Supporting Information

Nicotinamide Mononucleotide Synthetase is the key enzyme for an alternative route of NAD biosynthesis in *Francisella tularensis*

Leonardo Sorci^{*}, Dariusz Martynowski[†], Dmitry A. Rodionov^{*}, Yvonne Eyobo[†], Xhavit Zogaj[‡], Karl E. Klose[‡], Evgeni V. Nikolaev[¶], Giulio Magni[∥], Hong Zhang[†], Andrei L. Osterman^{*,**,††}

^{*}Burnham Institute for Medical Research, 10901 North Torrey Pines Road, La Jolla, CA 92037; [†]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390; [‡]South Texas Center for Emerging Infectious Diseases, University of Texas, San Antonio, TX 78249; [¶]Department of Pathology, Anatomy and Cell Biology, and Daniel Baugh Institute for Functional Genomics and Computational Biology, Thomas Jefferson University, Philadelphia, PA 19107; [¶]Istituto di Biotecnologie Biochimiche, Università Politecnica delle Marche, 60131 Ancona, Italy; ^{**}Fellowship for Interpretation of Genomes, Burr Ridge, IL 60527

- 1. Methods
- Chemicals and Strains
- Gene cloning, protein overexpression and purification
- Structural analysis of *ft*NadE*
- Enzyme assays and steady-state kinetic analysis
- Mathematical modelling of the Mixed Enzymatic solution system
- Francisella tularensis nadE* knockout mutant and complementation study
- *In vivo* assessment of NAD biosynthetic intermediates
- 2. Tables and Figures
 - SI Table 1. Functional reconstruction of NAD biosynthesis
 - SI Table 2. NAD biosynthetic machinery in F. tularensis strains
 - SI Fig. 1. Conservation of NadE among the 7 fully sequenced variants of *F. tularensis*.
 - SI Fig. 2. NAD biosynthesis subsystem diagram.
 - SI Fig. 3. SDS-PAGE of F. tularensis and B. anthracis enzymes
 - SI Fig. 4. HPLC-based analysis
 - SI Fig. 5. Plots of initial rates versus substrate concentrations
 - SI Table 3. C. glutamicum and H. pylori NadE substrate preference.
 - SI Table 4. Crystal data and refinement statistics of *ft*NadE* complex
 - SI Fig. 6. Ligand binding in the 3D structure of NadE active site
 - SI Fig. 7. Multiple sequence alignment of representative bacterial NadE enzymes.
 - SI Fig. 8. Maximum-likelihood phylogenetic tree of bacterial NAD synthetase family
 - SI Fig. 9. Direct verification of *ft*NadE* catalyzed conversion of NaMN to NMN.
 - SI Scheme 1. Enzymatic system described by mathematical model
 - SI Table 5. Kinetics rates for enzymatic reactions as in Fig. 1
 - SI Table 6. Enzymatic kinetic parameters for the rection rates as in SI Table 1
 - SI Fig. 10. Transient processes for the metabolite concentrations for organism F
 - SI Fig. 11. Transient processes for the metabolite concentrations for organism B
 - SI Table 7. Steady state concentrations and fluxes in organisms *F* and *B*.
 - SI Fig. 12. Complete HPLC traces of NAD pathway reconstitution experiments.
 - SI Fig. 13. In vivo assessment of NAD intermediates
 - SI Fig. 14. Measurement of NMN synthetase activity in crude extracts of *F. tularensis* strain U112: comparison between wild type and $\Delta NadE^*$ mutant.
- 3. Discussion
 - Phylogenetic distribution and a possible evolutionary scenario of the newly identified NMN synthetase
- 4. References

Methods

Chemicals and strains.

Buffers, salts, and reagents, including the assay components ATP, NMN, NaMN, NAD, NaAD, ADP-ribose, inorganic pyrophosphatase and alcohol dehydrogenase were purchased from standard commercial sources. Enzymes for PCR and DNA manipulations were from New England Biolabs Inc. (Beverly, MA). Plasmid purification kits were from Promega (Madison, WI). PCR purification kits were from Roche Diagnostics (Indianapolis, IN). Nickel-nitrilotriacetic acid resin was from Qiagen Inc. (Valencia, CA). Oligonucleotides for PCR and sequencing were synthesized by Sigma-Genosys (Woodlands, TX). *E. coli* strains DH5α (Invitrogen) and BL21/DE3 (Stratagene, La Jolla, CA) were used for gene cloning and protein overexpression, respectively.

Gene cloning, protein overexpression and purification

Generation of *F. tularensis* and *B. anthracis* expression constructs. The genes for *fi*NadE* and *fi*NadM (locus tags FTL_0685 and FTL_0452, respectively) were amplified by PCR from *F. tularensis* subsp. Holarctica LVS genomic DNA. The genes for *ba*NadD and *ba*NadE (BA_1998 and BA_4558, respectively) were PCR-amplified from *B. anthracis* strain Ames genomic DNA (gifted by Dr. R. Liddington). The following primers were used: *ft*NadE*, 5'-GGTCAGATCATGAAAATAGTTAAAGA-3' (forward) and 5'-ACTTATGTCGACTCAGAAAATAGGTATGATATTTCAGTTT-3' (forward) and 5'- cTATCACCCTGCAGTTATAGTATAGTTTTTACC -3' (reverse); *ft*NadD, 5'-GCACTCATGAGAAAATTGGCATCATTG -3' (forward) and 5'-GAAGTCTGCAGTCACGATTCATAAAATTGGCATCATTG -3' (forward) and 5'-GAAGTCTGCAGTCACGATTCATACAACC-3' (reverse); and *ba*NadE, 5'-GGACACACTCATGACATTACAATACAACAGAT-3' (forward) and

5'-TTTATGGTCGACTTATTTCCACCAATC-3' (reverse).

The gel-purified PCR products were inserted into a pET-derived vector (1) containing the T7 promoter, His_6 -tag, and TEV-protease cleavage site and were sequenced to confirm PCR accuracy.

Production and purification of *ft***NadE***,*ft***NadM**, *ba***NadD**, and *ba***NadE**. All recombinant proteins were expressed as N-terminal fusions with a His₆-tag and a TEV-protease cleavage site in E. coli strain BL21/DE3. Cells were grown to an optical density at 600 nm of 0.8–1.0 at 37°C of LB medium. IPTG was added to 0.8 mM, and harvesting was performed after 12 h of shaking at 20°C. Proteins were purified from 6-liter cultures by chromatography on a Ni-NTA agarose column followed by gel filtration on a HiLoad Superdex 200 16/60 column (Pharmacia) with an AKTA FPLC system. The full protocol was similar to a procedure previously described (2). Proteins were purified to homogeneity as assessed by SDS-PAGE. Purified recombinant NadE enzymes from *Corynebacterium glutamicum* and *Helicobacter pylori* were obtained using the same techniques (unpublished results).

Structural analysis of ftNadE*

For crystallization of ftNadE*, the 6xHis-tag was removed by treatment with TEV protease and the tag-removed protein was further purified with an ion exchange column Resource Q (GE Healthcare Life Sciences). The cocrystals of ftNadE* complexed with AMP, pyrophosphate (PPi), and Mg²⁺ were grown in a sitting-drop vapor diffusion setting at 20°C with 8 mg/ml protein mixed with an equal volume of reservoir solution containing 0.2M K/Na tartrate and 16% PEG 3350 in the presence of 5 mM ATP, 10 mM MgCl₂, and 20 mM NaMN. For data collection, ftNadE* crystals were transferred stepwise to a cryoprotection solution containing the original reservoir solution and PEG 3350 increased to a final concentration of 40%. The diffraction data were collected on an R-AXIS IV image plate detector with X-ray generated by Rigaku FR-E SuperBright generator equipped with Osmic's VariMac focusing device and were processed with HKL2000 (3). The structure was solved by the molecular-replacement method of Molrep (4) using the *Bacillus subtilis* (*bs*NadE) structure (5) as the initial search model (pdb code 1kqp). The refinement and manual model building were performed with the program Refmac (6) of the CCP4 program suite(7) and COOT (8). The crystal data and current refinement statistics of *ft*NadE* structures are listed in SI Table 4.

Enzyme assays and steady-state kinetic analyses.

These assays were used to assess apparent values of kinetic parameters toward pyridine nucleotide substrates (NaMN, NMN, and NaAD at 10 μ M–8 mM)) at constant saturating concentration of other substrates (2 mM ATP for all enzymes and 4 mM NH₄Cl for NadE enzymes). The amount of enzyme in the reaction mixture was maintained at a level between 0.1 and 500 μ g/ml, leading to 1–10% substrate consumption within the incubation time. For HPLC assays, the linearity of response was assessed by the analysis of aliquots taken at 2–3 time points over the course of reaction. Apparent values of K_m and k_{cat} were calculated by fitting initial rates to a standard Michaelis–Menten model using the software Prism 4 (GraphPad), unless otherwise stated.

Continuous coupled assays - NaMN and NMN adenylyltransferase activity

The respective activities of NadD and NadM enzymes were measured with continuous assay coupling NAD formation to the alcohol dehydrogenase-catalyzed conversion of NAD to NADH as described (9) with slight modifications. Reactions were performed in a reduced volume of 100 uL using UV-transparent micro-cuvettes (BrandTech Scientific, Essex, CT) at 37 °C and monitored by the change in UV absorbance at 340 nm using a Beckman DU-800 spectrophotometer. Assay mixtures contained 2 mM ATP, 10 mM MgCl₂, 20 mM semicarbazide, 6 U/ml alcohol dehydrogenase, and 60 mM ethanol in 50 mM Hepes (pH 7.5). The reaction was started by adding either NMN or NaMN to 1 mM. To evaluate the NaMN-specific activity, an excess of pure recombinant NAD synthetase from *C. glutamicum* and NH₄Cl to 5 mM were also added. An NADH extinction coefficient of 6.22 mM⁻¹ cm⁻¹ was used for initial rate calculations using Multimode Detection software (Beckman).

Continuous coupled assay - NAD synthetase activity

The measurement of NAD synthetase (NADS) activity was determined as reported previously(10). Briefly, the reactions containing 2 mM ATP, 10 mM MgCl₂, 20 mM semicarbazide, 6 U/ml alcohol dehydrogenase, 60 mM ethanol, 5 mM NH₄Cl in 50 mM Hepes (pH 7.5) were initiated by adding NaAD to 1 mM.

Discontinuous (2-step) coupled assay - NMN synthetase activity

We developed a two-step spectrophotometric assay coupling the NMN synthetase activity to NADH formation via conversion of NMN to NAD by added excess of the purified recombinant *ft*NadM, followed by alcohol dehydrogenase-catalyzed conversion of NAD to NADH. The first step reaction (in 200 uL: 100 mM Hepes pH 7.5, 10 mM MgCl₂, 1 mM NaMN, 2 mM ATP, 5mM NH₄Cl) was started by the addition of NadE enzyme. After incubation at 37°C, the reaction was stopped by rapid cooling and removing of the active enzyme by microultrafiltration with Microcon YM-10 centrifugal filters (Amicon, Bedford, Mass.). A complete conversion of NMN into NADH was then accomplished by addition of one volume of the coupling buffer: 50 mM Hepes (pH 7.4), 20 mM Semicarbazide, 120 mM EtOH, 1.4 U of alcohol dehydrogenase and pure recombinant *ft*NadM. The mixture was

incubated at 25 °C for 20 min and NADH concentration was determined by absorbance at 340 nm (molar extinction of NADH used for all calculations was 6,220 M⁻¹ cm⁻¹).

HPLC-based assay - NMN and NAD synthetase activities, and in vitro pathway reconstitution

A direct HPLC-based assay was used to directly assess and compare NMN synthetase and NAD synthetase activities of all NadE enzymes (including those from C. glutamicum and H. *pvlori*), to determine steady-state kinetic parameters (for NaMN and NaAD) of *ft*NadE* and baNadE enzymes and to monitor pathway reconstitution using mixtures of purified recombinant enzymes. Kinetic assays were carried out in 0.1 ml reaction mixtures in 96-well plates containing 50 mM Hepes (pH 7.5) with 10 mM MgCl₂ and 4 mM NH₄Cl. The enzymes ftNadE* and baNadE were incubated with 2 mM ATP and varying concentrations (from 50 microM to 12.5 mM NaMN or NaAD at 37°C. For in vitro pathway reconstitution, mixtures of purified *ft*NadE* and *ft*NadM or *ba*NadD and *ba*NadE (3-100 μ g ml⁻¹ each) were incubated with 2 mM ATP, 1 mM NaMN, 10 mM MgCl₂ and 4 U/ml inorganic pyrophosphatase in 50 mM Hepes, pH 7.5 at 37 °C for 0.5-1 h and analyzed by ion-pair HPLC on LC-18T column (Supelco). Reactions were stopped by adding half-volume of ice-cold 1.2M perchloric acid (HClO₄). After 10 min at 4 °C, samples were centrifuged and brought to pH 6 by 0.8M K₂CO₃ addition. Proteins were removed from the supernatant by microultrafiltration and the filtrates were analyzed by RP-HPLC using a Shimadzu Prominence HPLC system with LC-10AD solvent delivery system, CBM-20AD system controller, SPD-20AD UV/Vis detector, SIL-20A auto sampler, CTO-20A column oven and DGU-20A3 on-line degasser. The separation was performed at room temperature on a Supelco LC-18-T (15 cm × 4.6 mm, 3 µm particle size) column equipped with Supelguard LC-18-T (2.0 cm \times 4.0 mm, 5 μ m particle size) guard column (Supelco, Bellefonte, PA), monitoring absorbance at 254 nm. Data were collected and processed by EZstart software package (Version 7.3). Chromatographic conditions and a standard separation are illustrated in Supplementary Fig.4. Molar areas were determined using calibration with standard solutions under the same chromatographic conditions.

NMN synthetase activity in cleared cellular lysates. Crude extracts of *F. tularensis* strain U112 wild-type and NadE* knockout mutant were prepared from mid log-phase grown cells as described above for protein purification. Briefly, cells were harvested, and resuspended for 20 min in a buffer with 1 mg/ml lisozyme, 2 mM PMSF, and 2 mM β -merchaptoethanol. After a cycle of thawing and sonication, cell debris was removed by centrifugation and the supernatant was probed for NMN synthetase activity (HPLC-based assay).

Mathematical Modelling of the Mixed Enzymatic Solution System Mathematical model

A mathematical model describing a well-mixed enzymatic solution system, depicted in Scheme 1, is a set of standard kinetic mass balances [1]

$$\frac{d[A]}{dt} = V_0 - V_1 - V_3, \tag{1}$$

$$\frac{d[B]}{dt} = V_1 - V_2,$$
 (2)

$$\frac{d[C]}{dt} = V_3 - V_4, (3)$$

$$\frac{d[D]}{dt} = V_2 + V_4 - k_D \cdot [D].$$
(4)

Here t is time, [A] is the concentration of NaMN, [B] is the concentration of NMN, [C] is the concentration of NaAD, and [D] is the concentration of NAD (the nomenclature of metabolite and enzyme names is provided in the main text and in Figure 1). V_0 is the constant influx of NaMN into the system and k_D is the degradation/utilization kinetics rate constant for NAD. The reaction rates V_1 , V_2 , V_3 , and V_4 , and the corresponding kinetic constants are provided in SI Table 5 and SI Table 6, respectively. The mathematical expressions for the kinetic rates (i.e. V_1 , V_2 , V_3 , and V_4) are obtained using a standard Michaelis-Menten approach (11, 12).

Below the following two different modelling situations are considered in detail, (1) simulation of transient time courses in a *closed* well-mixed "test-tube" enzymatic solution system (i.e. when $V_0 = k_D = 0$) and (2) the analysis of the steady-state fluxes and the metabolite concentrations in an *open* "cellular subsystem" with a non-zero mass flow across the subsystem's systematic boundary. In this case, the nonzero values of V_0 and k_D are set so that [D] = 1 (mM) at the steady-state.

