Figure S1

>gi|226354810:682860-684161 (R) *Deinococcus deserti* VCD115, complete genome

TACACTCGGAAC**ATG**TCGAC**CCT**TGATCCGCTGGGCGTGCCGGGCGCCCCCTTGGAGACC**CTC**AATGCCGCCCTGTCCGGCGCCCAGACG

2 **M** S T L **D** **P** **L** **G** **V** **P** G **A** **P** **L** **E** **T** **L** N **A** A **L** S G A Q T

>gi|94984109:2131647-2132093 *Deinococcus geothermalis* DSM11300, complete genome

ATGCAGACGCCC**ATA**CACAT**CGC**TGAGCCGTTCGGCGTTCCAGGAGCACCGCTGGAAACG**TTG**CAGGCCGTCCTCGCACATCTGGAGAAG

1 **M** Q **A** V **L** A H **L** E K

2 M Q T P I H I A E **P** F **G** **V** **P** G **A** **P** **L** **E** **T** **L** Q **A** V **L** A H **L** E K

>gi|15805042:907631-908080 (R) *Deinococcus radiodurans* R1, complete genome

ATGCTGAGGGCC**ATG**AACAT**CGC**TGATCCGCTGGGGGTGCCCGCCGCGCCGCTGGAAACC**TGG**CAGGCCATGCAGGCGCGGCTGAACGAC

1 M L R A **M** N I A **D** **P** **L** **G** **V** **P** A **A** **P** **L** **E** **T** W Q **A** M Q A R **L** N D

2 **M** N I A **D** **P** **L** **G** **V** **P** A **A** **P** **L** **E** **T** W Q **A** M Q A R **L** N D

>gi|46198308:1131069-1131476 (R) *Thermus thermophilus* HB27, complete genome

TAAGATGAGGCC**ATG**GTACG**CGT**GGGCATGCGCGCCGCGCCCCGGGTGAGCCTCGAGGCC**CTG**AAGGCGGCCCTGGGCGGCCTGAAGCTT

1 **M** V R V G M R A A **P** R V S **L** **E** A **L** K **A** A **L** G G **L** K L

2 **M** V R V G M R A A **P** R V S **L** **E** A **L** K **A** A **L** G G **L** K L

>gi|227986969:242693-243124 (R) *Meiothermus silvanus* DSM9946, draft genome

CTTACCCCGTAG**AAT**ACCGTT**ATG**GAAGTCCTCGGCGTTCGTGGTGCACCCAAGCAGACC**CTC**GCGGCTCTCATGGAAGCACTCAAAGGT

1 **M** E V **L** **G** **V** R G **A** **P** K Q **T** **L** A **A** L M E A **L** K G

2 **M** E V **L** **G** **V** R G **A** **P** K Q **T** **L** A **A** L M E A **L** K G

Figure S1: Alignment of Deide_04870 orthologs. *Deide_04870, Dgeo_2021, DR_0900, TTC_1164* and *MesilDRAFT_02440* loci are shown from top to the bottom. Previously annotated and corrected protein sequences are labelled (1) and (2), respectively. This does not apply for Deide_04870 that was not annotated in (de Groot et al. 2009). Previously annotated initiation codons and their counterparts in other genomes are shown in red and blue, respectively. The new proposed initiation codons are underlined. Conserved amino acids are indicated in bold. It is worth to mention the ATA non-canonical translation initiation codon for *Dgeo_2021* as a possible alternative (see other results regarding non-canonical initiation codons in the manuscript). The underlined ATG for *DR_0900* may probably be the correct initiation codon start based on homology with Deide_04870.

Figure S2

>gi|226354810:754223-754855 *Deinococcus deserti* VCD115, complete genome

GTGATAAAC**ATA**GCGAAAGAACACAAGGTCAACGAGCAAATCCGCGTCCGCCAG**ATC**AGGCTTATCGGCGGCGAAGGCGAGCAGATCGGCATC**ATC**GAC

1 M I N I A K E H K V N E Q I R V R Q I R L I G G E G E Q I G I I D

2 **M** A **K** **E** H K V **N** **E** Q **I** **R** V R **Q** I **R** L **I** **G** G E G E **Q** I **G** **I** I **D**

>gi|94984109:1603384-1603998 (R) *Deinococcus geothermalis* DSM11300, complete genome

