Supplementary Figures

(for "RNA Sequencing and Proteogenomics Reveal the Importance of Leaderless mRNAs in the Radiation-tolerant Bacterium *Deinococcus deserti*" by de Groot *et al.*)

Contents:

Figure S1. Codon usage of leaderless and leadered genes. Relative synonymous codon usage (RSCU) is shown for the 1174 leaderless genes (339156 codons) and 784 leadered genes (277460 codons). RSCU values are the number of times a particular codon is observed, relative to the number of times that the codon would be observed for a uniform synonymous codon usage.

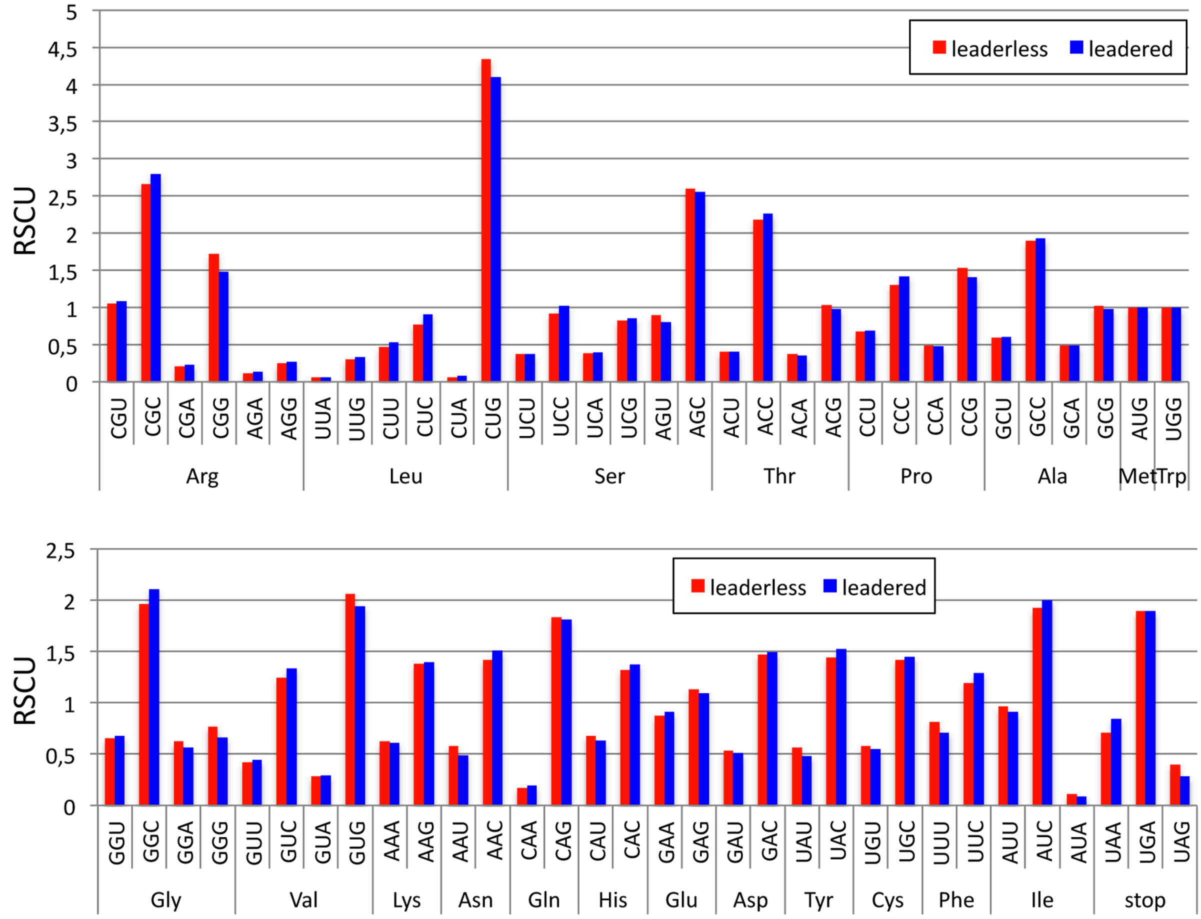


Figure S2. Amino acid composition of leaderless and leadered gene products. The average amino acid composition, in percentages, is indicated for the products of the 1174 leaderless genes (337982 residues) and 784 leadered genes (276676 residues).

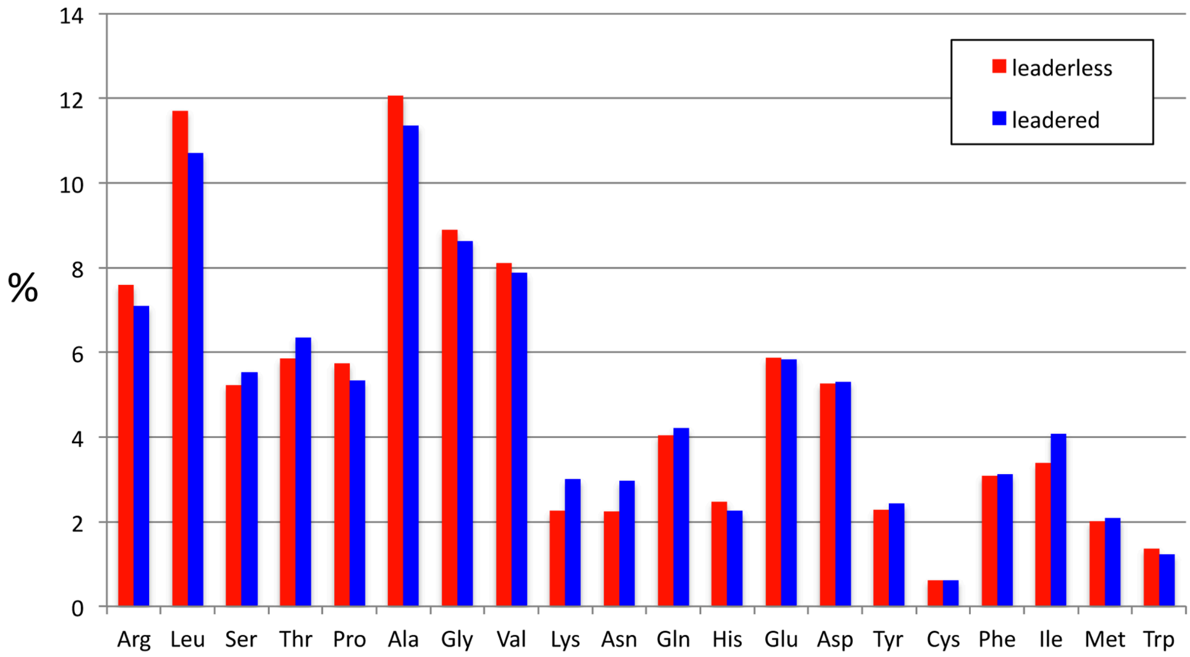


Figure S3. Read coverage of a region with leaderless *clpP* and leadered *lon*. The proteases ClpP (Deide_19570, Clp protease proteolytic subunit) and Lon (Deide_19590, Lon protease) are produced from leaderless and leadered mRNA, respectively. Coverage (in blue) of reads that map to the forward genomic DNA strand is shown above the genes (results in non-irradiated and irradiated samples were similar for these genes; only RD19 IR and RD19 IR + TEX samples are shown). Panels B-D are zoomed parts of the region shown in panel A. Transcription start sites (TSSs) for *clpP* and *lon* are indicated with arrows in panels B and C, respectively. Panel D is a zoom at the translation initiation codon of *lon*. Start codons, -10 motifs (upstream of TSSs) and SD sequence (upstream of start codon in leadered *lon* mRNA) are boxed. Treatment (+) or not (–) of RNA with TEX is indicated.

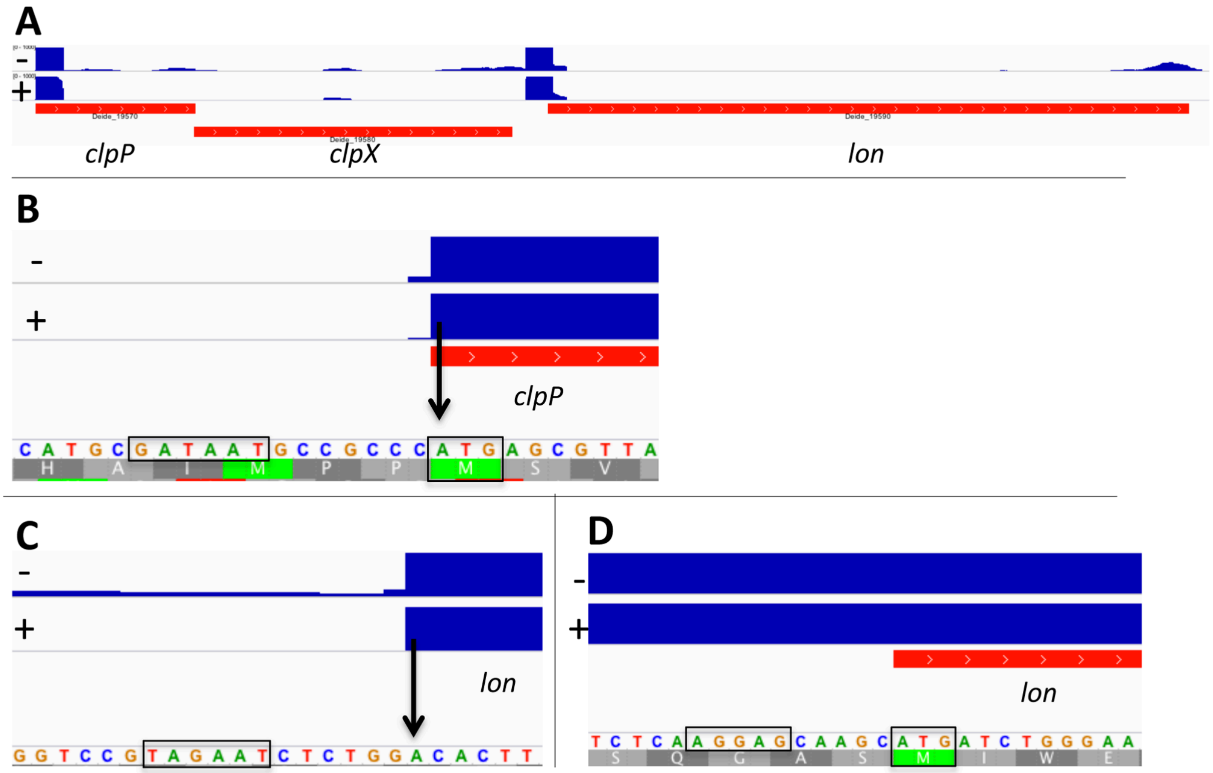
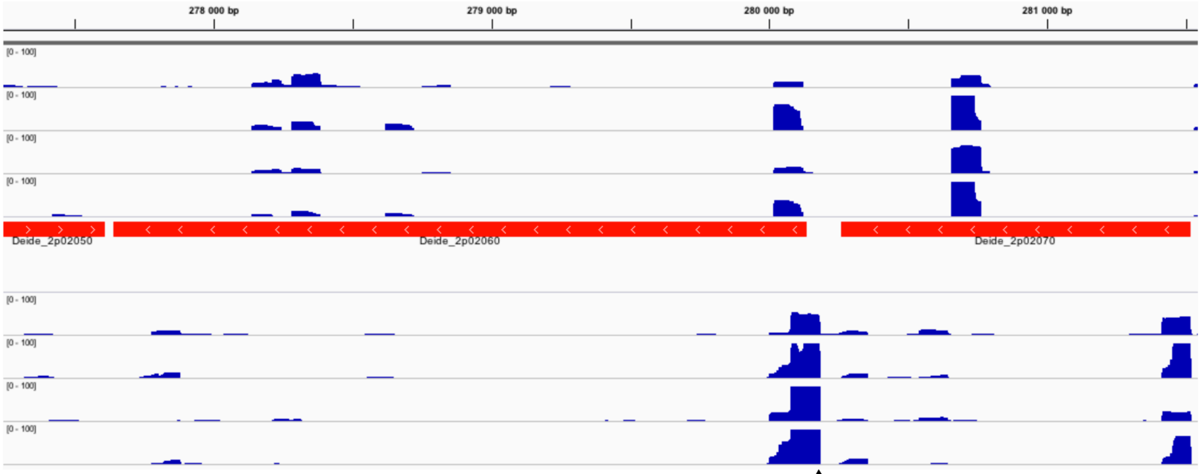


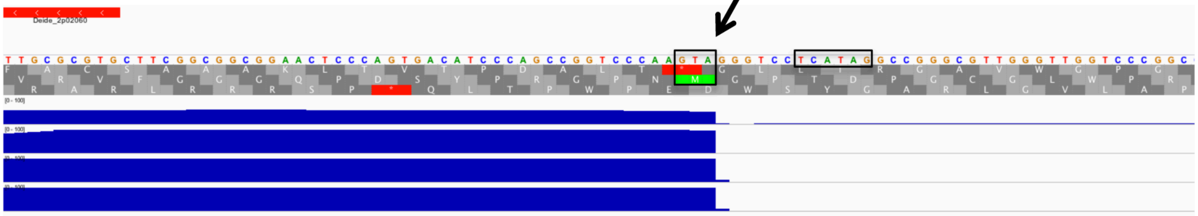
Figure S4. Start codon re-annotations of DNA repair genes in *D. deserti*. RNA-seq read coverage for *uvrA2* (*Deide_2p02060*), *recN* (*Deide_12310*), *rarA* (*Deide_04980*), *ruvA* (*Deide_09360*), *ruvC* (*Deide_20630*) is shown. Coverage (in blue) of reads that map to the forward genomic DNA strand is shown above the genes, and those on the reverse strand below the genes (independent of gene annotation and orientation). Above and below the genes, the order of the samples is: RD19 NI, RD19 NI +TEX, RD19 IR, RD19 IR +TEX. In each figure, the maximum height to show coverage is set at the same value for each sample (= in all 8 "lines"), but this value can be different between the figures. New start codons and -10 motifs are boxed.

Deide_2p02060 (UvrA2)

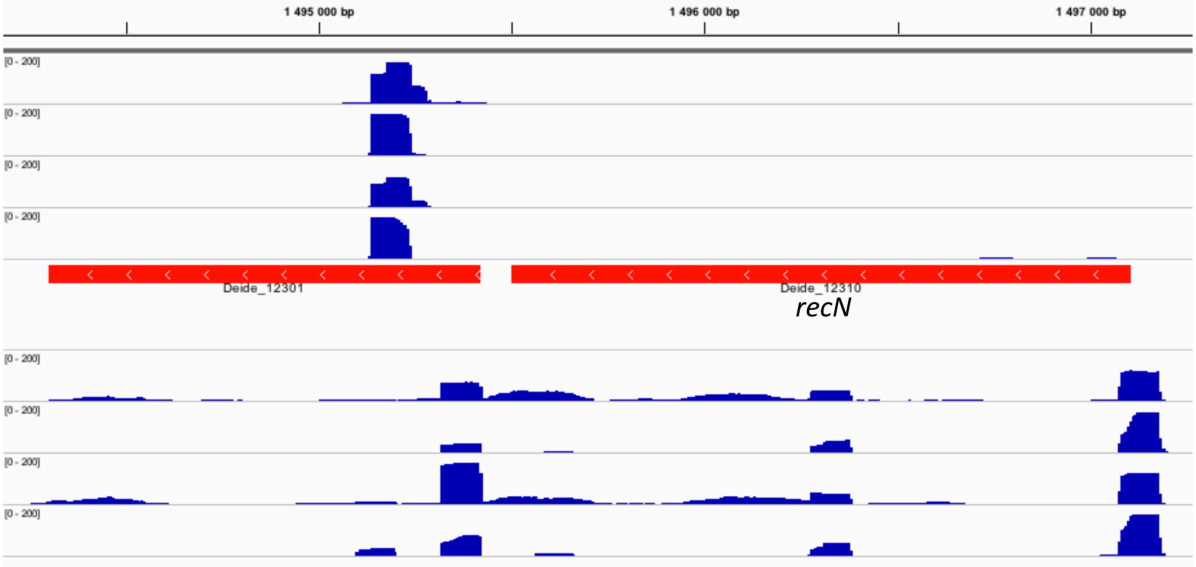


TSS & new start codon

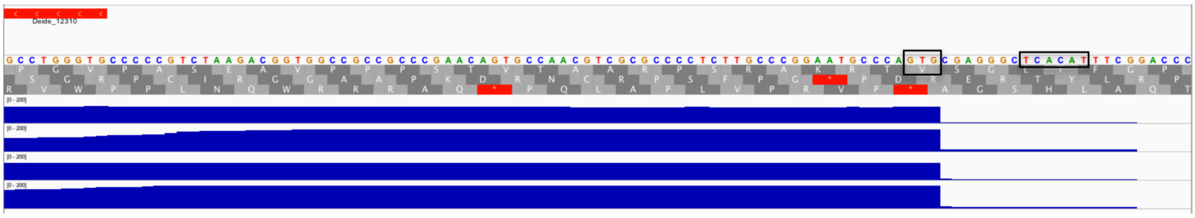
uvrA2, zoom at TSS and initial and new translation start


Deide_2p02060
TTGCGCGTGCTTCGGCGGCGGAACTCCCAGTGACATCCCAGCCGGTCCCAAGTAGGGTCCTCATAGGCCGGGGCGTTGGGTTGGTCCCGGC
M
[0 - 100]

Deide_12310 (recN) & flanking

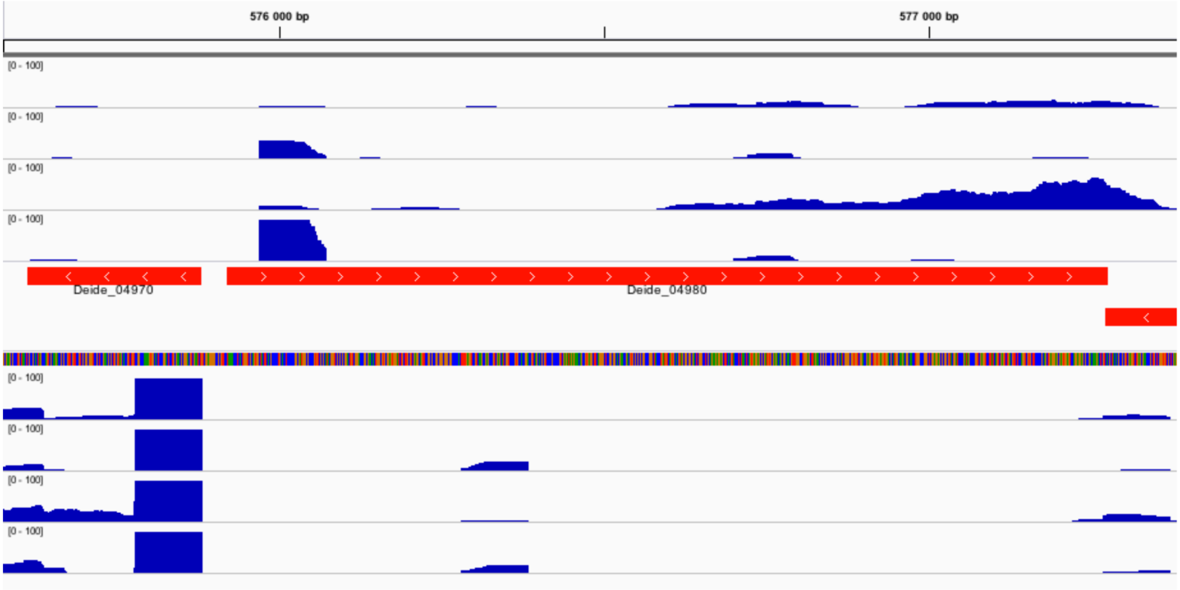


recN, zoom at TSS and new translation start

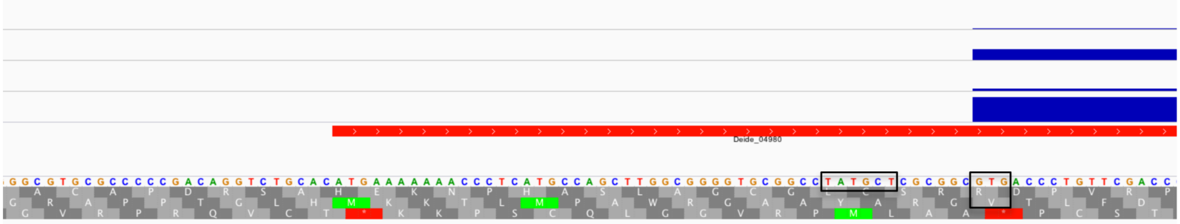

Deide_12310
GCCTGGGTGCCCCCGTCTAAGACGGTGGCCGCCGCCCGAACAGTGCCAACGTCGCGCCCCTCTTGCCCGGAATGCCCAGTGCGAGGGCTCACATTTCGGACCC
[0 - 200]
[0 - 200]
[0 - 200]
[0 - 200]

