Universities of Leeds, Sheffield and York
http://eprints.whiterose.ac.uk/

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83423

Published paper

Sequence divergence and diversity suggests ongoing functional diversification of vertebrate NAD metabolism

Toni I. Gossmann [a],*, Mathias Ziegler [b]

[a] Department of Animal and Plant Sciences, University of Sheffield, Alfred Denny Building, S10 2TN Sheffield, United Kingdom
[b] Department of Molecular Biology, University of Bergen, Postbox 7803, 5020 Bergen, Norway

A B S T R A C T

NAD is not only an important cofactor in redox reactions but has also received attention in recent years because of its physiological importance in metabolic regulation, DNA repair and signaling. In contrast to the redox reactions, these regulatory processes involve degradation of NAD and therefore necessitate a constant replenishment of its cellular pool. NAD biosynthetic enzymes are common to almost all species in all clades, but the number of NAD degrading enzymes varies substantially across taxa. In particular, vertebrates, including humans, have a manifold of NAD degrading enzymes which require a high turnover of NAD. As there is currently a lack of a systematic study of how natural selection has shaped enzymes involved in NAD metabolism we conducted a comprehensive evolutionary analysis based on intraspecific variation and interspecific divergence. We compare NAD biosynthetic and degrading enzymes in four eukaryotic model species and subsequently focus on human NAD metabolic enzymes and their orthologs in other vertebrates. We find that the majority of enzymes involved in NAD metabolism are subject to varying levels of purifying selection. While NAD biosynthetic enzymes appear to experience a rather high level of evolutionary constraint, there is evidence for positive selection among enzymes mediating NAD-dependent signaling. This is particularly evident for members of the PARP family, a diverse protein family involved in DNA damage repair and programmed cell death. Based on haplotype information and substitution rate analysis we pinpoint sites that are potential targets of positive selection. We also link our findings to a three dimensional structure, which suggests that positive selection occurs in domains responsible for DNA binding and polymerization rather than the NAD catalytic domain. Taken together, our results indicate that vertebrate NAD metabolism is still undergoing functional diversification.

1. Introduction

Understanding how selection and genetic drift have shaped the composition of proteins is a central problem in population genetics and molecular evolution. Considering the complex interplay of enzymatic proteins and the critical homeostatic equilibrium they have to maintain it is not surprising that most amino acid changing mutations affecting enzymes will be deleterious [1,2]. Due to recent advances in DNA sequencing technology it is now possible to obtain whole genome information from numerous non-model organisms and mutational variation of multiple individuals of the same species at reasonable costs [3,4]. Patterns of sequence divergence between species and variation in natural populations can be used to reveal underlying selective fitness effects of mutations and can help to understand how natural selection has shaped the functions of enzymes and other proteins [2,5,6].

Enzymes are often involved in a complex network of interactions [7]. Various theories exist which aim to explain how complex enzyme networks can evolve [8]. A currently favored model, the patchwork model [9,10], proposes that pathways evolve from a system of enzymes with a broad substrate specificity to enzymes with specialized functions through gene duplications. The rate at which a particular enzyme is evolving within a network depends on a number of factors, including how its products are involved in interactions with other gene products and where these products are located within the network [11,12]. It has been proposed that the position of an enzymes within a metabolic pathway (e.g. upstream or at a branch point) may play a role, even though the models are restricted to relatively simple pathways [13,14]. In brief, evolutionary theory suggests that important enzymes tend to accumulate fewer deleterious substitutions and undergo positive selection more frequently [13].

* Corresponding author. +44 (0) 1142220062.
E-mail address: toni.gossmann@googlemail.com (T.I. Gossmann).
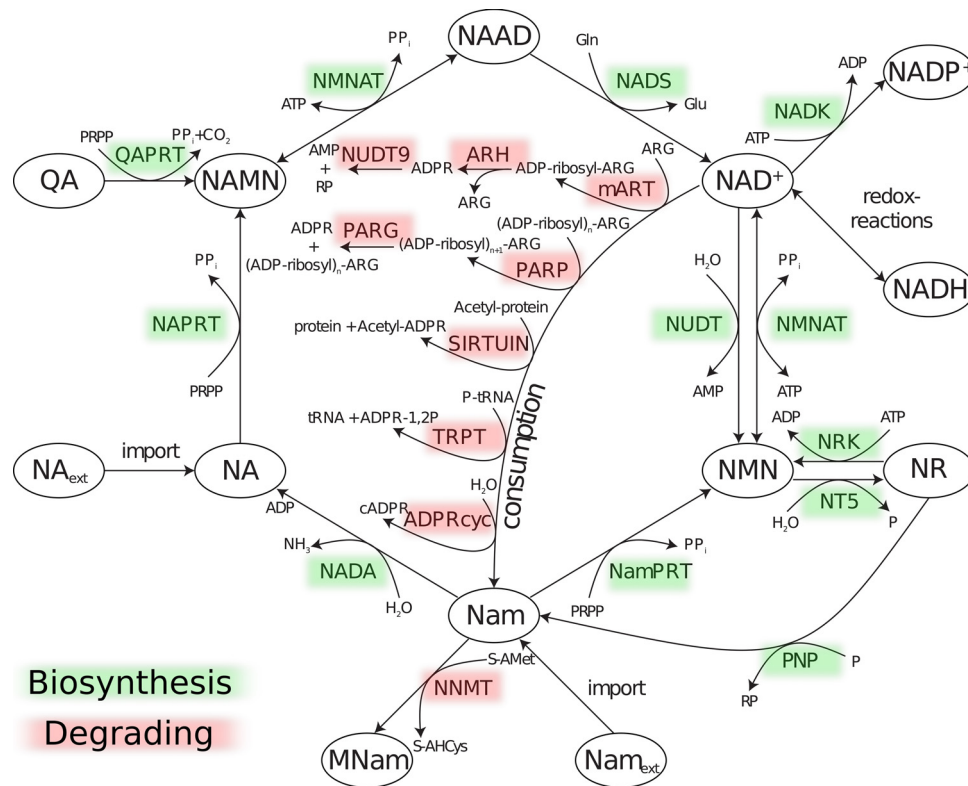
Fig. 1. Schematic representation of the NAD pathway model. Enzymes are subdivided into biosynthetic (green) and degrading enzymes (red). NAD key metabolites are circled, remaining substrates and products are indicated by reaction arrows. For further details and abbreviations see [16]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

NAD (nicotinamide adenine dinucleotide) is a key metabolite [15] as it participates in a vast number of redox reactions, but is also involved in a variety of signaling reactions. In contrast to redox reactions, these signaling reactions include degradation of NAD and require a permanent replenishment of cellular NAD pools. In yeast and human, NAD can be synthesized via a number of possible routes [15] which illustrates the fact that NAD metabolism can be considered as a relatively complex metabolic network (Fig. 1). From a phylogenetic perspective the occurrence of NAD biosynthetic enzymes is remarkably conserved across taxa [16], with only a few exceptions [17,18]. However the diversification of vertebrates species is associated with an diversification of NAD degrading enzymes [16]. In particular, mammals, including humans, have numerous NAD degrading enzymes. Because of the physiological importance of NAD signaling, the effect of the constant turnover of NAD on NAD biosynthesis has gained increased attention over the past years [19–21].

ADP-ribosyltransferases constitute the largest group of NAD degrading enzymes. They transfer the ADP-ribose moiety of NAD onto acceptor proteins coincident with the release of nicotinamide [22]. ADP-ribosyltranferases can be distinguished as either mono-ADP-ribosyltransferases (mARTs) or poly-ADP-ribosyltransferases (PARPs), even though this classification is inaccurate [22], since it is unclear if all PARPs truly involve a poly-ADP-ribose catalytic function. However, different PARPs have been experimentally proven to be involved in enzymatic processes such as DNA repair, maintenance of genomic stability and response to oxidative stress [23]. The PARP family consists of 17 family members, but the majority of poly-ADP-ribose is synthesized by PARP1 [24] and some PARP family members remain poorly characterized experimentally [22].

