209

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

DETERMINACION DE DOMINIOS CONSERVADOS EN LOS TRES LINAJES CELULARES DE DNA POLIMERASAS.

T E S I S

QUE, PARA OBTENER EL TITULO DE:

B I O L O G O

P R E S E N T A

HECTOR GILBERTO VAZQUEZ LOPEZ

DIRECTOR DE TESIS:

DR. ANTONIO EUSEBIO LAZCANO-ARAUJO REYES

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"Determinación de dominios conservados en los tres linajes celulares de DNA polimerasas"

realizado por Héctor Gilberto Vázquez López

con número de cuenta 9650458-5 , pasante de la carrera de Biología

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario Dr. Antonio Eusebio Lazcano-Araujo Reyes

Propietario Biol. Luis José Delaye Arredondo

Propietario Dr. Victor Manuel Valdés López

Suplente Dr. Luis Felipe Jiménez García

Suplente Dr. Arturo Carlos II Becerra Bracho

FACULTAD DE CIENCIAS
U.N.A.M.

Consejo Departamental de Biología

Dra. Patricia Ramos Morales

DEPARTAMENTO
DE BIOLOGIA

Agradecimientos:

Agradezco a mis padres, familiares y amigos que resistieron mis desveladas y continua charla de mi tema de tesis, gracias por su paciencia y opiniones objetivas hechas a mi trabajo

A Sara, Luis y a todos compañeros de laboratorio que estuvieron siempre apoyándome y compartiendo conmigo su conocimiento y su amistad, por escucharme como compañero colega y amigo.

Por último agradezco a mi director de Tesis, Antonio Lazcano por sabia tutela y amistad siempre presente durante la realización de este proyecto y a las instituciones que me apoyaron económicamente para poder realizar este proyecto como son PROBETEL y Fundación UNAM.

Dedicatoria:

Quiero dedicar este trabajo de tesis a todas las personas que me han ayudado a crecer y armonizar mi mente, mi trabajo y mi espíritu; que han hecho posible que me desarrolle a la par de que aprenda de mis vivencias y experiencia; a todos los forjadores de almas que me han acompañado en mi vida, ayudándome a enfrentar las adversidades y enseñarme la delicada labor de ser feliz. Que la Sagrada Presencia nos depare un futuro de cambio y de avance en todos aspectos. Para todos aquellos que hemos viajado juntos en esta senda de conocimiento les doy las gracias. Nunca hay que olvidar que siempre es posible aprender de nosotros mismos y de los demás. V Ad G JL, JLS S Lu Ch Ab R M IzTa. Gracias...

INDICE

Resumen:

Las DNA polimerasas son un grupo de enzimas esenciales para los procesos de replicación y reparación del genoma, estas pueden dividirse en cuatro grupos principales, los cuales se distribuyen dentro de los tres dominios, Archaea, Bacteria y Eucarya. A pesar de los intentos por comprender los posibles eventos de evolución que dieron origen a esta gran diversidad, no había sido posible evidenciar rasgos que permitan aclarar como se originaron cada uno de los grupos de DNA polimerasas. Recientemente, por comparaciones llevadas a cabo con las estructuras terciarias con representantes de tres de los cuatro grupos de polimerasas (Steitz, 1999) ha sido posible el evidenciar un rasgo común que permite relacionar a las DNA polimerasas tipo I con las DNA polimerasas tipo II: el dominio palm. Posteriormente dentro del reporte de Delaye, Vázquez y Lazcano, (2000), expuesto para este trabajo de titulación fue posible el evidenciar este rasgo, con un muestreo con DNA polimerasas tipo I y II de los tres linajes celulares identificando por estructura terciaria a una región del dominio palm homóloga para los dos grupos de proteínas-. Con esto fue posible suponer por su conservación y distribución que los mecanismos de polimerización se pudieron haber presentado en momentos evolutivos tempranos. En la continuación de este trabajo, se anexan los avances y resultados en donde se comprueba la homología del dominio 3' –5' exonucleasa para ambos grupos de DNA polimerasas a partir de una búsqueda de homólogos. Del mismo modo se muestran los avances del análisis evolutivo de una muestra propuesta de DNA polimerasas tipo II representadas por varias clases de enzimas reportadas para los genomas de Eucarya y Archaea las cuales se podrían ver relacionadas a eventos de duplicación y divergencia.

Introducción:

La reconstrucción de caracteres ancestrales esta basada en una metodología en donde a partir de rasgos tanto comunes como divergentes detectados en una serie de organismos, se deduce la existencia de posibles eventos de especiación y adaptación que resultan en la diversidad del linaje cuya historia se pretende reconocer. Bajo ésta misma óptica se ha abordado la evolución de las etapas tempranas de la vida, intentando reconocer caracteres y rasgos que relacionen a todos los seres vivos conocidos, para poder detectar características conservadas en todos ellos y que posiblemente ya estaban presentes en el último ancestro común.

Una de las primeras propuestas para estudiar la diversidad biológica y definir simultáneamente características ancestrales fue llevada a cabo por Haeckel en 1866, quien propuso la separación de todos los seres vivos en tres reinos. Uno de estos reinos, el Monera, agrupaba a las bacterias y a otros organismos unicelulares. A partir de aquí se propuso que las bacterias presentes en este reino podían ser similares a las de las primeras formas de vida, vistas entonces como sistemas elementales, unicelulares y autotróficos.

No fue sino hasta 1938 cuando Edouard Chatton diferenció dos grandes tipos de células. Uno de estos grupos, los eucariontes, se caracterizan por presentar arreglos intramembranales que constituyen organelos celulares y un núcleo-citoplasma. El otro que carecía de núcleo se le denominó procarionte. Esta propuesta fue luego retomada por Stanier & Van Niel en 1962, por un lado, y paralelamente por Margulis en 1970.

Pero fue hasta el trabajo de Margulis en 1970 en donde se proponen argumentos estables para poder suponer que los eucariontes son el resultado de procesos de endosimbiosis producidos entre un organismo hospedero, unas bacterias con metabolismo aerobio, unas cianobacterias y bacterias similares a

las espiroquetas, que pudieron haber dado origen a las actuales mitocondrias, cloroplastos y estructuras de microtúbulos de arreglos 9+2, respectivamente.

Paralelamente a estos estudios, se desarrolló la evolución molecular, que a partir del trabajo de Zuckerkandl y Pauling (1965) se mostró que el estudio de algunas de las secuencias de nucleótidos y aminoácidos podrían dar mas información con respecto a la evolución de las secuencias y de la evolución de los seres vivos que las presentan respaldando caracteres, funciones y eventos de evolución que no son detectables por caracteres fenotípicos. Este estudio de secuencias nos permitiría deducir las probables características de organismos ancestrales y asomarnos a su vez a eventos de la evolución temprana de la vida.

Uno de los aportes centrales basados en esta metodología es el trabajo de Woese & Fox (1977), en donde a partir de un banco de secuencias de fragmentos de restricción, primero, y posteriormente por las secuencias completas de las subunidades 16S / 18S rRNA fue posible proponer la separación de todos los organismos en tres dominios: Bacteria, división representada por procariontes generalmente mesofílicos, en donde se presentan una amplia gama de metabolismos y adaptaciones a diferentes ambientes y tipos de vida; el dominio Archaea, que separa a dos grandes grupos de procariontes: Crenarchaeota, conformado por procariontes hipertermofílicos y especies relacionadas filogenéticamente que presentan rasgos comunes con los eucariontes; y los Euryarchaeota, que incluye a los procariontes metanógenos y halofílicos. El tercer dominio es el Eucarya, que incluye a todos los organismos con células nucleadas. A partir de la propuesta de que los tres dominios tienen un origen común, Woese propone como último ancestro común de los tres grandes linajes celulares, la existencia de una entidad biológica hipotética llamada progenote, describiéndolo con un organismo de naturaleza menos compleja a la de un procarionte actual y con un fenotipo y genotipo indiferenciados y en continua evolución (Woese, 1982; Woese, 1998), con una probable localización temporal dentro del mundo del RNA (Alberts, 1986; Gilbert, 1986; Lazcano, 1986).

Sin embargo, Fitch & Upper (1987) consideran que por la ausencia de sistemas celulares cuya molécula bioinformacional sea de RNA así como por la complejidad operacional e informacional presente en cada representante de los tres dominios proponen como último ancestro común de los tres dominios propuestos por Woese a una entidad hipotética llamada cenancestro, que se define como un sistema celular mucho más complejo de lo que se pensaba en los primeros reportes, al identificar probables elementos tales como un sistema de ATPasas, múltiples rutas biosintéticas, con proteínas ribosomales, factores de elongación transcripcional, sistemas relacionados al metabolismo de tRNA, así como elementos metabólicos relacionados a la síntesis del mismo DNA como lo son rutas de salvamento de purinas y enzimas de polimerización de DNA y RNA (Lazcano, *et al.* 1988a; Lazcano *et al.* 1988b; Becerra *et al.*, 1997; Lazcano & Forterre, 1999)

A partir de este enfoque molecular de la evolución, fue posible evidenciar cierta relación entre el dominio Archaea y Eucarya, lo que permite suponer su pasado compartido. Para lograr comprender como se ha analizado esta relación entre los dominios es necesario aclarar algunos conceptos en el estudio de las secuencias. Los caracteres se pueden definir como homólogos cuando se les reconoce un pasado evolutivo común, originándose en un mismo ancestro (Fitch, 2000). A su vez, en las secuencias de nucleótidos es posible deducir tres tipos de homología: (a) los genes ortólogos, los cuales son de un llamado transporte vertical en donde es posible detectarlos relacionados a un proceso de especiación de un linaje; (b) los genes parálogos que resultan de un evento de duplicación de un gen, posterior a un evento de especiación; por último, (c) los genes xenólogos, que son secuencias compartidas entre dos o más organismos y que resultan de eventos de transferencia horizontal de genes.

La evolución de los caracteres y secuencias ortólogas han permitido el establecer la separación de los tres dominios a partir de diferentes filogenias, mientras que algunos genes parálogos han permitido enraizar árboles filogenéticos universales y reconocer la relación entre estos tres grandes grupos. Esta metodología fue aplicada por Iwabe *et al.*, (1989); Gogarten *et al.*, (1989) en donde a partir de las

secuencias que se les adjudica una antigüedad mayor a la del cenancestro es posible reconocer una relación estable entre los tres dominios. Así, al comparar las secuencias ya sea de las subunidades α y β de la ATPasa, o de los factores de elongación y sus homólogos, es posible proponer árboles filogenéticos que se relacionan en dos subarreglos que describan la estructura y relación entre los dominios. Ello ponía la raíz del árbol entre los grupos Archaea y Eucarya separándolos del grupo Bacteria.

Los resultados de ambos trabajos apoyan la información bioquímica, molecular e informacional que apoyaba la relación del dominio Eucarya y Archaea como grupos hermanos y que había sido deducida a partir del análisis de la estructura secundaria de las subunidades de la RNA polimerasa (Zillig, 1991). Sin embargo, cuando se incrementó el conocimiento de diferentes marcadores moleculares fue posible detectar variaciones en estos arreglos: así, hay filogenias basadas en genes operacionales (definidos en: Jain, Rivera & Lake, 1999) en donde se relacionan a los dominios Archaea y Bacteria como grupos hermanos o al dominio Bacteria y Eucarya como grupos que comparten una historia evolutiva común (Doolittle & Brown, 1994; Edgell & Doolittle, 1997).

Debido a la disponibilidad de genomas secuenciados se han podido detectar diversas familias de proteínas y de genes que se relacionan en forma a veces confusa, siendo difícil establecer si la relación de homología es ortóloga o paráloga, o si en algún momento la distribución de esta secuencia se deba a procesos de transferencia horizontal (Gogarten *et al.*, 1996; Doolittle & Longsdon, 1998). Otra hipótesis que podría explicar la presencia de posibles elementos moleculares antiguos que no reflejan la historia evolutiva de todos los organismos asumiría que el último ancestro común (LCA), mas que ser una sola entidad celular única, era un conglomerado de células primitivas y luego se separó en comunidades de domnde divergieron los tres diferentes linajes celulares (Woese, 1998).

Gracias a las evidencias obtenidas por diferentes trabajos de investigación, es posible suponer la conservación y la probable presencia de mecanismos moleculares en el cenancestro tales como la

transcripción, traducción, un sistema de reparación del genoma y rasgos conservados del metabolismo de desoxirribonucleótidos (Woese, 1987; Lazcano, 1992; Lazcano *et al.*, 1999; Penny & Poole, 1999; Doolittle, 2000), dando con esto la posibilidad de que fuese una entidad compleja en algunos sistemas y que pudiese tener ya un genoma de DNA.

Sin embargo, la ausencia aparente de caracteres homólogos dentro de los sistemas de replicación de los tres dominios han hecho que se argumenten otras posibilidades sobre la naturaleza del cenancestro, suponiendo un genoma de RNA (Mushegian & Koonin, 1996) o con una estructura combinada de RNA y DNA (Leipe, Aravind & Koonin, 1999), o bien suponiendo que la estructura genómica pudo haber evolucionado en forma independiente dos veces en la historia evolutiva. El empleo de estos argumentos supone escenarios con un mayor número de eventos evolutivos, que no toman en consideración todas las evidencias presentes, como lo son todos los elementos de los sistemas de informacionales que hablan de un sólo origen común y apoyándose en la poca conservación de los elementos conservados en el sistema de replicación.

Cuando uno comienza a estudiar las características de los sistemas de replicación es posible reconocer genes homólogos que relacionan a los dominios de Archaea y Eucarya, pero pocos elementos comunes a los tres dominios. Entre estos últimos se incluye la conservación del complejo "clamp loader", la ribonucleasa H y la presencia conservada de las DNA polimerasas tipo II (Edgell & Doolittle, 1997). Para poder estudiar con mayor detenimiento la conservación de las polimerasas tipo II es necesario reconocer la similitud en base a la estructura, funciones y similitudes, que hace posible dividir a las DNA polimerasas en cuatro grandes grupos (Braithwaite & Ito, 1993). La clasificación de estas enzimas se ha propuesto y apoyado por medio de comparaciones de sus secuencias, al arreglo de sus dominios (Braughtiman & Steitz, 1999), al reconocimiento que hacen de su molde y a diferentes motivos detectados en estructura primaria (Sousa, 1996), así como en base a sus funciones.

A partir de los primeros estudios llevados a cabo en las DNA polimerasas de bacteriófagos fue posible reconocer varios dominios bien definidos, lo cual permite identificar en la enzima cuatro regiones funcionales. Basándose en las primeras estructuras terciarias reportadas, cuya arquitectura y geometría se ha comparado con la de una mano derecha, es posible reconocer a cuatro dominios: (a) el dominio de los dedos (fingers), encargado de la especificidad del reconocimiento del molde, así como de promover la posición adecuada de los dNTP que se unen a las bandas recién sintetizadas; (b) el dominio pulgar (thumb) que permite el movimiento del molde dentro de la enzima y que posee motivos de reconocimiento específicos para el molde; (c) un dominio denominado 3' - 5' exonucleasa, encargado de detectar posibles errores en la polimerización y capaz de separarlos de la misma molécula; y (d) el dominio palma (palm) de la mano, que contiene el sitio activo que hace posible la formación del enlace fosfodiester entre los nucleótidos y que al parecer es uno de los dominios que se presenta conservado en función y arreglo dentro de todas las polimerasas conocidas.

Las DNA polimerasas se reportan como enzimas presentes en los tres linajes celulares estableciéndose una clasificación de cuatro grupos principales: las DNA polimerasas tipo I que tienen representantes dentro de los sistemas de reparación bacterianos, eucariontes y en la replicación de los genomas mitocondriales o de cloroplastos. El grupo de las DNA polimerasas tipo II que tienen un papel importante en la reparación y replicación de genomas eucariontes y posiblemente en arqueobacterianos y en algunos genomas de bacterias. el tercer grupo lo conforman solamente los sistemas multiméricos que conforman a las DNA polimerasas replicativas denominadas como las DNA polimerasas III y el cuarto y último grupo lo conforman las DNA polimerasas tipo IV que presentan actividad de nucleotidil transferasas, estas solamente se presentan dentro del dominio Eucarya.

Dentro de diferentes reportes ha sido posible detectar una gran diversidad de DNA polimerasas tipo II, cuatro tipos para los genomas eucariontes y hasta tres tipos dentro algunos genomas arqueobacterianos. Las clasificaciones que intentan englobar a todas estas enzimas cuya naturaleza podría

ser polifilética se enfrentan a la problemática de localizar rasgos comunes dentro de dos o mas grupos para así establecer posibles eventos de evolución entre cada uno de estos grupos.

Con el propósito de abordar el problema de la presencia de un genoma de DNA en el cenancestro, nosotros nos hemos concentrado al estudio de las DNA polimerasas a partir de la búsqueda de regiones conservadas dentro de estas enzimas en las cuatro familias (Braithwaite & Ito, 1993), llevando a cabo comparaciones en estructura primaria e intentando reconocer en forma común y estable posibles motivos que pudiesen ser comunes a mas de una familia.

Posteriormente analizamos las estructuras terciarías de las DNA polimerasas incluidas en la lista de Brautigam & Steitz (1998) y en la revisión de Steitz (1999). En este último trabajo se insiste en que la región palm de las polimerasas tipo I y II es una región homóloga a nivel de estructura terciaria, y que su conservación probablemente se ha debido al papel central que juega en la formación del enlace covalente entre los nucleótidos, a pesar de ser una secuencia con una alta tasa de cambio (Patel & Loeb, 2000).

A partir de estos antecedentes se intentó corroborar que la homología del dominio palm puede mantenerse constante frente a una muestra de polimerasas de la familia I y II con una distribución filogenética más extensa. Ello fue demostrado primero por Delaye, Vazquez & Lazcano (2000) a partir de una alineación en estructura secundaria, que apoya la conservación de esta homología en representantes de los tres dominios. Posteriormente, en la continuación de este proyecto hemos llevado a cabo comparaciones en estructura primaria de las secuencias de los diferentes dominios identificados dentro de una polimerasa tipo I y una tipo II. Así en este trabajo hemos estudiado si se presenta mas información por parte de otros dominios de DNA polimerasas analizando por separado a cada uno de los dominios y la distribución de sus homólogos encontrados para determinar si alguno de los dos linajes se presentan con una historia evolutiva mas antigua y suponer así su presencia en el último ancestro común. De la misma forma, se analizará de una forma posterior a los homólogos comunes a los dominios para cada uno de los linajes de DNA

polimerasas y enumerando las posibles consecuencias de la información obtenida dentro de una propuesta filogenética en forma de un árbol.

Tanto por el arreglo del mismo dominio palm como por la distribución que muestra dentro de los diferentes representantes de las familias de RNA y DNA polimerasas en donde se presenta, se busca obtener mas información para respaldar a este elemento como una parte de las primeras polimerasas peptídicas más antiguas así como el buscar dentro de los diferentes dominios de las DNA polimerasas otros rasgos que podrían ser ancestrales, y que podrían formar parte de los primeros sistemas replicativos del mundo del RNA/proteína.

Estructura de la tesis:

Esta tesis esta formada por el trabajo: Delaye, L. Vazquez, H., and Lazcano, A. 2000. The cenancestor and its contemporary biological relics: the case of nucleic acids polymerases. En: Chela-Flores, J. 2000. *Proceedings of the Trieste conference on the early evolution of life* ed. Kluwier Academic Press; en donde se demuestra la conservación del dominio palm dentro de los tres linajes celulares en las DNA polimerasas tipo I y II. Se incluye el trabajo en proceso: "Relics of the ancestral DNA polymerase lineage" que es continuación del proyecto de análisis de DNA polimerasas en donde se intentan evidenciar otros probables rasgos conservados dentro de las DNA polimerasas tipo I y II frente a la diversidad de secuencias y genomas completos detectados hasta mayo del 2001.

THE CENANCESTOR AND ITS CONTEMPORARY BIOLOGICAL RELICS: THE CASE OF NUCLEIC ACID POLYMERASES

Luis Delaye, Héctor Vázquez, and Antonio Lazcano
Facultad de Ciencias, UNAM
Apdo. Postal 70-407
Cd. Universitaria, 04510 México D.F.
MÉXICO
E-mail: alar@hp.fciencias.unam.mx

1. Introduction

The recognition that different macromolecules may be uniquely suited as molecular chronometers in the construction of nearly universal phylogenies has widened the range of phylogenetic studies to previously unsuspected heights. In particular, the use of small subunit ribosomal RNAs (16/18S rRNA) as molecular markers led to the construction of a trifurcated, unrooted tree in which all known organisms can be grouped in one of three major monophyletic cell lineages: the eubacteria, the archaebacteria, and the eukaryotic nucleocytoplasm, now referred to as the domains *Bacteria*, *Archaea*, and *Eucarya*, respectively (Woese et al., 1990). The construction of the rRNA tree showed that no single major branch predates the other two, and all three derive from a common ancestor. It was thus concluded that the latter was a progenote, which was defined as a hypothetical entity in which phenotype and genotype still had an imprecise, rudimentary linkage relationship (Woese and Fox, 1977). According to this view, the differences found among the transcriptional and translational machineries of eubacteria, archaebacteria, and eukaryotes, were the result of evolutionary refinements that took place separately in each of these primary branches of descent after they have diverged from their universal ancestor (Woese, 1987).

From an evolutionary point of view it is reasonable to assume that at some point in time the ancestors of all forms of life must have been less complex than even the simpler extant cells, but our current knowledge of the characteristics shared between the three lines suggests that the conclusion that the last common ancestor was a progenote may have been premature. Pending the issue of horizontal gene transport (Figure 1), a partial description of the last common ancestor (LCA) of eubacteria, archaebacteria, and eukaryotes may be inferred from the distribution of homologous traits among its descendants. Ten years ago, the set of such genes that had been sequenced and compared was still small, but the sketchy picture that had emerged suggested that the most recent common ancestor of all extant organisms, or *cenancestor*, as defined by Fitch and Upper (1987), was a rather sophisticated cell (Lazcano, Fox and Oró, 1992) with at least (a) DNA polymerases endowed with proof-reading activity; (b) ribosome-

mediated translation apparatus with an oligomeric RNA polymerase; (c) membrane-associated ATP production; (d) signalling molecules such as cAMP and insulin-like peptides; (e) RNA processing enzymes; and (f) biosynthetic pathways leading to amino acids, purines, pyrimidines, coenzymes, and other key molecules in metabolism (cf. Lazcano, 1995).

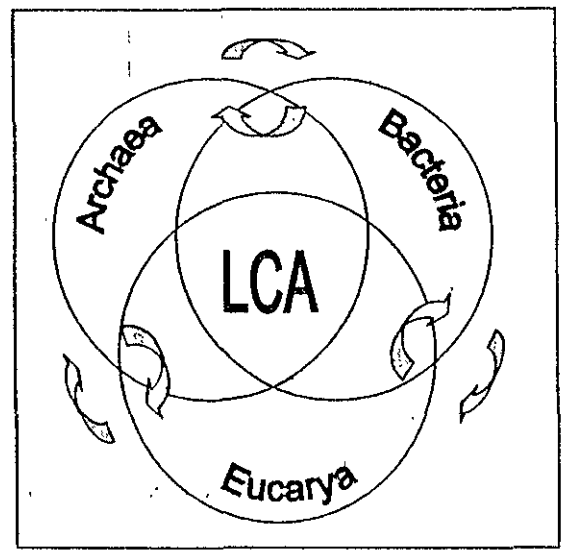


Figure 1. The gene complement of the LCA is defined by the intersection of the complete genomes of the three domains. The arrows represent the horizontal gene transfer between cellular domains.

Recent results have confirmed the above conclusions. These traits are far to numerous and complex to assume that they evolved independently or that they are the result of massive multidirectional horizontal transfer events which took place before the earliest speciation events recorded in each of the three lineages. Their presence suggests that the cenancestral population was not a direct, immediate descendant of the RNA world, a protocell or any other pre-life progenitor system (Lazcano, 1995). Very likely, the LCA was already a complex organism, much akin to extant bacteria, and must be considered the last of a long line of simpler earlier cells for which no modern equivalent is known. Moreover, the universal distribution of the same essential features of genome replication, gene expression, basic anabolic reactions, and membrane-associated ATPase-mediated energy production in all known organisms not only provide direct evidence of the monophyletic origin of all extant forms of life, but also imply that the sets of genes encoding the components of these complex traits were frozen a long time ago, i. e., major changes in them are very strongly selected against.

While trees based on whole genome information have confirmed at a broad level the rRNA-based phylogenies (Snel et al., 1999; Tekaia et al., 1999), it is also true that the congruence between rRNA genes and other molecules is not always ideal. A large variety of phylogenetic trees constructed from DNA and RNA polymerases, elongation factors, F-type ATPase subunits, heat-shock and ribosomal proteins, and an increasingly large set of genes encoding enzymes involved in biosynthetic pathways, have confirmed the existence of the three primary cellular lines of evolutionary descent (Doolittle and Brown, 1994), but there is evidence of extensive horizontal transfer events that have taken place in the past (Doolittle, 1999). In fact, in addition to lateral gene transfer (Figure 1), insights into cenancestral states can be strongly hindered by inadequate biodiversity sampling, polyphyletic gene losses, unequal rates of molecular evolution,

convergence, polyphyly, and secondary loss of organelles. These factors clearly limit our ability to recognize the extant molecular relics of the cenancestor.

1.1 THE SEARCH FOR THE ANCESTRAL NUCLEIC ACID POLYMERASE

Replication of genetic material must have been one of the oldest functions to evolve (Figure 2). Ideally, abiotic laboratory polymerization of nucleotides should provide insights into the transition from the prebiotic broth to the extant enzyme-mediated replication of nucleic acids.

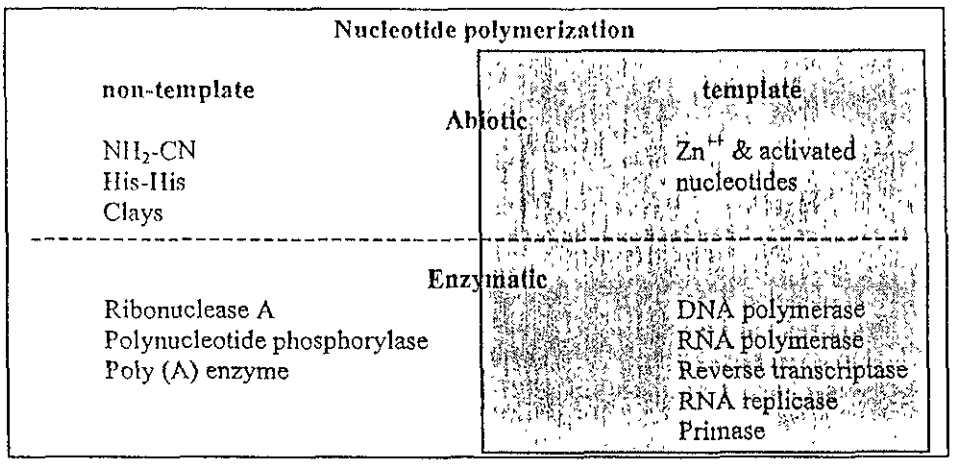


Figure 2. The abiotic and enzymatic polymerization of nucleotides.

In principle, this could also explain the evolutionary development of polymerases, an issue directly related to the chemical composition of the cenancestral genome. Since all extant cells are endowed with DNA genomes, the most parsimonious conclusion is that such genomes were already present in the cenancestral population. However, this hypothesis has been contested by suggestions of an RNA- (Mushegian and Koonin, 1996) or even a mixed DNA-RNA genome for the LCA (Leipe et al., 1999). These proposals are based, at least in part, on the low level of conservation of the primary structure of DNA polymerases (Olsen and Woese, 1996; Edgell and Doolittle, 1997), as well as on the striking differences in their phylogenetic distribution compared with rRNAs, aminoacyl-tRNA-synthetases, and other molecules involved in transcription and translation. This has led to suggestions that DNA genomes, together with the corresponding polymerases, may have been invented independently in the different cell domains (Mushegian and Koonin, 1996; Leipe, *et al.* 1999).

Evolution of enzymes in biological systems often involves the acquisition of new catalytic or binding properties by an existing protein scaffold. However, identification of several non-homologous classes of nucleic acid polymerases (primase, reverse transcriptase (RT), RNA polymerase and DNA polymerase) shows that this is not the

case for these enzymes, and demonstrates the polyphyletic origin of template-dependent enzyme-mediated synthesis of phosphodiester bonds (Steitz, 1999).

Based on sequence similarity and crystal structure analysis (Steitz, 1999) DNA polymerases have been classified into five families (Table 1). Three dimensional structures are available for the DNA polymerase families defined by the DNA pol I, DNA pol α, RT, and rat DNA pol β prototypes.

Family	Representatives
DNA polymerase I family (A polymerase family)	- Klenow fragment of *Escherichia coli* DNA polymerase I - Klenow fragment of *Bacillus* DNA polymerase I - *Thermus aquaticus* DNA polymerase - T7 RNA and DNA polymerases
DNA polymerase α (B family DNA polymerase or family II)	- All eukaryotic replicating DNA polymerases (α,δ,ε) - Phage T4 DNA polymerase - RB69 Phage polymerase
Reverse transcriptase family	- HIV reverse transcriptase - RNA-dependent RNA polymerase - Telomerase
Rat DNA polymerase β	- DNA polymerase β (rat)
Bacterial DNA polymerase III	- Bacterial DNA polymerase III, on the basis of amino acid sequence comparisons.

Table 1. Classification of DNA polymerases into five families according to sequence similarity and tertiary structure criteria (cf. Steitz, 1999).

All DNA polymerases whose tertiary structure has been determinated appear to share a common overall architectural feature comparable to a right hand shape. This structure is not so evident, however, in the case of rat DNA pol β and its homologues. The structure of the other polymerases has been described as consisting of "thumb", "palm", and "finger" domains (Kohlstaedt, et al, 1992). Detailed analysis of the three dimensional structure of DNA polymerases from the pol I, pol α, and RT families suggest that their palm sub-domain has a single origin, i.e., it is homologous in all of them, while the fingers and the thumb sub-domains are different in all four of the families for which structures are known (Brautigam and Steitz, 1998). The complex evolutionary history of nucleic acid polymerases, combined with the wide sequence space explored by these enzymes during biological evolution, strongly hinders the identification of the ancestral polymerase.

As argued here, the three-dimensional homology between the palm domains of DNA polymerase I and DNA polymerases B, which includes all eukaryotic replicating DNA polymerases (Steitz, 1999), can be extended to suggest that such domain, which catalyses the phosphodiester bond, was already present in the cenancestor. As shown here, the structural multiple alignment of the palm sub-domain of DNA polymerases belonging to the pol I and pol α families from the tree cellular domains of life strongly

suggests that this sub-domain is the most ancient protein segment found within these enzymes and could have been present in the LCA.

