

BLASTn (nucleotide blast) - מנוע חיפוש המאתר רשומות של רצפי נוקלאוטידים הדומים לרצף נוקלאוטידים נתון (רצף שאילתה).



BLASTn – כלי ה-BLAST, Basic Local Alignment Search Tool, משמשים לחיפוש רצפים דומים. BLASTn הוא מנוע חיפוש המאתר במאגרים של רצפי נוקלאוטידים רשומות של רצפים הדומים לרצף נוקלאוטידים נתון. בעוד שבמנועי חיפוש כמו Google או Entrez השאילתה מורכבת ממילות מפתח, הרי שבמנוע החיפוש BLASTn השאילתה מורכבת מרצף נוקלאוטידים. מנוע החיפוש משווה את רצף השאילתה לרצפים במאגרי נתונים של רצפי נוקלאוטידים ומציג רשומות בעלות רצף דומה לרצף השאילתה. רצף השאילתה מוצג מול רצף הרשומה הדומה לו, עמדה כנגד עמדה, בשיטה הנקראת "העמדת רצפים" (Sequence Alignment).

להלן מספר מקרים שבהם נשתמש במנוע החיפוש זה כדי למצוא רצפי נוקלאוטידים דומים לרצף השאילתה:

- השוואת רצף גן שפעילותו אינה ידועה מול מאגר רצפי הנוקלאוטידים עשויה לזהות גנים שנחקרו בעבר ושנמצאו דומים לרצף השאילתה. דמיון ברצף יכול להעיד על דמיון אפשרי במבנה או בתפקוד תוצר הגן, כלומר, החלבון.
- השוואת רצף גן מאורגניזם אחד למאגר רצפי הנוקלאוטידים עשויה לגלות גנים הומולוגיים באורגניזמים נוספים. לדוגמה, מדען זיהה גן שגורם להתקרחות בעכבר. האם גן הומולוגי נמצא גם באדם? השוואת רצף הגן העכברי מול רצפי הגנים בגנום האנושי תיתן מענה לשאלה זו.

BLASTp (protein blast) - מנוע חיפוש המאתר רשומות של רצפי חומצות אמינו הדומים לרצף חומצות אמינו נתון (רצף שאילתה).



BLASTp – כלי ה-BLAST, Basic Local Alignment Search Tool, משמשים לחיפוש רצפים דומים. BLASTp הוא מנוע חיפוש המאתר במאגרים של רצפי חומצות אמינו רשומות של רצפים הדומים לרצף חומצות אמינו נתון. בעוד שבמנועי חיפוש כמו Google או Entrez השאילתה מורכבת ממילות מפתח, הרי שבמנוע החיפוש BLASTp השאילתה מורכבת מרצף חומצת אמינו. מנוע החיפוש משווה את רצף השאילתה לרצפים במאגרי נתונים של רצפי חומצות אמינו ומציג רשומות בעלות רצף דומה לרצף השאילתה. רצף השאילתה מוצג מול רצף הרשומה הדומה לו, עמדה כנגד עמדה, בשיטה הנקראת "העמדת רצפים" (Sequence Alignment).

להלן מספר מקרים שבהם נשתמש במנוע החיפוש זה כדי למצוא רצפי חומצות אמינו הדומים לרצף השאילתה:

