Visual Neuroscience (2016), **33**, e002, 13 pages.
Copyright © Cambridge University Press, 2016 0952-5238/16
doi:10.1017/S0952523815000322

Comparative sequence analyses of rhodopsin and RPE65 reveal patterns of selective constraint across hereditary retinal disease mutations

FRANCES E. HAUSER,[1] RYAN K. SCHOTT,[1] GIANNI M. CASTIGLIONE,[2]
ALEXANDER VAN NYNATTEN,[2] ALEXANDER KOSYAKOV,[1,2] PORTIA L. TANG,[1,2]
DANIEL A. GOW,[1,2] AND BELINDA S.W. CHANG[1,2,3]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario, Canada
[2]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada
[3]Centre for the Analysis of Genome Evolution and Function, University of Toronto, Toronto, Ontario, Canada

(RECEIVED September 1, 2015; ACCEPTED October 23, 2015)

Abstract

Retinitis pigmentosa (RP) comprises several heritable diseases that involve photoreceptor, and ultimately retinal, degeneration. Currently, mutations in over 50 genes have known links to RP. Despite advances in clinical characterization, molecular characterization of RP remains challenging due to the heterogeneous nature of causal genes, mutations, and clinical phenotypes. In this study, we compiled large datasets of two important visual genes associated with RP: rhodopsin, which initiates the phototransduction cascade, and the retinoid isomerase RPE65, which regenerates the visual cycle. We used a comparative evolutionary approach to investigate the relationship between interspecific sequence variation and pathogenic mutations that lead to degenerative retinal disease. Using codon-based likelihood methods, we estimated evolutionary rates (d_N/d_S) across both genes in a phylogenetic context to investigate differences between pathogenic and nonpathogenic amino acid sites. In both genes, disease-associated sites showed significantly lower evolutionary rates compared to nondisease sites, and were more likely to occur in functionally critical areas of the proteins. The nature of the dataset (e.g., vertebrate or mammalian sequences), as well as selection of pathogenic sites, affected the differences observed between pathogenic and nonpathogenic sites. Our results illustrate that these methods can serve as an intermediate step in understanding protein structure and function in a clinical context, particularly in predicting the relative pathogenicity (i.e., functional impact) of point mutations and their downstream phenotypic effects. Extensions of this approach may also contribute to current methods for predicting the deleterious effects of candidate mutations and to the identification of protein regions under strong constraint where we expect pathogenic mutations to occur.

Keywords: Molecular evolution, Retinitis pigmentosa, Molecular phylogenetics, Rhodopsin, RPE65, Phylomedicine, Likelihood-based codon models of evolution

Introduction

Retinitis pigmentosa (RP) encompasses several heritable diseases that involve the degeneration of photoreceptor cells. Individuals with RP exhibit symptoms such as night blindness, followed by decreasing visual fields and ultimately progressive visual impairment that can result in legal or complete blindness (Hartong et al., 2006). On the basis of inheritance, nonsyndromic RP can be divided into autosomal dominant, recessive, and X-linked forms. Multiple factors contribute to the heterogeneous nature of RP (reviewed in Rivolta et al., 2002; Anasagasti et al., 2012). Clinical RP phenotypes vary considerably with respect to severity, age of onset, and causal genes, and are thought to reflect the effects of particular mutations of RP-associated proteins (Mendes et al., 2005; Hartong et al., 2006; Anasagasti et al., 2012). Currently, mutations in over 50 genes have known links to RP, and recent advances in high-throughput sequencing are frequently uncovering novel causative genes and mutations (e.g., Bowne et al., 2011; Nishiguchi et al., 2013; Yang et al., 2014). Despite these advances in diagnosis and clinical characterization, molecular characterization of RP remains challenging due to the diversity of causal genes and their numerous disease-associated mutations. These disease genes are involved in a variety of processes both within and outside of the visual cycle, such as vitamin A metabolism, cell–cell interactions, cell structure, and signaling (reviewed in Hartong et al., 2006).

One of the most prominent and best characterized RP disease genes investigated to date is rhodopsin (RHO), the visual pigment found in rod photoreceptor cells that initiates the visual transduction cascade. Mutations in RHO account for roughly 30–40% of dominant RP cases (Ferrari et al., 2011). Of the variety of RP-linked genes studied to date, RHO is particularly amenable to studies

Address correspondence to: Dr. Belinda Chang, Department of Cell & Systems Biology, Department of Ecology & Evolutionary Biology, University of Toronto, 25 Harbord St, Toronto, Ontario M5S 3G5, Canada. E-mail: belinda.chang@utoronto.ca

exploring the link between variation in amino acid residues and protein function. First, it has been established that disease-causing mutations in RHO often interfere with proper protein folding or transportation, resulting in the aggregation of defective proteins and subsequent photoreceptor cell death (Breikers et al., 2002; Mendes et al., 2005). Second, the crystal structure of RHO has been resolved for both the inactive and active conformations (Palczewski et al., 2000; Choe et al., 2011). Third, numerous mutant RHO proteins (comprising both naturally occurring and pathogenic mutants) have been characterized with respect to structure and function (e.g., Janz et al., 2003; Piechnick et al., 2012; McKeone et al., 2014; Morrow & Chang, 2015). Due to its importance in visual transduction, natural variation in RHO sequence, structure, and function has been extensively investigated across a diversity of animals (Zhao et al., 2009; Sugawara et al., 2010; Porter et al., 2011; Schott et al., 2014).

Outside of the visual transduction cascade, the visual cycle gene RPE65 has also been experimentally investigated with respect to clinically relevant mutations: over 60 point mutations have been linked to both RP and an early-onset form of retinal degeneration, Leber's congenital amaurosis. RPE65 is a large membrane-bound retinoid isomerase located in the retinal pigment epithelium and is responsible for the conversion of activated all-*trans*-retinal to 11-*cis*-retinal. While less extensively studied relative to RHO, recent structural and clinical studies have begun to elucidate the effect of point mutations on RPE65 function in greater detail (Moiseyev et al., 2005; Redmond et al., 2005; Philpa et al., 2009; Cideciyan, 2010; Li et al., 2014; Takahashi et al., 2014; Li et al., 2015), and a crystal structure has also been resolved (Kiser & Palczewski, 2010).

Despite numerous studies examining RP mutations in RHO and other causal RP genes in human populations, molecular characterization of this disease remains challenging, and accurately validating disease mutations is not always straightforward, particularly in genes for which there is little functional information. Comparative analyses combined with studies of protein structure and function can lend insight into the molecular determinants of disease and help to inform which mutations are more likely to result in a disease phenotype. Intuitively, mutations at highly conserved amino acid sites are more likely to be deleterious, often resulting in the disease phenotypes observed in RP (e.g., Briscoe et al., 2004; Iannaccone et al., 2006). However, despite the high conservation of RHO and RPE65 and their critical role in phototransduction and visual regeneration, there is abundant natural sequence variation in both genes across vertebrate groups. This variation presents an opportunity to conduct an evolutionarily informed (phylomedicine) study of retinal protein function in the context of hereditary disease.

The genetic complexity of RP combined with recent advances in sequencing technology means that accurate and reliable molecular diagnosis of disease-causing RP mutations is particularly important (reviewed in Anasagasti et al., 2012). Through the analysis of natural sequence variation, we can make inferences about the evolutionary constraints operating on amino acid residues critical for proper protein function. In addition to identifying the location of conserved residues, codon-based methods of sequence analysis can estimate the degree of selective constraint imposed on a given site, as indicated by the evolutionary rate ratio of nonsynonymous to synonymous amino acid substitutions, or d_N/d_S. This approach assumes that functionally important residues will be subject to greater selective constraint, and therefore lower d_N/d_S, and that mutations at these sites are more likely to be deleterious and potentially disease-causing. Numerous studies have harnessed interspecific sequence analysis to examine known disease mutations, encompassing a variety of diseases and proteins (Greenblatt et al., 2003; Gaucher et al., 2006; Stover & Verrelli, 2010; Rishishwar et al., 2012; Kirwan et al., 2013). Beyond an explicitly disease-focused context, these comparative methods have also addressed areas of medical relevance, such as longevity (Morgan et al., 2013) and immune response (Webb et al., 2015), using estimates of positive selection to interpret and identify potential changes in protein function. In addition to known disease sites, cross-species data may aid predictions of other conserved amino acid residues critical for proper protein function that would likely generate a defective product if mutated (Miller & Kumar, 2001; Mooney & Klein, 2002; Greenblatt et al., 2003). To take advantage of the diversity of sequences available, an appropriate phylogenetic scope for coding sequence sampling must be established. For analyses of human disease genes, although it may be desirable to use only mammalian sequences, if these sequences have high similarity there may be insufficient variation to detect conserved *versus* nonconserved sites (but see Cooper et al., 2003).

In this study, we calculated the evolutionary rate ratio (d_N/d_S) in vertebrate and mammalian datasets of the visual transduction gene RHO and the visual (retinoid) cycle gene RPE65, and compared the evolutionary conservation of codon sites harboring disease-associated (pathogenic) mutations with sites that have not been implicated in disease. Our aim was to assess the extent to which codon-based evolutionary analyses can inform studies of human diseases caused by defective proteins. Metrics of evolutionary conservation using orthologous sequences are implemented in several disease prediction programs, such as PolyPhen (Adzhubei et al., 2010) and SIFT (Ng & Henikoff, 2003), but the appropriate method of homologous sequence analysis varies depending on the nature of the alignment, disease, and protein in question (Flanagan et al., 2010; Hicks et al., 2011). Incorporating measures of d_N/d_S as estimated by the methods outlined in this study may provide a useful complement to programs designed to specifically identify disease-causing protein defects. Using large sequence datasets, we investigate how varying the level of taxonomic divergence, the selection of pathogenic sites, and the molecular evolutionary method can impact analyses of human visual disease genes. Overall, pathogenic sites had significantly lower d_N/d_S values in comparison to nonpathogenic sites. Many of the RHO mutations known to cause RP have been examined through mutagenesis experiments; however, new disease-associated RHO mutations are frequently discovered (Hollingsworth & Gross, 2013; Liu et al., 2013; Opefi et al., 2013; Pierrottet et al., 2014; Yang et al., 2014). In genes such as RPE65, for which there are fewer studies of structure and function, computational analyses of this nature may provide a crucial first step in understanding novel mutations in a clinical context.

