

ORF Finder (Open Reading Frame Finder) – כלי לחיזוי מסגרת קריאה פתוחה, כלומר, כלי לחיזוי רצף הנוקלאוטידים המקודד לחלבון.



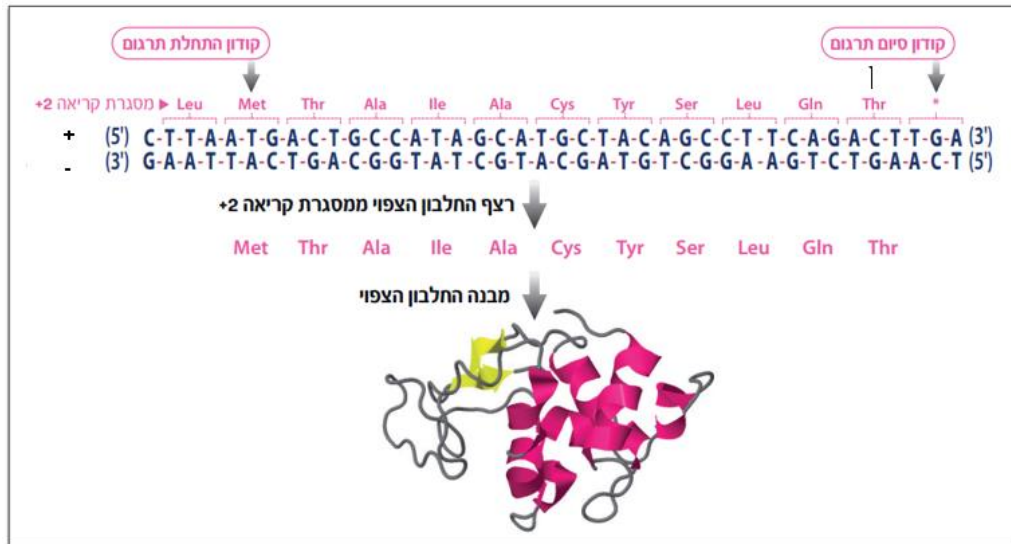
ORF Finder (Open Reading Frame Finder) – כלי לחיזוי מסגרת הקריאה הפתוחה, כלומר, רצף הנוקלאוטידים המקודד לחלבון. כל מקטע DNA מכיל 6 מסגרות קריאה, 3 בכל גדיל, הנבדלות זו מזו בעמדה שבה מתחיל הקודון הראשון. הכלי ORF Finder מקבל כקלט רצף נוקלאוטידים ומזהה בכל אחת ממסגרות הקריאה רצפים מקודדים שניתן להגדירם כמסגרות קריאה פתוחות (ORFs). בהתחשב ובאורך של כל מסגרת קריאה פתוחה, ניתן לזהות את מסגרת הקריאה הפתוחה הסבירה ביותר לקודד לחלבון, כלומר לחזות את הרצף המקודד (CDS) לרצף חומצות האמינו של החלבון שעתיד להיווצר. בעזרת הקישוריות לכלי ה-BLAST, הכלי מאפשר גם לחפש גנים או חלבונים ידועים בעלי רצף דומה לרצף הנחקר, וכך נוכל ללמוד על זהות הרצף הנחקר ועל תפקודו.



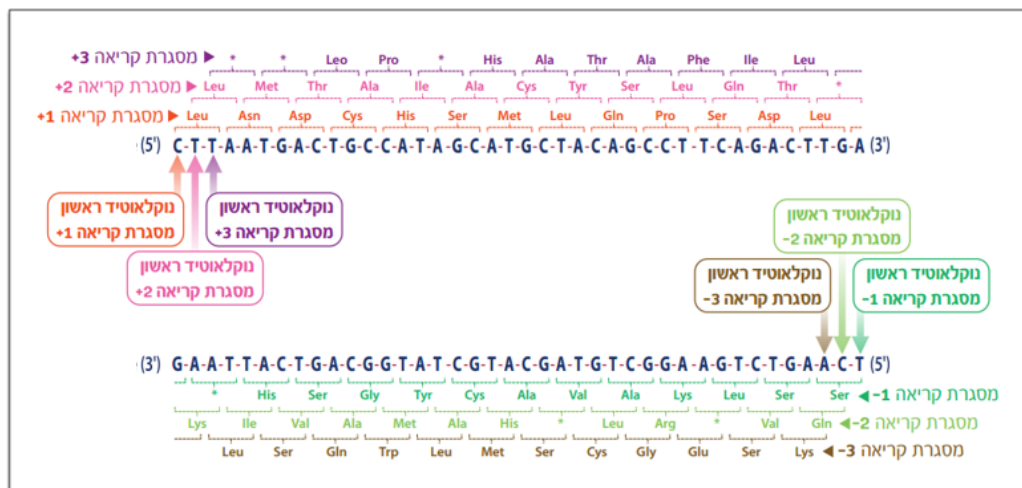
ORF Finder (Open Reading Frame Finder) – גישה לכלי בכתובת:
<https://www.ncbi.nlm.nih.gov/orffinder/>