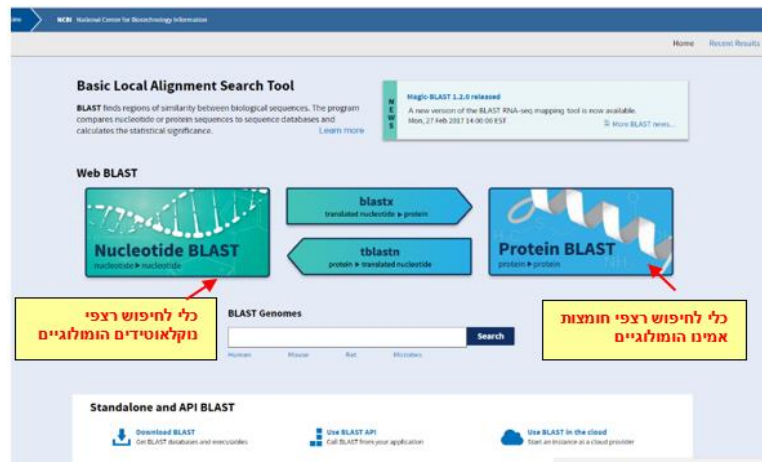
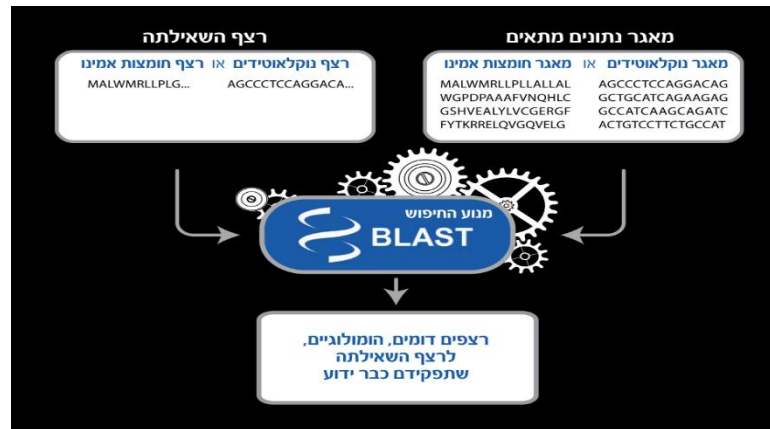
- השוואת רצף חלבון שמבנהו או פעילותו אינם ידועים, למאגר רצפי חומצות האמינו עשויה לזהות חלבונים שנחקרו בעבר ושנמצאו דומים לרצף השאילתה. דמיון ברצף יכול להעיד על דמיון אפשרי במבנה, בפעילות או בתפקוד החלבון.
- השוואת רצף חלבון מאורגניזם אחד למאגר רצפי חומצות האמינו עשויה לגלות חלבונים הומולוגיים באורגניזמים נוספים. לדוגמה, מדען זיהה חלבון בעל פעילות אנזימתית מסוימת בחיידק. האם חלבון הומולוגי נמצא גם ביונקים? השוואת רצף החלבון מהחיידק לרצפי החלבונים במאגר הנתונים תיתן מענה לשאלה זו.

גישה לכלים בכתובת: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>



ברוכים הבאים לסיוור המודרך של BLAST (Basic Local Alignment Search Tool), הכלי לשימוש במנועי חיפוש לרצפים.

כאשר אנו רוצים לבדוק אם רצף של נוקלאוטידים הוא גן המקודד לחלבון, או אם לרצף של חומצות אמינו יש תפקיד מסוים או מאיזה אורגניזם הרצף לקוח - ניעזר במנוע החיפוש BLAST להמשך מחקרנו. בעזרת מנועי החיפוש BLAST נוכל למצוא במאגרי הנתונים רצפים דומים, הומולוגיים, לרצף השאילתה שברשותנו. מאחר שהתפקיד והמבנה של רצפי התוצאה הנמצאים במאגר הנתונים, ידועים ומאורגנים בִּרְשומות, נוכל ללמוד מהן על התפקיד ועל המבנה של רצף השאילתה.



תפריט כלי ה-BLAST

זהו תפריט כלי ה-BLAST. תחילה נבחר את מנוע החיפוש הרצוי: כאשר השאילתה היא רצף נוקלאוטידים - נבחר ב-nucleotide blast; וכאשר השאילתה היא רצף חומצות אמינו - נבחר ב-protein blast. בדוגמה זו עומד לרשותנו רצף שאילתה של חומצות אמינו שאיננו יודעים את מקורו ואת החלבון שהוא מקודד. מאחר שבידינו רצף חומצות אמינו - נבחר ב-protein blast.

ממשק הכלי protein blast

זהו ממשק התוכנה protein blast. נוזן את רצף השאילתה בחלון החיפוש. כעת נבחר באיזה מאגר של רצפי חלבונים אנו מעוניינים לבצע את החיפוש. נבחר במאגר רצפי החלבונים swissprot, שהוא מאגר אמין ואיכותי. נלחץ על לחצן ה-BLAST ונמתין מספר שניות.

