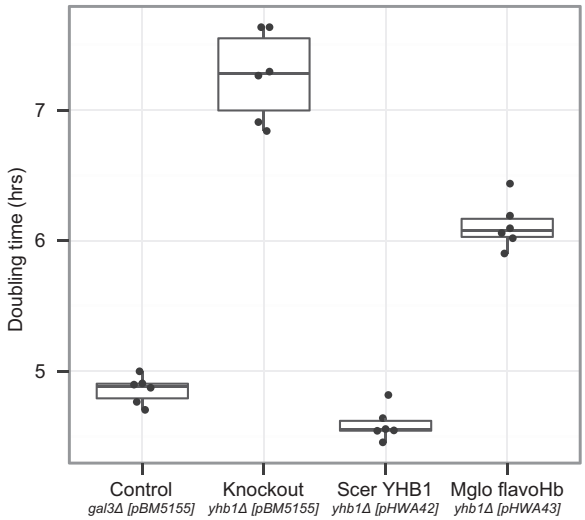

Fig. 2. Horizontally transferred flavoHb from *Malassezia globosa* confers NO resistance in a laboratory strain of *Saccharomyces cerevisiae*. All cultures were exposed to extracellular NO in the form of DPTA NONOate (see Materials and Methods). Boxplots indicate the doubling time (in h) of the different experimental cultures from left to right: control cultures in which a gene that is not required for NO metabolism was knocked out and a blank plasmid was inserted (*gal3Δ [pBM5155]*), cultures in which the gene encoding the native flavoHb was knocked out and a blank plasmid was inserted (*yhb1Δ [pBM5155]*), cultures in which the *yhb1Δ* mutation was complemented with a plasmid expressing Yhb1 from *S. cerevisiae* (*yhb1Δ [pHWA42]*), and *yhb1Δ* cultures complemented with a plasmid expressing the flavoHb from *M. globosa* (*yhb1Δ [pHWA43]*). The doubling times of all four strains were statistically significantly different from each other (Wilcox rank sum tests, Bonferroni adjusted *P* value < 0.05; supplementary table S5, Supplementary Material online). All data are plotted as jittered points. The middle of each boxplot represents the median doubling time for each strain, and the bottom and top of each box represent the first and third quartiles (the 25th and 75th percentiles), respectively.

Recurrent HGT of FlavoHb

The average number of HGT events predicted by the gene tree-species phylogeny reconciliation varied from 1, when the HGT cost was high (11–15 times the cost of GD), to 26.9, when the cost of HGT was low (five times the cost GD) (table 2). However, these numbers are likely underestimates of the actual number HGT events that have occurred during the course of flavoHb evolution. This is due to the fact that the reconciliation analysis was by necessity limited to species present in the reduced flavoHb phylogeny (i.e., organisms without flavoHbs were ignored), which grossly underestimated the number of GLs and thereby overestimated GD and underestimated HGT.

In agreement with the AU topological constraint test rejecting the hypothesis that eukaryotic flavoHbs were monophyletic, the reconciliation analysis supports the hypothesis that several eukaryotic flavoHbs, including those in Amoebozoa Clade 1 and in Fungi Clades 1 and 5, were independently acquired via HGT from bacteria (supplementary table S4, Supplementary Material online). Other reconciliation-predicted HGT events included inter- and intradivision HGT between bacteria (supplementary table S4, Supplementary Material online). However, ambiguities stemming from polytomies in the species phylogeny limited our ability to infer donor and recipient lineages of HGT events in some cases (supplementary table S4, Supplementary Material online). Here, we further examined three of these candidate HGT events that were all readily apparent from the flavoHb gene tree and had clear HGT donor and recipients from distantly related lineages (fig. 1).

