

Exploring Nitrilase Sequence Space for Enantioselective Catalysis†

Dan E. Robertson,* Jennifer A. Chaplin, Grace DeSantis, Mircea Podar, Mark Madden, Ellen Chi,‡
Toby Richardson, Aileen Milan, Mark Miller, David P. Weiner, Kelvin Wong, Jeff McQuaid,
Bob Farwell, Lori A. Preston, Xuqiu Tan, Marjory A. Snead, Martin Keller, Eric Mathur,
Patricia L. Kretz, Mark J. Burk, and Jay M. Short

Diversa Corporation, San Diego, California 92121

Received 18 September 2003/Accepted 22 December 2003

Nitrilases are important in the biosphere as participants in synthesis and degradation pathways for naturally occurring, as well as xenobiotically derived, nitriles. Because of their inherent enantioselectivity, nitrilases are also attractive as mild, selective catalysts for setting chiral centers in fine chemical synthesis. Unfortunately, <20 nitrilases have been reported in the scientific and patent literature, and because of stability or specificity shortcomings, their utility has been largely unrealized. In this study, 137 unique nitrilases, discovered from screening of >600 biotope-specific environmental DNA (eDNA) libraries, were characterized. Using culture-independent means, phylogenetically diverse genomes were captured from entire biotopes, and their genes were expressed heterologously in a common cloning host. Nitrilase genes were targeted in a selection-based expression assay of clonal populations numbering 10^6 to 10^{10} members per eDNA library. A phylogenetic analysis of the novel sequences discovered revealed the presence of at least five major sequence clades within the nitrilase subfamily. Using three nitrile substrates targeted for their potential in chiral pharmaceutical synthesis, the enzymes were characterized for substrate specificity and stereospecificity. A number of important correlations were found between sequence clades and the selective properties of these nitrilases. These enzymes, discovered using a high-throughput, culture-independent method, provide a catalytic toolbox for enantiospecific synthesis of a variety of carboxylic acid derivatives, as well as an intriguing library for evolutionary and structural analyses.

An inherent macromolecular asymmetry, determined by the primordial choice of L-amino acids as protein structural determinants, results in a bias in the populations of chiral molecular moieties produced by stereoselective enzymatic catalysis and in the specificities of those enantiomers in their subsequent interactions with proteins in complex natural systems (16). The principle of chirality and the natural predisposition for shape and handedness in molecular binding by receptors, pumps, and enzymes have been recognized as essential principles for effective drug design (6). In turn, asymmetric enzymatic catalysis, unlike scalar chemical methods, is the optimal tool for synthesis of these enantiopure, pharmaceutically optimal molecules (24). However, the synthetic potential of biocatalysts has yet to be fully realized due to the paucity of available enzymes.

The protein sequence space parsed by modern science represents a small fraction of the genetic information available in the biosphere. This has become apparent from recent microbiological efforts targeting a spectrum of physical and chemical environments. These studies have shown that biotopes at extremes of temperature, pressure, pH, salinity, etc., are rich in microbial biodiversity. It is also clear that the physical properties and chemical specificities of the associated gene products

reflect the extrinsic and intrinsic physical properties of the biotope and its own participation in a unique microscopic chemical cycle. It has been estimated that <0.1% of all eubacteria and archaea have been cultured and characterized (1, 20, 27). The geography of this potential sequence space promises the existence of new enzymes with fitness parameters compatible with highly selective industrial transformations.

This study focused on the discovery of new nitrile-hydrolyzing enzymes from various environmental sources with the goal of catalyzing efficient chemo- and enantioselective carboxylic acid synthesis. Due to the favorable economics of their preparation, nitriles are attractive starting points in fine chemical manufacturing. However, their conversion to corresponding amides and carboxylic acids requires harsh conditions and generates significant waste (31). Enzymatic hydrolysis of nitriles can be performed under mild conditions and can also leverage the advantages of stereoselective enzymatic hydrolysis and nitrile racemization to potentially realize 100% enantioselective conversions.

A broad spectrum of biotopes could be targeted for nitrile-hydrolyzing enzymes due to microbially assisted chemical cycles involving the mineralization of organocyanides. These compounds are intermediates in the metabolism of cyanide by plants, animals, and microbes. Cyanogenic glycosides and cyanolipids are found in plants, aminonitriles and cyanohydrins are found in fungi, mandelonitriles are found in arthropods, and a variety of nitrile compounds are found in microorganisms (15). It has been speculated that microbial nitrile metabolism is, at least in part, a consequence of eukaryotic-prokary-

* Corresponding author. Mailing address: Diversa Corporation, 4955 Directors Place, San Diego, CA 92121. Phone: (858) 526-5149. Fax: (858) 526-5649. E-mail: drobertson@diversa.com.

† We dedicate this paper to Mark Madden, whose enthusiasm inspired and propelled our nitrilase research.

‡ Present address: Syrrx Corporation, San Diego, CA 92121.

otic ecological relationships and that members of the nitrilase superfamily of genes have been horizontally transferred from plants to bacteria and archaea (18). Microbes metabolize cyanosugars, cyanohormones, such as plant auxins, and other organocyanide compounds produced for defense or storage by bacteria and eukaryotes. Also, some pathogenic fungi have cyanide hydratases that allow them to infect plants that synthesize large amounts of defensive alkenyl glucosinates, which break down into isothiocyanates and nitriles (25).