GTGATAGAC**ATA**GCGAAAGAACATAAGGTCAACGAGCAGATTCGCGTTCGTCAG**ATT**CGCCTGATCGGCGCTGAAGGCGAGCAGATCGGGATC**ATC**GAC

1 M I D I A K E H K V N E Q I R V R Q I R L I G A E G E Q I G I I D

2 **M** A **K** **E** H K V **N** **E** Q **I** **R** V R **Q** I **R** L **I** **G** A E G E **Q** I **G** **I** I **D**

>gi|15805042:2099093-2099731 (R) *Deinococcus radiodurans* R1, complete genome

GTGATAACC**ATA**GCGAAAGATCTTAAGGTCAACGAGCAGATTCGCGTCCGGCAG**GTC**CGCCTGATCGGCGCGGAAGGGGAGCAGATCGGGATT**ATC**GAT

1 M I T I A K D L K V N E Q I R V R Q V R L I G A E G E Q I G I I D

2 **M** A **K** D L K V **N** **E** Q **I** **R** V R **Q** V **R** L **I** **G** A E G E **Q** I **G** **I** I **D**

>gi|46198308:164771-165304 *Thermus thermophilus* HB27, complete genome

GTGATGTAG**ATA**AAGGAGTACCTGACCAACGAACGCATACGCGCCAAACAG**GTC**CGCGTCGTCGGTCCCGACGGCAAGCAGCTCGGCATC**ATG**GACACC

1 M D T

2 V M * **M** **K** **E** Y L T **N** **E** R **I** **R** A K **Q** V **R** V V **G** P D G K **Q** L **G** **I** M **D** T

>gi|227988584:72588-73112 (R) *Thermus aquaticus Y51MC23*, draft genome

GTGATGTAG**ATA**AAGGAGTATCTGACCAACGAGCGCATCCGTGCCAAACAG**GTG**AGGGTGATCGGTGCCGACGGCCAGCAGCTGGGCATC**ATG**GACACC

1 M R V I G A D G Q Q L G I M D T

2 V M * **M** **K** **E** Y L T **N** **E** R **I** **R** A K **Q** V **R** V **I** **G** A D G Q **Q** L **G** **I** M **D** T

Figure S2: Alignment of IF-3 orthologs. *Deide_06380, Dgeo_1509, DR_2087, TT_C0183* and *TaqDRAFT_3586* loci are shown from top to the bottom. Previously annotated and corrected protein sequences are labelled (1) and (2), respectively. Previously annotated initiation codons (as they appeared in Genbank) and their counterparts in other genomes are shown in bold red and blue, respectively. The new proposed non-canonical initiation codons are underlined. The Deide_06380 amino acid sequence uncovered by LC-MS/MS data is underlined. Conserved amino acids are indicated in bold. Note that the TT_C0183 protein (Q9ACJ8) is correctly annotated in the current swissprot database release based on remarks done by Wolfrum A et al. (2003) Protein Expression and Purification 29:15-23.