2. Material and Methods

The crystal structures from the following DNA polymerases sequences were downloaded from Protein Data Bank (www.rcsb.org/pdb/): DNA polymerases A family: 1KLN, *Escherichia coli*; 1TAQ, *Thermus aquaticus*; 1XWL *Bacillus stearothermophilus*; and from DNA polymerases B family: 1TGO, *Thermococcus gorgonarius*; 1D5A, and *Desulfurococcus sp.* Tok;.

The palm sub-domains of all of them, following the classification of CATH database (www.biochem.ucl.ac.uk/bsm/cath_new/index.html), were aligned manually using the program SPDBV (Guex, and Peitsch, 1997) (www.expasy.ch/spdbv/text/refs.htm) to construct a structural multiple alignment.

The sequence of the palm domain from *T. gorgonarius* (1TGO) was used as a query against the SwissProt database in the NCBI server (www.ncbi.nlm.nih.gov/BLAST/), using Blast (Altschul, et al, 1997). Sequences from eukaryoticDNA polymerases thus identified using this method, were added to the structural multiple alignment using the program ClustalX v1.81 (Thompson, et al, 1997).

The multiple structural alignment was performed by first aligning the two archaeal and the three bacterial palm sub-domains separately, in order to identify the conserved residues in each of the families. This was followed by the manual alignment of all structures looking for the 3-dimensional conserved residues identified before.

3. Results

The multiple structural alignment of the primary structure of the different palm sub-domains in, is shown in Figure 3.

The Blast search found eight eucaryotic DNA polymerases: DPOD_HUMAN DNA polymerase delta catalytic subunit (Expect = 4e-04); DPOD_BOVIN DNA polymerase delta catalytic subunit (Expect = 5e-04); DPOD_MESAU DNA polymerase delta catalytic subunit (Expect = 7e-04); DPOD_RAT DNA polymerase delta catalytic subunit (Expect = 8e-04); DPOZ_HUMAN DNA polymerase zeta catalytic subunit (Expect = 9e-04); DPOD_MOUSE DNA polymerase delta catalytic subunit (Expect = 0.001); DPOZ_MOUSE DNA polymerase zeta catalytic subunit (Expect = 0.001); DPOD_SOYBN DNA polymerase delta catalytic subunit (Expect = 001).

Figure 3. Multiple structural alignment of the palm subdomains. Enclosed in boxes are the residues which are structurally homologous in all the domains studied here. The sequences are as follows: DNA polymerases A family: 1KLN, *Escherichia coli*; 1XWL, *Bacillus stearothermophilus*, 1TAQ, *Thermus aquaticus*; and from DNA polymerases B family: 1TGO, *Thermococcus gorgonarius*; 1D5A, and *Desulfurococcus sp.* Tok. Above each sequence the secondary structure is shown: h, α-helix, and s, for β-strand.

4. Discusion and conclusions

Although the availability of completely sequenced cellular genomes has enhanced the likehood of more accurate reconstructions of ancestral states, horizontal gene transfer can strongly hinder our ability to understand the characteristics of the last common ancestor. As shown by the current discussions on the chemical nature of the LCA genome (Mushegian and Koonin, 1996; Leipe, *et al.*, 1999), our attempts to understand the distant past can also be limited by the polyphyletic origin of DNA polymerases, whose classification into different families reflects a case of convergence. Nevertheless, the evidence presented here clearly shows that the palm subdomains of the I and II

DNA polymerase families, which are found in all three cell lineages, have a common origin that has conserved the same tertiary structure and is thus an indication of the monophyletic origin of these enzymes. The lack of a crystalized eukaryotic replicative DNA polymerase has not allowed the recognition of the common origin of these polymerases. As shown here however, their monophyletic origin is recognizable even at the primary structure level (Blast search). The evolutionary conservation of this subdomain, which is involved in the catalysis of the phosphoribosyl transfer reaction (Steitz, 1999), is probably due to the central role it plays in the synthesis of polynucleotides.

On the other hand, the lack of homology between the other subdomains (i.e., the thumb and finger) indicates the easyness by which nucleotide-binding motifs can evolve. A possible evolutionary sequence of nucleotide polymerization agents, starting from the prebiotic synthesis of phosphodiester bonds (and omitting the existence of possible preRNA worlds) is shown in Figure 4. This scheme is based on Steitz (1999) suggestion of a stepwise-emergence of functional peptides in an ribozymic replicase, and on the evolution of polymerization agents discussed elsewhere (Lazcano et al., 1988).

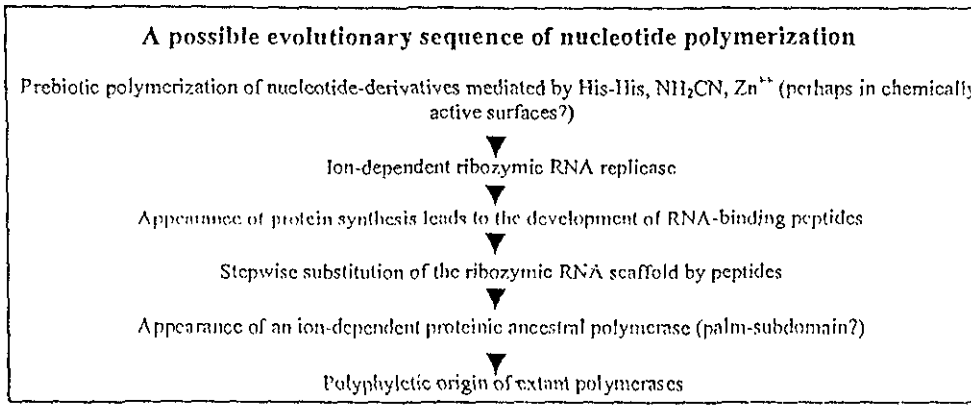


Figure 4. Possible evolutionary sequence of nucleotide polymerization agents, starting from the prebiotic synthesis of phosphodiester bonds (and omitting the existence of possible preRNA worlds).

Given the lack of absolute chemical specificity that polymerases exhibit for both template and substrate (Lazcano et al., 1988), it is quite possible that the conserved ion-dependent palm-subdomain discussed here was part of an ancestral replicase and transcriptase during the RNA/protein world stage (Figure 4). This possibility is supported by the homology between the viral T7 RNA and DNA polymerase. However, the highly conserved sequences of the β and β' subunits of the DNA-dependent RNA polymerase which are found in all three cellular domains, indicate that by the time the LCA had evolved, a modern type of oligomeric RNA polymerase had already evolved. Why polymerases have originated independently several times and why the level of divergence within each family of DNA polymerases is so high, are still open questions that deserve further attention.

Acknowledgements

Support from DGAPA-UNAM/PAPIIT IN213598 project is gratefully acknowledged.

5. References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25, 3389-3402

Brautigam, C.A., and Steitz, T.A. (1998) Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes, *Curr. Opin. Struct. Biol.*, 8, 54-63

Doolittle, W.F. (1999) Phylogeny classification and the universal tree, *Science*, 284, 2124-2128

Doolittle, W.F. and Brown, J.R. (1994) Tempo, mode, the progenote and the universal root, *Proc. Natl Acad. Sci. USA* 91, 6721-6728

Edgell, R.D. and Doolittle, W.F. (1997) Archaea and the origin(s) of DNA replication proteins, *Cell*, 89, 995-998

Fitch, W.M., Upper, K. (1987) The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code, *Cold Spring Harb Symp Quant Biol.* 52, 759-767

Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling, *Electrophoresis*, 18, 2714-2723

Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice,, P.A., and Steitz, T.A. (1992) Crystal structure at 3.5 Å of HIV-1 reverse transcriptase complexed with an inhibitor, *Science*, 264, 1781-1790

Lazcano, A., Fastag, J., Gariglio, P., Ramírez, C. and Oró, J. (1988) On the early evolution of RNA polymerase, *J. Mol Evol* 27, 365-376

Lazcano, A., Fox, G.E., and Oró, J. (1992) Life before DNA: the origin and evolution of early Archean cells, in R.P. Mortlock (ed), *The Evolution of Metabolic Function*, CRC Press, Boca Raton, pp 237-295

Lazcano, A. (1995) Cellular evolution during the Early Archaean: what happened between the progenote and the cenancestor? *Microbiologia SEM*, 11, 185-198

Leipe, D.D., Aravind, L., Koonin, E.V. (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res* 27, 3389-3401

Mushegian, A.R., and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes, *Proc. Natl. Acad. Sci. U.S.A.* 93, 10268-10273

Olsen, G.J. and Woese, C.R. (1996) Lessons from an Archaeal genome: what are we learning from *Methanococcus jannaschii*? *Trends Genet.* 12, 377-379

Snel, B., Bork, P., and Huynen, M.A. (1999) Genome phylogeny based on gene content, *Nature Genetics* 21, 108-110

Steitz, T.A. (1999) DNA polymerases: structural diversity and common mechanisms, *J. Biol. Chem* 274, 17395-17398

Tekaia, F., Lazcano, A., and Dujon, B. (1999) The genomic tree as revealed from whole proteome comparisons, *Genome Research* 9, 550-557

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res.* 24, 4876-4882

Woese, C.R. (1987) Bacterial evolution, *Microbiol. Reviews* 51, 221-271

Woese, C.R. and Fox, G.E. (1977) The concept of cellular evolution, *Jour. Mol Evol.* 10, 1-6

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms, proposal for the domains Archaea, Bacteria, and Eucarya, *Proc. Natl. Acad. Sci. USA* 87, 4576-4579

Relics of the ancestral DNA polymerase lineage

Sir,

It is clearly known that DNA polymerases have a main role in the replication and repair of genetic material in all extant organisms. Attempts to classify and understand the evolutive history of those enzymes have been tried by many research groups (Ito and Braithwaite, 1991; Forterre, 1993; Edgell, Klenk and Doolittle, 1997; Edgell Malik and Doolittle, Forterre, 1998 ; Villarreal and DeFilipis, 2000). These groups have accorded in the establishment of four main groups of DNA polymerases: The DNA polymerases type I with repair function in Bacteria, Eucarya and replication in some organelles. The DNA polymerases type II family with enzymes involved mainly in repair and replication. This role is played in the Archeobacterial and Eucaryal genomes, and is only detected in some cases for the Bacterial species. The DNA polymerase type III family are the replicative enzymes for the bacterial genomes and are specific for this domain. The forth group are the DNA polymerase type IV exclusive for the Eucaryotic genomes and are related into a nucleotydil transferase activity. In recent research works and according to similarity in primary and tertiary structure Steitz (1999) had classified polymerases in five main groups (Table 1)

The evolutionary history of these different groups seems to be a complex topics for research because its phylogenetic distribution and their probable polyphyletic origins.

All DNA polymerases whose tertiary structure appears to share a common overall architectural feature comparable to a right hand shape, consisting of "thumb", "palm", and "finger" domains (Kohlstaedt, *et al*, 1992). This structure is not so evident, however, in the case of rat DNA pol β and its homologues.

Family	Representatives
DNA polymerase type I or A	- Klenow fragment of *Escherichia coli* DNA polymerase I - Klenow fragment of *Bacillus* DNA polymerase I - *Thermus aquaticus* DNA polymerase - T7 RNA and DNA polymerases
DNA polymerases type II o B	- All eukaryotic replicating DNA polymerases (α,δ,ε) - Phage T4 DNA polymerase - RB69 Phage polymerase
Reverse transcriptases family (RT)	- HIV reverse transcriptase - RNA-dependent RNA polymerase - Telomerases
Rat DNA polymerase β	- Type β DNA polymerases
Bacterial DNA polymerases III	- Bacterial DNA polymerase III, on the basis of amino acid sequence comparisons

Table 1. Classification of DNA polymerases into five families according to sequence similarity and tertiary structure criteria (cf Steitz, 1999)

With the arrival of tertiary structures of DNA polymerases from I, II, RT, and β families it was possible to identify a common domain (i.e. the palm domain) between *Thermus aquaticus* DNA polymerase (family I), RB69 Phage polymerase (family II), and HIV reverse transcriptase (family RT) (Brautigam and Steitz, 1998; Steitz, 1999). This domain catalyze the formation of the fosfodiester bond using two metal ions of Mg joined in two aspartic residues that make possible the relation between the dNTP and the oxygen present in the last nucleotide bonded in the template (Steitz, 1998).

In a previews work, we corroborated the proposition made by Steitz (1999) that part of the palm domain is homologous between DNA polymerases family I and II (Figure 1), and showed that this region is conserved across the three cellular lineages of life, i.e. Archaea, Bacteria, and Eucarya (Delaye, Vázquez and Lazcano, 2000). The presence of part the palm domain from DNA pol I and II suggest that this could be an element of the ancestral polymerase. In fact, it is likely that this domain was already present in the last common ancestor (LCA). And the lack of homology between the finger and thumb domains from pols I and II it is likely an example of how easy is the evolution of nucleotide binding sites, for instance, in the CATH database, one of the most common topology that different homologous superfamilies of domains adopt is the Rossmand topology which is a nucleotide binding fold (Orengo, et al, 1997). Different domains within a superfamily are thought to be evolutionary related, but it is uncertain if different proteins within the same topology are related homologous or are the result of convergent evolution.

Because, DNA polymerases I and II shares part of the palm domain, but not finger and thumb domains, in this work we are attempting to study the evolution of the individuals domains (i.e. palm, finger, thumb, and 3'-5' exonuclease) from DNA polymerases I and II in order to know if there have been a process of mosaic evolution in these enzymes. We are also trying to provide a model for the early evolution of these enzymes to understand more features of the ancestral polymerase. Understanding the early evolution of DNA polymerization will also help us to improve our picture of the last common ancestor (LCA), (i.e., if it had a genome of RNA or DNA).

As show in table 2, we found that the palm, finger, thumb, and 3'-5' exonuclease domains of DNA pol II (1D5A) are conserved in Eucarya, Archaea, and Proteobacteria. Such pattern of phylogenetic distribution suggests that the proteobacterial polymerase II could be originated by means of an horizontal transfer event. It was not possible to detect the homology between the palm domains form pol I and II at the level of primary structure. Because the domains of DNA pol I and DNA pol II matched only domains

from their respective families, the whole molecules seems to have evolved as a unit (there is no mosaic evolution between both types of DNA polymerases).

As mentioned above, we couldn't find any homology-relationship between the two type of the DNA polymerases 3'-5'exonuclease domains at the level of primary structure. However, as in the case of part of the palm domain, analyzing the 3D structures we found similarities that we interpret as evidence of homology as suggested by Forterre et al. (1993) (Figure 2). Both domains are classified inside the same superfamily of homologous according to the CATH classification. It is intriguing that only one part of the palm domain and the 3'-5 exonuclease domain from pol I and II are homologous, while the thumb and fingers domains are not. Both domains show a similar level of conservation, that is, it is only possible to detect the homology at the level of tertiary structure and not with PSI-BLAST searches. Anyway, it is intriguing to note that the 3'-5' exonuclease domain of bacterial DNA pol III appears in the PSI-BLAST searches with the 3'-5' exonuclease domain of pol I and II (data not shown) but it doesn't appear any mitochondrial exonuclease domain.

The obtained phylogenetic proposal of the evolution of DNA polymerase type II can make us suppose that possibly the divergence between all the DNA polymerase type II could be a separed event in the three studied cellular lineages, to make more clear the study of these enzyme group is necessary to make an extensive sample to identify more properties about with more stability the different group identified in this first approximation. Some references report a divergence and possible relation between the cellular DNA polymerases and viral enzymes (Villareal and Defilipis, 2000).

To establish the robustness in this hypothesis we need to understand and analyze in posterior works the evolution of the viral polymerases and make an profile to establish common regions that could be related by convergence of by homology between a separed profile for all cellular DNA polymerases type II. In the same way is necessary to make a better and complete sample of cellular DNA polymerases

to identify some traits in the evolution of DNA such as duplication and divergence to try to identify a conserved and ancient DNA polymerase type II for each domain first and after to try to identify the common origins and properties of the Eucaryal/Archaeobacterial last common ancestor polymerase system.

It is likely that one of the earliest polymerases that ever existed was a ribozyme cation dependent, and that this ancestral molecule was replaced domain by domain during evolution (Steitz, 1999). If the similarities that we found between the pol I and pol II are due to common ancestry rater than to convergence, it is possible that both domains (palm and 3'-5' exonuclease) are the relics of an ancestral polymerase which evolved prior the last common ancestor, likely in the RNA-protein world and were the first domains to replace the ribozymic polymerase. Of course, this polymerase was already endowed with editing activity, which certainly represented an advantage on such evolutionary stage. This ancestral polymerase eventually gave rise by duplication to the ancestors of DNA polymerase I and II before the LCA, recruiting each different thumb and finger domains. If both types of enzymes were already present in the LCA the extant distribution of polymerases I and II in the tree cellular domains (Table 2) could be explained by differential gene losses.

Future perspectives about this work is to compare new DNA polymerase tertiary structures to corroborate and try to evidence more elements in the evolution between the different domains of the DNA polymerases and try to establish more paths and features for the ancestral DNA polymerase lineage represented in actual sequences of DNA polymerases type I and II.

Maybe some properties and clues about how the ancestral RNA polymerase evolved into an ancestral DNA polymerase could be present in the evolution and sequences of the RNA and DNA polymerases of the type I group, this study would focus in the fingers and thumb domain: regions of the polymerase with nucleic acid binding sites with recognition of deoxyribonucleic or ribonucleic acids.

Another topic to study is the properties and structure of the elements for the DNA polymerase type I. In this case we could try of evidence the possible characters that could have regulated the changes between an RNA and DNA polymerase, using as models to compare the DNA and RNA polymerases represented by some phage and viral DNA polymerases.

References:

Altschul, S.F., Madden, T. L., Schäffer, A. A., Zhang, A. A., Zhang, Z. Miller, W. Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 1997. **25**: p. 3389-3402

Braithwaite D.K. and Ito J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research* **21**:787-802

Brautigam, C.A. and Steitz, T.A. 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology*, **8**: 54-63.

Edgell, D., H. Klenk, and Doolittle, W.F. 1997. Gene duplications in evolution of archaeal family B DNA polymerases. *Journal of Bacteriology*, **179**: 2632-2640.

Edgell, D.R. and Doolittle, W.F. 1997 Archaea and the origin(s) DNA replication proteins. *Cell*,. **89**: 995-998.

Edgell, D.R., Malik, S. and Doolittle, W.F. 1998. Evidence of independent gene duplications during the evolution of archaeal and eukaryotic familyB DNA polymerases. *Molecular Biology and Evolution*, **15**: 1207-1217.

Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. and Labedan, B. 1993. The nature of the last universal ancestor and the root of the tree of life, still open questions. *BioSystems*. **28**: 15-32.

Guermeur, Y. Geourjon, C., Gallinari, P., and Deleage, G. 1999. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* **15**: 413-421

Guex, N. and Peitsch, M.C. 1997 SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714-2723.

Jain, R., Rivera, M.C. and Lake, J.A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proceedings of the National Academy of Sciences USA*, **96**: 3801-3806.

Kohlstaedt L. A., Wang, J., Friedman J. M., Rice P. A., and Steitz T. A. 1992. Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with and inhibitor. *Science* **256**: 1783-1790.

Levin, J.M. 1997. Exploring the limits of nearest neighbour secondary structure. *Protein Engineering*, **10**: 771-776

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thorton, J.M. 1997. CATH-a hierarchic classification of protein domain strucutres, *Structure*, **5**: 1093-1108

Sayle R.A., Milner-White E.J. 1995. RASMOL: biomolecular graphics for all. *Trends in Biochemical Science*, **20**: 374

Steitz, T.A., 1999. DNA polymerases: Structural diversity and common mechanisms. *Journal of Biological Chemistry*, **274**: 17395-17398.

Schuler, G.D., Altschul, S.F., and Lipman, D.J. 1991. A workbench for multiple alignment construction and analysis. *Proteins*, **9**: 180-190

Steitz, T.A., 1998. A mechanism for all polymerases. *Nature*, **391**: 231-232.

Pearson, W.F., 1994. Using the FASTA program to search protein and DNA sequence databases. *Methods in Molecular Biology*, **24**: 307-331.

Pearl, F.M.G., Lee D., Bray J. E., Silitoe I., Todd A. E., Harrison A. P., Thornton J.M. and Orengo C. A. 1999. Assigning genomic sequences to CATH. *Nucleic Acids Research*, **28**: 277-282.

Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., and Higgins D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876-4882

Villarreal L.P., DeFilippis V.R. 2000. A hypothesis for DNA viruses as the origin of eukaryotic replication proteins. *Journal of Virology*, **74**: 7079-7084

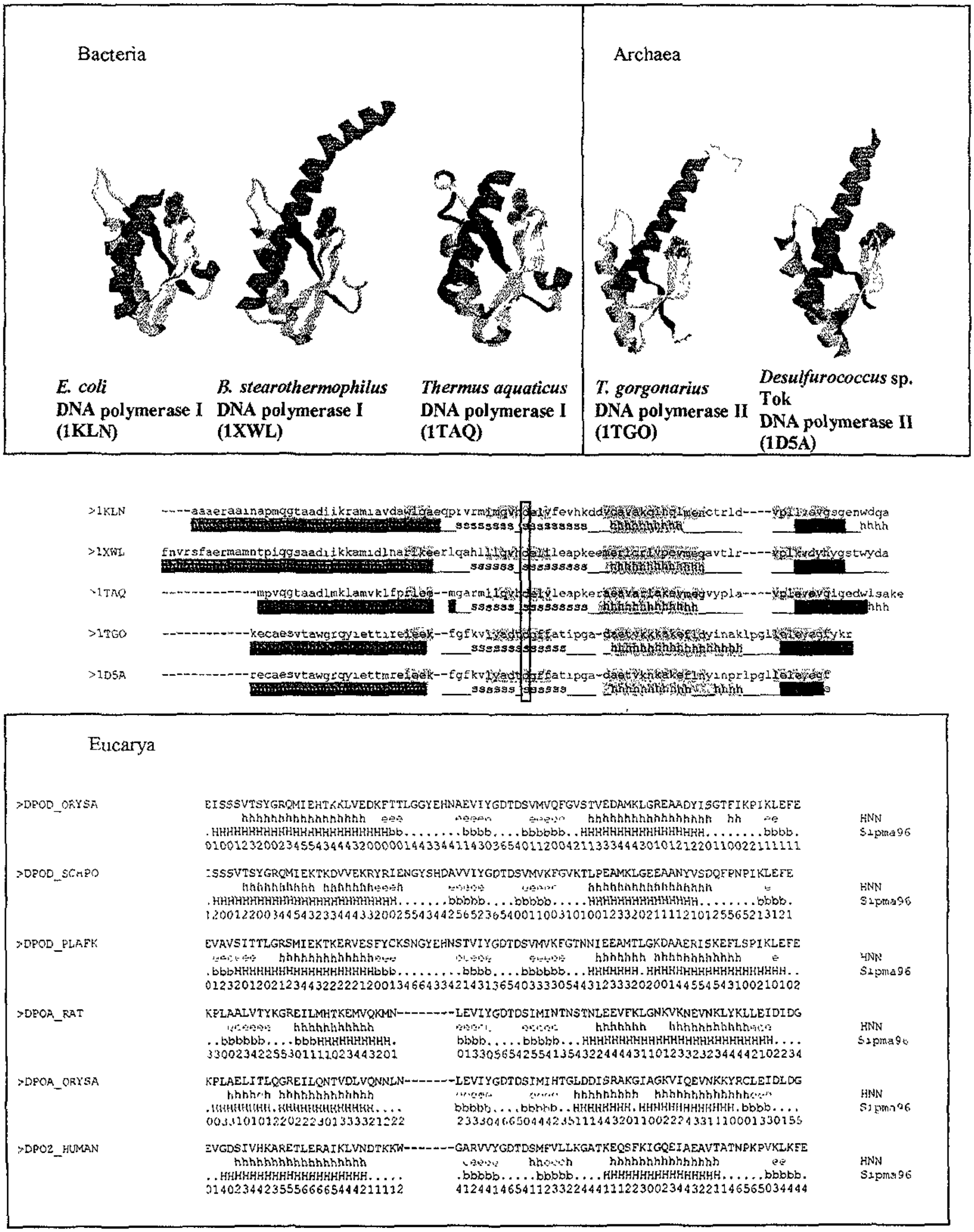
a)

E. coli DNA polymerase Type I VS Swiss Prot database sequences in primary and tertiary structure						
Domains	Bacteria		Archaea		Eucarya	
	Type	Function	Type	Function	Type	Function
"Thumb" *Escherichia coli* 1.10.152.220 CATH	Type I (Pol I)	DNA repair in gaps	----------	----------	Type I DNA polymerase Theta, and DNA pol A	DNA repair θ (interstrand crosslinks)
"Fingers" *Escherichia coli* 1.10.473 10 CATH	Type I (Pol I)	DNA repair in gaps	----------	----------	Type I DNA polymerase Theta, Gamma mitochondrial, and DNA pol A	DNA repair θ (interstrand crosslinks), mitochondrial replication γ
"Palm" *Escherichia coli* 3.30.70 370 CATH	Type I (Pol I)	DNA repair in gaps	----------	----------	Type I DNA polymerase Theta, and Gamma mitochondrial	DNA repair θ (interstrand crosslinks), mitochondrial replication γ
	Type II DNA polymerases	***DNA Replication and maybe DNA repair***	***Type II DNA polymerases (Archaea)***	***DNA Replication and maybe DNA repair***	***Type II DNA polymerases***	***DNA Replication and maybe DNA repair***
"3'-5' exonuclease" *Escherichia coli* 3.40.453.10 CATH	Type I (Pol I)	DNA repair in gaps	----------	----------	YASB_SCHPO PMC2_MOUSE Ribonuclease D	Hypothetical (YasB); Polymyositis/ Scleroderna auto antigen (Pmc2) Cleaves tRNA precursor and double strand DNA (RNAse - D)
	Type II DNA polymerases	***DNA Replication and maybe DNA repair***	***Type II DNA polymerases (Archaea)***	***DNA Replication and maybe DNA repair***	***Type II DNA polymerases***	***DNA Replication and maybe DNA repair***

b)

Desulfurococcus sp. Tok DNA polymerase type II sequence comparison VS Swiss Prot database sequences in primary and tertiary structure.						
Domains	Bacteria		Archaea		Eucarya	
	Type	Function	Type	Function	Type	Function
"Thumb" *Desulfurococcus* unassigned CATH	Type II DNA Polymerase	DNA repair and damaged-primer replication	Type II DNA polymerases	DNA Replication and maybe DNA repair	Type II DNA polymerases Alpha, Delta , Zeta and Epsilon	Replicative polymerase α, elongation polymerase δ, translesion synthesis ζ, participates in chromosomal DNA replication and repair ε
"Fingers" *Desulfurococcus* unassigned CATH	Type II DNA polymerases	DNA repair and damaged-primer replication	Type II DNA polymerase	DNA Replication and maybe DNA repair	Type II DNA polymerases Alpha, Delta , Zeta and Epsilon	Replicative polymerase α, elongation polymerase δ, translesion synthesis ζ, participates in chromosomal DNA replication and repair ε
"Palm" *Desulfurococcus* 3.30 70.510 CATH	Type II DNA polymerase	DNA repair and damaged-primer replication	Type II DNA polymerases	DNA Replication and maybe DNA repair	Type II DNA polymerases Alpha, Delta and Zeta	Replicative polymerase α, elongation polymerase δ, translesion synthesis ζ
	Type I DNA polymerase	***DNA repair in gaps***	***Type II DNA polymerases***	***DNA Replication and maybe DNA repair***	***Type II DNA polymerases***	***DNA Replication and maybe DNA repair***
3'-5' Exonuclease 3.40.453.10 CATH	Type II DNA polymerase	DNA repair and damaged-primer replication	Type II DNA polymerases	DNA Replication and maybe DNA repair	Type II DNA polymerases Alpha , Delta and Zeta	Replicative polymerase α, elongation polymerase δ, translesion synthesis ζ
	Type I DNA polymerase	***DNA repair in gaps***	***Type II DNA polymerase***	***DNA repair in gaps***	***Type II DNA polymerase***	***DNA repair in gaps***

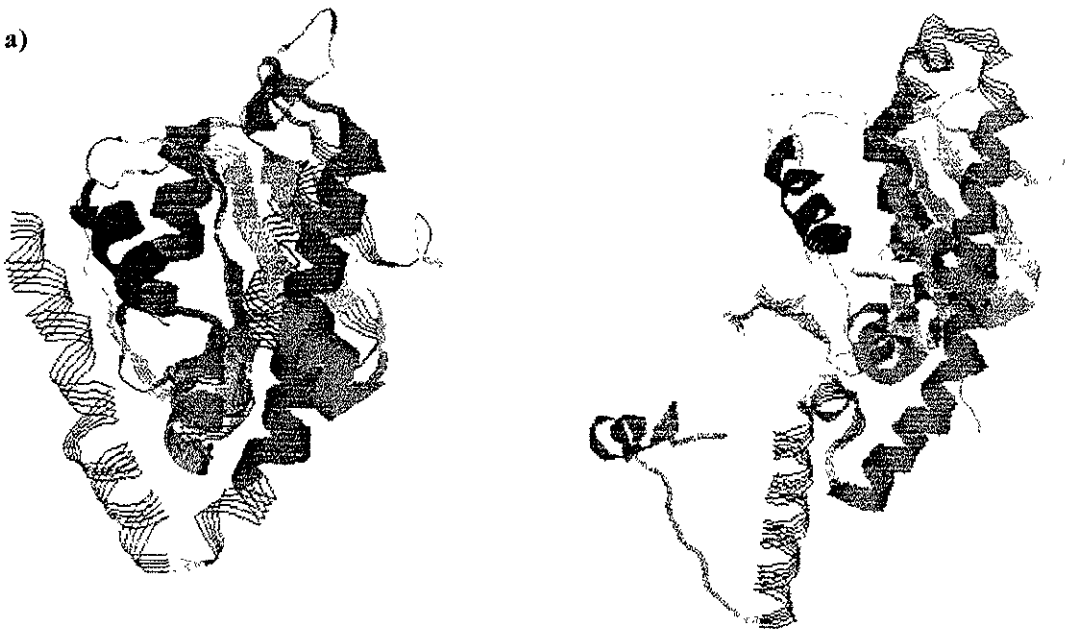
Table 2. Type and function of the different homologous sequences for the peptide domains of a) *Escherichia coli* and b) *Desulfurococcus* sp. Tok DNA polymerases type I and II respectively found with PSI-BLAST search (Altschul, *et al*, 1997) and Fasta33 (Pearson, 1994). Their CATH number classification is also shown in the first column. The thumb, palm, finger and 3'-5' exonuclease domains of DNA polymerases 1KLN and 1D5A were identified following the CATH database (www.biochem.ucl.ac.uk/bsm/cath_new/index.html). Then, we searched for homologous sequences for each of the protein domains in the swiss-prot database using the PSI-BLAST algorithm as implemented in NCBI (www.ncbi.nlm.nih.gov/BLAST/), cutoff value 0.001, matrix BLOSUM 62, default parameters, until convergence. We created eight databases consisting each one in homologous sequences for each of the domains found with PSI-BLAST and Fasta33. The domains were searched using Fasta33 algorithm against the following partial and complete set of genomes downloaded from the KEGG (http://www.genome.ad.jp/kegg/); bacterial complete genomes: *Aquifex aeolicus* VF5, *Bacillus halodurans* C-125, *Bacillus subtilis* 168, *Borrelia burgdorferi* B31, *Buchnera sp. Campylobacter jejuni, Chlamydia muridarum, Chlamydia pneumoniae* AR39, *Chlamydia trachomatis* (serovar D), *Deinococcus radiodurans* R1, *Escherichia coli* K-12, *Haemophilus influenzae* Rd, *Helicobacter pylori* 26695, *Helicobacter pylori* J99, *Mycobacterium leprae, Mycobacterium tuberculosis* H37Rv (lab strain), *Mycoplasma genitalium* G-37, *Mycoplasma pneumoniae* M129, *Neisseria meningitidis* MC58 (serogroup B), *Pseudomonas aeruginosa* PA01, *Rickettsia prowazekii* Madrid E, *Synechocystis sp.* PCC6803, *Thermotoga maritima* MSB8, *Treponema pallidum, Ureaplasma urealyticum, Vibrio cholerae, Xylella fastidiosa* 9a5c; archaeobacterial genomes: *Aeropyrum pernix* K1, *Archaeoglobus fulgidus* DSM4304, *Halobacterium sp.* NRC-1, *Methanobacterium thermoautotrophicum* delta H, *Methanococcus jannaschii, Pyrococcus abyssi, Pyrococcus horikoshii* OT3, *Thermoplasma acidophilum, Thermoplasma volcanium* GSS1; and eucaryal genomes: *Saccharomyces cerevisiae;* and the eucaryal fragmental genomes: *Mus musculus, Arabidopsis thaliana, Caenorhabditis elegans, Drosophila melanogaster.* According to the secondary structure and tertiary structure comparisons and, it was possible to identify homologous domains between the DNA polymerase I and II showed in shadowed lines.