Materials and methods

Dataset acquisition

RHO and RPE65 genes were selected for comparative analyses of coding sequence evolution in vertebrates. These genes were selected based on an abundance of nonsyndromic disease-associated missense or nonsense mutations in humans, expression in the retina, association with hereditary retinal degeneration, and a large number of available sequences for robust molecular evolutionary analyses. Since there are fewer RPE65 sequences available compared to RHO, an exhaustive search was performed to identify and acquire all vertebrate RPE65 sequences available on GenBank.

To maintain similar taxon sampling and ensure that the datasets were comparable, we acquired RHO sequences from these same species where possible. We also created a second RHO dataset that contained only mammals but with a similar number of sequences as the vertebrate datasets. As we already included all available RPE65 sequences, our mammalian dataset for RPE65 is simply the subset containing only mammals. Species list and accession numbers for all sequences used in this study are provided in Supplementary Table 1.

Phylogenetic analyses

Sequences from each of the four datasets were aligned separately using PRANK codon alignment, which has been found to reduce coding sequence alignment errors for evolutionary analysis by incorporating phylogenetic information into gap placement (Löytynoja & Goldman, 2005; Löytynoja & Goldman, 2008). Gene trees were estimated in MrBayes 3 (Ronquist & Huelsenbeck, 2003) and by maximum likelihood (ML) using PhyML 3 (Guindon et al., 2010) (Supplemental Figs. 1–8). The Bayesian analyses were run for five million generations with a 25% burn-in. Convergence was determined by verifying that the standard deviations of split frequencies approached zero and that there was no obvious trend in the log likelihood plot. The ML analyses were run under the GTR+G+I model with a BioNJ starting tree, the best of NNI (nearest neighbour interchange) and SPR (subtree pruning and regrafting) tree improvement and aLRT SH-like branch support (approximate likelihood ratio test with Shimodaira-Hasegawa-like branch support) (Anisimova & Gascuel, 2006).

Molecular evolutionary analyses

To estimate the strength of selection acting across RHO and RPE65, as well as the strength of selection acting on individual amino acid sites, the alignment and gene trees were analyzed in the HYPHY software package (Kosakovsky Pond et al., 2005), as implemented on the Datamonkey webserver (Delport et al., 2010) and in the codeml package of PAML 4 (Yang, 2007). Specifically, we used the PAML random sites models (M0, M1a, M2a, M3, M7, M8), and the FEL and FUBAR methods in HYPHY (Scheffler et al., 2006; Yang, 2007; Murrell et al., 2013). The PAML random sites models estimate d_N/d_S as a single value for a prespecified number of site classes. The posterior probability distribution of the assignment of individual sites to these site classes is then calculated using a Bayes empirical Bayes analysis and d_N/d_S values are calculated as weighted averages from this distribution. For comparison to the HYPHY models, we chose M8, as this was most often the best-fitting model and was most comparable in terms of the numbers of site classes. For M8, site categories are calculated from a β distribution ranging from 0 to 1, discretized into 10 categories, with an additional site category with $d_N/d_S \geq 1$. Unlike the PAML models, FEL and FUBAR estimate d_N and d_S separately. FEL estimates d_N and d_S independently for each site, while FUBAR estimates d_N and d_S using alignment-wide information and 400 rate categories. All analyses were carried out using both the Bayesian and ML trees.

Analyses of disease-associated mutations

Disease-associated mutations were identified for both RHO and RPE65. For the purpose of this study, we defined pathogenic mutations as being any mutation reported in the Human Gene Mutation Database (Stenson et al., 2013), dbSNP (Sherry et al., 2001), UniProt (Magrane & UniProt Consortium, 2011), or Online Mendelian Inheritance in Man (Hamosh, 2005) disease databases (referred to as "reported" pathogenic mutations in our RHO dataset). Sites linked to congenital stationary night blindness were included in our list of RHO pathogenic sites, as well as sites linked to Leber congenital amaurosis in RPE65. As substantial verification of disease-associated mutations has been performed for RHO, we also created a second subset of disease sites (hereafter, referred to as "confirmed pathogenic sites") that included only those sites with either experimental validation or that were clinically verified with familial cosegregation studies (Supplemental Table 2). We used these criteria to categorize sites as either pathogenic or nonpathogenic (and for RHO, confirmed pathogenic and nonpathogenic) and compared d_N/d_S values between these two groups. Because the HYPHY models estimate d_N and d_S separately, we were also able to compare a second metric, $d_N - d_S$, as well as d_N and d_S separately. To identify statistically significant differences between estimates at pathogenic *versus* nonpathogenic sites, we used the nonparametric Mann–Whitney *U* test, due to substitution rates being highly nonnormal. Comparisons were made using the M8 (PAML), FUBAR (HYPHY), and FEL (HYPHY) models for the vertebrate and mammalian datasets of RHO and RPE65 using both the Bayesian and ML tree topologies. Values of d_N/d_S for RHO and RPE65 from vertebrate FUBAR analyses were scaled and converted into color gradient space (where white represents the highest d_N/d_S, and dark blue represents the lowest d_N/d_S) and mapped onto the 3D crystal structures of RHO (IU19; Okada et al., 2004) and RPE65 (3FSN; Kiser et al., 2009) using Chimera (Pettersen et al., 2004). This created a "heat map" of conservation values (d_N/d_S) that were visualized in the context of the 3D structure. To infer how structural features correspond to conservation and pathogenicity, pathogenic sites in RHO and RPE65 were highlighted in both structures.

Results

Alignments and trees

The alignment length and number of sequences for the vertebrate and mammalian datasets of RHO and RPE65 are summarized in Table 1, as well as the number of disease-associated sites

Table 1. *Summary of RHO and RPE65 datasets*

Gene	Taxonomic scale	Number of sequences	Alignment length (codons)	Number of pathogenic sites	Overall d_N/d_S (PAML M0)
RHO	Vertebrate	102	355	53 (confirmed)/84 (reported)	0.055
	Mammal	110	353	53 (confirmed)/84 (reported)	0.041
RPE65	Vertebrate	115	533	66	0.065
	Mammal	66	533	66	0.055

Table 2. *Results of random sites models (PAML) for RHO performed on the ML tree*

						Parameters					
Tree	Model	np	ln *L*	κ	ω_0/p	ω_1/q	ω_2/ω_p	Null	LRT	df	*p*
Vertebrate	M0	203	−25219.9	2.53		0.055		N/A			
	M1a	204	−24968.3	2.68	0.04 (93.3%)	1 (6.7%)		M0	508.0	1	0.00
	M2a	206	−24968.3	2.68	0.04 (93.3%)	1 (6.7%)	5.7 (0%)	M1a	0.0	2	1.00
	M3	207	−24323.5	2.54	0.01 (62.7%)	0.09 (26.8%)	0.31 (10.5%)	M0	1795.4	4	0.00
	M7	204	−24305.8	2.54	0.267	3.312		N/A			
	M8a	205	−24305.8	2.54	0.280	3.98	1 (0.5%)	N/A			
	M8	206	−24305.8	2.54	0.280	3.98	1 (0.5%)	M7	11.0	2	0.004
								M8a	0.0	1	1.00
Mammalian	M0	219	−18018.3	4.21		0.041		N/A			
	M1a	220	−17758.5	4.39	0.03 (95.8%)	1 (4.2%)		M0	519.7	1	0.00
	M2a	222	−17758.5	4.39	0.03 (95.8%)	1 (3.1%)	1 (1.1%)	M1a	0.0	2	1.00
	M3	223	−17319.6	4.31	0.01 (76.7%)	0.13 (19.3%)	0.46 (3.9%)	M0	1397.5	4	0.00
	M7	220	−17315.2	4.31	0.163	2.05		N/A			
	M8	222	−17315.2	4.31	0.163	2.05	1 (0%)	M7	0.0	2	1.00

Note: np, number of parameters; ln *L*, ln likelihood; κ, transition/transversion ratio; df, degrees of freedom. For models M0–M3, the ω values for each site class (ω_0–ω_2) are shown with their proportions in parentheses. For models M7–M8, *p* and *q* describe the shape of the beta distribution, and ω_p refers to the positively selected site class (with proportion in parentheses) for models M8 and M8a (where it is constrained to one).

analyzed in each gene. Results from the phylogenetic analyses of the vertebrate and mammalian RHO and RPE65 alignments were generally congruent with species relationships (Meredith et al., 2011; Crottini et al., 2012; Fong et al., 2012). Bayesian and ML phylogenies for RHO and RPE65 were similar, and both phylogenies were used in subsequent molecular evolutionary analyses. The results of the analyses were consistent regardless of the phylogenetic reconstruction method used (Supplemental Tables 3–9).