The screenshot shows the NCBI BLAST web interface. The main heading is "BLAST Basic Local Alignment Search Tool". The navigation menu includes "Home", "Receiv...", "Strategies", and "Help". The "blastp" tab is selected. The "Enter Query Sequence" section has a text input field with the annotation "חלון הזנת רצף השאילתה" (Query sequence input window). Below it is a "Browse..." button. The "Choose Search Set" section has a "Database" dropdown menu with the annotation "בחירת מאגר מידע" (Database selection). The dropdown menu is open, showing options like "Non-redundant protein sequences (nr)", "Reference proteins (refseq_protein)", "Swissprot protein sequences (swissprot)", "Patented protein sequences (pat)", "Protein Data Bank proteins (pdb)", and "Environmental samples (env_nr)". The "blastp" algorithm is selected in the "Program Selection" section. At the bottom, the "BLAST" button is highlighted with the annotation "לחצן BLAST" (BLAST button).

דף תוצאה: נתונים טכניים ותיאור גרפי

דף התוצאות שהתקבל מכיל נתונים כלליים של רצף השאילתה בחלקו העליון וכן ארבע לשוניות הכוללות השוואות של רצפים הדומים לשאילתה, המבוצעות בדרכים שונות תוך התמקדות בפרמטרים שונים של השוואה.

החלק העליון של דף התוצאה מתקבל מפרט נתונים טכניים לגבי שאילתת החיפוש הכוללים את: שם הרצף, סוג הרצף, אורכו, מאגר נתונים בו התבצע החיפוש וכיוצא בזה. כמו כן, ניתן לבצע בחלק זה (מצד ימין) סינון של התוצאות המתקבלות על פי מספר פרמטרים.

נתונים כלליים על שאילתת החיפוש הכוללים את מספר הזיהוי של הרצף, מאגר הנתונים, סוג הרצף ואורכו

אפשרויות לסינון הרצפים בעלי הדמיון

דף תוצאה: תיאור הרשומות – Descriptions

בלשונית הראשונה של דף התוצאה, מופיעה טבלה הכוללת את תיאור הרשומות בעלות הומולוגיה לרצף המוזן. לכל רשומה מצוין קוד זיהוי הרשומה המקושר לדף הרשומה במאגר הנתונים, שם הרשומה ותיאור קצר שלה, ציון הדמיון לרצף השאילתה, אחוז הכיסוי הדומה לאורך הרצף לשאילתה, אחוז הזהות בין הרצפים ועוד.