Two general but evolutionarily distinct mechanisms have been described for enzymatic nitrile hydrolysis (5, 18). A two-step nitrile hydrolysis is catalyzed by the concerted action of a nitrile hydratase, producing an amide and a coexpressed amidase, which completes the transformation to the carboxylic acid. A direct hydrolysis of a nitrile to its corresponding carboxylic acid, the mechanism of interest here, is catalyzed by a nitrilase. The nitrilases are a subcategory of the carbon-nitrogen hydrolase superfamily, whose members in general catalyze nonpeptide carbon-nitrogen bond hydrolysis (18). Sequence analysis and subdivision of the superfamily results in the categorization of 13 distinct branches that further sort by bond specificity. Enzymes in the superfamily catalyze amidase, carbamylase, 9*N*-acyltransferase, and nitrilase reactions. The inference from sequence and structure data for two of the superfamily members, one an *N*-carbamyl-D-amino acid amidohydrolase from *Agrobacterium* (17) and the other the NitPhit Rosetta Stone protein from *Caenorhabditis* (19), is that the enzymes utilize a thiol mechanism, contain a Glu-Lys-Cys catalytic triad, and fold into an α - β - β - α sandwich conformation (18). There are no structures for the nitrilase subfamily, though their mechanisms and structures are presumed to conform to that of the amidohydrolase.

This study sought to leverage the synthetic utility of the nitrilase subfamily (EC 3.5.5.1) by searching exhaustively through large numbers of microbial genomes for novel sequences. Using a novel culture-independent approach, environmental DNA samples derived from >600 biotopes were screened exhaustively for nitrilase gene products with a view toward development of a synthetic toolbox for chiral carboxylic acid synthesis. Each of the newly discovered nitrilases was screened for activity on three nitrile substrates which are hydrolyzed to yield mandelic acid, phenyl lactic acid, and 4-cyano-3-hydroxybutyric acid. The collection of a large and diverse group of homologous sequences and the characterization of their activities also provides an opportunity to explore principles of protein structure and evolution and their relationship to catalytic specificity.

MATERIALS AND METHODS

DNA libraries. The samples corresponding to screened environmental-DNA (eDNA) libraries were collected under agreements negotiated with the political entities in control of these natural resources. All collecting was conducted under the guidelines of the Convention on Biological Diversity signed in 1992.

High-quality, high-molecular-weight DNA was isolated directly from soil or water samples following separation of cells from the environmental matrix (27–29). Highly purified suspensions of microbial consortia were obtained by isopycnic density gradient centrifugation with Nycodenz (22). The resulting cell pellet

was lysed by enzymatic and chemical digestions, followed by the isolation and purification of genomic DNA (33). After isolation from environmental contaminants, the genomic DNA was either digested using restriction enzymes or sheared to provide clonable fragments 1 to 10 kb in length.

The gene libraries were constructed in λ ZAP-based cloning vectors (Stratagene, La Jolla, Calif.). The library contained members with insert sizes of 1 to 10 kb. The libraries were propagated in the form of bacteriophage lambda and amplified to produce high-titer stocks. The lambda library constructed in λ ZAP was then converted into a phagemid library through *in vivo* excision (26).

Nitrilase selection. To prepare clones for screening, each eDNA phagemid library was combined with the *Escherichia coli* host (*X'* Sup⁰ F'), suspended at an A_{600} of 1 in 10 mM MgSO₄, allowed to adsorb for 15 min, and then incubated at 37°C for 45 min to allow expression of the kanamycin antibiotic resistance gene. Infected cells, now designated by kanamycin resistance, were plated on kanamycin Luria-Bertani (LB) plates and allowed to grow overnight at 30°C. Titer plates were also made to determine the infection efficiency. After overnight growth, the cells were pooled, washed, and resuspended in 10 mM MgSO₄. The cell suspension was then used to inoculate 1 ml of liquid M9 medium (without nitrogen) supplemented with 10 mM nitrile substrate. M9 media consisted of 1× M9 salts (with NH₄Cl omitted), 0.1 mM CaCl₂, 1 mM MgSO₄, 0.2% glucose, and ~10 mM either adiponitrile or (*R,S*)-4-chloro-3-hydroxyglutaronitrile. These selection cultures were then incubated at 30°C with shaking at 200 rpm for up to 5 weeks. Libraries containing positive nitrilase clones were identified by visual observation of growth. Growth resulted from a clone's ability to hydrolyze the nitrile substrate (which was the sole nitrogen source), thus generating ammonia to support the growth of that clone. Positive clones were isolated from this liquid culture by streak purification onto solid LB medium. Several colonies from the streak plate were subsequently recultured in the same defined liquid medium. The phagemid DNA from the secondary cultures exhibiting growth was then isolated and sequenced to confirm the discovery of a nitrilase gene and to establish the unique nature of that gene. In general, one to three unique nitrilases were identified from each iteration of this process on a particular library. Unique nitrilase genes were then subcloned into an expression vector. The nitrilase genes were cloned in pSE420 (Invitrogen, San Diego, Calif.) and expressed in *E. coli* XL1 Blue MR (Stratagene). Expression was performed in LB or Terrific Broth (Difco) medium and induced with 0.1 mM IPTG (isopropyl- β -D-thiogalactopyranoside) at 30°C.

Phylogenetic analysis. For phylogenetic analysis, 150 nitrilase amino acid sequences were aligned using Clustal W, followed by manual refinement. The small regions in the global alignment that could not be reliably aligned were masked out, resulting in a data set containing 256 positions. A maximum-likelihood analysis was performed in ProML (Phylip version 3.6) (9), using the JTT substitution model (12) with equal rates, global rearrangements, and three random sequence addition replicates. The large data set made the use of more complex evolutionary models computationally prohibitive. We also performed a Bayesian phylogenetic inference using MrBayes (11). Four Monte Carlo Markov chains were run for 700,000 generations after stabilization of the likelihood values, generating 7,000 trees. A majority rule consensus tree was generated, and the percentage of the time a particular clade occurred, i.e., its posterior probability, was recorded at the nodes. Values of >80 to 85% were considered strong support and were considered equivalent to high-confidence values obtained by bootstrap analysis. The topology of the maximum-likelihood tree and that of the Bayesian consensus tree were virtually identical. The same topology was also obtained by a neighbor-joining analysis. The internal topologies of the individual clades were unchanged regardless of whether two plant (*Arabidopsis* and *Oryza*) nitrilase genes were included as outgroups.