Figure S3

```
    >gi|226354810:682860-684161 (R) Deinococcus deserti VCD115, complete genome
    ATGTTCCGCCGTCGCCCGCCCCTGCCTCCATTCCCACCCAGCGCGCTGCTGGTTGGTGGCGCGGCGCGTGA//GTGCCGGACCCG
1                                                                            M  P  D  P
2    M  F  R  R  R  P  P  L  P  P  F  P  P  S  A  L  L  V  G  G  A  A  R  D//V  P  D  P

    >gi|94554390:654070-655371(R) Deinococcus geothermalis DSM11300, complete genome
    ATGTTCCGGCGCCGCCCTCCCCTCCCCCCCTTCCCGACGGGCGCGCTGCTGGTCGGCGGTGCTGCGCGACTGGCTGCGG
1    M  F  R  R  R  P  P  L  P  P  F  P  T  G  A  L  L  V  G  G  A  A  R  D  W  L  R

    >gi|11612676:1199855-1201159 Deinococcus radiodurans R1, complete genome
    ATGTTTCGTCGCCGTCCGCCCCTGCCGCCGTTTCCTCCAGGCGCTGCCCTGGTCGGCGGCGCGGTGCGGGACTGGCTGCGC
1    M  F  R  R  R  P  P  L  P  P  F  P  P  G  A  A  L  V  G  G  A  V  R  D  W  L  R

    >gi|227986969 :466647-467912 (R) Meiothermus silvanus DSM9946, draft genome
    ATCCGCAGGGTTCATCTGTCCCCTTTTCCCTTTCCTGCTCAAGGATACCTCGTAGGTGGCGCGGTGCGGGATCTGATGCTG
1                                                                   M  R  D  L  M  L
2    M  R  R  V  H  L  S  P  F  P  F  P  A  Q  G  Y  L  V  G  G  A  V  R  D  L  M  L

    >gi|46198308:542569-543873 (R) Thermus thermophilus HB27, complete genome
    GTGGCCCATATGGACTTCCCCTTTTATACCCCCAAGGACGCCTTCCCCGTGGGCGGGGCAGTGCGGGACCTCCTCCTGGGA
1    M  A  H  M  D  F  P  F  Y  T  P  K  D  A  F  P  V  G  G  A  V  R  D  L  L  L  G

    >gi|206900094:1348597-1350006 Dictyoglomus thermophilum H-6-12, complete genome
    ATGGAGCAAGA//AAGAGTATTCAAAAATTTACGAGATTACACCCTATATAGTAGGTGGCTTCATCAGAGATTATCTATTA
1    M  E  Q  E//K  E  Y  S  K  I  Y  E  I  T  P  Y  I  V  G  G  F  I  R  D  Y  L  L

    >gi|52696111 PcnB from Aquifex aeolicus (PDB code: 1VFG)
1             V  G  Q  I  A  K  E  M  G  L  R  A  Y  I  V  G  G  V  V  R  D  I  L  L
```

Figure S3: N-termini of PcnB from the *Deinococcus-Thermus* phylum. *Deide_05850*, *Dgeo_0612*, *DR_1191*, *MesilDRAFT_04540*, *TTC_0556* and *DICTH_1410* loci are shown from top to the bottom. The N-terminal sequence of PcnB from *Aquifex aeolicus* is also shown, as well as the corresponding secondary structure elements. Previously annotated and corrected protein sequences are labelled (1) and (2), respectively. Previously wrongly annotated initiators are shown in blue while correct translation initiation condons are in red. Conserved amino acids are indicated in bold.

