

# ביתוח השוואתי של מוטיבים גנטיים

ליהי מוטיבים ושימור גנטי ביונקים  
באמצעות מבני נתונים למחוזות

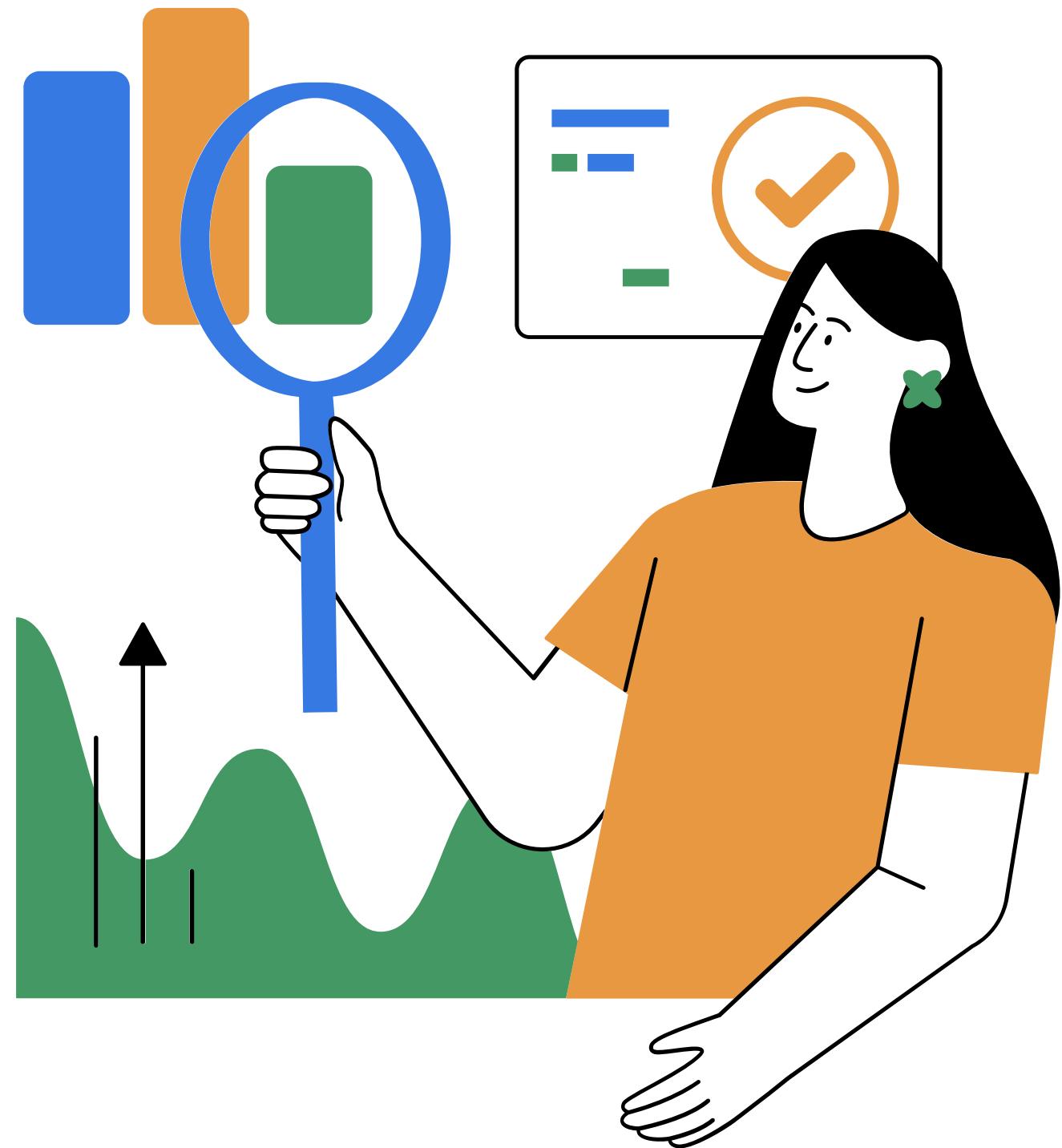
מגישות:

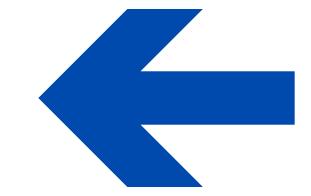
ימית מорנו

שקד טדסה

מרצה מנהה:

סמאח אידריס גלאו

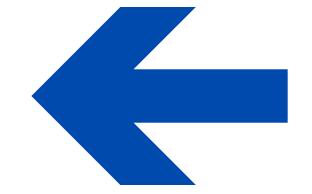




# הכרת הנזtones



- אדם (Human): ~ 5.5 מיליון בסיסים.
- שימפנזה (Chimpanzee): ~ 3.2 מיליון בסיסים.
- כלב (Dog): ~ 1.6 מיליון בסיסים.
- האלבבית: {A, C, G, T}.
- מהות הנזtones: רצפי DNA מקודדים (CDS) המחולקים למחולקות תפקודיות (Classes).
- מסר עיקרי: המקורות שונים באורכם, מה שמחיב ניתוח סטטיסטי מנורמל (ממוצעים וחיצונים).