ברוכים הבאים לסיור המודרך של הכלי ORF Finder, כלי המשמש לחיזוי רצפי נוקלאוטידים המקודדים לחלבונים. מקטע DNA המקודד לחלבון עובר תחילה תעוק למולקולת RNA-שליח אשר מתורגמת בהמשך לחלבון. אחד האתגרים הניצבים בפני החוקרים הוא לנבא אילו מקטעים בגנום הם גנים המקודדים לחלבונים, וכן לחזות על סמך רצף הנוקלאוטידים את רצף חומצות האמינו של החלבון שעתיד להיווצר.



כדי לעשות זאת, יש לאתר את מסגרת הקריאה הפתוחה מבין 6 מסגרות הקריאה בכל מקטע DNA. כל אחד משני גדילי ה-DNA מכיל 3 מסגרות קריאה: אלו מהגדיל האחד מסומנות ב-"+" (פלוס), ואלו מהגדיל השני מסומנות ב-"-" (מינוס). בכל גדיל מסגרות הקריאה נבדלות בעמדה שבה מתחיל הקודון הראשון, ומסומנות במספר "1", "2" ו-"3". מסגרת קריאה פתוחה נקראת Open Reading Frame או בקיצור ORF. כדי למצוא את מסגרת הקריאה הפתוחה המקודדת לחלבון, יש לחפש רצף נוקלאוטידים ארוך ככל האפשר, שתחילתו בקודון התחלת תרגום, המשכו בקודונים המקודדים לחומצות אמינו, וסופו באחד מקודוני עצירת התרגום.



ממשק הכלי ORF Finder

הכלי מקבל כקלט רצף נוקלאוטידים. כיוון שהכלי מתייחס לרצף הנוקלאוטידים כולו כאל רצף מקודד אפשרי, מקובל לנתח באמצעותו רצפים חסרי אינטרונים, כמו רצפי גנים מיצורים פרוקריוטים ורצפים של DNA משלים מיצורים אאוקריוטים. את הרצף ניתן להזניק בחלון הרצפים או לציין את קוד זיהוי הרשומה Accession number בחלון מתאים. לשם הפעלת הסריקה יש ללחוץ על לחצן Submit. כעת הכלי סורק את הרצף שהוזן ומחפש בכל אחת משש מסגרות הקריאה, מסגרות קריאה פתוחות.

The screenshot shows the NCBI ORFfinder web interface. A yellow box with a red arrow points to the 'תיאור הכלי' (Tool Description) section, which contains the following text:

תיאור הכלי

Open Reading Frame Finder

ORF finder searches for open reading frames (ORFs) in the DNA sequence you enter. The program returns the range of each ORF, along with its protein translation. Use ORF finder to search newly sequenced DNA for potential protein encoding segments, verify predicted protein using newly developed SMART BLAST or regular BLASTP.

This web version of the ORF finder is limited to the subrange of the query sequence up to 50 kb long. Stand-alone version, which doesn't have query sequence length limitation, is available for Linux x64.

Examples (click to set values, then click Submit button):

- NC_011604 Salmonella enterica plasmid pWES-1; genetic code: 11; 'ATG' and alternative initiation codons; minimal ORF length: 300 nt
- NM_000059; genetic code: 1; start codon: 'ATG only'; minimal ORF length: 150 nt

Enter Query Sequence

Enter accession number, gi, or sequence in FASTA format:

```
ACTGGGATGTGTTCCGGCTCTCGTCTGCTTCTGCTCGAGGCTATGCTCTGGCACTGGSCAAA
GAGGAGGA
CCGCTACC
TTTTGAGT
CAGAACCT
CCTGATCA
ACCGGGTC
AATGATAC
CTCCTACG
```