Even though the metabolic and biochemical properties of human NAD metabolic enzymes have been intensely studied, there

is currently a lack of a systematic evolutionary analysis of how natural selection has shaped sequence patterns of NAD metabolic enzymes. Because of the complexity of the NAD metabolic network (Fig. 1) it is difficult to make clear predictions on how evolutionary forces have shaped the sequence composition of NAD metabolic enzymes. Given the universality of NAD, its biosynthesis and the physiological importance of NAD degrading enzymes it seems plausible that mutations affecting NAD metabolic enzymes are likely to introduce harmful changes leading to enzymatic malfunction. However based on the distribution of NAD metabolic enzymes across taxa [16] it becomes apparent that vertebrates harbour NAD degrading enzymes that are absent in other species. These "new" enzymes are likely to be the result of duplication events and therefore may gain new or modified functions according to the patchwork model [25,26]. From an evolutionary perspective, phylogenetically young genes tend to evolve under relaxed selective constraint and possibly undergo adaptation more frequently [27]. We therefore postulate that among the genes encoding NAD degrading enzymes there are candidates that should show signatures of positive selection or the random accumulation of mutations as a part of subfunctionalization and therefore show a higher rate of protein evolution.

In this study we test the hypothesis that NAD biosynthetic and degrading enzymes show differences in their evolutionary patterns. We start our analysis on four model species, humans, mice, *Drosophila* and *Arabidopsis* and present evidence for differences in divergence and diversity between the two enzyme groups. We show that the direction of selective constraint is inconsistent between the species, which may reflect differences in their effective population sizes. Based on 50 human NAD metabolic enzymes, we subsequently calculate evolutionary rates and determine the role of positive selection for vertebrate enzymes involved in NAD metabolism. We particularly focus on two members of the PARP

family, PARP4 and PARP14. Human PARP4 is an enzyme with an unusually large number of high frequency segregating variation and shows a strong signature of population differentiation, as well as a possible signature of population specific positive selection. We reveal that the signature of population differentiation and positive selection coincides with the end of a homologous region which undergoes unequal recombination with an adjacent pseudogene. Secondly we provide strong evidence that PARP14 has been undergoing recurrent events of positive selection within the vertebrate clade. We identify potentially positively selected sites that lie in the functional domains of the PARP14 protein, but outside its catalytic domain, and are likely to have an impact on regulation and target substrate binding. We conclude that among NAD degrading enzymes there are fast evolving enzymes, illustrating an ongoing functional diversification in mammalian NAD metabolism.

2. Materials and methods

2.1. NAD pathway model

We focused on 50 human NAD metabolic enzymes based on a previous study and the availability of sequence data [16], a detailed list of gene identifiers can be found in Supporting dataset S1. We did not include human PARG in the study as we could not identify the gene in the HapMap dataset. We classify enzymes as either NAD biosynthetic or NAD degrading (Fig. 1) depending on whether they are part of a biosynthetic pathway or not [15].

2.2. Model species comparisons

We obtained sequence divergence and diversity of four model species: humans, *Arabidopsis thaliana*, *Drosophila melanogaster* and mice [16]. Sequence information was retrieved as follows: For humans we obtained information of the human genome project [28], coding SNPs for each gene were extracted from > 1000 human individuals and were resampled to 20 individuals using binomial sampling which is a reasonable approximation due to the large sample size. We also use this information to identify SNPs that segregate at high frequency (> 20%) within the human population. Divergence was obtained using the genome of macaque [29] to avoid misinference by segregating polymorphisms [30]. For *Drosophila* we used sequencing information from 17 sub-Saharan individuals [31] and outgroup information to *D. yakuba* [32], for *Arabidopsis* we used sequence information from 80 *Arabidopsis thaliana* accessions [33] and outgroup information to *A. lyrata* [34] and sequence information from ten wild house mice (*Mus musculus castaneus*) individuals [35] using rat as outgroup [36]. We focus on protein coding parts and obtain the number of nonsynonymous polymorphisms (P_n) and synonymous polymorphims (P_s) as well as the number of nonsynoymous substitutions (D_n) and synonymous substitution (D_s) for each gene. Because some genes are absent in a species or have little or no polymorphism or substitutions we sum data across genes for the two enzyme groups, NAD biosynthetic or NAD degrading. Estimates for D_n and D_s were calculated using codeml package of PAML (runmode = −2) or in case of mouse-rat comparison provided from the authors [35] as the number of fourfold and zerofold divergent sites. Significance for the contrast of P_n/P_s and D_n/D_s between enzyme groups was addressed by permutation tests. A classic McDonald Kreitman (MK) test [5] was conducted by contrasting patterns of polymorphisms and divergence. Within the MK framework an excess of nonsynonymous divergence to nonsynonymous polymorphisms suggests the action of adaptive evolution.
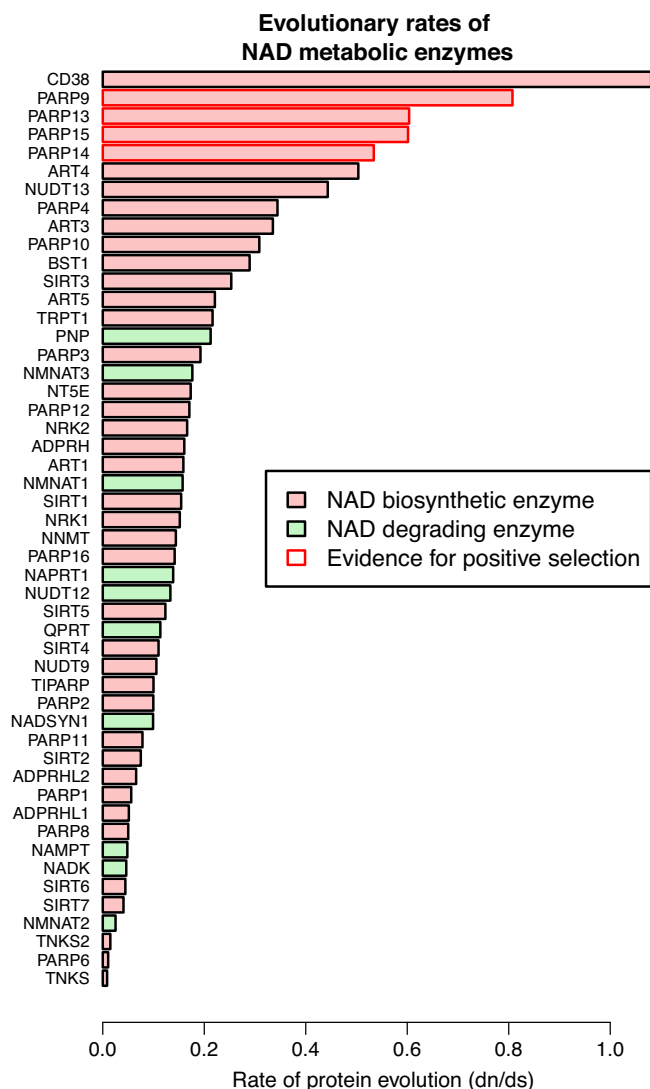
Fig. 2. Evolutionary rates for 50 NAD metabolic enzymes measured as d_n/d_s based on human NAD enzymes and their orthologs in other vertebrate species. Genes that show significant evidence for positive selection based on site test (see Section 2) are boxed in red. Parameter estimates for the site test results are shown in Table 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.3. Rates of molecular evolution of vertebrate NAD metabolic enzymes

We identified orthologous sequences for the 50 human enzymes using whole genome information from 22 vertebrate species which we obtained from Ensembl [37] (full species list in Supporting dataset S1). To identify homologs we used the best BLAST hit [38] followed by a filtering procedure. We excluded sequences which were too long or too short (>20% in comparison to the human sequence) and for which too many positions did not have sequence information (>5 nt missing). Therefore the number of outgroup sequences varies for the 50 enzymes, either because a particular gene is absent or poorly sequenced. The remaining sequences were aligned with MAFFT [39] and unreliable alignment positions were removed with GUIDANCE [40]. The final alignment was processed with PAL2NAL [41] and used as input for PAML [42]. Synonymous mutations can be regarded as neutral or nearly neutral [43], and this characteristic may be used to infer selective effects of amino acid altering mutations [5,44–47]. The rate of molecular protein

evolution can be measured by the ratio $\omega = d_n/d_s$, where d_n and d_s are, respectively, the numbers of nonsynonymous and synonymous differences per nonsynonymous and synonymous site between two sequences [48–50]. PAML uses a maximum likelihood approach to obtain ω estimates for the provided phylogeny using certain model assumptions. To obtain ω for each set of orthologous enzymes we assumed a constant ω for the whole phylogeny (model M0, one-ratio model). We also conducted site tests which assumes heterogeneity in ω within genes. For this a likelihood ratio test (LRT) is conducted to address a significant fit of a more complex model in comparison to a nested simpler model. Evidence of positive selection is based on the comparison of the site model models M1a and M2a [51,47] by applying an LRT with degrees of freedom (d.f.) = 2 and $P < 0.001$ which corresponds to a P-value of $P = 0.05$ with Bonferroni correction for multiple testing. Models M1a and M2a assume two classes of sites, either purifying selection ($\omega < 1$) or neutrality ($\omega = 1$), but M2a additionally allows for sites with $\omega > 1$. To identify specific sites with $\omega > 1$ we used the Naive Emprical Bayes (NEB) procedure as implemented in PAML and selected sites with a posterior probability > 0.5 to have $\omega > 1$.