(Figure S4, continued)

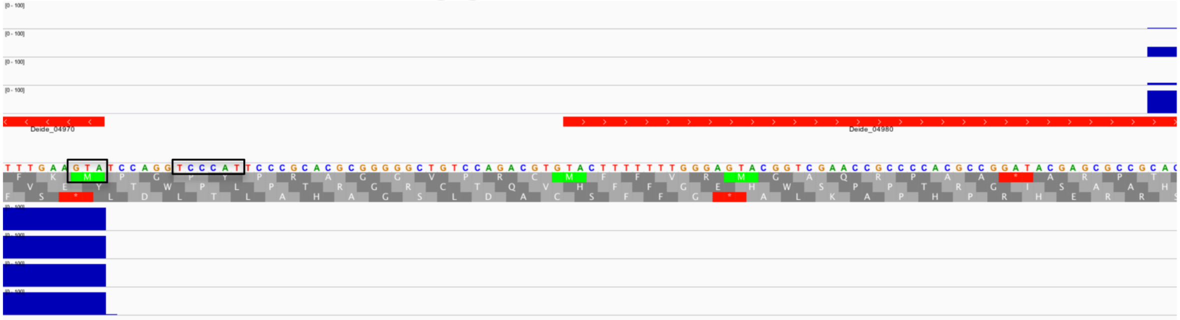
Deide_04980 (RarA) & flanking



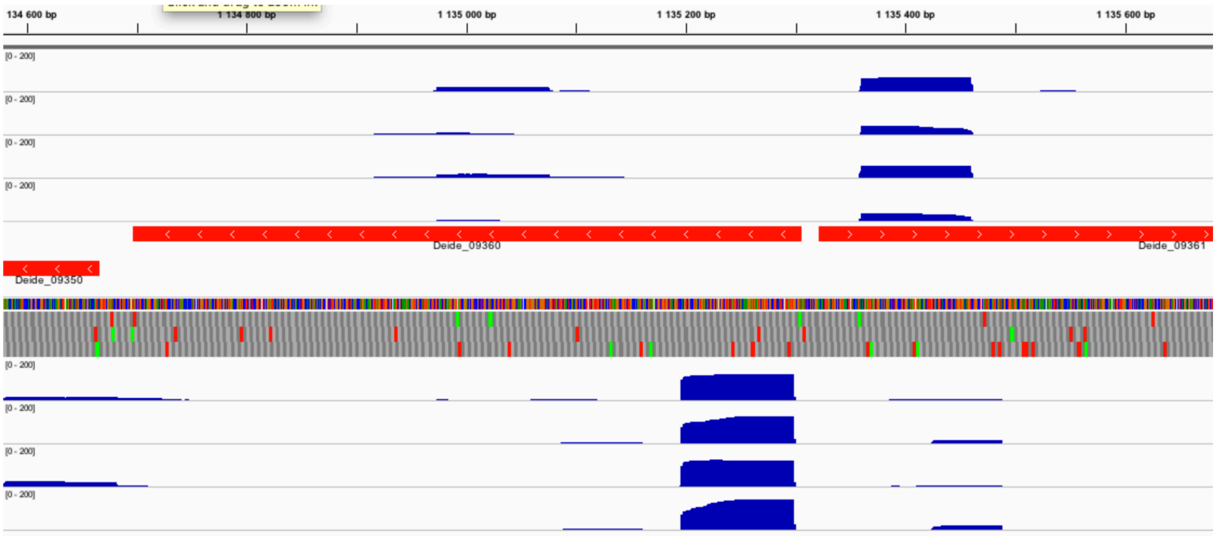
RarA, zoom at TSS and new translation start


Deide_04980
GGCGTGCGCCCCCGACAGGTCTGCACATGAAAAAAACCCTCATGCCAGCTTGGCGGGGTGCGGCCTATGCTCGCGGCGTGACCCTGTTCGACC

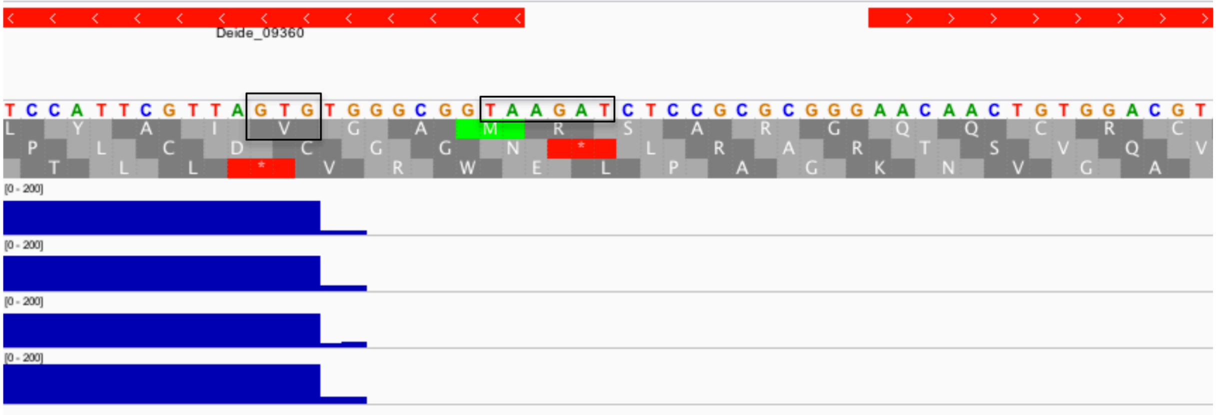
Zoom at TSS of flanking gene


[0 - 100]
[0 - 100]
[0 - 100]
[0 - 100]
Deide_04970
Deide_04980
TTTGAAGTATCCAGGTCCCATTCCCGCACGCGGGGGCTGTCCAGACGTGTACTTTTTTTGGGAGTACGGTCGAACCGCCCCACGCCGGATACGAGCGCCGCAC
[0 - 100]
[0 - 100]
[0 - 100]
[0 - 100]

Deide_09360 (ruvA)

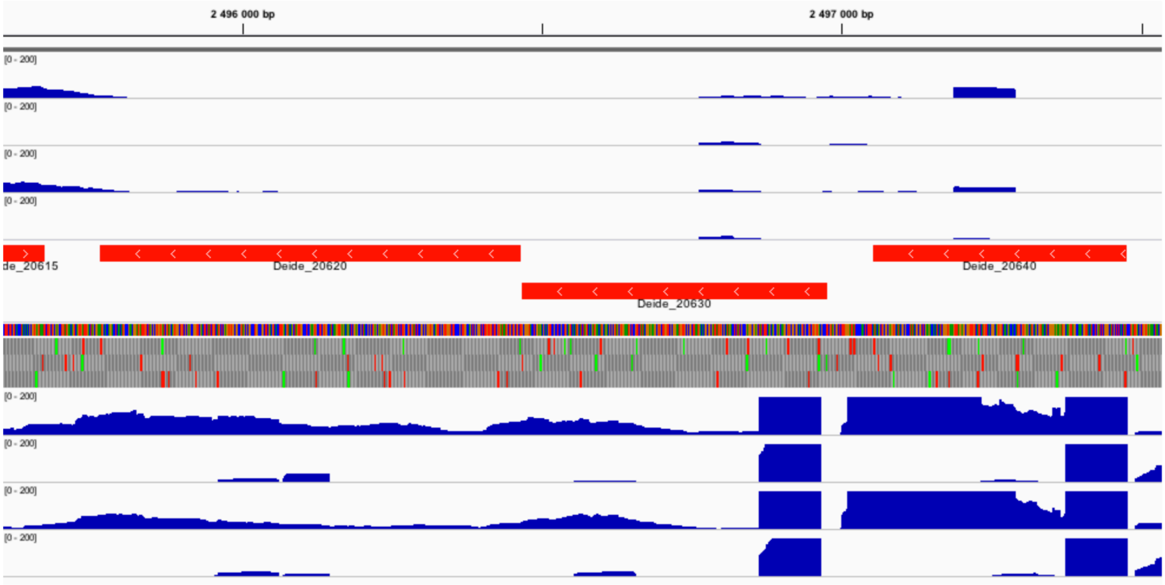


ruvA, zoom at TSS and new translation start

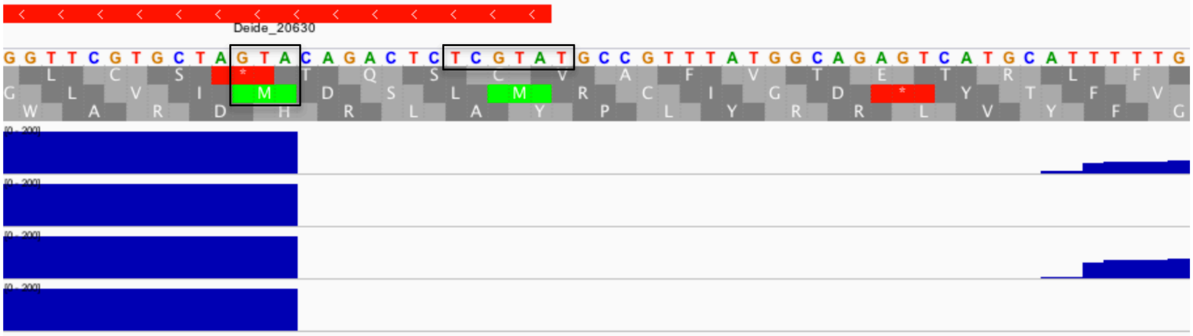

Deide_09360
T C C A T T C G T T A G T G T G G G C G G T A A G A T C T C C G C G C G G G A A C A A C T G T G G A C G T
[0 - 200]

(Figure S4, continued)

Deide_20630 (ruvC) & flanking



ruvC, zoom at TSS and new translation start



Deide_20630
G G T T C G T G C T A G T A C A G A C T C T C G T A T G C C G T T T A T G G C A G A G T C A T G C A T T T T T G
M
M
[0 - 200]

Figure S5. Examples of predicted start codon corrections in Deinococcal homologs of *D. deserti* proteins. In each example, the TSS was at the first nucleotide of the new start codon of the *D. deserti* protein. Blast and multiple alignments indicate that start codon re-annotation may also be required in several homologs, mainly from other *Deinococcus* species. Only the N-terminal regions of the proteins are shown. New annotated starts of proteins (*D. deserti*), or possible new starts (others), are in green. Initial wrong starts (*D. deserti*), or possible wrong starts (others), in cyan.

Shorter *D. deserti* proteins:

Deide_04020 (7aa) conserved protein of unknown function

```
Deide_04020      -----MPPRMARMTDLPSHWQPAPAGYKHVVSVSLGDSKRNAREEINVLGQPFILERIGT
Dgeo_0492        MPQMRSGCSRLHRMSLLRSWQPAPAGFKHVVSVSLGASKRNAREEISVLGQPFVLERIGT
Deipr_0163       ------------MSDLLRSWKPAPPGYKHVVSVSLGGSKRNAREEIEVLGQPFVLERIGT
Deima_2120       ------------MTDLLKNWQPAPSGVRHVVSVSLGSSKRNAREETEVLGQRFILERLGT
DR_1392          -----------------------------------------------MLGQPFILERLGT
                                                                :*** *:***:**
```

tblastn on *D. radiodurans*: DR_1392 should be longer

```
Query  1        MTDLPSHWQPAPAGYKHVVSVSLGDSKRNAREEINVLGQPFILERIGTDGDSRKAAQLFQ  60
                MTD  S WQPAPAG+KHVVSVSLG+SKRNAREEINVLGQPFILER+GTDGDS  AA+LF+
Sbjct  1396664  MTDPLSGWQPAPAGFKHVVSVSLGNSKRNAREEINVLGQPFILERLGTDGDSALAARLFR  1396843
```

Deide_04980 (17aa) RarA, TSS at GTG; V in *D.geothermalis* is also GTG

```
tr|C1D0F9|C1D0F9_DEIDV      ---MKKTLMPAWRGAAYARGVTLFDPPAPLAERLRPRTVAEVVGQTHLLG
tr|E8U6J5|E8U6J5_DEIML      --------------------MTLFEPPAPLAERLRPRTIEEVVGQRHLLG
tr|Q9RT67|Q9RT67_DEIRA      --------------------MTLFDPPAPLAERLRPRTVAEVAGQSHLLG
tr|Q1IYI8|Q1IYI8_DEIGD      MCKMLIRASLPGLPPAILPPVTLFDPPAPLPERLRPRTLAEVVGQGHLLG
tr|F0RJT4|F0RJT4_DEIPM      --------------------MTLFDPPAPLAERLRPRTVAEVVGQTHLLG
                                                :***:*****.*******: **.** ****
```

Deide_08770 (3aa) SsrA-binding protein [smpB], TSS at GTG (others also GTG at V); start of E8U8U7_DEIML is also GTG.

```
tr|C1D1M8|C1D1M8_DEIDV      MPRVYTNRRAHYEYELLERFEAGISLTGSEVKSIRAGGVDFRDAFARLHG
sp|Q9RUC1|SSRP_DEIRA        MRRVYTNRRAHHEYELLERFEAGISLTGSEVKSVRAGGVDFRDAFARING
tr|F0RLA9|F0RLA9_DEIPM      MPRVYVNRRAGYEYELLDRYEAGLSLTGSEVKSIRAGGVDFRDAFARLNG
sp|Q1J063|SSRP_DEIGD        MRRVYTNRRAHHEYELLERFEAGIALTGSEVKSVRAGGVDFRDAFARLNN
tr|H8GVQ1|H8GVQ1_9DEIO      MRGVYTNRRAHYEYELLERYEAGISLTGSEVKSIRAGGVDFRDAFARLTN
tr|E8U8U7|E8U8U7_DEIML      ---MYTNRRAHYEYELLERFEAGIQLTGSEVKSVRAGGVDFRDAFARVTN
tr|D7CY66|D7CY66_TRURR      ---MIQNRRASFDYELLERFEAGLVLTGSEVKALRQGGVTLGEAYARVRG
                               :  **** .:****:*:***: *******::* *** : :*:**: .
```

Deide_09360 (3aa) RuvA, TSS at GTG; V in E8U7P4_DEIML is also GTG

```
tr|C1D1T9|C1D1T9_DEIDV      MAGVIAYLSGVVREVRENSAVIVAGGVGYEVQCPAGTLGKLVVGQNAELS
sp|Q1J0F6|RUVA_DEIGD        ---MIAYLSGAVREVREASAVIVAGGVGYEVFCPASTLGRLVPGQPAELN
sp|Q9RUV7|RUVA_DEIRA        ---MIAYLSGVVREVREGSAVVVAGGVGYEVQCPAGMLARLKPGEAAEFS
tr|E8U7P4|E8U7P4_DEIML      MPGVIAYLTGTVRDVRDTSAVIVAGGVGYEVLCPAPTLAKLRVNDTAELH
tr|F0RMI0|F0RMI0_DEIPM      ---MIVYLSGTVREVRTSSAVLQTGGLGYEVFCPQSTLARLKPGEAAELH
                               :*.**:*.**:**  ***: :**:**** **   *.:*  .: **:
```

Deide_09750 (17aa) *Deinococcus*+*Truepera*-specific

```
Deide_09750      --------------MLRRQSRRPQQGTRLPGMAHPEFVGLVNSLQATAEAALGDLNAASA
DR_0889          MLPAASCTGGGFFVTPSFSLFPVRTPASIAGMAHPEFVGLVNSLHATAEAALGDLNAATA
Dgeo_1518        -----MCIEECSEEYARGQTKSRVPRPTLPHMPNLEFVGLVNSLQATAEAALGDLNAATS
Deipr_1288       -------------------------------MANPDFVGLVTSVQATAEAALGQLNAATS
Deima_2235       -------------------------------MSSPEFMGLVQSLQASAEAALGDLNAASA
Trad_1274        -------------------------------MADPRFIGLVHSLLSSAEAALGEEHSPMA
                                                *.   *:*** *: ::******: ::. :
```

Deide_09860 (16aa) Skp (OmpH)

```
Deide_09860      MSCFIRLRKAHDTVTVMKMNAKVLAPLAVVAAFGLGTVSPSAQTPAQKIGFVDVAKLISS
Dgeo_0715        MKGFIRLRKRRASVPAMKMNAKALAPLALVAAFGLGTVAPHAQTAPQKIGFVDVQKLLSA
DR_0989          MTCFIRLRNRHASVAVMKITAKALAPVTLAAAFGLGTLAPHAQTPAQKVGFVNVDALFAA
Deima_1344       ------------------MNVKQMLPVAVVAAFAVGTLAPHAQTAPQKVGFVNVQTVLEA
Deipr_1168       ----------------MNKAAKVLLPLSAVAAVAVATVAPSAQTPAQKVGFVDVDRVFAA
                                     .* : *:: .**..:.*::* ***..**:***:*  :: :

tr|C1CUK0|C1CUK0_DEIDV      MSCFIRLRKAHDTVTVMKMNAKVLAPLAVVAAFGLGTVSPSAQTPAQKIG
tr|Q1J0G7|Q1J0G7_DEIGD      MKGFIRLRKRRASVPAMKMNAKALAPLALVAAFGLGTVAPHAQTAPQKIG
tr|Q9RVN8|Q9RVN8_DEIRA      MTCFIRLRNRHASVAVMKITAKALAPVTLAAAFGLGTLAPHAQTPAQKVG
tr|E8U7F5|E8U7F5_DEIML      ------------------MNVKQMLPVAVVAAFAVGTLAPHAQTAPQKVG
tr|F0RNJ4|F0RNJ4_DEIPM      ----------------MNKAAKVLLPLSAVAAVAVATVAPSAQTPAQKVG
tr|Q5SK25|Q5SK25_THET8      ------------------MKRLPLIGVLLALGALLTPMLAQNKTVASRVG
tr|E8PM02|E8PM02_THESS      ------------------MKRFPLAALLLALGALLTPMLAQNKNVATRLG
tr|F6DG85|F6DG85_THETG      ------------------MKRLPLIGALLALGALLTPMLAQNKTVASRVG
tr|G8N8Y8|G8N8Y8_9DEIN      ------------------MKRFSLAALLLALGALLTPMLAQNKTLSTRVG
tr|B7A6H3|B7A6H3_THEAQ      ------------------MKRLPLAALFLALGALLTPMLAQNKNVATRVG
                                                   :     . .  : .: .  :. . ::*
```