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

Select all 49 sequences selected GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
RecName: Full=Isopenicillin N synthase, Short=IPNS [Penicillium chrysogenum]	626	626	100%	0.0	91.95%	P08703.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Aspergillus nidulans FGSC A4]	574	574	100%	0.0	83.28%	P05326.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Acremonium chrysogenum]	530	530	97%	0.0	77.19%	P05189.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Streptomyces jamaonensis]	379	379	97%	7e-131	57.46%	P18286.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Amycolatopsis lactamdurans]	377	377	98%	2e-130	57.23%	P22744.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Flavobacterium sp. SC12154]	370	370	97%	2e-127	55.38%	P16020.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Lysobacter lactamgenus]	369	369	97%	4e-127	55.38%	Q48739.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Streptomyces clavuligenus]	369	369	97%	6e-127	55.38%	P10621.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Streptomyces microflavus]	366	366	97%	5e-126	56.15%	P12438.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Streptomyces griseus]	363	363	98%	5e-125	55.28%	Q54243.1
RecName: Full=Isopenicillin N synthase, Short=IPNS [Streptomyces cattleya]	283	283	97%	2e-93	46.37%	Q53932.1
RecName: Full=Probable iron/ascorbate oxidoreductase DOB_G0283291 [Dictyostelium discoideum]	106	106	96%	6e-25	27.30%	Q54RA4.1
RecName: Full=1-aminocyclopropane-1-carboxylate oxidase, Short=ACC oxidase, Short=Dmaco, AltName: Full=Ethylene-forming enzyme, Sho	85.9	85.9	99%	1e-17	25.71%	A68M06.1
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase DIN11, AltName: Full=Protein DARK-INDUCIBLE 11 [Arabidopsis thaliana]	72.4	72.4	100%	4e-13	25.07%	Q8H113.1
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase At3g520210 [Arabidopsis thaliana]	72.4	72.4	98%	4e-13	25.50%	Q84MP6.1
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase At3g49630 [Arabidopsis thaliana]	71.6	71.6	99%	6e-13	26.02%	F41Y03.1
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase At5g095600 [Arabidopsis thaliana]	70.1	70.1	86%	2e-12	25.26%	Q9FFF6.1
RecName: Full=Gibberellin 20 oxidase 1-B, AltName: Full=GA 20-oxidase 1-B, AltName: Full=Gibberellin C-20 oxidase 1-B, AltName: Full=TaGA	68.9	68.9	96%	6e-12	24.91%	Q04706.1
RecName: Full=Gibberellin 20 oxidase 1-D, AltName: Full=GA 20-oxidase 1-D, AltName: Full=Gibberellin C-20 oxidase 1-D, AltName: Full=Prote	66.2	66.2	86%	5e-11	24.91%	Q04705.1
RecName: Full=Gibberellin 2-beta-dioxygenase 8, AltName: Full=GA 2-oxidase 8, AltName: Full=Gibberellin 2-beta-hydroxylase 8, AltName: Full	65.5	65.5	83%	8e-11	23.83%	Q49561.2
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase JRG21, AltName: Full=Protein JASMONATE-REGULATED GENE 21 [Arabidops	63.5	63.5	86%	4e-10	23.26%	Q9LY48.1
RecName: Full=2-oxoglutarate-dependent dioxygenase imaE, AltName: Full=Imuzoquin biosynthesis cluster protein E [Aspergillus flavus NRRL3	63.5	63.5	82%	4e-10	23.31%	B8N02.1
RecName: Full=Gibberellin 2-beta-dioxygenase 1, AltName: Full=GA 2-oxidase 1, AltName: Full=Gibberellin 2-beta-hydroxylase 1, AltName: Full	63.2	63.2	78%	5e-10	25.57%	Q8LEA2.2
RecName: Full=Gibberellin 2-beta-dioxygenase, AltName: Full=GA 2-oxidase, AltName: Full=Gibberellin 2-beta-hydroxylase, AltName: Full=Gibt	62.8	62.8	78%	6e-10	26.14%	Q9XG83.1
RecName: Full=1-aminocyclopropane-1-carboxylate oxidase homolog 3 [Arabidopsis thaliana]	61.2	61.2	90%	2e-09	23.59%	Q6H1S4.1
RecName: Full=2-oxoglutarate-dependent dioxygenase htyE, AltName: Full=L-homotyrosine biosynthetic cluster protein E [Aspergillus rugulosus	59.7	59.7	59%	6e-09	28.64%	K0DZ40.1
RecName: Full=Probable 2-oxoglutarate-dependent dioxygenase At3g111800 [Arabidopsis thaliana]	59.3	59.3	86%	1e-08	24.39%	Q9SRM3.1