Bacteria
Archaea
E. coli
DNA polymerase I
(1KLN)
B. stearothermophilus
DNA polymerase I
(1XWL)
Thermus aquaticus
DNA polymerase I
(1TAQ)
T. gorgonarius
DNA polymerase II
(1TGO)
Desulfurococcus sp. Tok
DNA polymerase II
(1D5A)
>1KLN ----aaaeraainapmqgtaadiikramiavdawlqaeqpivrmimqvhdelvfevhkddvdavakqihqlmenctrld----vplivevgsgenwdqa
>1XWL fnvrsfaermamntpiqgsaadiikkamidlnarlkeerlqahlllqvhdelileapkeemerlcrlvpevmeqavtlr----vplkvdyhygstwyda
>1TAQ -------------mpvqgtaadlmklamvklfprleg--mgarmllqvhdelvleapkeraeavarlakevmegvypla----vplevevgigedwlsake
>1TGO ------------kecaesvtawgrqyiettireieek--fgfkvlyadtdgffatipga-daetvkkkakefldyinaklpglleleyegfykr
>1D5A ------------recaesvtawgrqyiettmreieek--fgfkvlyadtdgffatipga-daetvknkakeflnyinprlpglleleyegf
Eucarya
>DPOD_ORYSA EISSSVTSYGRQMIEHTKKLVEDKFTTLGGYEHNAEVIYGDTDSVMVQFGVSTVEDAMKLGREAADYISGTFIKPIKLEFE
.HHHHHHHHHHHHHHHHHHHHHHHHHHbb........bbbb....bbbbbb..HHHHHHHHHHHHHHHHH........bbbb.
0100123200234554344432000001443344114303654011200421133344430101212201100221111111
HNN
Sipma96
>DPOD_SCHPO ISSSVTSYGRQMIEKTKDVVEKRYRIENGYSHDAVVIYGDTDSVMVKFGVKTLPEAMKLGEEAANYVSDQFPNPIKLEFE
.HHHHHHHHHHHHHHHHHHHHHHHHHH.......bbbbb....bbbbb.....HHHHHHHHHHHHHHH........bbbb.
1200122003445432334443320025543442565236540011003101001233202111121012556521312
HNN
Sipma96
>DPOD_PLAFK EVAVSITTLGRSMIEKTKERVESFYCKSNGYEHNSTVIYGDTDSVMVKFGTNNIEEAMTLGKDAAERISKEFLSPIKLEFE
.bbbHHHHHHHHHHHHHHHHHHHHHHbbb.........bbbb....bbbbbb...HHHHHHH.HHHHHHHHHHHHHHHHHHH..
012320120212344322222120013466433421431365403333054431233320200144554543100210102
HNN
Sipma96
>DPOA_RAT KPLAALVTYKGREILMHTKEMVQKMN--------LEVIYGDTDSIMINTNSTNLEEVFKLGNKVKNEVNKLYKLLEIDIDG
..bbbbbb....bbbHHHHHHHHHHH. bbbb.....bbbbb...HHHHHHHHHHHHHHHHHHHHHHHHHHH....
3300234225530111023443201 01330565425541354322444431101233232344442102234
HNN
Sipma96
>DPOA_ORYSA KPLAELITLQGREILQNTVDLVQNNLN-------LEVIYGDTDSIMIHTGLDDISRAKGIAGKVIQEVNKKYRCLEIDLDG
HNN
Sipma96
>DPO2_HUMAN EVGDSIVHKARETLERAIKLVNDTKKW-------GARVVYGDTDSMFVLLKGATKEQSFKIGQEIAEAVTATNPKPVKLKFE
..HHHHHHHHHHHHHHHHHHHHH..... .bbbbb.....bbbbb....HHHHHHHHHHHHHHHH......bbbbb.
01402344235556666544421111 2 4124414654112332244411122300234432211465650344444
HNN
Sipma96

Figure 1. Multiple alignment of part of the palm domain from DNA polymerases I and II families. The colors in the primary structure alignment correspond to the colors in the 3D structures (analized with RasMol program (Sayle & Milner-White, 1995), and the catalytic Aspartic is enclosed in a box and in dots in the primary and tertiary structures respectively. In shaded are the regions which align properly between the three molecules. Also shown are the second region of the palm domains from several eucaryotes whit an estimation of its secondary using the Hierarchical Neural Network method (Guermeur, *etal.* 1999) available in (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_nn.html), (Alpha helix (h); Extended strand (Ee); Random coil ( )) and the program SIMPA96 (Levin, 1997) (Alpha helix (H); Extended strand (b); values are from 0 to 9, being 9 very strong, 5 strong and 0 weak). DNA pol Delta: DPOD_ORYSA, *Oryza sativa* (Rice); DPOD_SCHPO, *Schizosaccharomyces pombe* (Fission yeast); DPOD_PLAFK *Plasmodium falci*parum (isolate K1 / Thailand); DNA pol Alpha: DPOA_RAT, *Rattus norvegicus* (Rat); DPOA_ORYSA *Oryza sativa* (Rice); and DNA pol Zeta: DPOZ_HUMAN, *Homo sapiens* (Human). The crystal structures from the following DNA polymerases sequences were downloaded from the Protein Data Bank (www.rcsb.org/pdb/) and visualizing them with the RasMol tertiary structure protein viewer (Sayle R.A. & Milner-White, 1995): DNA polymerase A family: 1KLN, *Escherichia coli*; 1TAQ, *Thermus aquaticus*; 1XWL, *Bacillus stearothermophilus*; and from DNA polymerase B family: 1TGO, *Thermococcus gorgonarius*; and 1D5A, from *Desulfurococcus* sp. Tok. The thumb, palm, finger and 3'-5' exonuclease subdomains of DNA polymerases 1KLN and the palm, thumb and finger subdomains of DNA polymerase 1D5A were identified following the CATH database (www.biochem.ucl.ac.uk/bsm/cath_new/index.html). The second region of the palm subdomain of all five crystal structures were aligned manually using the program SPDBV (www.expasy.ch/spdb/text/refs.htm) (Guex, and Peitsch, 1997).

a)



E coli *(1kln)* Desulfurococcus *sp. Tok (1d5a)*

b)

```
1d5a  ----------------------------ELR----------T-LAHAGAA  A PIL----------E
1kln  VISYDN   ILDEETLKAWIAK-LEK----------S DN---ISAN-LV----------EP
                                    ::**  **    .    .*. :: :*:*

1d5a  E----------    NIDLPYVESV  EKEMIKRFLKVTQEKDP----------GDNFDFAYLKKRS
1kln  G----------  AHDYLDAP--DQIS-RERALELLKPLLEDE-K----------LKYDRGIL
        * *.      :* *  :.:* .:. :: :  :::::   : :  *      : * :

1d5a  EMLGVKFIL RD SEP            VKG IHFD--LYPVIRRTINL  YTLETVYEP
1kln  ANYG--------------------IELRG-IAFDTMLESYILNSV-A RHDMDSLAER
        *                    :*::* * **  * . * .::    : :::: *

1d5a  VFGQP  KVYAEEIAEAWA GEGL  ERVA---RYSMEDAKATYELGKEFFPMEAQLSRL
1kln  WL-KHKTITF-EEIAG    QLTFNQIALEEAGRYAAEDADVTLQLHLKMWP---DLQKH
       : :    .: ****    .   :    :    **: ***..* :*  :::*   :*.:

1d5a  VG SLWDVSRSSTGNLVEWFLLRKAYERNDVAPNKPDERELARRTESYA
1kln  KGP----------------LN-VFENIEM-PLVPVLSRIER
       *                  *. .:*. :: *  *   .: *
```

Figure 2. a) 3'-5' exonuclease domain from *Escherichia coli* DNA pol I (1kln) residues 326-542 and *Desulfurococcus* sp. TOK (1d5a) residues 133-355 visualized with the Ras Mol Program and coloured according to the . b) Structural based hand-made alignment of exonuclease domains sequences improved with ClustalX. The color code from the non conserved regions of the protein has the following pattern: yellow, beta-sheet; Fucsia, alpha-helix; blue, turn;. Boxed areas signalize the secondary structures that spatially and topologically have similar positions in bout structures. Underlined is a region found with Macaw (Schuler, et al, 1991) to be "maybe" statistically significative using Blossum 80 matrix.

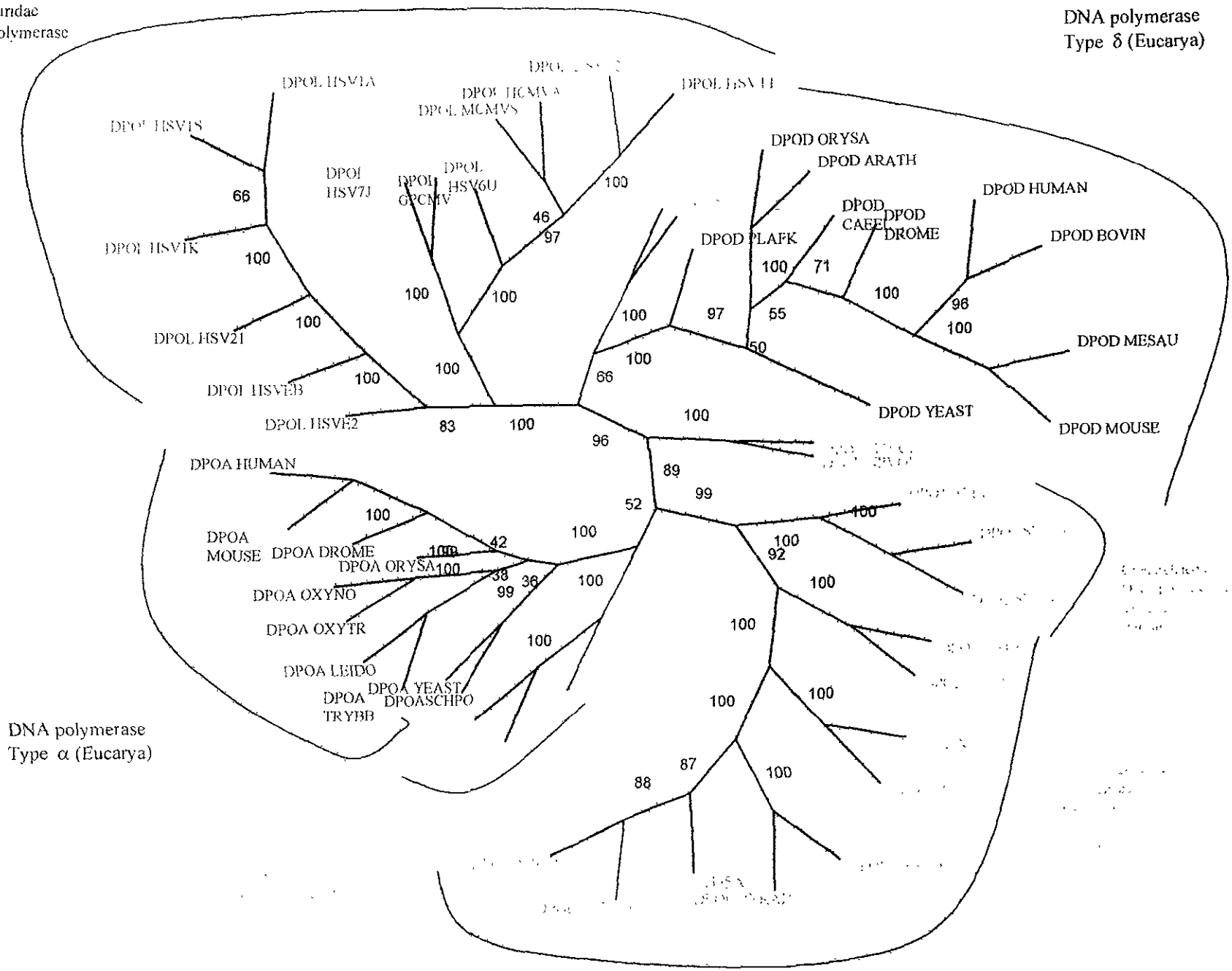

Herpesviridae
DNA polymerase
Type II
DNA polymerase
Type δ (Eucarya)
DNA polymerase
Type α (Eucarya)
DPOL HSV1A
DPOL HSV1S
DPOL HSV1K
DPOL HSV21
DPOL HSVEB
DPOL HSVE2
DPOL HSV7J
DPOL HSV6U
DPOL MCMVS
DPOL HSV1J
DPOD ORYSA
DPOD ARATH
DPOD PLAFK
DPOD CAEEL
DPOD DROME
DPOD HUMAN
DPOD BOVIN
DPOD MESAU
DPOD MOUSE
DPOD YEAST
DPOA HUMAN
DPOA MOUSE
DPOA DROME
DPOA ORYSA
DPOA OXYNO
DPOA OXYTR
DPOA LEIDO
DPOA TRYBB
DPOA YEAST
DPOASCHPO

Figure 3. Phylogenetic proposal for the DNA polymerases type II reported as homologous sequences for all the rhe structural domains of the *Desulfurococcus* sp. Tok polymerase. A multiple alignment was built for the sequences homologous for the DNA polymerase type II using ClustalX (1.81) (Thompson, et al, 1997). The alignment thus obtained was edited by hand to obtain the conserved regions among all the sequences. Then, a phylogenetic reconstruction was performed using the Neighbor-Joining algorithm with the PHYLIP program, using default parameters and 100 bootstrap replications. The results of this proposal is the construction of five groups of cellular DNA polymerases: Eucaryal DNA polymerase type α and DNA polymerase type δ, Bacterial DNA polimerase type II, Crearchaeota DNA polymerase type II and Crenarchaeota and Euryarchaeaota DNA polymerase type II; and three groups of viral DNA polymerases type II: Baculoviridae, Herpesviridae and Phycoviridae DNA polymerase type II. The name and classification of every sequence is based in the Swiss-prot code and the specie of every element of the tree is preesented in Appendix 3.

Conclusiones:

1 - La presencia del dominio palm de las DNA polimerasas I y II sugiere que éste podría ser parte de la polimerasa ancestral de donde se diversificaron la DNA polimerasas tipo I y II. De hecho, como lo indican las comparaciones de secuencia y estructura que reportamos aquí, este dominio palm probablemente ya estaba presente en el último ancestro común. La falta de similitud a nivel primario y terciario entre los dominios finger y thumb es probablemente un indicador de la facilidad con la que pueden surgir en la evolución sitios de unión a nucleóidos y a moldes de ácidos nucléicos.

2.- La conservación de los dominios palm, fingers, thumb y exonucleasa de la DNA pol II en Archaea, Eucarya y proteobacteria, pero su ausencia en otras bacterias, (incluyendo cianobacterias) y en cloroplastos puede ser explicada como el resultado de un evento de transporte horizontal del linaje Archaea /Eucarya hacia las proteobacterias, (antes de su diversificación) luego de la separación de las cianobacterias de los otros miembros del dominio Bacteria. Al parecer el posible eventos de transferencia horizontal de las DNA polimerasas tipo II dentro de las Proteobacterias detectadas podría determinarse al último ancestro común de las bacterias y que posiblemente no se reporten otros homologos a causa de el todavía bajo número de genomas completamente secuenciados .

3.- Que el dominio exonucleasa se encuentre localizado como homólogo solamente en estructura terciaria para la DNA polimerasa I y II es un evento que podría clasificarse, al igual que el palm, como un rasgo ancestral también conservado dentro de este linaje de enzimas extendido hacia los tres grandes dominios celulares. Esta característica es un rasgo que es necesario estudiar posteriormente para analizar su probable antigüedad como los posibles eventos de evolución que lo relacionan con la DNA polimerasa III y otras proteinas de representantes eucariontes con las que se relaciona en los anáisis de estructura primaria.

4.- Una limitante para estos estudios es el no tener dentro de los bancos de datos de estructuras terciarias a las DNA polimerasas eucariontes. Posiblemente cuando se presenten modelos de las DNA polimerasa α δ y ζ sea posible corroborar las homologías propuestas dentro de estos dos trabajos y sea posible el obtener mas información acerca de la evolución de las mismas DNA polimerasas.

5.- La conservación a nivel de estructura primaria de los diferentes dominios de las polimerasas I y II permite usarlas como marcadores filogenéticos y su utilización como identificador de dominios podríaser corroborada posteriormente cuando se dispongan de mas estructuras terciarias y estudios referentes a los diferentes dominios.

6.- La localización de estos homólogos distribuidos para las DNA polimerasa tipo II dentro de los tres dominios así como su actividad replicativa casi constante permite reconocer a a esta molécula como una de las mas importantes y probablemente ancestrales dentro de los sistemas replicativos comunes a los tres linajes celulares. Para poder dar mayor solidez a esta propuesta en el dominio Bacteria, es necesario disponer de un mayor número de homólogos tanto para comparar su estructura terciaria hasta para comparar y comprobar sus posibles capacidades replicativas como la reportada para la DNA polimerasa tipo II de *E. coli* (Rangarajan, *et al.* 1997).

7.- Gracias al dominio palm y exonucleasa es posible relacionar a las DNA polimerasas tipo I y II en un posible origen común y posterior reclutamiento de dominios. Este linaje parece presentar elementos para ser posiblemente un mecanismo catalítico y con capacidad correctiva como mecanismos más antiguos. Para el caso del palm es posible suponer su presencia en momentos de evolución temprana, ya que solamente se depende de los iones relacionados a los aspárticos para catalizar la reacción Con esto es posible ver que la homología de los palm presentados por las DNA polimerasas II y IV podrían presentar una historia evolutiva separada de su palm.

8.- La separación y diversificación de las DNA polimerasas tipo II analizada dentro de las referencias hace notar la presencia de 4 diferentes tipos de enzimas eucariontes relacionadas a la replicación (Hübcher, Nasheuer & Syväoja, 2000) y cuando menos tres tipos dentro de las Archaea (Edgell, Klenk & Doolittle, 1997; Edgell, Malik & Doolittle, 1998), el análisis realizado dentro de este estudio hizo posible evidenciar que posiblemente los eventos evolutivos que dieron lugar a la diversidad de DNA polimerasas II dentro de los dominios Archaea y Eukarya son eventos evolutivos separados y que se pudieron haber dado dentro del tiempo de su diversificación temprana. Sin embargo una de las características que permite pensar y relacionar a todas estas enzimas en un solo grupo es el presentarse como elementos fundamentales para los sistemas de reparación y replicación.

9.-Posiblemente los eventos de duplicación y divergencia para cuando menos las DNA polimerasas vitales para el sistema de reparación y replicación de los eucariontes (α, β y ε) se pudieron haber dado dentro de su último ancestro común mientras que los eventos de duplicación y diferenciación realizados dentro de los genomas de Archaea que se han estudiado de manera reciente e independiente a la separación de los grupos de Euryarchaeaota y Crenarchaeota (información pendiente a analizar dentro del árbol filogenético). Para entender la presencia de un tipo de DNA polimerasas tipo II de una Crearchaeota relacionada con el grupo de DNA polimerasas de Euryarchaeaota es necesario profundizar y realizar un estudio detallado de las DNA polimerasas de Archaea en donde se estudien todas las secuencias de DNA polimerasas tanto descritas como hipotéticas para asi intentar aclarar las posibles funciones de estas enzimas dentro de los sistemas de reparación y replicación de Archaea.

10.- Posiblemente también los sistemas de DNA polimerasas tipo I se necesiten estudiar mas, esto debido a los diferentes representantes de RNA y DNA polimerasas reportados de bacteriófagos y virus dentro de todo el grupo, posiblemente estas enzimas podrían ser buenos elementos de estudio para identificar posibles rasgos de especificidad a la síntesis de las bandas (ya sea de RNA o DNA) para así

suponer posibles eventos de evolución dentro de un linaje ancestral de una RNA polimerasa y como pudo dar origen a una DNA polimerasa ancestral.

11.-El estudio e importancia de las polimerasas de ácidos nucleicos de virus y bacteriófagos se presentan como un elemento pendiente a analizar dentro de este trabajo, las consecuencias de su estudio al analizar de forma separada cada uno de sus dominios podrían detectar posibles eventos de evolución para las DNA y RNA así como el poder detectar mas rasgos que nos podrían ayudar a entender como fue el paso de las RNA a las DNA polimerasas.

Referencias:

1. Alberts, B.M., 1986. The function of the heredity materials: biological catalyses reflect the cell's evolutionary history. *American Zoologist.* **26**: 781-796.

2. Becerra, A., Islas, S., Leguina, J.I., Silva, E. &Lazcano, A., 1997. Polyphyletic gene losses can bias backtrack characterizations of the cenancestor. *Journal of Molecular Evolution.* **45**: 115-118.

3. Braithwaite, D.K. & Ito, J., 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research.* **21**: 787-802.

4. Brautigam, C.A. & Steitz, T.A., 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology.* **8**: 54-63.

5. Delaye, L., Vazquez, H. & Lazcano, A., The cenancestor and its contemporary biological relics: the case of nucleic acids polymerases, en: *Proceedings of the Trieste conference on the early evolution of life*, Chela-Flores, J., Editor. 2000, Kluwier Academic Press.

6. Doolittle, W.F , 1999. Lateral genomics. *Trends in Biochemical Sciences.* **24**: M5-M8.

7. Doolittle, W.F. & Brown, J.R., 1994. Tempo, mode, the progenote, and the universal root. *Journal of Molecular Evolution.* **91**: 6721-6728.

8. Doolittle, W.F. & Logsdon, J.M.J., 1998. Archaeal genomics: do archaea have a mixed heritage? *Current Biology.* **8**: R209-R211.

9. Edgell, D., Klenk, H. & Doolittle, W.F., 1997. Gene duplications in evolution of archaeal family B DNA polymerases. *Journal of Bacteriology.* **179**: 2632-2640.

10. Edgell, D.R. & Doolittle, W.F., 1997. Archaea and the origin(s) DNA replication proteins. *Cell.* **89**: 995-998.

11. Edgell, D.R., Malik, S. & Doolittle, W.F., 1998. Evidence of independent gene duplications during the evolution of archaeal and eukaryotic familyB DNA polymerases. *Molecular Biology and evolution.* **15**: 1207-1217.

12. Fitch, W.M., & Upper, K. 1987. The phylogeny of tRNA sequences provides evidence for ambiguity reduction in the origin of the genetic code. *Cold Spring Harbor Symposia on Quantitative Biology* **52**: 759-767

13. Fitch, W.M., 2000. Homology, a personal view on some of the problems. *Trends in Genetics.* **16**: 227-231.

14. Forterre, P., Benachenhou-Lahfa, N., Confalonieri, F., Duguet, M., Elie, C. &Labedan, B., 1993. The nature of the last universal ancestor and the root of the tree of life, still open questions. *BioSystems*. **28**: 15-32.

15. Gilbert, W., 1986. The RNA world. *Nature*. **319**: 618.

16. Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. &Yoshida, M., 1989. Evolution of the vascular H^+-ATPase: Implications for the origin of eukaryotes. *Proceedings of the National Academy of Sciences USA*. **86**: 6661-6665.

17. Gogarten, J.P., Olendzenski, L., Hilario, E., Simon, C. &Holsinger, K.E., 1996. Dating the cenancestor of organisms. *Science*. **274**: 1750-1753.

18. Iwabe, N., 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proceedings of the National Academy of Sciences USA*. **86**: 9355-9359.

19. Kornberg, A. & Baker, T., *DNA Replication*. 1992: W.H. Freeman and Co. NY

20. Lazcano, A., 1986. Prebiotic evolution and the origin of the cells. *Treballs de la Societat Catalana de Biologia*. **39**: 73-103.

21. Lazcano, A., Fastag, J., Gariglio, P., Ramirez, C. & Oro, J., 1988. On the early evolution of RNA polymerase. *Journal of Molecular Evolution*. **27**: 365-376.

22. Lazcano, A., Guerrero, R., Margulis, L. & Oro, J., 1988. The evolutionary transition from RNA to DNA in early cells. *Journal of Molecular Evolution*. **27**: 283-290.

23. Lazcano, A., Valverde, V., Hernández, G., Gariglio, P., Fox, G.E. & Oro, J., 1992. On the early emergence of reverse transcription: theoretical basis and experimental evidence. *Journal of Molecular Evolution*. **35**: (524-536).

24. Lazcano, A. & Forterre, P., 1999. The molecular search for the last common ancestor. *Journal of Molecular Evolution*. **49**: 411-412.

25. Leipe, D.D., Aravind, L. & Koonin, E.V., 1999. Did DNA replication evolve twice independently? *Nucleic Acids Research*. **27**: 3389-3401.

26. Penny, D. & Poole, A., 1999. The nature of the last universal common ancestor. *Current Opinion in Genetic and Development*. **9**: 672-677.

27. Sousa, R., 1996. Structural and mechanistic relationships between nucleic acid polymerases. *Trends in Biochemical Sciences*. **21**.

28. Steitz, T.A., 1998. A mechanism for all polymerases. *Nature*. **391**: 231-232.

29. Steitz, T.A , 1999. DNA polymerases: Structural diversity and common mechanisms. *Journal of Biological Chemistry* 17395-17398.

30. Woese, C.R. & Fox, G.E., 1977. The concept of cellular evolution. *Journal of Molecular Evolution*. **10**: 1-6.

31. Woese, C.R., 1987. Bacterial evolution. *Microbiological Review*. **51**: 221-271.

32. Woese, C.R., Kandler, O. & Wheelis, M.I., 1990. Towards a natural system of organisms: Proposals for the domains Archaea, Bacteria and Eucarya. *Proceedings of the National Academy of Sciences USA*. **87**: 4576-4579.

33. Woese, C.R., 1998. The universal ancestor. *Proceedings of the National Academy of Sciences USA*. **95**: 6854-6859.

34. Zillig, W., 1991. Comparative biochemistry of Archaea and Bacteria. *Current Opinion in Genetic and Development*. **1**: 544-551.