2.4. PARP4: Fst and H-statistic

To investigate patterns of polymorphism of the human PARP4 gene we included noncoding regions (e.g. UTR and introns). We calculated Fst [52], a measurement of population differentiation, using the R-package Popgenome [53]. Additionally, we calculated Fay and Wu's H statistic [54] for each of the human population separately using the software package DHtest (http://zeng-lab.group.shef.ac.uk). Fay and Wu's H measures an excess of high frequency derived variants in comparison to intermediate alleles, which should be present if recombination occurs during a selective sweep [55]. We obtained genomic information covering all exons and introns of human PARP4 on chromosome 13 spanning a region of 85 kb for 1092 individuals from 14 human population using the Hapmap data [28]. Test statistics (Fst and Wu's H) were calculated for overlapping sliding windows (3 Kb window size and step size 0.5 Kb). Confidence intervals for Fst values were obtained by calculating minimum and maximum Fst of 100 random subsamples of 546 individuals. PARP4 is also undergoing gene conversion from a nearby pseudogene (Dumont et al, personal communication), e.g. as a result of ectopic recombination [56]. The PARP4 pseudogene is located around 5 Mb before the PARP4 gene on chromosome 13 covering a region of 32 Kb and shares 28 SNPs with PARP4, of which two are coding (Supporting dataset S1)

3. Results

Here we conducted a comprehensive evolutionary analysis of NAD metabolic enzymes by using intraspecific variation and sequence divergence to other species. We restricted our analyses to the most common type of mutations, point mutations. In particular, we tested the hypothesis that NAD degrading enzymes show faster rates of protein evolution in comparison to NAD biosynthetic enzymes. First, we compared the evolutionary patterns of NAD biosynthetic and degrading enzymes in four model species. We used a general metabolic model derived from humans and yeast [15] and classified enzymes as either being NAD biosynthetic or NAD degrading (Fig. 1). We subsequently focused on human NAD metabolism, based on sequence variation and divergence of 50 human NAD metabolic enzymes to elucidate candidate enzymes with sequence patterns that are consistent with an evolutionary functional transition.

3.1. Evolutionary patterns of NAD metabolic enzymes vary across model species

As an initial analysis we compared intra- and interspecific protein coding sequence variation of NAD metabolic enzymes in four model species (human, mouse, *Drosophila* and *Arabidopsis*) for which data is publicly available. A measurement of the currently acting selective forces can be obtained by contrasting the ratio of nonsynonymous polymorphism to synonymous polymorphisms (P_n/P_s). Assuming that nonsynonymous mutations are largely deleterious or neutral, a relaxed action of purifying selection or an increased number of effectively neutral mutations would lead to higher P_n/P_s values. Indeed, we found that in *Drosophila melanogaster* P_n/P_s is significantly higher for enzymes which consume NAD than for biosynthetic enzymes (Table 1). The same trend is observed in the other three species, however not significant. One possible reason may be a lack of statistical power because the number of polymorphisms is relatively low and the number of protein sequences is limited by the design of the study.

The ratio of nonsynonymous substitutions to synonymous substitutions (D_n/D_s) is a measurement of the evolutionary rate of protein evolution. Relaxed purifying selection or frequent positive selection lead to an elevation of D_n/D_s, because nonsynonymous mutations tend to become fixed, i.e., replace the initial allele in the whole population, more frequently. Therefore it is important to disentangle the predominant type of selection, which however is not trivial [57]. Regarding NAD metabolic enzymes, we found that D_n/D_s is significantly different in three species – humans, mice and *Arabidopsis* (Table 1). In humans and *Arabidopsis* there is an increased rate of D_n/D_s for NAD degrading enzymes in comparison to NAD biosynthetic enzymes, but for mice the trend is the opposite. A possible explanation for this pattern are differences in the effective population size (N_e, reviewed in [58]) between these species, because higher levels of N_e facilitate the action of selection. Humans and *Arabidopsis* have relatively small effective population sizes [30] and show little evidence for adaptive evolution [6,59]. In contrast, N_e of wild mice is relatively large and as a consequence selection is more efficient [60]. Indeed, mice undergo frequent adaptive evolution [60,35] which leads to increased levels of D_n/D_s. To test if positive selection has contributed to the divergence of NAD metabolic enzymes we contrasted patterns of polymorphisms and divergence (MK-test, [5]). We found evidence for positive selection in mice and *Drosophila* (Table 1), both species with large N_e and in the case of *Drosophila* this is only significant for NAD biosynthetic enzymes, but not NAD degrading enzymes. Theses results suggest that differences in D_n/D_s in mice are largely caused by increased rates of adaptation, rather than relaxed levels of purifying selection. In contrast, for humans and *Arabidopsis* the differences between NAD degrading and NAD biosynthetic enzymes are likely caused by differences in the intensity of purifying selection. As a consequence, the role of positive selection in human NAD degrading enzymes appears limited, however it may be obscured by the dominating action of purifying selection.