Molecular evolutionary analyses

To assess selective pressure across vertebrate and mammalian datasets of RHO and RPE65, we used a variety of molecular evolutionary approaches: the random sites models in the PAML software package (Yang, 2007), and FEL and FUBAR models in the HYPHY software package (Kosakovsky Pond et al., 2005). The random sites PAML models confirmed that there is variation in d_N/d_S in the vertebrate and mammalian datasets of both RHO and RPE65 (the ratio of nonsynonymous to synonymous substitution rate; M3 *vs.* M0, $P < 0.001$; Tables 2 and 3), as expected for protein-coding genes. These models did not find significant evidence for a positively selected class of sites (M2a *vs.* M1a, M7 *vs.* M8, M8 *vs.* M8a where applicable; Tables 2 and 3) for either gene, reflecting high evolutionary conservation across both vertebrates and mammals. No positively selected (i.e., $d_N/d_S > 1$) sites were found using any of the molecular evolutionary analyses implemented in this study.

Table 3. *Results of random sites models (PAML) for RPE65 performed on the ML tree*

						Parameters					
Tree	Model	np	ln *L*	κ	ω_0/p	ω_1/q	ω_2/ω_p	Null	LRT	df	*p*
Vertebrate	M0	229	−41143.6	2.08		0.065		N/A			
	M1a	230	−40979.2	2.17	0.06 (97.2%)	1 (2.8%)		M0	328.7	1	0.00
	M2a	232	−40979.2	2.17	0.06 (97.2%)	1 (2.8%)	20.3 (0%)	M1a	0.0	2	1.00
	M3	233	−40184.7	2.98	0.01 (51.5%)	0.09 (35.0%)	0.26 (13.6%)	M0	1917.8	4	0.00
	M7	230	−40166.8	2.98	0.396	4.736		N/A			
	M8a	231	−40156.3	2.98	0.411	5.181	1 (0.2%)	N/A			
	M8	232	−40155.5	2.98	0.411	5.168	1.3 (0.2%)	M7	22.6	2	0.004
								M8a	1.5	1	0.217
Mammalian	M0	131	−16150.7	2.99		0.055		N/A			
	M1a	132	−16051.4	3.16	0.04 (95.6%)	1 (4.4%)		M0	519.7	1	0.00
	M2a	134	−16051.4	3.16	0.04 (95.6%)	1 (4.4%)	13.1 (0%)	M1a	0.0	2	1.00
	M3	135	−15903.6	2.98	0.01 (73.6%)	0.14 (21.0%)	0.41 (5.4%)	M0	1397.5	4	0.00
	M7	132	−15904.8	2.98	0.211	3.09		N/A			
	M8	134	−15904.5	2.98	0.215	3.240	1 (0.2%)	M7	0.0	2	0.76

Note: np, number of parameters; ln *L*, ln likelihood; κ, transition/transversion ratio; df, degrees of freedom. For models M0–M3, the ω values for each site class (ω_0–ω_2) are shown with their proportions in parentheses. For models M7 and M8, *p* and *q* describe the shape of the beta distribution, and ω_p refers to the positively selected site class (with proportion in parentheses) for models M8 and M8a (where it is constrained to one).

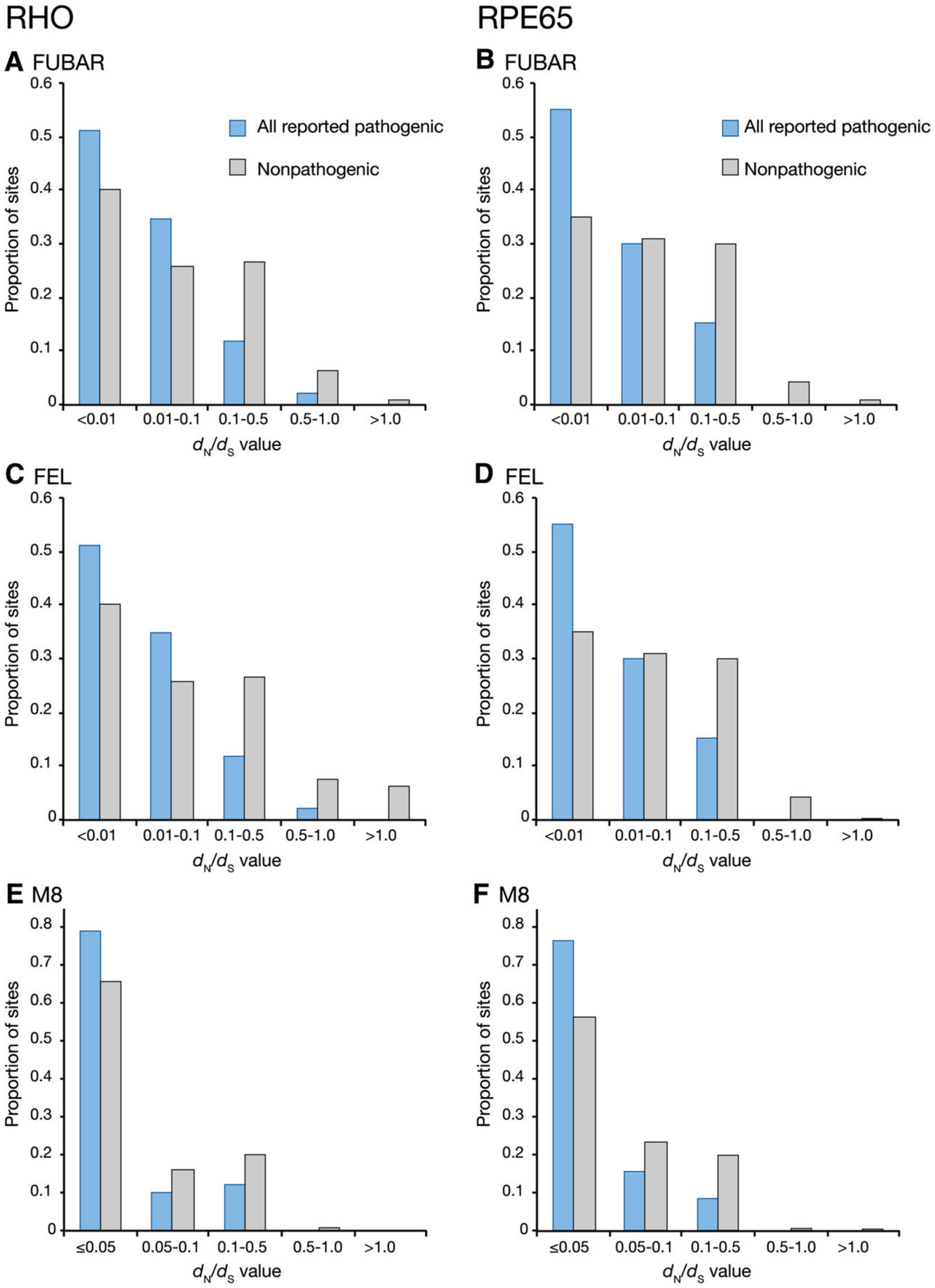

Fig. 1. Frequency histograms comparing the assignment of d_N/d_S ratios of pathogenic and nonpathogenic sites for FUBAR, FEL, and M8 analyses of vertebrate RHO and RPE65. For RHO, all 84 disease sites were used to generate histograms. Statistical comparisons of d_N/d_S values between pathogenic *versus* nonpathogenic sites for all three methods are summarized in Supplemental Table 6.

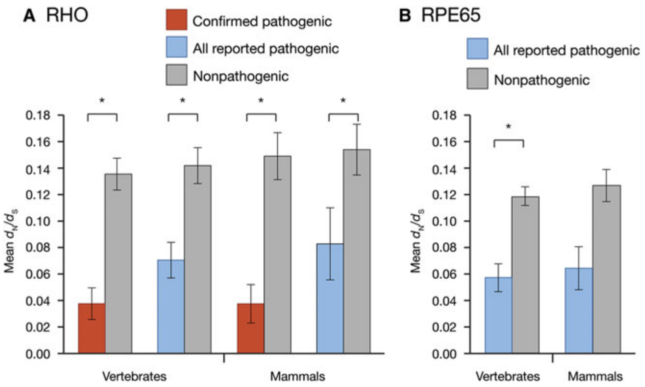

Fig. 2. Differences in mean d_N/d_S of pathogenic *versus* nonpathogenic sites in (A) RHO and (B) RPE65 datasets. Estimates of d_N/d_S were obtained with FUBAR. In all instances pathogenic sites had significantly lower d_N/d_S with the exception of the RPE65 mammal dataset (Mann–Whitney test, $P < 0.05$). Statistical comparisons are summarized in Supplemental Table 7. Error bars represent ±S.E.M.

Evolutionary constraint on disease-associated amino acid sites

In both RHO and RPE65, we compared estimates of d_N/d_S across individual amino acid sites generated by three different approaches: the M8 model in PAML, and the FEL and FUBAR methods in HYPHY (Kosakovsky Pond et al., 2005; Yang, 2005) (Murrell et al., 2013). Fig. 1 shows a series of histograms illustrating the distribution of pathogenic and nonpathogenic sites across d_N/d_S value categories, as estimated by FUBAR, FEL, and M8 in vertebrate RHO (all reported sites) and vertebrate RPE65. Regardless of the method used, pathogenic sites had significantly lower d_N/d_S when compared to nonpathogenic sites (Supplemental Table 6); however, Fig. 1 illustrates how the distribution of estimated d_N/d_S values in RHO and RPE65 differs depending on the method used.