Appendix 1

Fasta results: Type I DNA polymerase Fingers domain against Swiss Prot Database

FASTA (3.36 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, gap-pen: -12/ -2, width: 16
The best scores are

bsu:b	polA, dnaA, DNA polymerase I (POL I), [EC.	(948)	621	146	6e-47
pae:PA5495	polA; DNA polymerase I [EC:2.7.7.7]	(913)	636	148	3.7e-35
HI0856	polA; DNA polymerase I (POL I). [EC:2.7.7	(930)	624	145	2.3e-34
vch:VC0108	DNA polymerase I [EC 2.7.7.7]	(934)	605	141	4.1e-33
nme:NMB1982	DNA polymerase I [EC:2.7.7.7]	(938)	537	126	1.2e-26
xfa:XF1103	DNA polymerase I [EC:2.7.7.7]	(923)	518	122	2.1e-27
ML1381, polA, DNA polymerase I 1648220:1650955 re		(911)	435	104	5.7e-22
Rv1629	polA, polymerase I. [EC:2.7.7.7]	(904)	425	102	9.6e-22
bha:BH3154	polA; DNA polymerase I [EC:2.7.7.7]	(876)	423	102	3.4e-21
dra:DR1707	DNA-directed DNA polymerase [EC:2.7.7	(956)	415	100	1.2e-20
polA	DNA polymerase I. [EC:2.7.7.7]	(880)	408	98	3.3e-20
CT493	DNA Polymerase I. [EC:2.7.7.7]	(866)	406	98	4.4e-20
cmu:TC0780	DNA polymerase I [EC:2.7.7.7]	(866)	394	95	2.7e-19
tma:TM1619	DNA-directed DNA polymerase I [EC.2.7	(893)	387	94	7.9e-19
TP0105	DNA polymerase I (polA). [EC.2.7.7.7]	(997)	385	93	1.2e-18
RP776	polA; DNA polymerase I. [EC:2.7.7.7]	(867)	382	93	1.6e-18
CP0135	DNA polymerase I (polA)	(870)	370	90	1e-17
BB0548	polA; DNA polymerase I. [EC 2.7.7.7]	(908)	339	83	1.1e-15
cje:Cj0338c	polA; DNA polymerase I [EC:2.7.7.7]	(879)	319	79	2.3e-14
HP1470	polA; DNA polymerase I (POL I). [EC:2.7.7	(892)	316	78	3.6e-14
hpj:jhp1383	DNA polymerase i [EC.2.7.7.7]	(897)	315	78	4.2e-14
slr0707	polA; DNA polymerase I. [EC:2.7.7.7]	(986)	299	75	5.1e-13
aq_1967	polA; DNA polymerase I (PolI). [EC:2.7.7	(574)	293	73	7.9e-13
ath:AT4g02700	putative protein	(1548)	200	53	2.4e-06
cel:W03A3.2	DNA polymerase (CE14486) [EC:2.7.7.7	(1208)	146	41	0.0067
pae PA3678	PA3678, probable transcriptional regu	(212)	99	31	1.8
PAB1646 (thiD)	DE:hydroxymethylpyrimidine phospha	(446)	92	29	9.9

Sequences with E-value BETTER than threshold

Sequences producing significant alignments:	Score (bits)	E Value
gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)	208	1e-54
gi|2506365|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (...	206	6e-54
gi|416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I) >gi|4196. .	205	9e-54
gi|1706502|sp|P52028|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (...	204	1e-53
gi|2320101|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (T...	201	1e-52
gi|3913510|sp|O52225|DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (...	201	2e-52
gi|118828|sp|P19821|DPO1_THERQ DNA POLYMERASE I, THERMOSTABLE (T...	201	2e-52
gi|118827|sp|P13252|DPO1_STRPN DNA POLYMERASE I (POL 1) >gi|9802...	198	1e-51
gi|6015001|sp|O34996|DPO1_BACSU DNA POLYMERASE I (POL I) >gi|743...	196	5e-51
gi|1169402|sp|P43741|DPO1_HAEIN DNA POLYMERASE I (POL I) >gi|107...	196	5e-51
gi|118825|sp|P00582|DPO1_ECOLI DNA POLYMERASE I (POL I) >gi|6705...	192	6e-50
gi|6014999|sp|Q59156|DPO1_ANATH DNA POLYMERASE I (POL I) >gi|140...	191	1e-49
gi|6015003|sp|O32801|DPO1_LACLC DNA POLYMERASE I (POL I) >gi|228...	191	2e-49
gi|7404361|sp|P52027|DPO1_DEIRA DNA POLYMERASE I (POL I) >gi|747...	189	4e-49
gi|12229815|sp|Q9S1G2|DPO1_RHILE DNA POLYMERASE I (POL I) >gi|55...	187	3e-48
gi|1169403|sp|P46835|DPO1_MYCLE DNA POLYMERASE I (POL I) >gi|107...	186	5e-48
gi|585062|sp|Q07700|DPO1_MYCTU DNA POLYMERASE I (POL I) >gi|7434...	185	8e-48
gi|6106143|sp|P74933|DPO1_TREPA DNA POLYMERASE I (POL I) >gi|743...	184	1e-47
gi|5729984|ref|NP_006587.1| polymerase (DNA directed), theta; po...	179	5e-46
gi|6015002|sp|O08307|DPO1_CHLAU DNA POLYMERASE I (POL I) >gi|191...	177	3e-45
gi|6225284|sp|O51498|DPO1_BORBU DNA POLYMERASE I (POL I) >gi|743...	173	3e-44
gi|6015004|sp|O05949|DPO1_RICPR DNA POLYMERASE I (POL I) >gi|743...	167	2e-42
gi|9630428|ref|NP_046860.1| DNA polymerase; gp44 [Mycobacteriophi...	156	5e-38
gi|6225285|sp|Q55971|DPO1_SYNY3 DNA POLYMERASE I (POL I) >gi|743...	152	6e-38
gi|9789856|sp|P56105|DPO1_HELPY DNA POLYMERASE I (POL I)	144	2e-35
gi|9789748|sp|Q9ZJE9|DPO1_HELPJ DNA POLYMERASE I (POL I) >gi|743...	144	2e-35
gi|461960|sp|P30314|DPOL_BPSP1 DNA POLYMERASE	142	1e-34
gi|9625474|ref|NP_039708.1| predicted 66.2Kd protein [Mycobacter...	137	2e-33
gi|118855|sp|P29822|DPOL_BPT5 DNA POLYMERASE >gi|67055|pir||DJBP...	136	5e-33
gi|6015000|sp|O67779|DPO1_AQUAE DNA POLYMERASE I (POL I) >gi|751...	123	5e-29
gi|9633592|ref|NP_051006.1| P45 [Bacteriophage APSE-1] >gi|99106...	105	1e-23
gi|9627454|ref|NP_041982.1| gene 5, DNA polymerase [Bacteriophag ..	92	1e-19
gi|118854|sp|Pz03111|DPOL_BPT3 DNA POLYMERASE >gi|76915|pir||S075...	90	4e-19
gi|118852|sp|P0[illegible]|DPOL_BPSP2 DNA POLYMERASE >gi|67052|pir||DJB	82	9e-17
gi|2494179|sp|Q91684|DPOG_XENLA DNA POLYMERASE GAMMA (MITOCHONDR...	79	1e-16
gi|2494182|sp|Q12704|DPOG_SCHPO DNA POLYMERASE GAMMA (MITOCHONDR...	78	2e-16
gi|8567392|ref|NP_059490.1| mitochondrial polymerase gamma [Mus ...	78	2e-16
gi|4505937|ref|NP_002684.1| polymerase (DNA directed), gamma [Ho...	75	2e-14
gi|2494178|sp|Q92076|DPOG_CHICK DNA POLYMERASE GAMMA (MITOCHONDR...	75	2e-14
gi|2494180|sp|Q27607|DPOG_DROME DNA POLYMERASE GAMMA PRECURSOR (...	74	4e-14
gi|1706509|sp|P15801|DPOG_YEAST DNA POLYMERASE GAMMA (MITOCHONDR...	65	2e-11
gi|2494181|sp|Q01941|DPOG_PICPA DNA POLYMERASE GAMMA (MITOCHONDR...	62	1e-10
gi|14285433|sp|Q9Y7b7|DPOG_NEUCR DNA POLYMERASE GAMMA (MITOCHOND...	61	3e-10

Sequences with E-value WORSE than threshold

gi|6323193|ref|NP_013265.1| Yir460cp [Saccharomyces cerevisiae] ...	30	0.74

```
gi|6319945|ref|NP_010026.1|  Alcohol dehydrogenase, Ycr102cp [Sac...    29  1.0
gi|2500746|sp|Q53Z06|NIFA_RHISN  NIF-SPECIFIC REGULATORY PROTEIN ...    28  2.0
gi|3915752|sp|Q42476|KLP1_SCHPO  KINESIN-LIKE PROTEIN 1 >gi|74921. .    28  2.3
gi|3[illegible]|sp|P87146|IMP2_SCHPO  MITOCHONDRIAL IMPORT INNER MEMBR ..    28  .4
gi|3915220|sp|P94972|UVRA_MYCTU  EXCINUCLEASE ABC SUBUNIT A >gi|7 ..    28  . 5
gi|6319442|ref|NP_009524.1|  Ybl029wp [Saccharomyces cerevisiae] ...    28  2 7
gi|544130|sp|P32965|CYSK_CHICK  CYSTEINE-RICH PROTEIN 1 (CRP1) (C ..    28  3 4
gi|6324893|ref|NP_014962.1|  long chain fatty acyl-CoA synthetase...    28  3.2
gi|2494438|sp|Q49130|MAUM_METEX  METHYLAMINE UTILIZATION PERREDOX...    27  3.6
gi|6323737|ref|NP_013808.1|  Ymr090wp [Saccharomyces cerevisiae] ...    27  4 5
gi|9630045|ref|NP_046263.1|  unknown [Orgyia pseudotsugata nuclea...    27  4.5
gi|13626092|sp|Q9K9H0|ACEA_BACHD  ISOCITRATE LYASE (ISOCITRASE) (...    27  4 7
gi|2495247|sp|Q51472|IHFA_PSEAE  INTEGRATION HOST FACTOR ALPHA-SU  .    26  8 7
gi|4165[illegible]|sp|P32234|128U_DROME  GTP-BINDING PROTEIN 128UP >gi|422  .    26  9.2
gi|11024696|ref|NP_036304.1|  F-box only protein 24; F-box protei...    26  9.8
```

Alignments

```
>gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)
          Length = 876

 Score =  208 bits (530), Expect = 1e-54
 Identities = 65/128 (50%), Positives = 90/128 (69%)

Query: 1   IELRIMAHLSRDKGLLTAFAEGKDIHRATAAEVFGLPLETVTSEQRRSAKAINFGLIYGM 60
           IELR++AH++ D  L+ AF  G DIH  TA ++F +  E VT+  RR AKA+NFG++YG+
Sbjct: 657 IELRVLAHIAEDDNLITEAFRRGLDIHTKTAMDIFHVSEEDVTANMRRQAKAVNFGIVYGI 716

Query: 61  SAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDI 120
           S +GLA+ LNI RKEA ++++ YF  +PGV +YM+     +AK++GYV TL  RR YLPDI
Sbjct: 717 SDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTTLLHRRRYLPDI 776

Query: 121 KSSNGARR 128
           S N  R
Sbjct: 777 TSRNFNVR 784

>gi|2506365|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (TAC POLYMERASE 1)
 gi|1470115|gb|AAB81398.1|  (U62584) thermostable DNA polymerase [Thermus caldophilus]
          Length = 834

 Score =  206 bits (524), Expect = 6e-54
 Identities = 61/128 (47%), Positives = 92/128 (71%)

Query: 1   IELRIMAHLSRDKGLLTAFAEGKDIHRATAAEVFGLPLETVTSEQRRSAKAINFGLIYGM 60
           IELR++AHLS D+ L+  F EGKDIH  TA+ +FG+P E V    RR+AK +NFG++YGM
Sbjct: 616 IELRVLAHLSGDENLIRVFQEGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGM 675

Query: 61  SAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDI 120
           SA  L+++L IP +EA  ++ YF+ +P V  ++E+T  + +++GYVETL GRR Y+PD+
Sbjct: 676 SAHRLSQELAIPYEEAVAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDL 735

Query: 121 KSSNGARR 128
           +  + R
Sbjct: 736 NARVKSVR 743

>gi|416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I)
 gi|419652|pir||JX0256 DNA-directed DNA polymerase (EC 2.7.7.7) - Bacillus caldotenax
 gi|912445|dbj|BAA02361.1|  (D12982) DNA polymerase [Bacillus caldotenax]
          Length = 877

 Score =  205 bits (523), Expect = 9e-54
 Identities = 65/128 (50%), Positives = 88/128 (67%)

Query: 1   IELRIMAHLSRDKGLLTAFAEGKDIHRATAAEVFGLPLETVTSEQRRSAKAINFGLIYGM 60
           IELR++AH++ D  L+ AF    DIH  TA ++F +  + VT   RR AKA+NFG++YG+
Sbjct: 658 IELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDIFQVSEDEVTPNMRRQAKAVNFGIVYGI 717

Query: 61  SAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDI 120
           S +GLA+ LNI RKEA ++++ YFE +PGV  YME     +AK++GYV TL  RR YLPDI
Sbjct: 718 SDYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTTLLHRRRYLPDI 777

Query: 121 KSSNGARR 128
           + S N  R
Sbjct: 778 TSRNFNVR 785

>gi|170[illegible]|sp|P52026|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (TTH POLYMERASE 1)
 gi|[illegible]|dbj|BAA00[illegible]  (D[illegible]) thermostable DNA polymerase I [Thermus thermophilus]
 gi|[illegible]|dbj|BAA04[illegible]  (AB025288) Tth DNA polymerase [Expression vector pLEP-HB]
          Length = 834

 Score =  204 bits (521), Expect = 1e-53
 Identities = 61/128 (47%), Positives = 92/128 (71%)

Query: 1   IELRIMAHLSRDKGLLTAFAEGKDIHRATAAEVFGLPLETVTSEQRRSAKAINFGLIYGM 60
           IELR++AHLS D+ L+  F EGKDIH  TA+ +FG+P E V    RR+AK +NFG++YGM
Sbjct: 616 IELRVLAHLSGDENLIRVFQEGKDIHTQTASWMFGVPPEAVDPLMRRAAKTVNFGVLYGM 675

Query: 61  SAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDI 120
           SA  L+++L IP +EA  ++ YF+ +P V  ++E+T  + +++GYVETL GRR Y+PD+
Sbjct: 676 SAHRLSQELAIPYEEAVAFIERYFQSFPKVRAWIEKTLEEGRKRGYVETLFGRRRYVPDL 735

Query: 121 KSSNGARR 128
           +  + R
Sbjct: 736 NARVKSVR 743
```

Appendix 1
Fasta results: Type II DNA polymerase Fingers domain against Swiss Prot Database

```
FASTA (3.3x June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width:  16
The best scores are:                                      opt bits E(110722)
PAB1128 (polI) DE DNA polymerase I                    ( 771)  366   93   4e-19
PH1947  1235aa long hypothetical DNA-directed DNA     (1235)  263   70 8.8e-12
MTH1208  DNA-dependent DNA polymerase family B (P     ( 586)  231   62   8e-10
ape APE2098  DNA-directed DNA polymerase [EC:2.7.     ( 784)  199   55 1.7e-07
sce:YNL102W  POL1, CDC17; DNA polymerase alpha (D     (1468)  185   52 2.8e-06
AF0497  polB, DNA polymerase B1. [EC:2.7.7.7]         ( 781)  176   49 6.9e-06
tac Ta0907  DNA polymerase (PolB), large chain re     ( 796)  151   44 0.00039
sce:YDL102W  CDC2, POL3, TEX1, DNA polymerase del     (1097)  149   43  0.0007
TVG0859401 DNA polymerase                             ( 805)  144   42  0.0014
cel:F10C2.4  DNA polymerase family b (CE09308) [E     (1081)  140   41  0.0029
MJ0885  putative DNA polymerase. [EC:2.7.7.7] [SP     (1634)  133   40   0.013
mmu:99660  Solal; DNA polymerase alpha 1, 180 kDa     (1465)  130   39   0.019
ath:F12B7.5  putative DNA polymerase zeta catalyt     (1871)  122   37   0.083
ath:T1F15.3  Similar to putative DNA polymerase g     (1894)  122   37   0.083
sce:YPL167C  REV3, PSO1; DNA polymerase zeta cata     (1504)  106   33     0.9
hal:VNG0226G  polB1, DNA polymerase B1 (EC:2.7.7.     ( 901)   98   31     2.1

Sequences with E-value BETTER than threshold


                                                                    Score    E
Sequences producing significant alignments.                        (bits) Value

gi|6015025|sp|P56689|DPOL_THEGO  DNA POLYMERASE (TO POL) >gi|4699...   105  2e-23
gi|3913528|sp|P74918|DPOL_THEFM  DNA POLYMERASE (POL TFU) >gi|165...   104  2e-23
gi|3913540|sp|Q56366|DPOL_THES9  DNA POLYMERASE >gi|7434808|pir||...   102  1e-22
gi|399403|sp|P80061|DPOL_PYRFU  DNA POLYMERASE (PFU POLYMERASE) >...    99  1e-21
gi|9625739|ref|NP_039988.1|  DNA polymerase (8) [human herpesviru...    95  2e-20
gi|14251046|ref|NP_116408.1|  T54 [Tupaia herpesvirus] >gi|929696...    94  5e-20
gi|9296965|sp|Q9YU53|DPOL_HSVT1  DNA POLYMERASE >gi|4165073|gb|AA...    94  5e-20
gi|12644199|sp|P30316|DPOD_SCHPO  DNA POLYMERASE DELTA CATALYTIC ...    93  7e-20
gi|9625657|ref|NP_039908.1|  BALF5 DNA polymerase (early), homolo...    93  8e-20
gi|9628340|ref|NP_042931.1|  U38, DNA polymerase [Human herpesvir...    93  1e-19
gi|6015023|sp|O71121|DPOL_RHCMG  DNA POLYMERASE >gi|2944240|gb|AA...    93  1e-19
gi|13124219|sp|Q9LRE6|DPOD_ORYSA  DNA POLYMERASE DELTA CATALYTIC ...    92  1e-19
gi|13124199|sp|P90829|DPOD_CAEEL  DNA POLYMERASE DELTA CATALYTIC ...    92  1e-19
gi|6324227|ref|NP_014297.1|  DNA polymerase I alpha subunit, p180...    92  2e-19
gi|9628761|ref|NP_043792.1|  DNA polymerase [Human herpesvirus 7]...    91  3e-19
gi|118888|sp|P27172|DPOL_MCMVS  DNA POLYMERASE >gi|67043|pir||DJB...    90  6e-19
gi|13124220|sp|Q9LVN7|DPOD_ARATH  DNA POLYMERASE DELTA CATALYTIC ...    89  1e-18
gi|13124718|sp|P54358|DPOD_DROME  DNA POLYMERASE DELTA CATALYTIC ...    89  1e-18
gi|6015019|sp|Q69025|DPOL_GPCMV  DNA POLYMERASE >gi|459763|gb|AAA...    89  1e-18
gi|3913525|sp|O48901|DPOD_SCYBN  DNA POLYMERASE DELTA CATALYTIC S...    89  1e-18
gi|118838|sp|P28339|DPOD_BOVIN  DNA POLYMERASE DELTA CATALYTIC SU...    89  2e-18
gi|6320101|ref|NP_010181.1|  largest and catalytic subunit of DNA...    88  2e-18
gi|31220291|sp|P97283|DPOD_MESAU  DNA POLYMERASE DELTA CATALYTIC S...    88  3e-18
gi|4505933|ref|NP_002682.1|  polymerase (DNA directed), delta 1, ...    88  3e-18
gi|13124716|sp|P46588|DPOD_CANAL  DNA POLYMERASE DELTA CATALYTIC ...    88  3e-18
gi|3913530|sp|P77932|DPOL_PYRSE  DNA POLYMERASE >gi|1495770|emb|C...    87  4e-18
gi|14521919|ref|NP_127396.1|  DNA polymerase I [Pyrococcus abyssi...    87  4e-18
gi|1706505|sp|P52431|DPOD_MOUSE  DNA POLYMERASE DELTA CATALYTIC S...    87  5e-18
gi|6015022|sp|Q85428|DPOL_RCMVM  DNA POLYMERASE                        86  8e-18
gi|9625965|ref|NP_040211.1|  DNA  polymerase [Saimiriine herpesvi...    86  1e-17
gi|3913522|sp|O27276|DPOL_METTH  DNA POLYMERASE >gi|7482291|pir||...    86  1e-17
gi|6225286|sp|O93746|DPO2_AERPE  DNA POLYMERASE II >gi|7434801|pi...    85  2e-17
gi|118834|sp|P27727|DPOA_TRYBB  DNA POLYMERASE ALPHA CATALYTIC SU...    85  3e-17
gi|6015013|sp|O89042|DPOA_RAT  DNA POLYMERASE ALPHA CATALYTIC SUB...    84  4e-17
gi|1106738|ref|NP_006694.1|  DNA polymerase delta, catalytic sub...    84  4e-17
gi|6679409|ref|NP_032918.1|  DNA polymerase alpha 1, 180 kDa [Mus...    83  5e-17
gi|118831|sp|P26019|DPOA_DROME  DNA POLYMERASE ALPHA CATALYTIC SU ..    83  5e-17
gi|8393995|ref|NP_058633.1|  polymerase (DNA-directed), alpha; po...    83  6e-17
gi|9628011|ref|NP_042605.1|  DNA polymerase replicative subunit [...    83  7e-17
gi|9626767|ref|NP_041039.1|  DNA polymerase [Equine herpesvirus 1...    83  7e-17
gi|6015012|sp|O27752|DPOA_OXYTR  DNA POLYMERASE ALPHA CATALYTIC S...    82  1e-16
gi|6015011|sp|Q94636|DPOA_ORYNO  DNA POLYMERASE ALPHA CATALYTIC S...    82  2e-16
gi|6015010|sp|O48653|DPOA_ORYSA  DNA POLYMERASE ALPHA CATALYTIC S...    82  2e-16
gi|118833|sp|P28040|DPOA_SCHPO  DNA POLYMERASE ALPHA CATALYTIC SU...    81  4e-16
gi|118879|sp|P07917|DPOL_HSV1A  DNA POLYMERASE >gi|67036|pir||DJB...    81  4e-16
gi|9629411|ref|NP_044632.1|  DNA polymerase [human herpesvirus 1]...    81  4e-16
gi|118880|sp|P04292|DPOL_HSV1K  DNA POLYMERASE >gi|67034|pir||DJB...    81  4e-16
gi|118881|sp|P09854|DPOL_HSV1S  DNA POLYMERASE >gi|67035|pir||DJB...    81  4e-16
gi|130161|sp|P30307|DPOL_CHVN2  DNA POLYMERASE >gi|81076|pir||B4...    80  6e-16
gi|6015009|sp|O00874|DPOA_LEIDO  DNA POLYMERASE ALPHA CATALYTIC S...    80  6e-16
gi|118882|sp|P07918|DPOL_HSV2I  DNA POLYMERASE >gi|67039|pir||DJB...    80  7e-16
gi|13431465|sp|P77933|DPOL_PYRKO  DNA POLYMERASE [CONTAINS: ENDON...    79  1e-15
gi|6015024|sp|O70736|DPOL_RSIV  DNA POLYMERASE >gi|3176380|dbj|BA...    79  1e-15
gi|6015021|sp|Q84173|DPOL_ORFN2  DNA POLYMERASE >gi|1236947|gb|AA...    78  2e-15
gi|11498108|ref|NP_069333.1|  DNA polymerase B1 (pol8) [Archaeogl...    78  3e-15
gi|232011|sp|P30315|DPOD_PLAFK  DNA POLYMERASE DELTA CATALYTIC SU...    77  4e-15
gi|3913526|sp|O58610|DPOL_PYRHO  DNA POLYMERASE >gi|7446920|pir||...    77  4e-15
gi|9627571|ref|NP_042094.1|  E9L [Variola virus] >gi|1461961|sp|P3...    77  5e-15
gi|2494186|sp|Q51334|DPOL_PYRSD  DNA POLYMERASE (DEEP VENT DNA PO...    77  5e-15
gi|9790985|ref|NP_063712.1|  E9L; putative [Vaccinia virus] >gi|1...    77  8e-15
gi|118892|sp|P06856|DPOL_VACCV  DNA POLYMERASE >gi|335756|gb|AAB...    76  9e-15
gi|118859|sp|P23402|DPOL_FOWPV  DNA POLYMERASE >gi|67047|pir||DJV...    76  9e-15
gi|9629590|ref|NP_040151.1|  ORF28 (AA1-1196) [Human herpesvirus ...    76  1e-14
gi|1478475|sp|Q58884|DPOL_THEGO  DNA POLYMERASE [CONTAINS: ENDON...    76  1e-14
```

```
gi|9631753|ref|NP_048532.1|  PBVC-1 DNA polymerase [Paramecium bu..    78  1e-14
gi|232020|sp|P30317|DPOL_THELI  DNA POLYMERASE (VENT DNA POLYMERA...   74  4e-14
gi|3913534|sp|O43645|DPOL_THEST  DNA POLYMERASE >gi|2293389|emb|C     74  5e-14
gi|3913527|sp|O60673|DPOZ_HUMAN  DNA POLYMERASE BETA CATALYTIC SU     73  9e-14
gi|[illegible]|sp|P52025|DPOL_METVO  DNA POLYMERASI, >gi|449564|gb|AAA  68  [illegible]
gi|6970874|sp|Q61493|DPOZ_MOUSE  DNA POLYMERASE BETA CATALYTIC SU .   68  6e-13
gi|3913508|sp|O50607|DPO1_SULOH  DNA POLYMERASE I (DNA POLYMERASE .   67  1e-12
gi|6325090|ref|NP_015158.1|  DNA polymerase zeta subunit; Rev3p [...  66  7e-12
gi|126432?4|sp|P26811|DPO1_SULSO  DNA POLYMERASE I >gi|2052853|qb .   66  1e-11
gi|232019|sp|P30318|DPOL_NPVLD  DNA POLYMERASE >gi|484512|pir||JQ...  64  4e-11
gi|3913534|sp|P87154|DPOE_SCHPO  DNA POLYMERASE EPSILON, CATALYTI...  59  2e-09
gi|62252831|sp|O93745|DPO1_AERPE  DNA POLYMERASE I >gi|7434800|pir...  57  6e-09
gi|1169407|sp|P43139|DPOL_ASFL6  DNA POLYMERASE >gi|480553|pir||S...  56  8e-09
gi|3913538|sp|P95690|DPO1_SULAC  DNA POLYMERASE I >gi|2129430|pir...  56  1e-08
gi|1169408|sp|P42489|DPOL_ASFB7  DNA POLYMERASE >gi|457624|gb|AAA...  56  1e-08
gi|6015026|sp|Q90162|DPOL_NPVCF  DNA POLYMERASE >gi|747635|gb|AAC...  53  6e-08
gi|6324067|ref|NP_014137.1|  DNA polymerase II; Pol2p [Saccharomy...  53  6e-08
gi|9630008|ref|NP_046226.1|  DNA polymerase [Orgyia pseudotsugata...  53  1e-07
gi|3915679|sp|Q58295|DPOL_METJA  DNA POLYMERASE [CONTAINS. MJA PO...  53  1e-07
gi|5453926|ref|NP_006222.1|  polymerase (DNA directed), epsilon [...  50  7e-07
gi|9630872|ref|NP_047469.1|  DNA Polymerase=AcMNPV orf65 [Bombyx ...  48  2e-06
gi|9627808|ref|NP_054095.1|  DNA-dependant DNA-polymerase [Autogr...  48  2e-06
```

Sequences with E-value WORSE than threshold

```
gi|11467137|ref|NP_054438.1|  ORF207 [Marchantia polymorpha] >gi|...  37  0.005
gi|118829|sp|P21189|DPO2_ECOLI  DNA POLYMERASE II (POL II) >gi|67...  32  0.17
gi|14424453|sp|Q07635|DPO2_SULSO  DNA POLYMERASE II (DNA POLYMERA...  29  1.2
gi|232015|sp|P30319|DPOL_CBEPV  DNA POLYMERASE >gi|281209|pir||S2...  28  3.8
gi|9632637|ref|NP_049662.1|  DNA polymerase [Bacteriophage T4] >g...  28  3.9
gi|5174479|ref|NP_006022.1|  pericentrin B; pericentrin-B [Homo s...  27  4.4
gi|6225350|sp|O54243|FLRB_RHIME  FLAGELLAR BIOSYNTHETIC PROTEIN F...  27  5.1
gi|6094138|sp|O62951|RR3_PICAB  CHLOROPLAST 30S RIBOSOMAL PROTEIN...  27  7.5
gi|752469?|ref|NP_042447.1|  ribosomal protein S3 [Pinus thunberg...  26  8.6
```

Alignments

```
>gi|[illegible]|sp|P56089|DPOL_THEGO DNA POLYMERASE (TO POL)
 gi|4099806|pdb|1TGO|A Chain A, Thermostable B Type Dna Polymerase From Thermococcus
          Gorgonarius
         Length = 773

 Score =  105 bits (262), Expect = 2e-23
 Identities = 56/61 (91%), Positives = 59/61 (95%)

Query: 1   PGFIPSLLGDLLEERQKVKKKMKATVDPIERKLLDYRQRAIKILANSYYGYYAYANARWY 60
           PGFIPSLLGDLLEERQKVKKKMKAT+DPIE+KLLDYRQRAIKILANS+YGYY YA ARWY
Sbjct: 446 PGFIPSLLGDLLEERQKVKKKMKATIDPIEKKLLDYRQRAIKILANSFYGYYGYAKARWY 505

Query: 61  C 61
           C
Sbjct: 506 C 506

>gi|3913528|sp|P74918|DPOL_THEFM DNA POLYMERASE (POL TFU)
 gi|1655695|emb|CAA93738.1| (Z69882) DNA polymerase and endonuclease [Thermococcus fumicolans]
         Length = 1523

 Score =  104 bits (261), Expect = 2e-23
 Identities = 55/61 (90%), Positives = 58/61 (94%)

Query: 1   PGFIPSLLGDLLEERQKVKKKMKATVDPIERKLLDYRQRAIKILANSYYGYYAYANARWY 60
           PGFIPSLLGDLL+ERQKVKK MKATVDPIE+KLLDYRQRAIKILANS+YGYY YA ARWY
Sbjct: 806 PGFIPSLLGDLLDERQKVKKMMKATVDPIEKKLLDYRQRAIKILANSFYGYYGYAKARWY 865

Query: 61  C 61
           C
Sbjct: 866 C 866

>gi|3913540|sp|Q56366|DPOL_THES9 DNA POLYMERASE
 gi|7434808|pir||S67920 DNA-directed DNA polymerase (EC 2.7.7.7) - Thermococcus sp
 gi|1197452|gb|AAA88769.1| (U47108) DNA polymerase [Thermococcus sp. 9oN-7]
         Length = 775

 Score =  [illegible] bits (250), Expect = 1e-2[illegible]
 Identities = 54/61 (88%), Positives = 59/61 (96%)

Query: 1   PGFIPSLLGDLLEERQKVKKKMKATVDPIERKLLDYRQRAIKILANSYYGYYAYANARWY 60
           PGFIPSLLGDLLEERQK+K+KMKATVDP+E+KLLDYRQRAIKILANS+YGYY YA ARWY
Sbjct: 446 PGFIPSLLGDLLEERQKIKRKMKATVDPLEKKLLDYRQRAIKILANSFYGYYGYAKARWY 505

Query: 61  C 61
           C
Sbjct: 506 C 506

>gi|3994031|sp|P80061|DPOL_PYRFU DNA POLYMERASE (PFU POLYMERASE)
 gi|[illegible]|pir||[illegible] DNA-directed DNA polymerase (EC 2.7.7.7) - Pyrococcus furiosus
 gi|216918|dbj|BAA03762.1| (D12983) DNA polymerase [Pyrococcus furiosus]
 gi|2347063|gb|AAB67984.1| (U88155) DNA-dependent DNA polymerase [Pyrococcus woesei]
         Length = 775

 Score = [illegible] bits (247), Expect = 1e-21
```

