Biochemical and Biophysical Research Communications 440 (2013) 68–75

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Episodic positive selection during the evolution of naphthalene dioxygenase to nitroarene dioxygenase

Arindam Dutta, Joydeep Chakraborty, Tapan K. Dutta *

Department of Microbiology, Bose Institute, Kolkata 700054, India

ARTICLE INFO

Article history:
Received 29 August 2013
Available online 13 September 2013

Keywords:
Aromatic ring-hydroxylating oxygenase
Naphthalene dioxygenase
Nitroarene dioxygenase
Adaptive evolution
Episodic positive selection

ABSTRACT

Using different maximum-likelihood models of adaptive evolution, signatures of natural selective pressure, operating across the naphthalene family of dioxygenases, were examined. A lineage- and branch-site specific combined analysis revealed that purifying selection pressure dominated the evolutionary history of the enzyme family. Specifically, episodic positive Darwinian selection pressure, affecting only a few sites in a subset of lineages, was found to be responsible for the evolution of nitroarene dioxygenases (NArDO) from naphthalene dioxygenase (NDO). Site-specific analysis confirmed the absence of diversifying selection pressure at any particular site. Different sets of positively selected residues, obtained from branch-site specific analysis, were detected for the evolution of each NArDO. They were mainly located around the active site, the catalytic pocket and their adjacent regions, when mapped onto the 3D structure of the α-subunit of NDO. The present analysis enriches the current understanding of adaptive evolution and also broadens the scope for rational alteration of substrate specificity of enzyme by directed evolution.

1. Introduction

Microorganisms, particularly bacteria, play a crucial role in global bio-geochemical cycles of aromatic compounds using an array of catabolic enzymes. Aromatic ring-hydroxylating oxygenases (RHOs) constitute one such superfamily of enzymes that participate in the oxidative metabolism of a wide variety of aromatic compounds of pharmaceutical, agricultural and environmental significance [1,2]. Though the members of this superfamily work in conjugation with either one or two electron transport proteins (*viz.* ferredoxins and reductases), the substrate recognition and the subsequent catalysis is actually conferred by the large subunit (α) of the $\alpha_n\beta_n$ or α_n type terminal oxygenases [3,4].

The enormous structural variation of aromatics in the environment is responsible for the evolution of various RHO enzyme systems in such a way that they can transform a wide range of such compounds [1]. Several classification schemes have been proposed for RHOs [5,6], the most recent being that of Chakraborty et al. (2012) [7] where the RHOs have been classified into five 'Classes' and eleven 'Types' on the basis of their evolutionary and functional behaviors, in relation to structural configuration of substrates and preferred oxygenation site(s). According to that classification, naphthalene dioxygenases (NDO), together with four types of nitroarene dioxygenases (NArDO) *viz.* nitrobenzene dioxygenase (NBDO), 2-nitrotoluene dioxygenase (2NTDO), 2-chloronitrobenzene dioxygenase (2clNBDO) and 2,4-dinitrotoluene dioxygenase (2,4DNTDO), constitute the naphthalene family of RHOs, which belong to A-III$\alpha\beta$ type of RHOs.

It may be presumed that NArDOs must have been an outcome of relatively recent evolution since nitroaromatic compounds have been introduced into the environment by human activities. The hypothesis that a direct selective pressure operated during the evolution of NArDOs from NDO was proposed on the basis of the observation that nitroarene degrading bacteria are commonly found only in the nitroarene-contaminated sites [8,9]. A strong evolutionary relationship has indeed been observed among the catabolic operon(s) of nitroarene and naphthalene degradation pathways with frequent occurrence of pseudogenes in nitroarene degradation operons often characterized with deletions and/or frameshift mutations [10,11]. Though the core 3D structure of the catalytic α-subunit of both NDO (PDB: 1NDO) and NBDO (PDB: 2BMO) are quite conserved, nevertheless, the regio- and enantioselectivity of the reactions catalyzed by these enzymes exhibit distinct variations [8] and are anticipated to be the manifestation of variations in amino acid residues within their substrate cleft [8].

A comparison of rate of nonsynonymous substitutions per nonsynonymous site (d_N) to the rate of synonymous substitution per synonymous site (d_S) is an effective approach to detect the nature of selective pressure ($d_N/d_S < 1$, $d_N/d_S = 1$, $d_N/d_S > 1$ represents

* Corresponding author. Address: Department of Microbiology, Bose Institute, P-1/12 C.I.T. Scheme VII M, Kolkata 700054, India. Fax: +91 33 2355 3886.
E-mail address: tapan@jcbose.ac.in (T.K. Dutta).

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.09.029

negative (purifying), neutral, positive selection, respectively) acting on protein-coding genes. Statistical phylogenetics offers various tools to measure the d_N/d_S ratio (ω) for a set of protein-coding nucleotide sequences along the lineages and/or of individual sites [12,13]. The aim of this study is to unravel the signature of natural selection with a view to test the hypothesis that NArDO evolved from NDO and to identify amino acid residues and regions in NDO/NArDO that have evolved under positive selection.

