Protein Engineering, Design & Selection, 2015, vol. 28 no. 11, pp. 507–518
doi: 10.1093/protein/gzv038
Advance Access Publication Date: 13 August 2015
Original Article

Original Article

Assessing the prediction fidelity of ancestral reconstruction by a library approach

Hagit Bar-Rogovsky[1,4], Adi Stern[2,5], Osnat Penn[2,6], Iris Kobl[3], Tal Pupko[2], and Dan S. Tawfik[1,*]

[1]Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel, [2]Department of Cell Research and Immunology, George. S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel, and [3]Entelechon GmbH, Industriestraße 1, 93077 Bad Abbach, Germany

*To whom correspondence should be addressed. E-mail: tawfik@weizmann.ac.il
[4]Present address: Faculty of Medicine in the Galilee, Bar Ilan University, Safed 1311502, Israel
[5]Present address: Department of Molecular Microbiology and Biotechnology, George. S. Wise Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel
[6]Present address: Department of Genome Sciences, University of Washington, Seattle, WA, USA
Edited by Valerie Daggett

Received 16 July 2015; Revised 16 July 2015; Accepted 20 July 2015

Abstract

Ancestral reconstruction is a powerful tool for studying protein evolution as well as for protein design and engineering. However, in many positions alternative predictions with relatively high marginal probabilities exist, and thus the prediction comprises an ensemble of near-ancestor sequences that relate to the historical ancestor. The ancestral phenotype should therefore be explored for the entire ensemble, rather than for the sequence comprising the most probable amino acid at all positions [the most probable ancestor (mpa)]. To this end, we constructed libraries that sample ensembles of near-ancestor sequences. Specifically, we identified positions where alternatively predicted amino acids are likely to affect the ancestor's structure and/or function. Using the serum paraoxonases (PONs) enzyme family as a test case, we constructed libraries that combinatorially sample these alternatives. We next characterized these libraries, reflecting the vertebrate and mammalian PON ancestors. We found that the mpa of vertebrate PONs represented only one out of many different enzymatic phenotypes displayed by its ensemble. The mammalian ancestral library, however, exhibited a homogeneous phenotype that was well represented by the mpa. Our library design strategy that samples near-ancestor ensembles at potentially critical positions therefore provides a systematic way of examining the robustness of inferred ancestral phenotypes.

Key words: ancestral sequence reconstruction, inferred ancestor, predicted ancestor, serum paraoxonase

Introduction

Ancestor reconstruction, or resurrection (Benner, 2002), is an intriguing way of obtaining glimpses of long gone missing links. Phylogenetic analyses and ancestral reconstruction also appear to have a promising potential in protein engineering, in designing enzyme libraries (Goldsmith and Tawfik, 2013), and in providing highly stable proteins with broad specificity as starting points for further engineering (Risso *et al.*, 2013). The reconstruction does not necessarily address the ancestor's sequence, but primarily its function and structure (Dean and Thornton, 2007). This approach has therefore become instrumental in deciphering the mutational trajectories and mechanisms by which different protein families, each bearing a different function, diverged from one common ancestor (Adey *et al.*, 1994; Jermann *et al.*, 1995; Chandrasekharan *et al.*, 1996; Chang *et al.*, 2002;

Zhang and Rosenberg, 2002; Thornton *et al.*, 2003; Ugalde *et al.*, 2004; Thomson *et al.*, 2005; Kuang *et al.*, 2006; Gaucher *et al.*, 2008; Konno *et al.*, 2010; Cai *et al.*, 2011; Akanuma *et al.*, 2013; Bar-Rogovsky *et al.*, 2013; Harms *et al.*, 2013; Hobbs *et al.*, 2013; Ingles-Prieto *et al.*, 2013; Risso *et al.*, 2013; Williams *et al.*, 2013).

Based on sequence information and phylogenetic analysis, ancestral reconstruction reveals valuable information for protein design and engineering. Ancestor proteins tend to be more stable than their contemporary descents (Williams *et al.*, 2006). Individual back-to-ancestor mutations were also shown to increase stability (Watanabe *et al.*, 2006; Yamashiro *et al.*, 2010; Perez-Jimenez *et al.*, 2011; Sullivan *et al.*, 2012), and to boost a protein's evolvability (Bershtein *et al.*, 2008).

The reconstruction begins with the generation of a phylogenetic tree from the set of currently available sequences. Various statistical methods then calculate the probability that the set of currently existing sequences could have diverged from a given ancestral sequence. The number and diversity of sequences at hand, the assumed phylogenetic model and the topology of the tree, all affect the prediction (Maddison, 1995; Zhang and Nei, 1997; Chang *et al.*, 2005; Li *et al.*, 2008; Hanson-Smith *et al.*, 2010; Yang *et al.*, 2011; Ames *et al.*, 2013; Santander-Jimenez and Vega-Rodriguez, 2013). However, regardless of the statistical method used, the ancestral predictions result in alternative amino acids at most positions. The alternatives are ranked by their marginal probabilities (herein probabilities), and the highest probability is often significantly <1. What is often dubbed as 'the ancestor' is actually the most probable ancestor (mpa)—a sequence that comprises the amino acids predicted with the highest probability at all positions. However, in effect, the prediction is a 'cloud', or an ensemble, of sequences comprising near mpa probabilities. This work focuses on the exploration of ensembles of near-ancestor sequences by a gene library approach.

Ensembles of near-ancestor sequences

What comprises an ensemble of near-ancestor sequences? The sequences within this ensemble are identical at all positions predicted with a probability of 1, or practically, with a probability that is higher than a given threshold that comprises acceptably accurate predictions (C_{pred}, inhere). Each of the ensemble's sequences differs, however, at a subset of the ambiguously predicted positions. We define d' to be the average number of positions that differ between individual ensemble sequences and the mpa. The higher the number of ambiguously predicted positions, and the lower the probabilities for the most likely amino acid, the larger d' is. The larger d' is, the bigger the ensemble is, and the likelihood that the mpa represents the actual ancestor becomes smaller. It should also be noted that within near-ancestor ensembles, the mpa is extremely rare. The mpa's probability of occurrence corresponds to the multiplication of probabilities for the most probable amino acid at each ambiguous position. In N9's case, this probability is $\sim 10^{-6}$, and for the vertebrate paraoxonase ancestor, N6, it is $\sim 10^{-19}$.

Experimental exploration of prediction ambiguities

Not only is the mpa probability of occurrence extremely low, but also, ignoring certain alternatively predicted states might lead to two undesirable consequences: (i) the mpa sequence will be either misfolded or non-functional, but may become folded and/or functional when the alternative predictions are introduced at one or more positions. (ii) The ancestor's functional properties will differ when alternative predictions are introduced at certain positions.

In view of the above, several approaches have been described that attempt to maximize the chances of reconstructing a functional ancestor and minimize the ambiguity regarding the ancestral phenotype. These approaches explore alternatives in positions chosen based on structural and biochemical knowledge of the studied protein (Chandrasekharan *et al.*, 1996; Kuang *et al.*, 2006), or in ambiguously predicted positions defined by a certain threshold of marginal probability (Thomson *et al.*, 2005). Ambiguously predicted positions were also identified by comparing the ancestors inferred with different models and from phylogenetic trees generated from DNA, codons and amino acid sequences. Alternatives at these positions were explored by synthesizing several different ancestral sequences (Jermann *et al.*, 1995; Chang *et al.*, 2002, 2005; Thomson *et al.*, 2005; Gaucher *et al.*, 2008), or by small libraries that represented alternative ancestral states (Ugalde *et al.*, 2004). In contrast to these works, here, our aim was to systematically sample ensembles of near-ancestor sequences. We sought to examine how phenotypically diverse these ensembles might be, and consequently, how robust is the prediction of the ancestor's phenotype. Taking the devil's advocate's stand, a prediction is likely to be valid only if the ensemble proves to be phenotypically homogeneous—i.e. all sequences, including the mpa, encode proteins with essentially identical functional properties. In this case, the ancestral phenotype is likely to have been reconstructed with high fidelity, even though its exact genotype remains ambiguous. If the ensemble shows considerable heterogeneity, both the genotype and the phenotype remain obscure.

The experimental system

Our experimental system was the family of serum paraoxonases (PONs). PONs were first identified as mammalian organophosphate hydrolases, and later found to be lactonases (Teiber *et al.*, 2003; Draganov *et al.*, 2005; Khersonsky and Tawfik, 2005). Enzymes belonging to the three known mammalian PON families hydrolyze different lactones including lipophilic γ- and δ-lactones and *N*-(3-oxododecanoyl)-L-homoserine lactone (3OC12-HSL), a bacterial quorum sensing molecule. While examining the evolutionary history of PONs (Bar-Rogovsky *et al.*, 2013), we characterized the mammalian PON mpa"s (dubbed N9). Here, we describe the exploration of the ensembles of near-ancestor sequences of both the mammalian and vertebrate PON ancestors (N9 and N6, respectively). Libraries were designed to combinatorially explore the second alternatives at critical positions—positions that are ambiguously predicted and may affect the function and/or the stability of the predicted ancestors. By exploring only critical positions, the near-ancestor sequence ensembles that are too large to be sampled experimentally, could be examined by conventional medium-throughput assays. Our results show that there is a very high level of phenotypic variability of the vertebrate ancestor N6 and thus the validity of the reconstructed N6 ancestor is questionable. In contrast, our analyses support the evolutionary relevance of the mammalian PON ancestor (N9).

