

HW_6

нашли в NCBI геномы esoli и человека и скачали их в свою директорию
нашли HMM для днк полимеразы
с помощью хммер ищем белки, содержащие этот домен в геномах обоих организмов

ecoli

```
Query:      DNA_pol_A  [M=381]
Accession:  PF00476.25
Description: DNA polymerase family A
Scores for complete sequences (score includes all domains):
--- full sequence ---  --- best 1 domain ---  -#dom-
E-value  score  bias  E-value  score  bias  exp  N  Sequence  Description
-----
4e-171  567.1  0.0  5.2e-171  566.7  0.0  1.1  1  NP_418300.1  DNA polymerase I [Escherichia coli str. K-12 sub

Domain annotation for each sequence (and alignments):
>> NP_418300.1  DNA polymerase I [Escherichia coli str. K-12 substr. MG1655]
#    score  bias  c-Evalue  i-Evalue  hmmfrom  hmm to  alifrom  ali to  envfrom  env to  acc
---
1 !  566.7   0.0  1.2e-174  5.2e-171      3    380  ..    551    925  ..    549    926  ..  0.98

Alignments for each domain:
== domain 1  score: 566.7 bits; conditional E-value: 1.2e-174
      HHHHHHHHHHHHHHHHHHHHHSSCE-STTTCCCHHHHTCCCS.TSSBSTTS.BTTTHHHHHHTS.T-HHHHHHHHHHHHHHHHTTTTGGGG  CS
DNA_pol_A  3  aleelseeleeelkelekeiyelageefnlnSpkqLgeVlfekllpkkkkktktg.ystdeevLeklaeehpivkkileylrelaklkstyvdallkl 99
      l++ seel+ +l+elek+++e+ageefnl+S+kql+++Lfek +++++ kkt  g +st+eevLe+la ++p++k ileyr laklksty+d+l+ +
NP_418300.1 551 VLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGIKPLKKTGGAPSTSEEVLEELA-LDYPLPKVILEYRGLAKLKSTYTDKLP 647
      68999*****99*****9999*****5.69*****PP

      CBTTTSEE--EESSSSSSSS--EESS--GGGS--TSCCHHHHHHTBCGSTTCCEEEEEETTHHHHHHHHCT-HHHHHHHHTH-HHHHHHHHCTS  CS
DNA_pol_A 100 inpktgrihtsfngqvtvtaTGrLsssePNLqnipirteegreiRkafvaeegvlllsaDYsqiELrilaHlsgdeeleafkegeDiHtatAsevf 197
      inpktgr+hts+++q+vtatGrLss++PNLqnip+r+eeegr+iR+af+a+e +++++saDYsqiELri+Ahls+d+ L++af+eg+DiHtatA+evfg+
NP_418300.1 648 INPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPIVRNEEGRRIRQAFIAPE-DYVIVSADYSQIELRIMAHLSDKGLLTAFAGKDIHRATAAEV 744
      *****9.799*****PP

      -GGGHHHHHHHHHHHHHHHHHTT--HHHHHHHHHTSSHNNHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCESCEETTCTSEEEECTTTT-CCHCCCHHHHHH  CS
DNA_pol_A 198 pleevtkeqRrkaKavnFgiiYGisafgLaeqLgisrkeAkeliekYferypgvkeyleetveeakekgyvetllgRrrylpeinsknkalrsaaera 295
      ple+vt+eqRr+aKa+nFg+iYG+safgLa+qL+i+rkeA+++++ Yferypgv ey+e+t+++ake+gyvetl gRr ylp+i+s+n a+r+aaera
NP_418300.1 745 PLETVTSEQRRSKAINFGLIYGMSAFGLARQLNIPRKEAQYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRRLYLPDIKSSNGARRAAERA 842
      *****PP

      HHHHHHHHHHHHHHHHHHHHHHHHTT--SEEEEEETTEEEEEECCHHHHHHHHHHHHHHHHTTSSS--EEEEEESSCT  CS
DNA_pol_A 296 AiNapiQGsaADiiklamikvekalkeeklkarllllqvHDELvfevpeeeleevaevlveeemekeaaavklkVplkvevkvGknwg 380
      AiNap+QG+aADiik+ami+v+++l++e+ ++r+++qvHDELvfev+++++++va+ +++ me+ ++ l+Vpl vev++G+nw+
NP_418300.1 843 AINAPMQGTAADIKRAMIAVDLWLAQEQPRVRMIMQVHDELVFVHKDDVDVAVAKQIHQLMENCTR--LDVPLLEVVGSGENWD 925
      *****66666.*****8 PP
```