Appendix 1

Fasta results: Type I DNA polymerase Thumb domain against Swiss Prot Database

```
FASTA (3.36 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width:  16
The best scores are                                   opt bits E(110708)
b3863  polA, resA; DNA polymerase I (POL I). [EC    ( 928)  734  164   7e-40
vch:VC0108  DNA polymerase I [EC 2.7.7.7]           ( 934)  [illegible]  118  3.7e-26
HI0856  polA; DNA polymerase I (POL I). [EC:2.7.7   ( 930)  443  102  1.4e-21
nme:NMB1982  DNA polymerase I [EC:2.7.7.7]          ( 938)  431  100    1e-20
pae:PA5493  polA; DNA polymerase I [EC 2.7.7.7]     ( 913)  410   95  2.1e-19
xfa XF1103  DNA polymerase I [EC:2.7.7.7]           ( 923)  393   92  2.6e-18
tma.TM1019  DNA-directed DNA polymerase I [EC:2.7   ( 893)  369   87  8.2e-17
slr0707  polA; DNA polymerase I. [EC:2.7.7.7]       ( 986)  319   76  1.3e-13
dra:DR1707  DNA-directed DNA polymerase [EC:2.7.7   ( 956)  310   74  4.7e-13
polA  DNA polymerase I. [EC:2.7.7.7]                ( 880)  305   73    9e-13
bha.BH3153  polA; DNA polymerase I [EC.2.7.7.7]     ( 876)  300   72  1.9e-12
TP0105  DNA polymerase I (polA). [EC:2.7.7.7]       ( 997)  249   62  3.5e-09
BBD548  polA; DNA polymerase I. [EC:2.7.7.7]        ( 908)  246   61    5e-09
RP776  polA; DNA polymerase I. [EC.2.7.7.7]         ( 867)  238   59  1.5e-08
CT493  DNA Polymerase I. [EC:2.7.7.7]               ( 866)  226   57  8.8e-08
CPn0135 DNA polymerase I (polA)                     ( 870)  226   57  8.8e-08
cmu.TC0780  DNA polymerase I [EC 2.7.7.7]           ( 866)  219   55  2.4e-07
cje:Cj0338c  polA; DNA polymerase I [EC:2.7.7.7]    ( 879)  208   53  1.2e-06
HP1470  polA; DNA polymerase I (POL I). [EC:2.7.7   ( 892)  203   52  2.6e-06
hpj jhp1363  DNA polymerase I [EC 2.7.7.7]          ( 897)  197   51  6.2e-06
ML1881, polA, DNA polymerase I 1648220:1650955 re   ( 911)  177   47  0.00012
Rv1629  polA; polymerase I. [EC:2.7.7.7]            ( 904)  166   44  0.00057
cel:W03A3.2  DNA polymerase (CE14486) [EC:2.7.7.7   (1208)  128   36     0.18
aq_1967  polA; DNA polymerase I (PolI). [EC:2.7.7   ( 574)  113   33     0.87
PAB2359 (korA-1) DE.2-oxoglutarate ferredoxin ox    ( 408)   97   30      6.7
```

Sequences with E-value BETTER than threshold

Sequences producing significant alignments:	Score (bits)	E Value
gi|118826|sp|P19821|DPO1_THEAQ DNA POLYMERASE I, THERMOSTABLE (T...	177	2e-45
gi|2506365|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (...	175	1e-44
gi|232010|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (T...	174	2e-44
gi|1706502|sp|P52028|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (...	173	5e-44
gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)	172	5e-44
gi|118825|sp|P00582|DPO1_ECOLI DNA POLYMERASE I (POL I) >gi|6705...	171	2e-43
gi|3913510|sp|O52225|DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (...	164	1e-41
gi|6015001|sp|O34996|DPO1_BACSU DNA POLYMERASE I (POL I) >gi|743...	164	2e-41
gi|6014999|sp|Q59156|DPO1_ANATH DNA POLYMERASE I (POL I) >gi|140...	161	1e-40
gi|416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I) >gi|4196...	161	2e-40
gi|6015002|sp|O06307|DPO1_CHLAU DNA POLYMERASE I (POL I) >gi|191...	160	3e-40
gi|1169402|sp|P43741|DPO1_HAEIN DNA POLYMERASE I (POL I) >gi|107...	160	3e-40
gi|6225285|sp|Q55971|DPO1_SYNY3 DNA POLYMERASE I (POL I) >gi|743...	157	2e-39
gi|7404361|sp|P52027|DPO1_DEIRA DNA POLYMERASE I (POL I) >gi|747...	157	2e-39
gi|12229815|sp|Q9S1G2|DPO1_RHILE DNA POLYMERASE I (POL I) >gi|55...	157	3e-39
gi|6166143|sp|P74933|DPO1_TREPA DNA POLYMERASE I (POL I) >gi|743...	153	3e-38
gi|6225284|sp|O51498|DPO1_BORBU DNA POLYMERASE I (POL I) >gi|743...	149	9e-37
gi|6015003|sp|O32801|DPO1_LACLC DNA POLYMERASE I (POL I) >gi|228...	146	5e-36
gi|6015004|sp|O05949|DPO1_RICPR DNA POLYMERASE I (POL I) >gi|743...	141	1e-34
gi|585062|sp|Q07700|DPO1_MYCTU DNA POLYMERASE I (POL I) >gi|7434...	139	5e-34
gi|118827|sp|P13252|DPO1_STRPN DNA POLYMERASE I (POL I) >gi|9802...	139	5e-34
gi|9789856|sp|P56105|DPO1_HELPY DNA POLYMERASE I (POL I)	139	1e-33
gi|9789748|sp|Q9ZJC9|DPO1_HELPJ DNA POLYMERASE I (POL I) >gi|743...	137	3e-33
gi|1169403|sp|P46835|DPO1_MYCLE DNA POLYMERASE I (POL I) >gi|107...	135	1e-32
gi|9625474|ref|NP_039708.1| predicted 66.2Kd protein [Mycobacter...	131	2e-31
gi|118855|sp|P19822|DPOL_BPT5 DNA POLYMERASE >gi|67055|pir||DJBV...	130	2e-31
gi|9630428|ref|NP_046860.1| DNA polymerase; gp44 [Mycobacterioph...	130	3e-31
gi|5729984|ref|NP_006587.1| polymerase (DNA directed), theta, po...	105	8e-24
gi|461960|sp|P30314|DPOL_BPSP1 DNA POLYMERASE	102	7e-23
gi|6015000|sp|O67779|DPO1_AQUAE DNA POLYMERASE I (POL I) >gi|751...	88	2e-18
gi|118855|sp|P30311|DPOL_BPT3 DNA POLYMERASE >gi|76915|pir||S075...	86	7e-18
gi|9627454|ref|NP_041982.1| gene 5, DNA polymerase [Bacteriophag...	85	2e-17

Sequences with E-value WORSE than threshold

gi|6005709|ref|NP_009134.1| A kinase (PRKA) anchor protein 2 [Ho...	34	0.037
gi|120562|sp|P13419|FTHS_CLOAC FORMATE--TETRAHYDROFOLATE LIGASE ...	31	0.38
gi|2670471|sp|Q01903|SUBT_SYNY3 SULFATE-BINDING PROTEIN PRECURSOR...	30	0.48
gi|585157|sp|Q07064|FTHS_CLOCY FORMATE--TETRAHYDROFOLATE LIGASE ...	30	0.63
gi|1169148|sp|P23622|CY14_NCUCR SULFATE PERMEASE II	30	0.70
gi|2495758|sp|Q60347|Y042_METJA HYPOTHETICAL PROTEIN MJ0042 >gi|...	29	0.90
gi|2829827|sp|P96194|YIBL_AZOVI HYPOTHETICAL TRANSCRIPTIONAL REG...	29	1.00
gi|2501586|sp|Q94480|V136_DICDI VEG136 PROTEIN >gi|1513300|gb|AA...	29	1.2
gi|6978661|ref|NP_036662.1| creatine kinase, muscle form [Rattus...	29	1.4
gi|6671762|ref|NP_031736.1| creatine kinase, muscle [Mus musculu...	29	1.4
gi|[illegible]|sp|P00563|KCRM_RABIT CREATINE KINASE, M CHAIN (M-CK) ...	29	1.9
gi|[illegible]|sp|P05413|KCRM_CANFA CREATINE KINASE, M CHAIN (M-CK) ...	28	2.0
gi|5031651|ref|NP_005662.1| reproduction 8; Reproduction/chromos...	28	2.0
gi|4758964|ref|NP_004753.1| cytohesin 1, isoform 1, homolog of s...	28	2.1
gi|[illegible]|ref|NP_044217.1| IncC1 [Enterobacter aerogenes] >gi|...	28	2.1
gi|[illegible]|sp|Q9XSC6|KCRM_BOVIN CREATINE KINASE, M CHAIN (M-CK)...	28	2.1

```
gi|6478379|sp|Q9YB67|SYI_AERPE  ISOLEUCYL-TRNA SYNTHETASE (ISOLEU..     28   7.2
gi|1184527|sp|P00225|DPOL_BPT4  DNA POLYMERASE  gi|67051|pir||DJBP...   28   7.2
gi|1351041|sp|P00565|KCRM_CHICK  CREATINE KINASE, M CHAIN (M-CK) ...    28   4
gi|125305|sp|P06732|KCRM_HUMAN  CREATINE KINASE, M CHAIN (M-CK) ...     28   4.7
gi|1346615|sp|P49033|MYC_HYLLA  MYC PROTO-ONCOGENE PROTEIN (C-MYC...    28   4.
gi|1174548|sp|P43014|SYT_HAEIN  THREONYL-TRNA SYNTHETASE (THREONI ..    27   3.1
gi|1170168|sp|P46600|HAT1_ARATH  HOMEOBOX-LEUCINE ZIPPER PROTEIN ..     27   3.9
gi|9910840|sp|Q9X1G3|PRIM_THEMA  DNA PRIMASE >gi|7442849|pir||F72       27   4.1
gi|9632671|ref|NP_049644.1|  primase [Bacteriophage T4] >gi|1187...     27   4.8
gi|2492647|sp|Q92412|LYS4_EMENI  HOMOACONITASE PRECURSOR (HOMOACO...    27   6.2
gi|2501733|sp|Q01961|PEXC_PICPA  PEROXISOME ASSEMBLY PROTEIN PAS1..     27   6.2
gi|11134094|sp|Q9JX80|NUOC_NEIMA  NADH DEHYDROGENASE I CHAIN C (N...    27   6.6
gi|11134102|sp|Q9K1C1|NUOC_NEIMB  NADH DEHYDROGENASE I CHAIN C (N...    27   6.6
gi|13431701|sp|P57972|MUTS_PASMU  DNA MISMATCH REPAIR PROTEIN MUT...    27   7.1
gi|6679176|ref|NP_032791.1|  origin recognition complex, subunit ...    26   8.0
gi|8486982|sp|P78917|G6PI_SCHPO  GLUCOSE-6-PHOSPHATE ISOMERASE (G...    26   8.2
gi|6321749|ref|NP_011825.1|  Cytochrome B pre-mRNA processing pro...    26   8.4
gi|13632586|sp|O75747|PK3G_HUMAN  PHOSPHATIDYLINOSITOL 3-KINASE C ..    26   8.6
gi|5803187|ref|NP_006746.1|  transaldolase 1; dihydroxyacetone tr...    26   9.4
gi|6323515|ref|NP_013586.1|  NADH dehydrogenase (ubiquinone); Ndi...    26   9.9


                                  Alignments

>gi|118828|sp|P19821|DPO1_THEAQ DNA POLYMERASE I, THERMOSTABLE (TAQ POLYMERASE 1)
 gi|94709|pir||A33530 DNA-directed DNA polymerase (EC 2.7.7.7) I - Thermus aquaticus
 gi|3891938|pdb|1BGX|T Chain T, Taq Polymerase In Complex With Tp7, An Inhibitory Fab
 gi|7246019|pdb|1CMW|A Chain A, Crystal Structure Of Taq Dna-Polymerase Shows A New
           Orientation For The Structure-Specific Nuclease Domain
 gi|155129|gb|AAA27507.1| (J04639) DNA polymerase [Thermus aquaticus]
          Length = 832

 Score =  177 bits (451), Expect = 2e-45
 Identities = 52/115 (45%), Positives = 73/115 (63%), Gaps = 1/115 (0%)

Query: 2   GVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGIKPLKKT- 60
            GV++D    L    S E+    +A LE +   +AG  FNL+S  QL+  +LF++ G+  + KT
Sbjct: 448 GVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTE 507

Query: 61  PGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKSTYTDKLPLMINPKTGRVHT 115
             G  STS  VLE L   +P+ + IL+YR L KLKSTY D LP  +I+P+TGR+HT
Sbjct: 508 KTGKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLPDLIHPRTGRLHT 562

>gi|2506565|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (TAQ POLYMERASE 1)
 gi|1470115|gb|AAB81398.1| (U62584) thermostable DNA polymerase [Thermus caldophilus]
          Length = 834

 Score =  175 bits (444), Expect = 1e-44
 Identities = 49/115 (42%), Positives = 71/115 (61%), Gaps = 1/115 (0%)

Query: 2   GVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGIKPLKKT- 60
            GV++D    L    S EL    +  LE++   +AG  FNL+S  QL+  +LF++  +  L KT
Sbjct: 450 GVRLDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLNSRDQLERVLFDELRLPALGKTQ 509

Query: 61  PGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKSTYTDKLPLMINPKTGRVHT 115
             G  STS  VLE L   +P+ + IL++R L KLK+TY D LP  +++P TGR+HT
Sbjct: 510 KTGKRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPNTGRLHT 564

>gi|232010|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (TFL POLYMERASE 1)
 gi|281488|pir||S26675 DNA-directed DNA polymerase (EC 2.7.7.7) I - Thermus aquaticus
 gi|48166|emb|CAA46900.1| (X66105) DNA-directed DNA  polymerase [Thermus thermophilus]
          Length = 831

 Score =  174 bits (443), Expect = 2e-44
 Identities = 50/115 (43%), Positives = 73/115 (63%), Gaps = 1/115 (0%)

Query: 2   GVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGIKPLKKT- 60
            GV++D    L    S E+    + +LE++   +AG  FNL+S  QL+  +LF++ G+  + KT
Sbjct: 447 GVRLDVAYLQALSLEVEAFVRQLEEEVFRLAGHPFNLNSRDQLERVLFDELGLPAIGKTE 506

Query: 61  PGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKSTYTDKLPLMINPKTGRVHT 115
             G  STS  VLE L   +P+   IL+YR L KLK+TY D LP  +++PKTGR+HT
Sbjct: 507 KTGKRSTSAAVLEALREAHPIVDRILQYRELTKLKNTYIDPLPALVHPKTGRLHT 561

>gi|1706500|sp|P52028|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (TTH POLYMERASE 1)
 gi|466574|dbj|BAA06033.1| (D28878) thermostable DNA polymerase 1 [Thermus thermophilus]
 gi|6009435|dbj|BAA84909.1| (AB025786) Tth DNA polymerase [Expression vector pLED-HB]
          Length = 834

 Score =  173 bits (439), Expect = 5e-44
 Identities = 49/115 (42%), Positives = 71/115 (61%), Gaps = 1/115 (0%)

Query: 2   GVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTILFEKQGIKPLKKT- 60
            GV+ D    L    S EL    +  LE++   +AG  FNL+S  QL+  +LF++  +  L KT
Sbjct: 450 GVRRDVAYLQALSLELAEEIRRLEEEVFRLAGHPFNLNSRDQLERVLFDELRLPALGKTQ 509

Query: 61  PGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKSTYTDKLPLMINPKTGRVHT 115
             G  STS  VLE L   +P+ + IL++R L KLK+TY D LP  +++P+TGR+HT
Sbjct: 510 KTGKRSTSAAVLEALREAHPIVEKILQHRELTKLKNTYVDPLPSLVHPRTGRLHT 564

>gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)
          Length = 876

 Score =  172 bits (436), Expect = 5e-44
```

Appendix 1
Fasta results: Type II DNA polymerase Thumb domain against Swiss Prot Database

```
FASTA (3.36 June 2000) function [optimized, BL50 matrix (15 -5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width:  16
The best scores are:                                      opt bits E(110731)
PAB1128 (polI) DE:DNA polymerase I                      ( 771)  458  101 4.3e-21
PH1947  1235aa long hypothetical DNA-directed DNA       (1235)  456  101 8.1e-21
MJ0885  putative DNA polymerase. [EC:2.7.7.7] [SP       (1634)  297   69 3.7e-11
MTH208  DNA-dependent DNA polymerase family B (EC       (  94)  [illegible]
ape:APE2096  DNA-directed DNA polymerase [EC:2.7.       ( 744)  243   58 3.7e-08
AF0497  polB; DNA polymerase B1. [EC:2.7.7.7]           ( 781)  217   53 1.3e-06
sce:YDL102W  CDC2, POL3, TEX1; DNA polymerase del       (1097)  125   35     0.6
ape:APE2229  DNA-directed DNA polymerase [EC:2.7.       ( 637)  121   34    0.68
b0060  polB, dinA; DNA polymerase II (pol II). [E       ( 783)  120   34    0.92
mmu:99660  Pola1; DNA polymerase alpha 1, 180 kDa       (1465)  115   33       3
hal:VNG0521G  polB1; DNA polymerase B1 [EC:2.7.7.       ( 901)  111   32     3.6
pae:PA1886  polB, DNA polymerase II [EC:2.7.7.7]        ( 787)  110   32     3.7
vch:VC1212  DNA polymerase II [EC:2.7.7.7]              ( 787)  107   31     5.6
TVG0550709 ribosomal protein small subunit S9           ( 200)   97   29     7.6
AF1938  conserved hypothetical protein.                 ( 673)  103   30     8.6
ath:F12B7.5  putative DNA polymerase zeta catalyt       (1871)  108   32     9.7
ath.T1F15.3  Similar to putative DNA polymerase g       (1894)  108   32     9.8
MJ1630  conserved hypothetical protein.                 (1139)  105   31     9.9
cel:F10C2.4  DNA polymerase family b (CE09308) [E       (1081)  104   31      11

Sequences with E-value BETTER than threshold

                                                                      Score    E
Sequences producing significant alignments:                          (bits) Value

      gi|13124219|sp|Q9LRE6|DPOD_ORYSA  DNA POLYMERASE DELTA CATALYTIC ...   180  8e-46
      gi|13431465|sp|P77933|DPOL_PYRKO  DNA POLYMERASE [CONTAINS. FNDON...   180  8e-46
      gi|13878475|sp|Q9RH84|DPOL_THEGB  DNA POLYMERASE [CONTAINS. FNDON...   179  1e-45
      gi|60150025|sp|P96089|DPOL_THEGO  DNA POLYMERASE (TO POL) >gi|4499...   178  2e-45
      gi|3913510|sp|Q56360|DPOL_THES9  DNA POLYMERASE >gi|7434808|pir||...   178  4e-45
      gi|3913525|sp|O48901|DPOD_SOYBN  DNA POLYMERASE DELTA CATALYTIC S...   178  3e-45
      gi|3913528|sp|P74918|DPOL_THEFM  DNA POLYMERASE (POL TFU) >gi|165...   178  4e-45
      gi|3913530|sp|P77932|DPOL_PYRSE  DNA POLYMERASE >gi|1495770|embIC...   172  1e-43
      gi|14521919|ref|NP_127396.1|  DNA polymerase I [Pyrococcus abyssi...   172  1e-43
      gi|13124718|sp|P54358|DPOD_DROME  DNA POLYMERASE DELTA CATALYTIC ...   172  2e-43
      gi|3913526|sp|O59610|DPOL_PYRHO  DNA POLYMERASE >gi|7446920|pir||...   172  2e-43
      gi|232020|sp|P30317|DPOL_THELI  DNA POLYMERASE (VENT DNA POLYMERA...   171  4e-43
      gi|13124220|sp|Q9LVN7|DPOD_ARATH  DNA POLYMERASE DELTA CATALYTIC ...   170  5e-43
      gi|399403|sp|P80061|DPOL_PYRFU  DNA POLYMERASE (PFU POLYMERASE) >...   170  7e-43
      gi|2494186|sp|Q51334|DPOL_PYRSD  DNA POLYMERASE (DEEP VENT DNA PO...   170  7e-43
      gi|4505933|ref|NP_002082.1|  polymerase (DNA directed), delta 1,  ...   169  1e-42
      gi|1706505|sp|P52431|DPOD_MOUSE  DNA POLYMERASE DELTA CATALYTIC S...   168  2e-42
      gi|12644199|sp|P30316|DPOD_SCHPO  DNA POLYMERASE DELTA CATALYTIC ...   168  3e-42
      gi|118838|sp|P28339|DPOD_BOVIN  DNA POLYMERASE DELTA CATALYTIC SU...   168  3e-42
      gi|3913524|sp|O33845|DPOL_THEST  DNA POLYMERASE >gi|2293389|embIC...   168  3e-42
      gi|13124716|sp|P46588|DPOD_CANAL  DNA POLYMERASE DELTA CATALYTIC ...   168  4e-42
      gi|11067381|ref|NP_067694.1|  DNA polymerase delta, catalytic sub...   167  6e-42
      gi|31220029|sp|P97283|DPOD_MESAU  DNA POLYMERASE DELTA CATALYTIC S...   167  6e-42
      gi|1351199|sp|P90829|DPOD_CAEEL  DNA POLYMERASE DELTA CATALYTIC ...   163  6e-41
      gi|6320101|ref|NP_010181.1|  largest and catalytic subunit of DNA...   158  2e-39
      gi|6225286|sp|O93746|DPO2_AERPE  DNA POLYMERASE II >gi|7434801|pi...   157  9e-39
      gi|1149810A|ref|NP_069333.1|  DNA polymerase B1 (polB) [Archaeogl...   153  7e-38
      gi|3915679|sp|Q58295|DPOL_METJA  DNA POLYMERASE [CONTAINS MJA PO...   149  1e-36
      gi|6015009|sp|O00874|DPOA_LEIDO  DNA POLYMERASE ALPHA CATALYTIC S...   147  5e-36
      gi|1706513|sp|P52025|DPOL_METVO  DNA POLYMERASE >gi|495654|gb|AAA...   147  6e-36
      gi|3913502|sp|O05706|DPO3_SULSH  DNA POLYMERASE III (DNA POLYMERA...   146  1e-35
      gi|232011|sp|P30315|DPOD_PLAFK  DNA POLYMERASE DELTA CATALYTIC SU...   145  2e-35
      gi|118834|sp|P27727|DPOA_TRYBB  DNA POLYMERASE ALPHA CATALYTIC SU...   145  3e-35
      gi|3913515|sp|P95979|DPO3_SULSO  DNA POLYMERASE III (DNA POLYMERA...   140  9e-34
      gi|232016|sp|P30320|DPOL_CHVN2  DNA POLYMERASE >gi|281076|pir||B4...   136  1e-32
      gi|6015022|sp|Q85428|DPOL_RCMVM  DNA POLYMERASE                        136  1e-32
      gi|118838|sp|P28040|DPOA_SCHPO  DNA POLYMERASE ALPHA CATALYTIC SU...   135  2e-32
      gi|9631703|ref|NP_048517.1|  PRV-1 DNA polymerase [Paramecium bu...   132  2e-31
      gi|9629605|ref|NP_040111.1|  DNA polymerase [Epstein-Barr virus]...   131  4e-31
      gi|6015019|sp|Q96L51|DPOL_CMVM  DNA POLYMERASE >gi|1429764|gb|AAA...   130  7e-31
      gi|118808|sp|P27172|DPOL_MCMVS  DNA POLYMERASE >gi|167043|pir||DJB...   128  2e-30
      gi|9625607|ref|NP_039908.1|  BALF5 DNA polymerase (early), homolo...   128  3e-30
      gi|6015023|sp|O71121|DPOL_RHCM6  DNA POLYMERASE >gi|2944240|gb|AA...   126  8e-30
      gi|9629695|sp|Q91U53|DPOL_HSVT1  DNA POLYMERASE >gi|4165073|gb|AA...   125  2e-29
      gi|142510461|ref|NP_116408.1|  T54 [Tupaia herpesvirus] >gi|929696...   125  2e-29
      gi|3913527|sp|Q60673|DPOZ_HUMAN  DNA POLYMERASE ZETA CATALYTIC SU...   123  7e-29
      gi|9628761|ref|NP_043792.1|  DNA polymerase [Human herpesvirus 7]...   123  1e-28
      gi|9628340|ref|NP_042931.1|  U38, DNA polymerase [Human herpesvir...   123  1e-28
      gi|9630008|ref|NP_046226.1|  DNA polymerase [Orgyia pseudotsugata...   121  3e-28
      gi|9628011|ref|NP_042605.1|  DNA polymerase replicative subunit [...   121  5e-28
      gi|6325090|ref|NP_015158.1|  DNA polymerase zeta subunit, Rev3p [...   120  9e-28
      gi|9626767|ref|NP_041039.1|  DNA polymerase [Equine herpesvirus 1...   118  2e-27
      gi|8343965|ref|NP_058634.1|  polymerase (DNA-directed), alpha; po...   118  1e-27
      gi|6015013|sp|O89042|DPOA_RAT  DNA POLYMERASE ALPHA CATALYTIC SUB...   117  7e-27
      gi|6015020|sp|Q90162|DPOL_NPVCF  DNA POLYMERASE >gi|747638|gb|AAC...   117  7e-27
      gi|6677409|ref|NP_032916.1|  DNA polymerase alpha 1, 180 kDa [Mus...   115  2e-26
      gi|6015012|sp|Q07152|DPOA_OXYTR  DNA POLYMERASE ALPHA CATALYTIC S...   114  4e-26
      gi|7919874|sp|Q61493|DPOZ_MOUSE  DNA POLYMERASE ZETA CATALYTIC SU...   114  4e-26
      gi|6015011|sp|Q94636|DPOA_OXYNO  DNA POLYMERASE ALPHA CATALYTIC S...   114  5e-26
      gi|9629903|ref|NP_040151.1|  ORF18 (AA1-1194) [Human herpesvirus ...   109  2e-24
```

```
gi|[illegible]|ref|NP_047464.1|  DNA Polymerase-A MHBV [Bombyx ...   101   1e-[illegible]
gi|[illegible]|sp|P[illegible]|DPOL_[illegible]  DNA POLYMERASE I ...   100   [illegible]
gi|9627808|ref|NP_054095.1|  DNA-dependent DNA-polymerase [Autogr...   201   4e-[illegible]
gi|6324227|ref|NP_014297.1|  DNA polymerase I alpha subunit, p180...   100   2e-21
gi|118829|sp|P21189|DPO2_ECOLI  DNA POLYMERASE II (POL II) >gi|67...    99   2e-21
gi|118831|sp|P26019|DPOA_DROME  DNA POLYMERASE ALPHA CATALYTIC SU...    96   4e-21
gi|118881|sp|P03054|DPOL_HSV1S  DNA POLYMERASE >gi|67035|pir||DJB...    96   2e-20
gi|9629411|ref|NP_044632.1|  DNA polymerase [human herpesvirus 1]...    96   2e-20
gi|118880|sp|P04292|DPOL_HSV1K  DNA POLYMERASE >gi|67034|pir||DJB...    96   2e-20
gi|118879|sp|P07917|DPOL_HSV1A  DNA POLYMERASE >gi|67036|pir||DJB...    95   3e-20
gi|3913538|sp|P95690|DPO1_SULAC  DNA POLYMERASE I >gi|2129430|par...    94   9e-20
gi|118882|sp|P07918|DPOL_HSV2I  DNA POLYMERASE >gi|67039|pir||DJB...    93   1e-19
gi|[illegible]|sp|Q93749|DPO1_AERPE  DNA POLYMERASE I >gi|7448061|pir...    91   7e-19
gi|232019|sp|P30318|DPOL_NPVLD  DNA POLYMERASE >gi|484512|pir||JQ...    90   1e-18
gi|3913508|sp|O50607|DPO1_SULOH  DNA POLYMERASE I (DNA POLYMERASE...    88   4e-18
gi|6324067|ref|NP_014137.1|  DNA polymerase II, Pol2p [Saccharomy...    58   6e-09
gi|5453926|ref|NP_006222.1|  polymerase (DNA directed), epsilon [...    55   4e-08
gi|3913534|sp|P87154|DPOE_SCHPO  DNA POLYMERASE EPSILON, CATALYTI...    54   1e-07
```

Sequences with E-value WORSE than threshold

```
gi|6015024|sp|O70730|DPOL_RSIV  DNA POLYMERASE >gi|3176380|dbj|BA...    38   0.007
gi|1169407|sp|P43139|DPOL_ASFL6  DNA POLYMERASE >gi|480553|pir||S...    35   0.038
gi|1169408|sp|P42489|DPOL_ASFB7  DNA POLYMERASE >gi|457624|gb|AAA...    35   0.065
gi|9632637|ref|NP_049662.1|  DNA polymerase [Bacteriophage T4] >g...    31   0.74
gi|6325045|ref|NP_015113.1|  involved in ribosome biogenesis, Nip...    31   0.82
gi|6322305|ref|NP_012379.1|  Ssy5p [Saccharomyces cerevisiae] >gi...    30   1.3
gi|465671|sp|P34344|YK65_CAEEL  HYPOTHETICAL 73.3 KD PROTEIN C29E...    30   1.5
gi|6322923|ref|NP_012996.1|  Ykr070wp [Saccharomyces cerevisiae] ...    30   1.8
gi|417433|sp|P32871|P11A_BOVIN  PHOSPHATIDYLINOSITOL 3-KINASE CAT...    30   2.1
gi|6679317|ref|NP_032865.1|  phosphatidylinositol 3-kinase, catal...    29   2.2
gi|5453892|ref|NP_006209.1|  phosphoinositide-3-kinase, catalytic...    29   2.2
gi|232015|sp|P30319|DPOL_CBEPV  DNA POLYMERASE >gi|281209|pir||S2...    29   2.4
gi|3024676|sp|P54455|SYH_HELPY  HISTIDYL-TRNA SYNTHETASE (HISTIDI...    29   2.5
gi|[illegible]|sp|P46329|DHA3_BACSU  PROBABLE ALDEHYDE DEHYDROGENASE    29   2.9
gi|[illegible]|sp|Q9SZC9|AHM6_ARATH  POTENTIAL COPPER-TRANSPORTING A...    28   3.7
gi|[illegible]|sp|Q92K27|SYH_HELPJ  HISTIDYL-TRNA SYNTHETASE (HISTIDI...    28   3.8
gi|6015021|sp|Q84173|DPOL_ORFN2  DNA POLYMERASE >gi|1236947|gb|AA...    28   3.8
gi|6094269|sp|Q46455|SELB_MOOTH  SELENOCYSTEINE-SPECIFIC ELONGATI...    28   4.5
gi|8928039|sp|Q9X597|CQAA_BACTF  PESTICIDIAL CRYSTAL PROTEIN CRY2...    28   5.7
gi|14424453|sp|Q07635|DPO2_SULSO  DNA POLYMERASE II (DNA POLYMERA...    28   7.5
gi|3122742|sp|O24689|RL3_BYNPS  50S RIBOSOMAL PROTEIN L3 >gi|2446...    27   8.8
```