The first case concerns the grouping of the flavoHb from the firmicute *Virgibacillus halodenitrificans* with Proteobacteria, including *Halomonas anticariensis* (supplementary table S4, Supplementary Material online). Surprisingly, close examination of the *V. halodenitrificans* genome using an alien index analysis (Gladyshev et al. 2008) flagged four entire contigs in the assembly as being of foreign origin, including the contig containing the flavoHb gene (supplementary fig. S4, Supplementary Material online). Rather than interpreting this signal as biological, we offer it as a cautionary tale of how genome misassembly artifacts stemming from the presence of more than one organism's genomic DNA can be mistakenly interpreted as HGT. In contrast, the other examples of HGT discussed below show the transferred flavoHb gene nested within assembly contigs that contain predominantly vertically inherited genes (supplementary fig. S4, Supplementary Material online).

The second case of HGT concerns the horizontal acquisition of the flavoHb in *Chondromyces apiculatus* (Proteobacteria) from *Bacillus* (Firmicutes) (supplementary table S4, Supplementary Material online). In addition to the reconciliation-based evidence of HGT, topological constraint tests significantly rejected the monophyletic origins for both proteobacterial and firmicute flavoHbs (table 1), providing additional support for this HGT event.

The third and most striking case of HGT involved the horizontal acquisition of the flavoHb in two fungal pathogens of the human skin (*M. globosa* and *M. sympodialis*) from Actinobacteria, and was present in all reconciled tree topologies regardless of transfer cost (table 2). The specific donor in this HGT event was likely a species of *Corynebacterium* (supplementary table S4, Supplementary Material online). As both *Malassezia* and *Corynebacterium* are some of the most common members of the human skin microbiome (Findley and Grice 2014), their niche overlap may have provided the opportunity for HGT.

Horizontally Acquired FlavoHb Confers NO Resistance

To test the hypothesis that the transferred flavoHb of *Malassezia* is functional and mediates NO resistance, we evaluated the activity of the flavoHb protein from *M. globosa* in a *YBH1* knockout strain of *S. cerevisiae* (fig. 2). A *yhb1Δ* strain expressing the *M. globosa* flavoHb was less sensitive to exogenous NO, growing 16% faster than a blank vector control strain (Wilcox rank sum test Bonferroni adjusted *P* value = 0.013, supplementary table S5, Supplementary Material online). These data indicate that the flavoHb in *M. globosa* is indeed functional and able to mediate NO

resistance, although the resistance conferred by the *M. globosa* flavoHb in this experimental system is not as high as that of the native Yhb1 protein from *S. cerevisiae* (fig. 2).

Discussion

Here, we mined the genomes of thousands of organisms spanning the tree of life for diverse flavoHbs, constructed a comprehensive phylogeny, and retraced the evolutionary history and function of this large and complex protein family. Although the last common ancestor of flavoHb-containing organisms was likely the last universal common ancestor, somewhat surprisingly, the taxonomic diversity of flavoHb-containing organisms is relatively limited; flavoHbs are extremely common in some bacterial divisions but uncommon or absent in others. Similarly, in eukaryotes, flavoHbs are abundant in fungi, but rare elsewhere. Finally, flavoHb evolutionary history is complex and appears to have been extensively sculpted by hundreds of GD and GL events, which appear coupled with either functional specialization or the evolution of novel functions, as well as by several episodes of HGT (table 2).

Previous studies evaluating flavoHbs in the larger context of globin evolution and diversification, have proposed that the flavoHb gene family arose from an ancestral globin in bacteria and later spread to eukaryotes via HGT, either once during mitochondrial endosymbiosis or via multiple independent events into different eukaryotic lineages (Vinogradov et al. 2005, 2007, 2013). The evolutionary pattern of flavoHb phylogeny in figure 1 provides strong support for the latter hypothesis. Rather than form a monophyletic group, as would be suggestive of a single ancestral acquisition of flavoHbs, eukaryotic flavoHbs were distributed across ten divergent clades. Similarly, although topological uncertainty, particularly near the base of the flavoHb phylogeny makes it difficult to speculate on the possible HGT donors for all eukaryotic flavoHbs, several episodes of relatively recent HGT events from bacteria to eukaryotes can be inferred. Most notably, the flavoHbs that comprise Amoebozoa Clade 1 and Fungi Clade 1 were both acquired from Firmicutes. Likewise, the *Malassezia* flavoHbs that comprise Fungi Clade 5 were acquired from Actinobacteria. Importantly, functional experiments confirmed that the transferred flavoHb of *Malassezia* is functional and mediates NO resistance.