2. Materials and methods

2.1. Sequence retrieval and construction of dataset

The coding sequence of catalytic α-subunit of NBDO (GenBank: AF379638) was used as a query sequence in BLASTn [14] search using the default parameters and 34 full length homologous sequences (showing $\geqslant$99% query coverage and $\geqslant$75% identity) belonging to the naphthalene family of RHOs were retrieved. Corresponding protein sequences of these genes were also downloaded for analysis. Redundant sequences as well as those bearing internal stop codon were excluded from the downloaded set of sequences. Sequences (GenBank: JN655512, GQ184726) showing recombination signals based on recombination analysis using RDP3 (Recombination Detection Program Version3) [15] were discarded from the analysis. Moreover, absence of saturation of synonymous sites in our refined set of sequences was ensured following the methods described by Lynn et al. [16] where pairwise d_N and d_S values were calculated using the maximum-likelihood method implemented in codeml program in PAML4.5 package [13] followed by a correlation analysis using SPSS [17]. Two different datasets, one containing 9 sequences (small dataset) and another containing 21 sequences (large dataset), were finally built (Table S1) to examine the robustness of analysis against dataset size. Both small and large datasets contained 5 NArDOs, apart from 4 and 16 evolutionarily distinct NDOs, respectively.

2.2. Phylogenetic analyses

Nucleotide sequences of both the datasets were aligned and phylogenetic trees were constructed using neighbor-joining algorithm as implemented in MEGA5 [18]. Confidence values for the nodes were obtained by bootstrapping (100 replicates). The nucleotide sequences for each of the datasets were also aligned as codons using the same program. Codon alignments and trees generated for each of the datasets were used in the subsequent analyses.

2.3. Lineage-specific selection analysis

Lineage-specific selection analysis was performed individually for both datasets to detect variation in selection pressure across lineages and to identify lineages with elevated ω values indicating episodes of positive Darwinian selection. Two different approaches were utilized for this purpose *viz*. Genetic Algorithm (GA)-branch method [19] and maximum-likelihood method [12,13].

The GA-branch analysis was performed using the DATAMONKEY web server [20]. This method comes up with optimum branch partition pattern according to ω values but sheds no light on other evolutionary aspects, such as κ value (transition–transversion rate ratio), branch lengths (number of nucleotide substitutions per codon) and number of synonymous/non-synonymous substitutions along the lineages. To overcome these limitations, two different branch models (referred to here as GA-models), one for each dataset, were constructed by codeml program [13] following the identical branch partition pattern, as obtained from the best-fitting models using GA-branch method. During construction of the GA-models, branch labels were set to define various branch-rate classes. The GA-model, one- and free-ratio models were fitted by maximum-likelihood to each dataset followed by comparison of GA-model to one- and free-ratio model by likelihood ratio test (LRT) [21].

2.4. Site-specific selection analysis

In order to infer sites potentially under diversifying selection pressure during the evolution of NDO enzyme family, four site-specific likelihood models, namely M1a, M2a, M7 and M8, were fitted by maximum-likelihood [22,23] and the likelihood values under model M1a and M7 were compared with M2a and M8 respectively by LRT. M1a and M7 are the null models in which positive selection is not allowed while M2a and M8 are the alternative models that allow positive selection. Comparisons of M1a and M7 with M2a and M8 respectively are performed to assess gene evolution under positive selection, which is ensured by the acceptance of alternative models.

2.5. Branch-site specific selection analysis

Both datasets were considered for this analysis to detect positive selection affecting sites along specific lineage(s) (designated as foreground branch) with ω value greater than one as found in the lineage specific analysis. This was performed by fitting modified Branch-site model A and its corresponding null model and subsequent comparison of their likelihood values using LRT. In this study, branch-site analysis was performed in combination with site-stripping method [24] to stringently identify sites, contributing to the significance of the modified branch-site model A over its null model, among the positively selected sites with Bayesian posterior probabilities (*P*-value) >0.95. Before performing this method, the positively selected sites identified by branch-site model A were arranged in descending order of their *P*-values and then the sites were removed sequentially from the alignment starting from the site for which highest *P*-value was obtained. Each resulting stripped alignment was reanalyzed using the same models and parameters. This process was continued until the LRT failed. Sites, whose removal from the alignment resulted in failure of the LRT, were taken to be the sites making the alternate model significant.

3. Results and discussion

3.1. Lineage-specific selection analysis

3.1.1. GA-branch analysis

GA-branch analyses [19] allowed us to identify best-fitting models that led to the classification of lineages within phylogenetic tree into four and six branch-rate classes, for small and large dataset respectively, suggesting the presence of heterogeneity in the ω value among lineages (Fig. 1). Two branch-rate classes, each for small and large datasets, exhibited $\omega > 1$, indicating that some of the lineages have experienced positive selection (Fig. 1). The lineages exhibiting high percentage support in favor of positive selection are summarized in Table S2.