An alignment of the sequences used in the analysis is available on request.

Biochemical assays for nitrilase specificity. 3-Hydroxyglutaronitrile was purchased from TCI America and was used as received. (*R*)-3-hydroxy-4-cyanobutyric acid was obtained from Gateway Chemical Technology (St. Louis, Mo.). Mandelonitrile, (*R*)- and (*S*)-mandelic acid, and (*R*)- and (*S*)-phenyllactic acid were purchased from Sigma Aldrich. The synthesis of phenylacetaldehyde cyanohydrin was described previously (8).

High-performance liquid chromatography assay methods for determination of reaction products have been described previously (8). Briefly, activity assays for mandelonitrile and phenylacetaldehyde cyanohydrin were conducted using 25 mM substrate and 0.6 mg of enzyme (lyophilized cell lysate)/ml in 0.25 ml of assay solution (0.1 M sodium phosphate buffer, pH 8, 10% MeOH) for 24 h. For 3-hydroxyglutaronitrile, assay conditions of 50 mM substrate, 1.2 mg of enzyme/ml, and pH 7 were used, with an incubation time of 48 h.

Nucleotide sequence accession numbers. The nitrilase sequences were deposited in GenBank (accession numbers AY487426 to AY487562).

RESULTS AND DISCUSSION

Nitrilases belong to a subfamily of the carbon-nitrogen hydrolase superfamily. Members of the superfamily share an α - β - β - α sandwich fold, similar topologies, and characteristic sequence motifs (2). About 20 nitrilases have been identified in bacteria, half of them having been identified in sequenced genomes and not experimentally tested for activity. While members of the superfamily have been identified in archaea, no archaeal enzyme with nitrile-hydrolyzing activity has been reported.

A total of 651 environmental samples collected worldwide from terrestrial and aquatic microenvironments were processed into genomic eDNA libraries containing fragments of 1 to 10 kb. These fragments were cloned into a common bacterial expression vector. The process results in a population of 10^6 to 10^9 unique clones, each with the potential to express the individual or multiple genes on its discrete eDNA insert. Each environmental sample resulted in a single eDNA library, and one selection was performed per library. A high-throughput selection screen was designed in which clones were rescued if they were able to hydrolyze an added nitrile substrate and liberate ammonia in nitrogen-free growth medium. Selection substrates were chosen with regard to the substituent character. Adiponitrile was chosen as a small aliphatic dinitrile, and (*R,S*)-4-chloro-3-hydroxyglutaronitrile was chosen as a chiral substitute for 3-hydroxyglutaronitrile, a prochiral molecule that is useful after desymmetrization in the synthesis of the cholesterol-lowering drug Lipitor (8). Over 200 nitrilase expression hits were sequenced and shown to be unique at the DNA level. In addition, two nitrile hydratase-amidase gene combinations were identified by this screening strategy. They were not further characterized as part of this study; 137 nitrilase gene products were expressed and further characterized.

The collection of a large and diverse group of homologous sequences and characterization of their activities provide an opportunity to begin exploration of principles of evolution and protein structure and their relationship to catalytic specificity. The nitrilase sequences were subjected to phylogenetic analysis using maximum-likelihood and Bayesian inference methods (11). In addition to the 137 novel sequences, 9 bacterial and 2 plant nitrilase sequences, as well as 2 related fungal cyanide hydratase sequences, were retrieved from GenBank for comparison. The Glu-Lys-Cys catalytic triad was identified in all of the sequences. The nitrilases were between 304 and 385 amino acids in length, most falling in the 320- to 340-amino-acid range; the variation in length is due primarily to extensions of C termini.

Inspection of the resulting unrooted tree, shown in Fig. 1, reveals the presence of several distinct and highly supported sequence clades. All previously known bacterial sequences belong to clades 1 and 2. The degree of amino acid sequence similarity varies by 40 to 60% between clades and by 75% within the individual clades. The four panels in Fig. 1 overlay biogeographical data, as well as substrate and enantioselectivity data, on the tree. Quantitative specificity data, correlated with the tree topology, are presented in detail in Fig. 2.