To relate the data in Table 1 and 2 here (i.e. k_i , K_i , i = 1, 2, 3, 4) to the corresponding kinetic parameters in Table 1 of the main text, we note that $(k_1, K_1) = (k_{cat}, K_m)$ for ftNadE* (NaMN) and baNadE (NaMN), respectively; $(k_2, K_2) = (k_{cat}, K_m)$ for ftNadM (NaMN) and baNadD (NMN), respectively; $(k_3, K_3) = (k_{cat}, K_m)$ for ftNadM (NMN) and baNadD (NaMN), respectively; and $(k_4, K_4) = (k_{cat}, K_m)$ for ftNadE* (NaAD) and baNadE (NaAD).

Transient time courses in the closed "test-tube" enzymatic system

In this case, the zero values of the parameters V_0 and k_D are used, $V_0 = k_D = 0$. At time t = 0, initial conditions $[A_0] = 1$ (mM) and $[B_0] = [C_0] = [D_0] = 0$ (mM) were always used. By summing up the equations (1) through (4) with zero V_0 and k_D , it can be shown that the total sum of metabolite concentrations is preserved and does not change in time t,

$$[A] + [B] + [C] + [D] = [A_0].$$
(5)

Conservation relationship (5) is called a total *moiety* conserved pool. Mathematically, (5) means that differential equations (1) - (4) are linearly dependent and, therefore, to calculate the corresponding time courses in (1) - (4), only three differential equations, combined with the algebraic equation (5), are required. Specifically, equations (2) - (4) were integrated to calculate transient concentration profiles [B], [C], and [D], while concentration profile [A] of metabolite A was obtained from (5) using equations (6) - (9),

$$[A] = [A_0] - [B] - [C] - [D], (6)$$

$$\frac{d[B]}{dt} = V_1 - V_2,$$
(7)

$$\frac{d[C]}{dt} = V_3 - V_4,\tag{8}$$

$$\frac{d[D]}{dt} = V_2 + V_4.$$
(9)

The dependence of $V_1, ..., V_4$ on [A] is excluded from (7) – (9) using (6) substituted into the corresponding mathematical expressions for rates $V_1, ..., V_4$ given in Table 1.

To estimate the contribution of the two different routes, **Route II** utilizing enzymes ftNadE and ftNadM and **Route I** utilizing enzymes baNadD and baNadE, into the formation of the final product D (i.e. NAD), equation (9) was additionally replaced by two equations

$$\frac{d[D_1]}{dt} = V_2,\tag{10}$$

$$\frac{d[D_2]}{dt} = V_4. \tag{11}$$

Here $[D] = [D_1] + [D_2]$, D_1 can be interpreted as a "fraction" of metabolite *D* formed through flux V_2 (i.e. $A \xrightarrow{E_1} B \xrightarrow{E_2} D$) and D_2 is a "fraction" of *D* formed through V_4 (i.e. $A \xrightarrow{E_2} C \xrightarrow{E_1} D$). Equations (6) – (8), (10) and (11) were integrated using the Matlab[©] procedure *ode15s* with relative and absolute tolerances RelTal = AbsTol = 10⁻⁷, respectively. All integration calculations were automatically terminated when the concentration [D] of metabolite D reached ~90-99% of the initial concentration $[A_0]$ of metabolite A using a simple *event* procedure available through the Matlab[©] interface. The corresponding time courses are depicted in SI Figures 10 and 11. To relate SI Figure 10 and 11 to Fig. 3 (*a*, *b*) in the main text, we note that Figure 2(*a*) here corresponds to Fig. 11(*a*) in the main text, while Figure 11(c) here corresponds to Fig. 3(b) of the main text, respectively. We also note that vanishing concentrations of the intermediates do not assume that the corresponding fluxes leading to and from a vanishing intermediate are also vanishing. Specifically, although the concentration of intermediate B depicted in SI Figure 10(c) and (e) is negligible small compared to the other concentrations, the flux through B (i.e. fraction D1) is not small meaning that intermediate B is not accumulated in the mixture and it is instead converted to final product D rapidly.

From SI Figure 10, we find that in the case of organism F 100% of NAD is produced utilizing Route II when the concentration of E_1 is greater than the concentration of E_2 , that is $[E_1]:[E_2] = 10:1$ (see SI Figure 10 (a) and (b)). In the other two cases, depicted in SI Figure 10 (c) – (f), both enzymatic routes are involved in the formation of NAD. However, such scenarios may be unlikely to happen in or could be highly inefficient for organism F as in this case it can take 15-150 times longer to produce NAD. Note that for organism B, Route I always prevails over Route II in the formation of NAD as seen from SI Figure 11.

The analysis of steady states in the open enzymatic system

To evaluate steady-state fluxes in the open enzymatic system, the non-zero values of parameters V_0 and k_D were set. Specifically, we used $V_0 = 10^{-3} \text{ (mM} \cdot \text{s}^{-1})$ and $k_D = 10^{-3} \text{ (s}^{-1})$. After summing up equations (1) through (4) with non-zero values of V_0 and k_D , we obtain

$$\frac{d[D]}{dt} = V_0 - k_D \cdot [D].$$
⁽¹²⁾

It follows from (12), that the steady state concentration [D] of metabolite D is uniquely defined by the values of V_0 and k_D and [D] is independent of the other kinetic parameters,

$$[D] = \frac{V_0}{k_D}.$$
 (13)

Due to the choice of the specific values for V_0 and k_D , we have [D] = 1 (mM).

The steady-state values of metabolite concentrations and fluxes for organisms F and B with different ratios of enzyme concentrations E_1 and E_2 were obtained using a Matlab[©] numerical solver *fsolve* utilizing the Gauss-Newton method applied to the system of algebraic equations

$$V_0 - V_1 - V_3 = 0, (14)$$

$$V_1 - V_2 = 0, (15)$$

$$V_3 - V_4 = 0, (16)$$

$$V_2 + V_4 - k_D \cdot [D] = 0. \tag{17}$$

The equations (14) - (17) are obtained from the differential equations (1) - (4) by equating the time derivatives to zero values, d[A]/dt = d[B]/dt = d[C]/dt = d[D]/dt = 0 (mM·s⁻¹). Appropriate approximations for the solutions to (14) - (17), used as initial conditions for the Newton-like solver fsolve, were generated by the direct integration of equations (1) - (4) over a long time interval, $t_{end} = 10^4$ (s). The stability of the corresponding steady-state solutions was calculated using Matlab[©] function *eig* utilized to calculate the eigenvalues of the Jacoby matrix of partial derivations of the linearization of (14) - (17) at the solution. The Jacoby matrix was directly available from *fsolve* function. The corresponding steady state concentrations and fluxes are summarized in Table 3 for different values of enzyme concentrations ratios in organisms *F* and *B*. An example of eigenvalues (i.e. λ_k , k = 1, ..., 4) calculated to evaluate the stability is given below for organism *F* with $[E_1]:[E_2] = 10:1$. The negative real parts of the eigenvalues are indicative of the asymptotical stability of the algebraic solutions to (14) - (17) viewed as the corresponding steady states in the full system (1) - (4),

$$\lambda_1 = -1.00 \cdot 10^{-3}, \\ \lambda_2 = -2.51 \cdot 10^{-3}, \\ \lambda_3 = -3.05 \cdot 10^{-4}, \\ \lambda_4 = -8.58 \cdot 10^{-5}.$$

From the comparison of the steady-state fluxes, we find that for organism F, Route II, $A \xrightarrow{E_1} B \xrightarrow{E_2} D$, is exclusively utilized in the formation of NAD when $[E_1]:[E_2] = 10:1$ and both routes 1 and 2 are active when $[E_1]:[E_2] = 1:1$. We could not calculate any reasonable distributions of steady-state metabolite concentrations and fluxes in the case of $[E_1]:[E_2] = 1:10$ which may be indicative of the biological infeasibility of this case for organism F. The modelling data for organism B, provided in Table 3, show that in all the three cases, NAD is solely produced utilizing Route I only.

Francisella tularensis nadE* knockout mutant and complementation study

To address the in vivo function of ftNadE* we used a nadE transposon mutant of F. *tularensis* subsp. *novicida* strain U112. The transposon T20 in the mutant tnfn1_pw060328p05q173 is inserted in the *nadE* gene (FTN_1278) between nucleotide (nt) 414 and nt 415. The insertion was confirmed by PCR and sequencing using the transposon-specific primer KAN-125 and the flanking primer NADE-ECO. The same approach was used to analyze the viable *nadM* transposon mutant. The transposon T20 in the mutant tnfn1_pw060328p06q193 is inserted in the *nadM* gene (FTN_0483) between nt 565 and nt 566. The insertion was confirmed by PCR and sequencing using the transposon-specific primer KAN-125 and the flanking primer NADE. The mutants tnfn1_pw060328p05q173 and tnfn1_pw060328p06q193 were obtained from a comprehensive sequenced-defined transposon mutant library(13). The PCR reactions were performed with these primers:

KAN-125, 5'-AACGCAGACCGTTCCGTGGC-3',

NADE-*Eco*, 5'-ATGAATTCTGGGCAAAGAATTGCTGAAATAACTAAGG-3', and NADM-*Sal*, 5'-ATGTCGACGGGTCATTAAAGCTTGTATTTAGGGAGC-3'. For complementation study the *nadE* gene was PCR-amplified from *F. tularensis* subsp. *novicida* genomic DNA with primers

NADE-*Ncol* (CGCGCGCCATGGGCATGAAAATAGTTAAAGATTTTAGTCCTAAAGA) and NADE-*EcoRI* (CGGAATTCTCAGAAATTAGGAGTTAAAGCTAATTTTCTCT). The

PCR fragment was digested with *NcoI* and *EcoRI*, and ligated into pKEK894 (14) digested similarly. The new resulting plasmid was named pKEK1210 (pnadE). Cloning was performed into *E.coli* strain DH5 α . *F. tularensis* subsp. *novicida* strains were grown on TSM (tryptic soybean media), TSAP (tryptic soybean agar plates) supplemented with 0.1% cysteine, 25 mg/ml ferrous sulfate, 25 mg/ml sodium pyruvate, and 25 mg/ml sodium metasulfite, or on Chamberlain's defined medium (CDM)(15). Antibiotics were used at following concentrations: kanamycin 50 µg/ml, and tetracycline 10 µg/ml. Plasmid pKEK1210 was transformed into Δ nadE mutant (tnfn1_pw060328p05q173) (13) via electroporation(16). Transformed bacteria were then plated on media selective for tetracycline. A selected clone in which the plasmid pKEK1210 was confirmed (PCR, plasmid digestion and sequencing) was denoted KKF302 (Δ nadE + pnadE).

In vivo assessment of NAD biosynthetic intermediates.

F. tularensis strain U112 and B. anthracis strain Ames were grown in rich media (Trypticase Soy Broth and Luria Bertani, respectively) to log phase ($OD_{600nm} \sim 1$) and harvested by centrifugation. To assess relative abundance of the NMN intermediate in the F. tularensis wild type and NadE* knockout mutant, cells were grown in defined medium (Chamberlain) with 200 uM nicotinamide. Perchloric acid (PCA) extraction of pyridine nucleotides was adapted from Maharjan et al. (17). The cells were washed twice with cold PBS, followed by two cycles of freezing and thawing. Cells were then resuspended in 1:50 volume of MilliQ water, centrifuged, and extracted with 1:500 volume of 50% (V/V) ice-cold perchloric acid. After brief vortexing, the mixture was kept on ice for 15 min. Cell debris was removed by centrifugation, and supernatant was neutralized with 5N KOH; the resulting supernatants were removed and the extracts adjusted to pH 6.5. Extracts were aliquoted and stored at -20°C. NMN and NaAD intermediates were identified by HPLC analysis as described for the HPLCbased assay. In order to provide an accurate quantisation of these metabolites, these samples were subjected to enzymatic depletion of NMN and NaAD content by using an excess amount of *ft*NadM and *ba*NadE that would convert both intermediates to NAD in the proper buffer conditions. Total enzymatic depletion of intermediates was tested and optimized by adding exogenous NMN and NaAD at 100 uM in the presence or absence of the enzymes. Net NMN and NaAD concentrations were determined by subtracting samples' peaks from the corresponding peaks of the enzymatically depleted samples (see SI Fig. 11 for an example of NMN quantisation). For the enzymatic depletion of NMN and NaAD, 80 uL of extracts were diluted in 200 uL of buffer containing 50 mM Hepes (pH 7.5), 10 mM MgCl₂, 4 mM NH₄Cl, 2 mM ATP, and supplemented with an excess amount of ftNadM and baNadE recombinant enzymes.

2 Tables and Figures SI Table 1. Functional reconstruction of NAD biosynthesis and salvage pathways in a set of selected bacterial genomes.

					Salvage I: from		From NaMN to NAD								
Genome	Variant	De fron	novo path n Asp to N	iway: NaMN	nicotin nicoti	amide or nate to	Rou	ıte I	Ro	oute II	Salva nic	nicotinamide ribose to NMN			
		ASPOX		QAPRT nadC	NAM ppcA	NAPRT	NaMNAT nadD	NADS	NMNS	NMN nadM	AT nadR	RNK nadR	PnuC	NMPRT	
Conventional NaMN-NaAD-N	AD pathwa	ay with d	e novo k	piosynthe	sis (a)	p							pilde		
Escherichia coli K12	1.1.1.1	2545	736	109	1851	916	640	1723			<u>4298</u>	4298	737		
Erwinia carotovora	1.1.1.1	668	3569	2442	1996	1526	3358	844			<u>4163</u>	<u>4163</u>	3636		
Pseudomonas aeruginosa	1.1.1.1	762	1005	4524	4915	4916,	4914	4917	_			1957	1958		
Pseudomonas putida	1.1.1.1	1414	1221	778	2212	4803	4745	4804				2833	2834		
Pseudomonas fluorescens Bacillus anthracis	1.1.1.1	4142	/219	2968	155	3208	/530	3207	_			7505	2549		
Vibrio cholerae	1.1.1.0	980	2724	872	1633	639	1344	1343					671		
Bacillus clausii	1.1.1.0	1509	1511	1510	429	3941	1602	2455					011		
Bacillus halodurans	1.1.1.0	1218	1220	1219	3777	3249	1326	2285							
Bacillus subtilis	1.1.1.0	2790	2788	2789	3179	3178	2567	314							
Carboxydothermus.	1.1.1.0	2555	2556	2554	539		604	1486							
hydrogenoformans	4440	0477	0476	0470	2510	2067	2009	1050	-						
Desulfitobacterium sp	1.1.1.0	2477	2470	2470	1173	916	2098	017							
Exiquobacterium sp.	1.1.1.0	271	269	270	1390	1389	2783	2691							
Geobacillus kaustophilus	1.1.1.0	2735	2733	2734	669	668	2614	2693							
Listeria monocytogenes	1.1.1.0	2014	2016	2015	2561	1084	1480	1085							
marine actinobacterium	1.1.1.0	2539	2540	2538	1403	1721	2605	1679							
Neisseria gonorrhoeae	1.1.1.0	1646	1645	1643	1749	472	1328	554							
Nocardia farcinica	1.1.1.0	1846	1845	1847	1119	1120	1404	4155							
Corynebacterium efficiens	1.1.1.0	?	366	365	2385	2412	2257	2428							
Eusobacterium nucleatum	1.1.1.0	602	601	603	1004	032	1708	1778							
Bacillus B-14905	1.1.1.0	1935	1933	1934		853	1337	852							
Leifsonia xyli	1.1.1.0	?	1096	1054		810	193	748							
Shewanella oneidensis	1.0.1.1	1244	2121	395			1087	1842						1809	
Alteromonas macleodii	1.0.1.0	2200	843	3806			1341	2177							
Helicobacter hepaticus	1.0.1.0	17	1839	1833			351	1008							
Thiomicrospira denitrificans	1.0.1.0	1610	2028	2029			1383	1142							
Pseudoalteromonas haloplanktis	1.0.1.0	790	1975	370			881	144							
Shewanella sp. PV-4	1.0.1.0	3517	3630	2462			3656	327							
Colwellia psychrerythraea	1.0.1.0	4266	1627	4670			1581	3712							
Candidatus Pelagibacter ubique	1.0.1.0	?	1302	1301			1185	93							
Helicobacter mustelae	1.0.1.0	?	1532	1533			1585	1209							
Helicobacter pylori	1.0.1.0	?	1344	1343			1325	325							
Vibrio fischeri	1.0.1.0	2094	814	2184			3340	3341					3210		
Conventional NaMN-NaAD-N	AD pathwa	ay withou	ut de nov	vo biosyn	thesis (b)			-						
Oenococcus oeni	0.1.1.1				703	47	65	240	_		<u>53</u>	<u>53</u>	1055	813	
Lactococcus lactis	0.1.1.1				966	1132	703	1134	_		2041	2041	2343		
Streptococcus pneumoniae	0.1.1.1			2093	1437	120	1590	121			90	90	1672		
Streptococcus pyogenes	0.1.1.0			150	1368	1275	226	1274					1022		
Lactobacillus plantarum	0.1.1.0				2161	473	1268	474					468		
Campylobacter coli	0.1.1.0				1049	2028	1423	127							
Campylobacter jejuni	0.1.1.0				198	533	1624	851							
Campylobacter lari	0.1.1.0				468	45	682	331							
Campylobacter upsaliensis	0.1.1.0				977	1412	936	218							
	0.1.1.0	ASPOX	QSYN	QAPRT	2960 NAM	2450 NAPRT	NaMNAT	NADS	NMNS	NMN	AT	RNK	PnuC	NMPRT	
Genome	Variant	nadB	nadA	nadC	pncA	pncB	nadD	nadE	nadE	nadM	nadR	nadR	pnuC	nadV	
Mesoplasma florum	0.1.1.0				386	591	421	193							
Mycoplasma mycoides	0.1.1.0			ļ	996	653	417	415							
Lactobacillus acidophilus	0.1.1.0		L	<u> </u>	1338	148	1525	149							
Oceanobacillus ineyensis	0.1.1.0				568	391	1985	392							
Stanbylococcus aureus	0.1.1.0				2675	1782	1469	1781							
Tropheryma whipplei str. Twist	0.1.1.0				692	623	469	405							
Tropheryma whipplei TW08/27	0.1.1.0				677	611	277	343							
Buchnera aphidicola	0.1.1.0					347	426	173							
Mycoplasma hyopneumoniae	0.1.1.0					491	375	372							
Mycoplasma penetrans	0.1.1.0					939	520	195							
Nycoplasma pulmonis	0.1.1.0			<u> </u>		483	335	623		└──┤					
oreapiasma parvum serovar	0.1.1.0			<u> </u>		280	4/0	40/	+	├					
wycopiasma genitalium	0.1.1.0						246	398						37	
Mycoplasma pneumoniae	0.1.1.0			1004			336	563				0405	2100	47	
riavobacteria sp.	0.1.1.0	uou (-)	L	1034	I	I	2549	1800	<u> </u>			2195	2196		
Francisella tularonsis Schu S4		1160	1162	1161			1		1324	022		1	164	1292	
Mannheimia succiniciproducens	0,1.2.1	1100	1102	1101		1776			1775	555	164	164	185	1202	
Actinobacillus succinogenes	0.1.2.1					ZP_00732388			ZP_00732389		ZP_00732632	ZP_00732632	ZP_00733245		