3.1.2. Maximum-likelihood analysis

For both the datasets, ω values displayed by the branch-rate classes were compared with those exhibited by the corresponding GA-models, constructed using the codeml program [13] (Table S3). The values were found to be quite similar indicating that the best-fitting models and their corresponding GA-models are equivalent

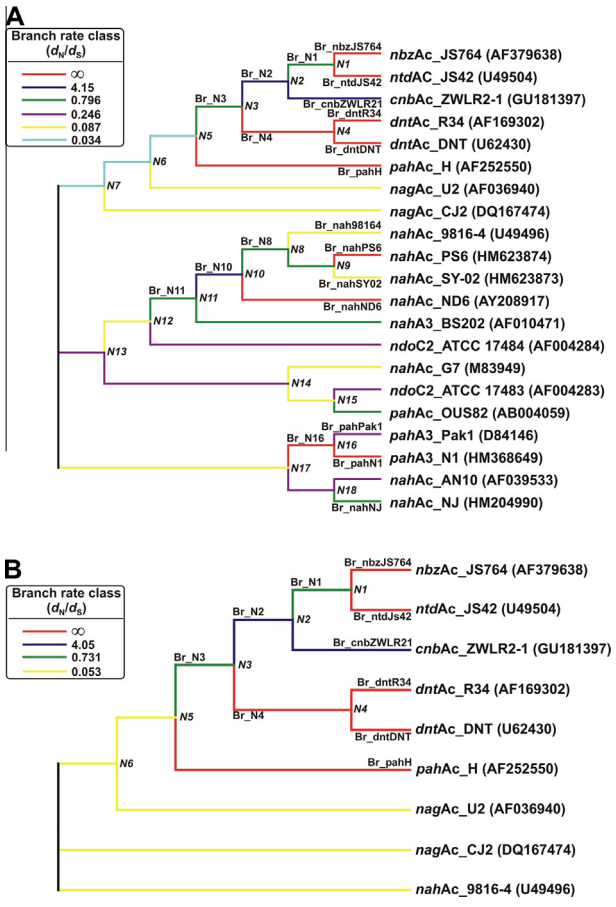

Fig. 1. Dendogram showing branch partitioning pattern based on d_N/d_S value obtained for large (a) and small (b) dataset in GA-branch analysis. Lineages with different d_N/d_S values have been indicated by color codes, shown in boxes. Extant sequences are represented by gene name followed by corresponding strain designation and NCBI accession number (within parentheses). *Nbz*Ac_JS764 (AF379638), *ntd*Ac_JS42 (U49504), *cnb*Ac_ZWLR2-1 (GU181397), *dnt*Ac_R34 (AF169302), *dnt*Ac_DNT (U62430) refer to sequence of NBDO, 2NTDO, 2clNBDO, 2,4DNTDO from *B. cepacia* R34 and 2,4DNTDO from *Burkholderia* sp. DNT respectively. Lineages with $\omega > 1$ are labeled. Each node is represented as N_x (x = 18 & 6 for large and small dataset respectively). (For interpretation of the reference to color in this figure legend, the reader is referred to the web version of this article.)

to each other. Comparison of lineage-specific likelihood models revealed that the GA-model was significantly better than the one-ratio model, but not significantly different from the free-ratio model (Table 1).

The estimates of the ω values under the one-ratio model for both datasets were less than 1 (0.17 for large and 0.16 for small dataset) indicating that, on average, synonymous substitutions were more frequent than nonsynonymous substitutions and that purifying selection pressure dominated the evolutionary history of NDO family. The maximum-likelihood estimates of ω values along the lineages demonstrated that the same set of lineages within the NArDO clade (Br_nbzJS764, Br_ntdJS42, Br_dntR34, Br_dntDNT, Br_cnbZWLR21, Br_N2 and Br_N4) exhibited $\omega > 1$ under both the GA and free-ratio models across both datasets (Table 2 and Fig. 1). Among them, the ω values along five lineages (Br_nbzJS764, Br_ntdJS42, Br_dntR34, Br_dntDNT and Br_N4) were estimated to be infinite (Table 2 and Fig. 1), indicating the absence of synonymous substitution along these lineages. The likelihood estimates of ω value for the other two lineages *viz.* Br_CnbZWLR21, Br_N2, were found to fall within the range of 3.41–3.78 and 3.59–4.00, respectively (Table 2 and Fig. 1), indicating the presence of both synonymous and non-synonymous substitutions along these lineages with the dominance of non-synonymous rate over the synonymous rate. In this respect, it is worth mentioning that some of the lineages (Br_pahH, Br_pahN1, Br_N16, Br_nahND6, Br_nahPS6 and Br_N10, Br_nah98164, Br_nahNJ, Br_pahPak1, Br_nahSY02, Br_N8 and Br_N11) outside the NArDO clade showed $\omega > 1$ (Table 2 and Fig. 1). The maximum-likelihood estimate of the number of synonymous and non-synonymous substitutions along the lineages demonstrated that none of the substitutions took place along any of the six of the lineages, *viz.* Br_N10, Br_nah98164, Br_pahPak1, Br_nahSY02, Br_N8 and Br_N11 (Table 2 and Fig. 1). Hence, the infinite value of ω along these lineages was solely due to the absence of synonymous substitutions and had nothing to do with positive selection.

Thus, both the GA-branch and the maximum-likelihood analyses unanimously indicated that although purifying selection pressure prevailed throughout the evolution of naphthalene family of dioxygenases, but there were several instances of positive Darwinian selection, mostly along the lineages within the nitroarene clade (a total of 7 out of 8 lineages) (Fig. 1). This indicates that adaptive evolution has played a crucial role during the evolution of NArDOs from NDO. Since our analysis did not show any statistically significant difference between the GA- and free-ratio model, all the positively selected lineages where evidence of substitutions were found were further considered for branch-site analysis [25,26] to find out positively selected sites along these lineages.