The variety of newly discovered nitrilases represents a broad

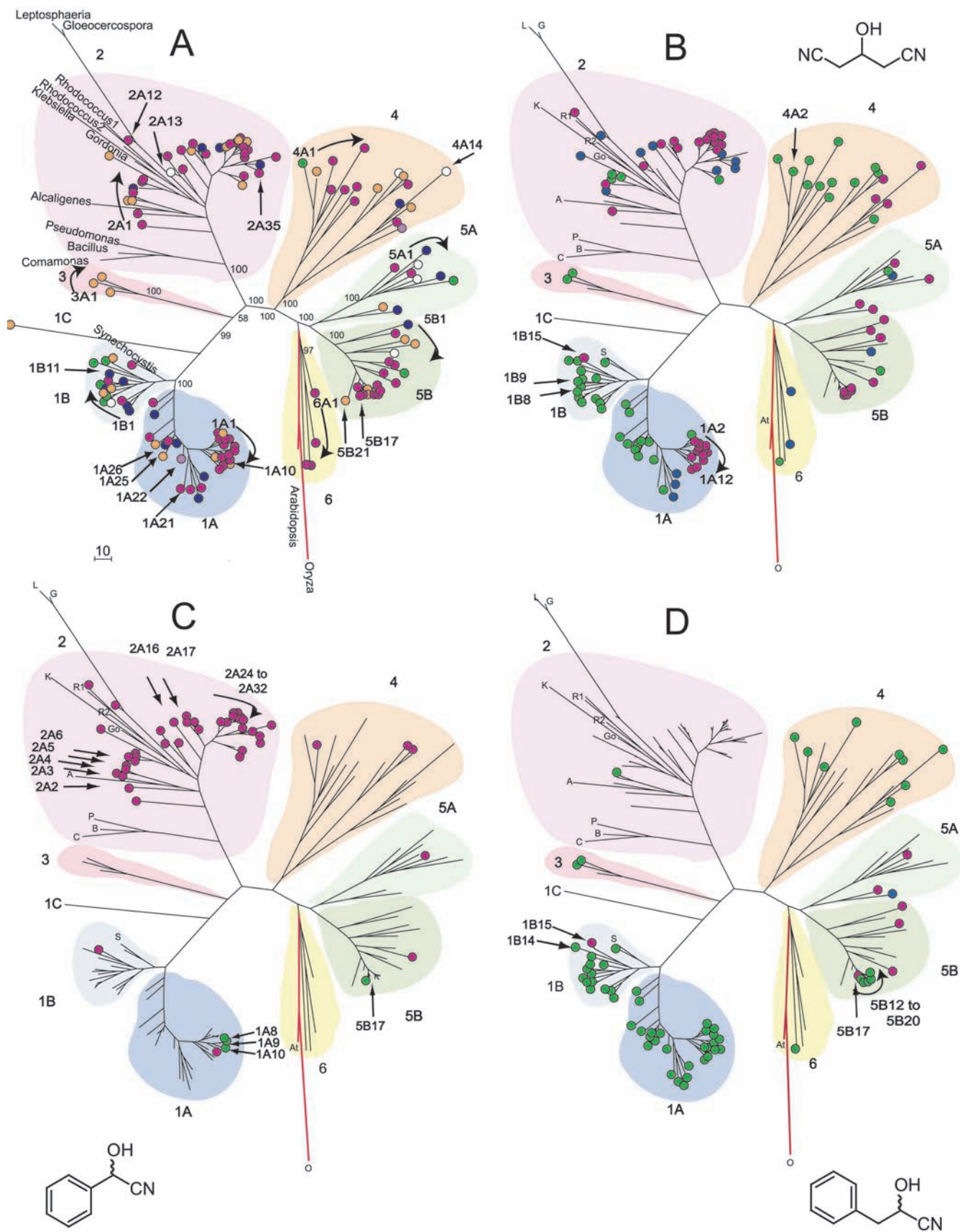
diversity of nitrilase sequence space. While most of the nitrilases were discovered in soil or in aquatic or marine sediments, there are no clear biogeographic correlations between the sample locations and the phylogenetic positions of the individual enzymes (Fig. 1A). In four cases (1A10 and 5B17, 2A12 and 5B21, 1B11 and 2A35, and 2A13 and 4A14), the same environmental library yielded two very different nitrilases, with sequences clustering in different clades. In contrast, examples of enzymes with similar sequences were discovered in both tropical and arctic samples. For example, nitrilase 1A26, which originates from Iceland, is 92% identical to 1A25 from Southeast Asia. In addition, 1A21, from a South Asian rainforest, is 71% identical to 1A22 from a deep-ocean location near Antarctica. Several clusters of highly similar sequences, such as those in clades 1A, 2, and 5, all of which are from tropical locations, were discovered. This may reflect episodes of rapid enzyme evolution resulting from the high microbial biodiversity associated with the multitude of microenvironments in tropical rainforests.

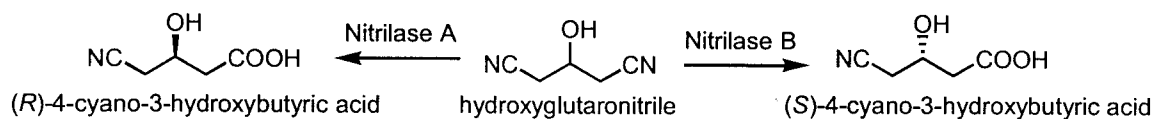
The natural substrates for nitrilases and the nature of the metabolic selection pressure that shapes the evolution of this family of enzymes are not understood. Activities have been reported, however, on a number of nitrile substrates that have chemical or pharmaceutical importance. Nitrilases from cultured species of the genera *Pseudomonas*, *Alcaligenes*, *Rhodococcus*, and *Acinetobacter* have previously been used in stereoselective syntheses of substituted (*R*)-mandelic acids (32), carbohydrate acids (14), (*S*)- α -phenylglycine (3), (*S*)-naproxen, and (*S*)-ibuprofen (13).

To determine the range of enzymatic specificity in the new nitrilase library, we tested the abilities of the 137 enzymes to catalyze stereoselective hydrolysis of three structurally distinct and industrially important nitrile substrates. The resultant chemo- and enantioselectivity of each enzyme was mapped to its position in the phylogenetic tree. Figure 1B to D illustrates the unrooted nitrilase tree with color-coded enantioselectivity indicated at the branch tips for each of the three substrates, 3-hydroxyglutaronitrile (Fig. 1B), mandelonitrile (Fig. 1C), and phenylacetaldehyde cyanohydrin (Fig. 1D). Substrate chemo- and enantioselectivities generally cluster with the phylogenetically defined clades. While none of these substrates is likely to be the natural substrate for these enzymes, the resulting specificities can provide clues regarding the evolution of the nitrilase active site to accommodate and enantioselectively hydrolyze structurally distinct molecules.

The nitrilase library was screened for the ability to desymmetrize the prochiral substrate 3-hydroxyglutaronitrile. This compound, with no reported natural occurrence, is an intermediate in the enantioselective synthesis of the cholesterol-lowering drug Lipitor (8) (see reaction scheme on page 2433, top).