Deide_10430 (8aa) acetylglutamate kinase

```
Deide_10430      -----------------MSYAKVRTMIIVKVGGSAGIDYDAVCADLAARWKAGERLVLVH
Dgeo_0678        -------------MPALLFTCYLLTMIVVKVGGSAGIDYDAVCADLAALWQGGQRFVLVH
DR_1420          MLSRDQHCFTFAKRFSFLVCIRIVNMIVVKVGGSDGIDYDAVCADLAERWQAGEKLILVH
Deima_1346       -------------------------MIVVKVGGSAGIDYDAVCADLAARVQAGERFVLVH
Trad_1399        -------------------------MIVVKVGGSTGIDYDALCEDVAALWREGQRLVLVH
Mrub_2721        -------------------------MIVVKVGGSEGINYEAVAKDAASLWKSGQKLILVH
Ocepr_1796       ---------------------MEDGLIVVKVGGSEGIDYAAVARDAAALWKQGRRLVLVH
Mesil_0435       -------------------------MIVVKVGGSEGINYEAVAKDAASLWKEGQRLVLVH
                                          :*:****** **:* *:. * *   : *.:::***
```

Deide_11030 (8aa) *Deinococcus*-specific

```
Deide_11030      --MWIATLLGMVLWLVVVFILLSATLILALSFGPLKTAENIRVIRMFAAVQYLAALLLAL
Dgeo_1349        ----------MVLWLVVAFIVLSATLILALTLGPLRKAANVRVIQLFAAVQYAAAVLLVG
Deima_2050       ----------MALWLLFAFILMSATLILALTLGPLKTAANVRTIRAFAYVQYAAALLLAG
DR_1429          MRAATRYPPPMLLWFVVIFILLSATGILYLTLGPLKTAANVSTLRAFAAVQYLCAAILAL
Deipr_0258       ----------MLLWVLVGFIVLSASVVLSLTFGALRTSPQVGLFRLIAGVQFLAAAVLAG
                           * **.:. **::**: :* *::*.*:.: ::  :: :* **: .* :*.
```

Deide_14250 (3aa) RecF, TSS at GTG

Start codon in RecF from DEIGI, DEIPD, TRURR, MARHT, MEIRD and MEISD is GTG, and V in RecF from DEIGD, DEIRA, DEIPM, DEIML is GTG; RecF start in OCEP & THET is ATG.

```
RECF_DEIDV      --------MRGVQLESLSTLNYRNLAPCTLSFPAGVTGVFGENGAGKTNLLEAAYLALTG
RECF_DEIGD      --------MSGVQLSSLSTLNYRNLAPGTLHFPAGVTGVFGENGAGKTNLLEAAYLALTG
RECF_DEIGI      -----------MLLSGLSTLNYRNLAPDTLEFPAGVTGVFGENGAGKTNLLEAAYLALTG
RECF_DEIRA      --------MGDVRLSALSTLNYRNLAPGTLNFPEGVTGIYGENGAGKTNLLEAAYLALTG
RECF_DEIPM      --------MAPVRLSKLSTLNYRNLAPDTLEFPAGVTGVWGENGAGKTNLLEAAYLALTG
RECF_DEIML      MPLPRHAYNAHVHLRALTTLHYRNLSPATLDLPRGITSIWGENGAGKTNLLEAAYLALTG
RECF_DEIPD      -----------MRLRALTTLNFRNLTPDTLELPAGLVSVSGANGAGKTNLLEAAYLVLTG
RECF_TRURR      -----------MRLLSLQQLNYRNLNTPRVTFGGGVTAIVGRNAAGKSNLLEAVYLGLTG
RECF_OCEP5      -----------MILTRLRQQNFRNLTSLELVLPPGPLALVGPNASGKTNLLEAIFLALGG
RECF_MARHT      -----------MRLLRFRQRHFRNLRSSELTLAGGPLAVVGANAQGKTNLLEALYLALGG
RECF_MEIRD      -----------MRLLRLRQKNFRNLFTPVFAPGPGLTTVVGGNAQGKTNLLEAIELALGG
RECF_MEISD      -----------MRLLRLRQTHFRNLKSPEFAPAPGLTTVVGGNAQGKSNLLEAIYLALGG
RECF_THET2      -----------MRLLLFRQRNFRNLALEAYRPPPGLSALVGANAQGKTSLLLGIHLALGG
                           : *  :   ::***         *   : * *. **:.** .  * * *
```

Deide_16430 (12aa) (uncharacterized *Deinococcus*-specific)

```
Deide_16430      ----------MCPAASPAYAAAMNVTRHFSDTRANTGRVRFLLQSGRVRLVAEGESWQHH
Dgeo_1997        MPTQDLIILTDAPLTSPAYAERMNVTRHFSDTRTGEGRVRFLITGGRVRLVAEGPGWQLE
DR_0600          ----------------------MNVIRHFSDTRTGEGRVRFLITQGRVRLVAEGPGWSHE
DGo_CA0379       --------------MSARYARGMNVTRHFSDTRTDEGRVRFLVLSGRVVLVAEGQGWQSS
Deipr_0112       ----------MAQLIPRYYAVNMNTARHFSDTRTEEGRVRFLLSDRCVQLVAEGPGWQHC
Deima_1143       ------------------------MFRHLSDTRTDEGRVRVLIDGASVLLRAETHAWQHD
                                           **:****:  ****.*:    * * **  .*.
```

Deide_19140 (19aa**)** aminomethyltransferase or folate-binding protein YgfZ

```
Deide_19140      MLFGTGIAHAWRCIPYPDPMWTFLPSSSLRITGADRVDFVHGQMTGDLRGAPTPGLVPCA
DR_0358          ----MPPTLCPVFSPYPDLMWTRLPSSGLRVTGADRTDFVHGQMTGDLRGAPTPGLVPCA
Dgeo_2127        -------------------MWTRIPSSALRLTGADRVDFVQGQMTNDLRGAPTPGMVACA
DGo_CA2386       -------------------MWTRIPSSSLRVTGPDRVDFVQGQMTGDLRGAPTPGLVAAC
Deima_0616       -------------------MWTRIPSSALRLTGADRVDFVQGQMTNHLKAAPTPGMVPCA
Deipr_1428       ----------------MSTFFTLVPSGALRVTGADRLDFVQGQMTNDLRGCPTPGYVAAC
                                    ::* :**..**:**.** ***:****..*:..**** *...
```

Deide_20630 (4aa) RuvC

```
tr|C1CYQ3|C1CYQ3_DEIDV     MLSDMIVLGIDPGLANLGLGLVDGDIRKARHLHHVCLTTESAWVMPRRLQ
sp|Q1J1K4|RUVC_DEIGD       ----MIVLGVDPGLANLGLGLVEGDVRKARHLYHVCLTTESAWLMPRRLQ
tr|H8GUN8|H8GUN8_DEIGI     -MTGMIVLGIDPGLANLGLGLIEGDIRKARHLHHVCLTTQSAWIMPRRLA
sp|Q9RX75|RUVC_DEIRA       ----MRVLGIDPGLANLGLGLVEGDVRRAKHLYHVCLTTESAWLMPRRLQ
tr|E8U2Z2|E8U2Z2_DEIML     ----MIVLGIDPGLANLGIGVVDGDARKARHLHSVCLFTASAWELPRRLR
tr|F0RMA9|F0RMA9_DEIPM     ----MRVLGVDPGLANLGLGVVDGDVRRAVCLHQECVTTPSSQEMGQRLL
                               * ***:********:*:::** *:*  *:  *: * *:  : :**
```

Deide_22690 (4aa) *Deinococcus*-specific

```
Deide_22690      --------------------------MLKGMNHTRTWSDVYGSACATFEGRAGGHRWLVA
Dgeo_2325        MKTALSFERHRPRGGFFVPARRSIFRILAGMEHKRTWTDVYGSAHAGFEGRAGGHRWLVA
DR_2558          ------------------------MPDDDAMQHTRTWTDVYGSPRASFEGRAGGHRWLVA
Deima_0019       ------------------------------MNLARTWEDAYGSVHAQYEGRAGGHSWLVV
Deipr_0036       ------------------------------MRHISNWTDLHGAVHACFEGRSGGHRWLLA
                                               *.   .* * :*:  * :***:*** **:.
```

Longer *D. deserti* proteins:

Deide_04180 (8aa) ferredoxin; *D. radiodurans* probably longer, too.

```
Deide_04180                ---------MTQGVTITVTGFGEIQAQEGERLVLALERGGVDMLHRCGGV
tr|F0RM49|F0RM49_DEIPM     MTESVIAQQEAQPITLTVEGFGEIEAKSGERLTTALERGGVDILHRCGGV
tr|Q9RSP9|Q9RSP9_DEIRA     MTEPLIAQQQGQPVTIAVEGYGEIQAHSGERLVTALERGGVDILHRCGGV
tr|E8U4J5|E8U4J5_DEIML     ------------MATITIEGFGTVEAHADERLVLALERAGTGVLHRCGGV
                                            : *:* ::*: .***. ****.*..:*******
```

Deide_17480 (30aa) (uvrD/rep helicase)

```
Deide_17480                MPAASPRRPENETHPEFDLEAAHLDGTVAAMLRQIEFWEDRERNAGADLE
tr|H8GW29|H8GW29_9DEIO     ------------------------------MLRQIETWEDRDRNAGADLE
tr|E8U4R4|E8U4R4_DEIML     -----MTTAPHPTHPDHPQEAAHLQGTIQAMLERINAWEDRDRNVGADLE
tr|F0RQR3|F0RQR3_DEIPM     ---------MTHPHPDFPAEEARLGSTVQAMIRRIQILEDRERHGGADEH
tr|F0RR58|F0RR58_DEIPM     ---------MTHPHPDFPAEEARLGSTIQAMIRRIRILEDRERHGGADEH
                                                         *:.:*.  ***:*: *** .
```

```
tblastn on D. gobiensis (H8GW29_9DEIO):
Query  1       MPAASPRRPENETHPEFDLEAAHLDGTVAAMLRQIEFWEDRERNAGADLETSVIMADEAG  60
               MP A PR P   THP+F+ E  HL GTVAA+LRQIE WEDR+RNAGADLETSV MAD A
Sbjct  491664  MPVAEPRTP---THPDFEAEQQHLAGTVAAVLRQIETWEDRDRNAGADLETSVTMADTAE  491834
```

Deide_2p02060 (15aa) uvrA2; *D. radiodurans* homolog is probably shorter than annotated, and that of *D. gobiensis* longer.

```
Deide_2p02060   --------------------MNPGRPYSDPQGGGFVRVRGAREHNLKDISVDVPRDALVV
Deipr_2182      ------------------------MTFSPDLSDGFVRVRGAREHNLKDISVELPRDALVV
Deipe_1722      -------------------------MTFSPESSGFVQVRGAREHNLKDISVSIPRDALVV
DGo_CA0556      -------------------------MTDHPGFSGFVRVRGAREHNLKDISVEVPRDALVV
DR_A0188        ..PTLRLFPPSRRRRSYCRAMTPSRPSPDFPDGGFVQVRGARQHNLKDISVKVPRDALVV
Deima_0190      ---------------------MPRRPRPHTARPGFVQVRGAREHNLKNLHVEFPRDALVV
                                                   ::*****:****:: *..*******
```

Deide_2p02180 (231aa) putative transcriptional regulator

```
Deide_2p02180   MDTLPARLRGRLHEGLHGVVAPVEVGAGQPALRKWARAASWTVLDALPGP-GARRWLWCP
DGo_PB0322      -------------------MAPVEVGAGQPELLAWAQARGWRPTRAAPVH-GTRGWVWWP
Deima_0584      MVLLPARVRARLAVRPTVVVAPVGVGFAQDALLQWATAHGYTVTRDLTGD-ETGLTVWWP
Mrub_0345       MSVLSPSALELLATAPAVVVAPVGIGFAQTDLEAWAHKTRRRILRDPAEFSGLAPTLILP
                                   :*** :* .*  *  **           .        :  *

Deide_2p02180   ATHTDLRSLSG---EHYAALILSMSDLSPDEDDWNAALSGASPDWRQQTFAAFGRWPAAM
DGo_PB0322      RYRSEVQAWAAGAGHAAKVLILSGDELLLDAGEWAAALG--DSDWGRVTFAQSSGWPAAL
Deima_0584      RSRGALETLGN-HADPRAALLLEEADLTYHLEDWANALPGLSADKAADSHAQAEGWPAAL
Mrub_0345       QRRSDLERLNS--SDPRDFLFLRESDLLFSHDEWQQAVS------TQQTYAETGGWPEAL
                  :  :.       .    *:*   :*     :*  *:          :.*    ** *:

Deide_2p02180   ELLARLLAQQGTENLPPVEELHRHPLMSVLVAPYQPSGSLRAAAVQLAAAALVTPAVADG
DGo_PB0322      DPVRGLAGRAAGD-------WAAHPQLQAALAPLLP-DVEQDNYAQLARTPLVTPPVQAL
Deima_0584      PLAAALADHPATD-------LAAHPLAPALLGPLLPPMPLRSAFERLAPAPLVTPDVARL
Mrub_0345       ALLQRIVLQPGES-------LVRHPLCVARLGSLLPKDIPREILAKAAQSPLLIPELYGL
                     :  : . .          **   . :..  *    :    : * :.*: * :

Deide_2p02180   LDVERHHLETLSDEGWLWPSPGGWAFPELLRRTLAPVPDPRRAIRAAQALQAAGHMPEAL
DGo_PB0322      LGVDGAALATLADGGWLWPAPGGWRVPALLRRLLVPALDVTLSAQVAAALSSAGHVGEAL
Deima_0584      LGTSADDVRALVDGGWLTPIPGGWRAPTLLRHLAAPAATARTAERIARALHDAGHTDAAL
Mrub_0345       LGLDDTSVAELYDRGLLYAQGSGLAMPKLLRLYLRGSIPAEVARFIEITLLASGHVTAVL
                *. .   :  * * * * .  .*   * ***           :     :*  :**   .*
```

tblastn on D. gobiensis:

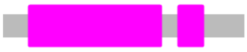
```
Query  1       MDTLPARLRGRLHEGLHG----VVAPVEVGAGQPALRKWARAASWTVLDALPGPGARRWL  56
               M  +  RLR RL    +G    +VAPVEVGAGQP L  WA+A  W    A P  G R W+
Sbjct  259075  MSEVSQRLRQRLEGSGNGPWAGIVAPVEVGAGQPELLAWAQARGWRPTRAAPVHGTRGWV  258896
```

Figure S6. Detected homologs for peptides and proteins from 17 new leaderless transcripts. New gene labels are mentioned. The TSS for these 17 *D. deserti* genes is at the first nucleotide of the start codon. Results of BLASTP, and TBLASTN (if any), and alignments are shown. Non-annotated homologs in *D. radiodurans*, *D. geothermalis* and/or *D. gobiensis* were found for Deide_07364, Deide_14766, Deide_15148 and Deide_23068. Amino acid composition and domains found by SMART are shown for several proteins (horizontal red and pink bars represent signal peptide and low complexity region, respectively; vertical blue bar represents transmembrane helix).

Deide_00694 conserved protein of unknown function (63aa)
MTDKGNEAEQMQEAYAERQEQEQATGKTSAGGAGSTGTPGNQHTGTETTEENDNGPRSGPTEN

Blastp: only one homolog (Deipe_2139)
Tblastn: no more hits

```
Deide_00694     MTDKGNEAEQMQEAYAERQEQEQATGKTSAGGAGSTGTPGNQHTGTETTEENDNGPRSGP
Deipe_2139      MTDKHNEAEEMRDAYAQRQQHEAEGAPTSAGGAGSTSTPGNTEAGTEDTGKNQAGSDQSP
                **** ****:*::***:**::*   . *********.**** .:*** * :*: *. ..*

Deide_00694     TEN------
Deipe_2139      IDPTADAGR
                 :
```



Number of amino acids: 63
Molecular weight: 6570.7
Theoretical pI: 4.26
Amino acid composition:

Ala (A)	6	9.5%	Arg (R)	2	3.2%
Asn (N)	5	7.9%	Asp (D)	2	3.2%
Cys (C)	0	0.0%	Gln (Q)	6	9.5%
Glu (E)	**10**	**15.9%**	**Gly (G)**	**10**	**15.9%**
His (H)	1	1.6%	Ile (I)	0	0.0%
Leu (L)	0	0.0%	Lys (K)	2	3.2%
Met (M)	2	3.2%	Phe (F)	0	0.0%
Pro (P)	3	4.8%	Ser (S)	3	4.8%
Thr (T)	**10**	**15.9%**	Trp (W)	0	0.0%
Tyr (Y)	1	1.6%	Val (V)	0	0.0%

Deide_02488 conserved protein of unknown function (85aa)
MLILDGKYQVQQNKRLTILAEAGHLPKGTLQSDIDALHDDCQAHGRCDVQVNTQHGLMQGTLVEKKPLKFSLWQF EGHLSFPART

Blastp: homolog only of other new leaderless gene Deide_11736 and DGo_CA2102 (86aa)
Tblastn: not more hits

```
Deide_02488     MLILDGKYQVQQNKRLTILAEAGHLPKGTLQSDIDALHDDCQAHG-RCDVQVNTQHGLMQ
Deide_11736     MIILQGTYQVAPTKRLTILAESGHQGKGTLATDIDALSKGCAQGGGKCDITVTTQHGPMT
DGo_CA2102      MLTLEGQYHVAPNKRLTISADTTGLPKGGSLTDLEALSRACLLNNGRCEVQVTTQNGVMQ
                *: *:* *:*  .***** *::    **   :*::**   *   . :*:: *.**:* *

Deide_02488     GTLVEKKPLKFSLWQFEGHLSFPART
Deide_11736     GTLYEKKPRKLSLWQFEGHLSFPQRS
DGo_CA2102      GTLTERPSRQFHRRLFEGYLAFPSRS
                *** *: . ::    ***:*:** *:
```

Deide_11736 conserved protein of unknown function (86aa)
MIILQGTYQVAPTKRLTILAESGHQGKGTLATDIDALSKGCAQGGGKCDITVTTQHGPMTGTLYEKKPRKLSLWQ
FEGHLSFPQRS

Blastp: homolog only of other new leaderless gene Deide_02488 and DGo_CA2102 (86aa)
Tblastn: not more hits

```
Deide_02488      MLILDGKYQVQQNKRLTILAEAGHLPKGTLQSDIDALHDDCQAHG-RCDVQVNTQHGLMQ
Deide_11736      MIILQGTYQVAPTKRLTILAESGHQGKGTLATDIDALSKGCAQGGGKCDITVTTQHGPMT
DGo_CA2102       MLTLEGQYHVAPNKRLTISADTTGLPKGGSLTDLEALSRACLLNNGRCEVQVTTQNGVMQ
                 *: *:* *:*  .***** *::    **   :*::**   *   . :*:: *.**:* *

Deide_02488      GTLVEKKPLKFSLWQFEGHLSFPART
Deide_11736      GTLYEKKPRKLSLWQFEGHLSFPQRS
DGo_CA2102       GTLTERPSRQFHRRLFEGYLAFPSRS
                 *** *: . ::    ***:*:** *:
```

Deide_04802 conserved protein of unknown function (36aa)
MEFLLAGLTIVGSLILASIQHRPQQGRVSVRTSRKG