Table 1
Comparison of the GA-models with one- and free-ratio models.

Dataset	Models	$2\Delta l$ [a]	df[b]	*P*-value
Small	GA-model vs. one-ratio model	305.46	3	P[c] $\ll 0.01$
	GA-model vs. free-ratio model	3.18	11	$0.99 > P > 0.95$
Large	GA-model vs. one-ratio model	374.8	5	P[c] $\gg 0.01$
	GA-model vs. free-ratio model	18.44	33	$0.99 > P > 0.95$

[a] Twice the log-likelihood difference between the nested models.
[b] Degrees of freedom.
[c] Statistically significant *P*-value.

3.2. Site-specific analysis

M2a and M8 models could not identify any statistically significant positively selected sites. Moreover, comparison between site-specific likelihood models (Table S4) revealed that models that allow positively selected sites were not significantly better than the neutral models except in one case, that of the large dataset, where M8 fitted better than the M7 model (Table S4). This observation suggested that none of the sites have experienced any diversifying selection pressure along all the lineages during evolution of the NDO family. This inference lead us to hypothesize that the observed positive selection along the lineages in the NArDO clade (Fig. 1), as obtained in the lineage-specific selection analysis, may have resulted from episodic selection pressure affecting only a few sites, along these lineages.

3.3. Branch-site specific analysis

A flowchart shown in Fig. 2 describes the sequential steps that were followed for each NArDO. Branch-site tests were performed twice for each dataset using two different trees (Figs. 1 and 3) and the lineages considered as foreground for each NArDO are shown in Table 3. The foreground branches, taken into consideration for branch-site test I using the tree shown in Fig. 1 for one NArDO, were within the evolutionary pathway of other NArDOs. Therefore it is logical to assume that among the positively selected sites, some may not have played any role in the evolution of the corresponding NArDO for which the branch-site test was executed. To screen out these false positives, branch-site test II was

Table 2
Maximum-likelihood estimates of various parameters along the lineages showing ω>1 under the lineage-specific models.

Dataset	Branch	ω		Branch length		Synonymous substitutions number		Non-synonymous substitutions number	
		GA-model	Free-ratio model	GA-model	Free-ratio model	GA-model	Free-ratio model	GA-model	Free-ratio model
Small	Br_nbzJS764[a]	∞	∞	0.04	0.04	0.0	0.0	18.9	18.9
	Br_ntdJS42[a]	∞	∞	0.01	0.01	0.0	0.0	5.8	5.8
	Br_cnbZWLR21[a]	3.59	3.41	0.03	0.03	1.1	1.2	13.6	13.6
	Br_N2[a]	3.59	3.77	0.03	0.03	1.1	1.0	12.8	12.8
	Br_dntR34[a]	∞	∞	0.01	0.01	0.0	0.0	3.5	3.5
	Br_dntDNT[a]	∞	∞	0.07	0.07	0.0	0.0	29.4	29.4
	Br_N4[a]	∞	∞	0.03	0.03	0.0	0.0	14.3	14.3
	Br_pahH	∞	∞	0.002	0.002	0.0	0.0	1.0	1.0
Large	Br_nbzJS764[a]	∞	∞	0.04	0.04	0.0	0.0	18.7	18.7
	Br_ntdJS42[a]	∞	∞	0.01	0.01	0.0	0.0	5.8	5.8
	Br_cnbZWLR21[a]	3.78	3.59	0.03	0.03	1.1	1.2	13.4	13.4
	Br_N2[a]	3.78	4.0	0.03	0.03	1.1	1.0	12.6	12.7
	Br_dntR34[a]	∞	∞	0.01	0.01	0.0	0.0	3.5	3.4
	Br_dntDNT[a]	∞	∞	0.07	0.07	0.0	0.0	29.1	29.1
	Br_N4[a]	∞	∞	0.03	0.03	0.0	0.0	14.1	14.1
	Br_pahH	∞	∞	0.00	0.00	0.0	0.0	1.0	1.0
	Br_pahN1	∞	∞	0.01	0.01	0.0	0.0	2.0	2.0
	Br_N16	∞	∞	0.02	0.02	0.0	0.0	6.9	7.0
	Br_nahND6	∞	∞	0.01	0.01	0.0	0.0	3.1	3.1
	Br_nahPS6	∞	∞	0.01	0.01	0.0	0.0	4.1	4.1
	Br_N10	3.78	2.5	0.0	0.0	0.0	0.0	0.0	0.0
	Br_nah98164[b]	0.09	2.58	0.0	0.0	0.0	0.0	0.0	0.0
	Br_nahNJ[b]	0.72	1.31	0.02	0.02	3.2	2.0	7.1	8.2
	Br_pahPak1[b]	0.23	2.58	0.0	0.0	0.0	0.0	0.0	0.0
	Br_nahSY02[b]	0.09	2.56	0.0	0.0	0.0	0.0	0.0	0.0
	Br_N8[b]	0.72	2.2	0.0	0.0	0.0	0.0	0.0	0.0
	Br_N11[b]	0.72	1.54	0.0	0.0	0.0	0.0	0.0	0.0

[a] Lineages within the NArDO clade.
[b] Lineages exhibiting $\omega > 1$ exclusively in free-ratio model.

performed using the trees shown in Fig. 3, consisting of any one of the NArDOs and the NDO sequences. As depicted in Fig. 2, positively selected sites obtained from these two branch-site tests were screened and a list of positively selected sites for each dataset was constructed. Sites common to both the lists (one for each dataset) were finally selected to identify the positively selected residues within each NArDO (Table 3). All the branch-site tests performed during this analysis favoured the modified branch-site model A over its null model, suggesting that episodic selection pressure operated along these lineages. The summary of the branch-site tests has been shown in Table 3.