HGT between the domains of life, and in particular between bacteria and eukaryotes, is infrequent (Andersson 2009); recurrent HGT of the same gene family between distantly related organisms, such as that described in this study for flavoHbs, is likely rarer still. We are aware of only two other known cases. The first involves pore-forming aerolysins proteins that disrupt target (e.g., host or prey) cell membranes, which have undergone repeated HGT from bacteria to eukaryotes (Moran et al. 2012). The second involves antibacterial lysozymes, which appear to have been transferred multiple times from bacteria to diverse eukaryotes, as well as to extremophile archaea (Metcalf et al. 2014). Intriguingly, all three proteins, including flavoHb, are ecologically important proteins with functions related to interspecific competition, predation, and defense. A second characteristic shared between flavoHbs, aerolysins, and lysozymes is that they function as self-contained metabolic units rather than as part of a protein complex or larger metabolic pathway. We suggest that their functional independence is exactly what renders them especially prone to evolution via GD and HGT as well as what enables them to quickly assimilate (i.e., "plug-and-play") in new and different cellular environments (Jain et al. 1999; Cohen et al. 2011). This functional independence and propensity for GD and HGT exhibited by single-protein metabolic modules has likely favored, over evolutionary time, their frequent involvement in host-microbe arms races.

Materials and Methods

Detection of FlavoHbs

The flavoHb protein encoded by the *hmp* gene of *E. coli* (UniProt ID: P24232) was queried against NCBI's nonredundant protein database (nr) using phmmer, a member of the HMMER3 software suite (web server accessed May 8, 2015) (Finn et al. 2011). All proteins with the same Pfam domain architecture as *E. coli* Hmp, characterized by an N-terminal globin domain [ID:PF00042] and C-terminal FAD- and NAD-binding oxidoreductase domains [ID:PF00970 and ID:PF00175], were considered flavoHbs and downloaded for further analysis. To check for additional eukaryotic flavoHbs, as well as sequences homologous to flavoHbs (i.e., missing one or more Pfam domains) that may have been missed using a bacterial query, eukaryotic flavoHbs were queried against a local copy of the nr database (last updated November 24, 2014) using BLASTp, and full-length proteins corresponding to the top 100 hits to each query sequence were extracted and included in downstream analyses (E-value cutoff 0.001). BLAST-identified sequences were considered putative flavoHbs if they had significant hits to all three Pfam domains (using a hmmsearch E-value inclusion threshold of 0.001). Identical HMMER- and BLAST-identified flavoHb homologs were collapsed with CD-HIT version 4.5.8 (Fu et al. 2012) using a sequence identity threshold of 1.0 and a word length of 4. Identification of N-terminal targeting peptides was performed using the TARGETP 1.1 web server (Emanuelsson et al. 2007).

An earlier study of globin domain-containing proteins reported the presence of flavoHbs in marine algae (haptophyte *Emiliania huxleyi* and five species of heterokonts, e.g., *Phaeodactylum tricornutum*), choanoflagellates (*Monosiga brevicollis* and *Sphaeroforma arctica*), and the apusomonad *Thecamonas trahens* (Vinogradov et al. 2013). Unfortunately, that study did not include bacterial representatives of the different families of globin-containing proteins, rendering accurate classification challenging. Our HMMER search did not identify flavoHbs from any of these species (supplementary table S6, Supplementary Material online), nor did our BLAST search identify any of these sequences as being homologous to the functionally characterized flavoHbs in Fungi, Amoebozoa, and Fornicata (taking the top 100 BLAST hits with E-values $\leq$ 0.001; supplementary table S1, Supplementary Material online), suggesting that these globin

domain-containing proteins are not members of the flavoHb family (supplementary table S6, Supplementary Material online).

