

BREVE INTRODUCCION A `TeXshade`, UN PAQUETE `TeX/LaTeX` PARA UNA EDICION RAPIDA, EFICIENTE Y FLEXIBLE DE ALINEAMIENTOS DE SECUENCIAS NUCLEICAS Y PEPTIDICAS

Mikel Egaña Aranguren (pik)

Diciembre-2003

Resumen

En este breve artículo se analiza por encima el paquete `TeXshade`, que se utiliza para editar alineamientos de secuencias, pudiendo colorear zonas o residuos, poner etiquetas, y un largo etc. Para ello se dan unas instrucciones básicas de instalación y se repasan por encima algunos comandos y opciones. También se facilitan lugares donde encontrar información mucho más extensa.

Índice

1. Introducción e información complementaria	2
1.1. ¿Qué es <code>TeXshade</code> ?	2
1.2. ¿De qué trata este artículo?	2
1.3. ¿Hacen falta conocimientos previos?	2
1.4. El archivo <code>ejemplos.pdf</code>	2
1.5. Actualizaciones	3
1.6. Contacto y agradecimientos	3
1.7. Licencia	3
2. Instalación: pasos generales y soluciones <i>ad hoc</i>	3
3. Uso y disfrute: estructura y funcionamiento	4
3.1. Estructura general	4
3.2. Archivos aceptados	5
3.3. Algunos comandos y opciones	6
3.3.1. Tipo de secuencia	6
3.3.2. Los colorines	6
3.3.3. Incluir información sobre estructuras secundarias	8
3.3.4. Añadir etiquetas	8

1. Introducción e información complementaria

1.1. ¿Qué es TeXshade?

TeXshade es un paquete TeX/L^AT_EX que sirve para editar, anotar y añadir características de todo tipo a alineamientos de secuencias de nucleótidos y aminoácidos, utilizando archivos de salida de los programas usuales para este tipo de procesos (por ejemplo ClustalX [1]), sin modificarlos. La edición de alineamientos es un proceso largo y laborioso que de hacerse con un procesador normal de gráficos, puede hacer perder un tiempo precioso al usuario. La ventaja de TeXshade reside en que ofrece una manera mucho más rápida de editar los alineamientos, manipulándolos directamente en el archivo `.tex` en el que estemos trabajando, mediante una serie de comandos que describiremos en este artículo.

1.2. ¿De qué trata este artículo?

Este artículo es simplemente una breve introducción a TeXshade *para salir del paso*; hace un par de semanas me encontré en una situación en la que tenía que entregar un artículo con un alineamiento, no sabía cómo hacerlo y descubrí TeXshade, que me permitió hacerlo en un lapso razonable de tiempo dejándome tiempo libre para el artículo en sí. Por esta razón te recomiendo que leas el artículo original aparecido en *Bioinformatics* [2] y la extensa y detallada documentación de la página web oficial [3] para obtener un idea real de todo el potencial de este impresionante paquete.

1.3. ¿Hacen falta conocimientos previos?

Evidentemente, hacen falta unos conocimientos básicos de TeX o L^AT_EX [4, 5, 6] para sacar algún provecho de este artículo. Espero que tengas unos cimientos en forma de cono en casa. El entorno en el que trabajo es GNU/Linux, así que este artículo probablemente servirá para cualquier distribución y otros sistemas tipo UNIX, aunque desconozco su validez para otros entornos y versiones de TeX.

1.4. El archivo ejemplos.pdf

Desgraciadamente, hay un problema con el paquete `babel` que sólo está solucionado para el alemán (lengua nativa del desarrollador del paquete): el entorno TeXshade no funciona correctamente en textos que no sean inglés o alemán, de modo que en este caso he puesto los ejemplos en un archivo aparte (`ejemplos.pdf`) que está configurado como inglés.

1.5. Actualizaciones

La versión más reciente de este documento puede encontrarse en:
<http://www.sindominio.net/~pik/texshade.es.pdf> (El texto)
<http://www.sindominio.net/~pik/ejemplos.pdf> (Los ejemplos)

1.6. Contacto y agradecimientos

Para cualquier duda o crítica puedes contactar conmigo en [pik\[at\]sindominio\[dot\]net](mailto:pik[at]sindominio[dot]net) (clave GnuPG disponible en: <http://www.sindominio.net/~pik>). Me gustaría agradecer públicamente al autor de **T_EXshade**, Eric Beitz, la ayuda y la atención prestadas. Me pareció una persona encantadora.