The present analysis detected more than one positively selected residues for each nitroarene dioxygenase (Table 3). This result is consistent with a previous hypothesis that multiple amino acid substitutions are required to effectively orient nitroaromatic compound inside the active site pocket of naphthalene dioxygenase for efficient catalysis [27]. In addition, branch-site analysis yielded different sets of positively selected residues for different NArDOs, in agreement with their differential catalytic properties. Interestingly, a difference was observed in the number of positively selected residues in the DNTDO sequences from *Burkholderia cepacia* R34 and *Burkholderia* sp. DNT (Table 3). In addition, the branch lengths for the dioxygenase from strain R34 was also found to be longer than that of strain DNT (Table 2). Although both the genes have been annotated as dinitrotoluene dioxygenase, but the above observations suggest a plausible difference in their evolutionary rates, and their differential range of substrate preferences.

In the branch-site test, conducted individually for each of the six lineages (Br_pahH, Br_pahN1, Br_N16, Br_nahND6, Br_nahPS6 and Br_nahNJ) outside NArDO clade, it was observed that for two of the lineages (Br_nahNJ, Br_N16) the modified branch-site model A was significantly better than the corresponding null models, suggesting the presence of positively selected sites along the lineages (Table S6 and Fig. 1). Though the sequences corresponding to these lineages have been reported as NDOs, it might be possible that the observed positive selection have led to alteration of their substrate specificities and modification of activity towards other structurally homologous compounds.

3.4. Positively selected sites mapped onto the 3D structure

Residues in NDO, corresponding to the positively selected sites for each NArDO were mapped onto the 3D structure of the α-subunit of NDO (PDB: 1NDO, Chain: A). Residue-specific distances, between C_α atoms and the active site iron were also calculated (Table S7). In addition, ASA (accessible surface area) ratios of these residues were measured using the GETAREA program [28]. The latter analysis showed most of them (12 out of 17) to be buried (Table S7). The mapping showed that six residues (*viz.* Ala206, Pro198, Trp316, Phe352, Leu253, His295) corresponding to the structure of NDO α-subunit (PDB: 1NDO) were located within the substrate channel (Fig. S1)[3]. Among them, Ala206 is located above the active site iron; Pro198 and Trp316 are located below the iron atom while Phe352 lines the narrowest region of the channel (Fig. S1). Seven other residues were found to be located within a diameter of 15 Å from the active site iron while three more residues were within a range of 15–20 Å. These observations indicate that residues within the 20 Å diameter of the active site iron have largely undergone positive selection during the evolution of NArDOs from NDO. Changes in physio-chemical properties (size, hydrophobicity, charge, polarity and aromaticity) of all the sites under positive selection were studied (Table S7) and the observed differences likely played a major role in altering the activity of NDO towards nitroaromatic compounds. Residues present within the catalytic pocket are known to directly influence the substrate promiscuity, regiospecificity and enantioselectivity of the enzymes while those outside the catalytic pocket modulate the activity of the enzymes by indirect effects such as through subtle changes in the complex architecture of tertiary and/or quaternary structure [29,30]. Another residue, Ser102, detected to be positively selected,