Phylogenetic Analyses

Sequences were aligned with MAFFT v7.023b using the E-INS-i strategy (Katoh and Standley 2013) and trimmed with TRIMAL v1.4.rev11 using its automated1 strategy (Capella-Gutierrez et al. 2009). Phylogenetic trees were constructed using FASTTREE (Price et al. 2010) initiated with 100 random starting trees using a Whelan and Goldman + GAMMA amino acid model of substitution, 1,000 resamples, four rounds of minimum-evolution subtree-prune- regraft moves (-spr 4), and the more exhaustive ML nearest-neighbor interchange option enabled (-mlacc 2-slownni). Tests of monophyly were performed in CONSEL version 1.2 (Shimodaira and Hasegawa 2001) using the AU test to compare the unconstrained best tree and the best tree given a constrained topology (Shimodaira 2004). To create constrained topologies, taxa of interest were forced to be monophyletic, and all other branches were resolved to obtain the maximum negative log likelihood using FASTTREE as described above.

Gene Tree-Species Tree Reconciliation

The species phylogeny of all flavoHb-containing organisms was extracted from the NCBI taxonomy database using phyloT (http://phylot.biobyte.de, last accessed April 19, 2016). However, 57.4% of internodes in this 2,567 taxon phylogeny were polytomous (multifurcating), making reconciliation intractable with current algorithms (Stolzer et al. 2012). Therefore, we pruned the species tree by allowing no more than eight descendants per node and no more than two terminal descendants (i.e., leaves) per node; descendant nodes were first ordered based on their subtree size, such that descendants with smaller subtrees (i.e., less leaves) were pruned first. This pruning step restricted our analysis to 693 organisms whose species phylogeny had far fewer polytomies (29.5% internal nodes were multifurcating). The ML gene tree for the 835 flavoHbs in this reduced taxa set was created using the same alignment, trimming, and tree-building software as described above. An additional ML gene tree for the 175 flavoHbs in Fungal Clade 3 was created using to same methods and used to evaluate support for intrafungal HGT (supplementary table S3, Supplementary Material online). All trees and alignments are available from the figshare repository (https://dx.doi.org/10.6084/m9.figshare.1591868.v1, last accessed April 19, 2016).

Gene tree-species phylogeny reconciliation was performed in NOTUNG using its duplication, transfer, loss, and incomplete lineage sorting aware parsimony-based algorithm (Chen et al. 2000; Stolzer et al. 2012). Ambiguity in the species phylogeny and low branch support in the reduced flavoHb gene tree were handled through a multistep approach (supplementary fig. S5, Supplementary Material online). First, flavoHb tree branches with less than 0.80 SH-like local support were collapsed and resolved against a bifurcating species tree, which was created using the multi2di function in the ape R package (Paradis et al. 2004). The resolved flavoHb tree was then rooted and reconciled with the multifurcating, consensus species phylogeny. All optimal (i.e., equally parsimonious) solutions from the resolve, root, and reconcile steps were retained. GD and GL costs were set to one, and transfer costs of 5, 7, 9, 11, 13, and 15 were all evaluated.

Alien Index Calculations

For putative HGT recipients, the genome containing the suspected HGT gene was screened for contamination using a modified alien index (AI) approach (Gladyshev et al. 2008). For each genome, two taxonomic lineages were first specified: the RECIPIENT into which possible HGT events may have occurred (e.g., *Malassezia*, NCBI taxonomy ID 55193), and a larger GROUP of related taxa (e.g., Fungi, NCBI taxonomy ID 4751). All predicted proteins in the genome assembly were queried against the nr database using uBLAST using a dbaccel parameter of 80, accel parameter of 1, and an E-value cutoff of 0.001 (http://drive5.com/usearch, last accessed April 19, 2016) (Edgar 2010). The AI score is given by the formula: $AI = \left(\frac{bbhO}{maxB} - \frac{bbhG}{maxB}\right)$, where *bbhO* is the bitscore of the best BLAST hit to a species outside of the GROUP lineage, *bbhG* is the bitscore of the best BLAST hit to a species within the GROUP lineage (all hits to the RECIPIENT lineage were skipped), and *maxB* is the maximum bitscore possible for the query sequence (i.e., the bitscore of the query aligned to itself). *AI* can range from 1 to −1. *AI* is greater than zero if the gene has a better BLAST hit to a species outside of the GROUP lineage and is suggestive of either HGT or contamination. Contamination was inferred in cases when all or the majority of genes on a contig had an $AI \geq 0$; in contrast, HGT was inferred when the transferred gene was located within a well assembled contig, and the majority of genes on that contig had an $AI < 0$ (supplementary fig. S4, Supplementary Material online). The perl script for calculating AI scores is freely available from the figshare repository, http://dx.doi.org/10.6084/m9.figshare.1593040, last accessed April 19, 2016.