3.2. Analysis of substitution rate variation mammalian species

To measure the rate of evolution we use nonsynonymous to synonymous substitution rates (ω) which provides a sensitive measure of selective pressure at the protein level. ω values <1, = 1, and >1 indicate purifying selection, neutral evolution, and positive selection, respectively. We obtained orthologous sequences from vertebrate species for 50 human NAD metabolic enzymes (see Section 2) and calculated evolutionary rates assuming a constant ω between species (one-ratio model). ω ratios vary greatly between enzymes (Fig. 2) from $d_n/d_s = 0.008$ for Tankyrase to $d_n/d_s = 1.08$ for CD38. The top 14 enzymes with the highest ω ratios are NAD
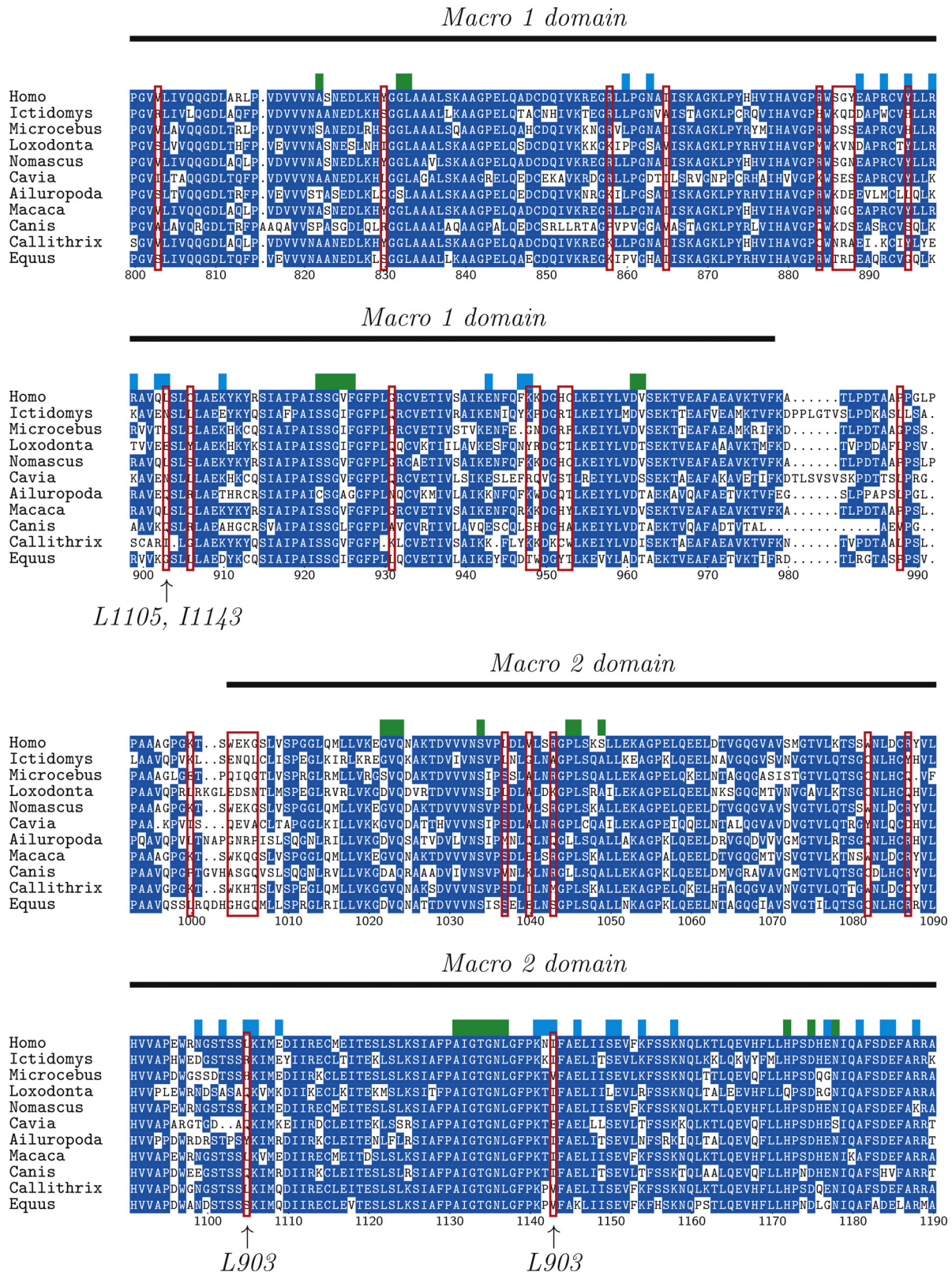
Fig. 3. Alignment of PARP14 homologs covering macro domain 1 and macro domain 2 (human residues 800–1200). Red boxes indicate sites that have been identified to be evolving under positive selection. Green blocks indicate sites for substrate binding and cyan blocks indicate macro domain 1 and 2 interacting sites [62]. Site 903 has nonpolar interactions with sites 1105 and 1143. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

degrading enzymes. Regarding the NAD biosynthetic enzymes PNP and NMNAT3 show the highest evolutionary rates, NADK and NMNAT2 show the lowest rates. Since variation in ω ratio can be the result of differences in the type of selection, we investigated the role of positive selection. To identify whether increased rates of evolution are caused by events of positive selection, we conducted site-specific tests with PAML and found evidence for positive selection in four enzymes of which all belong to the NAD degrading

Table 1

Diversity (P_n/P_s) and divergence (D_n/D_s) information for NAD biosynthetic and NAD degrading genes in four species. A test of positive selection (McDonald Kreitman (MK) test) was conducted, which contrasts the level of P_n/P_s and D_n/D_s.

Species		NAD biosynthesis		NAD degradation
Homo sapiens				
	P_n/P_s	0.28	≈	0.45
	D_n/D_s	0.26	<	0.39
	MK-test	n.s.		n.s.
Mus musculus				
	P_n/P_s	0.10	≈	0.12
	D_n/D_s	0.32	>	0.26
	MK-test	**		**
Drosophila melanogaster				
	P_n/P_s	0.03	<	0.06
	D_n/D_s	0.06	≈	0.06
	MK-test	*		n.s.
Arabidopsis thaliana				
	P_n/P_s	0.21	≈	0.24
	D_n/D_s	0.12	<	0.16
	MK-test	n.s.		n.s.

n.s. – not significant; <, > signs indicate significant differences between the two enzyme groups using a permutation test; ≈ indicates no significant difference.

  * $P < 0.05$.
  ** $P < 0.001$.

Table 3

Amino acid changing mutations of human NAD metabolic enzymes currently segregating at high frequency (> 20%) in the human population. Evolutionary rare changes based on negative PAM250 scores are in bold.

Enzyme	Amino acid change (position)	Frequency
NRK	V ↔ M (358)	0.31
NADK	N ↔ K (262)	0.28
NT5E	T ↔ A (376)	0.76
PNP	G ↔ S (51)	0.22
ART1	**L ↔ P (257)**	0.55
ART3	**S ↔ L (341)**	0.47
ART4	D ↔ N (265)	0.28
PARP13	H ↔ Q (565)	0.26
	R ↔ K (485)	0.25
PARP15	**G ↔ R (628)**	0.23
PARP4	A ↔ P (1656)	0.42
	P ↔ T (1328)	0.38
	G ↔ R (1280)	0.38
	G ↔ A (1265)	0.38
	V ↔ A (1065)	0.64
	A ↔ T (899)	0.66
	S ↔ N (873)	0.23
PARP9	Y ↔ C (528)	0.54
PARP10	V ↔ A (630)	0.54
	L ↔ P (395)	0.35
	I ↔ V (249)	0.38

enzymes (Fig. 2 and Table 2). Interestingly, all four enzymes are members of the PARP family (see 1), and in particular, PARP9, PARP14 and PARP15 form the group of macro domain containing PARPs. The macrodomain was initially described in histones and is involved in epigenetic regulation as a metabolite sensor on chromatin [61]. Recently a three dimensional structure of PARP14 has been published with focus on its macrodomains including binding sites and interactions between macro domains 1 and 2 [62]. We pinpoint 27 sites that show a signature of positive selection and lie in the macro domains 1 and 2 (Fig. 3), which is a significant enrichment compared to the rest of the protein (Fisher exact test, $P = 0.016$).

3.3. Amino acid altering variation segregating in human populations

We retrieved natural variation of human individuals from the Hapmap Genome project [28] to identify amino acid altering single nucleotide polymorphisms (SNPs) in NAD metabolic enzymes that are currently segregating in the human population at high frequencies (> 20%, Table 3). High frequency SNPs are likely to be polymorphisms which are either selectively neutral, under balancing selection or currently undergoing positive selection, but unlikely to have strong deleterious effects. Altogether we determined 21 high frequency amino-acid changing variants of which 4 belong to NAD biosynthetic enzymes (Table 3). The remaining 17 changes occur in ARTs and PARPs. Among the 21

Table 2

Human NAD metabolic enzymes with evidence for positive selection based on substitution rate analysis using vertebrate orthologs. Likelihood ratio test *P*-values are derived from the comparison of model M1a (nearly neutral model) and M2a (neutral+positive selection) to test for positive selection assuming hetereogeous ω among sites. p denotes the proportion of sites with $w > 1$, the number of orthologous sequences is denoted by # Seqs.

Enzyme	LRT *P*-value[a]	# Seqs	Parameter estimates
PARP9	0.0	7	$p = 0.13, \omega = 4.33$
PARP13	3.13×10^{-13}	3	$p = 0.05, \omega = 15.05$
PARP14	0.0	11	$p = 0.06, \omega = 3.41$
PARP15	3.46×10^{-7}	5	$p = 0.07, \omega = 4.65$

  [a] Site models M1a vs M2a.

high frequency mutations are 5 that are evolutionarily uncommon based on the PAM250 substitution score matrix [63] and they all occur in NAD degrading enzymes. Most strikingly we find that PARP4 (also dubbed vault-PARP [64]) has numerous high frequency nonsynonymous polymorphisms. A possible cause for this pattern is that PARP4 is currently undergoing a selective sweep [55] resulting in co-segregating SNPs. Another explanation could result from the fact that PARP4 has experienced gene conversion from a pseudogeneized paralog (Dumont et al., personal communication). Pseudogenes are genes that lost their protein coding ability and are therefore dysfunctional [65]. However, there are numerous polymorphisms at homologous positions that are segregating in both PARP4 and the PARP4 pseudogene likely caused by recombination events of the two different loci (ectopic recombination [56]). However, the region of non-homologous recombination does not entirely cover the PARP4 gene, but a stretch of around 38 Kb that includes the last 10 of the 33 PARP4 exons (e.g., the region encoding the C-terminus of the PARP4 protein, Fig. 4). To investigate these hypotheses in greater detail we retrieved haplotype information for the entire PARP4 gene along with the genomic alignment to chimpanzee. Based on a sliding window approach (see Section 2) we elucidated two remarkable patterns (Fig. 4). First there is a peak in the population structure, as measured by Fst, which is near the end of the nonhomologous recombining region. Second, we observe a signature of positive selection based on Fay and Wu's *H* statistic, which however appears to differ among the 14 populations. These population specific extreme values of *H* are also observed near the end of recombination region. Interestingly, we do not find an excess of diversity or divergence in the recombining region in comparison to the rest of the enzyme (results not shown).