Materials and methods

Identification of PON sequences and tree reconstruction

The human PON1 was used as query for protein-alignment BLAST (blastp) with default settings. PON sequences were collected from the NCBI non-redundant protein sequence database (nr), as well as from available ESTs and recently sequenced insect, fungal, protozoa and hydra genomes. All sequences were examined for the presence of key, absolutely essential residues based on the known structure

and catalytic mechanism of PON1 (H115, H134, C43, C353, E53, D54, N168, D169, N224, D268, N269) (Harel *et al.*, 2004). Sequences in which these residues positions are substituted were discarded. To minimize potential biases in the alignment, the redundancy was eliminated by using the Cd-hit program (Li and Godzik, 2006), such that the remaining sequences shared 30–90% identity. Sequence alignment was performed using PRANK (Loytynoja and Goldman, 2008) with the F+ option, and also using MUSCLE (Edgar, 2004). The evolutionary models most appropriate for analyzing our PON sequences were determined by using the ProtTest analysis (Abascal *et al.*, 2005). The alignments were used to generate trees with the 'PhyML' program (Guindon and Gascuel, 2003) by using the LG substitution matrix (Le and Gascuel, 2008) combined with +I and +G parameters as selected by ProtTest. Bootstrap resample tests were performed 100 times to confirm reliability of the reconstructed trees, and the resulting bootstrap values for the corrected tree were calculated by the RaxML program (Stamatakis, 2006). Further analysis was conducted based on the PRANK derived tree (Fig. 1) as it was in a better accordance with the currently accepted species tree (http://tolweb.org/tree/).

To establish the PONs tree topology, the Drp35 (PDB: 2DG1) and gluconolactonase (PDB: 3DR2) families, which are distinct is sequence (≤20% sequence homology) but share PON's fold and catalytic calcium, were added as outgroups. Their sequences were structurally aligned with PON (1V04) using Mustang (Konagurthu *et al.*, 2006). Additional PON sequences were added to the structural profile by Clustalw2 (Larkin *et al.*, 2007) after several manual refinements mostly at gaps, the alignment was used for tree construction by 'PhyML' as described above. Gaps were assigned as insertions or deletions depending on the length of the sequences that comprise an outgroup to any given node, and although those positions that were

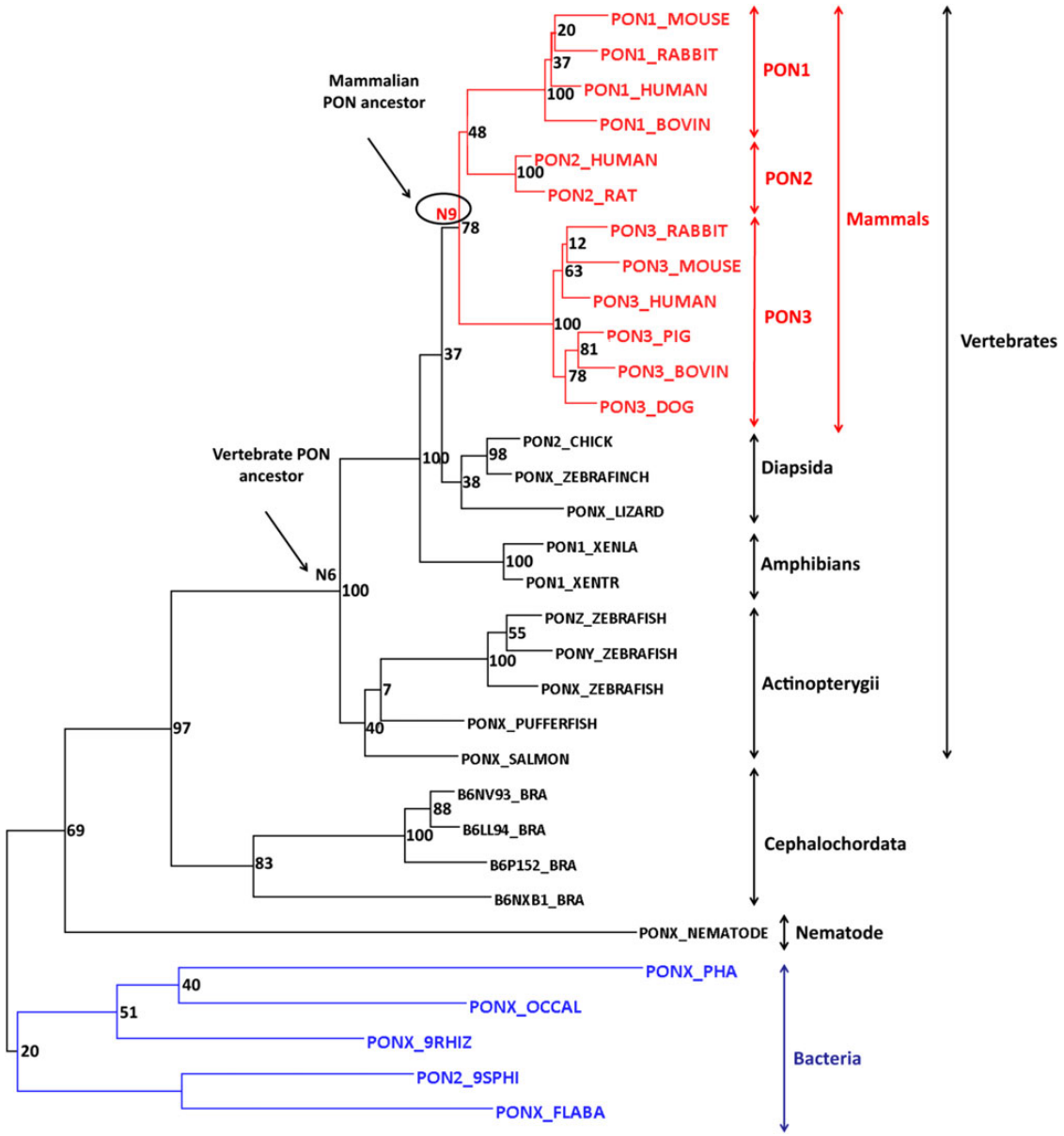

Fig. 1 The phylogenetic tree of PONs. Red branches mark the mammalian PONs. Bootstrap values (from 100 runs) are noted for each node.

evidently present in the N6 and/or N9 ancestor and deleted in some of the diverging lineages were included in mpa-N6/N9. The current version of FastML reconstructs InDels following the above consideration and using a probabilistic approach.

Inference of ancestral sequences and identification of critical positions

Ancestral sequences were inferred using the maximum likelihood methodology as implemented in FastML (Ashkenazy *et al.*, 2012). Using the alignment and the phylogenetic tree as input, FastML computes the posterior probability that a certain amino acid occupies a given position in the ancestral sequence, and also reconstructs ancestral indels. Positions in which alternatively predicted amino acids could affect the protein's structure and/or function were identified. As elaborated in the 'Results' section, their identification was based on the two main criteria: (i) sequence conservation, i.e. highly conserved positions that are under strong selection yet are predicted with low probability. (ii) Structural conservation, i.e. buried positions that are predicted with low probability, and on predefined threshold values. A simple computer program that performs this identification was developed and implemented on a webserver http://fastml.tau.ac.il/reconst/.

Construction of near-ancestor ensemble libraries

The N9 ancestral library was synthesized by Entelechon (http://www.entelechon.com/). This library comprised the gene encoding the mpa-N9 whereby, in codons encoding the identified critical positions, a mixture of nucleotides encoding the first and the second predicted amino acids was included at a ratio dictated by the marginal probability of the first prediction. To reduce the cost, and enable more flexibility in library making, the N6 library was generated by the oligo spiking method (Herman and Tawfik, 2007). By this method, the gene encoding the mpa-N6 was synthesized, as well as 14 mutagenesis oligonucleotides. These oligos encoded the amino acid predicted with the second highest posterior probability at one critical position with mpa-N6''s sequence flanking it (30 bp long. 15 bp from each side of the mutation). The oligo mix was composed of all the different oligos in amounts that corresponded to the marginal prediction of each diversified position in the library (i.e. the lower the marginal prediction for the first amino acid, the larger was the amount of the oligonucleotide encoding the second prediction). This mix was added to ~100 ng of DNaseI generated fragments of the mpa-N6 gene (~70–150 bp length) thus resulting in the combinatorial incorporation of the second predicted amino acids at their target sites. The library was retrieved by a nested PCR with external primers flanking the N6 gene, and the purified PCR product was digested and ligated into the expression vector. Both the N6 and N9 libraries were cloned into the pASK-TRX expression vector using NcoI and HindIII cloning sites. This pASK-TRX vector was constructed from pASK-IBA3plus vector (IBA) by replacing its ribosome binding site and the multiple cloning site (spanning from the XbaI and to the HindIII restriction sites) of the pASK-IBA3plus vector (IBA) with the ribosome binding site, the Trx-tag, and the multiple cloning site of the pET32b(+) vector (Invitrogen). Ligated DNA was transformed into *Escherichia coli* DH5α (Invitrogen) and plated on LB-Amp (LB medium supplemented with 100 mg/ml ampicillin) agar plates. The plasmid was extracted and used for transformation into *E. coli* Origami B (DE3) cells (Novagene) for expression and screening.