חלון להזנת רצף נוקלאוטידים בתצורת FASTA או מספר סידורי של הרצף

Choose Search Parameters

Minimal ORF length (nt): 75

Genetic code: 1, Standard

ORF start codon to use:

- "ATG" only
- "ATG" and alternative initiation codons
- Any sense codon

Ignore nested ORFs:

Start Search / Clear

Submit Clear

A second yellow box with a red arrow points to the 'Submit' button, with the text: **לחצן לאיתור מסגרות קריאה פתוחות**

ניתוח דף התוצאה

התוצאות של מסגרות הקריאה הפתוחות (ORFs) מוצגות באופן גרפי ועל ידי טבלת נתונים מספריים המגדירים פרמטרים שונים של מסגרות הקריאה הפתוחות.

בחלק העליון של דף התוצאה מוצגות מסגרות קריאה פתוחות באופן גרפי כך שכל מלבן כתום מייצג מסגרת קריאה פתוחה אחת ומיקומה היחסי ברצף המוזן. בנוסף, מצוין בחלק הגרפי ליד כל מלבן המספר הסידורי של מסגרת הקריאה הפתוחה (על פי סדר הופעתה בטבלה) וכיוון קריאתה (חץ לימין מצוין קריאה של רצף הנוקלאוטידים שהוכנס כקלט לממשק הכלי, חץ לשמאל מצוין קריאה של רצף הנוקלאוטידים הנמצא על הגדיל המשלים לרצף הקלט).

בחלק המרכזי של דף התוצאה נמצאת טבלה המסדרת את מסגרות הקריאה הפתוחות כך שמסגרת הקריאה הפתוחה הראשונה היא הארוכה ביותר ועל כן משערים שהיא הסבירה ביותר להיות הרצף המקודד. עבור כל מסגרת קריאה פתוחה מופיע בטבלה המידע הבא: הגדיל עליו היא נמצאת (סימון "+") מציין את גדיל רצף הנוקלאוטידים שהוכנס כקלט לממשק הכלי, סימון "-" מציין את הגדיל של רצף הנוקלאוטידים המשלים לו), מספר מסגרת הקריאה (1, 2 או 3), מספר הנוקליאוטיד הראשון והאחרון ברצף, אורך הרצף המקודד בנוקליאוטידים (bp) ואורך החלבון המקודד בחומצות אמינו (aa) מאותה מסגרת קריאה.

The screenshot shows the Open Reading Frame Viewer interface. At the top, it displays 'Sequence' and 'ORFs found: 7'. Below this is a sequence viewer with a scale from 0 to 1,000. A yellow box labeled 'ת'אור גרפי' (Graphical representation) points to the ORF bars. Below the viewer is a table of ORFs with columns for Label, Strand, Frame, Start, Stop, and Length (bp | aa). A yellow box labeled 'מסגרת הקריאה' (Reading frame) points to the 'Frame' column. To the right of the table is a 'k-frame translation track' with a yellow box labeled 'טבלה עם פרטים על מסגרות הפתוחות הקריאה' (Table with details on open reading frames). Below the table is a 'Mark subset' section with a yellow box labeled 'המסגרת סימול' (Mark subset) pointing to the 'Mark subset' dropdown. To the left of the table is a 'Mark' section with a yellow box labeled 'רצף החלבון של מסגרת הקריאה ORF1' (Protein sequence of ORF1) pointing to the amino acid sequence. At the bottom left, there are 'SmartBLAST ORF1' and 'BLAST ORF1' buttons, and a 'BLAST Database' dropdown set to 'UniProtKB/Swiss-Prot (swissprot)'.

Label	Strand	Frame	Start	Stop	Length (bp aa)
ORF1	+	1	61	1032	972 323
ORF2	+	2	503	637	135 44
ORF5	-	1	465	334	132 43
ORF3	+	2	701	811	111 36
ORF6	-	2	719	624	96 31
ORF4	+	2	932	1024	93 30
ORF7	-	2	620	537	84 27

חשוב לציין בשלב זה שתוצאות הכלי הן רק ניבוי ראשוני של רצפים מקודדים תיאורטיים. סביר להניח שחלק ממסגרות הקריאה הפתוחות שהכלי הציג אינן מקודדות כלל לחלבון.

בתוצאת החיפוש של מסגרות הקריאה של רצף נתון, מופיע משמאל לטבלה, באופן אוטומטי על ידי התוכנה, רצף חומצות האמינו של מסגרת הקריאה הפתוחה הארוכה ביותר, המופיעה בטבלה בשורה הראשונה.