4. Discussion

Taking into consideration the centrality and universality of NAD and the fact that there is no known mechanism of extracellular uptake of NAD, it is clear that there is the need for cellular NAD synthesis during cell growth and a constant recycling of NAD to satisfy the needs of NAD dependent reactions. The importance of NAD metabolism is also reflected in evolutionary terms, because of the widespread abundance of NAD biosynthetic enzymes in almost all genomes that have been sequenced so far [66,16]. However, within one organism multiple enzyme pathways may
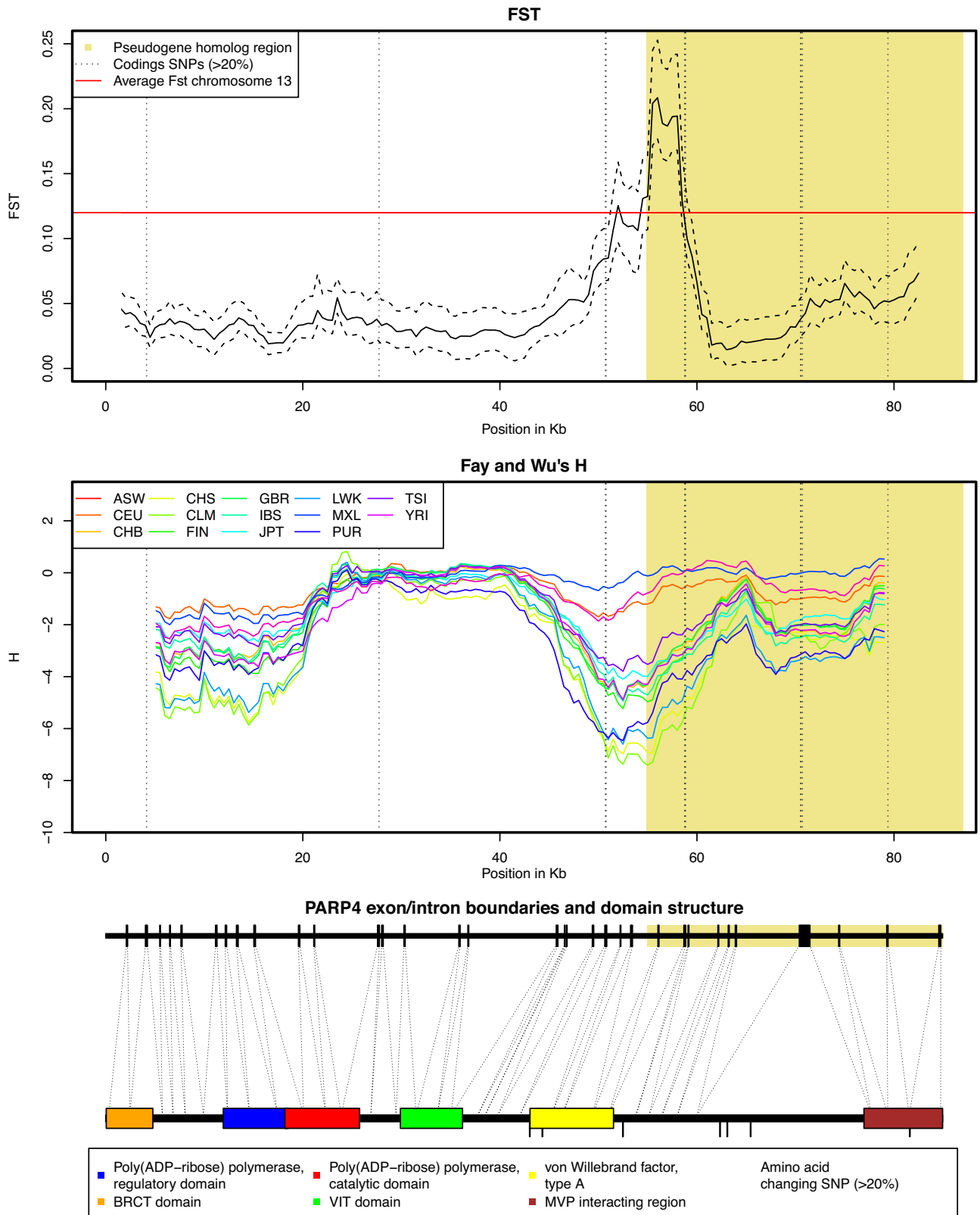
Fig. 4. Sliding window analysis (3 Kb overlapping windows for FST and 10 Kb for H, step size 0.5 Kb) of human PARP4 variation. The upper panel shows FST, a measurement of population differentiation along with confidence intervals (dashed black line, obtained from extreme values of 100 subsamples) and average FST of human chromosome 13 [82]. The middle panel shows Fay and Wu's H statistic, a test statistic for positive selection in presence of frequent recombination. The lower panel indicates the PARP4 exon/intron distribution along with the proposed protein domains (InterPro Accession Q9UKK3 [80]) and MVP interacting region [64,81]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

biosynthesize NAD, and the relative contribution of the different pathways to the overall cellular NAD content is still unknown [15]. Attempts to model pathways taking compartmentalization or even interactions between different tissues into account have been restricted to simple model systems as a consequence of lack of data due to experimental limitations [67–69]. Flux control analysis can predict the importance of an enzyme within a metabolic network [70]. However, at the current state there is clearly the need for a detailed analysis to link the outcome of flux distributions with an evolutionary interpretation [71].

A major predictor of an enzyme's evolutionary rate is its functional importance, but there are other factors which can influence patterns of sequence evolution of proteins, both within and between species [72]. This includes the local recombination rate, the linkage to nearby variation, epistatic effects and the effective population size (N_e) [73,30]. It is therefore difficult to derive clear predictions about how NAD metabolic enzymes are currently evolving and which evolutionary factors may have driven sequence divergence of NAD metabolic enzymes between vertebrate species. As a general assumption, we hypothesized that NAD biosynthetic enzymes should show a pattern of stronger selective constraint in contrast to NAD degrading enzymes which should reveal signatures of relaxed purifying selection and potentially show signs of functional diversification. In agreement with our prediction P_n/P_s, the number of nonsynonymous polymorphisms over synonymous polymorphisms, is consistently higher for NAD degrading enzymes in four model species. This illustrates that selection against deleterious mutations is less efficient or that more mutations tend to be effectively neutrally evolving in NAD degrading enzymes. The latter scenario is possible as a consequence of duplication events, under a model of subfunctionalization. Under such a scenario duplicated enzymes gain a specialized function and will therefore tend to evolve neutrally in the parts of the protein that do not exhibit a function anymore [26,74]. NAD degrading enzymes, such as PARPs, SIRTUINs and ARTs form protein families that are likely to have originated from duplication events. We also observe significant differences in the evolutionary rates between the two enzyme classes in three of the four model species. However, they are inconsistent among species which is possibly a consequence of the differences in their N_e. We showed that biosynthetic enzymes in humans and *Arabidopsis* exhibit reduced evolutionary rates, but appear to be evolving faster in mice, which may be attributed to the action of positive selection for which we find evidence in mice and *Drosophila*.