1.7. Licencia

Este texto tiene una licencia **Creative Commons Attribution-ShareAlike**. Por lo tanto, se permite la copia, distribución, muestra pública o privada, uso comercial y creación de trabajos derivados, bajo la siguientes condiciones: los trabajos derivados o basados en este texto deben tener también esta licencia y se debe citar al autor (Mikel Egaña Aranguren -pik-). En cualquier distribución o uso posterior de este texto se deben mantener dicha licencia y esta nota. El texto completo de dicha licencia puede encontrarse en:

<http://creativecommons.org/licenses/by-sa/1.0/legalcode>

2. Instalación: pasos generales y soluciones *ad hoc*

En este apartado vamos a ver los pasos generales para una instalación rápida y los problemas con los que me encontré y cómo los solucioné. Probablemente tendrás que hacerlo todo como root.

Lo primero que hay que hacer es bajarse los archivos de la última versión (<http://homepages.uni-tuebingen.de/beitz/dltse.html>): `texshade.dtx` (el manual), `texshade.ins` (el archivo de instalación), `texshade.txt` (las instrucciones de instalación) y `tsfaq.tex` (las FAQ).

Con todos los archivos en un directorio, hacemos:

```
yo@mysshell: latex texshade.ins
```

Esto genera unos cuantos archivos: la mayoría son ejemplos para el manual (que compilaremos en un momento) y hay uno que nos interesa de verdad: `texshade.sty` es el paquete `texshade` en sí (ver más adelante).

Lo siguiente es compilar el manual en sí: aunque no uses el manual

(???), es importante compilarlo para probar la memoria de L^AT_EX: T_EXshad_e requiere más memoria de lo normal, y si al final de la compilación te aparece un mensaje en plan "Tex capacity exceeded" o algo así es que no tienes suficiente memoria. Para cambiar la memoria, cambia los campos `main_memory` a 5000000 (no debe exceder los 8000000 en esa sección, así que mira también los otros dos campos) y `stack_size` a 5000 en el archivo `/var/lib/texmf/web2c/texmf.cnf` (En Debian; puede que en otras distribuciones y UNIX esté en otro sitio, en todo caso el archivo fijo que es `texmf.cnf`). Después de cambiar estos campos:

```
yo@myshell: texconfig init
```

La memoria ya está cambiada, vuelve a compilar el manual. Si el manual compila correctamente y cuando lo ves con `xdvi` no aparecen los alineamientos, es problema de `xdvi`, así que haz:

```
yo@myshell: dvips texshade.dvi
```

Esto genera un archivo PostScript (`texshade.ps`) que puedes ver fácilmente con `GV`; seguro que los alineamientos y el logo de T_EXshad_e (T_EXshad_e) aparecen perfectos.

Ahora sólo falta instalar el paquete en sí: en la FAQ de CervanTeX [6] explican la manera idónea de instalar paquetes en detalle, pero nosotros nos lo vamos a hacer sucio y rápido: copia el archivo `texshade.sty` a `/usr/share/texmf/tex/latex/` y haz:

```
yo@myshell: texshash
```

Y santas pascuas. Cuando digo sucio y rápido es por algo.

3. Uso y disfrute: estructura y funcionamiento

3.1. Estructura general

T_EXshad_e es un paquete que provee un *entorno* en el que trabajar. Así que lo primero que hay que hacer es añadir el paquete a nuestro archivo `.tex` al principio, así:

```
\usepackage{texshade}
```

El uso general del entorno es así:

```
\begin{texshade}[archivo_parametros]{archivo_alineamiento}
Otros comandos y opciones
\end{texshade}
```

Entre `\begin{texshade}` y `\end{texshade}` van todos los comandos que queramos aplicar a nuestro alineamiento (`archivo_alineamiento`) para que resalten las características que queramos. Esta parte se explica en la sección 3.3. El archivo con los parámetros (`archivo_parametros`)

es opcional y en él podemos poner comandos. Esta opción es útil si queremos utilizar los mismos comandos para diferentes alineamientos: podemos invocar dichos comandos mediante el archivo `archivo_parametros` y así no tenemos que escribirlos entre `\begin{texshade}` y `\end{texshade}`.