SI Table 1 caption

This condensed subsystem spreadsheet (modified from "NAD and NADP cofactor biosynthesis global" subsystem at http://anno-3.nmpdr.org/anno/FIG/subsys.cgi) shows gene patterns of NAD biosynthesis and salvage routes in a subset of representative bacterial genomes corresponding to those used for phylogenetic analysis of NadE family ("short" form without a glutamine amidotransferase domain (10) (18) (19). Rows indicate species with completely sequenced genomes integrated in The SEED database (except Actinobacillus succinogenes 130Z, draft assembly at http://www.jgi.doe.gov), while genes and assigned functional roles within a metabolic pathway are shown as columns. Proteins are represented by abbreviated SEED identification numbers, except those for Actinobacillus succinogenes where IDs are provided as in NCBI (http://www.ncbi.nlm.nih.gov/); "?" - inferred by pathway analysis but a gene is unknown (cannot be projected by homology). Genes clustered on the chromosome (operons) are outlined by matching background colours. Genes (fusion proteins) encoding more than one functional role in the subsystem are underlined. The second column reflects a classification of organisms (genomes) using a four-number variant code. Each position of the variant code reflects the presence (usually 1 and in some cases "2" to capture alternatives) or absence (0) of a subsystem module (pathway) as it appears in the top row. Functional roles are abbreviated as follows: ASPOX, L-aspartate oxidase (EC 1.4.3.16); QSYN, Quinolinate synthetase (EC 4.1.99.-); QAPRT, Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19), NAMNAT, Nicotinate-nucleotide adenylyltransferase (EC 2.7.7.18); NMNAT, Nicotinamide-nucleotide adenylyltransferase of NadM or NadR family (EC 2.7.7.1); NMNAT M, Nicotinamide-nucleotide adenylyltransferase, NadM family (EC 2.7.7.1); ADPRP, ADP-ribose pyrophosphatase (EC 3.6.1.13); NADS, NAD synthetase (EC 6.3.1.5); NMNS, NMN synthetase (EC 6.3.1.-); NAM, Nicotinamidase (EC 3.5.1.19); NAPRT, Nicotinate phosphoribosyltransferase (EC 2.4.2.11); NMPRT, Nicotinamide phosphoribosyltransferase (EC 2.4.2.12); PNUC, Ribosyl nicotinamide transporter, pnuC-like; RNK, Ribosylnicotinamide kinase (EC 2.7.1.22).

SI Table 2 | Conservation of NAD biosynthetic machinery among the 7 sequenced variants of *F. tularensis* deposited in GenBank.

<i>Francisella</i> Tularensis	nome s	De A	e novo pathwa Asp → →→NaM	ay, N	(Novel NaMN→	Niacin Salvage, Nm -) NAD	REF	
Strains	Ge	ASPOX (NadB)	QSYN (NadA)	QAPRT (NadC)	NMNS (NadE)	NMNAT/ADPRP (NadM)	NMPRT (NadV)	
FSC 198	AM286280	FTF1467c	FTF1469c	FTF1468c	FTF1259 (100)	FTF0386 (99)	FTF5134c	(20)
FTA	CP000803	FTA_1475	FTA_1477	FTA_1476	FTA_0724 (99)	FTA_0479 (100)	FTA_0613	
LVS	AM233362	FTL_1388	FTL_1390	FTL_1389	FTL_0685 (100)	FTL_0452 (100)	FTL_0579	
OSU18	CP000437	FTH_1350	FTH_1352	FTH_1351	FTH_0687 (100)	FTH_0449 (100)	FTH_0579	(21)
Schu 4	AJ749949	FTT1467c	FTT1469c	FTT1468c	FTT1259 (100)	FTT0386 (99)	FTT1534c	(22)
U112	CP000439	FTN_0694	FTN_0692	FTN_0693	FTN_1278 (99)	FTN_0483 (98)	FTN_1443	(23)
WY96-3418	CP000608	FTW_0631	FTW_0629	FTW_0630	FTW_0685 (99)	FTW_1688 (100)	FTW_0395	(24)

Genes implicated in NAD biosynthesis/salvage/recycling pathways of the eight sequenced strains of *Francisella* are shown by locus tags and genomes by GenBank Accession numbers. Enzyme abbreviations are described in the caption of SI Table 1. Numbers in brackets are the amino acid sequence identities (%) of NadE and NadM with respect to the enzymes of our reference strain, *F. tularensis* subsp *holarctica* LVS. Genes clustered in the chromosome are highlighted by the gray background. Schu S4, FSC198, and WY96-3418 belong to subsp *tularensis;* LVS, OSU18, and FTA to subsp *holarctica,* U112 to subsp *novicida;* This analysis suggests that all of these strains implement the same set of NAD pathways, including the novel NaMN \rightarrow NMN \rightarrow NAD pathway validated in this study for *F. tularensis* ssp *holarctica.*



SI Fig. 1 | Conservation of NadE* enzymes among the 7 fully sequenced variants of *F. tularensis* deposited in GenBank. Multiple sequence alignment was performed using *Muscle* (25) and edited with *Jalview* (26). Positions with a percentage of identity above 90% are coloured-coded according to *Clustalx* (27)output standards. None of the few amino acid substitutions between the seven strains affect residues (marked by blocks, see text for description) that according to our 3D analysis could substantially contribute to catalysis or/and specificity.



SI Fig. 2. NAD biosynthesis subsystem diagram. Functional roles are shown by abbreviations in boxes (as defined in Supplementary Table 1). Intermediates are shown by abbreviations in circles. Asp, *L*-aspartate; Trp, *L*-tryptophane, NA, nicotinic acid; NM, nicotinamide; NMR, *N*-ribosylnicotinamide; NaMN, nicotinate mononucleotide; NMN, nicotinamide mononucleotide; NaAD, deamido-NAD. Dashed arrows indicate multistep transformations. Thick coloured bars outline individual pathways present in *F. tularensis* (A), *B. anthracis* (B) and *M. succinoproducens/A. succinogenes* (C) as inferred by the presence or absence of respective genes (SI Table 1)



SI Fig. 3 | SDS-PAGE of F. tularensis and B. anthracis enzymes.

Analysis in 10 - 20% gradient SDS-PAGE of purified recombinant enzymes used for kinetics and pathway reconstitution studies : *Lane M*, molecular weight standards; *Lane 1, ft*NadE*; , Lane 2, *ft*NadM; Lane 3, *ba*NadE; Lane 4, *ba*NadD (5 µg each)



SI Fig. 4 | HPLC-based analysis used for kinetic studies of NadE enzymes and to monitor the entire pathway reconstitution. Standard separation of NaMN, NMN, NAD, NaAD, AMP, ADP and ATP (0.2-1 nmoles). The flow rate was 1 ml/min. Solvent A was a mixture of 0.1 M potassium phosphate, pH 6.0 and 8 mM tetrabutylammonium bromide and solvent B was 30% methanol in buffer A. B,. The solvent gradient consisted of 3 min at 95% A, 2 min at up to 30 % B, 5 min at up to 60% B, 3 min at up to 100 % B, hold for 4 minutes, and then returned to starting conditions for 3 min.



SI Fig. 5. Plots of initial rates versus substrate concentration obtained for ftNadE* (A), baNadE (B), baNadD (C), ftNadM (D). Apparent kinetic parameters for the respective pyridine nucleotide substrates were determined at saturating ATP (for NMN-NaMN adenylyltransferase activity assays) or at saturating ATP and NH₄Cl (for NaMN-NaAD synthetase assays). Data were fit to the Michaelis-Menten equation. Note that up to 20-30% substrate inhibition was observed for ftNadE* NMN synthetase activity at NaMN concentration above 3 mM. Therefore a limited data set (up to 1.2 mM) was used for the approximation by a simple Michaelis-Menten model. Fitting the entire data set (up to 6.4 mM) using a substrate inhibition model resulted in similar K_m and k_{cat} values (data not shown).

SI Table 3 | C. glutamicum and H. pylori NadE substrate preference.

Enzyme	NAD synthetase (U/mg)	NMN synthetase (U/mg)	Ratio
<i>C. glutamicum</i> NadE	3.27 ± 0.04	$(0.76 \pm 0.08) \times 10^{-3}$	4302
<i>H. Pylori</i> NadE	0.19 ± 0.01	$(0.18 \pm 0.03) \times 10^{-3}$	1055

The relative substrate preferences were assessed for NadE enzymes from C. glutamicum and H. pylori by measuring their specific amidation activities at fixed 2 mM concentrations of NaMN and NaAD by HPLC analysis.

SI Table 4. Crystal data and refinement statistics of *ft*NadE* complex

Data statistics							
Space group	$P2_{1}2_{1}2_{1}$						
Cell dimensions	<i>a</i> =50.43Å,						
	<i>b</i> =126.45Å,						
	<i>c</i> =152.54Å						
Resolution (Å)	50-1.85						
Total observations	294774						
Unique observations	79468						
Completeness (%)	$94.0(70.3)^1$						
$R_{\rm sym}^2$ on I	0.044 (0.356)						
Mean I/σ	19.8 (2.6)						
Refinement							
$R_{\rm cryst}^{3}$	0.184						
R_{free}^{4}	0.233						
Number of Protein atoms	7564						
Ligands	4 AMP. 4 PPi.						
C	8 Mg, 4 Na						
Number of waters	812						
Ave. B-factors (Å ⁻¹)							
Protein	25.5						
Ligands	19.1						
Water	33.5						
Rmsd bond length (Å)	0.014						
Rmsd bond angle (°)	1.394						

L

¹ The values in the parenthesis are for the highest resolution shells 1.92-1.85Å. ${}^{2}R_{sym} = \sum_{hkl} \sum_{i} |I_{i}(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_{i} I_{i}(hkl)$

 ${}^{3}R_{cryst} = \Sigma_{hkl} ||F_{obs}| - k |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

⁴ R_{free} is calculated the same way as R_{cryst} using randomly selected 5% of the reflections that were omitted from the refinement.



SI Fig. 6 Ligand binding in the 3D structure of NadE active site. A, The difference (Fo-Fc) electron density for the bound ligand and metal ions in ftNadE* complex structure. The map is contoured at 3σ level. B, Stereo view of superimposed catalytic site in ftNadE* (cyan) and bsNadE (magenta).