Alignments

```
>gi|13124219|sp|Q9LRE6|DPOD_ORYSA DNA POLYMERASE DELTA CATALYTIC SUBUNIT
 gi|19188570|dbj|BAA99573.1| (AB037899) DNA polymerase delta catalytic chain [Oryza sativa]
          Length = 1105

 Score =  180 bits (457), Expect = 8e-46
 Identities = 43/175 (24%), Positives = 66/175 (37%), Gaps = 32/175 (18%)

Query: 1   YRRGFFVTKKKYAVID-----EEDKITTRGLEIVRRDWSEIAKETQARVLEAILKHGDVE 55
           Y      ++KK+YA +       + DK+ T+G+E VRRD    + K       L  IL   DV
Sbjct: 804 YFPYLLISKKRYAGLYWTNPEKFDKMDTKGIETVRRDNCLLVKNLVTECLHKILVDRDVP 863

Query: 56  EAVRIVKEVTEKLSRHEVPPEKLVIYEAG-----------PHVAAA--------------- 90
            AV+ VK      L  + V     LVI +                HV  A
Sbjct: 864 GAVQYVKNTISDLLMNRVDLSLLVITKGLTKTGEDYAVKAAHVELAERMKKRDAATAPTV 923

Query: 91  ATVISYIVLK--GPGRVGDRAIPFDEFDPAKHRYDAEYYIENQVLPAVERILRAF 143
              + Y+++K      +  +R+              D +YY+ENQ+    + RI
Sbjct: 924 GDRVPYVIIKAAKGAKAYERSEDPIYVLDNNIPIDPQYYLENQISKPLLRIFEPI 978

>gi|13431465|sp|P77933|DPOL_PYRKO DNA POLYMERASE [CONTAINS  ENDONUCLEASE PI-PKOI (IVS-A); ENDONUCLEASE
          PI-PKOII (IVS-B)]
 gi|6706000|dbj|BAA06142.1| (D29671) DNA-dependent DNA polymerase [Pyrococcus sp.]
          Length = 1671

 Score =  180 bits (457), Expect = 8e-46
 Identities = 135/174 (77%), Positives = 143/174 (82%), Gaps = 23/174 (13%)

Query: 1    YRRGFFVTKKKYAVIDEEDKITTRGLEIVRRDWSEIAKETQARVLEAILKHGDVEEAVRI 60
           Y+RGFFVTKKKYAVIDEE KITTRGLEIVRRDWSEIAKETQARVLEA+LK GDVE+AVRI
Sbjct: 1480 YKRGFFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDVEKAVRI 1539

Query: 61   VKEVTEKLSRHEVPPEKLVIYEA----------GPHVAAAAT-------------VISYI 97
           VKEVTEKLS++EVPPEKLVI+E             GPHVA A               VISYI
Sbjct: 1540 VKEVTEKLSKYEVPPEKLVIHEQITRDLKDYKATGPHVAVAKRLAARGVKIRPGTVISYI 1599

Query: 98   VLKGPGRVGDRAIPFDEFDPAKHRYDAEYYIENQVLPAVERILRAFGYRKEDLR 151
           VLKG GR+GDRAIPFDEFDP KH+YDAEYYIENQVLPAVERILRAFGYRKEDLR
Sbjct: 1600 VLKGSGRIGDRAIPFDEFDPTKHKYDAEYYIENQVLPAVERILRAFGYRKEDLR 1653

>gi|13878475|sp|Q8HIB4|DPOL_THEGE DNA POLYMERASE [CONTAINS: ENDONUCLEASE PI-TSPGE8I (TSPGE8 POL-1
          INTEIN 1); ENDONUCLEASE PI-TSPGE8II (TSPGE8 POL-1 INTEIN
          4)]
 gi|10799865|emb|CAC12850.1| (AJ250333) DNA polymerase and endonuclease [Thermococcus sp. GE8]
          Length = 1699

 Score =  179 bits (455), Expect = 1e-45
```

Appendix 1
Fasta results: Type I DNA polymerase Palm domain against Swiss Prot Database

```
FASTA (3.36 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width   16
The best scores are.                                      opt bits E(110714)
b3862  polA, tesA; DNA polymerase I (POL I)  [EC.  ( 928)  599  142 1.5e-33
vch.VC0108  DNA polymerase I [EC:2.7.7.7]         ( 934)  568   91 3.9e-18
pae:PA5493  polA; DNA polymerase I [EC:2.7.7.7]   ( 913)  366   91 5.2e-18
HI0556  polA; DNA polymerase I (POL I). [EC:2.7.7  ( 930)  344   86 1.5e-16
xfa:XF1103  DNA polymerase I [EC:2.7.7.7]         ( 923)  331   83 1.1e-15
polA  DNA polymerase I. [EC:2.7.7.7]              ( 880)  299   76 1.5e-14
RP776  polA; DNA polymerase I  [EC:2.7.7.7]       ( 867)  297   75   2e-14
tma:TM1619  DNA-directed DNA polymerase I [EC:2.7  ( 895)  292   74 4.4e-13
Rv1629  polA; polymerase I. [EC:2.7.7.7]          ( 904)  289   73   7e-13
ML1381, polA, DNA polymerase I 1648220:1650955 re  ( 911)  286   73 1.1e-12
bha:BH3153  polA; DNA polymerase I [EC:2.7.7.7]   ( 876)  275   70 5.9e-12
dra:DR1707  DNA-directed DNA polymerase [EC:2.7.7  ( 956)  273   70 8.6e-12
slr0707  polA; DNA polymerase I. [EC:2.7.7.7]     ( 986)  252   65 2.2e-10
CP013S DNA polymerase I (polA)                    ( 870)  246   64   5e-10
nme:NME1982  DNA polymerase I [EC:2.7.7.7]        ( 938)  236   62 2.5e-09
cmu:TC0780  DNA polymerase I [EC:2.7.7.7]         ( 866)  232   61 4.3e-09
CT493  DNA Polymerase I. [EC:2.7.7.7]             ( 866)  230   60 5.9e-09
TP0105  DNA polymerase I (polA). [EC:2.7.7.7]     ( 997)  224   59 1.7e-08
cje:Cj0338c  polA; DNA polymerase I [EC:2.7.7.7]  ( 879)  218   58 3.7e-08
BB0548  polA; DNA polymerase I. [EC:2.7.7.7]      ( 908)  215   57 6.1e-08
HP1470  polA; DNA polymerase I (POL I). [EC:2.7.7  ( 892)  182   50 9.6e-06
hpj:jhp1163  DNA polymerase ; [EC:2.7.7.7]        ( 897)  182   50 9.6e-06
aq_1967  polA; DNA polymerase I (PolI). [EC:2.7.7  ( 574)  174   48 2.3e-05
cel:W03A3.2  DNA polymerase (CE14486) [EC:2.7.7.7  (1208)  160   45 0.00036
ath:T16K5.210  similarity to various ADP-RIBOSYLA  ( 169)   96   30      1.3
ath:AT4g32700  putative protein                    (1548)   98   31      6.1
CT362  Aspartokinase III. [EC:2.7.2.4]            ( 431)   91   29      6.1
yrvM  similar to hypothetical proteins.           ( 161)   84   28      7.9
vch:VC1150  hypothetical protein                  ( 201)   84   28      9.5


                                                                      Score    E
Sequences producing significant alignments.                          (bits) Value

        gi|3913510|sp|O52225|DPO1_THEFI  DNA POLYMERASE I, THERMOSTABLE (...   133  6e-32
        gi|3041672|sp|P52026|DPO1_BACST  DNA POLYMERASE I (POL I)              132  8e-32
        gi|5850621|sp|Q07700|DPO1_MYCTU  DNA POLYMERASE I (POL I) >gi|7434..   131  1e-31
        gi|6015001|sp|O34996|DPO1_BACSU  DNA POLYMERASE I (POL I) >gi|741 ..   131  2e-31
        gi|4169131|sp|Q04957|DPO1_BACCA  DNA POLYMERASE I (POL I) >gi|4196..   130  3e-31
        gi|6014999|sp|Q59156|DPO1_ANATH  DNA POLYMERASE I (POL I) >gi|140 ..   130  5e-31
        gi|1169403|sp|P46835|DPO1_MYCLE  DNA POLYMERASE I (POL I) >gi|107 ..   129  7e-31
        gi|2320101|sp|O30313|DPO1_THEFL  DNA POLYMERASE I, THERMOSTABLE (T...   129  1e-30
        gi|118828|sp|P19821|DPO1_THEAQ  DNA POLYMERASE I, THERMOSTABLE (T...   128  1e-30
        gi|6015003|sp|O32802|DPO1_LACLC  DNA POLYMERASE I (POL I) >gi|228...   126  5e-30
        gi|1706502|sp|P52028|DPO1_THETH  DNA POLYMERASE I, THERMOSTABLE (...   126  5e-30
        gi|2506365|sp|P80194|DPO1_THECA  DNA POLYMERASE I, THERMOSTABLE (...   125  1e-29
        gi|6166143|sp|P74933|DPO1_TREPA  DNA POLYMERASE I (POL I) >gi|743...   124  3e-29
        gi|6225286|sp|O51498|DPO1_BORBU  DNA POLYMERASE I (POL I) >gi|743...   123  4e-29
        gi|6225285|sp|Q55971|DPO1_SYNY3  DNA POLYMERASE I (POL I) >gi|743...   123  4e-29
        gi|7404361|sp|P52027|DPO1_DEIRA  DNA POLYMERASE I (POL I) >gi|747...   122  7e-29
        gi|118827|sp|P13252|DPO1_STRPN  DNA POLYMERASE I (POL I) >gi|9802...   122  9e-29
        gi|1188251|sp|P00582|DPO1_ECOLI  DNA POLYMERASE I (POL I) >gi|6705...   121  2e-28
        gi|1169402|sp|P43741|DPO1_HAEIN  DNA POLYMERASE I (POL I) >gi|107 ..   120  4e-28
        gi|4619601|sp|P30314|DPOL_BPSP1  DNA POLYMERASE                        115  9e-27
        gi|6015002|sp|O08307|DPO1_CHLAU  DNA POLYMERASE I (POL I) >gi|191...   115  2e-26
        gi|12229815|sp|Q9S1G2|DPO1_RHILE  DNA POLYMERASE I (POL I) >gi|55...   113  4e-26
        gi|6015004|sp|O05949|DPO1_RICPR  DNA POLYMERASE I (POL I) >gi|743...   113  4e-26
        gi|9789748|sp|Q9ZJE9|DPO1_HELPJ  DNA POLYMERASE I (POL I) >gi|743...   100  3e-22
        gi|9789856|sp|P56105|DPO1_HELPY  DNA POLYMERASE I (POL I)              100  5e-22
        gi|5729984|ref|NP_006587.1|  polymerase (DNA directed), theta; po...    95  1e-20
        gi|118853|sp|P20311|DPOL_BPT3  DNA POLYMERASE >gi|76915|pir||DQ75. .    88  2e-18
        gi|9627454|ref|NP_041982.1|  gene 5, DNA polymerase [Bacteriophag...   87  6e-18
        gi|9625474|ref|NP_039708.1|  predicted 66.2Kd protein [Mycobacter...   85  2e-17
        gi|2494181|sp|O01941|DPOG_PICPA  DNA POLYMERASE GAMMA (MITOCHONDR...   82  2e-16
        gi|9630428|ref|NP_046860.1|  DNA polymerase; gp44 [Mycobacterioph...   82  2e-16
        gi|1706504|sp|P15801|DPOG_YEAST  DNA POLYMERASE GAMMA (MITOCHONDR...   81  3e-16
        gi|4505937|ref|NP_002684.1|  polymerase (DNA directed), gamma [Ho...   80  4e-16
        gi|8567392|ref|NP_059490.1|  mitochondrial polymerase gamma [Mus  ..   80  4e-16
        gi|2494179|sp|Q91684|DPOG_XENLA  DNA POLYMERASE GAMMA (MITOCHONDR...   78  1e-15
        gi|2494178|sp|O92076|DPOG_CHICK  DNA POLYMERASE GAMMA (MITOCHONDR...   77  3e-15
        gi|2494182|sp|Q12704|DPOG_SCHPO  DNA POLYMERASE GAMMA (MITOCHONDR...   77  6e-15
        gi|14285433|sp|Q9Y767|DPOG_NEUCR  DNA POLYMERASE GAMMA (MITOCHOND...   77  7e-15
        gi|6015000|sp|O67779|DPO1_AQUAE  DNA POLYMERASE I (POL I) >gi|751...   77  7e-15
        gi|2494180|sp|Q27607|DPOG_DROME  DNA POLYMERASE GAMMA PRECURSOR (...   65  2e-11
        gi|118855|sp|P19822|DPOL_BPT5  DNA POLYMERASE >gi|67055|pir||DJBP...   61  3e-10


                        Sequences with E-value WORSE than threshold

        gi|118852|sp|P06229|DPOL_BPSP2  DNA POLYMERASE >gi|67052|pir||DJB...    33  9.0e5
        gi|9633592|ref|NP_051006.1|  P45 [Bacteriophage APSE-1] >gi|99106...    30  0.67
        gi|730217|sp|P18071|OAT_BACSU  ORNITHINE AMINOTRANSFERASE (ORNITH...    29  1.3
        gi|3913524|sp|O33845|DPOL_THSST  DNA POLYMERASE >gi|2293389|emb|C...    28  2.0
        gi|6323414|ref|NP_013486.1|  mitochondrial leucyl tRNA synthetase...    28  2.1
        gi|4826690|ref|NP_004932.1|  DEAD/H (Asp-Glu-Ala-Asp/His) box pol...    27  4.2
        gi|3913424|sp|O17438|DD15_STRPU  PUTATIVE PRE-MRNA SPLICING FACTO...    27  5.1
        gi|6319257|ref|NP_009340.1|  Yal061wp [Saccharomyces cerevisiae] ...    27  6.3
        gi|6081155|ref|NP_031865.1|  DEAD/H (Asp-Glu-Ala-Asp/His) box pol...    26  9.5


>gi|3913510|sp|O52225|DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (TF1 POLYMERASE 1)
 gi|2739139|gb|AAC46079.1|  (AF030120) thermostable DNA polymerase [Thermus filiformis]
          Length = 832

 Score =  133 bits (335), Expect = 6e-32
 Identities = 62/90 (68%), Positives = 68/90 (75%), Gaps = 2/90 (2%)

Query: 2   AAERAAAINAPMQGTAADIIKRAMIAVDAWLQAEQPKVRMIMQVHDELVFEVHKDDVDAVA 61
           AAER A N P+QGTAAD++K AM+ +   L+       +++QVHDELV EV +D  +
Sbjct: 741 AAERMAFNMPVQGTAADLMKIAMVKLFPRLK--PLGAHLLLQVHDELVLEVPEDRAFFAK 800

Query: 62  KQINQLMENCTRLDVPILVEVGSGENWDQA 91
             + ++MEN   LDVPL VEVG G +W +A
```

```
>gi|6015001|sp|O34996|DPO1_BACSU DNA POLYMERASE I (POL I)
 gi|7434813|pir||E69680 DNA polymerase I polA - Bacillus subtilis
 gi|2293272|gb|AAC00350.1| (AF008220) DNA-polymerase I [Bacillus subtilis]
 gi|2635374|emb|CAB14869.1| (Z99118) DNA polymerase I [Bacillus subtilis]
          Length = 880

 Score =  131 bits (331), Expect = 2e-31
 Identities = 83/89 (88%), Positives = 86/89 (73%)

Query: 3   AERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAK 62
           AER A+N P+QG+AADIIK AMI + A L+ +Q + R+++QVHDEL+FE   + + V
Sbjct: 791 AERTAMNTPIQGSAADIIKKAMIDMAARLKEKQLKARLLLQVHDELIFEAPKEEIEILEK 850

Query: 63  QIHQLMENCTRLDVPLLVEVGSGENWDQA 91
           + ++ME+    LDVPL V+   SG +W  A
Sbjct: 851 LVPEVMEHALALDVPLKVDFASGPSWYDA 879

>gi|4416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I)
 gi|419652|pir||JX0256 DNA-directed DNA polymerase (EC 2.7.7.7) - Bacillus caldotenax
 gi|912445|dbj|BAA02361.1| (D12982) DNA polymerase [Bacillus caldotenax]
          Length = 877

 Score =  130 bits (328), Expect = 3e-31
 Identities = 39/89 (43%), Positives = 62/89 (68%)

Query: 3   AERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAK 62
           AER A+N P+QG+AADIIK+AMI ++A L+ E+ + R+++QVHDEL+ E  K+++ + +
Sbjct: 788 AERMAMNTPIQGSAADIIKKAMIDLNARLKEERLQARLLLQVHDELILEAPKEEMERLCR 847

Query: 63  QIHQLMENCTRLDVPLLVEVGSGENWDQA 91
           + ++ME+    L VPL V+   G  W  A
Sbjct: 848 LVKREMENAVRLKVPLVVEVKEGLNWYE 876

>gi|6014999|sp|Q59156|DPO1_ANATH DNA POLYMERASE I (POL I)
 gi|1405138|emb|CAA67184.1| (X98575) DNA-directed DNA polymerase; DNA-dependent DNA polymerase
          [Anaerocellum thermophilum]
          Length = 850

 Score =  130 bits (327), Expect = 5e-31
 Identities = 41/88 (46%), Positives = 59/88 (66%)

Query: 3   AERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAK 62
           AER A+N+P+QG+ ADI+K AMI V    L+     + ++I+QVHDEL+ E   ++ D V +
Sbjct: 761 AERIAMNSPIQGSPADIMKLAMIKVYQKLKENNLKSKIILQVHDELIIEAPYEEKDIVKE 820

Query: 63  QIHQLMENCTRLDVPLLVEVGSGENWDQ 90
           + + MEN  RL VPL+VEV  G NW +
Sbjct: 821 IVKREMENAVRLKVPLVVEVKEGLNWYE 848

>gi|1169403|sp|P46835|DPO1_MYCLE DNA POLYMERASE I (POL I)
 gi|1076026|pir||S49522 DNA polymerase I - Mycobacterium leprae
 gi|743482|pir||S77659 DNA-directed DNA polymerase (EC 2.7.7.7) - Mycobacterium leprae
 gi|559913|emb|CAA86364.1| (Z46257) DNA polymerase I [Mycobacterium leprae]
 gi|13093270|emb|CAC31762.1| (AL583921) DNA polymerase I [Mycobacterium leprae]
          Length = 911

 Score =  129 bits (326), Expect = 7e-31
 Identities = 47/90 (52%), Positives = 62/90 (68%)

Query: 2   AERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAK 61
           AERAA+NAP+QG+AADIIK AMIAVD  L+  +    RM++QVHDEL+FEV   + + +
Sbjct: 820 AERAALNAPIQGSAADIIKVAMIAVDKSLKQAKLASRMLLQVHDELLFEVAIGERQIE 879

Query: 62  KQIHQLMENCTRLDVPLLVEVGSGENWDQA 91
           + + M +    LDVPL V VG G +W  A
Sbjct: 880 AMVREQMGSAYPLDVPLEVSVGFGRSWGAA 909

>gi|232010|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (TFL POLYMERASE 1)
 gi|201488|pir||S26675 DNA-directed DNA polymerase (EC 2.7.7.7) I - Thermus aquaticus
 gi|48166|emb|CAA46900.1| (X66105) DNA-directed DNA  polymerase [Thermus thermophilus]
          Length = 831

 Score =  129 bits (324), Expect = 1e-30
 Identities = 46/90 (51%), Positives = 57/90 (63%), Gaps = 2/90 (2%)

Query: 2   AERAAINAIMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAK 61
           AER A N P+QGTAAD++K AM+ +    LQ      RM++QVHDELV L  KD  + VA
Sbjct: 747 AERMAFNMPVQGTAADLMKLAMVRLFPALQE--LGARMLLQVHDELVLEAPKDRAEHVA 804
```

Appendix 1
Fasta results: Type II DNA polymerase Palm domain against Swiss Prot Database

```
FASTA (3.36 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width:  16
The best scores are:                                      opt bits E(110735)
PAB1128 (polI) DE:DNA polymerase I                ( 771)  356   90 4.4e-18
PH1947  1235aa long hypothetical DNA-directed DNA (1235)  347   88 2.7e-17
MJ0985  putative DNA polymerase. [EC:2.7.7.7] [SP (1034)  [illegible]  70 1.6e-11
MTH1208  DNA-dependent DNA polymerase family B (P ( 586)  189   52 1.1e-06
AF0497  polB, DNA polymerase B1. [EC:2.7.7.7]     ( 781)  173   48 1.7e-05
ape:APE2098  DNA-directed DNA polymerase [EC:2.7. ( 784)  150   43 0.00067
sce:YPL167C  REV3, PSO1, DNA polymerase zeta cata (1504)  115   35    0.29
sce:YDL102W  CDC2, POL3, TEX1; DNA polymerase del (1097)  100   32     2.4
ape:APE0099  DNA-directed DNA polymerase (pfu pol ( 959)   99   32     2.5
MTH208  DNA-dependent DNA polymerase family B (Po ( 223)   90   29     3.2


Sequences with E-value BETTER than threshold


                                                                      Score    E
Sequences producing significant alignments:                          (bits) Value

gi|6015025|sp|P56689|DPOL_THEGO  DNA POLYMERASE (TO POL) >gi|4699...   123  4e-29
gi|13431465|sp|P77933|DPOL_PYRKO  DNA POLYMERASE [CONTAINS: ENDON...   123  5e-29
gi|3913540|sp|Q56366|DPOL_THES9  DNA POLYMERASE >gi|7434808|pir||...   119  1e-27
gi|2494186|sp|Q51334|DPOL_PYRSD  DNA POLYMERASE (DEEP VENT DNA PO...   115  1e-26
gi|399403|sp|P80061|DPOL_PYRFU  DNA POLYMERASE (PFU POLYMERASE) >...   111  2e-25
gi|13124219|sp|Q9LRE6|DPOD_ORYSA  DNA POLYMERASE DELTA CATALYTIC ...   106  9e-24
gi|3913525|sp|O48901|DPOD_SOYBN  DNA POLYMERASE DELTA CATALYTIC S...   104  3e-23
gi|14521919|ref|NP_127396.1|  DNA polymerase I [Pyrococcus abyssi...   103  4e-23
gi|13124220|sp|Q9LVN7|DPOD_ARATH  DNA POLYMERASE DELTA CATALYTIC ...   102  9e-23
gi|3913530|sp|P77932|DPOL_PYRSE  DNA POLYMERASE >gi|1495770|emb|C...   101  1e-22
gi|1706505|sp|P52431|DPOD_MOUSE  DNA POLYMERASE DELTA CATALYTIC S...   101  1e-22
gi|13122029|sp|P97283|DPOD_MESAU  DNA POLYMERASE DELTA CATALYTIC S...   101  1e-22
gi|13124716|sp|P46588|DPOD_CANAL  DNA POLYMERASE DELTA CATALYTIC ...   101  2e-22
gi|11067381|ref|NP_067694.1|  DNA polymerase delta, catalytic sub...   101  3e-22
gi|4505933|ref|NP_002682.1|  polymerase (DNA directed), delta 1, ...   100  5e-22
gi|3913526|sp|O59610|DPOL_PYRHO  DNA POLYMERASE >gi|7446920|pir||...   100  7e-22
gi|118838|sp|P28339|DPOD_BOVIN  DNA POLYMERASE DELTA CATALYTIC SU...    99  9e-22
gi|13124718|sp|P54358|DPOD_DROME  DNA POLYMERASE DELTA CATALYTIC ...    99  1e-21
gi|12644199|sp|P30316|DPOD_SCHPO  DNA POLYMERASE DELTA CATALYTIC ...    99  1e-21
gi|6320101|ref|NP_010181.1|  largest and catalytic subunit of DNA ..    93  7e-20
gi|13124199|sp|P90629|DPOD_CAEEL  DNA POLYMERASE DELTA CATALYTIC ...    92  1e-19
gi|6015023|sp|O71121|DPOL_RHCM6  DNA POLYMERASE >gi|2944240|gb|AA...    90  5e-19
gi|232011|sp|P30315|DPOD_PLAFK  DNA POLYMERASE DELTA CATALYTIC SU...    90  7e-19
gi|3913527|sp|O60673|DPOZ_HUMAN  DNA POLYMERASE ZETA CATALYTIC SU...    89  1e-18
gi|6015011|sp|Q94636|DPOA_OXYNO  DNA POLYMERASE ALPHA CATALYTIC S...    89  2e-18
gi|3915679|sp|Q58295|DPOL_METJA  DNA POLYMERASE [CONTAINS: MJA PO...    89  2e-18
gi|6015013|sp|O89042|DPOA_RAT  DNA POLYMERASE ALPHA CATALYTIC SUB...    89  2e-18
gi|6919874|sp|Q61493|DPOZ_MOUSE  DNA POLYMERASE ZETA CATALYTIC SU...    87  4e-18
gi|6015010|sp|O48653|DPOA_ORYSA  DNA POLYMERASE ALPHA CATALYTIC S...    87  5e-18
gi|8393995|ref|NP_058633.1|  polymerase (DNA-directed), alpha; po...    87  6e-18
gi|9631753|ref|NP_048532.1|  PBVC-1 DNA polymerase [Paramecium bu...    85  1e-17
gi|6325090|ref|NP_015158.1|  DNA polymerase zeta subunit, Rev3p [ ..    85  2e-17
gi|6679409|ref|NP_032918.1|  DNA polymerase alpha 1, 180 kDa [Mus...    85  2e-17
gi|232016|sp|P30320|DPOL_CHVN2  DNA POLYMERASE >gi|281076|pir||B4...    84  3e-17
gi|9296965|sp|Q9YUS3|DPOL_HSVT1  DNA POLYMERASE >gi|4165073|gb|AA...    84  5e-17
gi|14251046|ref|NP_116408.1|  T54 [Tupaia herpesvirus] >gi|929696...    84  6e-17
gi|9628761|ref|NP_043792.1|  DNA polymerase [Human herpesvirus 7]...    83  6e-17
gi|9625903|ref|NP_040151.1|  ORF28 (AA1-1194) [Human herpesvirus ...    83  7e-17
gi|6015019|sp|Q69025|DPOL_GPCMV  DNA POLYMERASE >gi|459763|gb|AAA...    83  8e-17
gi|6015012|sp|Q27152|DPOA_OXYTR  DNA POLYMERASE ALPHA CATALYTIC S...    82  2e-16
gi|9626767|ref|NP_041039.1|  DNA polymerase [Equine herpesvirus 1...    82  2e-16
gi|118882|sp|P07918|DPOL_HSV21  DNA POLYMERASE >gi|67039|pir||DJB...    81  3e-16
gi|9628011|ref|NP_042605.1|  DNA polymerase replicative subunit [...    80  6e-16
gi|118881|sp|P09854|DPOL_HSV1S  DNA POLYMERASE >gi|67035|pir||DJB...    80  8e-16
gi|118880|sp|P04292|DPOL_HSV1K  DNA POLYMERASE >gi|67034|pir||DJB ..    80  8e-16
gi|9625739|ref|NP_039968.1|  DNA polymerase (8) [human herpesviru...    79  1e-15
gi|118883|sp|P30401|DPOA_SCHPO  DNA POLYMERASE ALPHA CATALYTIC SU...    79  1e-15
gi|118888|sp|P27172|DPOL_MCMVS  DNA POLYMERASE >gi|67043|pir||DJB...    79  2e-15
gi|9629411|ref|NP_044632.1|  DNA polymerase [human herpesvirus 1]...    79  2e-15
gi|118879|sp|P07917|DPOL_HSV1A  DNA POLYMERASE >gi|67036|pir||DJB...    78  2e-15
gi|6324227|ref|NP_014297.1|  DNA polymerase I alpha subunit, p180...    77  4e-15
gi|9628340|ref|NP_042931.1|  U38, DNA polymerase [Human herpesvir. .    77  4e-15
gi|118831|sp|P26019|DPOA_DROME  DNA POLYMERASE ALPHA CATALYTIC SU...    77  5e-15
gi|118829|sp|P21189|DPO2_ECOLI  DNA POLYMERASE II (POL II) >gi|67...    76  1e-14
gi|6015022|sp|Q85428|DPOL_RCMVM  DNA POLYMERASE                         76  1e-14
gi|6015009|sp|O00874|DPOA_LEIDO  DNA POLYMERASE ALPHA CATALYTIC S...    75  3e-14
gi|13878475|sp|Q9HH84|DPOL_THEG8  DNA POLYMERASE [CONTAINS: ENDON...    73  6e-14
gi|118884|sp|P27727|DPOA_TRYBB  DNA POLYMERASE ALPHA CATALYTIC SU...    73  6e-14
gi|1706513|sp|P52025|DPOL_METVO  DNA POLYMERASE >gi|495654|gb|AAA...    73  6e-14
gi|9625965|ref|NP_040211.1|  DNA  polymerase [Saimiriine herpesvi...    73  8e-14
gi|3913528|sp|P74918|DPOL_THEFM  DNA POLYMERASE (POL TFU) >gi|165...    73  9e-14
gi|6015024|sp|O70736|DPOL_RSIV  DNA POLYMERASE >gi|3176380|dbj|BA...    72  1e-13
gi|9625657|ref|NP_039908.1|  BALF5 DNA polymerase (early), homolo...    71  4e-13
gi|3913524|sp|O33845|DPOL_THEST  DNA POLYMERASE >gi|2293389|emb|C...    70  4e-13
gi|11498108|ref|NP_069333.1|  DNA polymerase B1 (polB) [Archaeogl...    70  5e-13
gi|1232020|sp|P30317|DPOL_THELI  DNA POLYMERASE (VENT DNA POLYMERA...    70  6e-13
gi|9630872|ref|NP_047469.1|  DNA Polymerase-AcMNPV orf65 [Bombyx ...    67  5e-12
gi|[illegible]|sp|[illegible]  DNA POLYMERASE II >gi|7434801|pi...    67  6e-12
gi|[illegible]|sp|P95690|DPOL_SULAC  DNA POLYMERASE I >gi|[illegible]|pir...    66  9e-12
gi|9627571|ref|NP_042094.1|  E9L [Variola virus] >gi|[illegible]|sp|[illegible]...    65  2e-11
```

```
gi|9627808|ref|NP_054095.1|  DNA-dependant DNA-polymerase [Autogr...    65  7e-11
gi|9790985|ref|NP_063712.1|  E9L; putative [Vaccinia virus] >gi|1...    64  4e-11
gi|3913508|sp|O50607|DPO1_SULOH  DNA POLYMERASE I (DNA POLYMERASE...    64  4e-11
gi|118892|sp|P06856|DPOL_VACCV  DNA POLYMERASE >gi|83576|gb|AAB5...    64  4e-11
gi|9630008|ref|NP_046226.1|  DNA polymerase [Orgyia pseudotsugata...    63  5e-11
gi|3913522|sp|O27276|DPOL_METTH  DNA POLYMERASE >gi|7482291|pir||...    63  5e-11
gi|6015020|sp|Q90162|DPOL_NPVCF  DNA POLYMERASE >gi|747638|gb|AAC...    63  9e-11
gi|6225283|sp|O93745|DPO1_AERPE  DNA POLYMERASE I >gi|7434800|pir...    62  2e-10
gi|12643274|sp|P26811|DPO1_SULSO  DNA POLYMERASE I >gi|2052353|gb...    59  1e-09
gi|118859|sp|P21402|DPOL_IOWBV  DNA POLYMERASE >gi|67047|pir||DJV...    58  3e-09
gi|6015021|sp|Q84173|DPOL_OREN2  DNA POLYMERASE >gi|1236947|gb|AA...    55  2e-08
gi|232019|sp|P30318|DPOL_NPVLD  DNA POLYMERASE >gi|484512|pir||JQ...    50  7e-07
```