הרצף בעל הדמיון הגבוה ביותר

שמות הרצפים בעלי הדמיון וקישורים להעמדת הרצפים

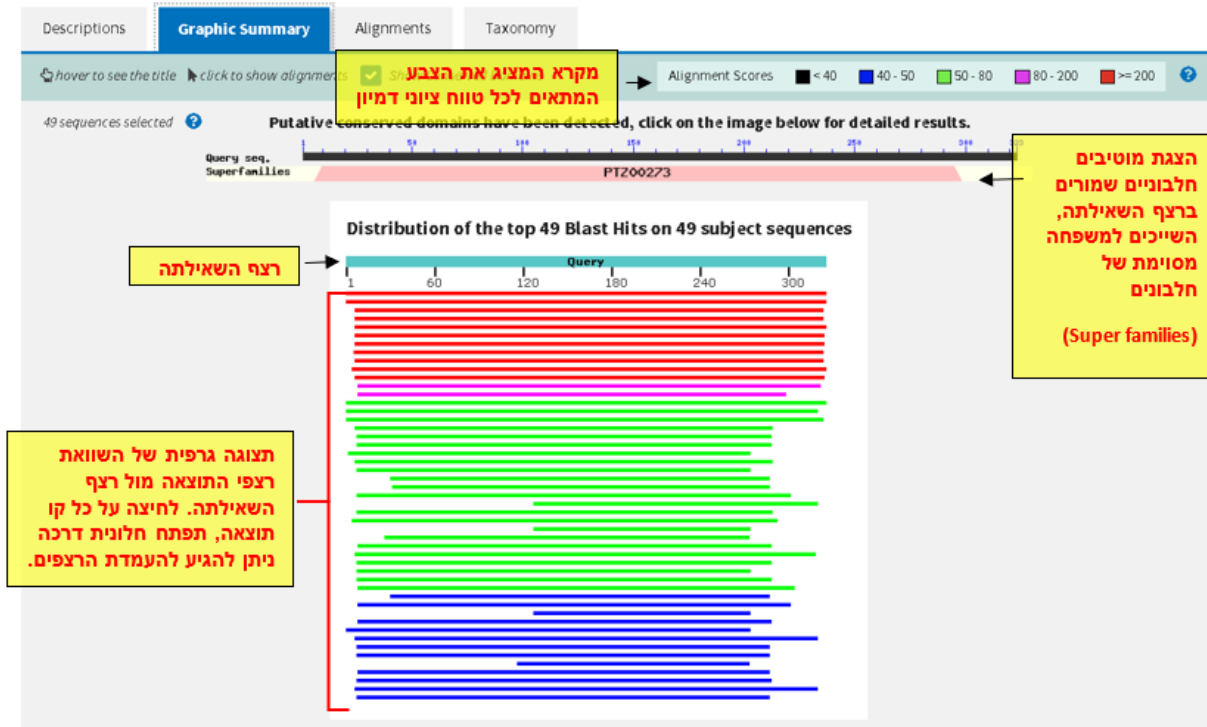
ציוני הדמיון אחוז הכיסוי הדומה לאורך הרצף אחוז הזהות קודי הדיהוי וקישור לרשומות

דף תוצאה: תיאור גרפי - Graphic Summary

הלשונית השנייה בדף התוצאות נקראת - Graphic Summary. בחלק העליון מוצג מקרא ובו ריבועים המשייכים צבע לדמיון. מתחת למקרא מוצגים על קנה המידה של רצף השאילתה, מוטיבים חלבונים שמורים ברצף השאילתה השייכים למשפחה מסוימת של חלבונים (Superfamilies). תצוגה זו מראה לאיזו משפחת חלבונים משתייך החלבון והיכן ממקומים המוטיבים היחודיים למשפחה זו בתוכו.

רצף השאילתה (query) בדף זה, מוצג באופן גרפי כקו טורקיז ובתחתיתו סרגל המתייחס לאורך הרצף והמשתנה בהתאם לאורך הרצף. מתחת לסרגל זה מוצגים הרצפים הדומים ביותר לרצף השאילתה שנמצאו במאגר הנתונים, בסדר יורד על פי מידת הדמיון לרצף השאילתה: תחילה מוצגים הרצפים הדומים ביותר ולאחר מכן הרצפים הדומים פחות. צבעו של כל קו נקבע על פי מידת הדמיון בינו לבין רצף השאילתה, מידת הדמיון מחושבת ומיוצגת בציון score. בדף התוצאות שלפנינו קיבלו רצף אחד שציונו גבוה מ-200, ולכן צבעו אדום.

ציון הדמיון של הרצפים הבאים נע בתחום שבין 80-200, ולכן צבעם ורוד וכך הלאה. מיקום הרצפים לאורך סרגל הקנה מידה של רצף השאילתה מעיד על האזור ברצף שבו נמצא הדמיון (cover). אם נגלול בעזרת העכבר מעל קו כלשהו, יופיעו פרטים מזהים אודות הרצף. לחיצה על הקו תוביל אותנו לחלונית בה יש ללחוץ על המילה Alignment לצורך הגעה להעמדת הרצפים של השאילתה ואותה רשומה. הסבר על העמדת הרצפים יפורט בהמשך.