FUBAR estimates the highest proportion of disease sites with d_N/d_S values less than 0.01 in both RHO and RPE65 (Fig. 1A and 1B), with an increasingly lower proportion estimated at higher d_N/d_S values. FUBAR does not estimate any pathogenic sites with d_N/d_S values greater than 0.5 in RPE65 (Fig. 1B). Nonpathogenic sites in both genes show a more even distribution across the d_N/d_S categories, with 6–10% of sites approaching neutral selection ($d_N/d_S = 1$) and a small proportion exceeding d_N/d_S of 1. FEL estimates of d_N/d_S reveal a similar trend to the FUBAR analyses (Fig. 1C and 1D). In contrast, for any given site in RHO or RPE65, M8 does not estimate a d_N/d_S value less than 0.05 (Fig. 1E and 1F). The overall d_N/d_S estimates show a narrower distribution in both pathogenic and nonpathogenic sites, with the majority of both types of sites having an estimated d_N/d_S value of 0.05.

To determine if pathogenic sites had statistically significantly lower d_N/d_S values than nonpathogenic sites, we compared values using the nonparametric Mann–Whitney test. For d_N/d_S values estimated with FUBAR at each amino acid site, vertebrate and mammalian RHO sequences showed significantly lower d_N/d_S values at all reported (84) pathogenic sites (Fig. 2A; Supplemental Table 7). The vertebrate datasets exhibited greater overall constraint (lower d_N/d_S) on pathogenic sites relative to the mammalian datasets and showed more significant differences when compared to the d_N/d_S values of nonpathogenic sites (Fig. 2A, Supplemental Table 7). This difference was even larger when the confirmed pathogenic subset of sites was compared to nonpathogenic sites, with these 53 sites having lower d_N/d_S relative to the 84 sites analyzed, and these d_N/d_S values were more significantly different from nonpathogenic sites (Fig. 2A, Supplemental Table 7). This pattern was found in both vertebrates and mammals. In RPE65, a similar finding was recovered using the FUBAR-estimated d_N/d_S values. Vertebrates exhibited significant differences in d_N/d_S at pathogenic *versus* nonpathogenic sites, while the differences detected in mammals were not significant, likely due to the lower available sample size (Fig. 2B; Supplemental Table 7).

We found that differences between pathogenic and nonpathogenic sites remained significant using both the M8 and FEL methods, and supported our FUBAR results. In general, we found that out of the three methods, the FUBAR method performed best (Supplemental Tables 6 and 7). At certain sites, FEL estimates either undefined or infinite values of d_N/d_S due to extremely low d_S values, so $d_N - d_S$ is also used to infer selection. When $d_N - d_S$ was compared between pathogenic and nonpathogenic sites, differences were not significant with the exception of the confirmed subset of pathogenic sites in mammalian and vertebrate RHO, and vertebrate RPE65 (Supplemental Table 8). Although the d_N/d_S estimates generated through M8 in PAML spanned a much narrower range of values relative to FUBAR and FEL, we still detected differences between pathogenic and nonpathogenic sites.

To explore the distribution of pathogenic *versus* nonpathogenic sites in more detail, we plotted each amino acid site with respect to the number of different amino acid residues found at that site in the alignment, against its assigned d_N/d_S value as estimated using FUBAR, for RHO and RPE65 (Fig. 3A and 3B). This shows that the majority of pathogenic sites are consistently found at invariant and highly conserved amino acid sites (with accompanying low d_N/d_S), while nonpathogenic sites show more variation in both dimensions. This pattern is even stronger in the “confirmed” subset of RHO pathogenic sites (Fig. 3A).

Since FUBAR allows d_N and d_S to be estimated independently, we were able to obtain individual values of d_N and d_S for each site in our vertebrate RHO and RPE65 alignments. Vertebrate RHO shows significant differences in d_N between all reported pathogenic sites and nonpathogenic sites (Fig. 4A and 4B; Supplemental Table 9), but no differences in d_S. The confirmed pathogenic set of RHO sites, however, shows significantly lower d_N and higher d_S. In vertebrate RPE65, d_N is significantly lower and d_S is significantly higher compared to pathogenic sites (Fig. 4C and 4D; Supplemental Table 9).

Disease sites and d_N/d_S on the crystal structure

Using FUBAR-estimated vertebrate d_N/d_S values, we created a d_N/d_S heat map of conservation (where dark blue = most conserved and white = least conserved) to illustrate the relationship between structure and disease in both RHO and RPE65 (Fig. 5). The interior views of RHO (Fig. 5A) and RPE65 (Fig. 5B) show conserved residues surrounding the retinal chromophore in RHO and the catalytically active iron cofactor in RPE65. Conversely, the exterior of the proteins (the smaller structures shown above the interior views of the respective proteins) has higher variability in d_N/d_S. Pathogenic sites, outlined in red, are distributed throughout both proteins, tending to occur in areas of low d_N/d_S.

Discussion

This study used a phylomedicine-based approach to investigate RP-associated mutations in two critical visual genes using interspecific sequence variation across vertebrates and mammals. In vertebrate and mammalian RHO sequences, we found that sites implicated in RP had significantly lower rates of nonsynonymous to synonymous substitution (d_N/d_S) relative to sites that were not associated with disease. We also found similarly significant lower rates for pathogenic sites in the vertebrate dataset for the visual cycle gene RPE65, but we did not find significant differences in the mammalian dataset, likely due to the smaller sample size and reduced variation in these sequences. In both genes, our vertebrate dataset yielded more statistically significant differences in mean d_N/d_S between pathogenic and nonpathogenic amino acid sites, as compared with a more restricted mammalian dataset. For RHO, these differences were even more striking when we only included disease-associated sites that had substantial biochemical or clinical support for pathogenicity. Our results were consistent regardless of the molecular evolutionary approach used, but here we discuss some important differences among these methods and their implications for future studies of this nature. Finally, we found that differences in the nonsynonymous substitution rate (d_N), rather than the synonymous substitution rate (d_S), between pathogenic and nonpathogenic sites, drive the overall differences in d_N/d_S rates in most instances. We discuss the importance of dataset properties, pathogenic site selection, and molecular evolutionary method for our results, and compare our approach to similar evolutionary studies of visual disease genes and human disease in general.

Implications of dataset properties on analyses of disease sites

In both visual disease genes, we consistently found significantly lower mean d_N/d_S in all analyses of pathogenic sites compared to nonpathogenic sites, with the single exception of the mammalian RPE65 dataset. This is most likely due to the absence of sufficient interspecific variation to generate (detectable) differences in d_N/d_S among sites. Though both genes are highly conserved among mammals, the additional mammalian sequences included in our RHO dataset may have provided sufficient additional variation to detect significant differences between pathogenic and nonpathogenic sites. In general, our analyses of the vertebrate dataset had greater differences between pathogenic and nonpathogenic sites and higher significance levels. Previous comparative studies of human disease genes have used various levels of evolutionary depth and taxonomic sampling (e.g., RP, Briscoe et al., 2004; deafness, Kirwan et al., 2013; and cystic fibrosis, Rishishwar et al. 2012). Past work has suggested that analyses of mammalian sequences are most appropriate for studies of human disease (Cooper et al., 2003). While this is feasible for extensively sequenced genes (e.g., RHO), genes with less sequence information, such as RPE65, may require a wider evolutionary scope for statistically robust analyses. In general, a wider scope when fewer interspecific sequences are available is likely to enhance the ability to detect differences between pathogenic and nonpathogenic sites, due to an increased amount of variation at nonpathogenic sites as a result of increased evolutionary divergence, when function is conserved. It is important to note that increased sequence divergence can also lead to functional divergence, potentially confounding studies aiming to investigate human disease. In the case of RHO and RPE65, where function is highly conserved across vertebrates, this is not likely to be an issue, but should be carefully considered for future studies on different genes. Fortunately, with the increasing number of both mammalian and vertebrate genomes available, evolutionary studies of human disease using interspecific variation can be undertaken at multiple taxonomic levels.

Beyond varying the taxonomic scale in our study, we also performed our RHO analyses on two separate sets of pathogenic sites. We found that both our more stringent pathogenic site dataset (53 confirmed pathogenic sites, selected based on substantial clinical or biochemical evidence for deleterious effects; Supplemental Table 2) and the “all reported” pathogenic sites showed significant differences between pathogenic and nonpathogenic sites. The confirmed pathogenic sites dataset was under greater constraint than all reported pathogenic sites and thus gave more significant results, suggesting that some of the “all reported” pathogenic sites may not actually be pathogenic, or may be less severe and therefore under weaker purifying selection. This result demonstrates that the method is robust to the inclusion of sites that do not have biochemical or clinical confirmation, since most disease genes lack substantial experimental investigation of point mutations and their impact on protein structure and function.

These results also highlight an application of the method for selecting putative disease-causing mutations that lack clinical or biochemical characterization but have low d_N/d_S similar to known pathogenic sites, for future *in vitro* mutagenesis experiments. Mutations implicated in disease, but of unknown effect, that show comparatively lower evolutionary constraint (i.e., more interspecific variation) may also warrant further clinical or experimental investigation to confirm that they are actually deleterious rather than benign polymorphisms (e.g., Vincent et al., 2013). For disease-associated genes with less functional data available, deleterious mutations found in these databases can be examined through comparative approaches outlined in this study to identify high priority sites (highly conserved sites) for further investigation.