Of the 137 nitrilases, 110 (>80% of the enzymes) hydrolyze this substrate (Fig. 1B and 2), with wide variation in the degree of enzyme enantioselectivity (Fig. 2). A few of the nitrilases exhibited the ability to hydrolyze both nitrile substituents of hydroxyglutaronitrile; however, this product is not commercially relevant and was therefore not pursued in the comprehensive analysis. The most selective enzymes generated products with enantiomeric excesses (ee) of >90% for (*S*)-4-cyano-3-hydroxybutyric acid (4A2) and >95% for (*R*)-4-cyano-3-hydroxybutyric acid (1A8 and 1A9). Several patterns of





enantioselectivity clustering can be discerned among the clades. While most of the enzymes in clade 1 are (*S*)-selective (27 enzymes), with ee values of 10 to 68%, there is a group of 12 related enzymes in subclade 1A (1A2 to 1A12) which show the opposite stereoselectivity, reaching values of 95% for 1A8 and 1A9. Outside of clade 1, (*R*)-selective enzymes predominate in clades 2 and 5, whereas (*S*)-selective enzymes constitute most of clade 4, although for the most part the ee values of these are modest. We can speculate that clades 1, 3, and 4 evolved from ancestral sequences with an enantioselective active site accommodating substrates structurally resembling the (*R*)-4-cyano-3-hydroxybutyric acid. Changes in the active-site configuration, resulting in reversal of stereoselectivity, occurred independently multiple times during the evolution of these enzymes, indicating possible switches in function following mutations or gene duplications.

(*S*)- and (*R*)-mandelic acids, the products of enantioselective mandelonitrile hydrolysis, are important in the production of pharmaceutical and agricultural intermediates (4) (see reaction scheme below).

Forty-eight nitrilases are active on mandelonitrile. The activity data fall into a radically different pattern from those of hydroxyglutaronitrile, with most active enzymes clustering in clade 2 (Fig.

1C). Activity was obtained only sporadically in other clades, including clades 1, 4, and 5. Forty-four of the active enzymes were (*R*)-selective, and 4 were (*S*)-selective, with ee values of up to 99% for (*R*)-mandelic acid and up to 30% for (*S*)-mandelic acid (Fig. 2). Most of the 37 (*R*)-specific enzymes of clade 2 were highly enantioselective, with ee values of >90%. Members of a clade 2 subgroup (2A24 to 2A32) of closely related sequences exhibit product ee values of $\geq 98\%$, as do several other more distant enzymes scattered throughout clade 2 (2A6, 2A16, and 2A17). Surprisingly, the two enzymes in this clade with the lowest stereoselectivity, 2A4 and 2A5 (ee = 9%), are 80% identical and are most closely related to the highly stereoselective enzymes 2A2 and 2A3 (ee > 90%; ~50% sequence identity). A group of three related enzymes in clade 1A (1A8 to 1A10; 82 to 95% identity) and one enzyme in clade 5 (5B17) have opposite enantioselectivities and generate (*S*)-mandelic acid with 25 to 30% ee. The activities of these enzymes on mandelonitrile are significantly lower than those of clade 2. (*S*)-Mandelic acid yields of $\leq 25\%$ were observed after 24 h of incubation, whereas most of the nitrilases of group 2 exhibited complete (*R*)-specific conversions within 6 h. While sequence clustering is observed relative to substrate specificity, the sequence, and consequently the structural factors that control enantioselectivity, are subtle, underscoring

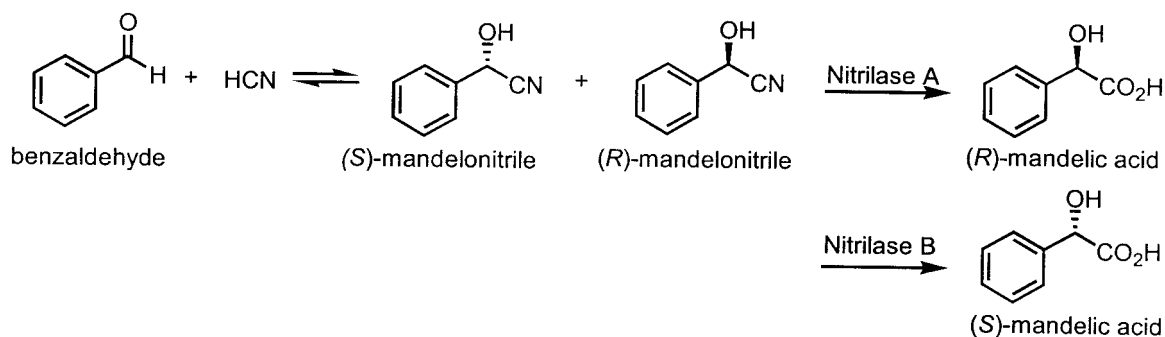
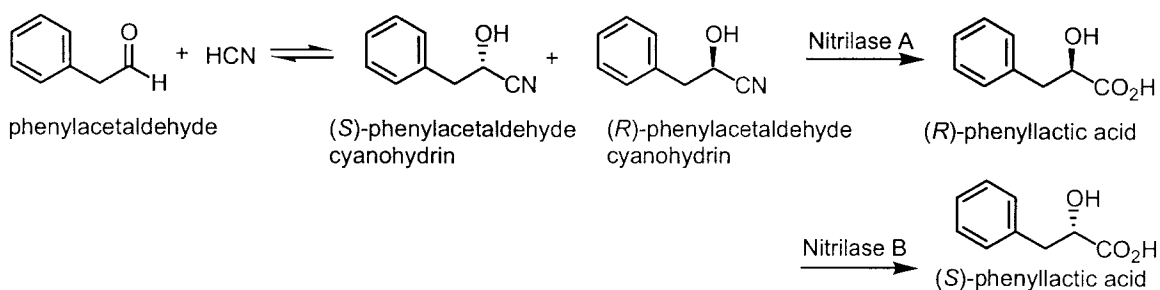


FIG. 1. Phylogenetic analysis of nitrilase sequences. The data are illustrated as a maximum-likelihood tree ($-\ln L = 43224.1$ [where L is the likelihood]) based on 137 new nitrilases and 13 previously known sequences from GenBank. The numbers at the nodes are percent bipartitions based on Bayesian phylogenetic inference (7,000 trees) and are shown for the major clades only. The scale bar represents the expected number of amino acid substitutions. (A) Geographic origins of environmental libraries from which individual nitrilases were discovered. Each geographic origin is represented by a colored dot: magenta, tropical Southeast Asia; mustard, tropical Central America; blue, Arctic; lilac, Antarctica; green, temperate; and white, other regions. (B to D) Reactivities and enantioselectivities of individual nitrilase enzymes on hydroxyglutaronitrile, mandelonitrile, and phenylacetaldehyde cyanohydrin. The (*R*)-enantioselectivity is represented by a magenta dot, (*S*)-enantioselectivity is represented by a green dot, and the absence of enantioselectivity is represented by a blue dot. Nonreactive enzymes do not have an associated dot. The public database enzymes were not tested.



the need to sample large portions of sequence space to uncover functional biocatalyst diversity.