בדוגמה זו, מסגרת הקריאה הפתוחה שסימולה ORF1, נמצאת על הגדיל שהוכנס כקלט לממשק הכלי ולכן סימונה + וכן היא נמצאת במסגרת הקריאה 1 ולכן היא מסומנת כ-1+. אורכה 972 נוקלאוטידים והיא מקודדת לחלבון שאורכו 323 חומצות אמינו.

כדי לבחון את רצף חומצות האמינו הצפוי בכל אחת ממסגרות הקריאה הפתוחות יחד עם רצף הנוקלאוטידים המקודד, יש ללחוץ על סימול הרצף בטבלה ולאחר מכן על המילה "MARK" הנמצאת מעל רצף החלבון הצפוי מצד שמאל. בעקבות כך תסומן מסגרת הקריאה הפתוחה שבחרנו בכחול בטבלה

וברקע ירוק בתיאור הגרפי. לאחר מכן יש ללחוץ על הכיתוב "Display ORF as..." ובחלון שנפתח לבחור ב-"CDS translation".

בדוגמה זו ניתן לראות את סימון מסגרת הקריאה ORF2 ובחירת התצוגה "CDS translation".

ORFfinder

Open Reading Frame Viewer

Sequence

ORFs found: 7 Genetic code: 1 Start codon: 'ATG' only

1: 1..1.0K (1.0Kbp) Find: ATG

ORFfinder_6.3.95529185

Display ORF as:

- Protein sequence
- Nucleotide sequence
- CDS translation

ORF2 (44 aa) Display ORF as... Unmark

>|c1|ORF2
MCSGSRLLPCCEAPLWHLWAKRRRTSAAATLRRTMPSRPSFSSATHF

Mark subset... Marked: 1 Download marked set as Protein FASTA

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	61	1032	972 323
ORF2	+	2	503	637	135 44
ORF5	-	1	465	334	132 43
ORF3	+	2	701	811	111 36
ORF6	-	2	719	624	96 31
ORF4	+	2	932	1024	93 30
ORF7	-	2	620	537	84 27

ORF2 Marked set (1) SmartFASTA SmartBLAST best hit titles...

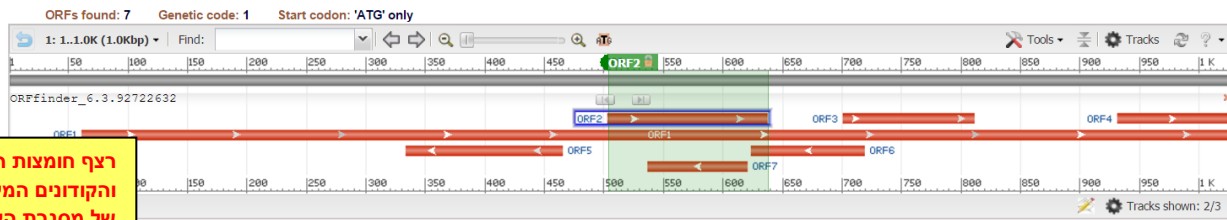
בחירת הצגת רצף הנוקלאוטידים ומתחתיו רצף חומצות האמינו של מסגרת הקריאה המבוקשת

בחירת מסגרת הקריאה המבוקשת

כתוצאה מבחירה זו ניתן לחזות מצד שמאל במסגרת הקריאה הפתוחה כרצף נוקלאוטידים מקודד ומתחתיו רצף חומצות האמינו – כל חומצת אמינו תחת הקודון המקודד לה.

Open Reading Frame Viewer

Sequence



רצף חומצות האמינו והקודונים המקודדים של מסגרת הקריאה ORF2

ORF2 (135 nt) Display ORF as... Unmark

ORF2 CDS translation

```
1 ATGTGTTCTGGGCTCTGCTGCTGCTGCTGCGAGGCTATGCTCTGG
  M C S G S R L P C C E A M L W
46 CACTGGGCAAGAGGAGGACTTCTTCAAGCCGACCTTAAAGAAGG
  H W A K R R T S S A A T L R R
91 ATGACGCCCTCTCCTCGGTCGTTCTCATCCGCTACCCATTTTAG
  M T P S P R S F S S A T H F *
```

Mark subset... Marked: 1 Download marked set as Protein FASTA

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	61	1032	972 323
ORF2	+	2	503	637	135 44
ORF5	-	1	465	334	132 43
ORF3	+	2	701	811	111 36
ORF6	-	2	719	624	96 31
ORF4	+	2	932	1024	93 30
ORF7	-	2	620	537	84 27

ORF2 Marked set (1)

SmartBLAST SmartBLAST best hit titles...