Functional Test for NO Resistance

To construct the *M. globosa flavoHb* pHWA43 and *S. cerevisiae YHB1* pHWA42 test plasmids, the pBM5155 plasmid (Alexander et al. 2016) was linearized with NotI (New England Biolabs, Ipswich, MA). The *S. cerevisiae YHB1* sequence was amplified from FM1282 (Hittinger and Carroll 2007), whereas the *M. globosa flavoHb* sequence was codon-optimized and ordered as a gBlock (IDT, Coralville, IA). Amplification of both was performed using primers that added approximately 40-bp regions of sequence to both ends of the sequence identical to a sequence on either side of the NotI restriction site of pBM5155. Gap repair cloning was performed by transforming the *S. cerevisiae* strain M22 (Mortimer 2000) with linearized plasmid and each of the polymerase chain reaction products. The transformations were plated to YPD +Nat plates (1% yeast extract, 2% peptone, 2% glucose, 1.8% agar, 100 μg/ml nourseothricin) after 3 h of recovery in YPD liquid (1% yeast extract, 2% peptone, 2% glucose). gDNA was extracted from the resulting colonies and used to electroporate *E. coli*. Bacterial colonies were screened for an insert of an

appropriate size, and the insert verified by was Sanger-sequencing. Primers used in construction and analysis are listed in supplementary table S7, Supplementary Material online. The yHWA396 (*gal3*Δ) and yHWA400 (*yhb1*Δ) strains of *S. cerevisiae* were obtained from the Yeast Genome Deletion Project (Giaever et al. 2002). Strains were transformed by the lithium acetate/polyethylene glycol method (Gietz and Woods 2002), and all strains used are listed in supplementary table S8, Supplementary Material online.

Each strain was struck to YPD + Nat agar and grown for 2 days at 30 °C. Three individual colonies for each strain were picked to YPD +Nat liquid with 10 μg/ml doxycycline added and incubated at 30 °C for 12 h. The strains were then used to inoculate two test tubes per strain of 3 ml of NO tester YPD (1% yeast extract, 2% peptone, 2% glucose, 0.001% uracil, 0.001% histidine, 0.001% leucine, 0.001% methionine, 0.005% adenine hemisulfate, 5 μM $CuCl_2$, 5 μM $(NH_4)Fe(SO_4)$, 100 mM potassium phosphate, 100 μg/ml ampicillin, 100 μg/ml nourseothricin, 10 μg/ml doxycycline, pH 7.4) with 1 mM dipropylenetriamine (DPTA) NONOate (Cayman Chemical, Ann Arbor, MI) in technical duplicates; a third tube of NO tester YPD without DPTA NONOate was also inoculated. Tubes were inoculated to an initial OD_{600} of 0.25–0.35, and they were incubated on a culture wheel for 12 h at 30 °C. After 12 h, the OD_{600} of each tube was read, and the doubling time for each replicate was calculated.