human

```

Query:      DNA_pol_A [M=381]
Accession:  PF00476.25
Description: DNA polymerase family A
Scores for complete sequences (score includes all domains):
--- full sequence ---    --- best 1 domain ---    -#dom-
E-value  score  bias    E-value  score  bias    exp  N  Sequence      Description
-----
3.3e-123  414.3  0.1    5e-123  413.7  0.1    1.3  1  NP_861524.2    DNA polymerase nu [Homo sapiens]
7.4e-104  350.6  0.1    1.7e-102  346.1  0.1    2.9  1  NP_955452.3    DNA polymerase theta [Homo sapiens]
1.2e-14   57.0  0.0    3.5e-06   29.1  0.0    2.2  2  NP_001119603.1 DNA polymerase subunit gamma-1 [Homo sapiens]
1.2e-14   57.0  0.0    3.5e-06   29.1  0.0    2.2  2  NP_002684.1    DNA polymerase subunit gamma-1 [Homo sapiens]

Domain annotation for each sequence (and alignments):
>> NP_861524.2 DNA polymerase nu [Homo sapiens]
#    score  bias  c-Evalue  i-Evalue  hmmfrom  hmm to  alifrom  ali to  envfrom  env to  acc
---
1 !  413.7  0.1  1.5e-127  5e-123   2       380  ..      448     853  ..      447     854  ..  0.91

Alignments for each domain:
== domain 1 score: 413.7 bits; conditional E-value: 1.5e-127
      HHHHHHHHHHHHHHHHHHHHHSSCE-STTTCCCHHHHTCCCCS.T.....SSBS..TTSBTTTHHHHHHHHTS.T-HHHHHHHHHHHHHHHHHH CS
DNA_pol_A 2 ealeelseeleelkelekeiyeIageefnLnSpkLgeVLfeklkIpk.....kkkt..ktgystdeevLeklaeehpivkkileyrElaklkst 91
      e+e++s+ l+++lkele+e++ +age+f ++S++qL+e+Lf kIkl+ +t ++ +st+e+vL++L ++ hp++k+ileyr+++k+st
NP_861524.2 448 EEMEKTsALLGARLKELEQEAHFVAGERFLITSNNQLREILFGKLLHLsqRnsLPRTgIqKYPSTSEAVLNAL-RDLHPLPRIILEYRQVHKIKST 544
      7899*****99*****665443432244545699*****.56*****PP

      TTTTGGGGCBTTTSEE--EEESSSSSSS-EEESS-GGGs-.....TTSCCHHHH....HHTBCGSTTCCEEEEEETTHHHHHHHHHCT- CS
DNA_pol_A 92 yvdallklipktgrihtsfngvtvtaTGrLsssePNLqnip.....irteegrei....RkafvaeegvlllsaDYsqiELriLhlsGd 173
      +vd+ll+++ k+g i +++nqt++TGrLs+++PN+q i+ + +e + i R++fv+++ g+++l+aD+sqiELriL hlsGd
NP_861524.2 545 FVDGLLACM--KKGSISSWNQGTGTGTGRLSAKHPQIQGIsKhpqiqtppknFKGKEDK-ItisprAMFVSSK-GHTFLAADFSQIELRIILTHLSGD 638
      *****.59*****9888887776653333333.235588*****.789*****PP

      HHHHHHHHT..TH-HHHHHHHHHCTS-GGGHHHHHHHHHHHHHHHHHTT--HHHHHHHHHTSSHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCESCCE CS
DNA_pol_A 174 eELieafke..geDihTatAsevfGvpleevtkEqRrkaKavnFgiiYGisafGLaeGLgisrkeAkeliekYferypgykeyleetveeakekgvve 269
      +eL++ f+e ++D+++++s++ +vp+e+vt+ +R++K+v+++++YG+++++La lg++ +eA++++e+++++y++k++ + ++++++g+v
NP_861524.2 639 PELLKLFQEsERDDVFSTLTSQWKDVPVEQVTHADREQTKVYVAVVYGAGKERLAACLGVPiQEAQFLESFLQKYKKIKDFARAAIAQCHQTGCVV 736
      *****995567*****PP

      ETTCEEEECTTTT-CCHCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHT.T-SEEEEEETTEEEEEEECCCHHHHHHHHHHHHHHHHT... CS
DNA_pol_A 270 fllgRrrylpeinsknkalrsaaeraAiNapiQGsaADiikLamikvekalkee.klkarlllvqVHDELvfevpeeeleevaelvkeemekeaaV... 363
      +++gRrr+lp+i+++++lr++aer+A+N+++QGsaAD++kLami+v ++++++ +L+arl++q+HDEL+fev++ ++ e+aLv+++me+ ++v
NP_861524.2 737 SIMGRRRLPLRIHANDDQLRAQAEQAVNFVVGQSAADLCKLAMIHVFTAVAAShtLTARLVAQIHDELLFEVEDPQIPECAALVRRTMESLEQVqal 834
      *****776377*****988555334 PP

      ..TSSS--EEEEEEESSCT CS
DNA_pol_A 364 ..kIkVpIkvekvGknwg 380
      +L+VpIkV+++G++wg
NP_861524.2 835 eLQLQVPLKVSLSAGRSWG 853
      4478*****7 PP

>> NP_955452.3 DNA polymerase theta [Homo sapiens]
#    score  bias  c-Evalue  i-Evalue  hmmfrom  hmm to  alifrom  ali to  envfrom  env to  acc
---
1 !  346.1  0.1  1.7e-102  346.1  0.1    2.9  1  NP_955452.3    DNA polymerase theta [Homo sapiens]