Library screening

Library transformations were grown on LB-Amp agar plates. Randomly picked clones were individually grown in 96 deep well plates at 37°C, overnight, with shaking, and were used to inoculate (1:50) 0.5 ml LB-Amp media containing 1 mM $CaCl_2$. The cells were grown at 30°C to OD^{600} of 0.8–1.0 and over-expression of the PON variants was induced by adding anhydrotetracycline (200 mg/l final). Cultures were grown at 20°C with shaking for ~16 h. Cells were spun down and subjected to two rounds of freezing at −20°C and thawing at 37°C. Thawed cells were lysed by shaking in lysis buffer (50 mM Tris pH 7.5, 1 mM $CaCl_2$, 50 mM NaCl, 1:500 protease inhibitor cocktail (PIC, Sigma), 5 U/ml benzonaze nuclease (Novagen), 1 mg/ml lysozyme, 0.1% Triton X-100) at 4°C for 6 h. Lysates were clarified by centrifugation for 10 min at 4000 rpm to remove the debris and 5 µl of clarified lysates were directly used for assaying enzymatic activities in 96 well plates, in a total volume of 200 µl activity buffer (50 mM Tris pH 8.0, 1 mM $CaCl_2$, 0.1% Triton X-100). In certain activity assays, undiluted lysates were used for activity measurements, whereas in other assays [for example for measuring the 4-thiobutyl-γ-butyrolactone (TBBL) and p-nitrophenylacetate (PNPA) activities] lysates were diluted in order to obtain measurable initial rates.

PAO-JP2 luminescence assay

The 3OC12-HSLase activity in crude lysates cannot be measured by the pH indicator assay (described in the 'Enzymatic assays' section). Therefore, the libraries were assayed by using the *Pseudomonas aeruginosa* PAO1-JP2 strain that harbors a reporter luciferase gene under the LasI promoter (Duan and Surette, 2007). This strain shows elevated luminescence levels that correspond to the concentration of 3OC12-HSL, but does not respond to the hydrolyzed lactone. PAO-JP2 cells were grown overnight at 37°C and used to inoculate (1:100) LB media plus Trimethoprim (300 µg/ml). From each library variant, 5 µl of clarified lysate (see above) was pre-incubated for 20 min in 50 µl activity buffer containing 200 nM 3OC12-HSL to allow its hydrolysis. The reaction mixtures were added to 450 µl of freshly inoculated PAO-JP2 cultures, and the cultures were shaken for 3 h at room temperature. Luminescence levels were measured in 460/40 nm to determine the concentration of intact (non-hydrolyzed) 3OC12-HSL in each sample. The latter were quantified by comparison to the luminescence values measured for a standard calibration curve generated from PAO1-JP2 cells incubated with different concentrations of 3OC12-HSL in the range of 0–20 nM. Relative specific activities were defined as: (20 nM − [intact 3OC12-HSL])/pre-incubation time/lysate volume.

Enzyme expression and purification

For expression of selected variants, single colonies of transformed *E. coli* Origami B (DE3) cells (Novagene) were grown in LB-Amp at 37°C overnight and were used to inoculate (1:50) LB-Amp media containing 1 mM $CaCl_2$. The cells were grown at 30°C to OD 0.8–1.0. Over-expression was induced by adding 200 mg/l anhydrotetracycline. Cultures were shaken at 20°C for ~16 h, harvested and subjected to two rounds of freezing at −20°C and thawing at 37°C. The thawed cells were lysed in lysis buffer [50 mM Tris pH 7.5, 1 mM $CaCl_2$, 50 mM NaCl, 1:500 PIC (Sigma), 5 U/ml benzonase nuclease (Novagen), 1 mg/ml lysozyme, 0.1% Triton X-100] at 4°C for 6 h. The clarified soluble fraction was bound to Ni-NTA agarose resin (Novagen). The resin was rinsed with IMAC buffer (50 mM Tris pH 7.5, 1 mM $CaCl_2$, 0.1% Triton X-100) containing 10, 20 and finally 40 mM imidazole. Elution was done with IMAC buffer containing 250 mM imidazole. The enzymatically active fractions were pooled and dialyzed against activity buffer (50 mM Tris pH 8.0, 1 mM

$CaCl_2$, 0.1% Triton X-100). Enzyme purity was assessed by SDS–PAGE and its concentration was measured with the BCA protein assay (Pierce). Enzyme preps were stored at 4°C.

Enzymatic assays

Enzymatic hydrolysis of TBBL and PNPA by library variants was determined in activity buffer as described (Draganov *et al.*, 2005; Khersonsky *et al.*, 2006). The final substrate concentration was 0.1 mM for TBBL and 1 mM for PNPA. The kinetics of δ- and γ-nonalactones and 3OC12-HSL were measured with pH indicator in BSCP buffer (1 mM bicine, 50 mM NaCl, 1 mM $CaCl_2$, 0.1 mM *m*-cresol purple, 0.1% Triton X-100) at 580 nm. The extinction coefficient (ε) was determined by titration with acetic acid. Stock solutions (100×) of all substrates were prepared in acetonitrile. Substrate concentrations varied according to their solubility in the reaction buffer and the detection limit of the products. The background hydrolysis with no enzyme was measured for each substrate concentration and was subtracted from the enzyme-catalyzed rates. All kinetic measurements were performed at room temperature and error ranges were derived from at least two independent measurements.

Results

Tree reconstruction

PON sequences were collected from the NCBI non-redundant protein sequence database (nr). While in mammals PON sequences were sufficiently represented, in vertebrates, invertebrates and other eukaryotes, they proved much more difficult to trace. Still, potential PONs were found in opossum (*Monodelphis domestica*), lizard (*Anolis carolinensis*), zebra finch (*Taeniopygia guttata*), platypus (*Ornithorhynchus anatinus*), as well as in several invertebrates such as sea urchin (*Strongylocentrotus purpuratus*), hydra (*Hydra magnipapillata*), *Drosophila melanogaster*, and in several fungi. Although all these sequences (except those from Drosophila, and certain fungal) included the essential residues for the catalytic mechanism described in the 'Materials and methods' section, they were excluded from the tree due to low sequence identity (<25%), or, as was the case with the fungal PONs, due to the fact that their addition to the tree resulted in their assignment as outgroup to all PONs regardless of their expected evolutionary position (Supplementary Fig. S1a and b). Despite being of much relevance, large insertions in the opossum PON (that may also reflect sequencing errors and mis-assignment of exon–inton junctions) and the lack of a fully sequenced PON in the platypus genome prevented their inclusion in the tree. The lizard and zebrafinch sequences exhibited high sequence identity to mammalian PONs (~55%) but were not included in the tree presented here since they appeared when this work was at an advanced stage.

The final set of selected PON sequences were aligned using PRANK (Loytynoja and Goldman, 2008) and MUSCLE (Edgar, 2004), and phylogenetic trees were constructed as described in 'Materials and methods' section. Both trees were in accordance with the currently accepted species tree. However, in the MUSCLE tree, the chicken and frog PONs (PON2_CHICK, PON1_XENTR and PON1_XENLA, respectively) were grouped within the mammalian PONs (Supplementary Fig. S2a). The PRANK tree, however, largely overlapped the species tree (Fig. 1).

Inference of N6 and N9

Ancestral inference was done by using FastML (Ashkenazy *et al.*, 2012). N9 and N6 were also inferred from the MUSCLE alignment, which was corrected according to the species tree. The predicted N9 sequences from both the PRANK and MUSCLE trees were identical. However, N6 predictions derived from the MUSCLE tree differed in 24 positions from the predicted ancestor of the PRANK tree, thus introducing another level of ambiguity (Supplementary Fig. S2b). Indel reconstruction was also done by Fastml based on a two-step approach as described in Ashkenazy *et al.* (2012). Nonetheless, in the case of the vertebrate ancestor N6, although deleted by FastML, the insertion at position 294–295 was kept in the reconstructed sequence since it appears in all the species preceding N6 (Supplementary Fig. S3). This short insertion, however, is situated far from PON's active site and is unlikely to affect the ancestor's enzymatic profile.

The inferred mpa-N9 and mpa-N6 (based on the PRANK alignment) gave soluble and enzymatically active proteins. Kinetic characterization of the mpa variants revealed that mpa-N9 is an efficient homoserinelactonase with a specific activity of 19.27 units/mg (Table I). This result correlated with our hypothesis that contemporary PONs original activity was homoserinelactonase. mpa-N6, however, exhibited much lower homoserinelactonase activity (0.53 units/mg) and higher aryl-esterase activity (25.8 vs. 2 units/mg for mpa-N9), which is a promiscuously hydrolyzed substrate of PONs (Table I).

The near-ancestor sequence ensembles

Positions with a posterior probability $C_{pred} > 0.9$ were considered as unambiguous. This threshold identified 39 positions in N9 where the ancestral prediction was considered ambiguous, resulting in a hypothetical number of 2^{39} different options (considering only the second predicted option) that comprise the near-ancestor ensemble. In the vertebrate N6, the overall prediction fidelity was lower, mainly due to the fact that apart from mammals, relatively few vertebrate sequences of PONs are available. This led to 105 ambiguous positions, and consequently, to 2^{105} different options. Clearly, both the N6 and N9 ensembles are impossible to screen. Thus, the distribution of all possible alternatively predicted ancestral sequences within the resulting ensemble was calculated, and analyzed as a function of their distance from the mpa (i.e. the number of positions in which the second prediction appears, *d*; Fig. 2). The distributions reveal that, on average, the N9 ensemble deviates by 10 positions from the mpa ($d' = 10$). The fraction of sequences with this average deviation is 0.153, and there are 4.73×10^8 such different sequences with various combinations of 10 alternatively predicted positions (39!/(39–10)!10!).