Detection of higher constraint in disease sites achieved with multiple methods

The three molecular evolutionary methods used to investigate RHO and RPE65 (FUBAR, FEL, and M8) yielded similar results, and mean d_N/d_S estimates of each method showed significant differences between pathogenic and nonpathogenic sites. The distribution of disease sites across varying levels of evolutionary conservation was comparable between RHO and RPE65 regardless of the method implemented. Using ratios derived from FEL may be problematic when d_S is zero, as d_N/d_S will be calculated as infinite (or undefined when d_N is also zero). In these instances, using normalized $d_N - d_S$ values for such comparisons between site categories could be used instead of d_N/d_S, or, the issue can be avoided *via* use of methods such as FUBAR that incorporate alignment-wide information rather than calculating d_N and d_S for each site independently (Kosakovsky Pond, 2005; Murrell et al., 2013). We found that in contrast to FUBAR and FEL, site-by-site d_N/d_S estimates from M8 were assigned a much narrower range of values, since these estimates are generally restricted to 10 rate categories. This discretization of the M8 β distribution into rate categories also results in d_N/d_S estimates that do not fall below a certain threshold (in the case of these analyses, 0.05). In contrast, FUBAR avoids this due to its larger number of site classes that are able to capture sequence variation that cannot otherwise be captured by a (discretized) β distribution. Despite these discrepancies across methods, we were still able to detect significant differences between pathogenic and nonpathogenic sites in RHO, and the M8 model in particular has been successfully used in

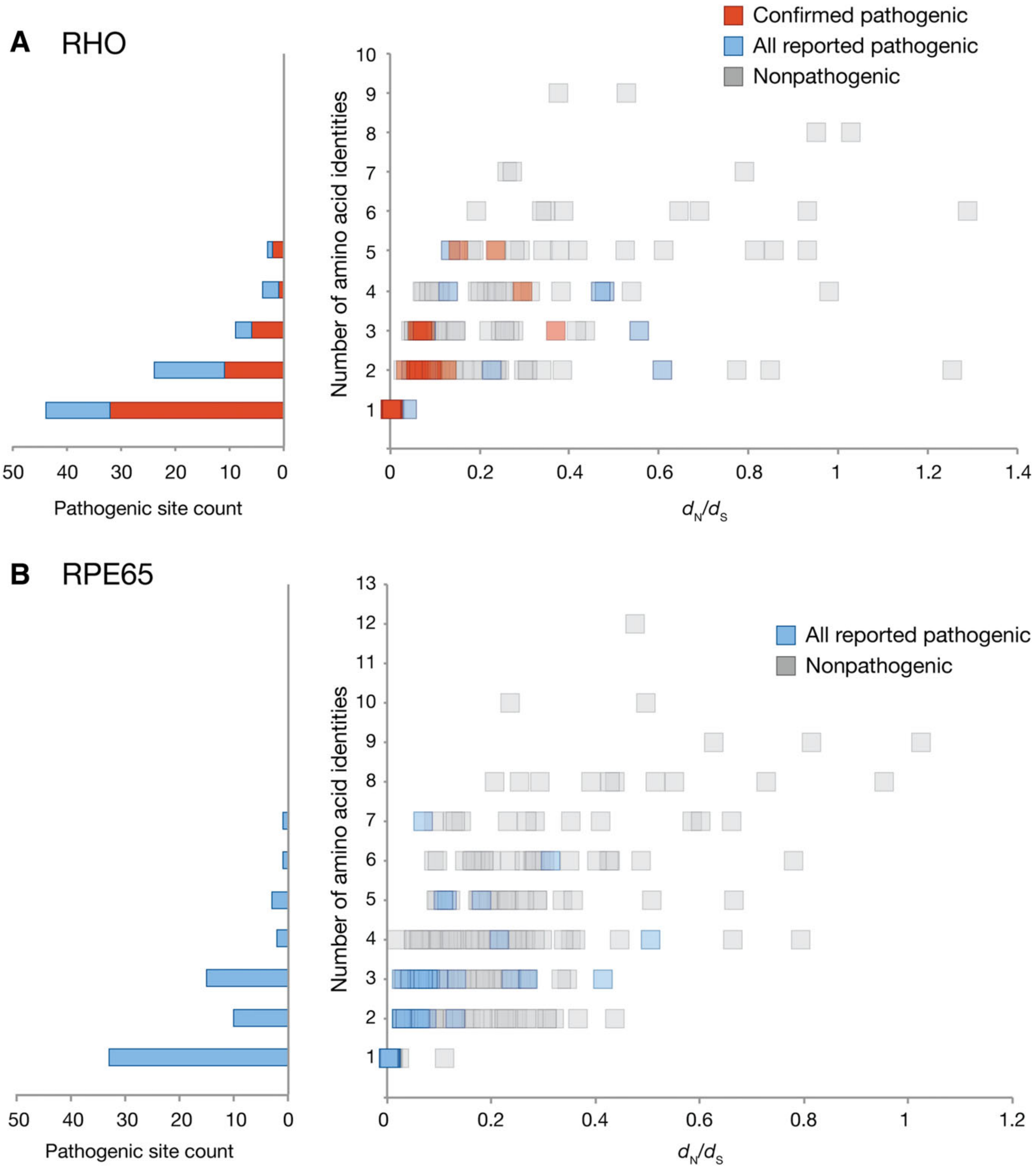

Fig. 3. Distribution of all alignment sites in (A) Vertebrate RHO and (B) Vertebrate RPE65 with respect to the number of different amino acid identities at a particular site and FUBAR-estimated d_N/d_S values. All reported RHO pathogenic sites are shown in blue, and confirmed pathogenic RHO sites are shown in red. To illustrate the number of disease-associated alignment sites with respect to amino acid variation (where 1 = one amino acid, i.e., no variation; 2 = two different amino acids, etc.), raw counts of the pathogenic sites relative to the number of different amino acid identities at a particular alignment site are shown to the left of each graph.

previous studies to investigate auditory disease mutations in mammals (Kirwan et al., 2013).

An advantage of the FUBAR and FEL analyses implemented in this investigation is the ability to separately examine d_N and d_S. Differing levels of evolutionary constraint on nonsynonymous substitutions at disease-associated amino acid sites are intuitive, since these mutations alter the amino acid sequence, and therefore the biochemical properties, of a particular protein (Li et al., 1985).

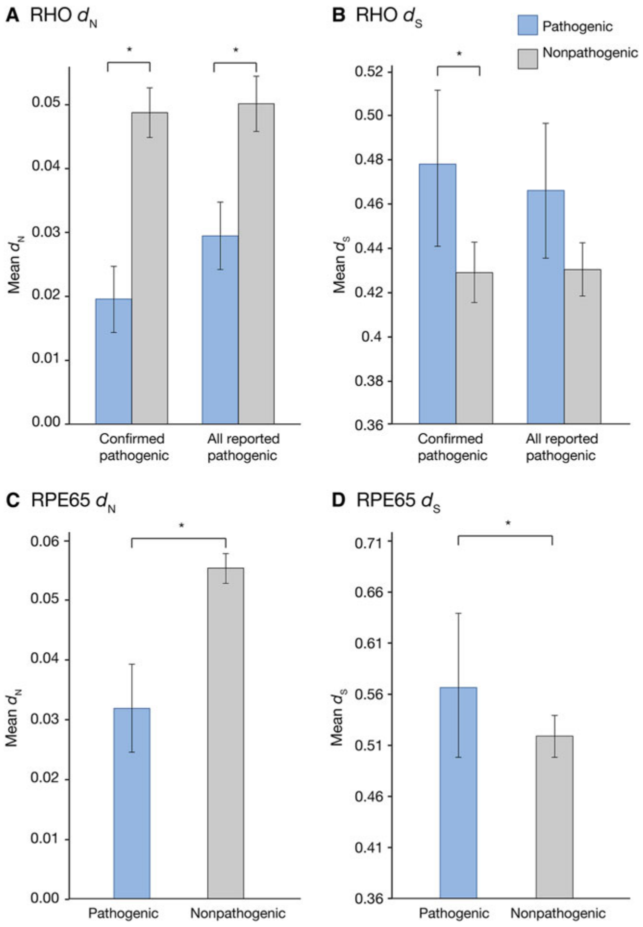

Fig. 4. Differences in mean nonsynonymous and mean synonymous substitution rates of pathogenic *versus* nonpathogenic sites in the RHO vertebrate datasets (A and B) and RPE65 (C and D), determined through FUBAR analyses. In both "all reported" and "confirmed pathogenic" site categories, d_N values were significantly lower relative to nondisease sites (Mann–Whitney test, $P < 0.05$). For confirmed sites, d_S was significantly higher in disease sites, but this was not the case for "all reported" sites. Statistical comparisons are summarized in Supplemental Table 9. Error bars represent ±S.E.M.

Elevated mean d_S at pathogenic sites in RHO and RPE65 may be driven by outlier pathogenic sites with high d_S or perhaps interspecific variation in codon usage bias at these sites (Plotkin & Kudla, 2011). The role of synonymous variation in human disease has gained prominence in recent years (e.g., Sauna et al., 2007; Sauna & Kimchi-Sarfaty, 2011), and work on mammalian RHO in particular has identified selection on synonymous sites, likely contributing to mRNA stability and translational efficiency (Du et al., 2014). A broader investigation of many protein-coding gene families also found that genes with synonymous rate variation were overrepresented in genetic diseases (Dimitrieva & Anisimova, 2014). Our finding that d_S may be consistently higher in pathogenic sites suggests that future work investigating the role of synonymous rate variation in genes involved in human disease is warranted.