Sequences with E-value WORSE than threshold

```
gi|1169407|sp|P43139|DPOL_ASFL6  DNA POLYMERASE >gi|480553|pir||S...    37  0.005
gi|1169406|sp|P42489|DPOL_ASFB7  DNA POLYMERASE >gi|457624|gb|AAA...    37  0.007
gi|118845|sp|P05664|DPOL_ADE07  DNA POLYMERASE >gi|58527|emb|CAA2...    35  0.023
gi|118844|sp|P04495|DPOL_ADE05  DNA POLYMERASE >gi|67027|pir||DJA...    34  0.036
gi|9626168|ref|NP_040516.1|  DNA polymerase [Human adenovirus typ...    34  0.036
gi|9626559|ref|NP_040853.1|  DNA polymerase [Human adenovirus typ...    34  0.047
gi|3913535|sp|P87503|DPOL_ADE04  DNA POLYMERASE >gi|1841694|emb|C...    33  0.11
gi|6015016|sp|O72540|DPOL_ADEB3  DNA POLYMERASE >gi|3128453|gb|AA...    33  0.12
gi|9629220|ref|NP_044409.1|  DNA polymerase [canine adenovirus ty...    32  0.14
gi|9626627|ref|NP_040915.1|  DNA polymerase [Human adenovirus typ...    32  0.16
gi|14424453|sp|Q07635|DPO2_SULSO  DNA POLYMERASE II (DNA POLYMERA...    32  0.17
gi|232015|sp|P30319|DPOL_CBEPV  DNA POLYMERASE >gi|281209|pir||S2...    32  0.24
gi|2494187|sp|Q65946|DPOL_ADECC  DNA POLYMERASE >gi|1477650|gb|AA...    32  0.25
gi|3023653|sp|Q38087|DPOL_BPR69  DNA POLYMERASE (GP43) >gi|643555...    30  0.52
gi|728801|sp|P39067|ACUC_BACSU  ACETOIN UTILIZATION ACUC PROTEIN ...    30  0.54
gi|9845478|ref|NP_064286.1|  DNA polymerase [bovine adenovirus 2]  .    30  0.63
gi|9628845|ref|NP_043878.1|  D2b pol [fowl adenovirus 1] >gi|2494. .    29  1 3
gi|1352306|sp|P09804|DPO1_KLULA  DNA POLYMERASE >gi|101132|pir||D...    29  1 3
gi|9790319|ref|NP_062890.1|  p130 structural polyprotein [Simdbis...    29  1.5
gi|130578|sp|P27265|POLS_SINDO  STRUCTURAL POLYPROTEIN (P130) (CO.     29  1.5
gi|6225462|sp|O66776|GLYA_AQUAE  SERINE HYDROXYMETHYLTRANSFERASE ...    29  1.5
gi|7993727|sp|O43184|ADA12_HUMAN  ADAM 12 PRECURSOR (A DISINTEGRIN  .    29  1 6
gi|14520810|ref|NP_126285.1|  DNA topoisomerase VI, subunit B [Py...    29  1.7
gi|9296955|sp|Q88469|DPOL_ADET1  DNA POLYMERASE >gi|1336626|gb|AA. .    28  2 1
gi|118830|sp|P05468|DPO2_KLULA  DNA POLYMERASE (PLASMID PGKL-2 PR...    28  2.2
gi|66861.7|sp|O34033|Y66X_CHLPS  HYPOTHETICAL 66.2 KDA PROTEIN (O..    28  2.5
gi|9629878|ref|NP_041148.1|  DNA polymerase [ictalurid herpesviru...    28  2.6
gi|13124601|sp|O74020|TPGB_PYRHO  TYPE II DNA TOPOISOMERASE VI SU...    28  2.7
gi|9626390|ref|NP_040719.1|  gene 2 product [Bacteriophage PZA] > .    28  2.9
gi|11750071|sp|P42018|WAPA_BACST  WALL-ASSOCIATED PROTEIN PRECURSO  .    27  4 8
gi|461963|sp|P33538|DPOM_NEUIN  PROBABLE DNA POLYMERASE >gi|10189...    27  4.8
gi|9626351|ref|NP_040682.1|  orf I [Enterobacteria phage PRD1] >g...    27  4.8
gi|127331|sp|P13819|MSP1_PLAFF  MEROZOITE SURFACE PROTEIN 1 PRECU...    27  4.9
gi|6226649|sp|P08569|MSP1_PLAFM  MEROZOITE SURFACE PROTEIN 1 PREC...    27  5.0
gi|118849|sp|P03680|DPOL_BPPH2  DNA POLYMERASE (EARLY PROTEIN GP2.     27  5.0
gi|13162353|ref|NP_077071.1|  complement factor I [Rattus norvegi ..    27  5.8
gi|6166016|sp|Q90125|COAT_GMDNV  COAT PROTEIN VP1 (STRUCTURAL PRO...    27  6.5
gi|6321346|ref|NP_011423.1|  nuclear pore protein; Nup145p [Sacch...    27  7.8
gi|118848|sp|P19894|DPOL_BPM2  DNA POLYMERASE >gi|76896|pir||JQ01...    27  8.1
gi|6015017|sp|Q37882|DPOL_BPB03  DNA POLYMERASE (EARLY PROTEIN GP...    27  8.2
gi|11132914|sp|P57203|HIS7_BUCAI  HISTIDINE BIOSYNTHESIS BIFUNCTI...    27  8.2
gi|12827751|sp|P19598|MSP1_PLAF3  MEROZOITE SURFACE PROTEIN 1 PREC...    27  8.3
gi|17300111|sp|P38426|TPS3_YEAST  ALPHA,ALPHA-TREHALOSE-PHOSPHATE ...    26  9.5
gi|1729957|sp|P51070|HS3_VITVI  STILBENE SYNTHASE 3 (RESVERATROL...    26  9.9
gi|116180|sp|P13952|CG2B_SPISO  G2/MITOTIC-SPECIFIC CYCLIN B >gi|...    26  10.0
```

Alignments

```
>gi|6015025|sp|P56689|DPOL_THEGO DNA POLYMERASE (TO POL)
 gi|4699806|pdb|1TGO|A Chain A, Thermostable B Type Dna Polymerase From Thermococcus
          Gorgonarius
         Length = 775

 Score =  123 bits (310), Expect = 4e-29
 Identities = 70/76 (92%), Positives = 74/76 (97%)

Query: 1   RECAESVTAWGRQYIETTMREIEEKFGFKVLYADTDGFFATIPGADAETVKNKAKEFLNY 60
           +ECAESVTAWGRQYIETT+REIEEKFGFKVLYADTDGFFATIPGADAETVK KAKEFL+Y
Sbjct: 507 KECAESVTAWGRQYIETTIREIEEKFGFKVLYADTDGFFATIPGADAETVKKKAKEFLDY 566

Query: 61  INPRLPGLLELEYEGF 76
           IN +LPGLLELEYEGF
Sbjct: 567 INAKLPGLLELEYEGF 582

>gi|13431465|sp|P77933|DPOL_PYRKO DNA POLYMERASE [CONTAINS- ENDONUCLEASE PI-PKOI (IVS-A), ENDONUCLEASE
          PI-PKOII (IVS-B)]
 gi|6706000|dbj|BAA06142.2| (D29671) DNA-dependent DNA polymerase [Pyrococcus sp.]
         Length = 1671

 Score =  123 bits (310), Expect = 5e-29
 Identities = 61/76 (80%), Positives = 69/76 (90%)

Query: 1    RECAESVTAWGRQYIETTMREIEEKFGFKVLYADTDGFFATIPGADAETVKNKAKEFLNY 60
           +ECAESVTAWGR+YI  T++EIEEK+GFKV+Y+DTDGFFATIPGADAETVK KA EFL Y
Sbjct: 1604 KECAESVTAWGREYITMTIKEIEEKYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKY 1663
```

Appendix 1
Fasta results: Type I DNA polymerase Exonuclease domain against Swiss Prot Database

>gi|118825|sp|P00582|DPO1_ECOLI DNA POLYMERASE I (POL I)
>gi|1169402|sp|P43741|DPO1_HAEIN DNA POLYMERASE I (POL I)
>gi|12229815|sp|Q9S1G2|DPO1_RHILE DNA POLYMERASE I (POL I)
>gi|6166143|sp|P74933|DPO1_TREPA DNA POLYMERASE I (POL I)
>gi|6015002|sp|O08307|DPO1_CHLAU DNA POLYMERASE I (POL I)
>gi|6225285|sp|Q55971|DPO1_SYNY3 DNA POLYMERASE I (POL I)
>gi|118855|sp|P19822|DPOL_BPT5 DNA POLYMERASE
>gi|6225284|sp|O51498|DPO1_BORBU DNA POLYMERASE I (POL I)
>gi|4619601|sp|P30314|DPOL_BPSP1 DNA POLYMERASE
>gi|118793|sp|P13267|DPO3_BACSU DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|8928564|sp|Q01780|PMC2_HUMAN POLYMYOSITIS/SCLERODERMA AUTOANTIGEN 2 (AUTOANTIGEN PM/SCL 2)
>gi|[illegible]|sp|P44442|RND_HAEIN RIBONUCLEASE D (RNASE D)
>gi|7710082|ref|NP_057908.1| polymyositis/scleroderma autoantigen 2 [Mus musculus]
>gi|6015001|sp|O34996|DPO1_BACSU DNA POLYMERASE I (POL I)
>gi|13959683|sp|Q53665|DPO3_STAAU DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|[illegible]|sp|P09156|RND_ECOLI RIBONUCLEASE D (RNASE D)
>gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)
>gi|13959318|sp|Q9KA72|DPO3_BACHD DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|6324574|ref|NP_014643.1| involved in 5.8S rRNA processing; Rrp6p [Saccharomyces cerevisiae]
>gi|6014995|sp|O67074|DP3E_AQUAE DNA POLYMERASE III, EPSILON CHAIN
>gi|416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I)
>gi|118827|sp|P13252|DPO1_STRPN DNA POLYMERASE I (POL I)
>gi|[illegible]|sp|Q9FDE9|DPO3_STRPY DNA POLYMERASE III POLC TYPE (POLIII)
>gi|6015000|sp|O67779|DPO1_AQUAE DNA POLYMERASE I (POL I)
>gi|[illegible]|sp|Q5LDI7|DPO3_LACLA DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|[illegible]|ref|NP_109722.1| DNA polymerase III (dnaE; alpha chain [Mycoplasma pneumoniae]
>gi|1169403|sp|P46835|DPO1_MYCLE DNA POLYMERASE I (POL I)
>gi|[illegible]|ref|NP_072691.1| DNA polymerase III, subunit alpha 'polC-1' [Mycoplasma genitalium]
>gi|[illegible]|ref|NP_046860.1| DNA polymerase, gp44 [Mycobacteriophage D29]
>gi|[illegible]
>gi|6015003|sp|O32801|DPO1_LACLC DNA POLYMERASE I (POL I)
>gi|6226742|sp|Q10146|YASE_SCHPO HYPOTHETICAL 89.6 KDA PROTEIN C3H8.11 IN CHROMOSOME I
>gi|1706437|sp|P54394|DING_BACSU PROBABLE ATP-DEPENDENT HELICASE DING HOMOLOG
>gi|9789748|sp|Q92JE9|DPO1_HELPJ DNA POLYMERASE I (POL I)
>gi|[illegible]|ref|NP_078211.1| DNA polymerase III alpha chain 1 [Ureaplasma urealyticum]
>gi|[illegible]
>gi|9789856|sp|P56105|DPO1_HELPY DNA POLYMERASE I (POL I)
>gi|[illegible]|ref|NP_389708.1| predicted 66.2Kd protein [Mycobacteriophage 13]
>gi|[illegible]
>gi|1722908|sp|Q10394|YL91_MYCTU HYPOTHETICAL 69.3 KDA PROTEIN RV2191
>gi|9789746|sp|P76281|EXOX_ECOLI EXODEOXYRIBONUCLEASE X (EXONUCLEASE X) (EXO X)
>gi|6225287|sp|Q9ZHF6|DPO3_THEMA DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|[illegible]|sp|Q9RC79|DP3E_RICPR DNA POLYMERASE III, EPSILON CHAIN
>gi|[illegible]|sp|O83649|DP3E_TREPA DNA POLYMERASE III, EPSILON CHAIN
>gi|6014999|sp|Q59156|DPO1_ANATH DNA POLYMERASE I (POL I)
>gi|[illegible]
>gi|6014992|sp|O60045|DP3A_RHOCA DNA POLYMERASE III ALPHA SUBUNIT
>gi|[illegible]
>gi|14424449|sp|P47729|DPO3_MYCPU DNA POLYMERASE III POLC-TYPE (POLIII)
>gi|[illegible]
>gi|2506367|sp|P14566|DP3E_SALTY DNA POLYMERASE III, EPSILON CHAIN
>gi|1169396|sp|P42745|DP3E_HAEIN DNA POLYMERASE III, EPSILON CHAIN
>gi|[illegible]
>gi|11132278|sp|P57337|DP3E_BUCAI DNA POLYMERASE III, EPSILON CHAIN
>gi|585062|sp|Q07700|DPO1_MYCTU DNA POLYMERASE I (POL I)
>gi|118805|sp|P03007|DP3E_ECOLI DNA POLYMERASE III, EPSILON CHAIN
>gi|[illegible]
>gi|2833216|sp|Q99980|DP3E_BUCAP DNA POLYMERASE III, EPSILON CHAIN
>gi|1706502|sp|P52028|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (TTH POLYMERASE 1)
>gi|2506365|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (TAC POLYMERASE 1)
>gi|118828|sp|P19821|DPO1_THEAQ DNA POLYMERASE I, THERMOSTABLE (TAQ POLYMERASE 1)
>gi|232010|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (TFL POLYMERASE 1)
>gi|3913510|sp|O53225|DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (TFI POLYMERASE 1)

>gi|7404361|sp|P52027|DPO1_DEIRA DNA POLYMERASE I (POL I)
****

Estas de aquí son un PSI-BLAST quitando a las polIII, en la última iteración (la séptima) salieron de todas formas 3 pol III. El archivo para clustal se llama exopolAPSIBLAST.txt

gi|118825|sp|P00582|DPO1_ECOLI DNA POLYMERASE I (POL I) >gi|6705...	**270**	**1e-72**
gi|1169402|sp|P43741|DPO1_HAEIN DNA POLYMERASE I (POL I) >gi|107...	251	6e-67
gi|12229815|sp|Q9S1G2|DPO1_RHILE DNA POLYMERASE I (POL I) >gi|55...	224	9e-59
gi|6166143|sp|P74933|DPO1_TREPA DNA POLYMERASE I (POL I) >gi|743...	215	5e-56
gi|118855|sp|P19822|DPOL_BPT5 DNA POLYMERASE >gi|67055|pir||DJBP...	195	4e-50
gi|6015002|sp|O08307|DPO1_CHLAU DNA POLYMERASE I (POL I) >gi|191...	195	6e-50
gi|6225285|sp|Q55971|DPO1_SYNY3 DNA POLYMERASE I (POL I) >gi|743...	191	5e-49
gi|6225284|sp|O51498|DPO1_BORBU DNA POLYMERASE I (POL I) >gi|743...	176	2e-44
gi|461960|sp|P30314|DPOL_BPSP1 DNA POLYMERASE	158	8e-39
gi|6015001|sp|O34996|DPO1_BACSU DNA POLYMERASE I (POL I) >gi|743...	156	3e-38
gi|1173094|sp|P44442|RND_HAEIN RIBONUCLEASE D (RNASE D) >gi|1075...	148	6e-36
gi|3041672|sp|P52026|DPO1_BACST DNA POLYMERASE I (POL I)	148	8e-36
gi|416913|sp|Q04957|DPO1_BACCA DNA POLYMERASE I (POL I) >gi|4196...	144	1e-34
gi|133152|sp|P09155|RND_ECOLI RIBONUCLEASE D (RNASE D) >gi|67284...	142	3e-34
gi|6015003|sp|O32801|DPO1_LACLC DNA POLYMERASE I (POL I) >gi|228...	142	4e-34
gi|1169403|sp|P46835|DPO1_MYCLE DNA POLYMERASE I (POL I) >gi|107...	139	2e-33
gi|9630428|ref|NP_046860.1| DNA polymerase; gp44 [Mycobacterioph...	138	6e-33
gi|6324574|ref|NP_014643.1| involved in 5.8S rRNA processing; Rr...	136	2e-32
gi|118827|sp|P13252|DPO1_STRPN DNA POLYMERASE I (POL I) >gi|9802...	135	5e-32
gi|8928564|sp|Q01780|PMC2_HUMAN POLYMYOSITIS/SCLERODERMA AUTOANT...	134	1e-31
gi|7710082|ref|NP_057908.1| polymyositis/scleroderma autoantigen...	130	2e-30
gi|6014999|sp|Q59156|DPO1_ANATH DNA POLYMERASE I (POL I) >gi|140...	125	6e-29
gi|9625474|ref|NP_039708.1| predicted 66.2Kd protein [Mycobacter...	125	6e-29
gi|6015000|sp|O67779|DPO1_AQUAE DNA POLYMERASE I (POL I) >gi|751...	124	7e-29
gi|6226742|sp|Q10146|YASB_SCHPO HYPOTHETICAL 89.6 KDA PROTEIN C3...	122	4e-28
gi|9789856|sp|P56105|DPO1_HELPY DNA POLYMERASE I (POL I)	119	4e-27
gi|9789748|sp|Q9ZJE9|DPO1_HELPJ DNA POLYMERASE I (POL I) >gi|743...	115	6e-26
gi|585062|sp|Q07700|DPO1_MYCTU DNA POLYMERASE I (POL I) >gi|7434...	107	1e-23
gi|1706502|sp|P52028|DPO1_THETH DNA POLYMERASE I, THERMOSTABLE (...	87	2e-17
gi|118828|sp|P19821|DPO1_THEAQ DNA POLYMERASE I, THERMOSTABLE (T...	87	2e-17
gi|2506365|sp|P80194|DPO1_THECA DNA POLYMERASE I, THERMOSTABLE (...	86	3e-17
gi|232010|sp|P30313|DPO1_THEFL DNA POLYMERASE I, THERMOSTABLE (T...	81	1e-15
gi|3913510|sp|O52225|DPO1_THEFI DNA POLYMERASE I, THERMOSTABLE (...	75	7e-14
gi|13959683|sp|Q53665|DPO3_STAAU DNA POLYMERASE III POLC-TYPE (P...	59	4e-09
gi|7404361|sp|P52027|DPO1_DEIRA DNA POLYMERASE I (POL I) >gi|747...	59	7e-09
gi|14424449|sp|P47729|DPO3_MYCPU DNA POLYMERASE III POLC-TYPE (P...	44	2e-04
gi|118793|sp|P13267|DPO3_BACSU DNA POLYMERASE III POLC-TYPE (POL...	42	6e-04
gi|14194673|sp|Q9FDF9|DPO3_STRPY DNA POLYMERASE III POLC-TYPE (P...	42	7e-04

>DPO1_ECOLI DNA POLYMERASE I (POL I)
VISYDNYVTILDEETLKAWIAKLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYI
PVAHDYLDAPDQISRERALELLKPLLEDEKALKVGQNLKYDRGILANYGIELRGIAFDTM
LESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIALEEAGRYAAEDAD
VTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRIER

>DPO1_HAEIN DNA POLYMERASE I (POL I)
IDRTKYETLLTQADLTRWIEKLNAAKLIAVDTETDSLDYMSANLVGISFALENGEAAYLP
LQLDYLDAPKTLEKSTALAAIKPILENPNIHKIGQNIKFDESIFARHGIELQGVEFDTML
LSYTLNS-TGRHNMDDLAKRYLGHETIAFESLAGKGKSQLTFNQIPLEQATEYAAEDADV
TMKLQQALWLKLQEEPTLVELYKTMELPLLHVLSRMER

>DPO1_RHILE DNA POLYMERASE I (POL I)
DHSAYVTIRDLVTLDRWIADARATGLVAFDTETTSLDAMQAELVGFSLAIADNTADPTGT
KIRAAYVPLVHKNGVGDLLGGGLADNQIPMRDALPRLKALLEDESVLKVAQNLKYDYLLL
KRYGIETRSFD-DTMLISYVLDAGTGAHGMDPLSEKFLGHTPIPYKDVAGSGKANVTFDL
VDIDRATHYAAEDADVTLRLWLVLKPRLAA-AGLTSVYERLERPLLPVLARME

>DPO1_TREPA DNA POLYMERASE I (POL I)
SAGHYRGVTDPVELKRIIDCACANGVVAFDCETDGLHPHDTRLVGFSICFQEAEAFYVPL
IVPDVSLHTESTQCTCARSTNVETEKECTEQHGVSASAVQDPA--YVQAVMHQLRRLWND
ETLTLVMHNGKFDYHVMHRAGVFEHCACNIFDTMVAAWLLDPDRGTYGMDVLAASFFQIR
TITFEEVVAKGQ---TFAHVPYECAVRYAAEDADITFRLYHYLKLRLE-TAGLLSVFETI

Appendix 1
Fasta results: Type II DNA polymerase Exonuclease domain against Swiss Prot Database

FASTA (3.46 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
join: 36, opt: 24, gap-pen: -12/ -2, width: 16

The best scores are:		opt	bits	E(110718)
PH1947 1235aa long hypothetical DNA-directed DNA	(1235)	1359	282	3.9e-75
PAB1128 (polI) DE-DNA polymerase I	(771)	1356	281	4e-75
MJ0885 putative DNA polymerase. [EC:2.7.7.7] [SP	(1634)	734	157	2.3e-37
MTH1208 DNA-dependent DNA polymerase family B (P	(586)	674	145	4.2e-34
AF0497 polB; DNA polymerase B1. [EC:2.7.7.7]	(781)	428	96	3.5e-19
ape:APE2098 DNA-directed DNA polymerase [EC:2.7.	(784)	356	81	7.7e-15
tac:Ta0907 DNA polymerase (PolB), large chain re	(796)	323	75	7.6e-13
TVG0859451 DNA polymerase	(800)	311	72	4e-12
ath:At2g27120 putative DNA polymerase epsilon ca	(2154)	265	63	5.3e-09
hal:VNG0521G polB1; DNA polymerase B1 [EC:2.7.7.	(901)	238	58	1.1e-07
sce:YNL262W POL2, DUN2; DNA polymerase epsilon,	(2222)	229	56	8e-07
ath:T23G18.11 putative DNA polymerase gi|3885342	(2271)	226	56	1.2e-06
cel.F33H2.5 DNA polymerase family b (3 domains)	(2144)	214	53	6.3e-06
pae:PA1886 polB; DNA polymerase II [EC:2.7.7.7]	(787)	196	49	3.4e-05
b0060 polB, dinA; DNA polymerase II (pol II). [E	(783)	193	49	5.2e-05
vch:VC1212 DNA polymerase II [EC:2.7.7.7]	(787)	193	49	5.2e-05
mmu:99660 Pola1; DNA polymerase alpha 1, 180 kDa	(1465)	180	46	0.00052
ape:APE0099 DNA-directed DNA polymerase (pfu pol	(959)	173	45	0.00097
ath:F12B7.5 putative DNA polymerase zeta catalyt	(1871)	137	38	0.25
ath:T1F15.3 Similar to putative DNA polymerase g	(1894)	137	38	0.25
bha:BH1765 unknown conserved protein in B. subti	(430)	121	34	0.7
nme:NMB1417 conserved hypothetical protein	(264)	115	33	1.1
sce:YDL102W *CDC2, POL3, TEX1, DNA polymerase del*	(1097)	122	34	1.3
sce:YPL167C REV3, PSO1; DNA polymerase zeta cata	(1504)	122	35	1.7
cel:F10C2.4 DNA polymerase family b (CE09308) [E	(1081)	115	33	3.4
hpj:jhp0628 putative	(276)	107	31	3.4
HP0688 H. pylori predicted coding region HP0688.	(166)	104	30	3.5
MTH491 conserved protein.	(303)	104	31	5.6
spsF ipa-68d, spore coat polysaccharide biosynth	(239)	99	29	9.2
aq_484 eno, enolase. [EC:4.2.1.11]	(426)	102	30	9.7

Sequences with E-value BETTER than threshold

Sequences producing significant alignments:	Score (bits)	E Value
gi|6015025|sp|P56689|DPOL_THEGO DNA POLYMERASE (TO POL) >gi|4699...	323	2e-88
gi|13431465|sp|P77933|DPOL_PYRKO DNA POLYMERASE (CONTAINS: ENDON...	322	3e-88
gi|13878475|sp|Q9HH84|DPOL_THEGB DNA POLYMERASE (CONTAINS: ENDON...	320	2e-87
gi|3913528|sp|P74918|DPOL_THEFM DNA POLYMERASE (POL TFU) >gi|165...	318	5e-87
gi|3913540|sp|Q56366|DPOL_THES9 DNA POLYMERASE >gi|7434808|pir||...	318	7e-87
gi|399403|sp|P80061|DPOL_PYRFU DNA POLYMERASE (PFU POLYMERASE) >...	317	9e-87
gi|3913526|sp|O59610|DPOL_PYRHO DNA POLYMERASE >gi|7446920|pir||...	312	4e-85
gi|2494186|sp|Q51334|DPOL_PYRSD DNA POLYMERASE (DEEP VENT DNA PO...	310	1e-84
gi|3913530|sp|P77932|DPOL_PYRSE DNA POLYMERASE >gi|1495770|emb|C...	306	3e-83
gi|14521919|ref|NP_127396.1| DNA polymerase I [Pyrococcus abyssi...	305	3e-83
gi|232020|sp|P30317|DPOL_THELI DNA POLYMERASE (VENT DNA POLYMERA...	293	2e-79
gi|13124219|sp|Q9LRE6|DPOD_ORYSA DNA POLYMERASE DELTA CATALYTIC ...	291	6e-79
gi|3913524|sp|O33845|DPOL_THEST DNA POLYMERASE >gi|2293389|emb|C...	291	7e-79
gi|1706505|sp|P52431|DPOD_MOUSE DNA POLYMERASE DELTA CATALYTIC S...	283	1e-76
gi|4505933|ref|NP_002682.1| polymerase (DNA directed), delta 1, ...	283	2e-76
gi|14251046|ref|NP_116408.1| T54 [Tupaia herpesvirus] >gi|929696...	283	3e-76
gi|6320101|ref|NP_010181.1| largest and catalytic subunit of DNA...	282	4e-76
gi|929696|sp|P07703|DPOL_HSVT DNA POLYMERASE >gi|416507|gb|AA...	281	6e-76
gi|11067381|ref|NP_067694.1| DNA polymerase delta, catalytic sub...	281	6e-76
gi|3122029|sp|P97283|DPOD_MESAU DNA POLYMERASE DELTA CATALYTIC S...	279	2e-75
gi|118838|sp|P28339|DPOD_BOVIN DNA POLYMERASE DELTA CATALYTIC SU...	279	2e-75
gi|12644199|sp|P30316|DPOD_SCHPO DNA POLYMERASE DELTA CATALYTIC ...	278	4e-75
gi|13124718|sp|P54358|DPOD_DROME DNA POLYMERASE DELTA CATALYTIC ...	277	1e-74
gi|13124716|sp|P46588|DPOD_CANAL DNA POLYMERASE DELTA CATALYTIC ...	276	2e-74
gi|13124220|sp|Q9LVN7|DPOD_ARATH DNA POLYMERASE DELTA CATALYTIC ...	275	6e-74
gi|6015023|sp|O71121|DPOL_RHCM6 DNA POLYMERASE >gi|2944240|gb|AA...	273	2e-73
gi|13124199|sp|P90829|DPOD_CAEEL DNA POLYMERASE DELTA CATALYTIC ...	268	5e-72
gi|9625657|ref|NP_039908.1| BALF5 DNA polymerase (early), homolo...	268	5e-72
gi|232011|sp|P30315|DPOD_PLAFK DNA POLYMERASE DELTA CATALYTIC SU...	267	1e-71
gi|3913525|sp|O48901|DPOD_SOYBN DNA POLYMERASE DELTA CATALYTIC S...	265	4e-71
gi|9628011|ref|NP_042605.1| DNA polymerase replicative subunit [...	263	2e-70
gi|6015019|sp|Q69025|DPOL_GPCMV DNA POLYMERASE >gi|459763|gb|AAA...	260	1e-69
gi|9625965|ref|NP_040211.1| DNA polymerase [Saimiriine herpesvi...	259	4e-69
gi|3913522|sp|O27276|DPOL_METTH DNA POLYMERASE >gi|7482291|pir||...	257	9e-69
gi|9628340|ref|NP_042931.1| U38, DNA polymerase [Human herpesvir...	254	1e-67
gi|118882|sp|P07918|DPOL_HSV21 DNA POLYMERASE >gi|67039|pir||DJB...	252	3e-67
gi|118888|sp|P27172|DPOL_MCMVS DNA POLYMERASE >gi|67043|pir||DJB...	250	1e-66
gi|9625739|ref|NP_039988.1| DNA polymerase (8) [human herpesviru...	250	2e-66
gi|6015022|sp|Q85428|DPOL_RCMVM DNA POLYMERASE	247	2e-65
gi|118880|sp|P04292|DPOL_HSV1K DNA POLYMERASE >gi|67034|pir||DJB...	246	2e-65
gi|9629411|ref|NP_044632.1| DNA polymerase [human herpesvirus 1]...	246	3e-65
gi|118879|sp|P07917|DPOL_HSV1A DNA POLYMERASE >gi|67036|pir||DJB...	245	4e-65
gi|118881|sp|P09854|DPOL_HSV1S DNA POLYMERASE >gi|67035|pir||DJB...	244	7e-65
gi|3915679|sp|O58295|DPOL_METJA DNA POLYMERASE (CONTAINS: MJA PO...	244	7e-65