In the sequence ensemble of N6, the most frequent sequences differed in 33 positions from the mpa ($d' = 33$; fraction from all sequences is 0.09), thus presenting 2.3×10^{27} different sequences with 33 positions bearing alternative predictions. Assuming that one of the sequences in the ensemble is the real ancestor, and that all the sequences in the ensemble exhibit equal probabilities of representing the ancestor, the probability that the mpa is the real ancestor in the N9 ensemble is 6×10^{-6}, and 1.29×10^{-19} for N6.

The above computations suggest that each mpa represents only one choice out of many probable ones. Thus, to maximize the chances of obtaining a folded and functional sequence, and to assess the reliability of the predicted phenotype, the ensemble of near-ancestor sequences should be explored. However, even the N9 ensemble cannot be explored experimentally, as standard assay methodologies by which a wide range of proteins and functions can be screened have limited throughput (in the order of 1000 variants). In most ambiguous positions, however, the alternatively predicted amino acid will have a negligible effect on

Table I. Specific enzymatic activities of purified variants from the N6 and N9 libraries

Variant	Mutations	3OC12-HSL	γ-Nonalactone	δ-Nonalactone	PNPA	TBBL
1D7	*P40S*, I142L,P164T	17.11 ± 0.48	36.27 ± 0.36	1.87 ± 0.27	2.05 ± 0.14	10.25 ± 0.56
1B12	I142L,E146Q,P164T,H200A,V281T,*F292S*	21.78 ± 1.29	24.30 ± 1.05	3.66 ± 0.31	3.89 ± 0.63	10.26 ± 0.15
2A4	I142L,*P164S*,M247L	14.16 ± 0.87	28.39 ± 0.35	4.09 ± 0.25	1.65 ± 0.56	9.42 ± 0.12
2H10	*T50S*,I234V,V236I	15.39 ± 0.44	48.47 ± 0.77	2.80 ± 1.04	0.49 ± 0.02	30.55 ± 0.95
2D1	*E146K*	16.78 ± 3.45	37.44 ± 0.2	2.66 ± 0.15	2.92 ± 1.93	12.15 ± 0.84
2G12	I142L	18.14 ± 0.48	55.80 ± 7.47	2.56 ± 0.47	1.29 ± 0.01	14.68 ± 1.13
3G5	NS	9.8 ± 0.03	13.15 ± 0.5	0.00	20.9 ± 0.6	7.14 ± 0.15
3A6	NS	12.7 ± 1.16	17.16 ± 0.06	1.4 ± 0.2	4.3 ± 0.2	14.4 ± 4.3
3B1	NS	8.9 ± 0	8.4 ± 2.23	0.00	37.3 ± 0.6	5.19 ± 0.1
3H4	NS	10.7 ± 0.4	2.5 ± 0.3	2.77 ± 0.14	2.8 ± 0.18	9.86 ± 1.09
3A5	NS	9.8 ± 0.75	11.4 ± 0.08	1.25 ± 0.83	3.5 ± 0.01	9.28 ± 0.26
3C8	NS	9.4 ± 0.75	11.2 ± 1.6	1.22 ± 0.01	7.7 ± 0.08	6.19 ± 0.07
mpa-N9		19.27 ± 0.1	36.53 ± 0.99	2.30 ± 0.5	2.00 ± 0.26	9.93 ± 0.1
2A12	L239I,H293Y	0.78 ± 0.21	31.31 ± 1.37	15.06 ± 3.17	7.00 ± 0.54	25.76 ± 0.09
4C12	T56S,T67S,I270L,F348Y	0.37 ± 0.005	10.36 ± 0.92	1.44 ± 0.77	14.07 ± 0.21	14.29 ± 0
4C3	*L239M*,L240F	0.38 ± 0.02	5.97 ± 0.24	0.10 ± 0.07	19.76 ± 1.14	3.22 ± 0.06
2B7	*L239M*,L240F,I270L,I281V,L329I	0.69 ± 0.97	5.24 ± 0.21	0.00	28.50 ± 0	2.82 ± 0.37
1E3	*E19G*,V195T,I270L,I281V,L329I	0.04 ± 0.001	0.62 ± 0.03	0.05 ± 0.04	0.82 ± 0.04	0.71 ± 0.02
1C7	*V6A*,E271S,L329I	0.11 ± 0.003	1.03 ± 0.05	0.08 ± 0.03	0.45 ± 0.02	0.54 ± 0.003
3A1	NS	0.54 ± 0.08	12.2 ± 0.38	0.91 ± 0.05	3.47 ± 0.06	13.77 ± 0.75
3D10	NS	0.00	5.65 ± 0.53	1.93 ± 0.39	4.55 ± 0.15	9.1 ± 0.7
3D3	NS	0.08 ± 0.02	14.04 ± 1.03	0.25 ± 0.04	54.03 ± 3.6	6.89 ± 0.2
3B6	NS	0.67 ± 0.05	5.84 ± 0.27	0.16 ± 0.06	21.24 ± 1.2	2.73 ± 0.18
3H1	NS	0.28 ± 0.04	4.24 ± 0.4	0.67 ± 0.05	6.04 ± 0.1	4.09 ± 0.03
3H8	NS	0.12 ± 0.04	3.08 ± 0.007	0.95 ± 0.06	8.52 ± 0.03	4.17 ± 0.14
mpa-N6		0.53 ± 0.03	2.98 ± 0.22	2.05 ± 0.54	25.83 ± 0.17	18.90 ± 1.04

Specific activities are provided in units (1 μmol of substrate hydrolyzed per minute per mg enzyme). The initial substrate concentrations applied were 0.1 mM for TBBL, 0.25 mM for 3OC12-HSL and 1 mM for the remaining substrates. Spontaneous mutations that accumulated during library making are marked in italics. The error ranges correspond to the maximal deviation seen in ≥2 independent measurements. NS, not sequenced.

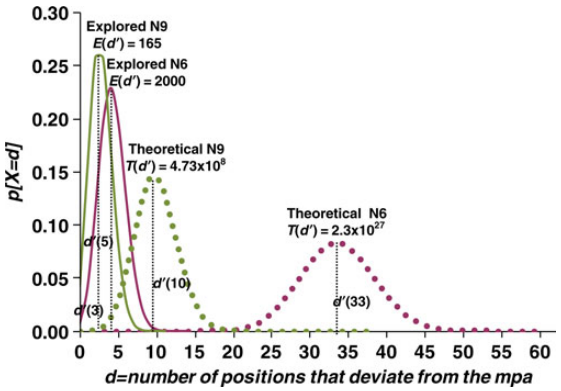

Fig. 2 Distributions and sizes of the N6 and N9 near-ancestor sequence ensembles. Plotted are histograms representing the fraction of sequences within an ensemble that carry a given number of positions with the alternatively predicted amino acid (this number, *d*, also represents the number of position that deviate from the mpa, i.e. for the mpa, $d = 0$). The *theoretical ensemble* explores all ambiguous positions (all position with marginal probabilities lower than 0.9). The *explored ensemble* includes only critical positions—i.e. positions with marginal probabilities <0.9 that are most likely to affect the ancestor's structure and/or function. The mean distance, *d'*, is the number of positions with the alternatively predicted amino acid in the most abundant subpopulation of a given ensemble. Ensemble sizes ($N_{(d')}$) correspond to the number of different sequences within the most abundant subpopulation.

the protein's structure or function. We therefore sought to focus the ensemble, and explore only ambiguous positions that are most likely to affect structure and/or function (hereafter 'critical positions').

Note that Fig. 2 takes the 'devil's advocate' view. Akin to the parsimony paradigm, we tested the robustness of the prediction while assuming that the alternative predictions are represented at equal likelihoods. However, as described below, the spiking approach implemented here yielded libraries whereby, per each gene, only few positions carrying the alternative prediction states, rather than *all* positions with low prediction accuracy carrying their alternative states. Thus, the fact that the second prediction usually comes with lower probability relative to the first one is accounted for.

Identification of critical positions

Our approach aims at identifying two types of potentially critical positions in the reconstructed protein: (i) positions under strong purifying selection that are predicted with low marginal probability. Such positions are more likely to affect structure and/or function than positions under weak selection. (ii) Positions that correspond to buried residues that are predicted with low marginal probability. Buried residues are far more crucial for protein folding and stability than surface ones. Mutations in buried residues (core residues) tend to be more deleterious than those on solvent exposed ones (surface residues) and are also more likely to affect function (Tokuriki *et al.*, 2008).

Accordingly, we have devised a systematic approach for detecting positions belonging to the above categories by considering the following three properties (Fig. 3):

1. *Marginal probabilities of the ancestral reconstruction*. The predicted amino acids and their marginal probabilities were determined by Fastml (Pupko *et al.*, 2000). A threshold probability is set that defines absolutely accurate predictions (C_{pred}). All

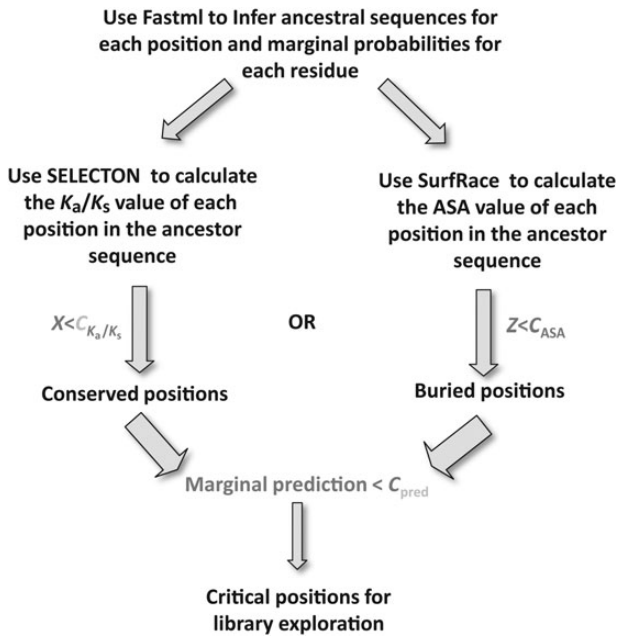

Fig. 3 The identification of potentially critical positions. Positions which are buried (core residues), or under strong selection, and are predicted with relatively low marginal probability are defined as critical. The thresholds define conserved positions (C_{K_a/K_s}), buried positions (C_{ASA}) and ambiguously predicted positions (C_{pred}) and can be tuned to fit the family and alignment under study. Location within or near the active site can serve as another criterion thereby further reducing the size of the explored ensemble.

positions in which the marginal probability for the most likely amino acid is lower than C_{pred} are assigned as ambiguous.