			Region	1″ "	P-loop	"		v	• P1 loo	р″			
			23 27		41								
SEC. STRUCTURE		hhhhhhhh	hhhhhhhh	eeee	hhl	hhhh	hhhh	eeee	ee	hhhhhhhhh	hhhh eeee		hhhhhhhhhhhh
<u>F tularensis</u>	MKIVKD	FSPKEYSQKLVN	ISDSCMNY.	. PAEG	GLSCCIDS	AVAAS	LAVKTG	LPTTAL	ILPSDNNQ	HQDMQDALEL	IEMLNIE.HYTI	s	QPAYEAFLASTQSFTNL
M_succiniciproducens A succinogenes		MRTAAYADYLIQ	ULENORTELY	. GMDGYTLC	GVSGGIDSA GVSGGIDSA	AVCAH	LAARTG	APVQAL	ILPAEVIS.	PEDVADAQAT	LESAGID.GQI	s	APWYDLIMQQLSPVLNS
			33 37		51								*
SEC. STRUCTURE	hhhhhhhhh	hhhhhhhhh	hhhhhhh	eee	hhh	hhhhh	hhhhhhhhhh	eeeeee	e	hhhhhhh	hhh eeeee	_	hhhhhhhhhhh
B_anthracis B_subtilis	MTLQEQIMKALHVQPV MSMOEKIMPELHVKPS	IDPKAEIRKRVD	LKDYVKKT.	GAKGEVL	GISCCODS:	PLAGE	LAQLAVEEIRNEGG	DAOFIAV	RLPYKV	OCDEDDAQLA	LQFIQADQSVAF	D	LASTVDAFSNQYENLLDE
E coli	MTLQQQIIKALGAKPO	INAEEEIRRSVD	LKSYLQTYF	F.IKSLVL	GISGCODS	LAGK	LCOMAINELRLETGNE	SLOFIA <mark>V</mark>	RLPYGV	OADEODCODA	IAFIOPDRVLT	N	KGAVLASEQALREAGI
H_Pylori	MQ	KDYQKLIVYLCD <mark>H</mark>	EK <mark>E</mark> VQK <mark>R</mark> G	FKK <mark>VVY</mark> (G <mark>LSGG</mark> LDS	A <mark>V</mark> VGV	LCQKVFK	ENAHA <mark>L</mark>	L <mark>M</mark> PSSVSM	PENKTDALNL	<mark>C</mark> EKFSIP.YTE <mark>Y</mark>	s	APYDAIFSSHFKD
L_xyli	MRDLQARIIHELNVRAT	IDPAAEVVERVD	LVRYVRAAG	ASG		LAGR	LCQLAVERLAEQGV	DATERA	RLPYAV	QNDEDDAQLA	LSFIRPERTVAV	N	QRGVEGVGNEYRDALGE
Bacillus B-14905	METMTLOOOIIEELRVOPT	IDAHEEIRKSID	LKEYAKHYS	F.VKGFVL		LTGK	LAQLAVDELNKEAG	.MKEYSFWAI	RLPYGV	OADEADAQIA	IDYIKPTGSYT <mark>V</mark>	N	KDAVDASARAVAEALGI [8]
B_clausii	MMKVNQDSIRRDLHTKAD	IDPKEEIRTRVD	<mark>71</mark> KN <mark>Y</mark> TRK <mark>A</mark> G	TKG <mark>FVL</mark> (G <mark>ISCC</mark> QDSI	r <mark>l</mark> aak	LAQMAVDELNQEAPG	TYEFHA <mark>V</mark>	R <mark>L</mark> PYGV	QHDEKDAQVA	<mark>L</mark> SFIGPDKIHT <mark>V</mark>	N	KPAVDAAVASFAEATGE
ExiGuobacterium_sP	MQQEIIQVTGVKPV	IDAEEEIKQRVQ	EKE <mark>Y</mark> LVH <mark>T</mark> G	AKGLVL	GISGGQDS	SLAGR	LCQIAVEELRSETNR	DYQFYA <mark>V</mark>	R <mark>L</mark> PYGQ	QQDESDAQLA	LSFIRPDHALR	D	KPAVAASMASFEQATGD
T whiPPlei	MCCDISKTLCVKPF	. IDPEETSHRVS	LADYLEHSE	ASGYVL		LAGR	LCOTAVESVRSIGE	DATLWAT	RUPYGO	OFDESDAOTA	MOFISPDEELSE	D	RSATDNLCVDLNRSLGS
C_Psychrerythraea	MNQQIIIAEMKVLPE	IDVQFEINRRVA	TIKK <mark>Q</mark> LVQ <mark>S</mark> G	LTNLVL	G <mark>ISGG</mark> VDSS	TCGR	LAQLAVNELNGQLDEGEI	DKTHYQFIA <mark>V</mark>	R <mark>L</mark> PYGI	QADEDDAQQA	VDFIQPSHCLT <mark>1</mark>	N	LAGADGIHHEVLQAMSKAQI
B_xenovorans_	MTHQDPAARQASISEEMHIGAT	FDAEYEIERRVT	LVS <mark>Y</mark> LRG <mark>S</mark> G	LKT <mark>YVL</mark> (GISGG VDS	TAGR	LAQLAVEQLRAEH	YEAQFVA <mark>I</mark>	RLPYGE	QKDEADAQQA	LRFIRADQNLA <mark>I</mark>	D	KPAADAMLAALDRSGVLF
P_Putida P fluorescens	MODRIARELNINRAL	UAAALEAEVARRVA VLGGEPEETORRTD	CIKDCLANAR	CKALVL		TAAL	LAQRAINELRAETGD	VAARTFIAV	RLPYQV	OADERDAQAC	LEVIKADEVHTV	D.	LAPAVRALAAEVAALKNG
W_Glossinidia	MNTQKKIIKEFNVQPN	FEEKHEFERIKN	FISTLNK	Y.LRTLVL	G <mark>V</mark> SGGQDS	r <mark>l</mark> TGK	ICYEVIKEMKKKNNN	LNYKFIAL	R <mark>L</mark> PYGK	QMDEIDCKLA	IKFINPDKVIK <mark>I</mark>	N	KSAVKASVFSLKKSGI
S_aureus	MSKLQDVIVQEMKVKKR	IDSAEEIMELKQ	TKN <mark>Y</mark> VQS <mark>H</mark> S	F.IKS <mark>LVL</mark> O	G <mark>ISGG</mark> QDSI	r <mark>l</mark> vgk	LVQMSVNELREEG	IDCTFIA <mark>V</mark>	KLPYGV	QKDADEVDQA	LRFIEPDEIVT <mark>V</mark>	N	KPAVDQSVQSLKEAGI
B_aPhidicola N Conorrhosse	MTLQKKIIELLGVKPA	MDTOAVITUTVD	UDENA AOAN	H. IKSLIV	GVSGGQDSI GVSGGTDC7	UUUCT	LCQMTAETLRKEKND	RTTLI	REPIGI	PGOLEDADDU	IRFIQPDQIFN	N	TREATESTERSLEES
Flavobacteria sP		MKVDKVTDHIVT	VLKDYAIKAN	VKG <mark>YV</mark> V	G <mark>V</mark> SGGIDS	VTSS	LCAMTG	LEVLC <mark>V</mark>	EMPIHQH	HDHVTRAQEH	IKQLKSR.FKNV	[5] DI	TPVFDTFIENMPDIKD
C_Pelagibacter		MKPLEKAQFISN	<mark>VI</mark> KD <mark>Y</mark> VNK <mark>M</mark> F	SKAQS <mark>LII</mark>	G <mark>ISGG</mark> IDSS	S <mark>V</mark> SST	LSAMTG	IKTIV <mark>L</mark>	S <mark>M</mark> PIKQKSSQI	HDLSLKHQEWL	vknfdnveaht <mark>i</mark>	NI	LDKLFETFESTLSNF
M_Pulmonis	MKSNTEIYKN	. LKLLEDYGTYLIE	VI KL <mark>K</mark> VQQ <mark>A</mark> N	KKG <mark>VIV</mark>	G <mark>ISGG</mark> IDS	ALVAC	LAKKAFP	ENSLGI	TMPIGNSM	KLDFDDIAKL	OKLTKLE.IIN <mark>I</mark>	DI	TLSYDALAKTLDV
C difficile		SNIKIOIDKTVE	LINKVNEAN	AKGLIV		VAN	LIKKAFP	ENSMGV	IMSIKSN	PODREDALKV	IEGCDIE.YLDI	DI	LIEPOSAILDMVVGNLKDKH.
Desulfitobacterium_sP		EELEIRINRAVE	vlre <mark>r</mark> vqe <mark>a</mark> r	AQG <mark>LVI</mark>	G <mark>V</mark> SGCVDS	A <mark>V</mark> VAG	LCKRAFP	HNSIG <mark>V</mark>	I <mark>L</mark> PAGSN	. PMDREDAWLA	realslk.ave <mark>i</mark>	DI	TQAHQGILASVKKALTAQ
C_hydroGenoformans	MR	VNWEEKTEKLVN	ILRE <mark>K</mark> TREAN	ASG <mark>LLV</mark> (G <mark>LSGG</mark> VDSP	A <mark>V</mark> VAT	LIKKAFP	EKSLG <mark>I</mark>	IMPCFSN	PEDEEDARMI	ANHLNLK.YIV <mark>V</mark>	NI	DEPYQALVSSLKNATP
M_hyoPneumoniae	MINK	KNILNYLNELEA	IRQEVKKAÇ	ACGVIF			LANKAFP	ONHLT	IMPIRDM	ELDHKATTAL	VEOLOLN NEOL	N	LEPPYRAMLOALTID
M Penetrans		KDYIKIIDKIAN	MNQTLTEAK	SNG <mark>FVY</mark>	G <mark>V</mark> SGGIDS	LICA	IASKFF	KDRSLAV	RLDIFNS	VNDTKDANLV	ISHFKVN.SVDF	N	LEQVFNTFIKDLP
W_succinoGenes	MTL	TNPSALVNQLVD	7LRQ <mark>E</mark> LAQ <mark>R</mark> G	FKK <mark>VVV</mark> (G <mark>LSGG</mark> VDS	A <mark>V</mark> VAR	LCQEAIG	ENLHA <mark>L</mark>	L <mark>M</mark> PSSVSS	. KESVEHALLL	<mark>C</mark> ERFNLS.HHI <mark>C</mark>	s	APLELAFRELHPE
F_nucleatum T_denitrificans	MNKLD	LNLKEVHNELVE	LRENFKKAG	FSKAVL	GLSGGIDS		LLRDALG	DNLLCV	MMPYKSSN	PDSLNHAKLV	VEDLKIN.SKTI	E	TDMIDAYFKNEKE
C uPsaliensis		MDCEKILTKIQH	IQERVANAK	AKG <mark>VIL</mark>	GLSGGIDS?	LVAT	LCKKAL	KDEVFAL	LMPTQHSN	EANLKDALKL	CEDLNLK.YKI	N	ETILQAFLKESEM
_ C_lari		MNYTKLQELLIN	TKEQARD.	EN <mark>LIL</mark>	G <mark>L</mark> SGGIDS	A <mark>L</mark> VAH	LCKKAVG	EKLFV <mark>L</mark>	l <mark>m</mark> ptkysk	. KENLDDALML	CDHLQIH.HKI <mark>I</mark>	Y	DEVLCAYEKICQD
H_mustelae	MQT	HFSPKSIQNILL	LKNTLQAKS	FHSVVL	GLSGG IDS ²	VAV	LCKHAFP	NTTLAI	SMPTLSSS	KOHLEDARIL	CEHFEIP.HIVH	s	IAPYEEIFTRNEKDFDT
T maritima		MKHFVNPCIH	ICKOLQERG	YRGAVV		VVAI	LCVOALGK	DRVFAL	ILPERDSS	KDSLKDAVDF	CERLGVE.YRKR		TPILRKIGVYRLFPPR [27]
marine_actinobacterium	MTPLQARIIYDLNTQPS	IDAAEQIRMRID <mark>H</mark>	<mark>elke</mark> ylkt <mark>s</mark> h	AKG <mark>LVL</mark> (G <mark>ISCC</mark> QDSS	S <mark>L</mark> AGR	LCQLAVAELRAEG	TEVQFVA <mark>V</mark>	r <mark>l</mark> phgv	QHDEDDAQLA	LDFIEADREVT <mark>F</mark>	N	QRAVDGIAAEFADALGE
C_michiGanensis	MRDIQSQIIDALEVRPT	IDPADEVRKRVD	LKAYLRSTG	AEG	G <mark>V</mark> SGGQDSS	LAGR	LAQLAVEELAAEG	LLAEFVAV	RLPYGV	QADEEDAQLA	LSFIQPKSSVV	D	KRAVDGFQAEYADAAGH
D radiodurans	MTPSPLPLSPLRSHIIRELHVOPD	IDPGAEVERRVA	LCDYLOSTF	TKGFVL		LAGR	LCOLAVERRRSOG	HGATFLAV	RLPIGV	OADEADAOOA	LDFIOADREVTV	N	KEAADASVAAAOAALGS
C_Glutamicum	MTNTQTEIINELKVŠPA	IDVAKEVEFRVQ	LVD <mark>Y</mark> LRASH	TKG <mark>FVL</mark>	G <mark>ISCC</mark> QDSI	r <mark>l</mark> agr	LTQLAVERIRAEENS	TDYVFYA <mark>V</mark>	R <mark>L</mark> PYAI	. QADEDDAQVA	<mark>l</mark> efiãpdksvt <mark>v</mark>	N	KDATDATEATVAAALEL
C_efficiens	MKTMTDTRDHIISQLSVVPS	IEPAAEVEARVQ	LVD YLRVSH	ARG <mark>YVL</mark>	GISGGQDSI	TAGR	LAQLAVERIRREDG	SDHQFVG <mark>V</mark>	RLPHGL	QADEDDALVA	LDFIQPDRSIS <mark>I</mark>	N	REATDVLSAATAAALGI
C diPhtheriae	MRNNITHRIGTKPT	IDPATEIRNRVE	TRUTRY TKHSC	TKGFAL	GISGGIDS GISGGODS	LAGR	LAOLAVEOLEVAG.	YPAEFWAT	RUPYGV	OADEDDAHIA	LEFIOPDHSVVI	N	KEATDAAARATAOALGL
S_Pneumoniae	MSLQETIIQELGVKPV	IDAQEEIRRSID	LKR <mark>Y</mark> LKK <mark>H</mark> F	F.LKT <mark>FVL</mark>	G <mark>ISGG</mark> QDSI	r <mark>l</mark> agr	LAQLAMEELRAETGD	DSYKFIA <mark>V</mark>	R <mark>L</mark> PYGV	QADEADAQKA	LAFIQPDVSLV <mark>V</mark>	N	KESADAMTAAVEATGS
S_Pyogenes	MTLQEEIIRQLGVKAS	IDPQEEIRKAVD	<mark>ELKA</mark> YLRK <mark>H</mark> S	F.LKT <mark>YVL</mark> (G <mark>ISGG</mark> QDSI	r <mark>l</mark> agk	LAQMAIAELREEASD	QAYQFIA <mark>V</mark>	R <mark>L</mark> PYGV	QADEADAQKA	LAFIAPDQTLT <mark>I</mark>	N	KAAVDGQVEALQAAGV
L_lactis E faecalis	MTLQDEIIKELGVKPV	IDPKEEIRVSVD	LKDYLKKYP	F.IKSFVL		LAGR	LAQIAIEEMRQETAD	MSVOFIAT	RLPYGV	OADEEDAQRA	LAFIQPDVSLT	N.	KAAVEGQVAALNEAGI
L_Plantarum	MRPLQAEIIKALHVAPT	IDPEVEIRRSID	LKA <mark>Y</mark> LTK <mark>N</mark> T	F.LKT <mark>YVL</mark>	G <mark>ISGG</mark> QDS	LAGK	LTEMAITEMRQETGD	DRYQFIA <mark>V</mark>	R <mark>L</mark> PYGN	QADEADAMAA	IDFMQADVTDR <mark>V</mark>	D	QPATDAMVTALEANQL
P_Pentosaceus	MRELQKEIIETLKVQPT	IDPKQEIRRSVD	MKN <mark>Y</mark> LKK <mark>V</mark> F	S.LRS <mark>LVL</mark>	G <mark>ISGG</mark> QDSI	r <mark>L</mark> AGA	LAEIAVRELREETGI	QDYRFIA <mark>V</mark>	R <mark>L</mark> PYGV	QADESDAMKA	IEFMQADEVVR <mark>I</mark>	D	KPAADAMVAALEANDL
L_acidophilus	METRERILADMOVART	IDPKKEIRRSID	LKDYLKANP	F. LKSTVL		LTGK	LOMAISEMREETGD	DEYRFFAV	SUPYGT	QADAADAADA OLDESDRODA	NEMEPDNRI TU	N	KASVDAMVKVVEATGQ
E_carotovora	MSLQNDIITALGVKSS	IDPAQEIRVSVD	LKNYLNAHF	F.ITSLVL	G <mark>ISGG</mark> QDSI	LTGK	LCQTAITELRNETGT	SRYQFIA <mark>V</mark>	R <mark>L</mark> PYGV	QADEADCQDA	IAFIQPDRVLT	N	KPAIESSEATLRAIGV
0_iheyensis	MSDLQKTIIEDLHVEPE	IDPKQEIRTRVD	LKS <mark>Y</mark> LKK <mark>H</mark> T	F.STG <mark>YVL</mark>	G <mark>MSGG</mark> QDS1	r <mark>l</mark> lsk	LTQIAVNELNEENST	NTYQFIG <mark>M</mark>	K <mark>L</mark> PYGV	QKDADDVDDA	IKFVEPSKVLT <mark>V</mark>	N	KESVDASVKALDGAGV
B_halodurans	MPRIOSKIIADLHVORV	VEPKEEIRKRVT		AKGYVLO		LAGK	LAQMAIDELNEEEQD	TSYVFIAV	RLPYGV	OCDENDAODA	IAFIKPSRSIT	N	KDAVDASTKSFEQATGE
V fischeri		IDPVEEIKKRVD	IKGKLLEAH	CKSLIL	GISGCVDS	TCGR	LAQLAVNELNLETOS	SDYOFIAV	RLPYGI	OODEDEAOLA	LOFIOPTHSISI	N	KNGVDGLHSANHIALKDTGL
V_cholerae	MEHKIREEMRVLPS	IDPQFEIERRVA	71 KR <mark>K</mark> LTE <mark>A</mark> R	CKS <mark>LVL</mark> O	G <mark>ISGG</mark> VDSI	r <mark>t</mark> cgr	LAQLAVEELNQQHNT	TEYQFIA <mark>V</mark>	R <mark>L</mark> PYGE	QKDEDEAQLA	LSFIRPTHSVS <mark>V</mark>	N	IKAGVDGLHAASHHALANTGL
Shewanella_sp	MKGQIIREMKVQPH	IEVEYEVQRRIA	IKAKLKEAR	ATSLVL	GISGGVDSS	LAGR	LCQLAVDELNSEGEYE.	GSYQFIAV	RLPFKV	QKDEDEAQMA	CQFIQPSKLVTV	N	GEGVEGIHHQTLAGLEVAGV
P haloplanktis		IDVNAEISRRVN	IKARLIAAH	ATSLVL	GISCCVDS	STAGR SVCGR	LCQLAVNELNQEQSTT	DYKFVA <mark>V</mark>	RLPYGV	QADENEAQQA	VDFIQPSSRMTV	N	KPATDALHEQTMAAIVGNGE
_ A_macleodii	MQKQAVIDEMKVLPE	IDVEYEVARRVS	71KK <mark>Q</mark> LLT <mark>S</mark> G	LNS <mark>LVL</mark>	G <mark>ISGG</mark> IDS	TLGR	LAQLAVDELNEEHHET.	YQFIA <mark>V</mark>	R <mark>L</mark> PYDT	QADEEDAQKS	<u>I</u> DFIQPTHSIA <mark>V</mark>	N	/KPGADAIHASTSQALSDANL
S_coelicolor	MSEPASIALQKEIVRELEVAET	FDAEREIERRVA	LAE <mark>R</mark> LTS <mark>T</mark> G	LRALVLO	GISGC VDS	TAGR	LCQLAVERARAAG	HEALFYA <mark>M</mark>	RLPHGV	QADEDDAQQA	LSFIRPDRVLT	D	KPASDAALDALLAADVAF
B_Pseudomallei B_cenacia	MSKPDQAARRAIAAELHVSPT	FDARDEAERRIG	ARYLR AC	LRACVLO	GISCCIDS GISCCVDS	TAGR	LAQLAVERLEASG	YDARFVAM	RLPYGA	ONDEADARRA	LAFVRADETLTV	D	KPAADAMLAALAAGGLAY
R_leGuminosarum	MTVLFDEQGEIIRELGVAAD	IDPEREIERRTA	LKD <mark>Y</mark> LVASG	MRG <mark>YVL</mark>	G <mark>ISGG</mark> VDSI	TAAL	LAQKAVRELRDSG	HAAEFIA	RLPYGV	QADEADAVKA	LETIGADRSMVV	N	KAPADAMLAAAQDGGLAF
P_syringae	MHAVQRQIAEQLKVQPPFA	DQNALQAEVARRVS <mark>I</mark>	7IKE <mark>C</mark> LQN <mark>A</mark> R	LKT <mark>LVL</mark> (G <mark>ISGG</mark> VDSI	. <mark>T</mark> AGL	LAQRAVKELRASTGD	NSYRFIA <mark>V</mark>	R <mark>L</mark> PYVV	QADEHEAQAS	VDFIEPDERHT <mark>I</mark>	N	GSSVKALAAEVKAFDGL
P_aeruginosa M florum	MQQIQRDIAQALQVQPPFQ	MELKEVIDVIJ	CLQQCLKDSG	LKTLVL	GISCOVDSI	AGL	LAQRAVEQLREQTGD	QAYRFIA <mark>V</mark>	RLPYQV	DEDVKCKOPT	LATIRADEEQT	N	LGPSVKALAEQLEALEGL
G_kaustoPhilus		MQEKIDKLVQ	VLRDQVSSAG	LNGAVV	G <mark>ISGG</mark> IDS	AVVAL	LIKRAFP	DDSLGL	IMPCKSN	PKDMEDALKV	VKSCGIR.HLVI	D	TEAHRTLFGAVEAELKAIG.
M_Genitalium	M	TNLIKYLKELQN	VLFD <mark>Y</mark> VKK <mark>S</mark> K	AKG <mark>VIF</mark> C	G <mark>LSGG</mark> IDS?	A <mark>V</mark> VAA	IAKETFG	FENHLA <mark>L</mark>	I <mark>M</mark> HINNS	KLDFQATSEL	<mark>V</mark> KKMQFN.SIN <mark>I</mark>	E <mark>I</mark>	LEESFNLLVKTLGID
C_jejuni		MDWQKITEKMCD		SQG <mark>VVL</mark>	GLSGGIDS ²	LVAT	LCKRALK	EMUEA	LMPTQISN	OENLNDALET	CADLNLE.YKI	E	LQS1LDAFIKQSEN
C_0011			N		AAA MA				A	A		· · · Þ	Soloon proops