Hydrolysis of cyanohydrin substrates to yield chiral arylactic acids (see reaction scheme above) is another important transformation providing valuable starting materials for fine chemical synthesis (4). Phenylacetaldehyde cyanohydrin does not appear in natural-product databases and may be a xenobiotic molecule.

Sixty-eight nitrilases were active on phenylacetaldehyde cyanohydrin; however, the patterns of activity tend to be opposed to those observed for mandelonitrile (Fig. 1D). Specifically, most of the enzymes that sort into clades 1, 3, 4, and 5 are active on this substrate, while they are, with a few exceptions, inactive on mandelonitrile. Enantioselectivity also correlates with individual clades, and unlike observations for hydroxyglutaronitrile, there appear to be fewer evolutionary events that lead to reversal of enantioselectivity. In two cases, minor sequence differences effected a reversal. The first is an (*S*)-specific enzyme (5B17; ee = 77%) within the cluster of highly similar enzymes in clade 5, whose members are otherwise (*R*)-selective (5B12 to 5B20). The second example is the (*R*)-selective enzyme 1B15 (ee = 76%) within clade 1, whose members are otherwise exclusively (*S*)-selective for this substrate. Strikingly, 1B15 is 97% identical to its neighbor, (*S*)-selective (ee = 43%) 1B14. 1B15 also shows an enantioselectivity switch for 3-hydroxyglutaronitrile relative to other members of its clade (Fig. 1C).

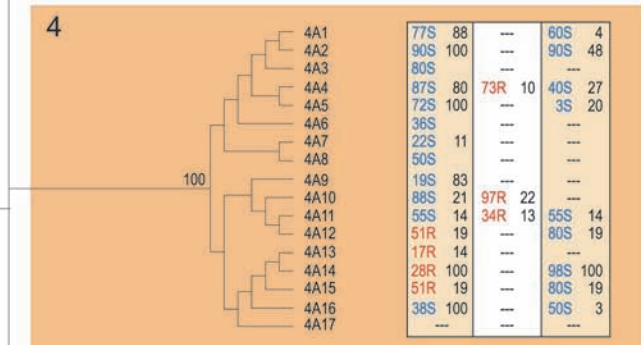
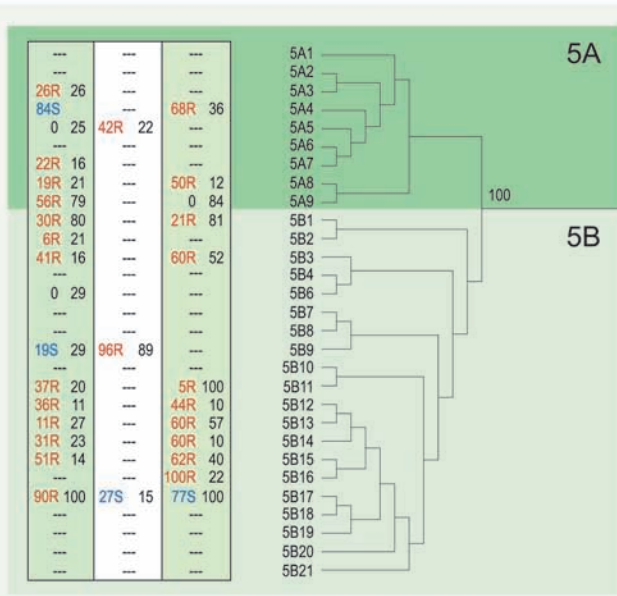
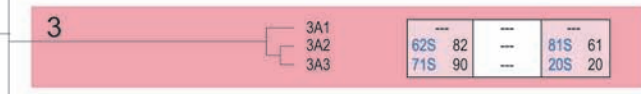
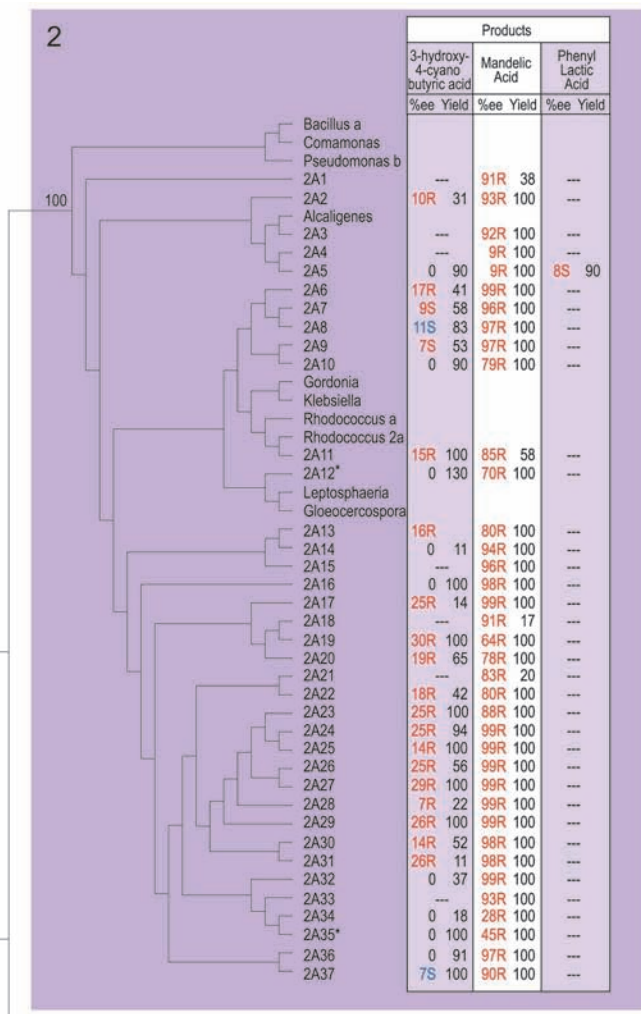
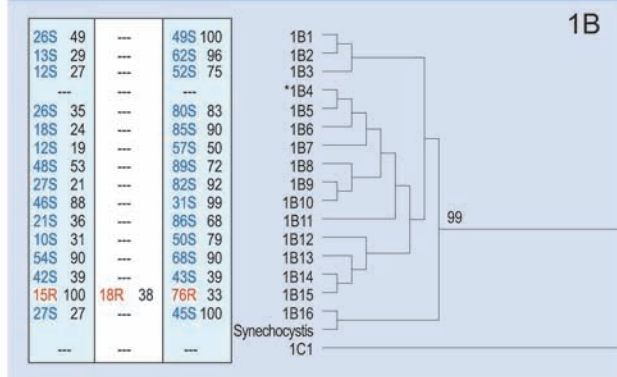
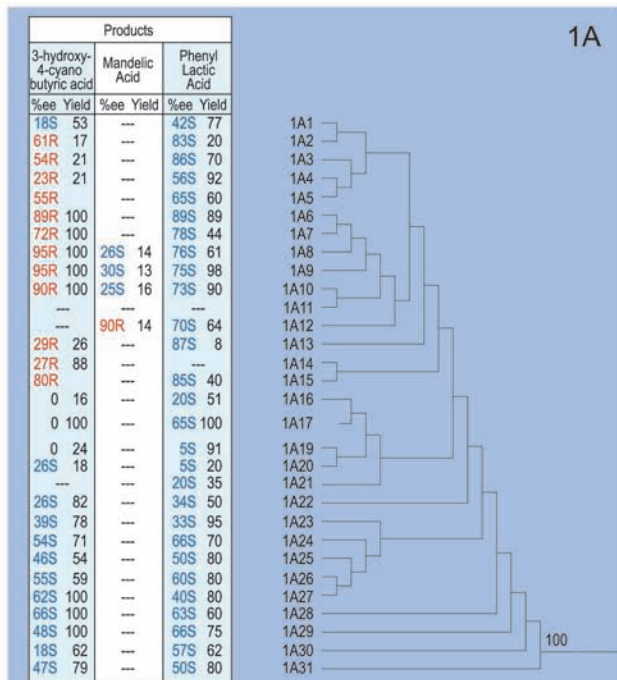
An important result of this study is the extraordinary multiplicity of novel nitrilase sequences discovered in diverse uncultured microbial populations. In a parallel to results from groups who have recorded new clades of bacteria and archaea in studies of microbial diversity using rRNA indexing (5), we have discovered several new clades of nitrilases, substantially