Based on substitution rate analysis of human metabolic enzymes along with their orthologs in other vertebrate species, we observe that the highest rates of protein evolution occur in NAD degrading enzymes, but it also becomes apparent that enzymes involved in biosynthesis are not consistently slowly evolving enzymes (Fig. 2). Even more surprisingly, there appears to be a substantial variation of evolutionary rate within both enzyme groups, NAD degrading and NAD biosynthesizing enzymes. One of the reasons for this heterogeneity may be caused by differences in the N_e of the different vertebrate species, as the number of species included differs for each of the enzymes due to technical (e.g. missing sequence information) or biological reasons (e.g. absence of an enzyme). Nevertheless, a cause for the observed heterogeneity in d_n/d_s is the action of positive selection. We find evidence for four enzymes where positive selection has driven sequence divergence within the vertebrate lineage. Another aspect which explains the substantial variation is the unlikely assumption of a constant evolutionary rate within the whole phylogeny, as the rate variation may be influenced by few species. A large scale study of >16,000 human genes found around 2.4% of genes to be evolving under positive selection [75] in six mammalian genomes (human, chimp, macaque, mouse, rat and dog), which is significantly lower than 8% (binomial test,

$P = 0.03$) identified in this study. This suggests two aspects, (1) NAD metabolism is enriched in genes undergoing positive selection and (2) positive selection may be in particular, acting in species that have not been included in the study of Kosiol et al. [75]. We find no evidence for positive selection based on branch specific test, as conducted previously [75], within the human lineage (results not shown).

Specific enzyme characteristics may cause evolutionary rate variation within proteins. There is striking evidence for evolutionary rate variation within vertebrate NMNATs [76], arguably the most central enzyme of NAD biosynthesis. The rate variation is caused by ISTIDs (isoform-specific targeting and interaction domains), which are interior stretches of the protein with an unknown three dimensional conformation. This potential conformational relaxation leads to elevated rates of protein evolution [76]. Indeed, NMNAT1 and NMNAT3 are among the fastest evolving NAD biosynthetic enzymes (Fig. 2), even though for two of the three ISTIDs a functional role has been experimentally verified [76]. However, protein families such as NMNATs and PARPs are not generally fast-evolving, e.g. NMNAT2 and PARP family members (Tankyrases and PARP6) are among the slowest evolving proteins. Among the slowest evolving enzymes is also NAMPT, the rate-limiting enzyme of NAD biosynthesis [77,78], which was believed to be only present in vertebrates until very recently. However there is accumulating evidence, that this enzyme is present in lower eukaryotes and has been lost several independent times in various non-vertebrate lineages [16,18]. However, enzymatic functionality of non-vertebrate NAMPT has yet to be shown.

Three macro-domain containing PARPs are among the enzymes with the highest evolutionary rates, and show a signature of positive selection. To identify potential functional roles of sites under positive selection we used a recently published three dimensional structure of PARP14 which includes binding sites as well as interaction sites of domains within the protein [62]. There are several sites subject to positive selection adjacent to binding sites, but more importantly, five macro domain interacting sites appear to have undergone positive selection (Fig. 3). Interestingly, three of the five sites form two nonpolar interactions between macro domain 1 and macro domain 2 [62], suggesting a potential coevolutionary mechanism. This suggest that not the NAD binding sites but the binding sites for the substrate on which the NAD moiety is transferred appears to be under evolutionary diversification. Taken together, the substitution rate analysis and the striking example of PARP14 shows a potential role of positive selection and illustrates that NAD degrading enzymes are under functional diversification. This also implies that experimentally proven functionality, e.g. in mouse, may not apply to other species, such as humans, and that a functional transfer from model species may be misleading.

Sequence variation in human NAD metabolic enzymes reveals that most amino acid changing SNPs are at very low frequencies suggesting strong deleterious effects of most of them. An exception is human PARP4 which shows an excess of high-frequency nonsynonymous polymorphisms. Based on a sliding window approach we observed a signal of population differentiation associated with a population specific negative value of Fay and Wu's H in PARP4, which may be interpreted as a signature of positive selection. Previous large scale analysis of human variation have not identified PARP4 as a target of selection [79] (http://haplotter.uchicago.edu/). One of the reasons may be that earlier approaches have been limited to the Hapmap2 dataset which was restricted to three human populations, while we use an extended dataset with 14 different populations. Interestingly there is evidence that PARP4 undergoes nonhomologous recombination with a nearby pseudogene (Dumont et al., personal communication) derived from common variation of PARP4 and the pseudogene. This pattern can, however,

also be found in the chimpanzee genome and therefore may reflect a distant evolutionary event rather than ongoing nonhomologous recombination events. The signature of population differentiation coincides with the end of the nonhomologous recombination region, however it is difficult to address the biological significance. The PARP4 homologous region that shows shared variation of the PARP4 pseudogene covers the last 10 exons (≈ 700 aa), which is not assigned to any domain function according to the InterPro database [80] but the major vault protein (MVP) interacting region [64,81].

5. Conclusion

Collectively, our results confirmed our general hypothesis that NAD biosynthetic enzymes are evolutionary more conserved while NAD degrading enzymes show less conservation and potentially are still undergoing functional diversification. However, we also showed that evolutionary patterns may differ between species and that there is a substantial evolutionary heterogeneity even within enzyme families (e.g. NMNATs, PARPs). We also stress that exceptional characteristics of particular enzymes may have an evolutionary impact, e.g. macro domains as targets for positive selection, the role of nonhomologous recombination or evolutionary heterogeneity within enzymes caused by ISTIDs. Taken together our findings provide a valuable example of how molecular evolution and population genetics can shed additional light on the functional characterization of enzymes and the networks they are involved in.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep. 2014.07.005.

References

[1] C. Pál, B. Papp, M.J. Lercher, An integrated view of protein evolution, Nat. Rev. Genet. 7 (5) (2006) 337–348, http://dx.doi.org/10.1038/nrg1838.

[2] A. Eyre-Walker, P.D. Keightley, The distribution of fitness effects of new mutations, Nat. Rev. Genet. 8 (8) (2007) 610–618, http://dx.doi.org/10.1038/nrg2146.

[3] E.R. Mardis, A decade's perspective on DNA sequencing technology, Nature 470 (7333) (2011) 198–203, http://dx.doi.org/10.1038/nature09796.

[4] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, et al., Comparison of next-generation sequencing systems, J. Biomed. Biotechnol. 2012 (2012) 251364, http://dx.doi.org/10.1155/2012/251364.

[5] J.H. McDonald, M. Kreitman, Adaptive protein evolution at the Adh locus in Drosophila, Nature 351 (6328) (1991) 652–654, http://dx.doi.org/10.1038/351652a0.

[6] A.R. Boyko, S.H. Williamson, A.R. Indap, J.D. Degenhardt, R.D. Hernandez, K.E. Lohmueller, et al., Assessing the evolutionary impact of amino acid mutations in the human genome, PLOS Genet. 4 (5) (2008) e1000083, http://dx.doi.org/10.1371/journal.pgen.1000083.

[7] A. Wagner, Metabolic networks and their evolution, Adv. Exp. Med. Biol. 751 (2012) 29–52, http://dx.doi.org/10.1007/978-1-4614-3567-9_2.

[8] S.C.G. Rison, J.M. Thornton, Pathway evolution, structurally speaking, Curr. Opin. Struct. Biol. 12 (3) (2002) 374–382.

[9] M. Ycas, On earlier states of the biochemical system, J. Theor. Biol. 44 (1) (1974) 145–160.

[10] R.A. Jensen, Enzyme recruitment in evolution of new function, Annu. Rev. Microbiol. 30 (1976) 409–425, http://dx.doi.org/10.1146/annurev.mi.30.100176.002205.

[11] M.W. Hahn, G.C. Conant, A. Wagner, Molecular evolution in large genetic networks: does connectivity equal constraint? J. Mol. Evol. 58 (2) (2004) 203–211, http://dx.doi.org/10.1007/s00239-003-2544-0.

[12] M.W. Hahn, A.D. Kern, Comparative genomics of centrality and essentiality in three eukaryotic protein–interaction networks, Mol. Biol. Evol. 22 (4) (2005) 803–806, http://dx.doi.org/10.1093/molbev/msi072.

[13] K.M. Wright, M.D. Rausher, The evolution of control and distribution of adaptive mutations in a metabolic pathway, Genetics 184 (2) (2010) 483–502, http://dx.doi.org/10.1534/genetics.109.110411.