3.2. Archivos aceptados

`TeXshade` acepta tres tipos de archivos: `.MSF`, `.ALN` y `.FASTA`. El autor del paquete recomienda el uso del formato `.MSF`, pero `.ALN` es más común ya que es lo que produce el programa Clustal. El archivo `.ALN` es realmente sencillo y editable antes de incluirlo en el entorno, por ejemplo para cambiar un nombre. El aspecto general es así:

```

CLUSTAL X (1.82) multiple sequence alignment

SRC_HUMAN      -----GSNKS KP-KDASQRRRSLEPAENVHGAGG--GAFP
YES_HUMAN      -----MGC IKS KENKSPA I KYR PENTPEPVSTSVSHYGAEP
FGR_HUMAN      -----MGC VFCKKLEPVATAKEDAGLEGDFRSYGA---ADH
ANALYSED       -----MGC VFCKKLEPAS--KEDVGLLEGDFRSQTA---EER
FYN_HUMAN      -----GC VQCKDKEATK----L TEERD GSLNQSS---GYR
YRK_CHICK      -----GC VHCKEKISGK----G QGSGGTGTPAHP---PSQ
HCK_HUMAN      GGRSSCEDPG CPRDEERAPRMGCMKSKFLQVGGNTFSK-----
LYN_HUMAN      -----GC IKS-----K GKDSLSDDGVDLK-----
LCK_HUMAN      -----GC GCS-----SHPEDDWMEN-----
BLK_HUMAN      -----GLVSS-----K KPDKEKPIKE-----
                *

SRC_HUMAN      KLFGGFNSSD TVTSPQRAGPLAGGVTTFVALYDYESRTETDLSF
YES_HUMAN      TPFGGASSSFSVVPSSYPAGLTGGVTIFVALYDYEARTTEDLSF
FGR_HUMAN      FLD SG-----TIRGVSGIGVTLFIALYDYEARTEDDLTF
ANALYSED       FLNTG-----NMRSISGTGVTIFVALYDYEARTGDDLTF
FYN_HUMAN      TVFGG--VNSSHTGTLRTRGGTGVTLFVALYDYEARTEDDLSF
YRK_CHICK      VPFSGPGFYPCNTLQAHSSITGGGVTLFIALYDYEARTEDDLSF
HCK_HUMAN      TPG-----I REAGS--ED IIVVALYDYEAIHHEDLSF
LYN_HUMAN      LPGQR-----FQTKDPEE--QGDIVVALYPYDGIHPDDLFSF
LCK_HUMAN      PLVTY-----EGSNPPASPLQDNLVIALHSYEPSHDGLGPF
BLK_HUMAN      T P P P P-----DEHLDEDK----HFVVALYDYTAMNDRDLQM
                . : ** : *      ** :

```

El aspecto de un archivo `.FASTA` también es muy sencillo y su uso está bastante extendido:

```

>BLK_HUMAN
GLVSSKKPDKEKPIKEKDKGQWSPLKVS AQDKDAPLPLPVFNHLTPPPPDEHLDEDKH
FVVALYDYTAMNDRDLQMLKGEKLQVLKGTGDWWLARSLVTGREGYVPSNFVARVESLEM
ERWFFRSQGRKEAERQLLAPINKAGSFLIRESETNKGAFSLSVKDVTTQGELIKHYKIRC
LDEGGYIISPRITFP SLQALVQHYSKKG DGLCQRLTLPCVRPAPQNPWAQDEWEIPRQSL
RLVRKLGSGGFGEVWVMGYKNNMKVAIKTLKEGTMSPEAFLEGEANMMKALQHERLVRLYA
VVTKEPIYIVTEYMARGCLLDFLKTDEGSRLSLPRLIDMSAQIAEGMAYIERMNSIHRDL
RAANILVSEALCCKIADFG LARIIDSEYTAQEGAKFP IKWTAPEAIHFGVFTIKADVWSF
GVLLMEVVTYGRVPYPGMSNPEVIRNLERGYRMPRPDTCPPELYRGVIAECWRSRPEERP
TFEFLQSVLEDFYTATERQYELQP
>FGR_HUMAN
MGCVFCKKLEPVATAKEDAGLEGDFRSYGAADHYGPDPTKARPASSFAHIPNYSNFSSQA
INPGFLDSGTIRGVSGIGVTLFIALYDYEARTEDDLTFTKGEKFHILNTEGDWWEARSL
SSGKTGCIPSNYVAPVDSIQAEWYFGKIGRKAERQLLSPGNPQGAFLIRESETTKGAY
SLSIRDWDQTRGDHVHKKIRKLDMGYYITTRVQFNSVQELVQHMEVNDGLCNLLIAP