Protein structure, d_N/d_S, and pathogenicity

Although an examination of specific structural defects in protein function is not readily available with the codon-based models used in this study, incorporating metrics of conservation from multiple methods can be an important first step for future analyses integrating protein structure and amino acid identity into disease site prediction and downstream phenotypic characterization of mutant proteins. For instance, RP-causing mutations in RHO are found throughout the protein, with the majority located within the transmembrane domains. Known pathogenic mutations disrupt protein folding, and several restricted to the C-terminus interfere with protein translocation but proper protein structure is maintained (Mendes et al., 2005). Critical structural sites, such as those that participate in glycosylation and those responsible for correct protein folding, are highly conserved across vertebrates and mutations at these sites invariably result in protein defects (Pope et al., 2014). Conserved d_N/d_S estimates mapped onto the RHO crystal structure coincide with these known highly conserved regions, as well as confirmed pathogenic sites. In some cases, however, a pathogenic site may be tolerant of a nonsynonymous substitution to a particular amino acid residue, while the true disease-associated residue differs considerably from the wild type with respect to size or charge. For instance, the A292S substitution in RHO occurs frequently across vertebrates (e.g., Sugawara et al., 2010), while the A292E substitution is not found in vertebrate sequences and is known to cause constitutive activation of RHO (Weitz & Nathans, 1992). Moreover, different amino acid substitutions at the same amino acid site in RHO may lead to different clinical disease phenotypes (Neidhardt et al., 2006). In our vertebrate analyses, we found that in both RHO and RPE65 certain pathogenic amino acid sites (including those with functional validation in RHO) exhibited some variability, supporting the use of amino acid characterization methods to determine which residue in particular is deleterious. In these instances, further *in silico* investigations of the mutation in question could be undertaken with disease prediction methods such as SIFT, PolyPhen, or Provean (Ng & Henikoff, 2003; Adzhubei et al., 2010; Flanagan et al., 2010; Choi et al., 2012), with additional *in vitro* characterization to corroborate the mutation.

Intuitively, sites in both RHO and RPE65 showed high evolutionary conservation in structurally important regions, particularly when contrasting the interior with the exterior of the protein. Much like RHO, pathogenic residues in RPE65 are located throughout the protein and their deleterious effects are mediated *via* a variety of mechanisms (Bereta et al., 2008). In RPE65, amino acid sites associated with isomerase activity, those pointing toward the catalytic interior of the protein (specifically, four highly conserved histidine residues), and those that may mediate membrane association are tightly constrained (Kiser & Palczewski, 2010; Li et al., 2015). Recent mutagenesis work has begun to elucidate the importance of these and other disease-associated residues for RPE65 function (Li et al., 2014). Since known pathogenic mutations are widespread in the RPE65 structure, rather than localized to a few key regions, further work investigating mutations *in vitro* will lend valuable insight to its function.

Evolutionary conservation and disease site prediction

The comparative approach used in this study is ideal for visual genes that are known to have many point mutations associated with RP. In some instances, however, only one or two pathogenic mutations in a given visual protein are known, and there are likely other deleterious mutations in such proteins yet to be discovered. Other techniques complementary to the ones employed in this study can be used to elucidate the nature of these mutations. Comparing observed and expected numbers of disease-associated

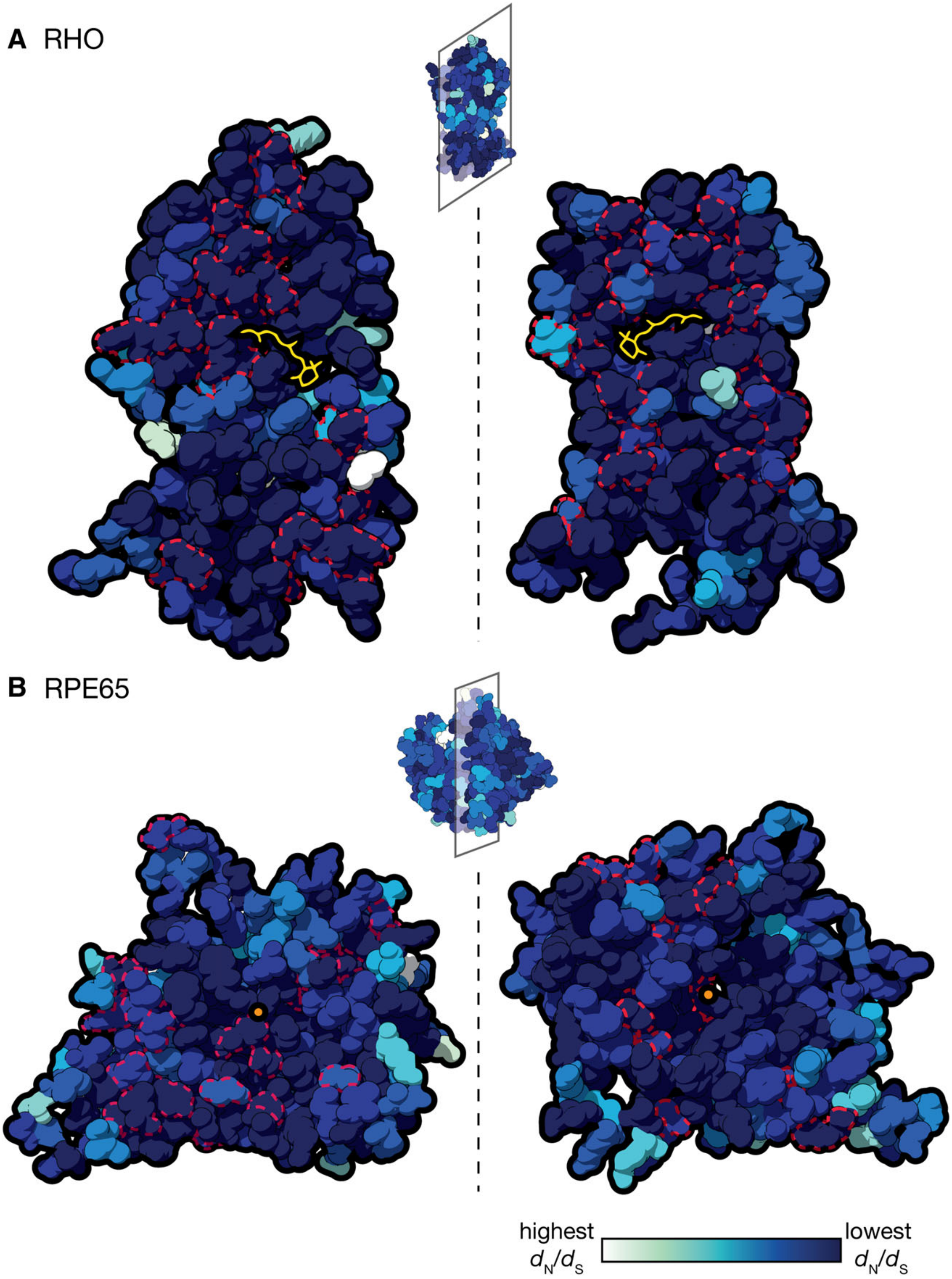

Fig. 5. d_N/d_S values estimated from vertebrate FUBAR analyses converted into a conservation heat map and mapped on the 3D crystal structures of (A) RHO and (B) RPE65. The structures of both proteins have been bisected and internal views are shown. The RHO chromophore is shown in yellow, and the RPE65 iron cofactor as an orange dot. Pathogenic sites are outlined in dotted red lines.

nonsynonymous substitutions across sites in a protein has been used to explore the spectrum of selective constraint across genes (reviewed in Miller & Kumar, 2001; Kumar et al., 2011). For these genes, even a small number of disease sites can be investigated, by testing whether disease mutations occur at highly conserved sites more frequently than expected by chance. An evolutionary prior, or "conservation index" ($1 - d_N/d_S$) has also been used to evaluate selective constraint across sites undergoing purifying selection (Burk-Herrick et al., 2006). In this approach, sites undergoing neutral or positive selection have an index set to 0, generating a spectrum of sites ranging from weak to strong purifying selection, wherein the most highly conserved sites have a conservation index approaching 1. Similar to our expectations for this study, pathogenic sites in the mammalian BRCA1 gene and 10 mammalian auditory genes were found to predominantly occur in regions undergoing strong purifying selection (Burk-Herrick et al., 2006; Kirwan et al., 2013). An additional caveat to the methods used in this study is that missense/ nonsense mutations were the sole target of our analyses. In addition to defects in protein structure/function, the role of pre-mRNA splicing mutations in human disease has been documented, including in hereditary visual diseases (e.g., Ueyama et al., 2012). Extending these existing phylomedicine approaches to incorporate splicing mutations would be a valuable avenue of future study.

Current tools aiming to predict the occurrence of novel disease sites often incorporate structural information for specific amino acids as well as entire proteins (Ng & Henikoff, 2003; Adzhubei et al., 2010) and can filter sequence variants that are of potential functional importance. In some cases, these methods may be sensitive to changes in alignment and taxonomic sampling and may not be sufficiently sensitive to resolve ambiguity surrounding particular disease sites (Flanagan et al., 2010; Hicks et al., 2011; Castellana et al., 2015); however, incorporating structural information and amino acid properties is an important aspect of disease prediction that is not explicitly addressed in the molecular evolutionary analyses undertaken here. Statistical estimates of selection obtained from several different molecular evolutionary methods could be integrated into existing techniques that use interspecific alignments to assess the severity of a given mutation (e.g., PolyPhen or SIFT) or complement these alignment-based analyses, improving the ability for interspecific variation to inform studies of human disease and predict deleterious mutations.