2. K_a/K_s *value*. For each position in the alignment, the ratio of amino acid altering (non-synonymous) substitutions (K_a) to synonymous substitutions (K_s) was determined. The resulting K_a/K_s ratio reflects the type and intensity of selection operating on individual positions. In cases of no selection, $K_a/K_s \approx 1$. In cases of strong purifying selection, K_a/K_s decreases, resulting in $K_a/K_s << 1$ (Massingham and Goldman, 2005). The calculations of K_a/K_s values were done with the SELECTON program (Stern *et al.*, 2007). The physio-chemical character of the replacements is also taken into consideration in the MEC evolutionary model used by SELECTON (Doron-Faigenboim and Pupko, 2007), which result in more accurate K_a/K_s estimates. All positions in which the K_a/K_s value was lower than a preset threshold (C_{K_a/K_s}) were defined as conserved.
3. *Surface accessibility*. Using the crystal structure, or structures, available for the family of interest (in our case, the structure of mammalian PON1 (Ben-David *et al.*, 2013)), the relative accessible surface area (ASA) was computed to determine how buried a position is. Since the structure is largely conserved within families, the same value was assigned to the corresponding positions in all sequences in the phylogenetic tree. The ASA values were calculated by using theSurfRace program (Tsodikov *et al.*, 2002). ASA value of 1 indicates a completely buried position, and a nil value indicates a completely exposed position. A threshold value was set that defines all positions with ASA values that are lower than C_{ASA} as buried positions.

A computer program implementing the above criteria was developed (see 'Materials and methods' section). The program applies the filtering steps described above and in Fig. 3, and provides a list of potentially critical residues. The threshold values, C_{K_a/K_s}, C_{pred} and C_{ASA}, determine the number of critical positions that are selected for diversification. For example, with relatively tight thresholds, i.e. $C_{pred} \leq 0.7$, $C_{K_a/K_s} \leq 0.4$ and $C_{ASA} < 0.05$, only 6 critical positions were defined for N9, and 11 positions for the N6. We applied, however, more relaxed values for exploring the PON ancestors. Specifically, the threshold for marginal probability (C_{pred}) was set to 0.9. C_{K_a/K_s} was set to 0.5, resulting in defining ~25% of the positions in the PON ancestors as conserved (Supplementary Fig. S4) and C_{ASA} was 0.25, which is the threshold differentiating surface from core positions in a wide range of proteins (Tokuriki *et al.*, 2007). Based on these thresholds, 11 critical predicted positions were identified for N9, and 42 for the N6. While in N9, this resulted in ensembles that can be easily sampled ($d' = 3$; $n = 165$; Fig. 2), for N6 the ensemble size was still too large ($d' \sim 12$; $n \sim 1.1 \times 10^{10}$). Because mpa-N6 was properly folded, stable and enzymatically active, our focus was the functional variability of the near-ancestor ensemble. We therefore further filtered the list of critical residues, and examined only positions that reside within or near the active site. All positions that reside within 12 Å from the catalytic calcium were examined. Indeed, 12 Å is a generously set threshold, and includes all positions that make the active-site wall, as well as all catalytic residues, and their second-shell residues (Alcolombri *et al.*, 2011) (Fig. 4; Supplementary Table SI). This criterion gave 14 critical positions in total, and an ensemble that could be experimentally sampled ($d' = 5$; $n = 2000$; Fig. 2).

Construction of libraries

The libraries were derived from the mpa sequence, with second predicted amino acids combinatorially spiked at their target sites at a frequency that corresponds to their marginal prediction probabilities. The N6 library was constructed by assembly PCR that combinatorially incorporated mutagenesis oligos encoding the second predicted amino acid (see 'Materials and methods' section for details). Sequencing of 62 randomly selected clones from this library indicated that, on average, each library gene carried 3.1 ± 1.3 positions with the second predicted amino acid. The N9 library was synthesized, and carried 3.0 ± 1.4 second prediction substitutions per gene. In both cases, no positional biases in the distribution of the second prediction substitutions were observed. The average number of spontaneous mutations per gene (due to synthesis errors, or PCRs) was low: 0.29 ± 0.24 in the N9 library and 0.4 ± 0.23 in the N6 library.

Library screens

The substrates used for enzymatic characterization of the libraries were: TBBL—a synthetic, chromogenic substrate that mimics PON's lipophylic γ-lactone substrates (Khersonsky and Tawfik, 2006); the quorum sensing lactone 3OC12-HSL that tested the ancestral homoserinelactonase activity (Bar-Rogovsky *et al.*, 2013); and PNPA that tested the promiscuous aryl-esterase activity (Supplementary Fig. S5). From each library, ~200 randomly picked clones were grown in 96 well plates and assayed for their lactonase activity using TBBL. The percentage of active variants was found to be 55% in N9 library, and 65% in the N6 library. However, the lower viability of the N9 library was primarily the result of a low ligation yield (~30% of the library transformants comprised a self-ligated vector vs. 10% in the N6 library). And, the sequences of inactive N9 variants that did contain a library gene revealed the presence of deletions that resulted in nonsense mutations or frameshifts. This was also the case for inactive variants from the N6 library (10/12 sequences). Two

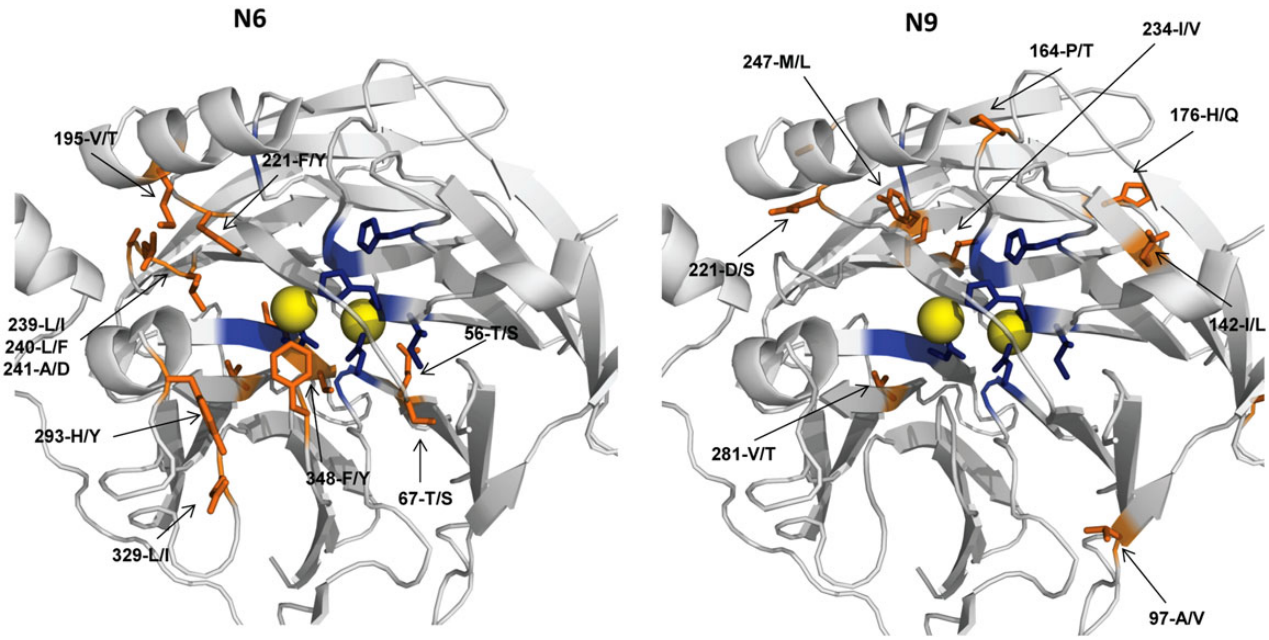

Fig. 4 The critical positions selected for experimental exploration mapped on the structure of PON1. The left panel represents the vertebrate ancestor N6, and the right panel represents the mammalian ancestor N9. Residues designated for diversification are marked in orange. The two calcium ions that are essential for maintaining PON's structure and catalytic activity are marked in yellow, and in blue, the residues that ligate these ions are denoted in sticks (E53, D54, N168, D169, N224, D268, N269) and the catalytic histidines (H115, H134).