```

```

CTIMKPQTLGLAKDAWEISRSSITLERRLGTGCFGDVWLGTWNGSTKVAVKTLKPGTMS
KAFLEEAQVMKLLRHDKLVQLYAVVSEPIYIVTEFMCHGSLDLFLKNPEGQDLRLPQLV
DMAAQVAEGMAYMERMNYIHRDLRAANILVGERLACKIADFLARLIKDDYVNPQQGSKF
PIKWTAPEAALFGRFTIKSDVWSFGILLTELITKGRIPYPMNKRVLQVEQGYHMPCP
PGCPASLYEAMEQTWRLDPEERPTFEYLSFLEDYFTSAEPQYQPGDQT

```

3.3. Algunos comandos y opciones

En esta sección vamos a *sobrevolar* algunas de las opciones más importantes que ofrece `TeXshade`, reiterar que el manual original ofrece todas las opciones y su uso. A partir de aquí se sobreentiende que todas las instrucciones van entre `\begin{texshade}` y `\end{texshade}`.

3.3.1. Tipo de secuencia

Antes de usar cualquier opción hay que decirle a `TeXshade` qué tipo de secuencia estamos usando, nucleótidos (N) o péptidos (P):

```
\seqtype{P}
```

3.3.2. Los colorines

La primera y más evidente herramienta que ofrece `TeXshade` es el sombreado: con él podemos colorear los residuos o bases¹ que queramos utilizando el criterio que más nos convenga. El sombreado se lleva a cabo mediante los *shadingmodes*, que no son más que criterios preestablecidos de qué residuos habrá que colorear y con qué color. Por supuesto los *shadingmodes* son completamente configurables e incluso podemos colorear los residuos de uno en uno con el criterio que más nos plazca sin usar un *shadingmode*.

Los *shadingmodes* tienen la siguiente estructura:

```
\shadingmode[option]{mode}
```

Donde *mode* es el *shadingmode* en sí y *option* es alguna opción específica del *shadingmode* en cuestión. Estos son los *shadingmodes* disponibles:

identical: los residuos idénticos en una posición concreta se colorean de un color. Si añadimos la opción `allmatchspecial` las posiciones en las que todos los residuos son idénticos se colorean diferente:

```
\shadingmode[allmatchspecial]{identical}
```

similar: se colorean los residuos *similares*. Aquí también se puede usar la opción `allmatchspecial`:

```
\shadingmode[allmatchspecial]{similar}
```

¹De aquí en adelante sólo se mencionarán los residuos y los ejemplos se referirán a proteínas.

diverse: utilizando una de las secuencias como consenso, enseña las *diferencias*. La secuencia consenso se define en el campo `option`²:

```
\shadingmode[3]{diverse}
```

functional: resalta alguna característica funcional de los residuos. La característica que queramos resaltar se elige mediante la entrada `option`:

- **charge:** colorea los residuos si tienen carga positiva (básicos) o negativa (ácidos).
- **hydropathy:** discrimina entre residuos ácidos, básicos, polares sin carga e hidrofóbicos no-polares.
- **structure:** enseña la posición de cada residuo en la estructura terciaria (externo, interno, ambivalente).
- **chemical:** los residuos se colorean teniendo en cuenta las propiedades químicas de sus grupos funcionales (**acidic**, **aliphatic**, **amide**, **aromatic**, **basic**, **hydroxyl**, **imino**, **sulphur**).
- **standard area:** los residuos se colorean según las diferencias en el espacio que ocupan las cadenas laterales.
- **accessible area:** los residuos se colorean según la accesibilidad que tengan para los disolventes.

Por ejemplo si queremos colorear los residuos según su carga sería así:

```
\shadingmode[charge]{functional}
```

Una opción muy útil es `fingerpint`, que permite comprimir toda la secuencia en una línea, convirtiendo los residuos a líneas verticales. Esto permite tener una visión general de toda la secuencia. Por ejemplo si queremos ver la carga de todos los residuos en un sola línea:

```
\shadingmode[charge]{functional}
\fingerpint{1000}
```

El número (en este caso 1000) representa cuántos residuos queremos por línea (en que medida queremos comprimir la secuencia).