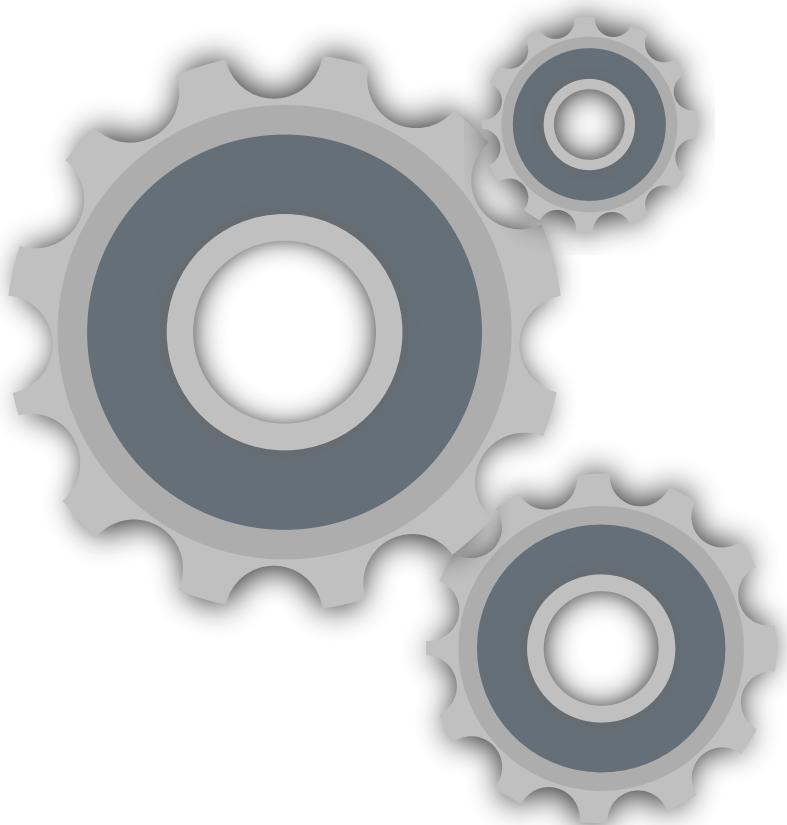
# הmethodולוגיה - איר חיפשו



שימוש ב-LCP Array-I Suffix Array.  
הסבר כללי: במקום לחפש "אות אחורי אות" (חיפוש נאיבי), אינדקסנו את כל הגנים למבנה מ민ין המאפשר למצוא חוזרות באורך  $T$  בשניות בודדות.

שימוש באלגוריתם Kasai למציאת התחלית המשותפת הארוכה ביותר (LCP).

מסר עיקרי: עילות אלגוריתמית מאפשרת לעבד מיליון בסיסי DNA בזמן אמיתי.



# חזרתיות פנימית - "טביעה האצבע" של המין

## ניתוח מבני של שכפולים עצמאיים בתוך הגנים



מה בדקנו?

סקירה משתנה: חיפשנו מוטיבים חוזרים בשלושה אורכים:

- $30 = L$  (רמת ה"אות"): מייצג "מילות מפתח" בגנים. אלו בדרך כלל אתרי קישור קטנים או בקרים שקבועים אם גן יפעל או לא.
  - $70 = L$  (רמת ה"פסקה"): מייצג דומיינים (Domains). אלו יחידות構筑ites עצמאיות בתוך חלבון. רצף באורך כזה כבר מוקודד לצורה פיזית מסוימת.
  - $120 = L$  (רמת ה"פרק"): מייצג משפחות גנים או מקטעים פונקציונליים שלמים. כאן אנחנו מחפשים את ה"העתק-הדבק" המשמעותי ביותר של האבולוציה.
- ספ רגישות: הגדרנו חזרתיות כרצף המופיע 3 פעמים ומעלה בתוך אותו המין.