Figure S4

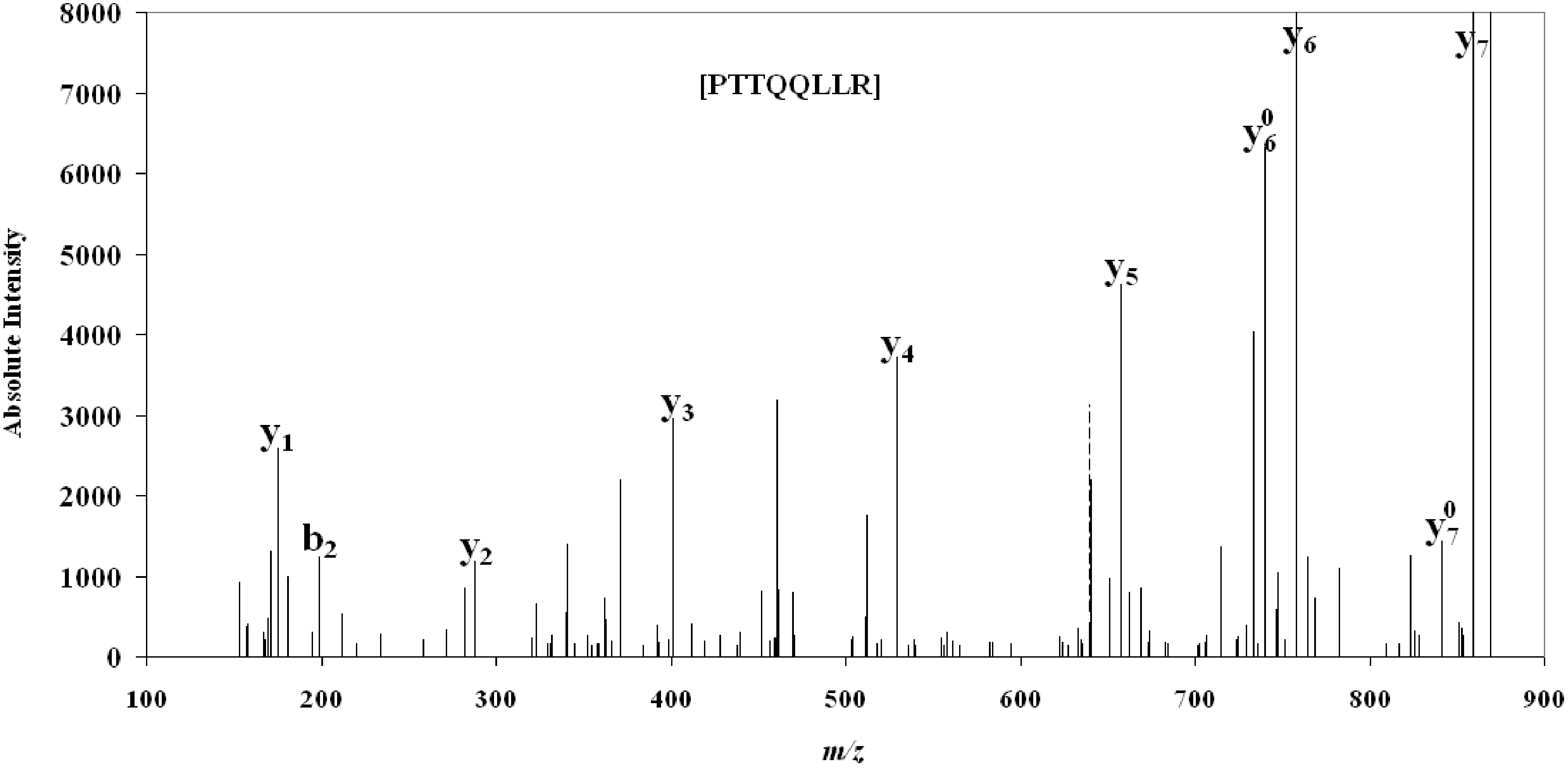

Figure S4: MS/MS spectrum of the [PTTQQLLR] peptide corresponding to the N-terminal-most peptide of Deide_19010 (RpsL).

Figure S5

```
     >gi|226354810:2291296-2291697 (R) Deinococcus deserti VCD115, complete genome
     TGATTTACTCTGCCTACCACCCAGCAGCTGCTCCGTAAGGGGCGCACCACGCTCCAGAAGAAGAGCAAAGTTCCGGCCCTGAAGGGC //GTGCGTCTG
1    *  F  T  L                                                                                M  R  L
2             M  P  T  T  Q  Q  L  L  R  K  G  R  T  T  L  Q  K  K  S  K  V  P  A  L  K  G  S //

     >gi|94984109:1970590-1971003 (R) Deinococcus geothermalis DSM11300, complete genome
     TGATTGACCCTGCCTACCACACAGCAACTGCTCCGTAAGGGGCGCACCACCCTGCAGAAGAAGAGCAAGGTTCCCGCTCTCAAGGGGAGCCCCTTCCGC
1    *  M  T  L  P  T  T  Q  Q  L  L  R  K  G  R  T  T  L  Q  K  K  S  K  V  P  A  L  K  G  S  P  F  R
2             M  P  T  T  Q  Q  L  L  R  K  G  R  T  T  L  Q  K  K  S  K  V  P  A  L  K  G  S  P  F  R

     >gi|15805042:306960-307364 Deinococcus radiodurans R1, complete genome
     AGTGATGTTTTGCCTACCACCCAGCAACTTCTGCGTAAGGGCCGCAAGGTTCTCCAGAAAAAGAGCAAGGTCCCGGCCCTGAAGGGCAGCCCCTTCCGC
1    S  D  V  M  P  T  T  Q  Q  L  L  R  K  G  R  K  V  L  Q  K  K  S  K  V  P  A  L  K  G  S  P  F  R
2             M  P  T  T  Q  Q  L  L  R  K  G  R  K  V  L  Q  K  K  S  K  V  P  A  L  K  G  S  P  F  R

     >gi|46198308:1270800-1271204 (R) Thermus thermophilus HB27, complete genome
     GTGGTGGCACTGCCGACGATCAATCAGCTCGTCCGAAAGGGCCGCGAGAAGGTCCGCAAAAAGAGCAAGGTTCCGGCGCTGAAGGGGGCGCCCTTCCGC
1    V  M  A  L  P  T  I  N  Q  L  V  R  K  G  R  E  K  V  R  K  K  S  K  V  P  A  L  K  G  A  P  F  R
2             M  P  T  I  N  Q  L  V  R  K  G  R  E  K  V  R  K  K  S  K  V  P  A  L  K  G  A  P  F  R

     >gi|218296332:348-785 Meiothermus silvanus DSM9946, draft genome
     GTGGTTACACTGCCAACGATCAACCAACTGCTACGCAAGGGTCGTGCTCCGGTCATCAAGAAGAGCAAAGTGCCCGCTTTGAAGGGGAGTCCCTTCCGC
1    V  V  T  L  P  T  I  N  Q  L  L  R  K  G  R  A  P  V  I  K  K  S  K  V  P  A  L  K  G  S  P  F  R
2             M  P  T  I  N  Q  L  L  R  K  G  R  A  P  V  I  K  K  S  K  V  P  A  L  K  G  S  P  F  R
```