						FZ 1	00₽		
	124	133	142 147		171	193	202		233 236 239
		100				100			
SEC. STRUCTURE	hhhhhhhhhhhhhhh	hhhhhhh eee	e hhhhh h	eee	hhhhhhh	hhhh	1	hhhhhhh	hhhhhhhhhhhh
F tularensis	QNNRQLVIKGNAQARLRMM	LYAYAQQYNRIVI	GTDNACEW. YM	GYFTKFGDGAADILPLV	N <mark>LKK</mark> SQVFE <mark>L</mark> GKY <mark>L</mark> I	VPKN <mark>IL</mark> DKAPSAGLWQ	GQTDEDEMGV	TYQEIDDFLDGK	QVSAKALERINFWHNRSHHKRKLALT[3]
M succiniciProducens	EPERVNVLKGNLMARLRMIA	ALFTTAOSHRSIVL	TONAAEW.LT	G <mark>YFTKFG</mark> DGAAD <mark>V</mark> LP <mark>L</mark> A	GLRKEQVFELGRYLG	VPOS <mark>VL</mark> DKKPSAGLWA	GOTDEAEMGV	TYAEIDAYLRGE	TVSPOALOOIRFWHNRSHHKRMLPPK[9]
A succinoGenes	ESERIN <mark>V</mark> LKGNLMARLRMI	AL <mark>F</mark> TTA <mark>O</mark> SHRS <mark>IVL</mark>	TDNAAEM. LT	G <mark>YF</mark> TK <mark>F</mark> GDGAAD <mark>V</mark> LP <mark>L</mark> A	R <mark>L</mark> RKEQVFE <mark>L</mark> GRY <mark>L</mark> G	VPKS <mark>VL</mark> EKKPSAGLWA	GQTDEGE <mark>MG</mark> V	S <mark>Y</mark> AE <mark>IDAYL</mark> RGE	TVSPQALKQIQF <mark>W</mark> HN <mark>R</mark> SHHKRMLPPT[8]
-	138	149	158 163		187	209	221		252 255 258
SEC. STRUCTURE	hhhhhhhhhhhhhh	hhhhhhh eee	e hhhhh h	eee	hhhhhhh	hhhh	hhhh	hhhhhhh	hhhhhhhhhhh
B anthracis	SLTD <mark>F</mark> NKGNVKARIR <mark>M</mark> VI	rQ <mark>Y</mark> AIG <mark>G</mark> QKGL <mark>LVI</mark>	GTDHAAEA. <mark>V</mark> T	G <mark>FF</mark> TK <mark>FG</mark> DGGAD <mark>L</mark> LP <mark>L</mark> T	G <mark>L</mark> TKRQGRA <mark>L</mark> LQE <mark>L</mark> G	ADER <mark>LY</mark> LKMPTAD <mark>LL</mark> D	EKPGQADETE <mark>LG</mark> I	T <mark>Y</mark> DQ <mark>LDDYL</mark> EGK	TVPADVAEK <mark>I</mark> EK <mark>R</mark> YT <mark>V</mark> SEHKRQVPAS[7]
B_subtilis	QLTD <mark>F</mark> NKGNVKARTRMI	AQ <mark>Y</mark> AIG <mark>G</mark> QEGL <mark>LVL</mark>	GTDHAAEA . <mark>V</mark> TO	G <mark>FF</mark> TK <mark>YG</mark> DGGAD <mark>L</mark> LP <mark>L</mark> T	3 <mark>L</mark> TKRQGRT <mark>L</mark> LKE <mark>L</mark> G	APER <mark>LY</mark> LKEPTAD <mark>LL</mark> D	EKPQQS DE TE <mark>LG</mark> I	S <mark>Y</mark> DE <mark>IDD</mark> YLEGK	EVSAKVSEA <mark>L</mark> EK <mark>R</mark> YS <mark>M</mark> TEHKRQVPAS [7]
E_coli	ELSD <mark>F</mark> VRGNEKARER <mark>M</mark> K#	AQ <mark>Y</mark> SIA <mark>G</mark> MTSG <mark>VVV</mark>	GTDHAAEA. I'T	G <mark>FF</mark> TK <mark>YG</mark> DGGTD <mark>I</mark> NP <mark>L</mark> Y	R <mark>L</mark> NKRQGKQ <mark>L</mark> LAA <mark>L</mark> A	CPEH <mark>LY</mark> KKAPTAD <mark>L</mark> ED	drpslp de va <mark>lg</mark> v	T <mark>Y</mark> DN <mark>ID</mark> DYLEGK	NVPQQVART <mark>I</mark> EN <mark>W</mark> YL <mark>K</mark> TEHKRRPPIT[6]
H_Pylori	ASL <mark>T</mark> RKGNFCARLRMAF	7L <mark>Y</mark> DYS <mark>L</mark> KSDS <mark>LVI</mark> (GTSNKSER. <mark>M</mark> LO	G <mark>YG</mark> TL <mark>FG</mark> DLACAINPIG	e <mark>l</mark> fktevye <mark>l</mark> arr <mark>l</mark> n	IIPKK <mark>IL</mark> NKP PS AD <mark>LF</mark> V	GQSDEKD <mark>LGY</mark>	P <mark>Y</mark> SV <mark>IDPLL</mark> KDIE [1]	5] GYDEILVKN <mark>I</mark> TS <mark>R</mark> IQ <mark>K</mark> NAFKLELPAI [7]
L_xyli	DMTD <mark>F</mark> AKGNVKARVR <mark>M</mark> VF	AQ <mark>Y</mark> AIA <mark>G</mark> QRRL <mark>LVV</mark>	GTDHAAEA . <mark>V</mark> TO	G <mark>FY</mark> TK <mark>YG</mark> DGGAD <mark>L</mark> LP <mark>L</mark> S	G <mark>L</mark> SKRQGRA <mark>L</mark> LQH <mark>L</mark> G	apar <mark>ly</mark> ekaptad <mark>ll</mark> d	QSPGQTDEAN <mark>LG</mark> L	R <mark>Y</mark> SD <mark>IDD</mark> FLEGK	DVAEKVAIA <mark>I</mark> EA <mark>R</mark> YL <mark>A</mark> TEHKRRVPAS[13
D_geothermalis	QPETTPHDPLRD <mark>F</mark> VRGNIKARERMV <i>F</i>	AQ <mark>Y</mark> AIA <mark>G</mark> QENL <mark>LVV</mark>	GTDHAAEA . <mark>L</mark> T	G <mark>FFTKY</mark> GDGGVD <mark>L</mark> TP <mark>L</mark> T	G <mark>L</mark> TKRQGAQ <mark>L</mark> LAF <mark>L</mark> O	GAPES <mark>TW</mark> RKVPTAD <mark>L</mark> ED	drpglp de va <mark>lgv</mark>	T <mark>Y</mark> AQ <mark>IDAYL</mark> EGR	AVSPEVAAR <mark>L</mark> ER <mark>L</mark> YL <mark>A</mark> TRHKRALPVT[8]
Bacillus_B-14905	QLND <mark>F</mark> VKGNEKARERMK	/Q <mark>Y</mark> SIA <mark>A</mark> MNGA <mark>VVL</mark>	GTDHAAEA . <mark>I</mark> T	G <mark>FY</mark> TK <mark>FG</mark> DGGAD <mark>L</mark> MP <mark>I</mark> F	R <mark>L</mark> NKRQGKQ <mark>L</mark> LAE <mark>L</mark> F	(CPEH <mark>LY</mark> MKVPTAD <mark>L</mark> EE	NRPSLP DE VA <mark>LGV</mark>	S <mark>Y</mark> DQ <mark>ID</mark> DYLEGK	EIPEEPRQL <mark>L</mark> EG <mark>Y</mark> YL <mark>R</mark> SQ <mark>HKRHMPIT[7]</mark>
B_clausii	ELSG <mark>F</mark> IKGNTKARERMKV	/Q <mark>F</mark> DLA <mark>A</mark> HYQC <mark>LVL</mark>	GTDHAAEA . <mark>V</mark> T	GFFTKFGDGACDVIPLY	G <mark>L</mark> TKRQGKA <mark>L</mark> LQE <mark>L</mark> O	SAPEA <mark>LY</mark> QKI PTAD <mark>L</mark> ED	DKPGLP DE EA <mark>LGM</mark>	T <mark>Y</mark> EQ <mark>IDDYL</mark> EGK	PIDSAIQEK <mark>L</mark> EQ <mark>R</mark> YA <mark>A</mark> TAHKRKDPVS[8]
ExiGuobacterium_sP	VLSD <mark>F</mark> SKGNTKARER <mark>M</mark> KV	/Q <mark>Y</mark> DIA <mark>A</mark> HYGC <mark>LVV</mark>	GTDHAAEF . <mark>V</mark> T	G <mark>FYTKHG</mark> DGACD <mark>L</mark> TPLT	G <mark>LNKRQGKQ</mark> LLRQ <mark>L</mark> Ç)APEG <mark>LI</mark> EKVPTAD <mark>L</mark> ED	NQPGLP DE QA <mark>LGM</mark>	T <mark>YNEIDDYL</mark> EGK	TISAESQAKLEAQYKRVGHKHHMPVS[7]
0_oeni		AQ <mark>Y</mark> AVAREHDG <mark>VVL</mark>	GTDHAAEA . <mark>F</mark> A	.G <mark>FFTKY</mark> GDGGTDLDPLW	R <mark>L</mark> DKSQGQQ <mark>M</mark> LKA <mark>L</mark> N	IAPES <mark>LY</mark> NKV PS AD <mark>L</mark> ED	KRPQLP DE VA <mark>LGV</mark>	KYKDIDKYLEGR	EVSEEAAKQIEKLYLTTKHKRHLPVT[7]
T_whiPPlei	KISD <mark>F</mark> NRGNIKARLR <mark>M</mark> VV	/Q <mark>Y</mark> AVAAHHDALVV	GTDHAAEA. <mark>V</mark> T	GFFTKFGDGAADILPLY	G <mark>L</mark> TKGQGRALLKALC	GACDSIIEKVPTADLLD	DLPCLPDETELGL	Q <mark>Y</mark> RDIDAFLEGK	PVSEDITMAITERYKSTLHKRMPPIT[8]
C_Psychrerythraea	LTSSDAQIDFSKGNVKARARMVS	SQ <mark>Y</mark> HIA <mark>G</mark> IIGALVI	GTDHSAEN. IT	GFFTKWGDGACDLAPLF	G <mark>L</mark> SKRQVRAIAKELC	GAPSI <mark>LV</mark> DKAPTADLEE	LEPGKTDEDALGI	SYEQLDNFLEGK	QVTTAVSEHIINIYKKTQHKRQAIPT[3]
B_xenovorans_	NDESHQDFVHGNIKARQRM17	AQYAVAGARAGVVI	GTDHAAES . VM	IGFFTKFGDGGADVLPLA	GLNKRRVRAVAKALC	SAPEELAHKVPTADLEM	LRPQRPDEDAYGI	PYDAIDDFLEGK	AVSDAARATILRYYDVTRHKRALPYT[1]
P_Putida	SPILVDFVVGNVKARIKMV	AQYTIAGARAGLVI	GIDHAABA. VM	IGFFIRFGDGACDLAPLS	GLVKNQVRALARSPO	JAPES LVEK VPTADLED	LEPGKPDEASHGV	TYAQIDAFLHGQ	PVDQAAFDIIVATYRKTQHKRELPFAP.
P_fluorescens	AEHVDFIKGNVKARTRMI	AQYAVANLHNGLVV	GTDHGAEA. LM	IGFFTKFGDGACDLAPLS	G <mark>L</mark> TKTQVRLLASALC	APVNLVHKPPTADLEE	LAPGKLDEHAYAC	TYEEIDAYLMGE	PVSERVKNIVERAYLKTAHKRALPIAPL
W_Glossinidia	NITDHIKGNEKSRERMK	QYSVAGMTSGIVV	GTCNAAES. TT	GFFTRHGDSAVDLNPIL	HUNKRQVKKILIYUP	CPKRLYSKI PTADLED	DRPGYPDERSLGI	KYDVIDDYLEGK	IINLKMKNIIENLFERTKHKRINQAL[7]
S_aureus	VLTDFQKGNEKARERMK	VQFSIASNRQGIVV	GTDHSAEN. TT	GFYTKYGDGAADIAPIF	GLNKRQGRQLLAYLC	GAPKELYEKTPTADLED	DKPQLPDEDALGV	TYEAIDNYLEGK	PVTPEEQKVIENHYIRNAHKRELAYT[7]
B_aPhildicola			GIGHAAEN. 1	GFFIRIGDSGIDINPIA	NUNKRQIRLLIKNLP		EHPQQDDESVLGV		DODDDDU DIVERSION DATE AND DE DE CONTRACTORIO DE C
N_GONOFFHOeae			GIGNKVEDFGV	GFFIRIGDGGVDISPIA	DLIKIQVIRLAHALO	VDEATOKAPPIDGLWD	IERIDEEQMGA		PGRRREVLEITIRLHRAMOHKINPIPV[9]
C DoloCiboator	DOPT CHANCEAD TEMP		GTONKVEDFOV	CPVTVVCDCCVDTCDTA	DONKORVIQIGKIVI	ITLORITDAAPBDGLIG	DCDTDECOLOL	VVEDI REAMIN	J NGREEVERSTEINENRINGEREINETEV [6]
C_relaGibacter M Bulmonia	VNVL AVANTEDDI DMA		GTONLORW VI	CVETEVCDCCVDLL DIS	VLTKERVICLACIV	UDKCIICAAPIDGLWD	NOPDERELCY		OTDOOTATETEROUOMTEURPOLASK [1]
M_Fuimonis	FYDGI I ATGNAKADI DMTT		CTONLORW UT		ULL KORVERNARTIN	NDEL TINEKERAGINE	COTDECETCE		NND DVI VKDTDVI UKT SKUKDSI A TK [6]
C difficilo	I VERVI VATENNI DA DUDACI		OTDINIDART UT	CVPTVPCDCCVDTIDIA	TT THE PULLED	WUEDI INKA DEACINE	COTDEDEDEMOT		EUROPERT TERIUR CEUROPERTAL
C_dillicite	DIRECIDAM DRARVENS	TTTTTM NDGI DVV	GIDNMAEL.HI	GILIKLODGCADIPLY	NTINGE VIEWAKELO	WIEDDIWKAPSAGLWE	GŐTDEDE <mark>MGI</mark>	TINNII MANDEGRE	Evennugei <mark>i</mark> en <mark>u</mark> dR <mark>u</mark> SEHKRNIPAQ[4]