expanding the sequence space previously described for this subfamily, and in doing so we have facilitated access to chiral monomers important in pharmaceutical manufacture. Even so, the full breadth of diversity of these enzymes remains unknown. This is undoubtedly the case for all enzyme classes, as well as classes of other biomolecules.

High-throughput enzyme discovery has very clear benefits, both for substantially expanding protein sequence space and for solving problems in industrial catalysis. This discovery paradigm may be used to access other important and previously underexplored enzyme classes for use not only in biocatalysis but also in therapeutics, as well as evolutionary research. The nitrilase phylogenetic network has locales of exquisite selectivity, and multiple solutions have evolved which address synthetic issues of chemo- and enantiospecificity. Efforts are under way to elucidate the tertiary structures of selected nitrilases to extend these conclusions and to better understand the complex contributions of enzyme structure to activity and selectivity.

Laboratory methods to create gene sequence diversity by generating synthetic gene libraries comprised of arrays of point mutations or chimeric recombinants of selected parental genes are now available (10, 21, 23, 30). The directed-evolution approach is complementary to discovery and is enhanced by the availability of novel sequence templates available via the high-throughput functional-discovery methods described here (7). Enzyme discovery can directly solve problems in industrial catalysis and can also provide mutivariant starting points for rapid and highly targeted directed evolution. Clearly, the combination of discovery of new sequences and laboratory evolution where an ideal functional sequence does not exist is a key to the rapid solution of problems in industrial transformation, particularly in the area of chiral catalysis.

FIG. 2. Linear representation of the maximum-likelihood tree topology overlaid with data for yield (percent) and enantioselectivity (% ee) for each of the individual nitrilases for 3-hydroxy-4-cyanobutyric acid, mandelic acid, and phenyllactic acid. The reactions were performed as described in Materials and Methods. *, enzymes discovered using (*R,S*)-4-chloro-3-hydroxyglutaronitrile as a growth selection substrate; all other nitrilases were identified through growth selection on adiponitrile.



Oryza
Arabidopsis

ACKNOWLEDGMENTS

We gratefully acknowledge the efforts of Mark Wall, Iris Saliba, Kevin Keegan, Karsten Zengler, Nancy Farid, Michael Kulwiec, Pei Chen, Kelly Chatman, Trina Bowden, John Huang, Armenia Flordeliza, Chelsea Cowden, Ronald Swanson, Steven Wells, Robert Melkus, John Macomber, Lisa Bibbs, Ana Pallares, Damon Baptista, Joslin Gensch, Flash Bartnek, Amit Vasavada, Don DiMasi, and Melvin Simon.