Conclusions

With the advancement of high-throughput sequencing technology, rapid and extensive recovery of putative genes and mutations implicated in visual disease is increasingly feasible (Neveling et al., 2012). This volume of data, particularly for newly recovered missense mutations, means that further validation of their clinical relevance is necessary. How likely are newly recovered mutations or genes to result in a disease phenotype? In instances where insufficient clinical or sequence information preclude the identification of important amino acid sites, computational approaches drawing on interspecific data found in vertebrate genomes, such as those employed in this study, may serve as an important first step in the characterization of novel genes and mutations. Genes implicated in hereditary visual disease are good candidates for these studies since there are abundant homologous sequence data for many genes, and a strong relationship between genotype and visual degeneration phenotype can be established. In this study, we used several molecular evolutionary methods and interspecific datasets to investigate disease-causing mutations in RHO and RPE65 and found that pathogenic sites in these genes were under significantly stronger evolutionary constraint than nonpathogenic sites. Models of molecular evolution may aid in the identification, prediction, and investigation of point mutations associated with human disease. Moreover, at the codon level, they may reveal patterns of selection not easily recovered with amino acid-based approaches. Current disease prediction methods using homologous sequences and amino acid properties may be complemented by codon-based molecular evolutionary analyses, advancing our understanding of conservation, protein function, and pathogenicity.

Acknowledgments

This work was supported by an NSERC Discovery Grant and a Foundation for Fighting Blindness seed grant to BSWC, Vision Science Research Fellowships to FEH, RKS, GMC, and AVN, and Ontario Graduate Fellowships to FEH and RKS. Amir Sabouhanian provided valuable feedback on the statistical analyses.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S. & Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249.

Anasagasti, A., Irigoyen, C., Barandika, O., López de Munain, A. & Ruiz-Ederra, J. (2012). Current mutation discovery approaches in retinitis pigmentosa. *Vision Research* **75**, 117–129.

Anisimova, M. & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology* **55**, 539–552.

Bereta, G., Kiser, P.D., Golczak, M., Sun, W., Heon, E., Saperstein, D.A. & Palczewski, K. (2008). Impact of retinal disease-associated RPE65 mutations on retinoid isomerization. *Biochemistry* **47**, 9856–9865.

Bowne, S.J., Sullivan, L.S., Koboldt, D.C., Ding, L., Fulton, R., Abbott, R.M., Sodergren, E.J., Birch, D.G., Wheaton, D.H., Heckenlively, J.R., Liu, Q., Pierce, E.A., Weinstock, G.M. & Daiger, S.P. (2011). Identification of disease-causing mutations in autosomal dominant retinitis pigmentosa (adRP) using next-generation DNA sequencing. *Investigative Ophthalmology & Visual Science* **52**, 494–503.

Breikers, G., Portier-VandeLuytgaarden, M.J.M., Bovee-Geurts, P.H.M. & DeGrip, W.J. (2002). Retinitis pigmentosa-associated rhodopsin mutations in three membrane-located cysteine residues present three different biochemical phenotypes. *Biochemical and Biophysical Research Communications* **297**, 847–853.

Briscoe, A.D., Gaur, C. & Kumar, S. (2004). The spectrum of human rhodopsin disease mutations through the lens of interspecific variation. *Gene* **332**, 107–118.

Burk-Herrick, A., Scally, M., Amrine-Madsen, H., Stanhope, M.J. & Springer, M.S. (2006). Natural selection and mammalian BRCA1 sequences: Elucidating functionally important sites relevant to breast cancer susceptibility in humans. *Mammalian Genome* **17**, 257–270.

Castellana, S., Rónai, J. & Mazza, T. (2015). MitImpact: An exhaustive collection of pre-computed pathogenicity predictions of human mitochondrial non-synonymous variants. *Human Mutation* **36**, E2413–E2422.

Choe, H-W., Kim, Y.J., Park, J.H., Morizumi, T., Pai, E.F., Krauß, N., Hofmann, K.P., Scheerer, P. & Ernst, O.P. (2011). Crystal structure of metarhodopsin II. *Nature* **471**, 651–655.

Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688.

Cideciyan, A.V. (2010). Leber congenital amaurosis due to RPE65 mutations and its treatment with gene therapy. *Progress in Retinal and Eye Research* **29**, 398–427.

Cooper, G.M., Brudno, M.; NISC Comparative Sequencing Program, Green, E.D., Batzoglou, S. & Sidow, A. (2003). Quantitative estimates of sequence divergence for comparative analyses of mammalian genomes. *Genome Research* **13**, 813–820.

Crottini, A., Madsen, O. & Poux, C. (2012). Vertebrate time-tree elucidates the biogeographic pattern of a major biotic change around the

K–T boundary in Madagascar. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5358–5363.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. (2010). Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455–2457.

Dimitrieva, S. & Anisimova, M. (2014). Unraveling patterns of site-to-site synonymous rates variation and associated gene properties of protein domains and families. *PLoS One* **9**, e95034.

Du, J., Dungan, S.Z., Sabouhanian, A. & Chang, B.S.W. (2014). Selection on synonymous codons in mammalian rhodopsins: A possible role in optimizing translational processes. *BMC Evolutionary Biology* **14**, 96.

Ferrari, S., Di Iorio, E., Barbaro, V., Ponzin, D., Sorrentino, F.S. & Parmeggiani, F. (2011). Retinitis pigmentosa: Genes and disease mechanisms. *Current Genomics* **12**, 238–249.

Flanagan, S.E., Patch, A-M. & Ellard, S. (2010). Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genetic Testing and Molecular Biomarkers* **14**, 533–537.

Fong, J.J., Brown, J.M., Fujita, M.K. & Boussau, B. (2012). A phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic lissamphibia. *PLoS One* **7**, e48990.

Gaucher, E.A., De Kee, D.W. & Benner, S.A. (2006). Application of DETECTER, an evolutionary genomic tool to analyze genetic variation, to the cystic fibrosis gene family. *BMC Genomics* **7**, 44.

Greenblatt, M.S., Beaudet, J.G., Gump, J.R., Godin, K.S., Trombley, L., Koh, J. & Bond, J.P. (2003). Detailed computational study of p53 and p16: Using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* **22**, 1150–1163.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307–321.

Hamosh, A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* **33**, D514–D517.

Hartong, D.T., Berson, E.L. & Dryja, T.P. (2006). Retinitis pigmentosa. *Lancet* **368**, 1795–1809.

Hicks, S., Wheeler, D.A., Plon, S.E. & Kimmel, M. (2011). Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Human Mutation* **32**, 661–668.

Hollingsworth, T.J. & Gross, A.K. (2013). The severe autosomal dominant retinitis pigmentosa rhodopsin mutant Ter349Glu mislocalizes and induces rapid rod cell death. *Journal of Biological Chemistry* **288**, 29047–29055.

Iannaccone, A., Man, D., Waseem, N., Jennings, B.J., Ganapathiraju, M., Gallaher, K., Reese, E., Bhattacharya, S.S. & Klein-Seetharaman, J. (2006). Retinitis pigmentosa associated with rhodopsin mutations: Correlation between phenotypic variability and molecular effects. *Vision Research* **46**, 4556–4567.

Janz, J.M., Fay, J.F. & Farrens, D.L. (2003). Stability of dark state rhodopsin is mediated by a conserved ion pair in intradiscal loop E-2. *Journal of Biological Chemistry* **278**, 16982–16991.

Kirwan, J.D., Bekaert, M., Commins, J.M., Davies, K.T.J., Rossiter, S.J. & Teeling, E.C. (2013). A phylomedicine approach to understanding the evolution of auditory sensory perception and disease in mammals. *Evolutionary Applications* **6**, 412–422.

Kiser, P.D. & Palczewski, K. (2010). Membrane-binding and enzymatic properties of RPE65. *Progress in Retinal and Eye Research* **29**, 428–442.

Kiser, P.D., Golczak, M., Lodowski, D.T., Chance, M.R. & Palczewski, K. (2009). Crystal structure of native RPE65, the retinoid isomerase of the visual cycle. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 17325–17330.

Kosakovsky Pond, S.L. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208–1222.

Kosakovsky Pond, S.L., Frost, S.D.W. & Muse, S.V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679.

Kumar, S., Dudley, J.T., Filipski, A. & Liu, L. (2011). Phylomedicine: An evolutionary telescope to explore and diagnose the universe of disease mutations. *Trends in Genetics* **27**, 377–386.

Li, S., Hu, J., Jin, R.J., Aiyar, A., Jacobson, S.G., Bok, D. & Jin, M. (2015). Temperature-sensitive retinoid isomerase activity of RPE65 mutants associated with Leber Congenital Amaurosis. *Journal of Biochemistry* **158**, 115–125.

Li, S., Izumi, T., Hu, J., Jin, H.H., Siddiqui, A-A.A., Jacobson, S.G., Bok, D. & Jin, M. (2014). Rescue of enzymatic function for disease-associated RPE65 proteins containing various missense mutations in non-active sites. *Journal of Biological Chemistry* **289**, 18943–18956.

Li, W.H., Wu, C.I. & Luo, C.C. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution* **2**, 150–174.

Liu, M.Y., Liu, J., Mehrotra, D., Liu, Y., Guo, Y., Baldera-Aguayo, P.A., Mooney, V.L., Nour, A.M. & Yan, E.C.Y. (2013). Thermal stability of rhodopsin and progression of retinitis pigmentosa: Comparison of S186W and D190N rhodopsin mutants. *Journal of Biological Chemistry* **288**, 17698–17712.

Löytynoja, A. & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10557–10562.

Löytynoja, A. & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632–1635.

Magrane, M. & Uniprot Consortium (2011). UniProt Knowledgebase: A hub of integrated protein data. Database 2011, bar009.

McKeone, R., Wikstrom, M., Kiel, C. & Rakoczy, P.E. (2014). Assessing the correlation between mutant rhodopsin stability and the severity of retinitis pigmentosa. *Molecular Vision* **20**, 183–199.