# הגןום האנושי

LRS

```
ATGGTGCACGTAGCCAGGCTGCTGCTGCTGCCAC  
TTTCTTCCTCCGCACGGATGCTGAGACACCTCCAAGGTT  
TACACGAACACCCGTTGATCAGACAGGGGTCTGGCG  
GAGTTGCCTCTTCATCTGCCAAGCTACGGGAGACCCA  
AGACCTAAAATTGTCTGGAACAAAAAAGGAAAGAAAGTC  
AGCAATCAGAGATTGAGGTAATAGAGTTGACGATGGG  
TCTGGATCAGTTCTCAGAACATAACCCCTACGGACTCCG  
AGGGATGAGGCCATTATGAATGTGTGGCCTCAAATAAT  
GTGGGAGAAATAAGTGTATCCACCAACTCACAGTTTG  
CGGGAAGATCAAATCCCAGGGGCTCCCTACCATTGA  
CATGGGCCACAGTTGAAGGTGGTGAGCGTACTCGCA  
CGGCCACCATGTTGTGCAGCCAGTGGTAATCCGGAT  
CCAGAAATCACTGGTTAAAGATTCTTACCTGTGGACA  
CAAGCAACACAATGGTCGTATTAAGCAGTTACGATCAG  
AATCTATTGGTGGTACACCAATAAGAGGGAGCCCTCAGA  
TTGAGCAGAGTGAAGAGTCTGACCAAGGAAAATGAGT  
GTGTTGCCACCAACAGCGCGGGCACTCGCTATTCCGCT  
CCTGCCAATTATATGTCAGAGAGCTGCGAGAAGTCGC  
CGTGTCCCACCAAGATTCTCTATCCCACCCACTAATCAT  
GAAATCATGCCAGGCGGAAGCGTTAATATCACCTGTGTG  
GCCGTGGGTCACCAATGCCTTATGTAAGTGGATGTTG  
GGGGCAGAAGATCTGACACCTGAAGATGATATGCCAATA  
GGAAGAAATGTCTAGAACTGAATGATGTAAGACAGTCA  
GCAAATTACACCTGTGTTGCTATGTCAACACTGGGTGTC  
ATTGAAGCAATAGCACAGATCACTGTCAAAGCCTTACCC
```

**ה-LRS הענקי (29,533 בסיסים)**

מצאנו שני מקטעי DNA זהים באורך של גנים שלמים  
זהרי תופעה שנקראת Segmental Duplication.

בניגוד לחיות פשוטות, הגןום האנושי עשיר בשכפולים כאלה, שנחשים מנו  
אבולוציוני: עותק אחד נשאר פעיל והouthק השני חופשי לשתנות ולפתח תכונות חדשות.