```

взяли белки с наименьшим E-value

и выравнивали с помощью muscle

```

>NP_955452.3 DNA polymerase theta [Homo sapiens]
MN-----LLRRSGKRRRS-ESGSDSFSGSGGDSSASPQ-----FLSGSVLSPPPGLGRCLKAAAAG
ECKPTVPDYERDKLLLANWGPKAV
>NP_418300.1 DNA polymerase I [Escherichia coli str. K-12 substr. MG1655]
MVQI-----PQNPLILVDGSSYLRYA-YNAFPPLTNSAGEPTGAMYGVNLMLRSLIMQYKPT-----HAAVVF
DAKGKT---FRDEL----FEHYKSH
>NP_861524.2 DNA polymerase nu [Homo sapiens]
MENYEALVGFDLCNTP--LSSVAQKIMSAMHSGDLVDSKTWGKSTETME-----VINKSSVK-----YSVQLE
DRKTQS---PEKKDL---KSLRSQT

```

ВЫВОД

На основе выравнивания видно, что ДНК-полимеразы человека и *ecoli* имеют как консервативные участки, так и значительные различия. Консервативные области (например, RA-YNAFPPLT у *ecoli* и аналогичные мотивы у человека) указывают на общность функций, таких как репликация ДНК. Однако человеческие белки содержат длинные вставки и дополнительные домены, которые могут быть связаны с регуляцией активности или специализированными функциями, например, репарацией ДНК