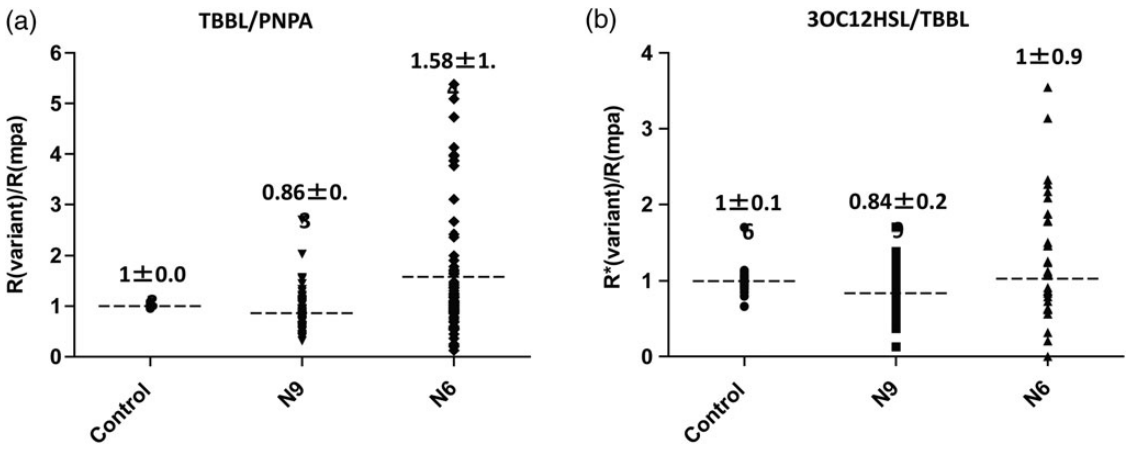

Fig. 5 Activities of randomly selected variants from the N9 and the N6 libraries. Presented are the ratios of activity for two different substrates, normalized for the ratio exhibited by the mpa: (**a**) normalized ratios of initial rates TBBL/PNPA ($n = 90$); and (**b**) like-wise for 3OC12-HS/TBBL ($n = 45$). Variants were grown, and their crude cell lysates were assayed in 96-well plates. Specific activities (μmol/min/μl crude lysate) were determined for TBBL, PNPA and 3OC12-HSL. To eliminate the effect of variations in expression levels, the ratios of specific activities for two substrates were calculated ($R_{variant}$) and then normalized to the ratio exhibited by the corresponding mpa (R_{mpa}). The *control* represents the variation in the assay, derived from 30 independent measurements (including growth and lysis) of mpa-N6 (or 45 measurements, in part b).

remaining inactive N6 variants carried three or more spontaneous mutations per gene. It was thus concluded that, in effect, ≥95% near-ancestor sequence variants encode properly folded and active enzymes.

Analysis of the active variants from the N9 library indicated that the TBBL activity of all the active variants was similar to that of mpa-N9. However, in the case of the N6 library, the TBBL activity exhibited high variation among library variants, and also relative to mpa-N6. Variations in activity in crude lysates can be the result of different expression levels, and/or of different specific activities. To eliminate the effect of expression level variations, both the lactonase (TBBL) and aryl-esterase activities (PNPA) were measured for each variant, and the ratio of these activities was determined and compared with that of mpa-N6 (Fig. 5a). Although the ratio of lactonase/aryl-esterase rates is independent of expression level, the ratio values were nearly as variable as the absolute activities, indicating that the variability in the N6 library stems from differences in specific enzymatic activities.

The variability in the N6 library could be over-estimated due to the fact that the tested substrates (TBBL and the aryl-ester PNPA) are synthetic, and that aryl-ester is a promiscuously hydrolyzed substrate [promiscuous activities tend to vary within variants that exhibit the same level of native activity (Amitai *et al.*, 2007)]. We therefore screened the libraries for homoserinelactonase activity using 3OC12-HSL. This substrate mediates quorum sensing in bacteria and relates to our hypothesis regarding the evolutionary origins of the PON family. Support of our hypothesis that PONs diverged from a homoserinelactonase [a 'quorum quenching lactonase' (Bar-Rogovsky *et al.*, 2013)], therefore, depends on the N6 and N9 ensembles showing homogeneity with respect to this activity. Forty-five variants from each library were tested for homoserinelactonase activity using the PAO-JP2 luminescence assay, and also with TBBL activity to address the variability in expression levels.

As shown in Fig. 5b, the N6 ancestral library also exhibited high variation in the 3OC12-HSL/TBBL activity ratio with a standard deviation of ±0.9 in comparison with ±0.29 measured for the N9 ancestral library. However, because the dynamic range of the PAO-JP2 luminescence assay is limited (complete 3OC12-HSL hydrolysis differs from no hydrolysis by only ~5× luminescence units), the variability within the N6 library is likely to be underestimated. The high variation in the activity ratio for TBBL vs. 3OC12-HSL demonstrates that the N6 library forms a rugged phenotypic ensemble with three different substrates. In contrast, the N9 library showed a significantly more homogeneous phenotype. The initial screen of 90 variants from the N6 library with TBBL/PNPA, and 45 variants with 3OC1HSL already revealed high phenotypic variability. Thus, there was no point in screening a larger number of variants with the aim of reliably sampling the N6 near-ancestor ensemble ($n = 2000$; Fig. 2). In the case of N9, the theoretical ensemble size is relatively small ($n = 165$), and unlike in the N6 library, an initial screening of 90 variants revealed high phenotypic similarity among the variants (Fig. 5).

Characterization of purified variants

To obtain a more detailed kinetic characterization of the near-ancestor library variants, 12 variants from each library were over-expressed and purified. For the N9 library, as the activities and specificities (ratio of activities with different substrates) of all library variants looked similar, we randomly selected 12 variants. For the N6 library, given the large variability, variants from the entire spectrum of specificities were selected, including variants with highest, lowest, as well as mpa-like activity ratios. The specific activities of the purified variants were determined for several substrates including 3OC12-HSL, the lipophylic δ- and γ-nonalactones (that cannot be assayed in crude lysates), TBBL and PNPA (Table I and Fig. 6).

The specific activities of the 12 purified variants from each library are presented in Table I. As observed in the 96-well screens of crude lysates, the purified variants from the N6 library showed high phenotypic variability, and in particular with respect to the homoserinelactonase activity—ca. 20-fold difference within the 12 purified variants (0.78 units for the most active variant vs. 0.04 units for the least active variant). In comparison, the same sample size from the N9 library (12 variants) showed <2-fold differences.

Most variants contained in addition to near-ancestor mutations, mutations that were not included in the library (marked in italics, Table I). In most cases, they appear in positions that were targeted, but with an amino acid that does not comprise the second prediction alternative. For example, E146K in variant 2D1 from the mpa-N9 library, whereby Q is the alternative state, or P164S in 2A4, whereby T is the targeted alternative state. Owing to codon similarity, the physio-chemical nature of the misincorporated amino acids was similar, and accordingly, their effect on the phoneotype was marginal (1D1 and mpa-N9 show the same specific activities within experimental error, and 2A4 that also carries two near-ancestor mutations deviated by max of 1.8-fold with one substrate).

It is also evident that the N6 variability stems from alternative, second predictions in active-site positions. Several variants vary in positions whereby mutations have been found to affect PON1's substrate specificity (e.g. 293, 240), and others carry mutations in residues that are in direct contact with specificity altering positions (e.g. 67, 348) (Harel *et al.*, 2004). Although present in some variants, the effect of spontaneous mutations on the functional variability of N6 variants appears to be negligible. This is supported by the observation that dispersed activity profiles were also exhibited by variants that contained only second alternative substitutions (for example, variant 4C3 that contains one random mutation exhibits a similar trend to 4C12 which contains only alternatively predicted amino acids). Further, N9 variants that also carry spontaneous mutations exhibit nearly identical specificities (Table I).

Discussion

The notion that the mpa represents a single sequence out of a wider ensemble of equally probable sequences has not been overlooked (Thornton, 2004; Ugalde *et al.*, 2004; Chang *et al.*, 2005; Gaucher *et al.*, 2008; Carroll *et al.*, 2011). Nonetheless, most reconstruction experiments explore only one, or few ancestral sequences (Jermann *et al.*, 1995; Chandrasekharan *et al.*, 1996; Thornton *et al.*, 2003; Gaucher *et al.*, 2008; Konno *et al.*, 2010; Carroll *et al.*, 2011). The alternative sequences were usually chosen based on considerations that are specific for the family of proteins at hand, and the criteria applied to define ambiguous positions that might affect the ancestral phenotype were inconsistent. Theoretically, the ensemble of near-ancestor sequences should be defined by one criterion only—the threshold marginal probability considered as an absolutely accurate prediction. However, in practice, the need to experimentally sample the ancestral ensemble imposes severe limitations. For example, 10 ambiguous positions form a reasonable ensemble size of 2^{10} 1024. In many cases, however, the number of ambiguously predicted positions is much larger than 10, certainly in ancestors that mark early divergence events. However, positions that are predicted with the lowest probability are usually highly diverged, and therefore have little impact on the structure and/or function. We therefore aimed at a systematic approach that uses evolutionary conservation (K_a/K_s), structure (the degree of burial, or ASA values, and location within or near the active site), as well as marginal probabilities for identifying the ambiguously predicted positions that are most likely to affect the ancestor's phenotype. These parameters are generally applicable provided that a structural model is available for at least one family member. The threshold values for these parameters can then be adjusted to obtain reasonably sized near-ancestor libraries.