Blastp: only DGo_CA2417
Tblastn: nothing

```
Deide_04802     MEFLLAGLTIVGSLILASIQHRP-QQGRVSVRTSRKG
DGo_CA2417      MELLLAALATLTALLLASRQAAPRKYARVPVRHHSRR
                **:***.*: : :*:*** *  * : .**.**   :
```

Deide_11672 protein of unknown function, partial (25aa) (pseudogene together with Deide_11671 and Deide_11670; DegV family protein)
MIAVLTDSTSDFSPEAARRGHTTSK

Blastp: N-terminal fragment of DegV family protein

```
Deide_11672      -MIAVLTDSTSDFSPEAARRGHTTSK-----------------------------------
Deipr_0962       -MIAVLTDSTCDLPPAALRDLGAGMLPLEVRLNGQTLRDWEEVTPQQVFGQLER…
LJ_1180          MKIALITDSTSDISPEEAKANDITVVPIPVIIGDKQYMDGVDITAEKLFELERD…
PF01_00648       MKIALITDSTSDISPEEAKANDITVVPIPVIIGDKQYMDGVDITAEKLFELERD…
Deide_12040      MTIAIVTDSTSDLSPELLDHYGIVSVPLYVLFDGKMHKDGIDLTPEELFAGLRA…
Deima_2005       -MIAVVTDSTCDLSPAQLQEQGVTVVPLHVQVGDQQFLDWVELDPDDLYRRMEQ…
                   **::****.*:.*
```

Deide_1p00954 conserved protein of unknown function (74aa)
MASSSLSAVATHVLEFLQQEHQKPRSADELAALLQRDRAEVNRALEELQAAGLVAPEAVSGYGGNDTVWSVTHS

The protein (74aa) has some homology with HTH_11 domain (a Pfam domain. Position: 9 to 63, E-value: 4.9e-05)
Blastp: one good homolog of similar length
Tblastn: not more good hits

```
Deide_1p00954      MASSSLSAVATHVLEFLQQEHQKPRSADELAALLQRDRAEVNRALEELQAAGLVAPEAVS
Deipr_2258         ----MTSSAATQVLEFLTREGPKAHSADELAALLNLDGETVQAALQELHAQGSAAPEEVS
                         *:.**:***** :*  *.:*********: *   *: **:**:* * .*** **

Deide_1p00954      GYGGNDTVWSVTHS-
Deipr_2258         GYGGSETVWRASQVN
                   ****.:*** .::
```

Deide_04426 conserved exported protein of unknown function (58aa)
MKRTIPLLIAALLLASCDDGAETETDTSTSTTTTTSTDQEDTTDTTQSTSTTTTEEEK

Blastp: several hits (lipoprotein signal peptide followed by T-rich region is found in various proteins), but only DR_1317 and DGo_CA2846 of similar size.
Tblastn: not clearly more hits

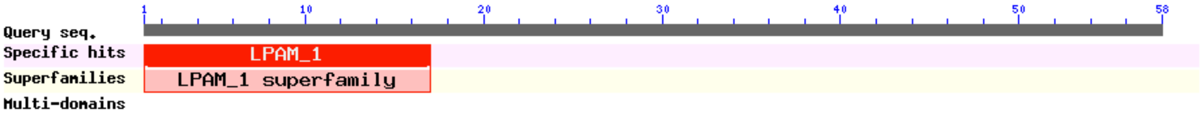
```
Deide_04426       ------------------MKR---TIPLLIAALLLASCDDGAETETDTSTSTTTTTSTDQ
DR_1317           MRCSRSHARSLAAFDNGAMKKAVLAVPALLLALSLSGCQKQADSNTSTSTTTTKSTDSTG
DGo_CA2846        ------------MLDTEAMKR---LLPLLAAALLLAGCSNQGSG-TSSSTTTTKTFDSSG
                                    **:    :* *  ** *:.*.. ..  *.:**:**.: .:

Deide_04426       EDTTDTTQSTSTTTTEEEK
DR_1317           QNTGTTTTSTTTTDTNNK-
DGo_CA2846        QPAGTSTTTTTESNK----
                  : :  :* :*: : .
```

Probable correct start of DR_1317 and DGo_CA2846 are in green.

The D at +2 after the cysteine indicates attachment of the mature protein to the inner membrane.

LPAM_1[pfam08139], Prokaryotic membrane lipoprotein lipid attachment site; In prokaryotes, membrane lipoproteins are synthesized with a precursor signal peptide, which is cleaved by a specific lipoprotein signal peptidase (signal peptidase II). The peptidase recognizes a conserved sequence and cuts upstream of a cysteine residue to which a glyceride-fatty acid lipid is attached.

```
                                  10
                             ....*....|....*..
lcl|local_MKRTIPLLIA       1 MKRTIPLLIAALLLASC 17
Cdd:pfam08139              1 MKKLLLLLLALLLLAGC 17
```



ProtParam of the mature protein starting with C:
Number of amino acids: 42
Molecular weight: 4455.3
Theoretical pI: 3.40
Amino acid composition:

Ala (A)	1	2.4%	Arg (R)	0	0.0%
Asn (N)	0	0.0%	**Asp (D)**	**6**	**14.3%**
Cys (C)	1	2.4%	Gln (Q)	2	4.8%
Glu (E)	**6**	**14.3%**	Gly (G)	1	2.4%
His (H)	0	0.0%	Ile (I)	0	0.0%
Leu (L)	0	0.0%	Lys (K)	1	2.4%
Met (M)	0	0.0%	Phe (F)	0	0.0%
Pro (P)	0	0.0%	**Ser (S)**	**5**	**11.9%**
Thr (T)	**19**	**45.2%**	Trp (W)	0	0.0%
Tyr (Y)	0	0.0%	Val (V)	0	0.0%

(Figure S6, continued)

Deide_14223 conserved protein of unknown function (56aa)
MRIDDHMDLNELAQHMGGATIEQARRMRELLLEKPRARTEDFTGKEWAELVLEATR

Blastp: only one good hit of similar size (DR_0413)
Tblastn: not more hits

```
Deide_14223      -----------------MRIDDHMDLNELAQHMGGATIEQARRMRELLLEKPRARTEDFT
DR_0413          MKKREWEKPERACHTAGMKIDERMDLQELATQMGSEDTAEAARLRDLLLGTGRERTEDFS
                                  *:**::***:*** :**.    :* *:*:*** . * *****:

Deide_14223      GKEWAELVLEATR---
DR_0413          GEEWAELLIRAGEQNK
                 *:*****::.* .
```

Possible correct start of DR_0413 in green

Deide_07364 conserved protein of unknown function (36aa)
MREFLNDWWRLGKLTAATLAIPVVLWGLLVLLGILR

Blastp: three *Deinococcus* homologs (V in Deima_1393 = GTG)
Tblastn: indicates homologs also in other *Deinococcus*, see below

```
Deide_07364      ------------------MREFLNDWWRLGKLTAATLAIPVVLWGLLVLLGILR
Deipr_1862       ------------------MPEWLNDWWRLLKLTVLSLAVPVLLWALLVWAGVLH
Deima_1393       MKVPWRTFAAGRPCYAEEVKEFFNDWWRLLKLIVLSLAVPVVLYLLLLWVGILK
Deipe_3104       ------------------MREFWDYWWRFLKFTVGALAVPVLLYLLLLWLGILR
                                   : *: : ***: *: . :**:**:*: **:  *:*:
```

tblastn:

```
Deinococcus gobiensis I-0, complete genome
Features:
87 bp at 5' side: Phosphoglucomutase, alpha-D-glucose phosphate-specific
194 bp at 3' side: putative Fructose-bisphosphatase

Query  1        MREFLNDWWRLGKLTAATLAIPVVLWGLLVLLGILR  36
                MREFLNDWWRL KLT ATLAIPV LW LLV  G+LR
Sbjct  2750143  MREFLNDWWRLIKLTVATLAIPVALWLLLVWAGVLR  2750250

Deinococcus geothermalis DSM 11300, complete genome
Features:
115 bp at 5' side: phosphoglucomutase, alpha-D-glucose phosphate-specific
187 bp at 3' side: fructose-1,6-bisphosphatase, class II

Query  1        MREFLNDWWRLGKLTAA  17
                MREFLNDWWRLGKL A
Sbjct  1973393  MREFLNDWWRLGKLIAG  1973443

Deinococcus radiodurans R1 chromosome 1, complete sequence
Features:
hypothetical protein (= DR_2037 on opposite strand)
Query  1        MREFLNDWWRLGKLTAATLAIPVVLWGLLVLLGILR  36
                MRE L D WR+ KL A     P+++WG LV +G+L+
Sbjct  2053590  MRELLTDLWRIFKLVAGICIGPLLIWGALVWMGVLK  2053483
```

Further analysis reveals the following entire non-annotated proteins:

```
Dgob MREFLNDWWRLIKLTVATLAIPVALWLLLVWAGVLR
Dgeo MREFLNDWWRLGKLIAGVLALPLLLWGLLVWAGILH
Drad MRELLTDLWRIFKLVAGICIGPLLIWGALVWMGVLK

Deide_07364      ------------------MREFLNDWWRLGKLTAATLAIPVVLWGLLVLLGILR
Deipr_1862       ------------------MPEWLNDWWRLLKLTVLSLAVPVLLWALLVWAGVLH
Deima_1393       MKVPWRTFAAGRPCYAEEVKEFFNDWWRLLKLIVLSLAVPVVLYLLLLWVGILK
Deipe_3104       ------------------MREFWDYWWRFLKFTVGALAVPVLLYLLLLWLGILR
Dgob             ------------------MREFLNDWWRLIKLTVATLAIPVALWLLLVWAGVLR
Dgeo             ------------------MREFLNDWWRLGKLIAGVLALPLLLWGLLVWAGILH
Drad             ------------------MRELLTDLWRIFKLVAGICIGPLLIWGALVWMGVLK
                                   : *     **: *: .     *: ::  *:  *:*:
```

Deide_12656 conserved exported protein of unknown function (70aa)
Protein detected by proteomics (Figure S4).
MTKLLKLLAFSAVLALPVNAGAQETNTTTETTNIEMNERGTDWGWLGLAGLLGLAGLAGRRHVETSTVRR

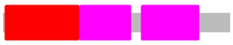
Blastp: many hits (note conserved C-terminal region, which is present in many more homologs of similar size)

```
Deide_12656      -MTKLLKLLAFSAVLALPVNAGAQETNTTTETTNIEMN----------------ERGTDW
Dgeo_1211        -MTRVLKALTLTALLALPVSALAQTDTTTTTATTTTTN------------------GFDW
DR_1067          --MKLLKTVAVVAALALPVAASAQDTNNTTGTTQTTTTTTTE------------KRGFDW
Deipr_2033       -MKKATYTLLLTGLLAAPITASAQTETTSETTTSTTSTPETTTT-----TVERENDGFDW
Deima_0510       MTQRMKHTLLALTLTFAATPAFAQDTTTTGTDTGTTQTTTNN---------DNDNDGFDW
Deipe_0721       MTKMTKTTLTALLLLLAPLPALAQTDTGTTGTTGTTDTTDTANTGTTNTATQNEDRGMDW
                         :        .  * **  . :   *    .                  * **

Deide_12656      GWLGLAGLLGLAGLAG-----RRHVETSTVRR-------------------
Dgeo_1211        GWLGLAGLIGLAGLAG---GSRRYVDTAPGRR-------------------
DR_1067          GWLGLLGLAGLLGLGRQQPAPTVHTTTTTTRR-------------------
Deipr_2033       GWLGLLGLAGLAGRRR---EPEHVVRTAPVHTTPTQTTHTTTTHTNDTTRR
Deima_0510       GWLGLLGLLGLAGLRRQEPPREVHLGGPTDGPRR-----------------
Deipe_0721       GWLGLLGLAGLAGLRKPTPTVVVPDN---TGARR-----------------
                 ***** ** ** *
```

with some more homologs:

```
Deide_12656           ------MTKLLKLLAFSAVLALPVN---AGAQETNTTTETTN----------------IE
Dgeo_1211             ------MTRVLKALTLTALLALPVS---ALAQTDTTTTTATT----------------TT
DR_1067               -------MKLLKTVAVVAALALPVA---ASAQDTNNTTGTTQTT------------TTTT
Deipr_2033            ------MKKATYTLLLTGLLAAPIT---ASAQTETTSETTTSTTSTPETTT-----TTVE
Nos7107_2275          -MKSNLTKALGAGVLTLGMAIMPLTTLPVQAQDN-----TTTTGDAPRTTT-----YD--
N9414_23183           -MTRNFTKAVGAGFLTLSMAMLPLT-LPVNAQVT-----DPRVETTPRTTV-----YE--
Glo7428_3587          MKRSQLSKIFGASVLGLSLAVLPST-LPVSAQTTNTAPGTTDTTTTTAPTT-----TTTT
Ava_2326              -MNRDFSKTVGAAVITLSMATLPLS-LPANAQVQ-----TAPRTDGTTTRT-----YDRT
FJSC11DRAFT_0576      MMKNNLTKMVGASVLTLGMTILPLT-IPAQAQTT-----TDPTINNPNPPN-----TG-V
Deima_0510            -----MTQRMKHTLLALTLTFAATP---AFAQDTTTTGTDTGTTQTTTNN---------D
Deipe_0721            -----MTKMTKTTLTALLLLLAPLP---ALAQTDTGTTGTTGTTDTTDTANTGTTNTATQ
                                   .        .     . **

Deide_12656           MNERGTDWGWLGLAGLLGLAGLAG---------------RRHVETSTVRR---------
Dgeo_1211             TN--GFDWGWLGLAGLIGLAGLAGGS-------------RRYVDTAPGRR---------
DR_1067               TEKRGFDWGWLGLLGLAGLLGLGRQQ------------PAPTVHTTTTTTRR-------
Deipr_2033            RENDGFDWGWLGLLGLAGLAGRRREPEHVVRTAPVHTTPTQTTHTTTTHTNDTTRR---
Nos7107_2275          --RNDFDWGWLGLLGLLGLAGLAGRK--------HNDETTRYRDPNAPG---ATSYRD-
N9414_23183           --RRDFDWGWLGLIGLFGLAGLAGRK--------RGEEPTAYREPTTPG---STTYRD-
Glo7428_3587          ETNDGFDWGWLGLIGLAGLAGLAGRK----------SEPTRYREPDTVGTTTSSTYREP
Ava_2326              ADRNDFDWGWLGLIGLLGLAGLAGKK--------RDDEPTRYRDPSAPG---ASSYRE-
FJSC11DRAFT_0576      YYDRGFDWGWLGLLGLLGLAGLAGRK--------RNDEPTRYRDPNAVG---SSTYRE-
Deima_0510            NDNDGFDWGWLGLLGLLGLAGLRRQE------------PPREVHLGGPTDGPRR-----
Deipe_0721            NEDRGMDWGWLGLLGLAGLAGLRKPT------------PTVVVPDN---TGARR-----
                          . ******* ** ** *
```

SP='YES' Cleavage site between pos. 22 and 23: AGA-QE D=0.848 D-cutoff=0.510 Networks=SignalP-TM.



ProtParam of mature protein starting with Q:
Number of amino acids: 48
Molecular weight: 5270.8
Theoretical pI: 5.65
Amino acid composition:

Ala (A)	3	6.2%	**Arg (R)**	**5**	**10.4%**
Asn (N)	3	6.2%	Asp (D)	1	2.1%
Cys (C)	0	0.0%	Gln (Q)	1	2.1%
Glu (E)	**5**	**10.4%**	**Gly (G)**	**7**	**14.6%**
His (H)	1	2.1%	Ile (I)	1	2.1%
Leu (L)	6	12.5%	Lys (K)	0	0.0%
Met (M)	1	2.1%	Phe (F)	0	0.0%
Pro (P)	0	0.0%	Ser (S)	1	2.1%
Thr (T)	**9**	**18.8%**	Trp (W)	2	4.2%
Tyr (Y)	0	0.0%	Val (V)	2	4.2%

Deide_14766 conserved protein of unknown function (34aa)
MKGLGEFIEWLREVLKGASQPQPQPVPVRVRQRR

Blastp: homology with Deipe_0001 and Deima_1530
Tblastn: also homology with other *Deinococcus*, see below

```
Deide_14766      -------------MKGLGEFIEWLREVLKGAS----QPQPQPVPVRVRQRR---
Deima_1530       -----MSAPYNEAMKALEDFLQKLRELIRAGT----TPKPALVPVPVRTRQPRR
Deipe_0001       MTLIRSERKLILVMDDLKKALSALREALERLLG--AKPQPVPVPVPVRRRR---
syc0626_d        ---------MGLVDQILDRLQDLARRLIEALFGPEAQPEPEPIPVPVRDRR---
                               . *    .  *. :.        *:*  :** ** *:
```

tblastn:

```
Deinococcus radiodurans R1 chromosome 1, complete sequence
Features:
111 bp at 5' side: GTP pyrophosphokinase
45 bp at 3' side: peptidyl-prolyl cis-trans isomerase, FKBP-type

Query  1        MKGLGEFIEWLREVLKGASQPQPQPVPVRVRQR  33
                +KGL EF+EWLR VL G  +P+PQPVP+ VR R
Sbjct  1867050  VKGLSEFLEWLRGVLTGLGEPRPQPVPIPVRTR  1866952


Deinococcus gobiensis I-0, complete genome
Features:
147 bp at 5' side: Peptidylprolyl isomerase FKBP-type
228 bp at 3' side: ppGpp synthetase I, SpoT/RelA

Query  1        MKGLGEFIEWLREVLKGASQPQPQPVP  27
                MKGL EFI+WLRE L+GA QP+P P+P
Sbjct  1110864  MKGLREFIDWLRETLQGAPQPKPVPIP  1110944


Deinococcus geothermalis DSM 11300, complete genome
Features:
242 bp at 5' side: sigma 54 modulation protein/ribosomal protein S30EA
354 bp at 3' side: ABC transporter related protein

Query  1        MKGLGEFIEWLREVLKGASQPQPQPV  26
                MKGL E I+WLRE LKG++ PQP PV
Sbjct  1466393  MKGLRELIDWLREALKGSASPQPVPV  1466316
```

Further analysis reveals the entire non-annotated proteins, included in the multiple alignment:

```
Deima_1530       --------MSAPYNEAMKALEDFLQKLRELIRA-GTTPKPALVPVPVRTRQPRR
Drad             ----------------MKGLSEFLEWLRGVLTG-LGEPRPQPVPIPVRTRERR-
Deide_14766      ----------------MKGLGEFIEWLREVLKG-ASQPQPQ--PVPVRVRQRR-
Dgob             ----------------MKGLREFIDWLRETLQG-APQPKP--VPIPVRVRDRR-
Deipe_0001       MTLIRSERKLILVMDDLKKALSALREALERLLG--AKPQPVPVPVPVRRRR---
syc0626_d        ---------MGLVDQILDRLQDLARRLIEALFGPEAQPEPEPIPVPVRDRR---
Dgeo             ----------------MKGLRELIDWLREALKG-SASPQPVPVPVRVRDRR---
                                 :.   .        : .    *.*   *: ** *
```

Deide_15148 conserved protein of unknown function (91aa)
MSGFSGGGFSFSRSSHGRGGFFAHSRSSGHRGGMVGGLLGHSHSSGRRGHYVQGGHYRQAKRRRSGGCLGAFLVT
AGLAGAGVMGLVSLIA