Entre otros refinamientos, se pueden definir grupos funcionales de residuos con el criterio que se quiera; antes de definir el grupo se define un shading mode `functional` sin el campo `option`:

```
\shadingmode{functional}
```

Luego se define el grupo funcional con el comando `\funcgroup`:

```
\funcgroup{descr}{residues}{res. color}{shad. color}
{case}{style}
```

²Las secuencias se cuentan empezando por arriba y los residuos se cuentan de izquierda a derecha sin tener en cuenta los huecos (como era de esperar): en la secuencia `RRRR_RR__K` el residuo `K` sería el número 7. Un programa muy útil para ver las posiciones absolutas y relativas en un alineamiento es `CINEMA5` [7].

Donde `descr` es un descripción del grupo, `residues` son los residuos que forman el grupo, `res. color` es el color con el que queremos que aparezca la letra del residuo, `shad. color` es el color con el que queremos colorear el residuo en sí (el fondo) y `case` y `style` se refieren al estilo. Por ejemplo:

```
\funcgroup{Acidos}{DE}{white}{red}{upper}{up}
```

También se pueden colorear residuos individualmente o en grupos, con el comando `shaderegion`:

```
\shaderegion{1}{257..279,355..355}{white}{Gray50}
```

En este caso queremos sombrear los residuos que van del 257 al 279 y el residuo 355 de la secuencia 1, con gris y que las letras aparezcan en blanco (blanco sobre gris).

3.3.3. Incluir información sobre estructuras secundarias

En `TeXshade` se pueden incluir predicciones de estructuras secundarias en los siguientes formatos: DSSP, STRIDE y PHD. Se hace así:

```
\includeDSSP{num_sec}{archivo}
```

Donde `num_sec` se refiere al número de la secuencia a la que hace referencia la predicción de la estructura secundaria, y `archivo` hace referencia al archivo donde se encuentra la predicción en sí (el archivo que nos da el algoritmo de predicción). Para los demás casos la sintaxis es similar, sólo cambia el `\include`:

```
\includeSTRIDE{num_sec}{archivo}
```

```
\includePHDsec{num_sec}{archivo} (para la estructura secundaria)
```

```
\includePHDtopo{num_sec}{archivo} (para la topología)
```

3.3.4. Añadir etiquetas

Muchas veces queremos llamar la atención del lector hacia una región concreta de un alineamiento, por ejemplo por que hay un motif importante o un residuo conservado, porque queremos delimitar un dominio, o porque sí. Para hacerlo se usa el comando `\feature`, que tiene los siguientes campos:

```
\feature{position}{num}{residues}{label style}{text}
```

`position` es el lugar en el que queremos la etiqueta (arriba o abajo, `top` o `bottom`), `num` es el número de la secuencia a la que se refiere la etiqueta (1, 2,...), `residues` son los residuos que comprenden la etiqueta en sí (la posición relativa de los residuos que queremos resaltar, en el formato ya descrito: 234..256,567..567,8900..8923), `label style` es el tipo de etiqueta que queremos y `text` es el texto de la etiqueta (por ejemplo *Dominio SH3*, *Y conservado*, *Loop*, etc.). Básicamente, el tipo de etiqueta se describe con combinaciones de tres caracteres:

- Izquierda: `--<' , |`

- Mitad: -=
- Derecha: ==>', |

Por ejemplo si queremos un flecha que apunta a la derecha y tiene el extremo izquierdo hacia arriba: '→. También se pueden utilizar otros comandos como `brace` para que la etiqueta sea una llave, etc.

Referencias

- [1] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., 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.
- [2] Beitz, E. (2000). **TeXshade: shading and labelling of multiple sequence alignments using L^AT_EX 2_ε.** *Bioinformatics.* 16, 135-139.
- [3] <http://homepages.uni-tuebingen.de/beitz/txe.html>
- [4] **¿Alergia al L^AT_EX?** (F. Javier Pueyo Mena):
<http://sindominio.net/ayuda/latex>
- [5] **The not so short introduction to L^AT_EX.** (Tobias Oetiker, Hubert Partl, Irene Hyna and Elisabeth Schlegl). Esta es la guía que uso yo, es buenísima:
<http://people.ee.ethz.ch/~oetiker/lshort/lshort.pdf>
- [6] La web de CernvaTeX está llena de recursos:
<http://filemon.mecanica.upm.es/CervanTeX/index.php>
- [7] Parry-Smith, D. J., Payne, A. W., Michie, A. D., Attwood, T. K. (1998). **CINEMA: a novel colour interactive editor for multiple alignments.** *Gene.* 221, 57-63:
<http://aig.cs.man.ac.uk/utopia/>