Supplementary Material

Supplementary figures S1–S5 and tables S1–S8 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank members of the Rokas lab for helpful discussions. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This material is based upon work supported by the National Science Foundation (http://www.nsf.gov) under Grants IOS-1401682 to J.H.W., DEB-1442148 to C.T.H., and DEB-1442113 to A.R.; in part by the DOE Great Lakes Bioenergy Research Center (DOE Office of Science BER DE-FC02-07ER64494 to C.T.H.), and the USDA National Institute of Food and Agriculture (Hatch project 1003258 to C.T.H.). C.T.H. is a Pew Scholar in the Biomedical Sciences and an Alfred Toepfer Faculty Fellow, supported by the Pew Charitable Trusts and the Alexander von Humboldt Foundation, respectively.

References

Alexander WG, Peris D, Pfannenstiel BT, Opulente DA, Kuang M, Hittinger CT. 2016. Efficient engineering of marker-free synthetic allotetraploids of *Saccharomyces*. *Fungal Genet Biol*. 89:10–17.

Andersson JO. 2009. Gene transfer and diversification of microbial eukaryotes. *Annu Rev Microbiol*. 63:177–193.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

Cassanova N, O'Brien KM, Stahl BT, McClure T, Poyton RO. 2005. Yeast flavohemoglobin, a nitric oxide oxidoreductase, is located in both the cytosol and the mitochondrial matrix: effects of respiration, anoxia, and the mitochondrial genome on its intracellular level and distribution. *J Biol Chem*. 280:7645–7653.

Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol*. 7:429–447.

Cohen O, Gophna U, Pupko T. 2011. The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol*. 28:1481–1489.

Cramm R, Siddiqui RA, Friedrich B. 1994. Primary sequence and evidence for a physiological function of the flavohemoprotein of *Alcaligenes eutrophus*. *J Biol Chem*. 269:7349–7354.

Crawford MJ, Goldberg DE. 1998. Role for the *Salmonella* flavohemoglobin in protection from nitric oxide. *J Biol Chem*. 273:12543–12547.

Danchin EGJ, Rosso M-N, Vieira P, de Almeida-Engler J, Coutinho PM, Henrissat B, Abad P. 2010. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A*. 107:17651–17656.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.

Emanuelsson O, Brunak S, Heijne von G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2:953–971.

Findley K, Grice EA. 2014. The skin microbiome: a focus on pathogens and their association with skin disease. *PLoS Pathog*. 10:e1004436.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 39:W29–W37.

Forrester MT, Foster MW. 2012. Protection from nitrosative stress: a central role for microbial flavohemoglobin. *Free Radic Biol Med*. 52:1620–1633.

Frey AD, Kallio PT. 2003. Bacterial hemoglobins and flavohemoglobins: versatile proteins and their impact on microbiology and biotechnology. *FEMS Microbiol Rev*. 27:525–545.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–391.

Gietz RD, Woods RA. 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Method Enzymol*. 350:87–96.

Gladyshev EA, Meselson M, Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.

Gogarten JP, Townsend JP. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol*. 3:679–687.

Gupta S, Pawaria S, Lu C, Hade MD, Singh C, Yeh S-R, Dikshit KL. 2012. An unconventional hexacoordinated flavohemoglobin from *Mycobacterium tuberculosis*. *J Biol Chem*. 287:16435–16446.

Gupta S, Pawaria S, Lu C, Yeh S-R, Dikshit KL. 2011. Novel flavohemoglobins of mycobacteria. *IUBMB Life* 63:337–345.

Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–6U1.

Hoogewijs D, Dewilde S, Vierstraete A, Moens L, Vinogradov SN. 2012. A phylogenetic analysis of the globins in fungi. *PLoS One* 7:e31856.

Hu Y, Butcher PD, Mangan JA, Rajandream M-A, Coates AR. 1999. Regulation of *hmp* gene transcription in *Mycobacterium tuberculosis*: effects of oxygen limitation and nitrosative and oxidative stress. *J Bacteriol*. 181:3486–3493.

Iijima M, Shimizu H, Tanaka Y, Urushihara H. 2000. Identification and characterization of two flavohemoglobin genes in *Dictyostelium discoideum*. *Cell Struct Funct*. 25:47–55.