		N NA	N	N	A M	NI	NN N	1	A			M	N M				~~		NN	
C coli		LSRGNFAARTRM	SLLYDYS	ALKKAL	VVGTSNKSEL	LLGYG	TIYGDLAYA	FNPIGE	LYKSEIYG	LARYL	NLNEKF	IOKAPSAD	LWIGOSDEKDLG	FSYELIDK	LKALF	7] ENLDKKI I.D	MLOS	IKN	NAFKRNMPET	[5]
C jejuni		VSLGNFAARTRM	SLLYDYS	ALKNSL	VIGTSNKSEL	LLGYG	TIYGDLACA	FNPIGS	LYKSEIVA	LAKYL	NLHENF	IKKAPSAD	LWENOSDEADLG	SYAKIDE	LKALF	71RTLDPSI.TA	MLKN	MOK	NAFKGKMPETT	LET
M Genitalium	PKKDF	LTAGNIKARLRM	ITLYAYA	OKHNFL	VLGTGNFVEY	TLGYF	TK <mark>WG</mark> DGACI	IAPLAW	LKEDVYK	LAKHF	NIPEIV	ITRAPTAS		TYKELDOY	LKGDL.	ILSSEKOK	IVLD	KAK	AEHKHNSPLK	191
G kaustophilus	EWSEERAR	<mark>L</mark> GDANTRARLR <mark>M</mark>	TTL <mark>Y</mark> AVA	NYGYL	VVGTDNAAEW.	HTGYF	tk <mark>y</mark> gdggve	DLVPLIH	TKGEVRE	MGRLL	GVPEE <mark>I</mark>	IKKAPSAG	LWEGQTDESE <mark>M</mark> G	TTYEMIDKy	LKGE	DIPERDRK	IIER	HER	SH <mark>H</mark> KRQLAIA [[4]
M_florum	TPEHK	<mark>L</mark> AIANAKARLR <mark>M</mark>	TTL <mark>Y</mark> TVA	O TNSYL	VLGTDNLDEW.	. <mark>H</mark> IG <mark>YF</mark>	TK <mark>FG</mark> DGGVI	MVPLVH	LKREVRE	<mark>A</mark> ARI <mark>L</mark>	GVPTS <mark>I</mark>	INRAPTAS	LWEDQTDESELG	ITYDQIDA <mark>Y</mark>	LAGE	INDENVKS	R <mark>V</mark> DH	HKI	SE <mark>H</mark> KRNGAVA [[7]
P_aeruginosa	EPAKSD	<mark>F</mark> VIGNIKARIR <mark>M</mark>	VAQ <mark>Y</mark> AIA	GARGG <mark>L</mark>	VIGTDHAAEA.	. <mark>V</mark> MGFF	TK <mark>FG</mark> DGACI	DLAPLSG	<mark>l</mark> akhqvra	<mark>L</mark> ARA <mark>L</mark>	gapen <mark>l</mark>	VEKI PTAD	<mark>L</mark> EDLRPGHP DE AS <mark>H</mark> G	VT <mark>Y</mark> AEIDA <mark>E</mark>	LHGQ	PLREEAAR	VIVD	YHK	tq <mark>h</mark> krelpkap	۶
P_syringae	PASSVD	<mark>F</mark> VLGNTKARMR <mark>M</mark>	VAQ <mark>Y</mark> TVA	GAYQG <mark>L'</mark>	VIGTDHAAEA.	. <mark>V</mark> MGFF	TK <mark>FG</mark> DGACI	DLAPLSG	<mark>l</mark> vknqvra	IARH <mark>F</mark>	gapes <mark>l</mark>	<mark>V</mark> EKVPTAD	<mark>L</mark> EDLSPGKP DE AS <mark>H</mark> G	VTYAEIDA <mark>B</mark>	LHGE	PVREEAFK	ICE	YAK	TQ <mark>H</mark> KRELPYAP	؛
R_leguminosarum	ADAGRQD	<mark>F</mark> ILGNIKARQR <mark>M</mark>	IAQ <mark>F</mark> ALA	ALGS <mark>L'</mark>	VIGTDHAAEA.	. <mark>V</mark> MGFF	TK <mark>F</mark> GDGAAI) <mark>ILPL</mark> AG	<mark>L</mark> NKRRVRL	<mark>L</mark> AKR <mark>L</mark>	GAPDE <mark>L</mark>	<mark>V</mark> FKV P TAD	<mark>l</mark> edqrplrp de ea <mark>y</mark> g	/SYDEIDDE	LEGK	PVGEIARR	RILA	YRA	TA <mark>H</mark> KRALPVA [[4]
B_cepacia	ETAAQQD	<mark>F</mark> VHGNIKARER <mark>M</mark>	IAQ <mark>Y</mark> AVA	GARRGI	VIGTDHAAES.	. LMGFF	TK <mark>FG</mark> DGGAI	D <mark>VLPL</mark> AG	<mark>L</mark> SKRRVRA	<mark>V</mark> ARA <mark>L</mark>	GGDEL <mark>I</mark>	<mark>V</mark> MKVPTAD	<mark>l</mark> eelrplrp de ha <mark>y</mark> g	VTYDEIDDE	LEGK	PVSDNVYE	T <mark>V</mark> LR	YDG	SR <mark>H</mark> KRALPYT [[9]
B_Pseudomallei	LDHAQQD	<mark>F</mark> VLGNI KARER <mark>M</mark>	I AQ <mark>Y</mark> AVA	A <mark>G</mark> ARNG <mark>V</mark>	VIGTDHAAES.	. <mark>V</mark> MGFF	TK <mark>FG</mark> DGGAI) <mark>V</mark> LP <mark>L</mark> AG	LTKRRVRA	LARM <mark>L</mark>	GADEP <mark>L</mark>	VLKTPTAD	<mark>l</mark> etlrpqrp de ha <mark>y</mark> g.	IT <mark>Y</mark> EQ <mark>IDD</mark> E	LEGK	PMDDAVAE	T <mark>V</mark> LR	YDA	TR <mark>H</mark> KRALPYT [[9]
S_coelicolor	RDPHHQD	<mark>F</mark> VHGNI KARQR <mark>M</mark>	I AQ <mark>Y</mark> AVA	A <mark>G</mark> AHGG <mark>L</mark>	VVGTDHAAEA.	. <mark>V</mark> SGFF	TK <mark>FG</mark> DGAAI	D <mark>V</mark> VP <mark>L</mark> TG	LTKRRVRA	VGDA <mark>L</mark>	GAPAA <mark>L'</mark>	VRKVPTAD	<mark>L</mark> ETLDPGKA DE DA <mark>L</mark> G	VTYEEIDDE	LEGK	PVTEQVFE	T <mark>IVT</mark>	YRQ	TD <mark>H</mark> KRRLPIAF	?
A_macleodii	LPESEAKQD	<mark>F</mark> VKGNVKARTR <mark>M</mark>	VIQ <mark>Y</mark> EIA	ACMVDGL	VLGTDHSAEN.	. <mark>ITGFY</mark>	TK <mark>Y</mark> GDGACI)LAPLFG	LSKRQVRA	VAAH <mark>L</mark>	GAPHN <mark>V</mark>	ITKAPTAD	<mark>l</mark> eslspqka de qa <mark>l</mark> gi	MSYDQIDD <mark>H</mark>	LEGK	HVSAEVEE	K <mark>L</mark> LY	YER	TQ <mark>H</mark> KRVPIPT [[5]
P_haloplanktis	SLPEQEKID	FIKGNVKARQR <mark>M</mark>	IAQ <mark>Y</mark> EIA	AFCQG <mark>L</mark>	VVGTDHSAEN.	. <mark>ITGFY</mark>	TK <mark>FG</mark> DGACI	DLAPLFG	LSKRQVRA	LGST <mark>L</mark>	GASSV <mark>L</mark>	V NKAPTAD	LESDRPGLTDEEALG	LSYEQIDDE	LEGK	PVTQQVEQ	T L SA	YQR	TQ <mark>H</mark> KRQPVPT [[5]
S_oneidensis	TTPDAAKVD	F IKGNVKARMR <mark>M</mark>	ltaQ <mark>Y</mark> ELÀ	C VGGL	VVGTDHSAEN.	. TTGFY.	TKWGDGACI	JLAPLFG	LN KRQVRQ	LAAY	GAPESL	V 1 KAPTAD	LEDNKPLLEDEVALG	LUAEOIDDE	BEGK	VVDKAVEE	KLIN	YKA	TQHKRQPIPTI	. YĎ
snewanella_sP	ISHPHSNVD	FVKGNVKARMRM	TAQYEIA	GLTGGL	VVGTDHSAEN.	TIGFY.	TRWGDGACL	JLAPLFG.	LSKRQVRQ	LAAAL	GAPSVL	VDKAPTAD	LECDKPQLEDEVALG.	LITYDQ IDDI	LEGK	PVDAAVEA	RLIA	YNA	TOHKKKPIPTL	JYD
V_Cholerae	IPSDPAKVD	PIRGNVKARARM			VLGIDHSAEN.	TEGPY	TREGDGACL	LAPLFG.	LINKKQVRL	LAKTL	GAPEQL	VIKIPTAD	LEELAPQKADEAALN	TROIDDE	BEGK	AVPAEVSQ		THA	TOHKROPIPTI	. ID
V_fischeri	UFIDSAKID	T V CON V KARARM	UNOVETA	GVUCCI	UL OTDUCARN.	TTOPI.	TRECOGACE	DAPLEG.	NUCOUDI		CADEOL	VARVPIAD	LEELAPQADEDALS		EDGK	AUDADAED		1QM	TOURDODIDTI	. 1 D
L_mesenceroldes		DUVCNUVADADM	TYNNA	OVUCCI I	UL OTDUCARN	TTOPY	TRECOCICI	TADI PC	NEDOVDE		CADEOL	UKKUDTAD	LEELADOKADEDAL		ELGR	DVANIDAE		VOM		. / J
L mesenteroides	vuou	MCDCCTVDVVDM		PEUNCA	VICTOURADA.	VACEE	TRACDGACL	TNDLWD	NIKDOGRO		UNDVUL	VDEVDTAD			LEGK	DUANTRA		VL		. / J [7]
B halodurans	VLSD	FNKGNTKARERM	KAOVDVG	AHYGOL	VIGTOHAARA	TTGFF	TKHGDGACI	VAPLEG	TKROGKS	LLKEL	GAPTHL	VTKAPTAD	LEDDRPGLPDEEALG	TYFOLDDY	LEGK	OVHDATEK	KTES	VIA	TEHKROLPVT [[7]
0 ihevensis	RISD	FLEGNERARERM	KAOVSVA	AFNOR	VIGTOHAARA	VTGEV	TKHGDGACI	LAPLEG	NKROGKO	MLVAL	NCPEHL	VNKKPTAD	LEDDRPALPDREALG	TYEOTDER	LEGK	EVAEDSKR	TEG	VI	TMHKREGEVT	[7]
E carotovora	FLSD	FVKGNEKARERM	KTONGTA	GMNAGL	VVGTDHAARA	VTGEE	TKYGDGGTI	TNPTER	NKROGKA	LLREL	GCPSHL	VTKAPTAD	LEEDRESLEDEVALC	TYEKTOD	LEGK	PTDANDAA'	TEN	VRK	TEHKERPETT	[7]
L monocytogenes	ELSD	FAKGNEKARERM	KVOVATA	AMHKGW	VVGTDHSARA	VTGEY	TKYGDGGTE		NKROGKA	LIKEL	GCPEHL	VLKKPTAD	LEDNKPALPDEVALG		LEGK	EVPADAAA	KTEN	FT	TEHKRHMATT	[71
L acidophilus	KTTD	FNKGNIKARORM	VVOYATA	GANNGA	VVGTDHAAEN	FSGFY	TKYGDGAAL	LTPLFR	DKROGKM	LLKEL	GCPKHL	YEKAPTAD	LEEEKPDLPDEVALG	TYEEVDDY	LEGK.	EVSEKAAD	DIEK	WNK	SKHKRHLPVT	(71
P Pentosaceus	KISD	FNKGNI KARÊR <mark>M</mark>	IAQ <mark>Y</mark> GIA	GSNAGI	VIGTDHAAES.	. VTGFY	TK <mark>F</mark> GDGGAI	D <mark>IIPL</mark> WR	LDKRQGKA	MLKEL	NAPVEL	YEKTPTAD	LEEERPALPDEVALG	VTYDDIDN <mark>Y</mark>	LEGR	EISSEAAD	K <mark>I</mark> EG	YKR	TA <mark>HK</mark> RVLPYT [(4)
L plantarum		FNKGNI KARQR <mark>M</mark>	IVQ <mark>Y</mark> GIA	GEMHGA	VVGTDHAAEA.	. <mark>V</mark> TG <mark>FY</mark>	TK <mark>YG</mark> DGGAI	DIVPLWR	LNKRQGKQ	MLAAL	dapkh <mark>l</mark>	YDKVPTAD	LEEDRPALPDEVALG	VRYDDIDD <mark>Y</mark>	LEGR	TVSDAAAE	K <mark>IEA</mark>	YLK	TA <mark>H</mark> KRHAAIT [[7]
E_faecalis	QISD	FNKGNMKARQR <mark>M</mark>	ITQ <mark>Y</mark> AVA	GENAG <mark>A'</mark>	VIGTDHAAEN.	. <mark>VTAFF</mark>	TK <mark>Y</mark> GDGGAI	D <mark>ILPL</mark> FR	<mark>L</mark> NKRQGKA	LLKE <mark>L</mark>	GAPEA <mark>L</mark>	YLKI PTAD	LEDDKPLVADEVALG	7TYDAIDD <mark>Y</mark>	LEGK	KVSETDQQ	T <mark>IEN</mark>	YK	GQ <mark>HK</mark> RHLPIT [[7]
L_lactis	EVSD	<mark>F</mark> NKGNIKARQR <mark>M</mark>	ITQ <mark>Y</mark> AVA	GQYQG <mark>A</mark>	VLGTDHAAEN.	. ITGFF	TK <mark>F</mark> GDGGAI	D <mark>LLPL</mark> FR	<mark>L</mark> NKRQGKA	LLAE <mark>L</mark>	GADPA <mark>I</mark>	YEKVPTAD	LEEGKPGLADEIA <mark>L</mark> G	VTYNDIDD <mark>Y</mark>	TEGK	VISEDAKA	K <mark>I</mark> EA	WKK	TQ <mark>H</mark> KRHLPIS [[7]
SPyogenes	EISD	<mark>F</mark> NKGNIKARQR <mark>M</mark>	ISQ <mark>Y</mark> AIA	GQMAG <mark>A</mark>	VIGTDHAAEN.	. <mark>I</mark> TGFF	TK <mark>F</mark> GDGGAI	D <mark>ILPL</mark> FR	<mark>L</mark> NKRQGKA	<mark>l</mark> lkv <mark>l</mark>	GADAA <mark>L</mark>	YEKVPTAD	<mark>L</mark> EDQKPGLA DE VA <mark>L</mark> G	VTYQDIDD <mark>Y</mark>	LEGK	LISKVAQA	T <mark>I</mark> EK	WHK	GQ <mark>H</mark> KRHLPIT [[7]
S_Pneumoniae	PVSD	<mark>F</mark> NKGNIKARCR <mark>M</mark>	IAQ <mark>Y</mark> ALA	GSHSGA	VIGTDHAAEN.	. ITGFF	TK <mark>F</mark> GDGGAI	D <mark>ILPL</mark> YR	<mark>l</mark> nkrqgkq	<mark>l</mark> lqk <mark>l</mark>	GAEPA <mark>L</mark>	YEKI PTAD	LEEDKPGLADEVA <mark>L</mark> G	VTYAEIDD <mark>y</mark>	LEGK	TISPEAQA	T <mark>IEN</mark>	WHK	GQ <mark>H</mark> KRHLPIT [[7]
C_diPhtheriae	QDVGD	<mark>F</mark> NKGNIKARQR <mark>M</mark>	IAQ <mark>Y</mark> ALA	GEKKLL	VIGTDHAAEN.	. <mark>V</mark> TG <mark>FF</mark>	TK <mark>FG</mark> DGGAI	D <mark>ILPL</mark> AG	<mark>L</mark> SKRQGAA	LLEH <mark>L</mark>	GAPAS <mark>T</mark>	WEKVPTAD	LEEDRPALPDEEALG	VT <mark>Y</mark> REIDA <mark>Y</mark>	IEGNE.	EVSPEAQQ	RIER	WKI	GQ <mark>H</mark> KRHLPVE [[7]
U_Parvum	TNDK	YVLYNLKPKIR <mark>T</mark>	'NYL <mark>Y</mark> AMA	NAYKG <mark>V</mark>	VVSNLNYDEY.	. <mark>I</mark> LGFF	TK <mark>Y</mark> GDSAAI	O <mark>Y</mark> YM <mark>L</mark> IG	<mark>L</mark> LKKHIYE	lgay <mark>y</mark>	hlpnk <mark>i</mark>	LNRAPTPA	NEDD.EHKTDESF <mark>F</mark> G	TYNDLDQ <mark>F</mark> TYND	LYR	KINPKIVS	MIKK	YET	NA <mark>H</mark> KHFVFDK [(14
C_efficiens	DELGD	<mark>F</mark> NKGNVKARQR <mark>M</mark>	VAQ <mark>Y</mark> AIA	CELGL <mark>L</mark>	VVGTDHAAEN.	. <mark>V</mark> TGFF	TK <mark>Y</mark> GDGAAI) <mark>V</mark> LP <mark>L</mark> AG	LTKRQGAL	LLQH <mark>L</mark>	GAPES <mark>T</mark>	WTKVPTAD	LEEDRPALPDEEA <mark>L</mark> G	/TYREIDTY	LENSG.	EVSPEAAA	RIEH	WK <mark>V</mark>	GQ <mark>H</mark> KRHMPVT [[8]
C_Glutamicum	PELTD	<mark>F</mark> NRGNI KARQR <mark>M</mark>	VAQ <mark>Y</mark> AIA	A <mark>G</mark> QLGL <mark>L</mark>	VIGTDHAAEN.	. <mark>V</mark> TG <mark>FF</mark>	TK <mark>FG</mark> DGAAI)LLPLAG	<mark>L</mark> SKRQGAA	ILEH <mark>L</mark>	GAPSS <mark>T</mark>	WTKVPTAD	<mark>l</mark> eedrpalp de ea <mark>l</mark> g	7SYADIDN <mark>Y</mark>	LENKP.	DVSEKAQQ	RIEH	WKV	GQ <mark>H</mark> KRHLPAT [[7]
D_radiodurans	EVRD	<mark>F</mark> VRGNVKARER <mark>M</mark>	VAQ <mark>Y</mark> ALA	A <mark>G</mark> QENL <mark>L</mark>	VVGTDHAAEA.	. <mark>L</mark> TG <mark>FY</mark> :	TK <mark>Y</mark> GDGGVI	D <mark>LTPL</mark> SG	<mark>L</mark> TKRQGAQ	LLAH <mark>L</mark>	GAPEG <mark>T</mark>	WRKVPTAD	<mark>l</mark> eddrpglp de va <mark>lg</mark>	VT <mark>Y</mark> AQIDA <mark>Y</mark>	LEGR	EVSDEAAA	R <mark>L</mark> ER	FLN	SR <mark>H</mark> KRALPVT [(16
N_farcinica	DALRD	<mark>F</mark> VRGNI KARER <mark>M</mark>	IIIQ <mark>Y</mark> AIA	GQENL <mark>L</mark>	<mark>VVGT</mark> DHAAEA.	. <mark>V</mark> TG <mark>FF</mark>	TK <mark>YG</mark> DGGVI	O <mark>LTP</mark> LTG	LTKRQGAA	LLQE <mark>L</mark>	GAPPS <mark>T</mark>	WSKVPTAD	<mark>l</mark> eddrpalp de ea <mark>l</mark> g	LRYSEIDD <mark>y</mark>	LEGK	EVTEAVAA	R <mark>V</mark> EQ <mark>I</mark>	YTA	TR <mark>HK</mark> RTVPVS [[3]
C_michiGanensis	AMTD	FTKGNVKARSR <mark>M</mark>	IAQ <mark>Y</mark> ALA	GQARL <mark>L</mark>	VIGTDHAAEA.	. <mark>V</mark> TGFF	TK <mark>YG</mark> DGGAI	O <mark>VLPL</mark> TG	LTKRQGRA	LLER <mark>L</mark>	GAPER <mark>L</mark>	YLKAPTAD	<mark>LL</mark> DETPGQT DE AN <mark>L</mark> G	LTYAD I DDE	LEGR	DVDDEVAE	AIEA	YAS	TE <mark>H</mark> KRRVPAS [[3]
marine_actinobacterium	PITD	<mark>F</mark> NKGNVKARER <mark>M</mark>	VAQ <mark>Y</mark> AIA	GQLGY <mark>L</mark>	VVGTDHAAEA.	. <mark>V</mark> TGFF	TK <mark>FG</mark> DGGAI) <mark>V</mark> LP <mark>L</mark> SA	LTKRQGKQ	LLKH <mark>L</mark>	NAPER <mark>L</mark>	YEKVPTAD	LLDHTPGQADEDN <mark>LG</mark>	TYEHIDD'	LEGR	DVPAEVAV	EIET	YLN	SR <mark>HK</mark> RTVPVS [[5]
T_maritima	LRNTGPEEFLKG	LAYYRIKHRIR <mark>M</mark>	ICLL <mark>Y</mark> FEA	EKRGY <mark>A</mark>	VVGTTNRTEY.	. <mark>L</mark> TG <mark>LY</mark>	VK <mark>W</mark> GDEAVI)IEPIMH	LYKTQVFE	LAKE <mark>M</mark>	NVPEK <mark>I</mark>	LKKP PS PD	LIPGITDEMA <mark>FN</mark>	MSYLE <mark>LDR</mark> I	LMKLE [3] SDEDPKKVE	R <mark>V</mark> KK	LEL	SE <mark>KY</mark> RRDIPI [[2]
H_hePaticus	EKLDMNQK	MRMGNFCARIR <mark>M</mark>	TML <mark>Y</mark> DCA	ADNAL ^I	VLGTSNKSEI.	. <mark>LLGY</mark> G	TI <mark>F</mark> GDLAYA	AINPIGG	LYKTQIFA	FARA <mark>L</mark>	NVPQE <mark>I</mark>	IAKKPSAD	LFANQSDETD <mark>L</mark> G	YNYAD I DTH	LEAFE [:	23]GFECNMIE	SLST	evwn.	NT F KRTKPTI [[2]
H_mustelae	PKPSA	LRLGNFLARIR <mark>M</mark>	NIL <mark>Y</mark> DYS	MQKNA <mark>L</mark>	VIGTSNKSEL.	. <mark>MLGY</mark> G'	TI <mark>Y</mark> GDLAYA	AINPIGG	FFKTEIFA	LAKA <mark>L</mark>	ELPDS <mark>I</mark>	LTKEPSAD	LYPDQSDAKE <mark>L</mark> G	YTYAQIDPI	LEAIH [18]NFDAKMVE	DITT	ILK	NCFKQKSPIIY	.′QΑ
C_lari	LNP	<mark>L</mark> RFGNLAARVR <mark>M</mark>	SLLYDYS	ALHNAL	VVGTSNKSEL.	. MLGYG:	TI <mark>YG</mark> DLACA	AFNPLAT	LYKSEVFE	LAKFI	GVHEN <mark>F</mark>	IQKAPSAD	LWPNQSDEKDLG	YKYEVLDEV	LKALE [6].NFDENLKN	LVLE	av Q N	NAFKRKLPTT [[8]
C_uPsaliensis	TNQ	INTGNYAARIRM	SLLYDYS	ALKNYL	VVGTSNKSEL.	. MLGYG.	TIYGDLACA	FNFIGK	JYKSELYT	LAKYL		LOKAPSAD	LWE NQSDEEDLG.	SYTLIDEC	ERALE [8 J TKLNPRLIS		T.KO	NAFKRVMPPI	101
T_denitrificans	MDN	LKKGNLSAKLRM	STLFDLS	AKHNAL	VLGTSNKSEL.	. mLGYG	ILYGDLACA	ALINPIGD.	LIKSEVYE	DAEYL	NVTNS 1	MAKAPSAD	LWAGQSDEADLG			⊥∠ J GYDAKMLDI	MILE	INLR	NOTKGKMPLI	161
F_HUCLEATUM	ATS	DODONI CADI DM	CTT PDI C	AMALAN	VICTONCIEL.	MI OVC		INPIGD	TRINIWD	LARVI	NUTING	MERRESAD		TREADQU		12] GENKULVDI		INT D	SETRERMPLI [.41 [0]
w_succinogenes	AIP	L DMCNIKMA DRDM	CTL PDVC	CKENAT	WUCTONKSEI.	VI CVC	TOPCDACA	LNPTCD	VETNIT	LCDVT	VIDNET	TERRESAD	LWF COTDFORMO			12] GRAKELVE		MNP	CEVEDDWDT T	[2]
M_Penetrans	DNK	DICNACADEDM	TTT VDFC	TINIT	UTOTONNADEL.	TTOYCE	TREBUSGSL	T NETCO	UNALDAKE	LAVEL	CIDDET	TOKADEAD		CVNDMDOT	LEDUT [101 CUARELVE		TOT	SERRANMPRI [. /]
M Depotrong	PQKEL	TAT MILLARLER	MCLIIHA		VEGIGNFIEY.	VTOVP	TRAGDGACL	T DI AN	TRADUCE	COVI	CUDCOL	TNERPIAS	LEE NOVDEDDIN	CVI PTDM	FQGHL.	QLSAIKQQ		RQS UV	CEURDNMDRT [./]
M Droumoniae	DOKET	MUACHI KADI DM	ACT VTUA		UL OPCNETRY	CI CVP	TRUCDCACE		LI KODUVA	LCOUR	GIPESI.	TEDADEAC	LEA COTDENEMO	TYPELDO	PIENPIN.	OI CATKOO		DOC	SOUNDEL DATE	.0]
C_HyuroGenorormans	HEPER	LANCHTODEL DM	TOTVARA	OPPDVI I	UL OTDNECEM	VI OVP	TRACDOCUL		TRUEVWE	MAART	GUPERI.	TVVVDCAN		PERCELDIN	MENDN	EGGAELVD		UPT	CONKERT PEVP	1911) [0]
C hudroConoformer	BIIFEEQ	LALANTYDR DA		ANT NUT	UNCTONAPEA.	PTOVP	TRIGDGGVL	TT DTCN	TRAEVRA	RADVT	GLPERI	TTENDERG			TTTC	GVPEERQE	PUVD		COUNTRY TO AND	. 7 J
Degulfitobagtorium an	DVTDDDO	L COCMI VADI DM	COTT VOTUA	MOT NUT	ALCONTRACTOR A	VTOVD	TRUCCOUL	TT DT AC	A DUD A DUD A		CI DEVI	MATD V DTD A	LUD COTDRODMO	TTY TOT TODA	TICE	CUDEEDOE		IIIOO	CETTROAT DDA [(01