[14] M.D. Rausher, The evolution of genes in branched metabolic pathways, Evolution 67 (1) (2013) 34–48, http://dx.doi.org/10.1111/j.1558-5646.2012.01771.x.

[15] L.F. de Figueiredo, T.I. Gossmann, M. Ziegler, S. Schuster, Pathway analysis of NAD+ metabolism, Biochem. J. 439 (2) (2011) 341–348, http://dx.doi.org/10.1042/BJ20110320.

[16] T.I. Gossmann, M. Ziegler, P. Puntervoll, L.F. de Figueiredo, S. Schuster, I. Heiland, NAD(+) biosynthesis and salvage-a phylogenetic perspective, FEBS J. 279 (18) (2012) 3355–3363, http://dx.doi.org/10.1111/j.1742-4658.2012.08559.x.

[17] A. Rongvaux, F. Andris, F.V. Gool, O. Leo, Reconstructing eukaryotic NAD metabolism, Bioessays 25 (7) (2003) 683–690, http://dx.doi.org/10.1002/bies.10297.

[18] J. Carneiro, S. Duarte-Pereira, L. Azevedo, L.F.C. Castro, P. Aguiar, I.S. Moreira, et al., The evolutionary portrait of metazoan NAD salvage, PLOS ONE 8 (5) (2013) e64674, http://dx.doi.org/10.1371/journal.pone.0064674.

[19] A. Nikiforov, C. Dölle, M. Niere, M. Ziegler, Pathways and subcellular compartmentation of NAD biosynthesis in human cells: from entry of extracellular precursors to mitochondrial NAD generation, J. Biol. Chem. 286 (24) (2011) 21767–21778, http://dx.doi.org/10.1074/jbc.M110.213298.

[20] A. Chiarugi, C. Dölle, R. Felici, M. Ziegler, The NAD metabolome – a key determinant of cancer cell biology, Nat. Rev. Cancer 12 (11) (2012) 741–752, http://dx.doi.org/10.1038/nrc3340.

[21] C. Dölle, R.H. Skoge, M.R. Vanlinden, M. Ziegler, NAD biosynthesis in humans – enzymes, metabolites and therapeutic aspects, Curr. Top. Med. Chem. 13 (23) (2013) 2907–2917.

[22] M.O. Hottiger, P.O. Hassa, B. Lüscher, H. Schüler, F. Koch-Nolte, Toward a unified nomenclature for mammalian ADP-ribosyltransferases, Trends Biochem. Sci. 35 (4) (2010) 208–219, http://dx.doi.org/10.1016/j.tibs.2009.12.003.

[23] S. Vyas, M. Chesarone-Cataldo, T. Todorova, Y.H. Huang, P. Chang, A systematic analysis of the PARP protein family identifies new functions critical for cell physiology, Nat. Commun. 4 (2013) 2240, http://dx.doi.org/10.1038/ncomms3240.

[24] M.L. Meyer-Ficca, R.G. Meyer, E.L. Jacobson, M.K. Jacobson, Poly(ADP-ribose) polymerases: managing genome stability, Int. J. Biochem. Cell Biol. 37 (5) (2005) 920–926, http://dx.doi.org/10.1016/j.biocel.2004.09.011.

[25] A. Force, M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, J. Postlethwait, Preservation of duplicate genes by complementary, degenerative mutations, Genetics 151 (4) (1999) 1531–1545.

[26] M. Lynch, A. Force, The probability of duplicate gene preservation by subfunctionalization, Genetics 154 (1) (2000) 459–473.

[27] J.J. Cai, D.A. Petrov, Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes, Genome Biol. Evol. 2 (2010) 393–409, http://dx.doi.org/10.1093/gbe/evq019.

[28] International HapMap 3 Consortium, D.M. Altshuler, R.A. Gibbs, L. Peltonen, D.M. Altshuler, R.A. Gibbs, et al., Integrating common and rare genetic variation in diverse human populations, Nature 467 (7311) (2010) 52–58, http://dx.doi.org/10.1038/nature09298.

[29] Rhesus Macaque Genome Sequencing and Analysis Consortium Human Genome Sequencing Center BCoMHTUa, R.A. Gibbs, J. Rogers, M.G. Katze, R. Bumgarner, G.M. Weinstock, et al., Evolutionary and biomedical insights from the rhesus macaque genome, Science 316 (5822) (2007) 222–234, http://dx.doi.org/10.1126/science.1139247.

[30] T.I. Gossmann, P.D. Keightley, A. Eyre-Walker, The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes, Genome Biol. Evol. 4 (5) (2012) 658–667, http://dx.doi.org/10.1093/gbe/evs027.

[31] J.E. Pool, R.B. Corbett-Detig, R.P. Sugino, K.A. Stevens, C.M. Cardeno, M.W. Crepeau, et al., Population genomics of sub-saharan Drosophila melanogaster: African diversity and non-African admixture, PLOS Genet. 8 (12) (2012) e1003080, http://dx.doi.org/10.1371/journal.pgen.1003080.

[32] A.G. Clark, M.B. Eisen, D.R. Smith, C.M. Bergman, B. Oliver, T.A. Markow, et al., Evolution of genes and genomes on the Drosophila phylogeny, Nature 450 (7167) (2007) 203–218, http://dx.doi.org/10.1038/nature06341.

[33] J. Cao, K. Schneeberger, S. Ossowski, T. Günther, S. Bender, J. Fitz, et al., Whole-genome sequencing of multiple Arabidopsis thaliana populations, Nat. Genet. 43 (10) (2011) 956–963, http://dx.doi.org/10.1038/ng.911.

[34] T.T. Hu, P. Pattyn, E.G. Bakker, J. Cao, J.F. Cheng, R.M. Clark, et al., The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change, Nat. Genet. 43 (5) (2011) 476–481, http://dx.doi.org/10.1038/ng.807.

[35] D.L. Halligan, A. Kousathanas, R.W. Ness, B. Harr, T.M. Keane, et al., Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents, PLOS Genet. 9 (12) (2013) e1003995, http://dx.doi.org/10.1371/journal.pgen.1003995.

[36] R.A. Gibbs, G.M. Weinstock, M.L. Metzker, D.M. Muzny, E.J. Sodergren, S. Scherer, et al., Genome sequence of the Brown Norway rat yields insights into mammalian evolution, Nature 428 (6982) (2004) 493–521.

[37] P. Flicek, I. Ahmed, M.R. Amode, D. Barrell, K. Beal, S. Brent, et al., Ensembl 2013, Nucleic Acids Res. 41 (D1) (2012) D48–D55, http://dx.doi.org/10.1093/nar/gks1236.

[38] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

[39] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (4) (2013) 772–780, http://dx.doi.org/10.1093/molbev/mst010.

[40] O. Penn, E. Privman, G. Landan, D. Graur, T. Pupko, An alignment confidence score capturing robustness to guide tree uncertainty, Mol. Biol. Evol. 27 (8) (2010) 1759–1767, http://dx.doi.org/10.1093/molbev/msq066.

[41] M. Suyama, D. Torrents, P. Bork, PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments, Nucleic Acids Res. 34 (Web Server issue) (2006) W609–W612, http://dx.doi.org/10.1093/nar/gkl315.

[42] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (8) (2007) 1586–1591, http://dx.doi.org/10.1093/molbev/msm088.

[43] J.L. King, T.H. Jukes, Non-Darwinian evolution, Science 164 (3881) (1969) 788–798.

[44] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (5) (1986) 418–426.

[45] A.L. Hughes, M. Nei, Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection, Nature 335 (6186) (1988) 167–170, http://dx.doi.org/10.1038/335167a0.

[46] N. Goldman, Z. Yang, A codon-based model of nucleotide substitution for protein-coding DNA sequences, Mol. Biol. Evol. 11 (5) (1994) 725–736.

[47] Z. Yang, R. Nielsen, N. Goldman, A.M. Pedersen, Codon-substitution models for heterogeneous selection pressure at amino acid sites, Genetics 155 (1) (2000) 431–449.

[48] W.H. Li, Molecular Evolution, Sinauer Associates Incorporated, 1997.

[49] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford University Press, New York, 2000.