Jain R, Rivera MC, Lake JA. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A*. 96:3801–3806.

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 42:D199–D205.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res*. 39:W475–W478.
Li L, Huang Y, Xia X, Sun Z. 2006. Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol*. 23:2467–2473.
Mastronicola D, Testa F, Forte E, Bordi E, Pucillo LP, Sarti P, Giuffrè A. 2010. Flavohemoglobin and nitric oxide detoxification in the human protozoan parasite *Giardia intestinalis*. *Biochem Biophys Res Commun*. 399:654–658.
Metcalf JA, Funkhouser-Jones LJ, Brileya K, Reysenbach A-L, Bordenstein SR. 2014. Antibacterial gene transfer across the tree of life. *Elife* 3:e04266.
Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U. 2012. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. *Mol Biol Evol*. 29:2223–2230.
Mortimer RK. 2000. Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Res*. 10:403–409.
Muto A, Kotera M, Tokimatsu T, Nakagawa Z, Goto S, Kanehisa M. 2013. Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *J Chem Inf Model*. 53:613–622.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
Prachumwat A, Li W-H. 2006. Protein function, connectivity, and duplicability in yeast. *Mol Biol Evol*. 23:30–39.
Price MN, Dehal PS, Arkin AP. 2010. Fasttree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490.
Rafferty S, Luu B, March RE, Yee J. 2010. *Giardia lamblia* encodes a functional flavohemoglobin. *Biochem Biophys Res Commun*. 399:347–351.
Runckel C, DeRisi J, Flenniken ML. 2014. A draft genome of the honey bee trypanosomatid parasite *Crithidia mellificae*. *PLoS One* 9:e95057.
Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. *Syst Biol*. 51:492–508.
Shimodaira H. 2004. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann Stat*. 32:2616–2641.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
Stevanin TM, Poole RK, Demoncheaux EAG, Read RC. 2002. Flavohemoglobin Hmp protects *Salmonella enterica* serovar Typhimurium from nitric oxide-related killing by human macrophages. *Infect Immun*. 70:4399–4405.
Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D. 2012. Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28:1409–1415.
Takaya N, Suzuki S, Matsuo M, Shoun H. 1997. Purification and characterization of a flavohemoglobin from the denitrifying fungus *Fusarium oxysporum*. *FEBS Lett*. 414:545–548.
Thakur N, Gupta S, Hade MD, Dikshit KL. 2014. Type I flavohemoglobin of *Mycobacterium smegmatis* is a functional nitric oxide dioxygenase. *IUBMB Life* 66:396–404.
Ullmann BD, Myers H, Chiranand W, Lazzell AL, Zhao Q, Vega LA, Lopez-Ribot JL, Gardner PR, Gustin MC. 2004. Inducible defense mechanism against nitric oxide in *Candida albicans*. *Eukaryot Cell*. 3:715–723.
Vinogradov SN, Bailly X, Smith DR, Tinajero-Trejo M, Poole RK, Hoogewijs D. 2013. Microbial eukaryote globins. *Adv Microb Physiol*. 63:391–446.
Vinogradov SN, Hoogewijs D, Bailly X, Arredondo-Peter R, Guertin M, Gough J, Dewilde S, Moens L, Vanfleteren JR. 2005. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc Natl Acad Sci U S A*. 102:11385–11389.
Vinogradov SN, Hoogewijs D, Bailly X, Mizuguchi K, Dewilde S, Moens L, Vanfleteren JR. 2007. A model of globin evolution. *Gene* 398:132–142.
Wapinski I, Pfeffer A, Friedman N, Regev A. 2007. Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449:54–61.
Zhou S, Fushinobu S, Kim S-W, Nakanishi Y, Maruyama J-I, Kitamoto K, Wakagi T, Shoun H. 2011. Functional analysis and subcellular location of two flavohemoglobins from *Aspergillus oryzae*. *Fungal Genet Biol*. 48:200–207.
Zhu H, Riggs AF. 1992. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proc Natl Acad Sci U S A*. 89:5015–5019.