SI Fig. 7 | **Multiple sequence alignment of representative bacterial NadE enzymes**. Complete organism names for all NadE representatives are provided in SI Table 1. Those with known 3D structure are underlined. The signature "P-loop" motif residues are highlighted in black. Locations of secondary structure elements are marked for *ft*NadE* and *bs*NadE with "h" for helix and "e" for strand. Conserved small or polar residues are highlighted in gray. Uncharged residues in mainly hydrophobic sites are highlighted in yellow. Residues that interact with ATP, NAD, or metal ions are indicated by letter A, N or M, respectively, at the bottom of the alignment. Residues that are different between *ft*NadE* and *bs*NadE in their adenosyl binding site (shaded by cyan) and nicotinosyl binding site (shaded by green) are indicated.



SI Fig. 8 | Maximum-likelihood phylogenetic tree of bacterial NAD synthetase family. Constructed based on 67 bacterial representatives of NadE family that do not contain a glutamine-amidotransferase domain (reduced from the original set of ~200 sequences by a single linkage clustering method with 75% identity threshold to decrease redundancy). Color coding denotes major taxonomic groups: α -proteobacteria (magenta), β -proteobacteria (red), γ -proteobacteria (cyan), ε -proteobacteria (orange), Bacillus/Clostridium group (dark blue), Mycoplasmatales (brown). Experimentally characterized NAD synthetases with solved tertiary structures are marked by "3D". A distinct NMN synthetase branch containing *F. tularensis* NadE and two orthologs from *M. succinoproducens* and *A. succinogenes* is outlined by grey background. An approximate branching point for a more divergent group of NAD synthetases containing an additional glutamine-amidotransferase domain (NADS-GAT) is shown by a dashed line.



SI Fig.9 | Direct verification of ftNadE* catalyzed conversion of NaMN to NMN. (A) direct verification of novel NMN synthase activity of *F. tulariensis nadE* gene product. Shown are the HPLC (traces at 254 nm) of reaction mixtures containing 2 mM ATP, 1 mM NaMN, 4 mM NH₄Cl, 10 mM Mg²⁺ in the absence or presence of 4 uM ftNadE* after 1 h of incubation. Positions corresponding to NMN and AMP products are indicated by arrows. (B) The proposed mechanism of reaction catalyzed by NMN synthetase by analogy with the known mechanism of NAD synthetase



Scheme 1. An enzymatic system described by the model (1) – (4). The system is analyzed for two important organisms, *Francisella tularensis* (organism *F*) and *Bacillus anthracis* (organism **B**), see the main text for more details. Here **A** is the universal nicotinic acid mononucleotide precursor (NaMN), **B** is nicotinamide mononucleotide intermediate (NMN), C – nicotinic acid dinucleotide (NaAD), D – nicotinamide adenine dinucleotide (NAD). Reaction 1 is catalyzed by NMN synthetase (ftNadE), reaction 2 is catalyzed by NMN adenylyl transferase (ftNadM), reaction 3 is catalyzed by NaMN adenylyl transferase (baNadD), and reaction 4 is catalyzed by NAD synthetase (baNadE). V_0 is an influx into the open system; the pathway is modelled as a closed system (e.g. a test-tube) when V_0 is absent (i.e. when $V_0 = 0$ and the utilization of D is absent). Route I, $A \rightarrow C \rightarrow D$, and Route II, $A \rightarrow B \rightarrow D$, corresponds to Route I and Route II depicted in Fig.1 of the main text, respectively.