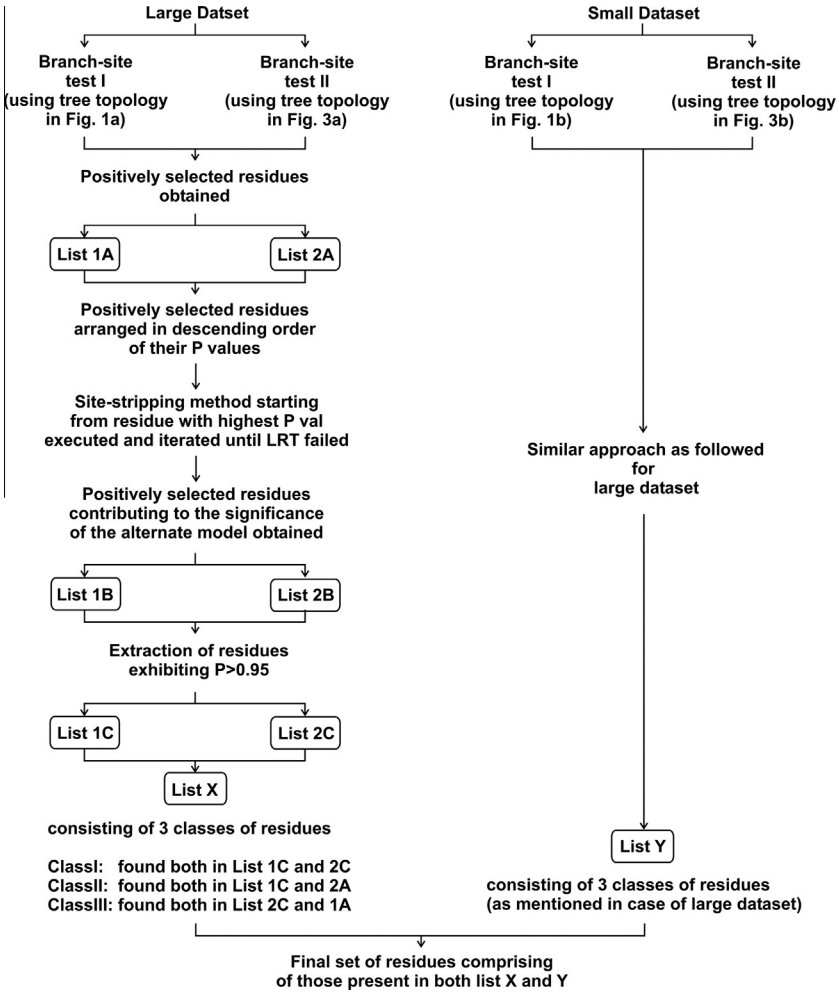

Fig. 2. Flowchart exhibiting the sequential steps carried out during branch-site analysis for each NArDO to find out the signature of episodic selection pressure affecting the subset of lineages within nitroarene clade during the evolution of NDO into NArDO and to identify the positively selected sites for each NArDO. Lists of positively selected residues (list 1A, 1B, 1C, 2A, 2B, 2C, X, Y) constructed in different steps are shown in Table S5.

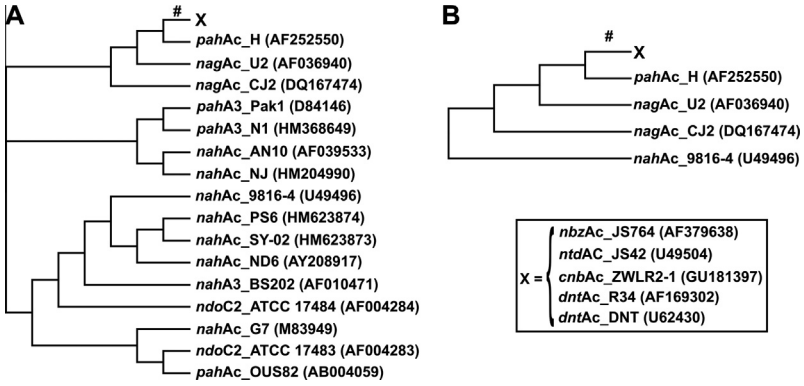

Fig. 3. Alternate tree topology ('a' for large dataset and 'b' for small dataset) used for branch-site test II. The branch selected as 'foreground' has been labeled as # and X represents the NArDO under consideration. Extant sequences are represented by gene name followed by corresponding strain designation and NCBI accession number (within parentheses).

belongs to the Rieske domain. The Rieske center of one α-subunit is structurally proximal to the active site of the neighbouring α-subunit. Therefore, Ser102 is likely to influence the enzyme activity of the neighbouring α-subunit. However, possibilities of distant residues exerting indirect effects on the catalytic domain of the same subunit cannot be ignored as has been previously observed in case of UDP-glucuronosyltransferase [31]. Friemann et al. [8] reported that although structures of NDO and NBDO α-subunits are similar, significant structural differences were observed along the loop (Leu253-Leu265 in 1NDO) located at the entrance of the substrate channel. Branch-site specific analysis for respective NArDO detected two residues (Leu253 and His261) under positive selection

Table 3
Summary of branch-site test of positive selection along the lineages within NArDO clade.

Dioxygenase	Model	l[a]		Foreground branches	$2\Delta l$[d]		$P_{(df=1)}$[e]		Positively selected sites[f,g]
		Large	Small		Large	Small	Large	Small	
NBDO	Modified branch-site Model A	−5390.11	−4023.95	(Br_nbzJS764, br_N1, br_N2, br_N3)[b]	62.8	54.36	≪0.01	≪0.01	Phe196(Pro198), Gly204(Ala206), Ala214(Ser216), Phe251(Leu253), Phe259(His261), Phe293(His295), Ala310(Val312), Met367(Ala369), Ser384(Asn386)
	Null model for modified branch-site Model A	−5421.51	−4051.13						
	Modified branch-site Model A	−4753.24	−3377.9	Ancestral branch of NBDO[c]	7.38	6.7	<0.01	<0.01	
	Null model for modified branch-site Model A	−4756.93	−3381.25						
2NTDO	Modified branch-site Model A	−5401.43	−4035.34	(Br_ntdJS42, br_N1, br_N2, br_N3)[b]	51.0	43.42	≪0.01	≪0.01	Gly100(Ser102), Phe196(Pro198), Ile204(Ala206), Ala214(Ser216), Thr251(Leu253), Phe259(His261), Ile293(His295), Thr310(Val312), Ser384(Asn386)
	Null model for modified branch-site Model A	−5426.93	−4057.05						
	Modified branch-site Model A	−4721.92	−3346.58	Ancestral branch of 2NTDO[c]	18.46	19	≪0.01	≪0.01	
	Null model for modified branch-site Model A	−4731.15	−3356.08						
2clNBDO	Modified branch-site Model A	−5422.47	−4051.53	(Br_CnbZWLR21, Br_N2, Br_N3)[b]	19.64	16.56	≪0.01	≪0.01	Asn100(Ser102), Phe196(Pro198), Ile204(Ala206),Ile214(Ser216), Ala406(Ile408)
	Null model for modified branch-site Model A	−5432.29	−4059.81						
	Modified branch-site Model A	−4683.67	−3307.38	Ancestral branch of 2clNBDO[c]	12.7	12.44	≪0.01	≪0.01	
	Null model for modified branch-site Model A	−4690.02	−3313.6						
2,4 DNTDO (*dnt*Ac_R34)	Modified branch-site Model A	−5420.83	−4049.43	(Br_DntR34, Br_N4, Br_N3)[b]	26.04	23.78	≪0.01	≪0.01	Leu196(Pro198), Ile214(Ser216), Phe314(Trp316),Asp384(Asn386)
	Null model for modified branch-site Model A	−5433.85	−4061.32						
	Modified branch-site Model A	−4654.12	−3278.99	Ancestral branch of 2,4 DNTDO[c]	8.56	8.44	<0.01	<0.01	
	Null model for modified branch-site Model A	−4658.40	−3283.21						
2,4 DNTDO (*dnt*Ac_ DNT)	Modified branch-site Model A	−5398.93	−4030.92	(Br_DntDNT, Br_N4, Br_N3)[b]	39.48	35.86	≪0.01	≪0.01	Val200(Ala197), Phe201(Pro198), Ile209(Ala206),Ile219(Ser216), Ala245(Met242), Gly312(Cys309), Phe319(Trp316), Thr355(Phe352), Val370(Thr368),Asp388(Asn386)
	Null model for modified branch-site Model A	−5418.67	−4048.85						