BLAST BLAST

BLAST Database:
UniProtKB/Swiss-Prot (swissprot)

בדוגמה זו, מסגרת הקריאה הפתוחה שסימולה ORF2 נמצאת במסגרת הקריאה +2, אורכה 135 בסיסים והיא מקודדת לחלבון שאורכו 44 חומצות אמינו.

מסגרת קריאה פתוחה המקודדת לחלבון חייבת להכיל קודון התחלת תרגום וקודון סיום תרגום. לעתים הכלי מציג גם רצף מקודד אפשרי שאינו מכיל קודונים אלו, מתוך הנחה שקודון התחלת תרגום נמצא במעלה הרצף שהוזן, או שקודון סיום תרגום נמצא במורד הרצף שהוזן. במקרים אלה יש לוודא כי הקודונים "החסרים" אכן קיימים בהמשך הרצף, שאם לא כן - אין לרצפים הללו משמעות.

ניתוח רצף החלבון הצפוי

לעתים לאחר שמצאנו את מסגרת הקריאה הפתוחה הסבירה ביותר לקודד לחלבון, ויש בידינו רצף חלבון צפוי, נרצה לבדוק אם אכן מסגרת קריאה זו מקודדת לחלבון אמתי, באיזה חלבון מדובר, ואם הוא דומה לחלבונים ידועים אחרים.

הכלי ORF Finder מאפשר לנו לבצע חיפוש של רצף החלבון הצפוי באמצעות כלי ה-Protein BLAST (BLASTp), שהינו מגוון לחיפוש והשוואת רצפי חלבונים. מכיוון שמסגרת הקריאה ORF1 היא הסבירה ביותר, נבצע את השוואת רצף החלבון המקודד ממסגרת קריאה פתוחה זו.

בעזרת לחיצה על "BLAST" מתחת לחלון בו מצוי רצף החלבון, נוכל לבצע השוואת הרצף המקודד ממסגרת הקריאה הפתוחה ORF1, לרצפי חלבונים ע"י הכלי BLASTp. לצורך הגבלת החיפוש ניתן לקבוע מאגר רצפי חלבונים הרצוי לנו. מאגר החלבונים swissprot נחשב מאגר אמין ולכן ישמש אותנו במקרים רבים לחיפוש זה.

Open Reading Frame Viewer

Sequence

ORFs found: 7 Genetic code: 1 Start codon: 'ATG' only

ORF1 (972 nt) Display ORF as... Unmark

ORF1 CDS translation

```

1 ATGGCTTCCACTCTCAAGGCCAATGTCCCAAGATCGATGTGTCG
  H A S T L K A H V P K I D V S
46 CCCCTGTTGGTGAACAATATGGAGGAGAAGATGAAGGTTGCCCGC
  P L F G D N H E E K M K V A R
91 GCGATTGACGCTGCCGCGGACACCGGCTCTTCTACGCGGTC
  A I D A A S R D T G F F Y A V
136 AACACGGETGTGATGTGAAGGACTCTCGAACAAAGACAGGGAG
  N H G V D V K R L S N K T R E
181 TTCCACTTTCTATCACAGGAGGAAGAGTGGGACTCGCGATT
  F H F S I T D E E K W D L A I
226 CGCGCTACAACAAGGAGCACAGGACGATCCGGCCGGGTAC
  R A Y N K E H Q D Q I R A G Y
  
```

Mark subset... Marked: 1 Download marked set as Protein FASTA

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF1	+	1	61	1032	972 323
ORF2	+	2	503	637	135 44
ORF5	-	1	465	334	132 43
ORF3	+	2	701	811	111 36
ORF6	-	2	719	624	96 31
ORF4	+	2	932	1024	93 30
ORF7	-	2	620	537	84 27

Add six-frame translation track

ORF1 SmartBLAST BLAST

Marked set (1) SmartBLAST best hit titles... BLAST

BLAST Database: UniProtKB/Swiss-Prot (swissprot)

לחצן הקישור לכלי BLAST עבור מסגרת הקריאה המסומנת

בחירת מאגר מידע לחיפוש בכלי BLAST

במידה וברצוננו להשוות רצף חלבון שמקורו במסגרת קריאה פתוחה אחרת, נסמן מסגרת קריאה פתוחה זו בטבלה ונלחץ על "BLAST".