Based on our approach, ancestral libraries were designed that sample the near-ancestor sequence ensembles of two PON ancestors. The construction of the near vertebrate ancestor was carried out by oligo spiking (Herman and Tawfik, 2007). This method provides a simple and reproducible way for construction of near-ancestor libraries, assuming that a screening method that enables a reasonable sampling of the library is available. Oligo spiking also allows to easily tune

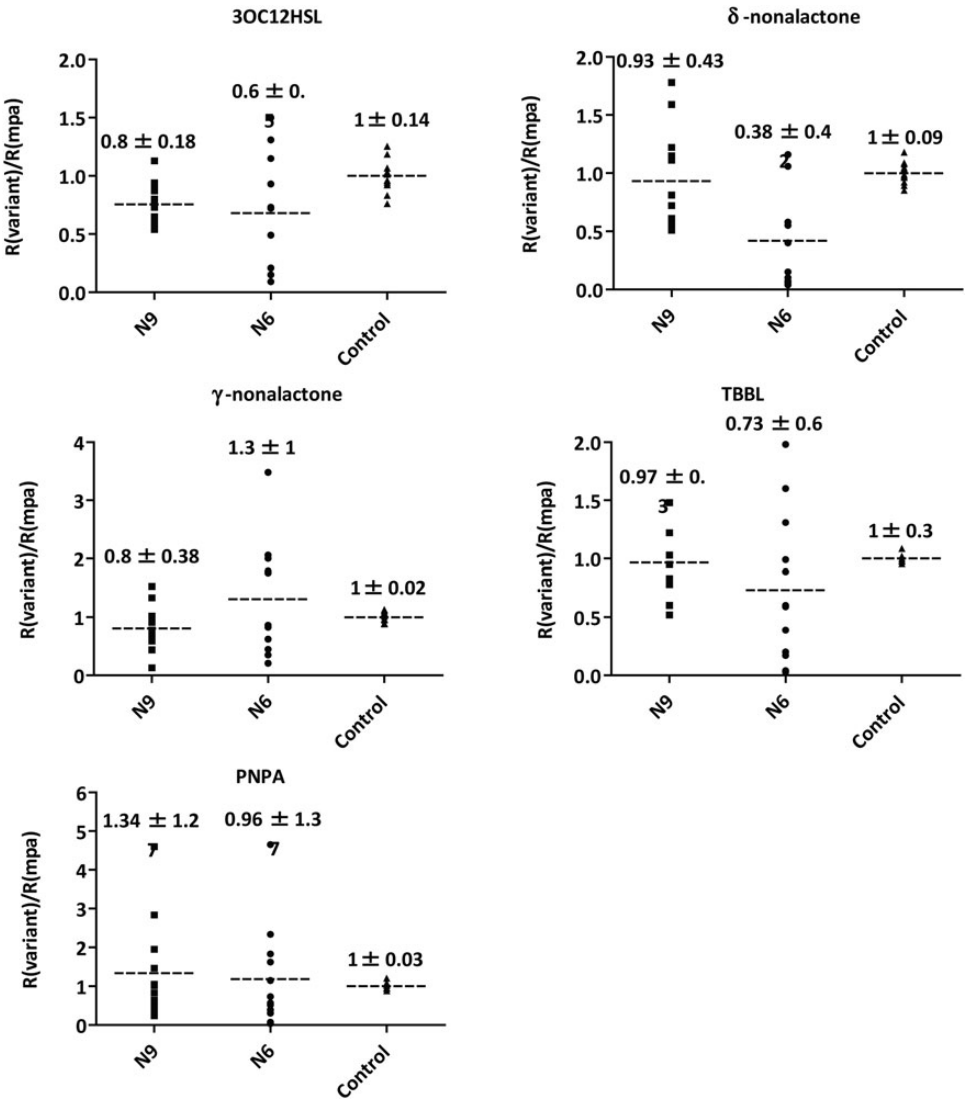

Fig. 6 The specific enzymatic activities of 12 purified variants from the N9 and N6 libraries. Shown for various substrates are dotplots of the normalized specific activities (activity of the mpa = 1).

the number of diversified positions (i.e. positions carrying the second predicted alternative). Library approaches are not applicable to every case, but medium-throughput activity screens (50–1000 variants) can be easily set for most binding or enzymatic activities (Arnold and Georgiou, 2003).

The sampling of second predictions addresses two equally undesirable scenarios: (i) the mpa exhibits diminished configurational stability and/or function; or (ii) the mpa exhibits functional properties that bare no historical relevance. In this case, both mpa-N6 and mpa-N9 were well expressed, stable, and enzymatically active at relatively high levels (Figs. 5 and 6) as described for other reconstructed mpa's (Williams *et al.*, 2006; Akanuma *et al.*, 2013; Risso *et al.*, 2013). Our findings are therefore not in full agreement with the view that reconstruction errors are most likely to reduce the performance of the ancestral gene product in functional assays or eliminate it altogether (Thornton, 2004). In general, it seems that as far as the reported cases go, the first scenario is uncommon. We therefore surmise that reconstruction uncertainties are more likely to lead to an erroneous ancestral phenotype than to a non- or poorly-functional ancestor. Indeed, the library screenings revealed that the second scenario takes place in the vertebrate ancestor N6. In spite of the fact that the theoretical ensemble size of the near mpa-N6 library is ~2000 (Fig. 2), high phenotypic variability was observed by screening as few as 90 variants. The incorporation of second predicted amino acids in N6 significantly changed the mpa's functional properties, suggesting that given the lack of vertebrate sequences, the ancestral phenotype cannot be determined in reasonable accuracy. The observed phenotypic variability in the N6 library also demonstrates that the ability of the applied conservation and structural criteria to identify positions in which incorporation of the alternatively predicted amino acid is likely to affect the function of the ancestor.

In contrast to N6, the phenotypic homogeneity observed in the sample of screened variants from the N9 library supports the notion that mpa-N9, and for that matter any other sequence within its near-ancestor ensemble, is likely to represent the phenotype of the actual ancestor from which the contemporary mammalian PONs have diverged (Bar-Rogovsky *et al.*, 2013), even though its exact genotype remains unknown. We note, however, that by definition, our approach

can certainly point out the low fidelity of certain predictions (e.g. N6), and can support the reliability of other predictions. But it cannot absolutely validate any predicted phenotype, including that of N9's.

It should also be emphasized that the validity of all ancestral predictions is strongly influenced by the number and quality of the available sequences, and foremost by their species distribution. The available sequences comprise a small and rather sporadic sampling of the actual diversity. The approach suggested here alleviates, to a certain extent, prediction uncertainties that stem from the stochasticity of the evolutionary process and limited samplings of sequence data. However, in this work, we do not consider other sources of error, such as the choice of the evolutionary model used to describe the substitution process. A significant progress towards more realistic evolutionary models was achieved in the past few decades. Models were developed that account for patterns of rate variation across sites reviewed in Pupko and Mayrose (2010), specific physio-chemical attributes of the protein review in Liberles *et al.* (2012), the non-stationarity of the evolutionary process see Susko and Roger (2013), across-site heterogeneities in the amino-acid replacement process (Lartillot and Philippe, 2004), to name a few. However, all these models approximate reality only to a certain extent and are usually over-simplified, i.e. they do not account for the complex structure–function relationships that dictate sequence evolution (Wilke, 2012). Thus, the development of ancestral sequence methodologies that account for uncertainty in the chosen model and in the inference of their parameters is an important topic for future research.

Foremost, inclusion of relevant sequences increases the prediction's fidelity, and may also change the ancestral phenotype (Carroll *et al.*, 2011). The representation of mammalian PON sequences is probably sufficient (as for other mammalian genes). Additional mammalian sequences are therefore unlikely to improve the prediction's fidelity (Li *et al.*, 2008) or alter the ancestral phenotype. However, the phenotype of the mammalian ancestor might be more sensitive to the inclusion of the currently unavailable platypus and/or opossum PON sequences (if these species are found to possess functional PONs). On the other hand, the inclusion of lizard and zebrafinch PONs did not induce a change in phenotype. The predicted ancestors from a tree that included these two sequences that appeared while this work was ongoing, differed in 11 positions. Nonetheless, phenotypically, the two mpa's, dubbed N9 and N9*, were found to be phenotypically identical (Bar-Rogovsky *et al.*, 2013). Thus, the ensemble approach presented here can unambiguously indicate that certain predictions are unlikely to reflect ancestors (e.g. N6), and support the likelihood of certain other predictions (e.g. N9).

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

Financial support by the Defense Threat Reduction Agency (HDTRA1-11-C-0026) and by the Israel Science Foundation is gratefully acknowledged. D.S.T. is the Nella and Leon Benoziyo Professor of Biochemistry.