Blastp: only two good homologs
Tblastn: also homology in *D. geothermalis*

```
Deide_15148     MSGFSGGGFS--FSRSSHGRGGFFA--HSRSSGHRGGMVGGLLGHSHSSGRRGHYVQGGH
DGo_CA2041      MSSHSGRGFFGHSRSSGHRRGGFVSRGHSHSSGHRRGGMMGGLMGGSSSGHRGHYAQGGH
Deima_0925      MSGFSGGSFS---FSRSSGRGARGFRGHSHSHSHSGG---------------HRYGHRGH
                **..** .*       .  **.     **:* .*  *                :* : **

Deide_15148     YRQAKRRRSGGCLGAFLVTAGLAGAGVMGLVSLIA---
DGo_CA2041      FRPQQRR-GLGCLGVFVVGAALLGGGVAGLVSLVA---
Deima_0925      ARHVVRR--GGCLGAFVVGVAVLSGAVAAVGGVFALLA
                 *   **   ****.*:* ..: ...* .: .:.*
```

tblastn:
Deinococcus geothermalis DSM 11300, complete genome
Features:
aminoglycoside phosphotransferase

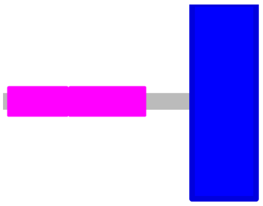
```
Query  18       RGGFFAHSRSSGHRGGMVGGLLGHSHSSGRRGHYVQGGHYRQAKRRRSGGCLGAFLVTAG  77
                R G+  HS SSGH  G +G  LG +H    RGH  + GHYR A  RR  GCLGAFLV  G
Sbjct  1766870  RYGYRGHSHSSGHGAGFLG--LGSAH----RGHDGRHGHYRHAAHRRGFGCLGAFLVGVG  1766709

Query  78       LAGAGVMGLVSLIA  91
                L GA V G++SL+A
Sbjct  1766708  LVGASVTGVLSLLA  1766667
```

Further analysis reveals the following non-annotated Dgeo protein:
MSGFSGGSFSFGHSHSHGRYGYRGHSHSSGHGAGFLGLGSAHRGHDGRHGHYRHAAHRRGFGCLGAFLVGVGLVGASVTGVLSLLA

```
Dgeo             MSGFSGGSFS---FGHSHSHGRYGYRGHSHSSGHGAGFLG------LGSAHRGHDGRHGH
Deima_0925       MSGFSGGSFS---FSRSSGRGARGFRGHSHSHSHSGGHR---------------YGHRGH
Deide_15148      MSGFSGGGFS--FSRSSHGRGGFFA--HSRSSGHRGGMVGGLLGHSHSSGRRGHYVQGGH
DGo_CA2041       MSSHSGRGFFGHSRSSGHRRGGFVSRGHSHSSGHRRGGMMGGLMGGSSSGHRGHYAQGGH
                 **..** .*       .  :*      **:* .*  *                   : **

Dgeo             YRHAAHRRGFGCLGAFLVGVGLVG---ASVTGVLSLLA
Deima_0925       ARHVVRRG--GCLGAFVVGVAVLSGAVAAVGGVFALLA
Deide_15148      YRQAKRRRSGGCLGAFLVTAGLAG---AGVMGLVSLIA
DGo_CA2041       FRPQQRR-GLGCLGVFVVGAALLG---GGVAGLVSLVA
                  *   :*   ****.*:* ..: .   ..* *:.:*:*
```

Amino acid composition <u>of the cytoplasmic domain</u>:
For the cytoplasmic domain only :
Number of amino acids: 67
Molecular weight: 7031.6
Theoretical pI: 12.37
Amino acid composition:

Ala (A)	2	3.0%	**Arg (R)**	**10**	**14.9%**
Asn (N)	0	0.0%	Asp (D)	0	0.0%
Cys (C)	0	0.0%	Gln (Q)	2	3.0%
Glu (E)	0	0.0%	**Gly (G)**	**19**	**28.4%**
His (H)	**7**	**10.4%**	Ile (I)	0	0.0%
Leu (L)	2	3.0%	Lys (K)	1	1.5%
Met (M)	2	3.0%	Phe (F)	5	7.5%
Pro (P)	0	0.0%	**Ser (S)**	**13**	**19.4%**
Thr (T)	0	0.0%	Trp (W)	0	0.0%
Tyr (Y)	2	3.0%	Val (V)	2	3.0%



(Figure S6, continued)

Deide_19985 conserved protein of unknown function (67aa)
MDLDSWTPDDNARRLATLIATAVGVFTFVALWLGASLHALLGLVLGAVLGVVVWFIARRLLVSWFRR

Blastp: only several homologs (35-50%) in *Deinococcus*

```
Deide_19985        ----------------------MDLDSWTPDDNARRLATLIATAVGVFTFVALWLGASLH
Deipr_0048         -------------------MSAMDFNSWRPEDTARRFAIMFATSLGTFGWLAAWLAYGQN
DGo_PC0138         ----------------------MDLNSWTPVDKARRWAVLVAGYLACFILLAVWLGLNWP
Deide_3p01320      ----------------------MDLESWTPKDKARRLAVLVALYLSTMLMVVSVLALKWP
DR_1299            ----------------MAQGVTMNLDSWTPTDKARRLATLIAAYLATSAGLIAALGLHWP
Deipe_2825         MGHWPAALTLHCRSSPERAGGSVDLDSWKPSDVYRRVSIALSVQLGIFVALALVMGFGWP
                                         ::::** * *  ** :  .:  :.    :   :.

Deide_19985        ALLGLVLGAVLGVVVWFIARRLLVSWFRR--
Deipr_0048         VWIGLLAGVAVAAVLYWPLYLILRQVFRR--
DGo_PC0138         WWLSLIAGIAGYFCTFYVVFTLLRSLFRR--
Deide_3p01320      WFVAPLVGAVGYAVAFYVAYAILRNTFRR--
DR_1299            WYLALLSVLVLYGVLYVVGYAVLKAVFRA--
Deipe_2825         WWLGWPLGVLLAAVLHWGAQRWYALRHRRSR
                     :.                      .*
```

Deide_20865 protein of unknown function, partial (72aa)
VPPATPGPEMPHSREWYARLARELGGYRLPWTRVLSGPDPELTFDQKAQCHRKWTEVGVVKRTAIPKAQLHI

Blastp: hits with larger proteins

```
Deide_20865      VPPATPGPEMPHSREWYARLARELGGYRLPWTRVLSGPDPELTFD-------------QK
DR_0468          ---------MNHSRESYDRLARELGGYRHPWARVLSGPDPELTFDLWLSRLLTPQTRVLE
DGo_CA2682       ------MSDLPHSRAWYARLGREQSVYAHPWRRVLSGPDPEETFDGLLAALLTPQAQVLE
                          : ***  * **.** . *  ** ********* ***              :

Deide_20865      AQCHR------------KWT------EVGVVKRTAIPKAQLHI-----------------
DR_0468          AGCGHGPDAARFGPQAARWAAYDFSPELLKLARANAPHADVYEWNGKGELPAGLGA…
DGo_CA2682       AGCGHGPDAARFGARAARWVAYDFVPEWVAAAQANAPHAEVHLWDGRGEVPAPLRG…
                 * * :            :*.       *     ::  *:*:::
```

```
Deide_20865          ------VPPATPGPEMPHSREWYARLARELGGYRLPWTRVLSGPDPELTFD---------
DR_0468              ---------------MNHSRESYDRLARELGGYRHPWARVLSGPDPELTFDLWLSRLLTP
DGo_CA2682           ------------MSDLPHSRAWYARLGREQSVYAHPWRRVLSGPDPEETFDGLLAALLTP
Deima_1024           -----------MSALTPHSREWYAALAARTGGYVHPWRQTLAGPSGEALFDALLEPLLTP
B14911_25565         ----MTKLTSIQGWLAPHSIEWYEQLGKLEGKYLYPWDSFINEPNGESIFDS-EAEELSV
PaelaDRAFT_3864      MSIKWFNPKTHTDWVRPHSIEWYAQLGRLTGQYSYSWKSTITEPNGELIFTN-EVSQMVP
Deipe_2589           -----------MTGKTEAGRAWSDDIARRPGGYSVTWTQWVEGPDAQAIFDA-LVFDRTA
                                       .      :.   . *  .*   :  *. :  *

Deide_20865          ----QKAQCHR------------KWT------EVGVVK---RTAIPKAQLHI--------
DR_0468              QTRVLEAGCGHGPDAARFGPQAARWAAYDFSPELLKLA---RANAPHADVYEWNGK…
DGo_CA2682           QAQVLEAGCGHGPDAARFGARAARWVAYDFVPEWVAAA---QANAPHAEVHLWDGR…
Deima_1024           DTRVLEAGCGHGVDAARFAPRVAHWTGYDFTPASLVRA---QRDVPGATFVEWDSS…
B14911_25565         NQKVLDVGCGEGRFTMHFASFAKEIVGVDASEAFIMEG---HRQRMPNVSFINANT…
PaelaDRAFT_3864      GKKVLDIGCGHGEFALQWSPVVKHIVGIDITSDFIKQG---NDAGRHNVTFITANT…
Deipe_2589           GKIALDCGCGDGAFTLAVARGASSVTGIDFSEGMLAHARVLAAERGMQNVVFVHAH…
                          .  *                .
```

(Figure S6, continued)

```
Deide_23068 conserved protein of unknown function (61aa)
MTDDQKKPQGHDPAEQSPAEGQSHAIPDAAQGKNPGVDPAAKGAPAEGGRDEVEGSSTPGQ

Blastp: several "good" (50%) homologs in Deinococcus only
Tblastn: more hits, see below

Deide_23068      ------------------MTDDQKKP----------------------------------
Dgeo_2289        MKHPMNSMTPSTRKAPVLSGHPKSGPGRRGGEVLSSHSHSSGLQDRPVQRRSESPSRDGL
Deima_0533       ------------------MTDETKQQ----------------------------------
Deipe_1333       ------------------MTHPDERP----------------------------------
                                     .  .

Deide_23068      ----------------------------QGHDPAEQSPAEG-------------------
Dgeo_2289        LPLNKPSRESRLLGRACRQAHVMSDDSKKPYDPANTAPAEGQSHPIPPQDQGNAPNFDPA
Deima_0533       ----------------------------NLPDPADKEQAEG-------------------
Deipe_1333       ---------------------------AEQHDPADTSPADG-------------------
                                             :  ***:   *:*

Deide_23068      -------QSHAIPDAAQGKNPGVDPAAKGAPAEGGRDEVEGSSTPGQ----
Dgeo_2289        NASPAEGQSHPIPPQDRGQNPGVDPAAKDQPAEGSRDDGLPGASTPTASRE
Deima_0533       -------DRQDTPTQDQGQSPHVDPAMNREPAEGGRDEVEGQNG-------
Deipe_1333       ------GNDRNISPNERGQSPHIDPADKDQPAEGGRSEGAAAGS-------
                        : :  .   :*:.* :*** :  ****.*.:
```

tblastn:

```
Deinococcus gobiensis I-0, complete genome
Features:
hypothetical protein
Query  1      MTDDQKKPQGHDPAEQSPAEGQSHAIPDAAQGKNPGVDPAAKGAPAEGGRDEVEGSSTPG  60
              M+DD KK  G+DPA  SPAEGQS  IP+  +GK P  DPAAK  PAEGGRDEVEGSSTPG
Sbjct  43089  MSDDPKK--GYDPANTSPAEGQSRPIPEEDRGKAPNADPAAKDEPAEGGRDEVEGSSTPG  43262

Deinococcus radiodurans R1 chromosome 1, complete sequence
Features:
21 bp at 5' side: endonuclease III
84 bp at 3' side: conserved hypothetical protein
Query  1        MTDDQKKPQGHDPAEQSPAEGQSHAIPDAAQGKNPGVDPAAKGAPAEGGRDEVE  54
                M+DDQKK  G+DPA QSPAEGQSHAIP   +GK+P +DPAAK  PAEGGR+E E
Sbjct  2439859  MSDDQKK--GYDPANQSPAEGQSHAIPAQDRGKDPNIDPAAKDQPAEGGREEAE  2439704
```

Further analysis reveals the following entire non-annotated proteins:

```
Dgob MSDDPKKGYDPANTSPAEGQSRPIPEEDRGKAPNADPAAKDEPAEGGRDEVEGSSTPGA
Drad MSDDQKKGYDPANQSPAEGQSHAIPAQDRGKDPNIDPAAKDQPAEGGREEAEDGAQQSS

Deide_23068       -MTDDQKKPQGHDPAEQSPAEGQSHAIPDAAQ--------------------------GK
Dgeo_2289mod      ---MSDDSKKPYDPANTAPAEGQSHPIPPQDQGNAPNFDPANASPAEGQSHPIPPQDRGQ
Deima_0533        -MTDETKQQNLPDPADKEQAEGDRQDTPTQDQ--------------------------GQ
Deipe_1333        MTHPDERPAEQHDPADTSPADGGNDRNISPNE-------------------------RGQ
Dgob              ---MSDDPKKGYDPANTSPAEGQSRPIPEED--------------------------RGK
Drad              ---MSDDQKKGYDPANQSPAEGQSHAIPAQD--------------------------RGK
                      .    :  ***:   *:*                                    *:

Deide_23068       NPGVDPAAKGAPAEGGRDEVEGSSTPGQ----
Dgeo_2289mod      NPGVDPAAKDQPAEGSRDDGLPGASTPTASRE
Deima_0533        SPHVDPAMNREPAEGGRDEVEGQNG-------
Deipe_1333        SPHIDPADKDQPAEGGRSEGAAAGS-------
Dgob              APNADPAAKDEPAEGGRDEVEGSSTPGA----
Drad              DPNIDPAAKDQPAEGGREEAEDGAQQSS----
                   *  *** :  ****.*.:
```

Deide_2p00483 conserved protein of unknown function (49aa)
MTKKKTGTTSPRVAKKASELLSNPKSAAKVKSVAASALANAADKPKQKK

Upstream of Deide_2p00480 (integrase).
Blastp: several homologs of similar size.
tblastn: no more hits.
Looks bit like N-terminus of HU (also histones among blastp hits).
In some others, gene adjacent to phage-associated genes.

```
Deide_2p00483        ---MTKKKTGTTSPRVAKKASELLSNPKSAAKVKSVAASALANAADKPKQKK-
BN541_00580          -----MGKNEKTSPKVASIASELLRNPKTPKKVKTVAASALTQTADKKKSKK-
Dsui_1484            -MSSKKPTNEHTSARVASTAAKLLSNPRTPASVKSVAASALTQKASSSKAKGK
OR214_01714          -MSSKKPTNEQTSARVASTAAKLLSNPRTPASVKSVAASALTQKASPSKSKGK
TIB1ST10_08240       ---MDTRNTKQTSRPVAKKASALLRDGRTSAKTKSVAASALAQAKPRKGK---
HMPREF9949_1121      ---MDTRNTKQTSRPVAKKASALLRDGRTSAKTKSVAASALAQAKPRKGK---
CDVA01_2128          MTQLAKQNSKQTSPNVARKASAALRDGRSSARTKSVAASALAQARPKRRK---
MHB_29408            -----MAKDEKTNESVASKAAKLLADPTTPPDVKSVAASALTQAPDKKKK---
                            .   *.  **  *:  * :  :.  .*:******::
```

```
Number of amino acids: 49
Molecular weight: 5066.9
Theoretical pI: 10.69
Amino acid composition:
Ala (A)  10     20.4%         Arg (R)   1      2.0%
Asn (N)   2      4.1%         Asp (D)   1      2.0%
Cys (C)   0      0.0%         Gln (Q)   1      2.0%
Glu (E)   1      2.0%         Gly (G)   1      2.0%
His (H)   0      0.0%         Ile (I)   0      0.0%
Leu (L)   3      6.1%         Lys (K)  12     24.5%
Met (M)   1      2.0%         Phe (F)   0      0.0%
Pro (P)   3      6.1%         Ser (S)   6     12.2%
Thr (T)   4      8.2%         Trp (W)   0      0.0%
Tyr (Y)   0      0.0%         Val (V)   3      6.1%
```

Deide_2p01755 conserved membrane protein of unknown function (92aa)
MNPVREWNWKSGAWLLGALLLVVLVYQLSGTHLEAYQVELSLISMILMVLYATDRTFVLWRRGDYRMALGNAFFC
TVALMLQARSLLMMVRS

Blastp: only one homolog
Tblastn: no other hits

```
Deide_2p01755      MNPVREWNWKSGAWLLGALLLVVLVYQLSGTHLEAYQVELSLISMILMVLYATDRTFVLW
Deide_15680        MNPVQEWDWKGGAWLLGALLLAVVIYQSFEAYLEEYQVPISVISTILLVVYMAHRTFVLW
                   ****:**:**.**********.*::**   ::** *** :*:** **:*:* :.******

Deide_2p01755      RRGDYRMALGNAFFCTVALMLQARSLLMMVRS
Deide_15680        RQGDHRMALISAGIMAVVLILRAFSLLVMYRY
                   *:**:**** .* : :*.*:*:* ***:* *
```

Figure S7. New proteins detected by proteomics. New gene labels are mentioned. Tryptic or chymotryptic peptides are indicated. Transcription start site, when found, is mentioned. Results of BLASTP and TBLASTN (if any) and alignments are shown.