(*continued on next page*)

Table 3 (*continued*)

Dioxygenase	Model	l[a]		Foreground branches	$2\Delta l$[d]		$P_{(df=1)}$[e]		Positively selected sites[f,g]
		Large	Small		Large	Small	Large	Small	
	Modified branch-site Model A	−4703.63	−3350.77	Ancestral branch of 2,4 DNTDO[c]	21.0	23.92	≪0.01	≪0.01	
	Null model for modified branch-site Model A	−4714.13	−3362.73						

[a] Log-likelihood value.
[b] Shown in Fig. 1.
[c] Shown in Fig. 3.
[d] Twice the log-likelihood difference between the nested models.
[e] Probability value at 1 degree of freedom .
[f] Positively selected sites refer to the respective NArDO and their corresponding sites α-subunit of 1NDO are mentioned within parentheses.
[g] Positively selected sites for each NArDO were obtained through robust screening method, described in Fig. 2.

for NBDO and 2NTDO which were found to be located in the loop region. This finding suggests the possibility of the loop region to be under selection pressure.

In summary, maximum-likelihood estimation of lineage- and branch-site specific ω values obtained from the above study nullifies the neutral model of sequence evolution for naphthalene family of dioxygenase. The analysis reveals that episodes of diversifying evolution, affecting only couple of sites along small subset of lineages within the nitroarene clade, were instrumental in the evolution of NDO into NArDO. Although positively selected residues are scattered throughout the protein structure, mapping of the positively selected residues suggests that the active site, the substrate channel and its adjacent regions have mainly been targeted by selection pressure to broaden the enzyme specificity towards nitro-aromatic compounds. As a whole the study thus reveals that with the addition of new compounds into the environment, bacteria adapt to the milieu, occasionally by altering and/or broadening substrate specificities of relevant enzyme system(s) for selective advantage. Information on the key amino acid substitution events detected in the present work lays the foundation on rational approaches to design site-directed mutagenesis experiments in order to engineer catabolic activity of the RHO enzyme system toward target pollutants. This may be enriched further by molecular dynamics approach for wider applications in the field of biocatalysis and bioremediation.

Acknowledgments

This work was supported in part by a Grant-in aid from Department of Biotechnology, Government of India (#BT/PR11831/BID/07/290/2009 to T.K.D.), and Bose Institute, Kolkata, India. We thank Gautam Basu for editing the manuscript.

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.bbrc.2013.09.029.