REFERENCES

- Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**:143–169.
- Brenner, C. 2002. Catalysis in the nitrilase superfamily. *Curr. Opin. Struct. Biol.* **12**:775–782.
- Choi, S. Y., and Y. M. Goo. 1986. Hydrolysis of the nitrile group in α -aminophenylacetone nitrile by nitrilase: development of a new biotechnology for stereospecific production of *S*- α -phenylglycine. *Arch. Pharmacol. Res.* **9**:45–47.
- Coppola, G. M., and H. F. Schuster. 1997. Chiral hydroxyacids in enantioselective synthesis. Wiley-VCH, Weinheim, Germany.
- Cowan, D., R. Cramp, R. Pereira, D. Graham, and Q. Almatawah. 1998. Biochemistry and biotechnology of mesophilic and thermophilic nitrile metabolizing enzymes. *Extremophiles* **2**:207–216.
- Crossley, R. 1992. The relevance of chirality to the study of biological activity. *Tetrahedron* **48**:8155–8178.
- DeSantis, G., K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz, and M. J. Burk. 2003. Creation of a productive, highly enantioselective nitrilase through Gene Site Saturation Mutagenesis (GSSM). *J. Am. Chem. Soc.* **125**:11476–11477.
- DeSantis, G., Z. Zhu, W. A. Greenberg, K. Wong, J. Chaplin, S. R. Hanson, B. Farwell, L. W. Nicholson, C. L. Rand, D. P. Weiner, D. E. Robertson, and M. J. Burk. 2002. An enzyme library approach to biocatalysis: development of nitrilases for enantioselective production of carboxylic acid derivatives. *J. Am. Chem. Soc.* **124**:9024–9025.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics* **5**:164–166.
- Gray, K. A., T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, R. Knowles, L. Kan, P. E. Swanson, and D. E. Robertson. 2001. Rapid evolution of reversible denaturation and elevated melting temperature in a microbial haloalkane dehalogenase. *Adv. Synth. Catal.* **343**:607–617.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kakeya, H., N. Sakai, T. Sugai, and H. Ohta. 1991. Microbial hydrolysis as a potent method for the preparation of optically active nitriles, amides and carboxylic acids. *Tetrahedron Lett.* **32**:1343–1346.
- Klempier, N., A. De Raadt, K. Faber, and H. Griengl. 1991. Selective trans-formation of nitriles into amides and carboxylic acids by an immobilized nitrilase. *Tetrahedron Lett.* **32**:341–344.
- Legras, J. L., G. Chuzel, A. Arnaud, and P. Galzy. 1990. Natural nitriles and their metabolism. *World J. Microbiol. Biotechnol.* **6**:83–108.
- Milton, R. C. D., S. C. F. Milton, and S. B. H. Kent. 1992. Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **256**:1445–1448.
- Nakai, T., T. Hasegawa, E. Yamashita, M. Yamamoto, T. Kumasaka, T. Ueki, H. Nanba, Y. Ikenaka, S. Takahashi, M. Sato, and T. Tsukihara. 2000. Crystal structure of *N*-carbonyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure* **8**:729–737.
- Pace, H., and C. Brenner. 2001. The nitrilase superfamily: classification, structure and function. *Genome Biol.* **2**:1–9.
- Pace, H. C., S. C. Hodawadekar, A. Draganescu, J. Huang, P. Bieganski, Y. Pekarski, C. M. Croce, and C. Brenner. 2000. Crystal structure of the worm NitFhit Rosetta Stone protein reveals a NIT tetramer binding two Fhit dimers. *Curr. Biol.* **10**:907–917.
- Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734–740.
- Richardson, T. H., X. Tan, G. Frey, W. Callen, M. Cabell, D. Lam, J. Macomber, J. M. Short, D. E. Robertson, and C. Miller. 2002. A novel, high performance enzyme for starch liquification: discovery and optimization of a low pH, thermostable α -amylase. *J. Biol. Chem.* **277**:26501–26507.
- Rickwood, D., T. Ford, and J. Graham. 1982. Nycodenz: a new nonionic iodinated gradient medium. *Anal. Biochem.* **123**:23–31.
- Schmidt-Dannert, C. 2001. Directed evolution of single proteins, metabolic pathways and viruses. *Biochemistry* **40**:13125–13136.
- Schoemaker, H. E., D. Mink, and M. G. Wubbolts. 2003. Dispelling the myths: biocatalysis in industrial synthesis. *Science* **299**:1694–1697.
- Sexton, A. C., and B. J. Howlett. 2000. Characterization of a cyanide hydratase gene in the phytopathogenic fungus *Leptosphaeria maculans*. *Mol. Gen. Genet.* **263**:463–470.
- Short, J. M., J. M. Fernandez, J. A. Sorge, and W. D. Huse. 1988. λ ZAP: a bacteriophage λ expression vector with in vivo excision properties. *Nucleic Acids Res.* **16**:7583–7600.
- Short, J. M. 1997. Recombinant approaches for assessing biodiversity. *Nat. Biotechnol.* **15**:1322–1323.
- Short, J. M. 28 September 1999. Protein activity screening of clones having DNA from uncultivated microorganisms. U.S. patent 5,958,672.
- Short, J. M. 28 August 2001. Gene expression library produced from DNA from uncultivated organisms and methods for making the same. U.S. patent 6,280,926.
- Stemmer, W. P. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751.
- Weiser, M., and T. Nagasawa. 2000. Stereoselective nitrile-converting enzymes, p. 461–486. *In* R. Patel (ed.), *Stereoselective biocatalysis*. Marcel Dekker, New York, N.Y.
- Yamamoto, K., K. Oishi, I. Fujimatsu, and K. Komatsu. 1991. Production of *R*-(-)-mandelic acid from mandelonitrile by *Alcaligenes faecalis* ATCC 8750. *Appl. Environ. Microbiol.* **57**:3028–3032.
- Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**:316–322.