Deide_05864 conserved protein of unknown function (76aa)
MLTVKMHLAGGDIIALNMTPSQKNRLSKTINQAQLPTLPFTANVDGVDVEIPWRSISYISSYPQVQSSPVLREAAM

Detected peptides: TINQAQLPTLPFTANVDGVDVEIPWR, MHLAGGDIIALNMTPSQK, TANVDGVDVEIPW, **SKTINQAQLPTLPF.**
Transcription start at -6 of start codon.
Blastp: only one homolog, Dgeo_2254
Tblastn: two additional homologs in *D. deserti*: new predicted Deide_05654 (no identified TSS), unpredicted Deide_11206 (TSS at -6 as for Deide_05864).

```
Deide_05864      MLTVKMHLAGGDIIALNMTPSQKNRLSKTINQAQLPTLPFTANVDGVDVEIPWRSISYIS
Deide_11206      MLSVKLHLAGGDVIALNMTLSQKNRISRTMNQNQLPTTPFTTQVNGLDIEIPWRSIAYLS
Deide_05654      MLTVHLHLAGGDTIALEMSPSQKDRLSRTINQEKLPPLPFVAAINGMTVEIPWRSIAYLS
Dgeo_2254        MLTLNLHLCNGDVVAIQVTSSQRDRISRTLNQAVLPTTPFEVQVAGGTLMIPWRSIGYLS
                 **:::**..** :*:::: **::*:*:*:**  **. ** . : *  : ******.*:*

Deide_05864      SYPQVQSSPVLREAAM--------------------------
Deide_11206      SGPQVQLAQMQQEAAD--------------------------
Deide_05654      SCPQVPN--VALEAAD--------------------------
Dgeo_2254        TQAQAEPELRATEAADCFPVCCRLVPGRVSPAGGVVAALVRRG
                 : .*.       ***
```

One detected peptide (low score) for Deide_11206: TTQVNGLDIEIPW

Deide_13059 protein of unknown function (90aa)
MPILVVTTSKGGRHTYRGSQEVLQEYVDTYQIFSSSGGPNILEFNNSDPNQPSDQISMNIITSMVIHPDDLQSVPEPITDAMVDNAMDKK

Detected peptides: SSSGGPNILEF, RGSQEVLQEY, RGSQEVLQEYVDTY.
Transcription start at 1st nt of start codon.
Blastp: no hits
Tblastn: no hits

Deide_1p00482 protein of unknown function (69aa)
MDEHQVQPYVEALQERGCLVTQHPDGRYSVTLPDGETIEPGAPSVHPQTPWSALIEACSRLNVTVPFGE

Detected peptides: LNVTVPFGE, SVTLPDGETIEPGAPSVHPQTPW, VEALQERGCL, VEALQERGCLVTQHPDGRY.
Transcription start at 1st nt of start codon.
Blastp & tblastn: nothing

Deide_2p01542 protein of unknown function (113aa)
MERPDEVNGLTFNAQRDHDGFRHVEAGFPMPLRPVFALLGQADRSPAHEETTRSSLLKQLRDLKTTAPEAFSKETSGFLTTATFMTNVSPEDEYFNRLLTFLVEAYRKHATSD

Detected peptides: MTNVSPEDEYFNRLL, ALLGQADRSPAHEETTRSSLL
Transcription start at 1st nt of start codon.
Blastp & tblastn: nothing

Deide_15253 conserved exported protein of unknown function (158aa)
MKTLLVTLALLSAPVAYAQTDTTTPETATETTTDVTGTETTGTDTTGTDTTGTDTAGTDATDTDAMGTDTTETDA
TDATGTETTDTTDATETTETTETTEATDTDAATTTSGTTVTETENRSGFPWGLLGLLGLAGLAGRNRATHAHTTT
QTTQTRRP

Detected peptides: SGFPWGLLGLLGLAGLAGR, MKTLL
Blastp: hits with DR_2344 (78/167 = 46%) and Deima_0987 (81/186 = 43%)
(DR_2344 probably too long: start at MKK.. gives predicted signal peptide).
Tblastn: nothing

```
Deide_15253          ----------MKTLLVTLALLSAPVAYAQTDTTTPETATETTTDVTGTETTGTDTTGTDT
DR_2344       MGFPLELTTHQPDKEITMKKSILALTILLGSVAYAQDTGTTTTDTSTTTTDTTGTGTTDTTGTG---
Deima_0987           MRHRLTLALTLALTAPALATVRGPSTLHFTQSTDTGTTTGTDTGTTGTDTSTGTGTTDTT
                               :     :*: : .. :      * .  :: * *..*** *:    *

Deide_15253          TGTDTAGTDATDTDAMGTDTTETDATDA------TGTETTDTTDATETTETTETTEATDT
DR_2344              -TTDTTGTDTTGTDTSGTTTDTSTTTDT------TSTDTTDTNVQNDAVTTSTEADGVPG
Deima_0987           TGTDSAGTTTETAPGTGDDASSATGTDTGTGTTGTDTGTTDTSTGTDDTTGTDTATDTSN
                       **::** :  :   *  :  :  **:      *.* ****.  .: .  :  :  .

Deide_15253          DAATTTSGTTVTETENRS-----------------------GFPWGLLGLLGLAGLAGRN
DR_2344              NEKEPAG-----------------------------------FPWGLLGLLGLAGLMNRG
Deima_0987           SSAGNGGAIPTTPTTTGTNGTVATVNSGNASTNPSDDNNGRGFPWGLLGLLGLLGLAGRR
                     .     .                                   *********** ** .*

Deide_15253          RATHAHTTTQTTQTRRP------------------
DR_2344              RPQPTPVVHTTTTEPRRDTTVVTGTTTTNNDPNRR
Deima_0987           R--HDTVVTPTRNDVR-------------------
                     *     ..  *    *
```

Deide_12656 conserved exported protein of unknown function (70aa)
MTKLLKLLAFSAVLALPVNAGAQETNTTTETTNIEMNERGTDWGWLGLAGLLGLAGLAGRRHVETSTVRR

Detected peptide: GTDWGWLGLAGLLGLAGLAGR
Transcription start at 1st nt of start codon.
Blastp: several hits (note conserved Cter corresponding to detected peptide, which is present in many more homologs of similar size)

```
Deide_12656           -------------------MTKLLKLLAFSAVLALPVNAGAQETNTTTETTN--------
Dgeo_1211             -------------------MTRVLKALTLTALLALPVSALAQTDTTTTTATT--------
DR_1067               --------------------MKLLKTVAVVAALALPVAASAQDTNNTTGTTQTT------
Deipr_2033            -------------------MKKATYTLLLTGLLAAPITASAQTETTSETTTSTTSTPETT
N9414_23183           --------------MTRNFTKAVGAGFLTLSMAMLPLTLPVNAQVT-DPRVE-T--T---
Ava_2326              --------------MNRDFSKTVGAAVITLSMATLPLSLPANAQVQTAPRTDGTTTR---
FJSC11DRAFT_0576      -------------MMKNNLTKMVGASVLTLGMTILPLTIPAQAQTTTDPTIN-NPNP---
asr5071               --------------MKRDFSKTVGAAVLSLSMATLPLSLPANAQVQTAPGTDGTTIR---
Npun_F0452            MVSNFGWHKELSQLMKSNFITAVGAGILTLSMGILPLTLSAQAQTTTDPGAN-T--A---
Deima_0510            ------------------MTQRMKHTLLALTLTFAATPAFAQDTTTTGTDTGTTQT----
                                                .        .    .:

Deide_12656           ---IEMNERGTDWGWLGLAGLLGLAGLAG-----------------RRHVETSTVRR---
Dgeo_1211             ---TTTN--GFDWGWLGLAGLIGLAGLAG---------------GSRRYVDTAPGRR---
DR_1067               -TTTTTEKRGFDWGWLGLLGLAGLLGLGRQQ------------PAPTVHTTTTTTRR---
Deipr_2033            TTTVERENDGFDWGWLGLLGLAGLAGRRREPEHVVRTAPVHTTPTQTTHTTTTHTNDTTR
N9414_23183           PRTTVYERRDFDWGWLGLIGLFGLAGLAGRKRG----------EEPTAYREPTTPGSTTY
Ava_2326              TYDRTADRNDFDWGWLGLIGLLGLAGLAGKKRD----------DEPTRYRDPSAPGASSY
FJSC11DRAFT_0576      PNTGVYYDRGFDWGWLGLLGLLGLAGLAGRKRN----------DEPTRYRDPNAVGSSTY
asr5071               TYDRTADRNDFDWGWLGLIGLLGLAGLAGKKRD----------DEPTRYRDPSAPGASSY
Npun_F0452            PRTTTYDRNDFDWGWLGLLGLFGLAGLAGKKRD----------NEPTAYRDPNAPGATTY
Deima_0510            TTNNDNDNDGFDWGWLGLLGLLGLAGLRRQE------------PPREVHLGGPTDGPRR-
                               . ******* ** ** *                      :

Deide_12656           --
Dgeo_1211             --
DR_1067               --
Deipr_2033            R-
N9414_23183           RD
Ava_2326              RE
FJSC11DRAFT_0576      RE
asr5071               RE
Npun_F0452            RD
Deima_0510            --
```

Deide_3p02615 **(Deide_23165)** conserved protein of unknown function (78aa)
MREFNSVTAF**FGDLAVPGRIEALEGGRGL**MRVSLNGAPDISEGAEAILEMHDGVRFRVAVTERLDDTNEVRMKLLARS

Detected peptides: VAVTERLDDTNEVR, MREFNSVTAFFGDLAVPGRIEALEGGR, FGDLAVPGRIEAL, **FGDLAVPGRIEALEGGRGL**
Blastp: two homologs, Deima_0424 & Deipe_2626
Tblastn: one almost identical protein (76/78 = 97%) in *D. deserti* itself (= new Deide_23165); RNAseq indicates better expression of *Deide_23165* than *Deide_3p02615*.

```
Deide_3p02615       MREFNSVTAFFGDLAVPGRIEALEGGRGLMRVSLNGAPDISEGAEAILEMHDGVRFRVAV
Deide_23165         MREFNSVTAFFGDIAVPGRIEALEGGRGLMRVSLNGAPDISEGAEAILEMHDGVRFRVAV
                    *************:**********************************************

Deide_3p02615       TERLDDTNEVRMKLLARS
Deide_23165         TERLDDTNEVRMKLLARA
                    *****************:


Deide_3p02615       --------------------MREFNSVTAFFGDLAVPGRIEALEGGRGLMRVSLNG---A
Deide_23165         --------------------MREFNSVTAFFGDIAVPGRIEALEGGRGLMRVSLNG---A
Deima_0424          --------------------MTNPQAFTAHFGDTAVPGEIQALEGRGGYMRVHLRAG--S
Deipe_2626          MNLFPPWRHFPATKTGYGDGVTDPNSITAHFDEVSIPATITALEGGGGYMRVTLNWQNTA
                                        : : ::.**.*.: ::*. * ****  * *** *.    :

Deide_3p02615       PDISEGAEAILEMHDGVRFRVAVTERLDDTN----EVRMKLLARS--
Deide_23165         PDISEGAEAILEMHDGVRFRVAVTERLDDTN----EVRMKLLARA--
Deima_0424          VPTAEGTPCELEMHDGARFRMVITEDLGDAGPGARNVRLKLVGRGE-
Deipe_2626          FSPAPGMESELEMHDGGRFRVTLLEQITDTGKTSAEFRMKLLGRGRG
                       : *  . ****** ***:.: * : *:.    :.*:**:.*.
```

Deide_11207 protein of unknown function (70aa)
MDISQVVRATAHHLFKLYWAMFANIENPEEALASAGQAVLLYLDDCGMPAQEAAMLRDEIMLSIPPTRKM

Detected peptide: ANIENPEEALASAGQAVLL
Transcription start at 1st nt of start codon.
Blastp & tblastn: no good hits

Deide_14224 conserved protein of unknown function (90aa)
MTAGLTGPQARVLGALRDGAALIMHTRTERGAFYTLGGRRLSVTLLKDLERLRYVSRSAGAGRTAVAYELTPGGSAALAQWESGNPASRG

Detected peptide: TAVAYELTPGGSAALAQWESGNPASRG
Blastp: two homologs, Deima_2324 & Deipe_1712
Tblastn: no other good hits

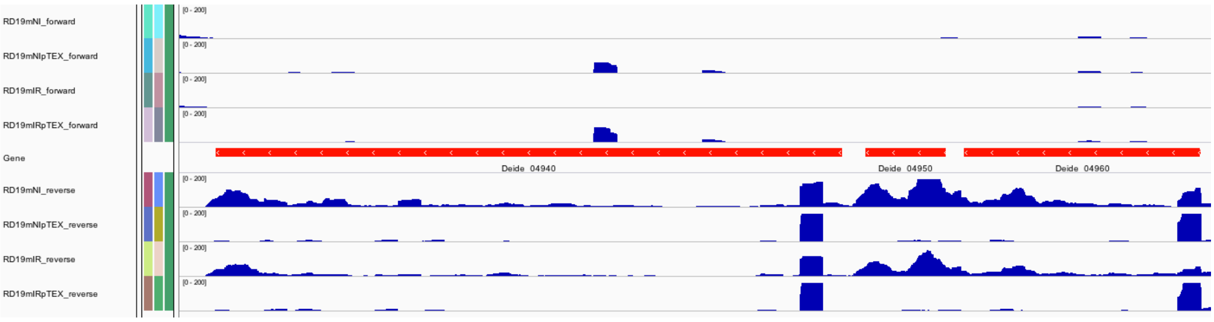
```
Deide_14224         MTAGLTGPQARVLGALRDGAALIMHTRTERGAFYTLGGRRLSVTLLKDLERLRYVSRSAG
Deipe_1712          MSVKASEHQVRVLKALRAGGLLVMHTRGERGPYYTLDGRWLSVTLVKGLEAARLIQREGS
Deima_2342          -MPTLTDAQARVLQALQDGAALTMHARGDRGPYYTLNGRRLSMPLVKSLEANRWIEREGA
                         :  *.*** **: *. * **:* :**.:***.** **:.*:*.**  * :.*...

Deide_14224         AGRTAVAYELTPGGSAALAQWESGNPASRG----
Deipe_1712          SGRVVASYQLTPAGESALAEWEAVRPPLTSEERR
Deima_2342          GRGAASAYQLTAEGESALQAWAAPTPPTH-----
                    .  .. :*:**. *.:**  * :  *.
```

Deide_3p02814 protein of unknown function (142aa)
LTRVPEKTLRRTLRGRGHRHPQKIILLIQGSALSAQNLQTLAIRALTANRLEQRVVLQAISDNLSVQTIKANLSDLKVVAAQVLLGLLAQKTQFIFSIQHGRLHRSCSTPENLKRIEHALDEFRVDRGLVTLTQSKATPLGN

Detected peptide: IQGSALSAQNLQTL
Blastp & tblastn : no good hits

C_571026_3
ORF opposite of *Deide_04940* (*gcvP*, glycine dehydrogenase)

In figure below, Deide_04940 to Deide_04960 are on reverse strand, ORF C_571026_3 is on the forward strand. Deide_04940 has a potential internal TSS (iTSS) and two potential antisense TSS (aTSS) (one with only few reads). The three peptides in C_571026_3 are downstream of these two aTSS. However, antisense reads mainly found with +TEX, without TEX hardly any reads.



C_571026_3 sequence (potential TSS positions, potential start codons and detected peptides):

```
KKTAPKAFGAVFLNSAGLRADVILNRWAGAHEVAVAVHIVHAVDGRPVLPALLMLAGVRGCFAAVGAVPLVGHQV
VLGVGCVLQRAFLRLQQTVLHLLNLTADLQHGVYEAVELGLRLGFGGLDHQGTGHREAHGGRVETVVHQPFGDVL
LGDAARLLEGADINDALVRHAVAFSLVQDGVGARELAGNVVGVEQRHLGGLAQAMGAQQFDVQVADREDARAAVG
RGTYGAGLAAADIAHHVVGQERGKVSLDPDGAHAGATATVRDGKGLVQVQVRDVTADQTWLGHADLCVHVGTVQV
HLTAVLVDQVAHVPDVLFVHTVGAGVGDHQGGEVLAVLLGLGGQVVKVDVAVRVGFHNHHLHAHHGGAGGVGAVC
TGGNQADVPVVLAAALEVLADHQQAGVLALCAAVGLHGNRVVAGDVGQPRFQFRQQLGVAFGLVGRGKGVQRPEF
RPGHRDHLGGGIELHGARTQWNHAVNEAVVAVFQHLQVAQHAVLTAVGLEDRVGQVLTGALEALRNAVHGLGVQR
QHVHSLPGNGFGHVCQISRSHGFVQAHADLVAFIAEVDAFGLRALADSRCVTFEGQGVEEGYAGRQARVLEGSGQ
NAGQAMHAAGNGAQAVGTVVHGVGRGHVGQQRLCGADVAGGLLTADVLLTGLHGHAQGSLAAHVLADADHAARHG
ALVGLLTRQECSVRSTEAHRHAQTLGAAYHDVRALLAGRGDQCAGQQISCDDQCAADCMHLLGHGRQVAQIAVGP
GILDQHAEHPFRQLGSRVAGHHFEAEVLGAGLDHVQGLRVHVVGHEKDVALAFCLTLGQGHGFGGGAFIQQRSV
GDRHASQVHHDLLEVQQHLQTPLTDFSLVRRVGGVPARIFQHVAQNHRRRVGAVIPHADIAAEHLVLLGDAFQVS
QSLSLGDALAGLQLSAELDGFRQRGFGQLVQAGNAQFGQHICLFRLAGAKVTGSEVVGLKQVRQRFHGTP
```

Detected peptides : AGAKVTGSEVVGL, GSRVAGHHFEAEVL, QVSQSLSLGDAL

In sequence below, part of Deide_04940 on reverse strand, part of C_571026_3 on forward strand, with indicated potential TSS positions, potential -10 motif, (potential) start codons, and detected peptides

```
     G  Q  G  V  E  E  G  Y  A  G  R  Q  A  R  V  L  E  G  S  G     F1
1741 GGTCAGGGTGTCGAAGAAGGTTACGCTGGGCGTCAGGCCCGCGTTCTGGAGGGCTCTGGC 1800
     ----:----|----:----|----:----|----:----|----:----|----:----|
1741 CCAGTCCCACAGCTTCTTCCAATGCGACCCGCAGTCCGGGCGCAAGACCTCCCGAGACCG 1800
       T  L  T  D  F  F  T  V  S  P  T  L  G  A  N  Q  L  A  R  A   F4

     Q  N  A  G  Q  A  M  H  A  A  G  N  G  A  Q  A  V  G  T  V     F1
1801 CAGAATGCCGGTCAGGCGATGCACGCGGCCGGCAATGGTGCGCAGGCCGTCGGCACCGTG 1860
     ----:----|----:----|----:----|----:----|----:----|----:----|
1801 GTCTTACGGCCAGTCCGCTACGTGCGCCGGCCGTTACCACGCGTCCGGCAGCCGTGGCAC 1860
       L  I  G  T  L  R  H  V  R  G  A  I  T  R  L  G  D  A  G  H   F4

     V  H  G  V  G  R  G  H  V  G  Q  Q  R  L  C  G  A  D  V  A     F1
1861 GTACACGGCGTAGGCCGCGGCCATGTTGGCCAGCAGCGCCTGTGCGGTGCAGATGTTGCT 1920
     ----:----|----:----|----:----|----:----|----:----|----:----|
1861 CATGTGCCGCATCCGGCGCCGGTACAACCGGTCGTCGCGGACACGCCACGTCTACAACGA 1920
       Y  V  A  Y  A  A  A  M  N  A  L  L  A  Q  A  T  C  I  N  S   F4
     G  G  L  L  T  A  D  V  L  L  T  G  L  H  G  H  A  Q  G  S     F1
1921 GGTGGCCTTCTCACGGCGGATGTGCTGCTCACGGGTCTGCATGGCCATGCGCAGGGCAGT 1980
     ----:----|----:----|----:----|----:----|----:----|----:----|
1921 CCACCGGAAGAGTGCCGCCTACACGACGAGTGCCCAGACGTACCGGTACGCGTCCCGTCA 1980
       T  A  K  E  R  R  I  H  Q  E  R  T  Q  M  A  M  R  L  A  T   F4

     L  A  A  H  V  L  A  D  A  D  H  A  A  R  H  G  A  L  V  G     F1
1981 CTTGCCGCGCACGTCCTTGCTGACGCCGATCACGCGGCCCGGCATGGAGCGCTGGTAGGC 2040
     ----:----|----:----|----:----|----:----|----:----|----:----|
1981 GAACGGCGCGTGCAGGAACGACTGCGGCTAGTGCGCCGGGCCGTACCTCGCGACCATCCG 2040
       K  G  R  V  D  K  S  V  G  I  V  R  G  P  M  S  R  Q  Y  A   F4

     L  L  T  R  Q  E  C  S  V  R  S  T  E  A  H  R  H  A  Q  T     F1
2041 CTCCTGACACGCCAGGAATGCAGCGTGCGGTCCACCGAAGCCCATCGGCACGCCCAGACG 2100
     ----:----|----:----|----:----|----:----|----:----|----:----|
2041 GAGGACTGTGCGGTCCTTACGTCGCACGCCAGGTGGCTTCGGGTAGCCGTGCGGGTCTGC 2100
       E  Q  C  A  L  F  A  A  H  P  G  G  F  G  M  P  V  G  L  R   F4

     L  G  A  A  Y  H  D  V  R  A  L  L  A  G  R  G  D  Q  C  A     F1
2101 CTGGGCGCTGCCTACCACGATGTCCGCGCCCTGCTCGCCGGGAGGGGTGACCAGTGCGCA 2160
     ----:----|----:----|----:----|----:----|----:----|----:----|
2101 GACCCGCGACGGATGGTGCTACAGGCGCGGGACGAGCGGCCCTCCCCACTGGTCACGCGT 2160
       Q  A  S  G  V  V  I  D  A  G  Q  E  G  P  P  T  V  L  A  C   F4

     G  Q  Q  I  S  C  D  D  Q  C  A  A  D  C  M  H  L  L  G  H     F1
2161 GGCCAGCAGATCAGCTGCGACGATCAGTGCGCCGCCGACTGCATGCACCTTCTCGGCCAT 2220
     ----:----|----:----|----:----|----:----|----:----|----:----|
2161 CCGGTCGTCTAGTCGACGCTGCTAGTCACGCGGCGGCTGACGTACGTGGAAGAGCCGGTA 2220
       A  L  L  D  A  A  V  I  L  A  G  G  V  A  H  V  K  E  A  M   F4

     G  R  Q  V  A  Q  I  A  V  G  P  G  I  L  D  Q  H  A  E  H     F1
2221 GGGAGACAGGTCGCGCAGATCGCCGTAGGTCCCGGGATACTGGACCAGCACGCCGAACAC 2280
     ----:----|----:----|----:----|----:----|----:----|----:----|
2221 CCCTCTGTCCAGCGCGTCTAGCGGCATCCAGGGCCCTATGACCTGGTCGTGCGGCTTGTG 2280
       P  S  L  D  R  L  D  G  Y  T  G  P  Y  Q  V  L  V  G  F  V   F4

     P  F  R  Q  L  G  S  R  V  A  G  H  H  F  E  A  E  V  L  G     F1
2281 CCCTTCAGGCAGCTCGGCAGTCGGGTCGCCGGTCACCACTTCGAAGCCGAAGTACTCGGC 2340
     ----:----|----:----|----:----|----:----|----:----|----:----|
2281 GGGAAGTCCGTCGAGCCGTCAGCCCAGCGGCCAGTGGTGAAGCTTCGGCTTCATGAGCCG 2340
       G  E  P  L  E  A  T  P  D  G  T  V  V  E  F  G  F  Y  E  A   F4

     A  G  L  D  H  V  Q  G  L  R  V  H  V  V  G  H  E  K  D  V     F1
2341 GCGGGTCTTGACCACGTCCAGGGTCTGCGGGTGCACGTCGTCGGCCATGAAAAAGACGTT 2400
     ----:----|----:----|----:----|----:----|----:----|----:----|
2341 CGCCCAGAACTGGTGCAGGTCCCAGACGCCCACGTGCAGCAGCCGGTACTTTTTCTGCAA 2400
       R  T  K  V  V  D  L  T  Q  P  H  V  D  D  A  M  F  F  V  N   F4

     A  L  A  F  C  L  T  L  G  Q  G  H  G  F  G  G  G  G  A  F     F1
2401 GCCCTTGCTTTTTGCCTGACGCTTGGCCAGGGTCATGGCTTCGGCGGCGGCGGTGCCTTC 2460
     ----:----|----:----|----:----|----:----|----:----|----:----|
2401 CGGGAACGAAAAACGGACTGCGAACCGGTCCCAGTACCGAAGCCGCCGCCGCCACGGAAG 2460
       G  K  S  K  A  Q  R  K  A  L  T  M  A  E  A  A  A  T  G  E   F4

     I  Q  Q  R  S  V  G  D  R  H  A  S  Q  V  H  H  D  L  L  E     F1
2461 ATCCAGCAGCGAAGCGTTGGAGATCGGCATGCCAGTCAGGTCCATCACGACCTGCTGGAA 2520
     ----:----|----:----|----:----|----:----|----:----|----:----|
2461 TAGGTCGTCGCTTCGCAACCTCTAGCCGTACGGTCAGTCCAGGTAGTGCTGGACGACCTT 2520
       D  L  L  S  A  N  S  I  P  M  G  T  L  D  M  V  V  Q  Q  F   F4
```