References

[1] R.-H. Peng, A.-S. Xiong, Y. Xue, X.-Y. Fu, F. Gao, W. Zhao, Y.-S. Tian, Q.-H. Yao, Microbial biodegradation of polyaromatic hydrocarbons, FEMS Microbiol. Rev. 32 (2008) 927–955.
[2] S. Mallick, J. Chakraborty, T.K. Dutta, Role of oxygenases in guiding diverse metabolic pathways in the bacterial degradation of low-molecular weight polycyclic aromatic hydrocarbons: a review, Crit. Rev. Microbiol. 37 (2011) 64–90.
[3] B. Kauppi, K. Lee, E. Carredano, R.E. Parales, D.T. Gibson, H. Eklund, S. Ramaswamy, Structure of an aromatic-ring-hydroxylating dioxygenase-naphthalene 1,2-dioxygenase, Structure 6 (1998) 571–586.
[4] M.D. Wolfe, J.V. Parales, D.T. Gibson, J.D. Lipscomb, Single turnover chemistry and regulation of O_2 activation by the oxygenase component of naphthalene 1,2-dioxygenase, J. Biol. Chem. 276 (2001) 1945–1953.
[5] J.W. Nam, H. Nojiri, T. Yoshida, H. Habe, H. Yamane, T. Omori, New classification system for oxygenase components involved in ring-hydroxylating oxygenations, Biosci. Biotechnol. Biochem. 65 (2001) 254–263.
[6] O. Kweon, S.-J. Kim, S. Baek, J.-C. Chae, M.D. Adjei, D.-H. Baek, Y.-C. Kim, C.E. Cerniglia, A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases, BMC Biochem. 9 (2008) 11.
[7] J. Chakraborty, D. Ghosal, A. Dutta, T.K. Dutta, An insight into the origin and functional evolution of bacterial aromatic ring-hydroxylating oxygenases, J. Biomol. Struct. Dyn. 30 (2012) 419–436.
[8] R. Friemann, M.M. Ivkovic-Jensen, D.J. Lessner, C. Yu, D.T. Gibson, R.E. Parales, H. Eklund, S. Ramaswamy, Structural insight into the dioxygenation of nitroarene compounds: the crystal structure of nitrobenzene dioxygenase, J. Mol. Biol. 348 (2005) 1139–1151.
[9] J.C. Spain, Biodegradation of nitroaromatic compounds, Annu. Rev. Microbiol. 49 (1995) 523–555.
[10] R.M. Jones, B. Britt-Compton, P.A. Williams, The naphthalene catabolic (*nag*) genes of *Ralstonia* sp. strain U2 are an operon that is regulated by NagR, a LysR-type transcriptional regulator, J. Bacteriol. 185 (2003) 5847–5853.
[11] K.S. Ju, R.E. Parales, Nitroaromatic compounds, from synthesis to biodegradation, Microbiol. Mol. Biol. Rev. 74 (2010) 250–272.
[12] Z. Yang, J.P. Bielawski, Statistical methods for detecting molecular adaptation, Trends Ecol. Evol. 15 (2000) 496–503.
[13] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591.
[14] S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[15] D.P. Martin, P. Lemey, M. Lott, V. Moulton, D. Posada, P. Lefeuvre, RDP3: a flexible and fast computer program for analyzing recombination, Bioinformatics 26 (2010) 2462–2463.
[16] D.J. Lynn, A.T. Lloyd, M.A. Fares, C. O'Farrelly, Evidence of positively selected sites in mammalian alpha-defensins, Mol. Biol. Evol. 21 (2004) 819–827.
[17] B. Wellman, Doing It Ourselves: The SPSS Manual as Sociology's Most Influential Recent Book, in: Clawson (Ed.), Required Reading: Sociology's Most Influential Books, University of Massachusetts Press, Amherst, 1998, pp. 71–78.
[18] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, S. Kumar, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739.
[19] S.L.K. Pond, S.D.W. Frost, A genetic algorithm approach to detecting lineage-specific variation in selection pressure, Mol. Biol. Evol. 22 (2005) 478–485.
[20] S.L.K. Pond, S.D.W. Frost, Datamonkey: rapid detection of selective pressure on individual sites of codon alignments, Bioinformatics 21 (2005) 2531–2533.
[21] Z. Yang, Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution, Mol. Biol. Evol. 15 (1998) 568–573.
[22] R. Nielsen, Z. Yang, Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene, Genetics 148 (1998) 929–936.
[23] Z. Yang, R. Nielsen, N. Goldman, A.K. Pedersen, Codon-substitutions models for heterogenous selection pressure at amino acid sites, Genetics 155 (2000) 431–449.
[24] S.A.A. Travers, M.J. O'Connell, G.P. McCormack, J.O. McInerney, Evidence for heterogeneous selective pressures in the evolution of the *env* gene in different human immunodeficiency virus type 1 subtypes, J. Virol. 79 (2005) 1836–1841.
[25] Z. Yang, R. Nielsen, Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages, Mol. Biol. Evol. 19 (2002) 908–917.

[26] J. Zhang, R. Nielsen, Z. Yang, Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level, Mol. Biol. Evol. 22 (2005) 2472–2479.
[27] B.G. Keenan, T. Leungsakul, B.F. Smets, M. Mori, D.E. Henderson, T.K. Wood, Protein engineering of the archetypal nitroarene dioxygenase of *Ralstonia* sp. Strain U2 for activity on aminonitrotoluenes and dinitrotoluenes through alpha-subunit residues Leucine 225, Phenylalanine 350, and Glycine 407, J. Bacteriol. 187 (2005) 3302–3310.
[28] R. Fraczkiewicz, W. Barun, Exact and efficient analytical calculation of the accessible surface areas and their gradients macromolecules, J. Comp. Chem. 19 (1998) 319–333.
[29] K.L. Morley, R.J. Kazlauskas, Improving enzyme properties: when are closer mutations better?, Trends Biotechnol 23 (2005) 231–237.
[30] R.N. McLaughlin, F.R. Poelwijk, A. Raman, S.W. Gosal, R. Ranganathan, The spatial architechture of protein function and adaptation, Nature 491 (2012) 138–142.
[31] R. Fujiwara, M. Nakajima, H. Yamanaka, T. Yokoi, Key amino acid residues responsible for the differences in substrate specificity of human UDP-glucuronosyltransferase (UGT)1A9 and UGT1A8, Drug Metab. Dispos. 37 (2009) 41–46.