תמונת תוצאת הקישור לכלי BLASTp

U.S. National Library of Medicine | NCBI National Center for Biotechnology Information | Sign in to NCBI

BLAST » blastp suite

Standard Protein BLAST

Enter Query Sequence

Enter accession number(s), g(i)s, or FASTA sequence(s)

Or, upload file

Choose Search Set

Database: UniProtKB/Swiss-Prot (swissprot)

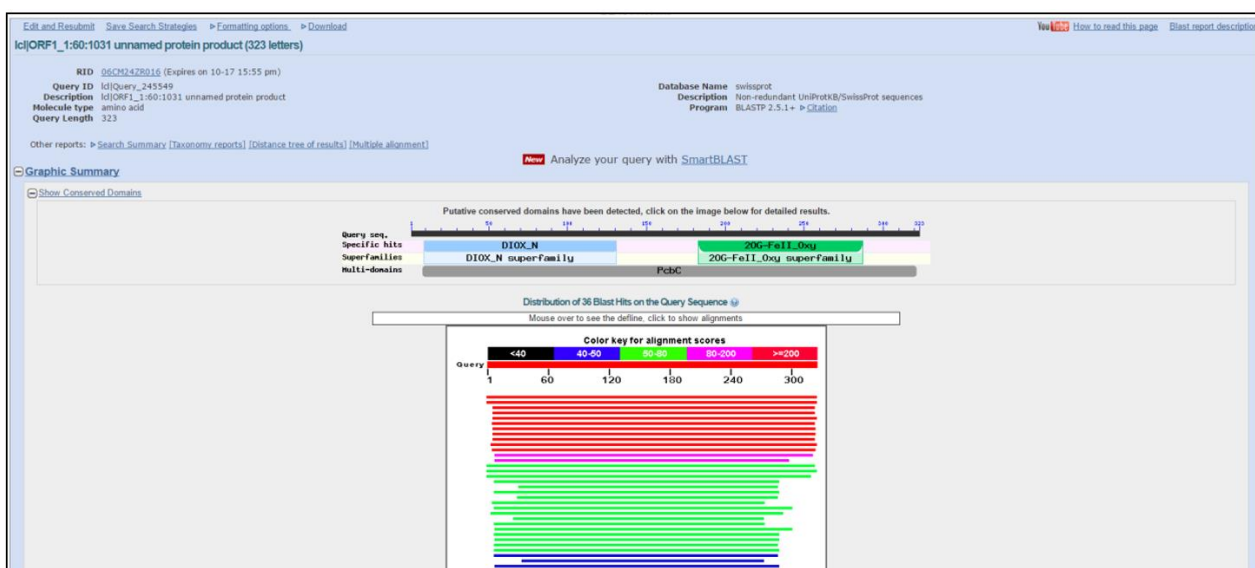
Program Selection

Algorithm: blastp (protein-protein BLAST)

BLAST Search database UniProtKB/Swiss-Prot (swissprot) using Blastp (protein-protein BLAST)

בכדי להפעיל את הכלי BLASTp ולבצע את השוואת רצפי החלבונים, יש ללחוץ על המילה BLAST

תוצאת החיפוש בכלי BLASTp



על מנת להבין כיצד ניתן לקרוא את תוצאות הכלי BLASTp יש להיכנס לארגז הכלים וללמוד על כלי זה.

לסיכום, הכלי Finder ORF מציג באופן גרפי ומילולי את מסגרות הקריאה הפתוחות (כלומר את הרצפים המקודדים התיאורטיים) במסגרות הקריאה השונות של הרצף המוזן. הכלי מציג לכל מסגרת קריאה פתוחה את מיקומה ברצף שהוזן, את הרצף המקודד ואת רצף החלבון הצפוי. כך המשתמשים יכולים לקבוע בסבירות גבוהה את מסגרת הקריאה הפתוחה שאכן מקודדת לחלבון. לרוב זו מסגרת הקריאה הפתוחה המייצגת את הרצף המקודד הארוך ביותר. הקישוריות לכלי BLAST, מאפשרת לחפש גנים או חלבונים ידועים בעלי רצף דומה לרצף הנחקר, וכך נוכל ללמוד על זהות הרצף הנחקר ועל תפקודו.

אתם מוזמנים להתנסות בכלי, ואנו מקווים שגם תיהנו. בהצלחה!