(Figure S7, continued)

```
     V  Q  Q  H  L  Q  T  P  L  T  D  F  S  L  V  R  R  V  G  G     F1
2521 GTTCAGCAGCATCTCCAGACGCCCCTGACTGATTTCAGCCTGGTACGGCGTGTAGGCGGT 2580
     ----:----|----:----|----:----|----:----|----:----|----:----|
2521 CAAGTCGTCGTAGAGGTCTGCGGGGACTGACTAAAGTCGGACCATGCCGCACATCCGCCA 2580
      N  L  L  M  E  L  R  G  Q  S  I  E  A  Q  Y  P  T  Y  A  T     F4

     V  P  A  R  I  F  Q  H  V  A  Q  N  H  R  R  R  V  G  A  V     F1
2581 GTACCAGCCCGGATTTTCCAGCATGTTGCGCAGAATCACCGGAGGCGTGTGGGTGCCGTG 2640
     ----:----|----:----|----:----|----:----|----:----|----:----|
2581 CATGGTCGGGCCTAAAAGGTCGTACAACGCGTCTTAGTGGCCTCCGCACACCCACGGCAC 2640
      Y  W  G  P  N  E  L  M  N  R  L  I  V  P  P  T  H  T  G  H     F4

     I  P  H  A  D  I  A  A  E  H  L  V  L  L  G  D  A  F  Q  V     F1
2641 ATACCCCATGCCGATATAGCTGCGGAACACCTTGTTCTTCTGGGCGACGCGTTTCAGGTC 2700
     ----:----|----:----|----:----|----:----|----:----|----:----|
2641 TATGGGGTACGGCTATATCGACGCCTTGTGGAACAAGAAGACCCGCTGCGCAAAGTCCAG 2700
      Y  G  M  G  I  Y  S  R  F  V  K  N  K  Q  A  V  R  K  L  D     F4

     S  Q  S  L  S  L  G  D  A  L  A  G  L  Q  L  S  A  E  L  D     F1
2701 AGCCAGAGCCTGAGCCTCGGTGACGCCCTCGCCGGCCTTCAGCTCTCCGCGGAACTGGAT 2760
     ----:----|----:----|----:----|----:----|----:----|----:----|
2701 TCGGTCTCGGACTCGGAGCCACTGCGGGAGCGGCCGGAAGTCGAGAGGCGCCTTGACCTA 2760
      A  L  A  Q  A  E  T  V  G  E  G  A  K  L  E  G  R  F  Q  I     F4

     G  F  R  Q  R  G  F  G  Q  L  V  Q  A  G  N  A  Q  F  G  Q     F1
2761 GGCTTCAGGCAGCGTGGTTTCGGTCAACTCGTCCAGGCTGGAAACGCCCAGTTCGGCCAG 2820
     ----:----|----:----|----:----|----:----|----:----|----:----|
2761 CCGAAGTCCGTCGCACCAAAGCCAGTTGAGCAGGTCCGACCTTTGCGGGTCAAGCCGGTC 2820
      A  E  P  L  T  T  E  T  L  E  D  L  S  S  V  G  L  E  A  L     F4

     H  I  C  L  F  R  L  A  G  A  K  V  T  G  S  E  V  V  G  L     F1
2821 CATATCTGCCTGTTCCGCCTCGCTGGGGCCAAGGTGACGGGCAGTGAAGTCGTGGGTTTG 2880
     ----:----|----:----|----:----|----:----|----:----|----:----|
2821 GTATAGACGGACAAGGCGGAGCGACCCCGGTTCCACTGCCCGTCACTTCAGCACCCAAAC 2880
      M  D  A  Q  E  A  E  S  P  G  L  H  R  A  T  F  D  H  T  Q     F4

     K  Q  V  R  Q  R  F  H  G  T  P  *                             F1
2881 AAGCAGGTCAGACAGCGGTTTCATGGAACTCCTTAG 2916
     ----:----|----:----|----:----|----:-
2881 TTCGTCCAGTCTGTCGCCAAAGTACCTTGAGGAATC 2940
      L  L  D  S  L  P  K  M  ~~S  S  R  L~~                        F4
```

Figure S8. Examples of several highly induced genes encoding small proteins. Multiple alignments with homologs are shown. For 5 of these *D. deserti* proteins, potential non-annotated homologs in *D. geothermalis* and/or *D. radiodurans* are also indicated. All proteins are of unknown function. Conserved cysteine residues in some of the proteins are indicated.

Deide_04721 conserved protein of unknown function (74aa)

MNRSFRMRRAGSEPAQAFPDSGRGYRHACPSCGQNLTLYDLRDGDQAYWCDPCGKGHRASDPPPGALQPLPDVS

Blastp & tblastn: no hits with *D. radiodurans* & *D. gobiensis*.

```
Deide_04721      ---------------MNRSFRMRRAGSEPAQAFPDSGRGYRHACPSCGQNLTLYDLRDGD
Dgeo_2035        MDWTPAVRRGNPTSLFCYHQNMKRAFRIKADPSPVSGRGYAHVCPFCGQSLALYDLRDGD
Deima_3010       ---------------------MKRRFRE--DPYPASGRGYVHVCPTCTAPMPLYDTRDGD
Deipe_1056       ---------------------MKSRFSE--D-FPASGRGYEHVCPHCGDLLQLHEMRDGD
Deipr_1808       ---------------MKQRYVKNVRKNWKEDPYPASGRSYEYVCPDCGRTMELHDLRDGD
                                       .       :  * ***.* :.** *   : *:: ****

Deide_04721      QAYWCDPCGKGHRASDPPPGALQPLPDVS
Dgeo_2035        QAYWCGPCGKGHRAGELPPGALRPLPEAS
Deima_3010       QAYWCHTCDRGHRASDPPPEALRPLQNAG
Deipe_1056       QAYWCQRCERGHRAGDLPVQALRPLQTAI
Deipr_1808       QGYWCHGCSHGHRAGQPPLAALRRGDDVA
                 *.***  * :****.: *  **:    .
```

with SSDG_06207 = Predicted protein (100aa) of *Streptomyces pristinaespiralis*:

```
Deide_04721      ---------------MNRSFRMRRAGSEPAQAFPDSGRGYRHACPSCGQN-------LTL
Dgeo_2035        MDWTPAVRRGNPTSLFCYHQNMKRAFRIKADPSPVSGRGYAHVCPFCGQS-------LAL
Deima_3010       ---------------------MKRRFRE--DPYPASGRGYVHVCPTCTAP-------MPL
Deipe_1056       ---------------------MKSRFSE--D-FPASGRGYEHVCPHCGDL-------LQL
Deipr_1808       ---------------MKQRYVKNVRKNWKEDPYPASGRSYEYVCPDCGRT-------MEL
SSDG_06207       ---------MGEHRKGAAVTDLDDWYRAYRTVYEDASRGRTVACPHCGARSLRLLFVVRN
                                                    :.*.   .** *          :

Deide_04721      YDLRDGDQAYWCDPCGKGHRASDPPPGALQPLPDVS-------------
Dgeo_2035        YDLRDGDQAYWCGPCGKGHRAGELPPGALRPLPEAS-------------
Deima_3010       YDTRDGDQAYWCHTCDRGHRASDPPPEALRPLQNAG-------------
Deipe_1056       HEMRDGDQAYWCQRCERGHRAGDLPVQALRPLQTAI-------------
Deipr_1808       HDLRDGDQGYWCHGCSHGHRAGQPPLAALRRGDDVA-------------
SSDG_06207       QDDEDGTAAFWCDACLHGLMPTRAPVPPTGERYVKGTESVPDYSLITGD
                  : .**  .:**  * :*  .   *  .
```

Deide_05260 conserved protein of unknown function (62aa)

MNDTEHQSMVGRCDATNCRFNDDMECTAGQIEVQMSGQMAQCITYTPTDGMGESYGATADNR

Tblastn: additional homolog in *D. geothermalis* (see below)

```
Deide_05260      -----MNDTEHQSMVGRCDATNCRFNDDMECTAGQIEVQMSGQMAQCITYTPTDGMGESY
DGo_CA1723       --------MNDTTTVSRCDATTCRFNSDMKCTAGQIEVSMSAHQAQCLTFSPAEGDQGQR
Deipe_3565       MTQNDQMGERQTSIVGACGATDCRYNEDRECHAGQIQVGMAGNMAQCMTYDPTGDQSGMT
Deima_2451       ------MTQQDTSIVGVCEAQDCRFNQERRCHAGQIEVSFSGTQAACMTYSPSGDAQGTG
Deima_2988       ------MTNDTTSIVGTCTAEHCRYNEAQRCTAGQIEVSMDGAHAACATFTPRTDATDQ-
                             : *. * *  **:*.  .* ****:* : .  * * *: *  .

Deide_05260      GATADNR-
DGo_CA1723       PTAQQ---
Deipe_3565       DMPRVNPS
Deima_2451       EQPQQRQ-
Deima_2988       PQPGTNA-
                   .
```

```
DUF1540 (pfam07561) (Cd Length: 40  Bit Score: 35.27  E-value: 7.15e-05)
lcl|local_MNDTEHQSMV 11 GRCDATNCRFNDDMECTAGQIEVQM-----SGQMAQCITY 45
Cdd:pfam07561         1 VACTVTNCAYNEGNECTADAITVGHgsnatTSEETDCATF 40
```

Tblastn on *D. geothermalis*:

123 bp at 5' side: hypothetical protein (Dgeo_0866 = Deide_05250 homolog)
105 bp at 3' side: cation diffusion facilitator family transporter (Dgeo_0867)

```
Query  1       MNDTEHQ----SMVGRCDATNCRFNDDMECTAGQIEVQMSGQMAQCITYTPTDGMGES  54
               M D  HQ    S+VGRCDAT+CR N++ EC AGQIEV +SGQMAQC+TYTP +GMG+S
Sbjct  925654  MEDRSHQNQQASIVGRCDATSCRHNENQECHAGQIEVALSGQMAQCLTYTPQEGMGDS  925827
```

Deide_19965 conserved protein of unknown function (63aa)

```
MQELACTWVPGTLDIVRLKVGTSTIELTSTRLARIFGPQALNDLYLKGRAVVKADARQVAMLA
```

Among the homologs are new *D. deserti* genes identified in this study: Deide_3p00225 and Deide_1p00514.
Tblastn: additional homolog in *D. geothermalis* (see below).

```
Deide_19965       -----MQELACT----------WVPGTLDIVRLKVGTSTIELTSTRLARIFGPQALNDLY
DGo_CA2814        -----MQELSCT----------WVPGTIDVVRLRLGTRNIELTSTRLGRIFGSQALNDLY
Deide_3p00225     -----MIEVKCT----------WIPGTLDMLQLRAGNRHGRLSVHELRRRFGMGAMNSMY
Mrub_0267         -----MQALQQGLVARGGIQCEWVPGTMNQVRVYLPDHQVQISLERLQQIAGIDAVHELY
Deima_1438        -----MTTASHALSVR------WVPGTMNEVQFMLGQALHRIHLSALHRTFGARSSDRLY
Deima_1206        -----MKNWQCS----------WVPGTMNRVQIKGENASTETTIDKLVRAFGAPILTDLY
Deipe_2935        -----MSQLQCS----------WVPGTFNRVRMNSIHDLIEITIERAERILGRGSLHDLY
Deipe_3286        MTQTPAVRFSCA----------WVPATLDRVRVSSPYGTFEVDLQLVRQVLGRPALQALY
Deide_1p00514     --MTASVSVTCS----------WVPGTLDRIRVTCAQHDEVWHIRDVANRYGREALNALY
Deipe_0580        ----MTNVLTCR----------WTLGTLDRVRITTPWVAGEVHVAHIVRLLGRNALEGLY
Deipe_1840        ----MIDTLTCH----------WVPNTVDRVLVVTTHEKAEVHIDRIRRVLGCEALEALY
                                        *   *.: : .               .  *      :*

Deide_19965       LKGRAVVKADARQVAMLA------------------------------------------
DGo_CA2814        LKGRVVLRANPQQIDLLA------------------------------------------
Deide_3p00225     LIGRFQTLAEPGALTGLALS----------------------------------------
Mrub_0267         LKGLVCLPCTSSLCEAFDKA----------------------------------------
Deima_1438        LQGHMTVQVSSSELRTLLR-----------------------------------------
Deima_1206        LRGRAVISTERDLLKILA------------------------------------------
Deipe_2935        LKGRVTLTVNEDVLQRLSA-----------------------------------------
Deipe_3286        LQGRYNVEKPELTLRAILDQLGIELAPQTIAPRTASTCEAQSASGLQALDDPGAPRASET
Deide_1p00514     LKGRYQTHVSRRELLAFP---------------------------FIARTEPKS------
Deipe_0580        LRGSYVLDADEDLLWDVT----------------------------QALFS-LESVASAA
Deipe_1840        LRGHFALSTSGRATSQAM----------------------------QALTASSITFAQAA
                  * *

Deide_19965       --------
DGo_CA2814        --------
Deide_3p00225     --------
Mrub_0267         --------
Deima_1438        --------
Deima_1206        --------
Deipe_2935        --------
Deipe_3286        HLKIRPVH
Deide_1p00514     --------
Deipe_0580        D-------
Deipe_1840        D-------
```

With 3 best blastp hits only :

```
Deide_19965      MQELACTWVPGTLDIVRLKVGTSTIELTSTRLARIFGPQALNDLYLKGRAVVKADARQVA
DGo_CA2814       MQELSCTWVPGTIDVVRLRLGTRNIELTSTRLGRIFGSQALNDLYLKGRVVLRANPQQID
Deima_1206       MKNWQCSWVPGTMNRVQIKGENASTETTIDKLVRAFGAPILTDLYLRGRAVISTERDLLK
Deipe_2935       MSQLQCSWVPGTFNRVRMNSIHDLIEITIERAERILGRGSLHDLYLKGRVTLTVNEDVLQ
                 *.:  *:*****:: *::.      * *  :  * :*   * ****:**..: .:   :

Deide_19965      MLA-
DGo_CA2814       LLA-
Deima_1206       ILA-
Deipe_2935       RLSA
                  *:
```

Tblastn on *D. geothermalis*:

179 bp at 5' side: a/b hydrolase superfamily protease and regulatory beta pr... = Dgeo_0365
87 bp at 3' side: peptidase M20 = Dgeo_0366 (Syntheny with *D. deserti*)

```
Query  1       MQELACTWVPGTLDIVRLKVGTSTIELTSTRLARIFGPQALNDLYLKGRAVVKADARQVA  60
               M EL CTWVPGTLDIVRLKV   TIELTSTRLARIFG QALN+LYLKGR  +KA+ +QVA
Sbjct  372429  MDELLCTWVPGTLDIVRLKVAGRTIELTSTRLARIFGQQALNELYLKGRTTLKANPQQVA  372608

Query  61      MLA  63
               +LA
Sbjct  372609  LLA  372617
```

Deide_2p00980 conserved protein of unknown function (64aa)
MTNERGGSSGSAGGRDPNGDDKTNNGLGDGRRDPGSNHGSPDDREGDGRRNGSESGGGKSQTKD