SI Table 5. Kinetic rates for enzymatic reactions as in Scheme 1.

Nº	Enzyme Name	Reaction Rate
1	NMN synthetase	$V_1 = \frac{k_1 \cdot [E_1] \cdot [A]}{K_1 \cdot (1 + [C]/K_4) + [A]}$
2	NMN adenylyl- transferase	$V_{2} = \frac{k_{2} \cdot [E_{2}] \cdot [B]}{K_{1} \cdot (1 + [C]/K_{4}) + [A]}$
3	NaMN adenylyl- transferase	$V_{3} = \frac{k_{3} \cdot [E_{2}] \cdot [A]}{K_{1} \cdot (1 + [C]/K_{4}) + [A]}$
4	NAD synthetase	$V_4 = \frac{k_4 \cdot [E_1] \cdot [C]}{K_1 \cdot (1 + [C]/K_4) + [A]}$

Parameters	Francisella tularensis	Bacillus anthracis
k1 (1/sec)	0.5	0.004
<i>K</i> 1 (mM)	0.2	1.18
k2 (1/sec)	2.8	0.014
K2 (mM)	0.034	0.94
k3 (1/sec)	0.16	25.6
K3 (mM)	0.81	0.04
k4 (1/sec)	0.25	2.64
K4 (mM)	5.8	0.29

SI Table 6. Enzymatic kinetic parameters for the reaction rates as in SI Table 5.



SI Figure 10. Transient processes for the metabolite concentrations for organism F. (*a*), (*b*) $[E_1] : [E_2] = 10 : 1$, $[E_1] = 4 \cdot 10^{-3}$ (mM); (*c*), (*d*) $[E_1] : [E_2] = 1 : 1$, $[E_1] = 4 \cdot 10^{-3}$ (mM) (*e*), (*f*) $[E_1] : [E_2] = 1 : 10$, $[E_1] = 4 \cdot 10^{-4}$ (mM). $\% D_k(t) = [D_k](t)/[D](t)$.



SI Fig. 11. Transient processes for the metabolite concentrations for organism *B*. (*a*), (*b*) $[E_1] : [E_2] = 10 : 1$, $[E_1] = 4 \cdot 10^{-3}$ (mM); (*c*), (*d*) $[E_1] : [E_2] = 1 : 1$, $[E_1] = 4 \cdot 10^{-3}$ (mM) (*e*), (*f*) $[E_1] : [E_2] = 1 : 10$, $[E_1] = 4 \cdot 10^{-4}$ (mM). $\% D_k(t) = [D_k](t)/[D](t)$ (k = 1 and 2).

Or	ganism F	Organism B						
1. E1=0.004 (mM) an	d E2 = 0.0004 (mM)							
concentrations	fluxes	concentrations	fluxes					
[A] = 0.20 (mM)	$V_1 = 9.99 \cdot 10^{-4} (\mathrm{mM} \cdot \mathrm{s}^{-1})$	$[A] = 4.38 \cdot 10^{-3} (\text{mM})$	$V_1 = 5.35 \cdot 10^{-8} (\mathrm{mM \cdot s^{-1}})$					
[B] = 0.35 (mM)	$V_2 = 9.99 \cdot 10^{-4} (\mathrm{mM \cdot s^{-1}})$	$[B] = 1.02 \cdot 10^{-2} (\mathrm{mM})$	$V_2 = 5.44 \cdot 10^{-8} \text{ (mM} \cdot \text{s}^{-1}\text{)}$					
[C] = 0.016 (mM)	$V_3 = 1.37 \cdot 10^{-6} (\mathrm{mM \cdot s^{-1}})$	$[C] = 3.04 \cdot 10^{-2} (\text{mM})$	$V_3 = 1.00 \cdot 10^{-3} (\mathrm{mM \cdot s^{-1}})$					
[D] = 1.00 (mM)	$V_4 = 1.37 \cdot 10^{-6} (\mathrm{mM \cdot s^{-1}})$	[D] = 1.00 (mM)	$V_4 = 1.00 \cdot 10^{-3} (\mathrm{mM \cdot s^{-1}})$					
2. E1=0.004 (mM) an	d E2 = 0.004 (mM)							
[A] = 0.20 (mM)	$V_1 = 8.83 \cdot 10^{-4} (\mathrm{mM \cdot s^{-1}})$	$[A] = 3.95 \cdot 10^{-4} (\mathrm{mM})$	$V_1 = 4.84 \cdot 10^{-9} (\mathrm{mM \cdot s^{-1}})$					
[B] = 0.36 (mM)	$V_2 = 8.83 \cdot 10^{-4} (\mathrm{mM \cdot s^{-1}})$	$[B] = 8.50 \cdot 10^{-5} (\mathrm{mM})$	$V_2 = 5.02 \cdot 10^{-9} \text{ (mM} \cdot \text{s}^{-1}\text{)}$					
[C] = 1.53 (mM)	$V_3 = 1.17 \cdot 10^{-4} (\mathrm{mM} \cdot \mathrm{s}^{-1})$	$[C] = 3.03 \cdot 10^{-2} (\text{mM})$	$V_3 = 1.00 \cdot 10^{-3} ((\text{mM} \cdot \text{s}^{-1}))$					
[D] = 1.00 (mM)	$V_4 = 1.17 \cdot 10^{-4} (\mathrm{mM \cdot s^{-1}})$	[D] = 1.00 (mM)	$V_4 = 1.00 \cdot 10^{-3} (\mathrm{mM \cdot s^{-1}})$					
3. E1=0.0004 (mM) a	and E2 = 0.004 (mM)							
The concentrations of	metabolite C becomes	$[A] = 3.94 \cdot 10^{-4} (\mathrm{mM})$	$V_1 = 2.84 \cdot 10^{-11} (\text{mM} \cdot \text{s}^{-1})$					
very large, $[C] > 100$	(mM)	$[B] = 4.81 \cdot 10^{-7} (\mathrm{mM})$	$V_2 = 2.84 \cdot 10^{-11} (\text{mM} \cdot \text{s}^{-1})$					
while $[D] = 1 \text{ (mM)}.$		[C] = 5.18 (mM)	$V_3 = 1.00 \cdot 10^{-3} \text{ (mM} \cdot \text{s}^{-1}\text{)}$					
		[D] = 1.00 (mM)	$V_4 = 1.00 \cdot 10^{-3} \text{ (mM} \cdot \text{s}^{-1}\text{)}$					

SI Table 7. Steady state concentrations and fluxes in organisms F and B.



SI Fig. 12.Complete HPLC traces of NAD pathway reconstitution experiments. Individual chromatograms of pathway reconstitution experiments presented in Fig. 3c,3d and performed as described in Methods. Upper, middle and lower traces show progression of *in vitro* NAD biosynthesis at time 0, 30 and 60 minutes in *Francisella tularansis* [(A), NaMN \rightarrow NMN \rightarrow NAD] and *Bacillus anthracis* systems [(B), NaMN \rightarrow NAD]. Peaks corresponding to pathway intermediates, NMN and NaAD, are indicated by dotted vertical lines. An extra peak corresponding to ADP (spontaneous hydrolysis of ATP) is marked by asterisk.



SI Fig. 13. In vivo assessment of NAD intermediates. Shown are partial HPLC profiles of NAD biosynthetic intermediates in extracts of *F. tularensis* strain U112 prior (blue) and after (red) enzymatic depletion of NMN (*) by NMN adenylyltrasferase. A subtracted profile (green) enables an accurate quantitation of cellular NMN amount. A similar strategy, with an excess of NAD synthetase as a metabolite-depleting enzyme was followed to quantitate NAAD (not shown).



SI Fig.14. Measurement of NMN synthetase activity in crude extracts of F. *tularensis* strain U112: comparison between WT and NadE* knockout mutant. Incubation of 70 ug total protein extract with 1 mM NaMN, 2 mM ATP and 4 mM NH₄Cl yelds accumulation of NMN intermediate over time, while no formation of NMN is detected for the *ft*NadE* knockout mutant. This result confirms the in vivo NMN synthetase function of *ft*NadE*.

Supplementary discussion

Phylogenetic distribution and a possible evolutionary scenario of the newly identified NMN synthetase

The phylogenetic distribution of the newly identified NMN synthetase, a signature enzyme of the Route II (NaMN \rightarrow NMN \rightarrow NAD) pathway, appears to be rather limited. A comparative sequence analysis of all representatives of the NadE family from the 650 completely sequenced genomes integrated in The SEED allowed us to identify only two additional representatives of the NadE* branch beyond the seven strains of *F*. *tularensis*. As can be seen from the simplified tree (SI Fig. 8) of single-domain members of the NadE family in bacteria (same as those included in the NAD subsystem in SI Table 1 and in the multiple alignment in SI Fig. 7), a relatively well-separated NadE* branch contains the proteins from two species of the Pasteurellaceae group, *Mannheimia succinoproducens* and *Actinobacillus succinogenes*. Amino-acid sequences of both Pasteurellaceae enzymes are generally close to the sequence of *fi*NadE* (54% identity) and share with it a number of residues listed above as part of an NMN synthetase signature contributing to its unique substrate preference.

Colocalization of *M. succinoproducens* and *A. succinogenes nadE** genes in the same operon with the *pncB* gene encoding nicotinate phosphoribosyltransferase suggests that together these two enzymes may compose a new variant of the nicotinic acid (Na) salvage pathway (SI Fig. 2). In this pathway, Na would be transformed to NaMN by the PncB enzyme, followed by the amidation of NaMN to NMN intermediate, which would then be converted to the NAD cofactor by the NMN adenylyltransferase of the NadR family. The latter enzyme was previously characterized as the only housekeeping adenylyltransferase in *H. influenzae* (9). The NMN adenylyltransferase and the RNm kinase (encoded in the C-terminal domain of the NadR protein) activities together with the PnuC transporter compose the RNm salvage pathway that is present in all Pasteurellaceae, including *M. succinoproducens* and *A.* succinogenes. However, the presence of pncB and nadE homologs is the unique feature of the latter two strains. Identification of a similarly organized *pncB-nadE* operon in some Gram-positive bacteria suggests a possible evolutionary scenario that involves its horizontal transfer to a common ancestor of *M. succinoproducens* and *A. succinogenes*. As most of NAD biosynthetic machinery (including the NaMN adenylyltransferase of

28

NadD family) was lost in the common ancestor of all Pasteurellaceae, the NAD synthetase activity of the acquired conventional NadE enzyme would not give any competitive advantage. It is tempting to speculate that a fully functional NMN synthetase could have evolved in this genetic background starting from a low side-activity of the acquired (and otherwise obsolete) NAD synthetase enzyme. It is conceivable that an evolved *nadE** gene could become a subject of further horizontal transfer events, e.g., to an ancestor of *F. tularensis*. A combination of NMN synthetase with NMN adenylyltransferase (of the NadM family) would enable an alternative NAD biosynthetic route and allow the elimination of the functionally redundant conventional pathway from *F. tularensis*. Although the proposed evolutionary scenario is highly speculative, it allows us to explain most of the presently available data.



Comparison of the chromosomal arrangement of the *nadE** and *nadE* genes. In both members of Pasteurellaceae group, the gene $nadE^*$ occurs in the operon-like chromosomal cluster containing gene *pncB*. A similar chromosomal arrangement can be found for the canonical *nadE* gene in a number of Gram-positive bacteria, as illustrated here for *Desulfitobacterium sp Y51*. Both genes, *pncB* and *nadE* from the latter organism, are close homologs of *pncB* and *nadE** genes from Pasteurellaceae (~43% and 48% identity, respectively). Genes surrounding this cluster and not conserved between any of these species are colored grey.

References

- 1. Osterman A, Grishin NV, Kinch LN, & Phillips MA (1994) Formation of functional cross-species heterodimers of ornithine decarboxylase. *Biochemistry* 33(46):13662-13667.
- 2. Daugherty M, Vonstein V, Overbeek R, & Osterman A (2001) Archaeal shikimate kinase, a new member of the GHMP-kinase family. *J Bacteriol* 183(1):292-300.
- 3. Otwinowski Z, Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276:307-326.
- 4. Vagin A, Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* 30:1022-1025.
- 5. Symersky J, Devedjiev Y, Moore K, Brouillette C, & DeLucas L (2002) NH3dependent NAD+ synthetase from Bacillus subtilis at 1 A resolution. *Acta Crystallogr D Biol Crystallogr* 58:1138-1146.
- 6. Murshudov GN, Vagin AA, & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240-255.
- 7. Collaborative Computational Project N (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760-763.
- 8. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126-2132.
- 9. Kurnasov OV, *et al.* (2002) Ribosylnicotinamide Kinase Domain of NadR Protein: Identification and Implications in NAD Biosynthesis. *J Bacteriol* 184(24):6906-6917.
- 10. Gerdes SY, *et al.* (2006) Comparative genomics of NAD biosynthesis in cyanobacteria. *J Bacteriol* 188(8):3012-3023.
- 11. Segel IH ed (1993) Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems (John Wiley & Sons, Inc., New York).
- 12. Bisswanger H (2002) Enzyme Kinetics: Principles and Methods. ed Bubenheim L (Darmstadt:Wiley-VCH).
- 13. Gallagher LA, *et al.* (2007) A comprehensive transposon mutant library of Francisella novicida, a bioweapon surrogate. *Proc Natl Acad Sci U S A* 104(3):1009-1014.
- 14. Zogaj X, S. Chakraborty, J. Liu, D. Thanassi, K. Klose (2008) Characterization of the Francisella tularensis subsp. novicida type IV pilus. *Microbiology* 154.
- 15. Chamberlain RE (1965) Evaluation of Live Tularemia Vaccine Prepared in a Chemically Defined Medium. *Appl Microbiol* 13:232-235.
- 16. Baron GS, Myltseva SV, & Nano FE (1995) Electroporation of Francisella tularensis. *Methods Mol Biol* 47:149-154.
- 17. Maharjan RP & Ferenci T (2003) Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. *Anal Biochem* 313(1):145-154.

- 18. Bieganowski P, Pace HC, & Brenner C (2003) Eukaryotic NAD+ synthetase Qns1 contains an essential, obligate intramolecular thiol glutamine amidotransferase domain related to nitrilase. *J Biol Chem* 278(35):33049-33055.
- 19. Bellinzoni M, *et al.* (2005) Glutamine amidotransferase activity of NAD+ synthetase from Mycobacterium tuberculosis depends on an amino-terminal nitrilase domain. *Res Microbiol* 156(2):173-177.
- 20. Chaudhuri RR, *et al.* (2007) Genome sequencing shows that European isolates of Francisella tularensis subspecies tularensis are almost identical to US laboratory strain Schu S4. *PLoS ONE* 2(4):e352.
- 21. Petrosino JF, *et al.* (2006) Chromosome rearrangement and diversification of Francisella tularensis revealed by the type B (OSU18) genome sequence. *J Bacteriol* 188(19):6977-6985.
- 22. Larsson P, *et al.* (2005) The complete genome sequence of Francisella tularensis, the causative agent of tularenia. *Nat Genet* 37(2):153-159.
- 23. Rohmer L, *et al.* (2007) Comparison of Francisella tularensis genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biol* 8(6):R102.
- 24. Beckstrom-Sternberg SM, *et al.* (2007) Complete genomic characterization of a pathogenic A.II strain of Francisella tularensis subspecies tularensis. *PLoS ONE* 2(9):e947.
- 25. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32(5):1792-1797.
- 26. Clamp M, Cuff J, Searle SM, & Barton GJ (2004) The Jalview Java alignment editor. *Bioinformatics* 20(3):426-427.
- 27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, & Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24):4876-4882.