Mendes, H.F., van der Spuy, J., Chapple, J.P. & Cheetham, M.E. (2005). Mechanisms of cell death in rhodopsin retinitis pigmentosa: Implications for therapy. *Trends in Molecular Medicine* **11**, 177–185.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S. & Murphy, W.J. (2011). Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524.

Miller, M.P. & Kumar, S. (2001). Understanding human disease mutations through the use of interspecific genetic variation. *Human Molecular Genetics* **10**, 2319–2328.

Moiseyev, G., Chen, Y., Takahashi, Y., Wu, B.X. & Ma, J-X. (2005). RPE65 is the isomerohydrolase in the retinoid visual cycle. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12413–12418.

Mooney, S.D. & Klein, T.E. (2002). The functional importance of disease-associated mutation. *BMC Bioinformatics* **3**, 24.

Morgan, C.C., Mc Cartney, A.M., Donoghue, M.T.A., Loughran, N.B., Spillane, C., Teeling, E.C. & O'Connell, M.J. (2013). Molecular adaptation of telomere associated genes in mammals. *BMC Evolutionary Biology* **13**, 251.

Morrow, J.M. & Chang, B.S.W. (2015). Comparative mutagenesis studies of retinal release in light-activated zebrafish rhodopsin using fluorescence spectroscopy. *Biochemistry* **54**, 4507–4518.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution* **30**, 1196–1205.

Neidhardt, J., Barthelmes, D., Farahmand, F., Fleischhauer, J.C. & Berger, W. (2006). Different amino acid substitutions at the same position in rhodopsin lead to distinct phenotypes. *Investigative Ophthalmology & Visual Science* **47**, 1630–1635.

Neveling, K., Collin, R.W., Gilissen, C., van Huet, R.A., Visser, L., Kwint, M.P., Gijsen, S.J., Zonneveld, M.N., Wieskamp, N., de Ligt, J., Siemiatkowska, A.M., Hoefsloot, L.H., Buckley, M.F., Kellner, U., Branham, K.E., den Hollander, A.I., Hoischen, A., Hoyng, C., Klevering, B.J., van den Born, L.I., Veltman, J.A., Cremers, F.P. & Scheffer, H. (2012). Next-generation genetic testing for retinitis pigmentosa. *Human Mutation* **33**, 963–972.

Ng, P.C. & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812–3814.

Nishiguchi, K.M., Tearle, R.G., Liu, Y.P., Oh, E.C., Miyake, N., Benaglio, P., Harper, S., Koskiniemi-Kuendig, H., Venturini, G., Sharon, D., Koenekoop, R.K., Nakamura, M., Kondo, M., Ueno, S., Yasuma, T.R., Beckmann, J.S., Ikegawa, S., Matsumoto, N., Terasaki, H., Berson, E.L., Katsanis, N. & Rivolta, C. (2013). Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene.

Proceedings of the National Academy of Sciences of the United States of America **110**, 16139–16144.

Okada, T., Sugihara, M., Bondar, A-N., Elstner, M., Entel, P. & Buss, V. (2004). The retinal conformation and its environment in rhodopsin in light of a new 2.2Å crystal structure. *Journal of Molecular Biology* **342**, 571–583.

Opefi, C.A., South, K., Reynolds, C.A., Smith, S.O. & Reeves, P.J. (2013). Retinitis pigmentosa mutants provide insight into the role of the N-terminal cap in rhodopsin folding, structure, and function. *Journal of Biological Chemistry* **288**, 33912–33926.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289**, 739–745.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. & Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **25**, 1605–1612.

Philpa, A.R., Jin, M., Li, S., Schindler, E.I., Iannaccone, A., Lam, B.L., Weleber, R.G., Fishman, G.A., Jacobson, S.G., Mullins, R.F., Travis, G.H. & Stone, E.M. (2009). Predicting the pathogenicity of RPE65 mutations. *Human Mutation* **30**, 1183–1188.

Piechnick, R., Ritter, E., Hildebrand, P.W., Ernst, O.P., Scheerer, P., Hofmann, K.P. & Heck, M. (2012). Effect of channel mutations on the uptake and release of the retinal ligand in opsin. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 5247–5252.

Pierrottet, C.O., Zuntini, M., Digiuni, M., Bazzanella, I., Ferri, P., Paderni, R., Rossetti, L.M., Cecchin, S., Orzalesi, N. & Bertelli, M. (2014). Syndromic and non-syndromic forms of retinitis pigmentosa: A comprehensive Italian clinical and molecular study reveals new mutations. *Genetics and Molecular Research* **13**, 8815–8833.

Plotkin, J.B. & Kudla, G. (2011). Synonymous but not the same: The causes and consequences of codon bias. *Nature Reviews Genetics* **12**, 32–42.

Pope, A., Eilers, M., Reeves, P.J. & Smith, S.O. (2014). Amino acid conservation and interactions in rhodopsin: Probing receptor activation by NMR spectroscopy. *Biochimica et Biophysica Acta* **1837**, 683–693.

Porter, M.L., Blasic, J.R., Bok, M.J., Cameron, E.G., Pringle, T., Cronin, T.W. & Robinson, P.R. (2011). Shedding new light on opsin evolution. *Proceedings of the Royal Society B: Biological Sciences* **279**, 3–14.

Redmond, T.M., Poliakov, E., Yu, S., Tsai, J-Y., Lu, Z. & Gentleman, S. (2005). Mutation of key residues of RPE65 abolishes its enzymatic role as isomerohydrolase in the visual cycle. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13658–13663.

Rishishwar, L., Varghese, N., Tyagi, E., Harvey, S.C., Jordan, I.K. & McCarty, N.A. (2012). Relating the disease mutation spectrum to the evolution of the Cystic Fibrosis transmembrane conductance regulator (CFTR). *PLoS One* **7**, e42336.

Rivolta, C., Sharon, D., DeAngelis, M.M. & Dryja, T.P. (2002). Retinitis pigmentosa and allied diseases: Numerous diseases, genes, and inheritance patterns. *Human Molecular Genetics* **11**, 1219–1227.

Ronquist, F. & Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Sauna, Z.E. & Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* **12**, 683–691.

Sauna, Z.E., Kimchi-Sarfaty, C., Ambudkar, S.V. & Gottesman, M.M. (2007). Silent polymorphisms speak: How they affect pharmacogenomics and the treatment of cancer. *Cancer Research* **67**, 9609–9612.

Scheffler, K., Martin, D.P. & Seoighe, C. (2006). Robust inference of positive selection from recombining coding sequences. *Bioinformatics* **22**, 2493–2499.

Schott, R.K., Refvik, S.P., Hauser, F.E., López-Fernández, H. & Chang, B.S.W. (2014). Divergent positive selection in rhodopsin from lake and riverine cichlid fishes. *Molecular Biology and Evolution* **31**, 1149–1165.

Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E.M. & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research* **29**, 308–311.

Stenson, P.D., Mort, M., Ball, E.V., Shaw, K., Phillips, A.D. & Cooper, D.N. (2013). The human gene mutation database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human Genetics* **133**, 1–9.

Stover, D.A. & Verrelli, B.C. (2010). Comparative vertebrate evolutionary analyses of type I collagen: Potential of COL1a1 gene structure and intron variation for common bone-related diseases. *Molecular Biology and Evolution* **28**, 533–542.

Sugawara, T., Imai, H., Nikaido, M., Imamoto, Y. & Okada, N. (2010). Vertebrate rhodopsin adaptation to dim light via rapid meta-II intermediate formation. *Molecular Biology and Evolution* **27**, 506–519.

Takahashi, Y., Moiseyev, G. & Ma, J.X. (2014). Identification of key residues determining isomerohydrolase activity of human RPE65. *Journal of Biological Chemistry* **289**, 26743–26751.

Ueyama, H., Miraki-Oda, S., Yamade, S., Tanabe, S., Yamashita, T., Shichida, Y. & Ogita, H. (2012). Unique haplotype in exon 3 of cone opsin mRNA affects splicing of its precursor, leading to congenital color vision defect. *Biochemical and Biophysical Research Communications* **424**, 152–157.

Vincent, A.L., Carroll, J., Fishman, G.A., Sauer, A., Sharp, D., Summerfelt, P., Williams, V., Dubis, A.M., Kohl, S. & Wong, F. (2013). Rhodopsin F45L allele does not cause autosomal dominant retinitis pigmentosa in a large caucasian family. *Translational Vision Science and Technology* **2**, 4.

Webb, A.E., Gerek, Z.N., Morgan, C.C., Walsh, T.A., Loscher, C.E., Edwards, S.V. & O'Connell, M.J. (2015). Adaptive evolution as a predictor of species-specific innate immune response. *Molecular Biology and Evolution* **32**, 1717–1729.

Weitz, C.J. & Nathans, J. (1992). Histidine residues regulate the transition of photoexcited rhodopsin to its active conformation, metarhodopsin II. *Neuron* **8**, 465–472.

Yang, G., Xie, S., Feng, N., Yuan, Z., Zhang, M. & Zhao, J. (2014). Spectrum of rhodopsin gene mutations in Chinese patients with retinitis pigmentosa. *Molecular Vision* **20**, 1132–1136.

Yang, Z. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution* **22**, 1107–1118.

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**, 1586–1591.

Zhao, H., Ru, B., Teeling, E.C., Faulkes, C.G., Zhang, S. & Rossiter, S.J. (2009). Rhodopsin molecular evolution in mammals inhabiting low light environments. *PLoS One* **4**, e8326.

Supplementary Data

Supplemental materials can be viewed in this issue of VNS by visiting http://journals.cambridge.org/VNS.