References

Abascal,F., Zardoya,R. and Posada,D. (2005) *Bioinformatics*, **21**, 2104–2105.
Adey,N.B., Tollefsbol,T.O., Sparks,A.B., Edgell,M.H. and Hutchison,C.A., 3rd. (1994) *Proc. Natl Acad. Sci. U.S.A.*, **91**, 1569–1573.
Akanuma,S., Nakajima,Y., Yokobori,S., Kimura,M., Nemoto,N., Mase,T., Miyazono,K., Tanokura,M. and Yamagishi,A. (2013) *Proc. Natl Acad. Sci. U.S.A.*, **110**, 11067–11072.
Alcolombri,U., Elias,M. and Tawfik,D.S. (2011) *J. Mol. Biol.*, **411**, 837–853.
Ames,R.M., Money,D., Ghatge,V.P., Whelan,S. and Lovell,S.C. (2013) *Bioinformatics*, **28**, 48–55.
Amitai,G., Gupta,R.D. and Tawfik,D.S. (2007) *HFSP J.*, **1**, 67–78.
Arnold,F.H. and Georgiou,G. (2003) *Methods in Molecular Biology*. Humana Press, New Jersey.
Ashkenazy,H., Penn,O., Doron-Faigenboim,A., Cohen,O., Cannarozzi,G., Zomer,O. and Pupko,T. (2012) *Nucleic Acids Res.*, **40**, W580–W584.
Bar-Rogovsky,H., Hugenmatter,A. and Tawfik,D.S. (2013) *J. Biol. Chem.*, **288**, 23914–23927.
Ben-David,M., Elias,M., Filippi,J.J., Dunach,E., Silman,I., Sussman,J.L. and Tawfik,D.S. (2013) *J. Mol. Biol.*, **418**, 181–196.
Benner,S.A. (2002) *Proc. Natl Acad. Sci. U.S.A.*, **99**, 4760–4761.
Bershtein,S., Goldin,K. and Tawfik,D.S. (2008) *J. Mol. Biol.*, **379**, 1029–1044.
Cai,R., Yan,S., Liu,H., Leman,S. and Vinatzer,B.A. (2011) *Infect. Genet. Evol.*, **11**, 1738–1751.
Carroll,S.M., Ortlund,E.A. and Thornton,J.W. (2011) *PLoS Genet.*, **7**, e1002117.
Chandrasekharan,U.M., Sanker,S., Glynias,M.J., Karnik,S.S. and Husain,A. (1996) *Science*, **271**, 502–505.
Chang,B.S., Jonsson,K., Kazmi,M.A., Donoghue,M.J. and Sakmar,T.P. (2002) *Mol. Biol. Evol.*, **19**, 1483–1489.
Chang,B.S., Ugalde,J.A. and Matz,M.V. (2005) *Methods Enzymol.*, **395**, 652–670.
Dean,A.M. and Thornton,J.W. (2007) *Nat. Rev. Genet.*, **8**, 675–688.
Doron-Faigenboim,A. and Pupko,T. (2007) *Mol. Biol. Evol.*, **24**, 388–397.
Draganov,D.I., Teiber,J.F., Speelman,A., Osawa,Y., Sunahara,R. and La Du,B. N. (2005) *J. Lipid Res.*, **46**, 1239–1247.
Duan,K. and Surette,M.G. (2007) *J. Bacteriol.*, **189**, 4827–4836.
Edgar,R.C. (2004) *Nucleic Acids Res.*, **32**, 1792–1797.
Gaucher,E.A., Govindarajan,S. and Ganesh,O.K. (2008) *Nature*, **451**, 704–707.
Goldsmith,M. and Tawfik,D.S. (2013) *Methods Enzymol.*, **523**, 257–283.
Guindon,S. and Gascuel,O. (2003) *Syst. Biol.*, **52**, 696–704.
Hanson-Smith,V., Kolaczkowski,B. and Thornton,J.W. (2010) *Mol. Biol. Evol.*, **27**, 1988–1999.
Harel,M., Aharoni,A., Gaidukov,L., *et al.* (2004) *Nat. Struct. Mol. Biol.*, **11**, 412–419.
Harms,M.J., Eick,G.N., Goswami,D., Colucci,J.K., Griffin,P.R., Ortlund,E.A. and Thornton,J.W. (2013) *Proc. Natl Acad. Sci. U.S.A.*, **110**, 11475–11480.
Herman,A. and Tawfik,D.S. (2007) *Protein Eng. Des. Sel.*, **20**, 219–226.
Hobbs,J.K., Shepherd,C., Saul,D.J., Demetras,N.J., Haaning,S., Monk,C.R., Daniel,R.M. and Arcus,V.L. (2013) *Mol. Biol. Evol.*, **29**, 825–835.
Ingles-Prieto,A., Ibarra-Molero,B., Delgado-Delgado,A., Perez-Jimenez,R., Fernandez,J.M., Gaucher,E.A., Sanchez-Ruiz,J.M. and Gavira,J.A. (2013) *Structure*, **21**, 1690–1967.
Jermann,T.M., Opitz,J.G., Stackhouse,J. and Benner,S.A. (1995) *Nature*, **374**, 57–59.
Khersonsky,O. and Tawfik,D.S. (2005) *Biochemistry*, **44**, 6371–6382.
Khersonsky,O. and Tawfik,D.S. (2006) *Chembiochem*, **7**, 49–53.
Khersonsky,O., Roodveldt,C. and Tawfik,D.S. (2006) *Curr. Opin. Chem. Biol.*, **10**, 498–508.
Konagurthu,A.S., Whisstock,J.C., Stuckey,P.J. and Lesk,A.M. (2006) *Proteins*, **64**, 559–574.
Konno,A., Kitagawa,A., Watanabe,M., Ogawa,T. and Shirai,T. (2010) *Structure*, **19**, 711–721.
Kuang,D., Yao,Y., Maclean,D., Wang,M., Hampson,D.R. and Chang,B.S. (2006) *Proc. Natl Acad. Sci. U.S.A.*, **103**, 14050–14055.
Larkin,M.A., Blackshields,G., Brown,N.P., *et al.* (2007) *Bioinformatics*, **23**, 2947–2948.
Lartillot,N. and Philippe,H. (2004) *Mol. Biol. Evol.*, **21**, 1095–1109.
Le,S.Q. and Gascuel,O. (2008) *Mol. Biol. Evol.*, **25**, 1307–1320.
Li,W. and Godzik,A. (2006) *Bioinformatics*, **22**, 1658–1659.

Li,G., Steel,M. and Zhang,L. (2008) *Syst. Biol.*, **57**, 647–653.
Liberles,D.A., Teichmann,S.A., Bahar,I., *et al.* (2012) *Protein Sci.*, **21**, 769–785.
Loytynoja,A. and Goldman,N. (2008) *Science*, **320**, 1632–1635.
Maddison,W.P. (1995) *Syst. Biol.*, **44**, 474–481.
Massingham,T. and Goldman,N. (2005) *Genetics*, **169**, 1753–1762.
Perez-Jimenez,R., Ingles-Prieto,A., Zhao,Z.M., *et al.* (2011) *Nat. Struct. Mol. Biol.*, **18**, 592–596.
Pupko,T. and Mayrose,I. (2010) *Probabilistic Methods and Rate Heterogeneity*. John Wiley & Sons, Hoboken, NJ, USA.
Pupko,T., Pe'er,I., Shamir,R. and Graur,D. (2000) *Mol. Biol. Evol.*, **17**, 890–896.
Risso,V.A., Gavira,J.A., Mejia-Carmona,D.F., Gaucher,E.A. and Sanchez-Ruiz, J.M. (2013) *J. Am. Chem. Soc.*, **135**, 2899–2902.
Santander-Jimenez,S. and Vega-Rodriguez,M.A. (2013) *Biosystems*, **114**, 39–55.
Stamatakis,A. (2006) *Bioinformatics*, **22**, 2688–2690.
Stern,A., Doron-Faigenboim,A., Erez,E., Martz,E., Bacharach,E. and Pupko,T. (2007) *Nucleic Acids Res.*, **35**, W506–W511.
Sullivan,B.J., Nguyen,T., Durani,V., Mathur,D., Rojas,S., Thomas,M., Syu,T. and Magliery,T.J. (2012) *J. Mol. Biol.*, **420**, 384–399.
Susko,E. and Roger,A.J. (2013) *Syst. Biol.*, **62**, 330–338.
Teiber,J.F., Draganov,D.I. and La Du,B.N. (2003) *Biochem. Pharmacol.*, **66**, 887–896.
Thomson,J.M., Gaucher,E.A., Burgan,M.F., De Kee,D.W., Li,T., Aris,J.P. and Benner,S.A. (2005) *Nat. Genet.*, **37**, 630–635.
Thornton,J.W. (2004) *Nat. Rev. Genet.*, **5**, 366–375.
Thornton,J.W., Need,E. and Crews,D. (2003) *Science*, **301**, 1714–1717.
Tokuriki,N., Stricher,F., Schymkowitz,J., Serrano,L. and Tawfik,D.S. (2007) *J. Mol. Biol.*, **369**, 1318–1332.
Tokuriki,N., Stricher,F., Serrano,L. and Tawfik,D.S. (2008) *PLoS Comput. Biol.*, **4**, e1000002.
Tsodikov,O.V., Record,M.T., Jr. and Sergeev,Y.V. (2002) *J. Comput. Chem.*, **23**, 600–609.
Ugalde,J.A., Chang,B.S. and Matz,M.V. (2004) *Science*, **305**, 1433.
Watanabe,K., Ohkuri,T., Yokobori,S. and Yamagishi,A. (2006) *J. Mol. Biol.*, **355**, 664–674.
Wilke,C.O. (2012) *PLoS Comput. Biol.*, **8**, e1002572.
Williams,P.D., Pollock,D.D., Blackburne,B.P. and Goldstein,R.A. (2006) *PLoS Comput. Biol.*, **2**, e69.
Williams,S.G., Harms,M.J. and Hall,K.B. (2013) *J. Mol. Biol.*, **425**, 3846–3862.
Yamashiro,K., Yokobori,S., Koikeda,S. and Yamagishi,A. (2010) *Protein Eng. Des. Sel.*, **23**, 519–528.
Yang,J., Li,J., Dong,L. and Grunewald,S. (2011) *BMC Bioinform.*, **12**, 18.
Zhang,J. and Nei,M. (1997) *J. Mol. Evol.*, **44**(Suppl 1), S139–S146.
Zhang,J. and Rosenberg,H.F. (2002) *Proc. Natl Acad. Sci. U.S.A.*, **99**, 5486–5491.