Figure S5: Alignment of RpsL orthologs. *Deide_19010, Dgeo_1873, DR_0305, TT_C1333* and *MesilDRAFT_16410* loci are shown from top to the bottom. Previously annotated and corrected protein sequences are labelled (1) and (2), respectively. Previously annotated initiation codons (as they appeared in genbank) and their counterparts in other genomes are shown in bold red and blue, respectively. The new proposed non-canonical initiation codons are underlined. The Deide_19010 amino acid sequence uncovered by LC-MS/MS data is underlined. Conserved amino acids are indicated in bold. Note that the TT_C1333 protein (P61941) is correctly annotated in the current swissprot database release based on a structural work reported by Laurberg M et al. (2008) Nature 454:852-857.

Figure S6

>gi|226316800:341409-341939 *Deinococcus deserti* VCD115, complete genome
GGCCAG**CTG**AGCGTCTCTAATGACGCCTACGTCGCGCTGGGAGCGAATCTGGGCGACGCCCACGGCACGCTGTGC//**GTG**AAGCCGTCACCA
1 **M** H A S **S**
2 **M** S V S N D **A** Y V **A L G A N L G** D A H G **T L** C // V H A S **S**

>gi|94984109:404426-404928 *Deinococcus geothermalis* DSM11300, complete genome
GCGGCCAGGCGA**ATG**AGCGCGCGGGCCTTCATCGCGTTGGGGGCGAACTTGGGCGAACCGCTCGTCACGCTGCGC //CTCAAGGGGGTCAGC
1 **M** S A **R A** F I **A L G A N L G** E **P** L V **T L** R // L K G V **S**

>gi|15805042:172108-172623 *Deinococcus radiodurans* R1, complete genome
CGAGAAGGAGCGCAC**GTG**ACGCGGGCTTTTATCGCGCTGGGCGCCAACCTGGGCGACCCCCAGGCCACCCTGCGC //GTGCGCGGCGTGTCG
1 **M** T **R A** F I **A L G A N L G** D **P** Q A **T L** R // **V** R G V **S**

>gi|22123922:879625-880104 (R) *Yersinia pestis* KIM, complete genome
GGGTATAATTTTACC**ATG**ATCCGGGTCTATATCGCGCTGGGTAGCAATTTGGCCATGCCACTACAACAGGTTAGC //TTAGTAGCGTGTTCG
1 M I **R** V Y I **A L G** S **N L** A M **P** L Q Q V S // L V A C **S**

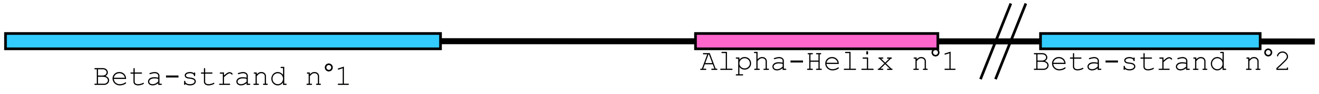

Figure S6: Alignment of Deide_03051 orthologs. *Deide_03051, Dgeo_0398, DR_0170,* and *y0787 (folK)* loci are shown from top to the bottom. Previously annotated and corrected protein sequences are labelled (1) and (2), respectively. Previously annotated initiation codons and their counterparts in other genomes are shown in red and blue, respectively. The new proposed initiation codon for Deide_03051 is underlined (no other possible initiation codon is found upstream). Conserved amino acids are indicated in bold. Secondary elements for FolK from *Yersinia pestis* are indicated (PDB code: 2QX0).