דף תוצאה: העמדת הרצפים - Alignments

בלשונית השלישית של דף התוצאה מוצגות השוואות הרצפים (Alignments). שורת הפתיחה מציגה את פרטי הרשומה ממאגר הנתונים, לאחר מכן מפורט בשדה Score ציון הדמיון, ובשדה Identities מצוין מספר חומצות האמינו הזהות שבין רצף השאילתה ובין הרצף ממאגר הנתונים. בדוגמה זו קיימת זהות של 92%, שהן 297 חומצות אמינו זהות מתוך 323 חומצות אמינו המהוות את הרצף בהשוואה זו. בשדה ה-Positives מצוין מספר חומצות האמינו הדומות, אך לא בהכרח זהות, בין רצף השאילתה ובין הרצף ממאגר הנתונים. חומצות אמינו דומות הן חומצות אמינו בעלות תכונות כימיות ופיזיקליות דומות. לדוגמה, ליזין וארגנין הן חומצות אמינו דומות. כיוון ששתיהן טעונות במטען חיובי. בשדה Gaps מצוין מספר הרווחים שהוכנסו בהעמדת הרצפים זה מול זה, כלומר, עמדות שבהן מול חומצה אמינית ברצף אחד לא קיימת חומצה אמינית ברצף האחר. לאחר מכן מפורטת השוואת הרצפים: בשורה העליונה מוצג רצף השאילתה שסימונו, Query ובשורה התחתונה מוצג הרצף

ההומולוגי ממאגר הנתונים, וסימונו. Sbjct בשורה האמצעית מוצגת תוצאת ההשוואה בין שני הרצפים: אם בעמדה מסוימת קיימת אותה חומצה אמינית בשני הרצפים, היא תופיע בשורה האמצעית, שהיא שורת ההשוואה. אם חומצות האמינו בעמדה מסוימת דומות אך לא זהות בין שני הרצפים, יופיע הסימן + בשורת ההשוואה, ואם החומצות שונות זו מזו ולא קיים דמיון בתכונות שלהן, לא יופיע כל סימן בשורת ההשוואה. אם קיים רווח (gap) בהעמדת הרצפים, המעיד על חסר של חומצות אמינו באחד הרצפים, יסומן הרווח בסימן – במקום חומצות האמינו החסרות. בתחילת הרצפים ובסופם מוצגים מספרי העמדות בכל רצף. כך לדוגמה, ניתן לראות שהעמדה התבצעה מהחומצה האמינית הראשונה ועד לעמדה מספר 323, הן ברצף השאלתה (Query) והן ברצף שנמצא במאגר הנתונים (Sbjct).

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise

49 sequences selected

Download GenPept Graphics

RecName: Full=Isopenicillin N synthase; Short=IPNS [Penicillium chrysogenum]

Sequence ID: **P08703.1** Length: **331** Number of M

Range 1: 1 to 323 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
626 bits(1615)	0.0	Compositional matrix adjust.	297/323(92%)	308/323(95%)	0/323(0%)

מספר ואחוז חומצות אמינו זהות
מספר ואחוז חומצות אמינו דומות
מספר ואחוז מרווחים בהעמדת הרצפים

Query	Sbjct	Score
1	1	60
61	61	120
121	121	180
181	181	240
241	241	300
301	301	323

רצף השאלתה
רצף התוצאה
דוגמאות לחומצות אמינו זהות
דוגמאות לחומצות אמינו דומות
דוגמאות לחומצות אמינו שונות

Download GenPept Graphics

סיכום

למדנו כיצד ניתן לחפש רצפים הומולוגיים לרצף שאילתה בעזרת הכלי BLAST. אם ברשותנו רצף נוקלאוטידים - נשתמש בכלי nucleotide blast; ואם ברשותנו רצף חומצות אמינו - נשתמש בכלי protein blast. תהליך החיפוש, כמו גם מבנה דף התוצאה, דומים מאוד בִּשְׁנֵי הכלים. החשיבות של השוואת רצפים במחקר ביואינפורמטי היא גדולה מאוד, והשימושים בה חולשים על פני תחומים רבים כמו איתור חלבונים דומים שמשתייכים לאותה משפחה, מחקר אבולוציוני המשווה רצפי חלבונים או רצפי גנים בין אורגניזמים שונים ועוד. אלו רק חלק מדוגמאות רבות שבהן חיפוש והעמדת רצפים באמצעות BLAST יכולים לתרום לקידום המחקר הביולוגי.

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