Tblastn indicates additional homolog in *D. geothermalis* (see below)

```
Deide_2p00980      MTNERGGSSGSAGGRDPNGDDKTNNGLGDGRRDPGSNHGSPDDREGDGRRNGSESGGGKS
DR_A0234           ------MTQAEKKNDPERSHERDNEPKSGGQRDPGDGQPSTDEKNGDGRRNGSESGGGKD
                          :...  .   ...:: *:  ..*:****..: *.*:::**************.

Deide_2p00980      QTKD
DR_A0234           GNS-
                    ..
```

with possible homolog YciG of *E. coli*:

```
Deide_2p00980      MTNERGGSSGSAGGRDPNGDDKTNNGLGDGRRDPGSNHGSPDDREGDGRRNGSESGGGKS
DR_A0234           ------MTQAEKKNDPERSHERDNEPKSGGQRDPGDGQPSTDEKNGDGRRNGSESGGGKD
YCIG_ECOLI         ----MAEHRGGSGNFAEDREKASDAGRKGGQHSGGNFKNDPQRASEAGKKGGQQSGGNKS
                            .   .     ..  :    .*::. *. : ..:  .  *::.*.:***.*.

Deide_2p00980      QTKD
DR_A0234           GNS-
YCIG_ECOLI         GKS-
                    ..
```

with GsiB of *Bacillus* (Glucose starvation-inducible protein B, General stress protein B) (Induction: Glucose or phosphate starvation, and addition of decoyinine. Also by heat shock, salt stress and oxidative stress):

```
Deide_2p00980      ---------------------------MTNERGGSSGSAGGRDPNGDDKTNNGLGDGRRD
DR_A0234           ---------------------------------MTQAEKKNDPERSHERDNEPKSGGQRD
YCIG_ECOLI         -------------------------------MAEHRGGSGNFAEDREKASDAGRKGGQHS
GSIB_BACSU         MADNNKMSREEAGRKGGETTSKNHDKEFYQEIGQKGGEATSKNHDKEFYQEIGEKGGEAT
                                                       .   .     .   :    .*.

Deide_2p00980      PGSNHGSPDDREGD------------------GRRNG-----------SESGGGKSQTKD
DR_A0234           PGDGQPSTDEKNGD------------------GRRNG-----------SESGGGKDGNS-
YCIG_ECOLI         GGNFKNDPQRASEA------------------GKKGG-----------QQSGGNKSGKS-
GSIB_BACSU         SKNHDKEFYQEIGEKGGEATSENHDKEFYQEIGRKGGEATSKNHDKEFYQEIGSKGGNAR
                     . . .                         *::.*            :. *.*. .

Deide_2p00980      ---
DR_A0234           ---
YCIG_ECOLI         ---
GSIB_BACSU         NND
```

for *E. coli* YciG and *B. subtilis* GsiB:
Family: *KGG* (PF10685)
Stress-induced bacterial acidophilic repeat motif.
This repeat is found in proteins which are expressed under conditions of stress in bacteria. The repeat contains a highly conserved, characteristic sequence motif, KGG, that is also recognised by plants and lower eukaryotes and repeated in their LEA (late embryogenesis abundant) family of proteins, thereby rendering those proteins bacteriostatic. An example of such an LEA family is LEA_5, PF00477. Further downstream from this motif is a Walker A, nucleotide binding, motif GXXXXGK(S,T), that in YciG of E coli, eg Q8X7B4 is QSGGNKSGKS [URL]. YciG is expressed as part of a three-gene operon, yciGFE, and this operon is induced by stress and is regulated by RpoS, which controls the general stress-response in E coli. YciG was shown to be important for stationary-phase resistance to thermal stress and in particular to acid stress.

tblastn on *D. geothermalis*:
88 bp at 5' side: UspA
247 bp at 3' side: Rhodanese-like protein

```
Query  21       DKTNNGLGDGRRDPGSNHGSPDDREGDGRRNGSESGGGK  59
                D+ N G  D    PG++H    D+ GDGRRNGSESGGGK
Sbjct  1361039  DRYNEGERDHSPKPGNHHRPDADKNGDGRRNGSESGGGK  1360923
```

Deide_09148 protein of unknown function (30aa)

```
MTRPTARQLQLAMATVLLLTLLGGALGRLL
```

Putative CDS directly downstream of *ddrA*.
Potential, non-annotated homologs are present downstream of *ddrA* in *D. radiodurans* and *D. geothermalis*.

```
Deide_09148      ----MTRPTAR-------------------------QLQLAMATVLLLTLLGGALGRLL
Drad             MTRSLTSAELRGGAAPSVTDPVMPARVSPARLPDTPHLGWAMVNLGLLTLLGGALSRLF
Dgeo             ----MSPATPR-------------------------QLAAVMLGVLTLTLLGGALAKL-
                     :: .  *                         :*  .*  :  ********.:*

# WEBSEQUENCE Length: 30
# WEBSEQUENCE Number of predicted TMHs:  1
# WEBSEQUENCE Exp number of AAs in TMHs: 18.12642
# WEBSEQUENCE Exp number, first 60 AAs:  18.12642
# WEBSEQUENCE Total prob of N-in:        0.65291
# WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE    TMHMM2.0       outside      1     9
WEBSEQUENCE    TMHMM2.0       TMhelix     10    29
WEBSEQUENCE    TMHMM2.0       inside      30    30
```

Deide_11446 protein of unknown function (57aa)

```
MAAARQDTSMHFHIELAQQRAQDLRQEAAQRRLIREAQARKRRKFRFPSLLGHLRLA
```

Putative CDS, or an ncRNA? Located downstream of, and in opposite direction (convergent) of *uvrC*.
Weak homology with DGo_CA1576 (55aa) (similarly present downstream and opposite strand of *uvrC*): Identities = 20/41 (48%), Positives = 25/41 (60%).

```
Deide_11446     VAAARQDTSMHFHIELAQQRAQDLRQEAAQRRLIREAQARKRR-KFRFPSLLG--HLRLA
DGo_CA1576      -----MNPTFHFELTFAAQRAADMRQEAARDRQARAAQPTPLRPTFRWPWTRTRLHPKPA
                      :.::**.: :* *** *:*****: *  * **.   * .**:*      * : *


Deide_11446     VAAARQDTSMHFHIELAQQRAQDLRQEAAQRRLIREAQARKRR-KFRFPSLLG--HLRLA
DGo_CA1576      -----MNPTFHFELTFAAQRAADMRQEAARDRQARAAQPTPLRPTFRWPWTRTRLHPKPA
ACPL_1607       -----MYP--EINLSLARQRGEQLRQEAEAYRRAREA--AGGRPRRRRSVRWPVPRPR--
Mesil_2337      -----MYPNPEMAKKLARERAEAIQQEANQRRLLQEAGLELARPFRHRLAFWLAQLAQRL
                       .  .:   :* :*.  ::***   *  : *     *   :          :

Deide_11446     -------------
DGo_CA1576      -------------
ACPL_1607       ----------PA-
Mesil_2337      DAEVMLKLKIPSR
```

ACPL_1607 and Mesil_2337 are not next to *uvrC*

(Figure S8, continued)

Deide_20580 conserved hypothetical protein (83aa)

MRALDTIAESIRIGFVHPTTVMNTLIQVENEGGLGAVRRIERQLHLGTNALRHRDHPNTALAQTWLSAARAYLIT
QAERRQAV

Located upstream and divergent of *ddrO*$_C$ (these two genes share the palindromic RDRM). Similar gene pair in others, except *D. radiodurans*.

```
Deide_20580      -----MRALDTIAESIRIGFVHPTTVMNTLIQVENEGGLGAVRRIERQLHLGTNALRHRD
Dgeo_0335        -----MRALDIIAESIRVGYVHPTTVLNTLIEAENEGGLGAIRRIERHLSLGLNALRDRQ
Deipe_1112       -----MRALETIAHTIKIGKVHPTTVVNTLIEVENEGGTGALRRVERHLALSEEALRERA
DGo_CA0309       -----MRALDQIAGSIRAGYVHSTTVLNTLIETENEGGLNAVRRVERHLDNGMQAMLQRQ
Deima_0559       MHTLNQRTLECIAETIRTGQAHHTVVLNTLIETENQGGSGALRQLERQLSRSADALQTRQ
Deipr_0092       -----MKALTVMADSLRAGYIHPTHVLNTLIELENAGGTAALREFEAHLTSGRQALTERG
Marky_0724       -----MNELLRIAHRLKHGRVHPNEALNLFIEVDNRAGLEGLHALEEALETALHRLQHRP
Mrub_2304        -----MNVSEIIWKSVGRGAAHPSEVLNALIELDNRKGQIGLWALENELRAKMPLLRPAA
Ocepr_1741       -----MKRIQRLFRTPP--QALPAAMLNLLIEVDNREGRAGLDRLEAEIKAALARLQAAG
Mesil_2926       -----MRDNPAVWRSLGRGKVHPSEVLNALIELDNRRGMLGLEALEAEINEHLPRLSPRA
                       .    :              :* :*: :*  *  .:  .*  :      :

Deide_20580      HPNTALAQTWLSAARAYLITQAERRQAV------------------ 100%
Dgeo_0335        HPHSRLAQTWLGAARAYLVTQAERKQAV------------------ 78%
Deipe_1112       HPHSRLARAWLDATRAYLVAHAECKRAV------------------ 63%
DGo_CA0309       HPRADLVQVWLGATRAYLVSRAEQRQAV------------------ 63%
Deima_0559       HPHTHVARTWLDATRAYLLVNATRKQAV------------------ 56%
Deipr_0092       HPHARLAEAWLQATRFYLQESQRGAA-------------------- 48%
Marky_0724       HPSTRLLARWLEALRVYRSAAYPSPKTPPPLSKEVRRVAY------ 38%
Mrub_2304        RP---LAQAWLEATVLYRTTFYSEGRLSRLFHRFVQPEQRPLPFAS 33%
Ocepr_1741       HPQAARLTLWLKALEAYRRTYHPRPRWTRFLRRPRAFRRAVPASAR 32%
Mesil_2926       QV---LANAWLEAISAYRAAYYPRSALAKIFARIVN—LEPLPKAG  32%
                 :        ** *   *


Deide_20580      -----MRALDTIAESIRIGFVHPTTVMNTLIQVENEGGLGAVRRIERQLHLGTNALRHRD
Dgeo_0335        -----MRALDIIAESIRVGYVHPTTVLNTLIEAENEGGLGAIRRIERHLSLGLNALRDRQ
Deipe_1112       -----MRALETIAHTIKIGKVHPTTVVNTLIEVENEGGTGALRRVERHLALSEEALRERA
DGo_CA0309       -----MRALDQIAGSIRAGYVHSTTVLNTLIETENEGGLNAVRRVERHLDNGMQAMLQRQ
Deima_0559       MHTLNQRTLECIAETIRTGQAHHTVVLNTLIETENQGGSGALRQLERQLSRSADALQTRQ
Deipr_0092       -----MKALTVMADSLRAGYIHPTHVLNTLIELENAGGTAALREFEAHLTSGRQALTERG
                       ::*  :* ::: *  * * *:****: ** **  *:*..* :*  . :*:  *

Deide_20580      HPNTALAQTWLSAARAYLITQAERRQAV
Dgeo_0335        HPHSRLAQTWLGAARAYLVTQAERKQAV
Deipe_1112       HPHSRLARAWLDATRAYLVAHAECKRAV
DGo_CA0309       HPRADLVQVWLGATRAYLVSRAEQRQAV
Deima_0559       HPHTHVARTWLDATRAYLLVNATRKQAV
Deipr_0092       HPHARLAEAWLQATRFYLQESQRGAA--
                 **.: :...** *:* **
```

Deide_07900 conserved protein of unknown function (63aa)

MSNDKNQPAQSDAPQGGDKDTQGLEGIKQVQDQGMQEKGRQVDQTPQDVTGELDGAQPINRRA

Only homologs in two *Deinococcus*.
Tblastn indicates homolog in *D. radiodurans* (see below)

```
Deide_07900      MSNDKNQPAQSDA-PQGGDKDTQGLEGIKQVQDQGMQEKGRQVDQTPQDVTGELDGAQPI
Deipr_1284       -MTDQNHPAQGQTSPEGGDKDTNSLHDIKDVQKQDMLEKGRQMDQTPESVKDTDQGQHPQ
DGo_CA1816       --MNDDHQVPGDK-PGGGQKDTNDLSDIKGVQDTGMARKGQDVQDLPKAVTGEMDGSQPQ
                    :.:: . .:  * **:***:.* .** **. .* .**::::: *: *..  :* :*

Deide_07900      NRRA-
Deipr_1284       QLPRR
DGo_CA1816       NQRR-
                 :
```

tblastn gives also hit with *D. radiodurans*, but no others

Features:

76 bp at 5' side: conserved hypothetical protein
68 bp at 3' side: hypothetical protein

```
Query  1        MSND-KNQPAQSDAPQGGDKDTQGLEGIKQVQDQGMQEKGRQVDQTPQDVTG  51
                MS+D K  P    +P+GG KDT  L  IK +QD GM EK +Q DQTP+ V G
Sbjct  1917919  MSDDAKAMPPAERSPEGGSKDTNDLSDIKGIQDTGMAEKAKQADQTPESVLG  1917764
```

(Figure S8, continued)

Deide_13590 conserved protein of unknown function (77aa)

```
MSDKSTAENMLDAAAAKVNETADRAREAGHNVAHAVTGDAHHKAEALEDRGKAELHNREANAEFHEGKHEATDGD
GH
```

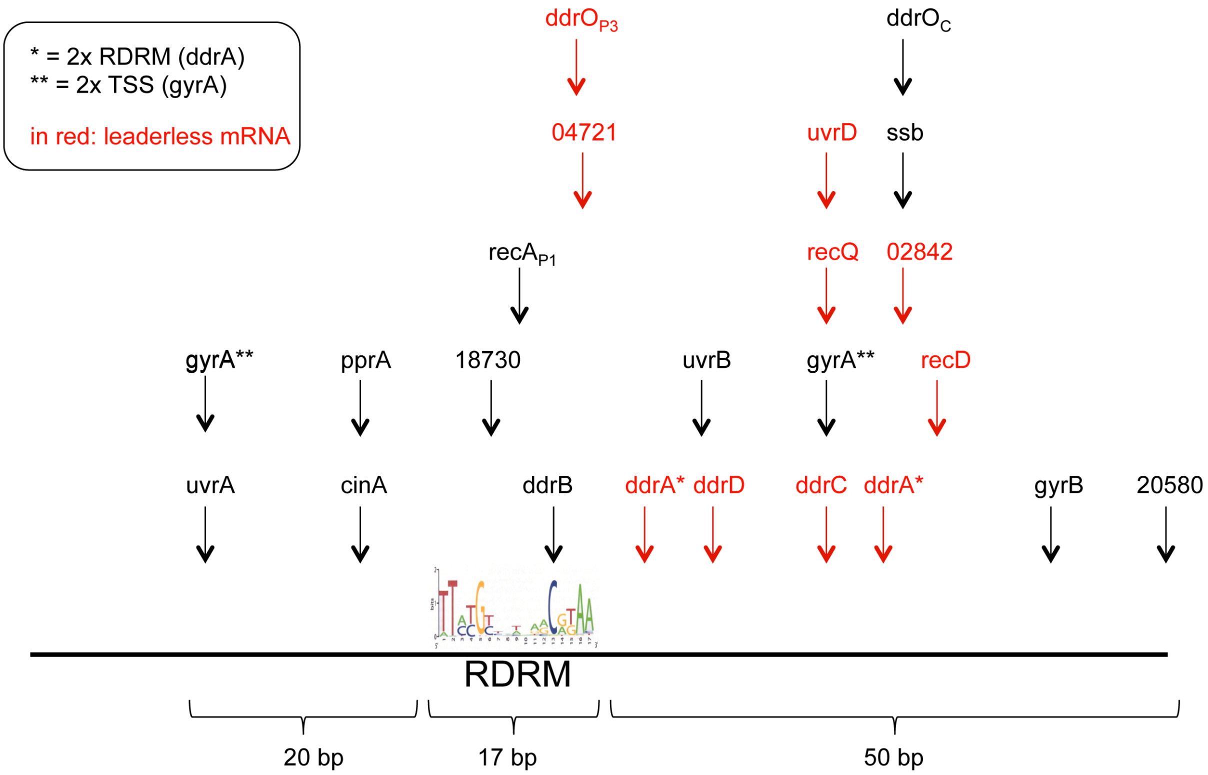
Deinococcus-specific

```
Deide_13590      -----------------------------MSDKSTAENMLDAAAAKVNETADRAREAGHN
Dgeo_1167        --------------------MPYTGGSPMSEDKSALENMVDAAKAKLQEGVDRARAAAHD
DGo_CA1692       -----------------------------------MQNLADAAKAKINEGADRLRAAGHD
DR_1539          MTLTAGRAPGRTPLGLPCRSGNSTSGGQFMSEKTTLDHLADAAGAKLNEVADRARAAGHE
Deipr_0475       ----------------------------MSEDKNVLENLADAAAAKINEGVDRASAAGHN
Deipe_3116       --------------------------MSDQSLGDKLGNAADAVKHKVNEAADRARAEGHD
Deide_13821      --------------------------MTEKSMGERLGEAVDSAKHKVNEMADRTRAEGHE
                                                      .  *:.  *::* .**    .*:

Deide_13590      VAHAVT----GDAHHKAEALEDRGKAELHNREANAEFHEGKHEATDGDGH-
Dgeo_1167        VASNFG-GTADNLKDKAQAAEDRAKAEVHNAQAHAAYNEGKREAQDGDGH-
DGo_CA1692       VASKVGNDHVDNAADKVKATEDRARAELHNREAHAEYNEGKRESKDGDGH-
DR_1539          VAARVSDSPLDTASEKVKAGVDRAKAGIHNAEAHASYDEGHREATDGDGH-
Deipr_0475       VASRDG-NLLDNAADKLHEGADRARAEANNVDARSSFDKAKDQISDALNGK
Deipe_3116       AKSQTSDNPIESLVEKGKAALDRGKAEAHEHQADRDARDAGR---------
Deide_13821      FKAETSDSPVERAVEKGKATVDHSKAELHEAASEKQARDAGR---------
                               .* .   *:.:*  ::  :     ..
```

without Deipe_3116 and Deide_13821:

```
Deide_13590      -----------------------------MSDKSTAENMLDAAAAKVNETADRAREAGHN
Dgeo_1167        --------------------MPYTGGSPMSEDKSALENMVDAAKAKLQEGVDRARAAAHD
DGo_CA1692       -----------------------------------MQNLADAAKAKINEGADRLRAAGHD
DR_1539          MTLTAGRAPGRTPLGLPCRSGNSTSGGQFMSEKTTLDHLADAAGAKLNEVADRARAAGHE
Deipr_0475       ----------------------------MSEDKNVLENLADAAAAKINEGVDRASAAGHN
                                                     ::: *** **::* .**   *.*:

Deide_13590      VAHAVTG----DAHHKAEALEDRGKAELHNREANAEFHEGKHEATDGDGH-
Dgeo_1167        VASNFGG-TADNLKDKAQAAEDRAKAEVHNAQAHAAYNEGKREAQDGDGH-
DGo_CA1692       VASKVGNDHVDNAADKVKATEDRARAELHNREAHAEYNEGKRESKDGDGH-
DR_1539          VAARVSDSPLDTASEKVKAGVDRAKAGIHNAEAHASYDEGHREATDGDGH-
Deipr_0475       VASRDGN-LLDNAADKLHEGADRARAEANNVDARSSFDKAKDQISDALNGK
                 **    .       .* .   **.:*  :* :*.: :.:.: :  *. .
```

Figure S9. TSS positions relative to RDRM (radiation-desiccation response motif). For the different radiation-induced genes, the arrows indicate the TSS position relative to the 17-bp RDRM. Either gene names or gene numbers (without “Deide_”) are shown.



* = 2x RDRM (ddrA)
** = 2x TSS (gyrA)
in red: leaderless mRNA
ddrO$_{P3}$
ddrO$_C$
04721
uvrD
ssb
recA$_{P1}$
recQ
02842
gyrA**
pprA
18730
uvrB
gyrA**
recD
uvrA
cinA
ddrB
ddrA*
ddrD
ddrC
ddrA*
gyrB
20580
RDRM
20 bp
17 bp
50 bp