gi|9628761|ref|NP_043792.1| DNA polymerase [Human herpesvirus 7]... 244 9e-65
gi|9626767|ref|NP_041039.1| DNA polymerase [Equine herpesvirus 1... 241 6e-64
gi|9625903|ref|NP_040151.1| ORF28 (AA1-1194) [Human herpesvirus ... 230 2e-60
gi|1706513|sp|P52025|DPOL_METVO DNA POLYMERASE >gi|495654|gb|AAA... 221 1e-57
gi|11498108|ref|NP_069333.1| DNA polymerase B1 (polB) [Archaeogl... 220 2e-57
gi|5453926|ref|NP_006222.1| polymerase (DNA directed), epsilon [... 214 8e-56
gi|9631750|ref|NP_048522.1| PBV-1 DNA polymerase [Paramecium bu... 213 2e-55
gi|232016|sp|P30320|DPOL_CHVN2 DNA POLYMERASE >gi|281076|pir||B4... 212 3e-55
gi|6225286|sp|O93746|DPO2_AERPE DNA POLYMERASE II >gi|7434801|pi... 212 5e-55
gi|6324067|ref|NP_014137.1| DNA polymerase II; Pol2p [Saccharomy... 207 1e-53
gi|3913534|sp|P87154|DPOE_SCHPO DNA POLYMERASE EPSILON, CATALYTI... 205 4e-53
gi|8393995|ref|NP_058633.1| polymerase (DNA-directed), alpha; po... 203 2e-52
gi|6015013|sp|O89042|DPOA_RAT DNA POLYMERASE ALPHA CATALYTIC SUB... 202 4e-52
gi|6015024|sp|O70736|DPOL_RSIV DNA POLYMERASE >gi|3176380|dbj|BA... 201 9e-52
gi|6679409|ref|NP_032918.1| DNA polymerase alpha 1, 180 kDa [Mus... 196 3e-50
gi|3913527|sp|O60673|DPOZ_HUMAN DNA POLYMERASE ZETA CATALYTIC SU... 192 4e-49
gi|3913502|sp|O05706|DPO3_SULSH DNA POLYMERASE III (DNA POLYMERA... 188 7e-48
gi|3913515|sp|P95979|DPO3_SULSO DNA POLYMERASE III (DNA POLYMERA... 186 3e-47
gi|118831|sp|P26019|DPOA_DROME DNA POLYMERASE ALPHA CATALYTIC SU... 182 5e-46
gi|118829|sp|P21189|DPO2_ECOLI DNA POLYMERASE II (POL II) >gi|67... 179 3e-45
gi|6015010|sp|O48653|DPOA_ORYSA DNA POLYMERASE ALPHA CATALYTIC S... 179 6e-45
gi|6015012|sp|Q27152|DPOA_OXYTR DNA POLYMERASE ALPHA CATALYTIC S... 175 5e-44
gi|3913538|sp|P95690|DPO1_SULAC DNA POLYMERASE I >gi|2129430|pir... 174 9e-44
gi|126432741|sp|P26811|DPO1_SULSO DNA POLYMERASE I >gi|2052353|gb... 174 9e-44
gi|6325090|ref|NP_015158.1| DNA polymerase zeta subunit; Rev3p [... 174 2e-43
gi|9630008|ref|NP_046226.1| DNA polymerase [Orgyia pseudotsugata... 172 5e-43
gi|6015011|sp|Q94636|DPOA_OXYNO DNA POLYMERASE ALPHA CATALYTIC S... 172 5e-43
gi|3913508|sp|O50607|DPO1_SULOH DNA POLYMERASE I (DNA POLYMERASE... 172 7e-43
gi|6919874|sp|Q61493|DPOZ_MOUSE DNA POLYMERASE ZETA CATALYTIC SU... 169 3e-42
gi|1169407|sp|P43139|DPOL_ASFL6 DNA POLYMERASE >gi|480553|pir||S... 167 1e-41
gi|6015020|sp|Q90162|DPOL_NPVCF DNA POLYMERASE >gi|747648|gb|AAC 167 1e-41
gi|118834|sp|P27727|DPOA_TRYBB DNA POLYMERASE ALPHA CATALYTIC SU... 166 2e-41
gi|9630872|ref|NP_047469.1| DNA Polymerase-AcMNPV orf65 [Bombyx ... 165 6e-41
gi|1169408|sp|P42489|DPOL_ASFB7 DNA POLYMERASE >gi|457624|gb|AAA... 162 3e-40
gi|9627808|ref|NP_054095.1| DNA-dependant DNA-polymerase [Autogr... 162 5e-40
gi|9632637|ref|NP_049662.1| DNA polymerase [Bacteriophage T4] >g... 160 2e-39
gi|3023653|sp|Q38087|DPOL_BPR69 DNA POLYMERASE (GP43) >gi|643555... 158 5e-39
gi|6225283|sp|O93745|DPO1_AERPE DNA POLYMERASE I >gi|7434800|pir... 156 4e-38
gi|118833|sp|P28040|DPOA_SCHPO DNA POLYMERASE ALPHA CATALYTIC SU... 154 1e-37
gi|6324227|ref|NP_014297.1| DNA polymerase I alpha subunit, p180... 150 2e-36
gi|232019|sp|P30318|DPOL_NPVLD DNA POLYMERASE >gi|484512|pir||JQ... 146 3e-35
gi|6015009|sp|O00874|DPOA_LEIDO DNA POLYMERASE ALPHA CATALYTIC S... 145 6e-35
gi|232015|sp|P30319|DPOL_CBEPV DNA POLYMERASE >gi|281209|pir||S2... 107 1e-23
gi|118859|sp|P21402|DPOL_FOWPV DNA POLYMERASE >gi|67047|pir||DJV... 103 3e-22
gi|6015021|sp|Q84173|DPOL_ORFN2 DNA POLYMERASE >gi|1236947|gb|AA... 101 1e-21
gi|9627571|ref|NP_042094.1| E9L [Variola virus] >gi|461961|sp|P3... 93 4e-19
gi|9790985|ref|NP_063712.1| E9L; putative [Vaccinia virus] >gi|1... 92 8e-19
gi|118892|sp|P06856|DPOL_VACCV DNA POLYMERASE >gi|335756|gb|AAB5... 91 1e-18
gi|1730955|sp|P50837|YPRB_BACSU HYPOTHETICAL 48.0 KDA PROTEIN IN... 88 1e-17
gi|12044881|ref|NP_072691.1| DNA polymerase III, subunit alpha (... 77 2e-14
gi|9626878|ref|NP_041148.1| DNA polymerase [ictalurid herpesviru... 71 2e-12
gi|13507773|ref|NP_109722.1| DNA polymerase III (dnaE) alpha cha... 64 2e-10
gi|13357937|ref|NP_078211.1| DNA polymerase III alpha chain 1 [U... 62 1e-09

Sequences with E-value WORSE than threshold

gi|461962|sp|P33537|DPOM_NEUCR PROBABLE DNA POLYMERASE >gi|28335... 41 0.002
gi|12959341|sp|Q57811|Y365_METJA HYPOTHETICAL PROTEIN MJ0365 >gi|... 39 0.006
gi|6225287|sp|Q9ZHF6|DPO3_THEMA DNA POLYMERASE III POLC-TYPE (PO... 38 0.011
gi|1352308|sp|P09804|DPO1_KLULA DNA POLYMERASE >gi|101132|pir||S... 34 0.26
gi|6014996|sp|O83649|DP3E_TREPA DNA POLYMERASE III, EPSILON CHAI... 34 0.29
gi|9626351|ref|NP_040682.1| orf I [Enterobacteria phage PRD1] >g... 33 0.52
gi|1730190|sp|P54804|GALC_CANFA GALACTOCEREBROSIDASE PRECURSOR (... 32 0.72
gi|2494182|sp|Q12704|DPOG_SCHPO DNA POLYMERASE GAMMA (MITOCHONDR... 32 0.89
gi|2496012|sp|Q57923|Y500_METJA HYPOTHETICAL PROTEIN MJ0500 >gi|... 32 1.1
gi|9627454|ref|NP_041982.1| gene 5, DNA polymerase [Bacteriophag... 31 2.0
gi|2498300|sp|Q54443|DEXT_STRMU DEXTRANASE PRECURSOR (ALPHA-1,6-... 31 2.1
gi|118853|sp|P20311|DPOL_BPT3 DNA POLYMERASE >gi|76915|pir||S075... 31 2.2
gi|6679927|ref|NP_032105.1| galactosylceramidase (Mus musculus) ... 31 2.2
gi|6014992|sp|O68045|DP3A_RHOCA DNA POLYMERASE III ALPHA SUBUNIT... 31 2.5
gi|8134550|sp|Q9WZ25|LEUD_THEMA 3-ISOPROPYLMALATE DEHYDRATASE SM... 30 3.4
gi|401241|sp|P31254|UBAY_MOUSE UBIQUITIN-ACTIVATING ENZYME E1 Y ... 30 5.5
gi|9627557|ref|NP_042080.1| C16L [Variola virus] >gi|465091|sp|P... 29 5.9
gi|6319436|ref|NP_009518.1| B subunit of DNA polymerase alpha-pr... 29 7.0
gi|129772|sp|P04094|PENK_RAT PROENKEPHALIN A PRECURSOR [CONTAINS... 29 7.1
gi|1730335|sp|P40803|PKSK_BACSU PUTATIVE POLYKETIDE SYNTHASE PKSK... 29 7.4
gi|6647814|sp|Q92CW8|SECD_RICPR PROTEIN-EXPORT MEMBRANE PROTEIN ... 29 7.4
gi|2494181|sp|Q01941|DPOG_PICPA DNA POLYMERASE GAMMA (MITOCHONDR... 29 7.5
gi|1267309|sp|P29889|VF12_VACCP PROTEIN F12 >gi|335701|gb|AAA4828... 29 8.1

Alignments

>gi|6015025|sp|P56689|DPOL_THEGO DNA POLYMERASE (TO POL)
gi|4699806|pdb|1TGO|A Chain A, Thermostable B Type Dna Polymerase From Thermococcus Gorgonarius
Length = 773

Score = 323 bits (828), Expect = 2e-88
Identities = 231/256 (90%), Positives = 241/256 (93%)

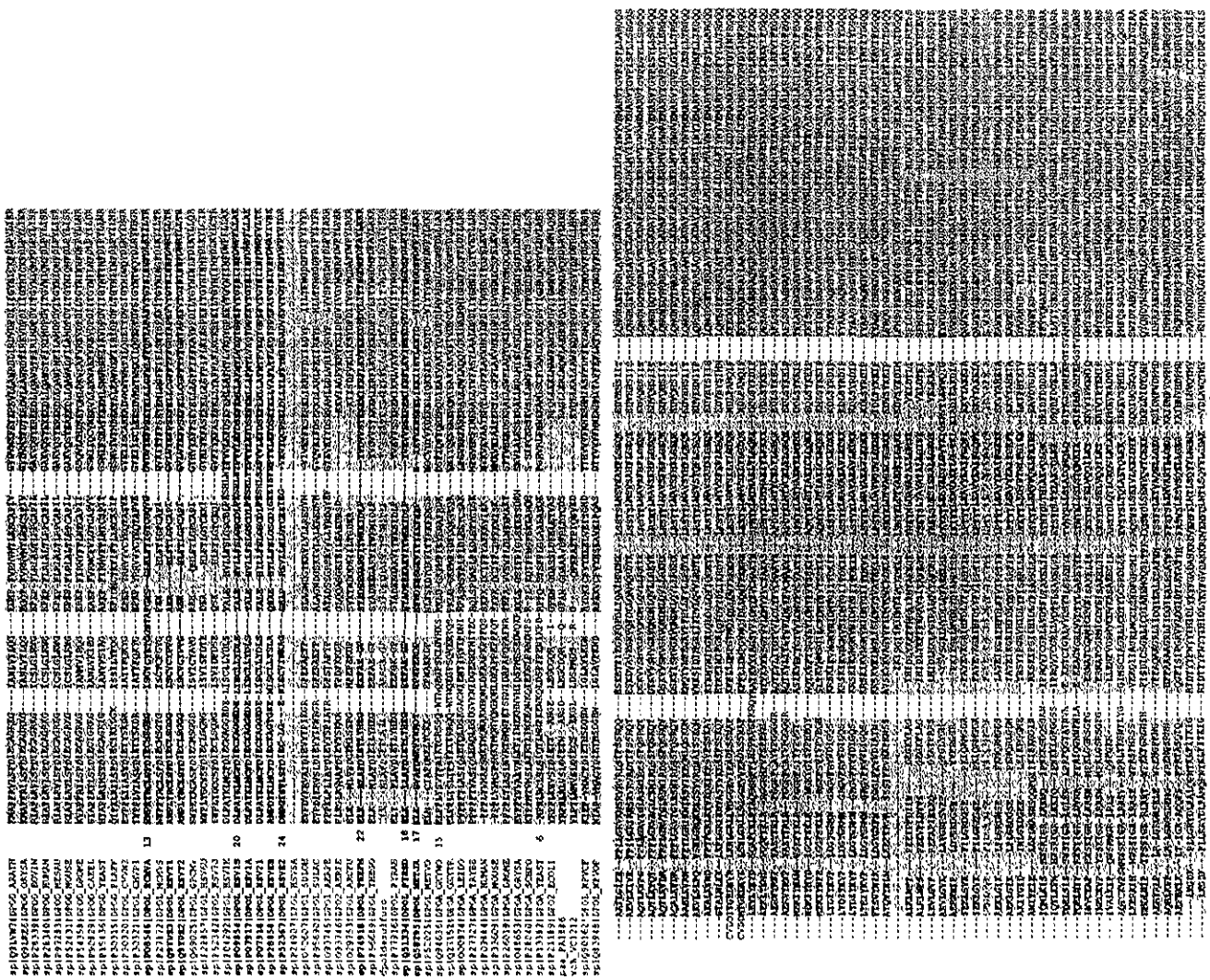
Query: 1 ELRMLAPDIETLAHAGAAAGAGPILMISYADEEGARVITWKNIDLPYVESVSTEKEMIKR 60

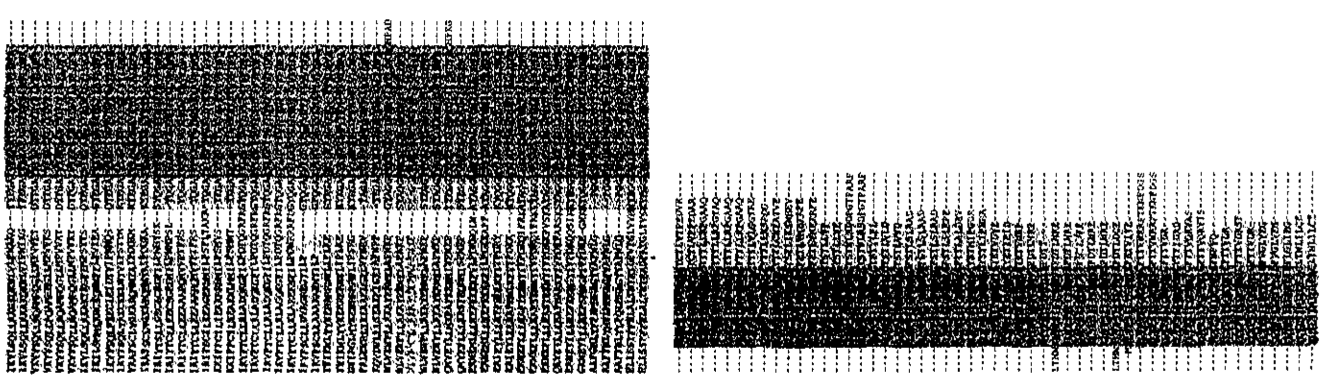
Appendix I
Fasta results: Type II DNA polymerase Thumb domain against Swiss Prot Database

```
FASTA (3.36 June 2000) function [optimized, BL50 matrix (15:-5)] ktup: 2
 join: 36, opt: 24, gap-pen: -12/ -2, width:  16
The best scores are:                                      opt bits E(110731)
PAB1128 (polI) DE:DNA polymerase I                  ( 771)  458  101 4.3e-21
PH1947  1235aa long hypothetical DNA-directed DNA   (1235)  456  101 8.1e-21
MJ0885  putative DNA polymerase. [EC:2.7.7.7] [SP   (1634)  297   69 3.7e-11
MTH208  DNA-dependent DNA polymerase family B [Po   ( 223)  282   66 6.2e-11
ape:APE2098  DNA-directed DNA polymerase [EC:2.7.   ( 784)  243   58 3.7e-08
AF0497  polB; DNA polymerase B1. [EC:2.7.7.7]       ( 781)  217   53 1.3e-06
sce:YDL102W  CDC2, POL3, TEX1; DNA polymerase del   (1097)  125   35     0.6
ape:APE2229  DNA-directed DNA polymerase [EC:2.7.   ( 637)  121   34    0.68
b0060  polB, dinA; DNA polymerase II (pol II). [E   ( 783)  120   34    0.92
mmu:99660  Polal; DNA polymerase alpha 1, 180 kDa   (1465)  115   33       3
hal:VNG0521G  polB1; DNA polymerase B1 [EC:2.7.7.   ( 901)  111   32     3.6
pae:PA1886  polB; DNA polymerase II [EC:2.7.7.7]    ( 787)  110   32     3.7
vch:VC1212  DNA polymerase II [EC:2.7.7.7]          ( 787)  107   31     5.6
TVG0550709 ribosomal protein small subunit S9       ( 200)   97   29     7.6
AF1938  conserved hypothetical protein.             ( 673)  103   30     8.6
ath:F12B7.5  putative DNA polymerase zeta catalyt   (1871)  108   32     9.7
ath:T1F15.3  Similar to putative DNA polymerase g   (1894)  108   32     9.8
MJ1630  conserved hypothetical protein.             (1139)  105   31     9.9
cel:F10C2.4  DNA polymerase family b (CE09308) [E   (1081)  104   31      11


Sequences with E-value BETTER than threshold


                                                                    Score    E
Sequences producing significant alignments:                        (bits) Value

        gi|13124219|sp|Q9LRE6|DPOD_ORYSA  DNA POLYMERASE DELTA CATALYTIC ...   180  8e-46
        gi|13431465|sp|P77933|DPOL_PYRKO  DNA POLYMERASE [CONTAINS: ENDON...   180  8e-46
        gi|13878475|sp|Q9HH84|DPOL_THEGS  DNA POLYMERASE [CONTAINS: ENDON...   179  1e-45
        gi|6015025|sp|P56689|DPOL_THEGO  DNA POLYMERASE (TO POL) >gi|4699...   178  2e-45
        gi|3913540|sp|Q56366|DPOL_THESS  DNA POLYMERASE >gi|7434808|pir||...   178  3e-45
        gi|3913525|sp|O48901|DPOD_SOYBN  DNA POLYMERASE DELTA CATALYTIC S...   178  3e-45
        gi|3913528|sp|P74918|DPOL_THEFM  DNA POLYMERASE (POL TFU) >gi|165...   178  4e-45
        gi|3913530|sp|P77932|DPOL_PYRSE  DNA POLYMERASE >gi|1495770|emb|C...   172  1e-43
        gi|14521919|ref|NP_127396.1|  DNA polymerase I [Pyrococcus abyssi...   172  1e-43
        gi|13124718|sp|P54358|DPOD_DROME  DNA POLYMERASE DELTA CATALYTIC ...   172  2e-43
        gi|3913526|sp|O59610|DPOL_PYRHO  DNA POLYMERASE >gi|7446920|pir||...   172  2e-43
        gi|232020|sp|P30317|DPOL_THELI  DNA POLYMERASE (VENT DNA POLYMERA...   171  4e-43
        gi|13124220|sp|Q9LVN7|DPOD_ARATH  DNA POLYMERASE DELTA CATALYTIC ...   170  5e-43
        gi|399403|sp|P80061|DPOL_PYRFU  DNA POLYMERASE (PFU POLYMERASE) >...   170  7e-43
        gi|2494186|sp|Q51334|DPOL_PYRSD  DNA POLYMERASE (DEEP VENT DNA PO...   170  7e-43
        gi|4505933|ref|NP_002682.1|  polymerase (DNA directed), delta 1, ...   169  1e-42
        gi|1706505|sp|P52431|DPOD_MOUSE  DNA POLYMERASE DELTA CATALYTIC S...   168  2e-42
        gi|12644199|sp|P30316|DPOD_SCHPO  DNA POLYMERASE DELTA CATALYTIC ...   168  3e-42
        gi|118838|sp|P28339|DPOD_BOVIN  DNA POLYMERASE DELTA CATALYTIC SU...   168  3e-42
        gi|3913524|sp|O33845|DPOL_THEST  DNA POLYMERASE >gi|2293389|emb|C...   168  3e-42
        gi|13124716|sp|P46588|DPOD_CANAL  DNA POLYMERASE DELTA CATALYTIC ...   168  4e-42
        gi|11067381|ref|NP_067694.1|  DNA polymerase delta, catalytic sub...   167  6e-42
        gi|3122029|sp|P97283|DPOD_MESAU  DNA POLYMERASE DELTA CATALYTIC S...   167  6e-42
        gi|13124199|sp|P90829|DPOD_CAEEL  DNA POLYMERASE DELTA CATALYTIC ...   163  6e-41
        gi|6320101|ref|NP_010181.1|  largest and catalytic subunit of DNA...   158  2e-39
        gi|6225286|sp|O93746|DPO2_AERPE  DNA POLYMERASE II >gi|7434801|pi...   157  9e-39
        gi|11496108|ref|NP_069333.1|  DNA polymerase B1 (polB) [Archaeogl...   153  7e-38
        gi|3915679|sp|Q58295|DPOL_METJA  DNA POLYMERASE [CONTAINS: MJA PO...   149  1e-36
        gi|6015009|sp|O00874|DPOA_LEIDO  DNA POLYMERASE ALPHA CATALYTIC S...   147  5e-36
        gi|1706513|sp|P52025|DPOL_METVO  DNA POLYMERASE >gi|495654|gb|AAA...   147  6e-36
        gi|3913502|sp|O05706|DPO3_SULSH  DNA POLYMERASE III (DNA POLYMERA...   146  1e-35
        gi|232011|sp|P30315|DPOD_PLAFK  DNA POLYMERASE DELTA CATALYTIC SU...   145  2e-35
        gi|118834|sp|P27727|DPOA_TRYBB  DNA POLYMERASE ALPHA CATALYTIC SU...   145  3e-35
        gi|3913515|sp|P95979|DPO3_SULSO  DNA POLYMERASE III (DNA POLYMERA...   140  9e-34
        gi|232016|sp|P30320|DPOL_CHVN2  DNA POLYMERASE >gi|281076|pir||B4...   136  1e-32
        gi|6015022|sp|Q85428|DPOL_RCMVM  DNA POLYMERASE                        136  1e-32
        gi|118833|sp|P28040|DPOA_SCHPO  DNA POLYMERASE ALPHA CATALYTIC SU...   135  2e-32
        gi|9631753|ref|NP_048532.1|  PBVC-1 DNA polymerase [Paramecium bu...   132  2e-31
        gi|9629965|ref|NP_040211.1|  DNA  polymerase [Saimiriine herpesvi...   131  3e-31
        gi|6015019|sp|Q69025|DPOL_GPCMV  DNA POLYMERASE >gi|459763|gb|AAA...   130  7e-31
        gi|118888|sp|P27172|DPOL_MCMVS  DNA POLYMERASE >gi|67043|pir||DJB...   128  2e-30
        gi|9625657|ref|NP_039908.1|  BALF5 DNA polymerase (early), homolo...   128  3e-30
        gi|6015023|sp|O71121|DPOL_RHCM6  DNA POLYMERASE >gi|2944240|gb|AA...   126  8e-30
        gi|9296965|sp|Q9YUS3|DPOL_HSVT1  DNA POLYMERASE >gi|4165073|gb|AA...   125  2e-29
        gi|14251046|ref|NP_116408.1|  T54 [Tupaia herpesvirus] >gi|929696...   125  2e-29
        gi|3913527|sp|O60673|DPOZ_HUMAN  DNA POLYMERASE ZETA CATALYTIC SU...   123  7e-29
        gi|9628761|ref|NP_043792.1|  DNA polymerase [Human herpesvirus 7]...   123  1e-28
        gi|9628340|ref|NP_042931.1|  U38, DNA polymerase [Human herpesvir...   123  1e-28
        gi|9630008|ref|NP_046226.1|  DNA polymerase [Orgyia pseudotsugata...   121  3e-28
        gi|9628011|ref|NP_042605.1|  DNA polymerase replicative subunit [...   121  5e-28
        gi|6325090|ref|NP_015158.1|  DNA polymerase zeta subunit; Rev3p [...   120  9e-28
        gi|9626767|ref|NP_041039.1|  DNA polymerase [Equine herpesvirus 1...   118  2e-27
        gi|8393995|ref|NP_058633.1|  polymerase (DNA-directed), alpha; po...   118  4e-27
        gi|6015013|sp|O89042|DPOA_RAT  DNA POLYMERASE ALPHA CATALYTIC SUB...   117  7e-27
        gi|6015020|sp|Q90162|DPOL_NPVCF  DNA POLYMERASE >gi|747638|gp|AAC...   117  7e-27
        gi|6679409|ref|NP_032918.1|  DNA polymerase alpha 1, 180 kDa [Mus...   115  2e-26
        gi|6015012|sp|Q27152|DPOA_OXYTR  DNA POLYMERASE ALPHA CATALYTIC S...   114  4e-26
        gi|6919874|sp|Q61493|DPOZ_MOUSE  DNA POLYMERASE ZETA CATALYTIC SU...   114  4e-26
        gi|6015011|sp|Q94636|DPOA_OXYNO  DNA POLYMERASE ALPHA CATALYTIC S...   114  5e-26
        gi|9625903|ref|NP_040151.1|  ORF28 (AA1-1194) [Human herpesvirus ...   109  2e-24
        gi|9625739|ref|NP_039988.1|  DNA polymerase (8) [human herpesviru...   108  4e-24
        gi|6015010|sp|O48653|DPOA_ORYSA  DNA POLYMERASE ALPHA CATALYTIC S...   104  5e-23
```

Appendix 2

Part of the multiple alignment to establish the phylogenetic tree, in this sequences is only showed only the most conserved areas, common for all the DNA polymerases.





Appendix 3. Species used for the analysis in the type II DNA polymerases' phylogenetic tree in Figure 3.

Herpesviridae type II DNA polymerases

DPOL CHVP1	*Paramecium bursaria* chlorella virus 1 (PBCV-1)
DPOL CHVN2	*Chlorella* virus NY-2A (CV-NY2A).
DPOL GPCMV	Guinea pig cytomegalovirus (strain 22122 / ATCC VR682)
DPOL HCMVA	Human cytomegalovirus (strain AD169)
DPOL HSV1A	Herpes simplex virus (type 1 / strain Angelotti)
DPOL HSV1K	Herpes simplex virus (type 1 / strain KOS)
DPOL HSV1S	Herpes simplex virus (type 1 / strain SC16)
DPOL HSV21	Herpes simplex virus (type 2 / strain 186)
DPOL HSV6U	Human herpesvirus (type 6 / strain Uganda-1102) (HHV6)
DPOL HSV7J	Human herpesvirus (type 7 / strain JI) (HHV7)
DPOL HSVE2	Equine herpesvirus type 2 (strain 86/87) (EHV-2)
DPOL HSVEB	Equine herpesvirus type 1 (strain Ab4p) (EHV-1)
DPOL HSVT1	Herpesvirus tupaia (Strain 1) (THV-1)
DPOL HSVT2	Herpesvirus tupaia (Strain 2) (THV-2)
DPOL MCMVS	Murine cytomegalovirus (strain Smith)

Type δ DNA polymerases (Eucarya)

DPOD ARATH	*Arabidopsis thaliana* (Mouse-ear cress)
DPOD BOVIN	*Bos taurus* (Bovine)
DPOD CAEEL	*Caenorhabditis elegans*
DPOD HUMAN	*Homo sapiens* (Human)
DPOD MESAU	*Mesocricetus auratus* (Golden hamster)

DPOD MOUSE *Mus musculus* (Mouse)

DPOD ORYSA *Oryza sativa* (Rice)

DPOD PLAFK *Plasmodium falciparum* (isolate K1 / Thailand)

DPOD YEAST *Saccharomyces cerevisiae* (Baker's yeast)

DPOL NPVCF *Choristoneura fumiferana* nuclear polyhedrosis virus (CfMNPV)

DPOL NPVOP *Orgyia pseudotsugata* multicapsid polyhedrosis virus (OpMNPV)

Crenarchaeota Type II DNA polymerases (Group II)

DPO1 AERPE *Aeropyrum pernix*

DPO1 SULAC (B1) *Sulfolobus acidocaldarius*

DPO1 SULOH *Sulfurisphaera ohwakuensis*

Creanarchaeaota and Euryarchaeota Type II DNA polymerases (Group I)

1D5A *Desulfurococcus* sp. Tok (C)

DPO2 AERPE *Aeropyrum pernix* (C)

DPOL ARCFU *Archaeoglobus fulgidus* (E)

DPOL METJA *Methanococcus jannaschii* (E)

DPOL METVO *Methanococcus voltae* (E)

DPOL PYRAB *Pyrococcus abyssi* (E)

DPOL PYRSD *Pyrococcus* sp. (strain GB-D) (E)

DPOL THEFM *Thermococcus fumicolans* (E)

DPOL THEGO *Thermococcus gorgonarius* (E)

Bacterial Type II DNA polymerases (Proteobacteria)

1KLN

pac PA1886 *Pseudomonas aeruginosa*

vch VC1212 *Vibrio cholerae*

Type α DNA polymerase (Eucarya)

DPOA DROME *Drosophila melanogaster*

DPOA HUMAN *Homo sapiens*

DPOA LEIDO *Leishmania donovani*

DPOA MOUSE *Mus musculus*

DPOA ORYSA *Oryza sativa*

DPOA OXYNO *Oxytricha nova*

DPOA OXYTR *Oxytricha trifallax*

DPOA SCHPO *Schizosaccharomyces pombe* (Fission yeast)

DPOA TRYBB *Trypanosoma brucei brucei*

DPOA YEAST *Saccharomyces cerevisiae*