[50] Z. Yang, Computational Molecular Evolution, vol. 284, Oxford University Press, Oxford, 2006.

[51] Z. Yang, R. Nielsen, M. Hasegawa, Models of amino acid substitution and applications to mitochondrial protein evolution, Mol. Biol. Evol. 15 (12) (1998) 1600–1611.

[52] R.R. Hudson, M. Slatkin, W.P. Maddison, Estimation of levels of gene flow from DNA sequence data, Genetics 132 (2) (1992) 583–589.

[53] B. Pfeifer, U. Wittelsbuerger, PopGenome: A Swiss Army Knife for Population Genetic & Genomic Analysis, 2013 http://CRAN.R-project.org/package=PopGenome, r package version 1.2.5.

[54] J.C. Fay, C.I. Wu, Hitchhiking under positive Darwinian selection, Genetics 155 (3) (2000) 1405–1413.

[55] J.M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene, Genet. Res. 23 (1) (1974) 23–35.

[56] L. Duret, N. Galtier, Biased gene conversion and the evolution of mammalian genomic landscapes, Annu. Rev. Genomics Hum. Genet. 10 (2009) 285–311, http://dx.doi.org/10.1146/annurev-genom-082908-150001.

[57] A.D. Cutter, B.A. Payseur, Genomic signatures of selection at linked sites: unifying the disparity among species, Nat. Rev. Genet. 14 (4) (2013) 262–274, http://dx.doi.org/10.1038/nrg3425.

[58] B. Charlesworth, Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation, Nat. Rev. Genet. 10 (3) (2009) 195–205, http://dx.doi.org/10.1038/nrg2526.

[59] T.I. Gossmann, B.H. Song, A.J. Windsor, T. Mitchell-Olds, C.J. Dixon, M.V. Kapralov, et al., Genome wide analyses reveal little evidence for adaptive evolution in many plant species, Mol. Biol. Evol. 27 (8) (2010) 1822–1832, http://dx.doi.org/10.1093/molbev/msq079.

[60] D.L. Halligan, F. Oliver, A. Eyre-Walker, B. Harr, P.D. Keightley, Evidence for pervasive adaptive protein evolution in wild mice, PLOS Genet. 6 (1) (2010) e1000825, http://dx.doi.org/10.1371/journal.pgen.1000825.

[61] M. Posavec, G. Timinszky, M. Buschbeck, Macro domains as metabolite sensors on chromatin, Cell. Mol. Life Sci. 70 (9) (2013) 1509–1524, http://dx.doi.org/10.1007/s00018-013-1294-4.

[62] A.H. Forst, T. Karlberg, N. Herzog, A.G. Thorsell, A. Gross, K.L.H. Feijs, et al., Recognition of mono-ADP-ribosylated ARTD10 substrates by ARTD8 macrodomains, Structure 21 (3) (2013) 462–475, http://dx.doi.org/10.1016/j.str.2012.12.019.

[63] M. Dayhoff, R. Schwartz, B. Orcutt, A model of evolutionary change in proteins, Atlas Protein Seq. Struct. 5 (1978) 345–352.

[64] V.A. Kickhoefer, A.C. Siva, N.L. Kedersha, E.M. Inman, C. Ruland, M. Streuli, et al., The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase, J. Cell Biol. 146 (5) (1999) 917–928.

[65] E.F. Vanin, Processed pseudogenes: characteristics and evolution, Annu. Rev. Genet. 19 (1985) 253–272, http://dx.doi.org/10.1146/annurev.ge.19.120185.001345.

[66] F. Gazzaniga, R. Stebbins, S.Z. Chang, M.A. McPeek, C. Brenner, Microbial NAD metabolism: lessons from comparative genomics, Microbiol. Mol. Biol. Rev. 73 (3) (2009) 529–541, http://dx.doi.org/10.1128/MMBR.00042-08, Table of Contents.

[67] M.J. Herrgård, N. Swainston, P. Dobson, W.B. Dunn, K.Y. Arga, M. Arvas, et al., A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology, Nat. Biotechnol. 26 (10) (2008) 1155–1160, http://dx.doi.org/10.1038/nbt1492.

[68] N. Klitgord, D. Segré, The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles, Genome Inf. 22 (2010) 41–55.

[69] A.K. Stavrum, I. Heiland, S. Schuster, P. Puntervoll, M. Ziegler, Model of tryptophan metabolism, readily scalable using tissue-specific gene expression data, J. Biol. Chem. 288 (48) (2013) 34555–34566, http://dx.doi.org/10.1074/jbc.M113.474908.

[70] J.D. Orth, I. Thiele, B.Ø. Palsson, What is flux balance analysis? Nat. Biotechnol. 28 (3) (2010) 245–248, http://dx.doi.org/10.1038/nbt.1614.

[71] C. Kaleta, L.F. de Figueiredo, I. Heiland, S. Schuster, Special issue: integration of OMICs datasets into metabolic pathway analysis, Biosystems 105 (2) (2011) 107–108, http://dx.doi.org/10.1016/j.biosystems.2011.05.008.

[72] S.S. Choi, S. Hannenhalli, Three independent determinants of protein evolutionary rate, J. Mol. Evol. 76 (3) (2013) 98–111, http://dx.doi.org/10.1007/s00239-013-9543-6.

[73] T.I. Gossmann, M. Woolfit, A. Eyre-Walker, Quantifying the variation in the effective population size within a genome, Genetics 189 (4) (2011) 1389–1402, http://dx.doi.org/10.1534/genetics.111.132654.

[74] T.I. Gossmann, K.J. Schmid, Selection-driven divergence after gene duplication in *Arabidopsis thaliana*, J. Mol. Evol. 73 (3-4) (2011) 153–165, http://dx.doi.org/10.1007/s00239-011-9463-2.

[75] C. Kosiol, T. Vinar, R.R. da Fonseca, M.J. Hubisz, C.D. Bustamante, R. Nielsen, et al., Patterns of positive selection in six mammalian genomes, PLOS Genet. 4 (8) (2008) e1000144, http://dx.doi.org/10.1371/journal.pgen.1000144.

[76] C. Lau, C. Dölle, T.I. Gossmann, L. Agledal, M. Niere, M. Ziegler, Isoform-specific targeting and interaction domains in human nicotinamide mononucleotide adenylyltransferases, J. Biol. Chem. 285 (24) (2010) 18868–18876, http://dx.doi.org/10.1074/jbc.M110.107631.

[77] J.R. Revollo, A.A. Grimm, S. Ichiro Imai, The NAD biosynthesis pathway mediated by nicotinamide phosphoribosyltransferase regulates Sir2 activity in mammalian cells, J. Biol. Chem. 279 (49) (2004) 50754–50763, http://dx.doi.org/10.1074/jbc.M408388200.

[78] A. Rongvaux, M. Galli, S. Denanglaire, F.V. Gool, P.L. Dréze, C. Szpirer, et al., Nicotinamide phosphoribosyl transferase/pre-B cell colony-enhancing factor/visfatin is required for lymphocyte development and cellular resistance to genotoxic stress, J. Immunol. 181 (7) (2008) 4685–4695.

[79] B.F. Voight, S. Kudaravalli, X. Wen, J.K. Pritchard, A map of recent positive selection in the human genome, PLOS Biol. 4 (3) (2006) e72, http://dx.doi.org/10.1371/journal.pbio.0040072.

[80] S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, et al., InterPro in 2011: new developments in the family and domain prediction database, Nucleic Acids Res. 40 (Database issue) (2012) D306–D312, http://dx.doi.org/10.1093/nar/gkr948.

[81] A. van Zon, M.H. Mossink, M. Schoester, G.L. Scheffer, R.J. Scheper, P. Sonneveld, et al., Structural domains of vault proteins: a role for the coiled coil domain in vault assembly, Biochem. Biophys. Res. Commun. 291 (3) (2002) 535–541, http://dx.doi.org/10.1006/bbrc.2002.6472.

[82] B.S. Weir, L.R. Cardon, A.D. Anderson, D.M. Nielsen, W.G. Hill, Measures of human population structure show heterogeneity among genomic regions, Genome Res. 15 (11) (2005) 1468–1476, http://dx.doi.org/10.1101/gr.4398405.