EXECUTIVE REPORT: human

=====

DATA: Total Length: 5543336

ACTUAL LRS: ATGGTGCACGTAGCCAGGCTGCT  
Length: 29533

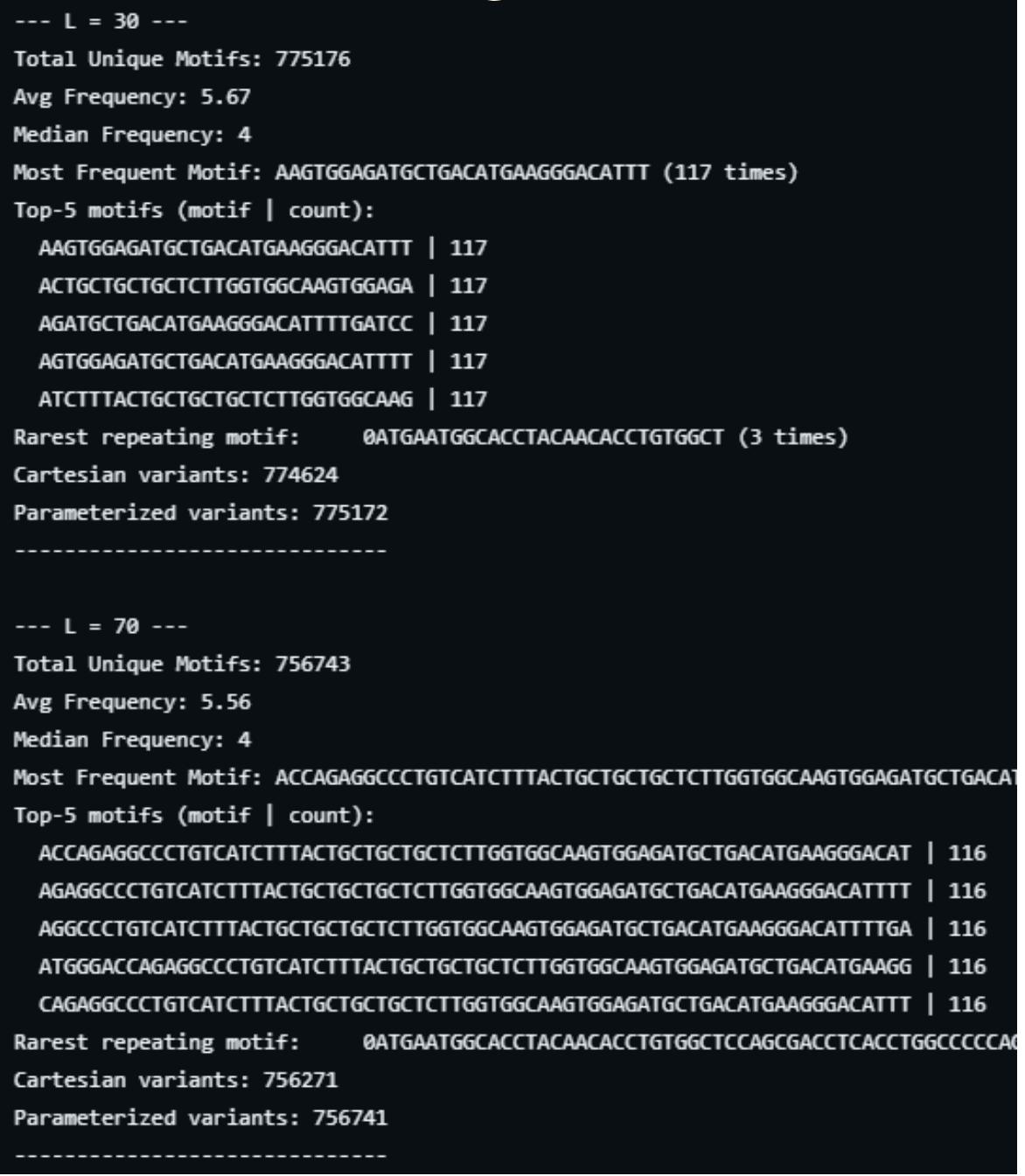
ANALYSIS RUNTIME: 151788.98 s

PEAK MEMORY: 4698907 KB



# הגןום האנושי

## חוות קצחות: המוח והמבנה של התא



### ממצא כללי

מצאנו חוזרות גבוהה מאוד (מעל 115 פעמים) ברצפים באורך **30 ו-70 בסיסים**. למרהן הדמיון בכמה חוות, כל אורך מייצג תפקוד ביולוגי שונה לחלווטין.

### רצפי 30 בסיסים – בקרה גנטית (Regulatory Motifs)

רצפים קצרים אלו משמשים כמתגים שמאפשרים או מכוונים גנים באמצעות קשרים חלבוניים לדNA. ריבוי החזרות מאפשר הפעלה מתואמת של קבוצות גנים שלמות בזמנית, למשל בתגובה להזע או מחלת.

### רצפי 70 בסיסים – מבנה חלבוניים (Protein Domains)

רצפים אלו מוקודדים לחומינים – יחידות תפקוד בסיסיות בתוך חלבוניים. חוות שלהן משקפות שימוש אבולוציוני עיל ב"חלקים מוצלחים", שמאפשר יצירה של חלבוניים מורכבים ורבתכלתיים.

# הגנוּם האנושי

## המוטיב הנפוץ (114 חזרות)

מחובר ברצף **אורך 120 בסיסים** שחוור מעל **100 פעמים**, מה שמעיד כנראה על משפחות גנים או דומיינים הלבוניים.

חזרות כאלה מאפשרות לגוף לייצר במהירות ובכמות גדולה לבוני חיוניים, למשל המערכת החיסונית או לשריר הלב.



```
--- L = 120 ---  
Total Unique Motifs: 733418  
Avg Frequency: 5.44  
Median Frequency: 4  
Most Frequent Motif: ACCAGAGGCCCTGTCATCTTACTGCTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGGACCAT (114 times)  
Top-5 motifs (motif | count):  
ACCAGAGGCCCTGTCATCTTACTGCTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGGACCAT | 114  
ACTGCTGCTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGGACCATCCCAGACAGTGACATCTCTGC | 114  
ATCTTTACTGCTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGGACCATCCCAGACAGTGACAT | 114  
ATGGGACCAAGAGGCCCTGTCATCTTACTGCTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGG | 114  
CTGCTCTTGGTGGCAAGTGGAGATGCTGACATGAAGGGACATTTGATCCTGCCAAGTGCCGCTATGCCCTGGGCATGCAGGGACCGGACCATCCCAGACAGTGACATCTCTGCTTCCAGC | 114  
Rarest repeating motif: 0ATGAATGGCACCTACAACACCTGTGGCTCCAGCGACCTCACCTGGCCCCAGCGATCAAGCTGGGCTTCTACGCCCTACTTGGGCGTCTGCTGGTGCAGGCCTGCTAACAGCC (3 times)  
Cartesian variants: 732981  
Parameterized variants: 733416
```

# מיזמים

## דפוס ה"צברים" (Clustering)

החוויות מופיעות בקבוצות צפופות למרחוקים קצרים, מה שمعد על Gene Clusters. אזורים כאלה מכילים עותקים רבים של אותו רכיב, ונפוצים בגנים מבניים או במערכת החיסון, שבהם נדרש מגון גדול באותה מקום אחד.

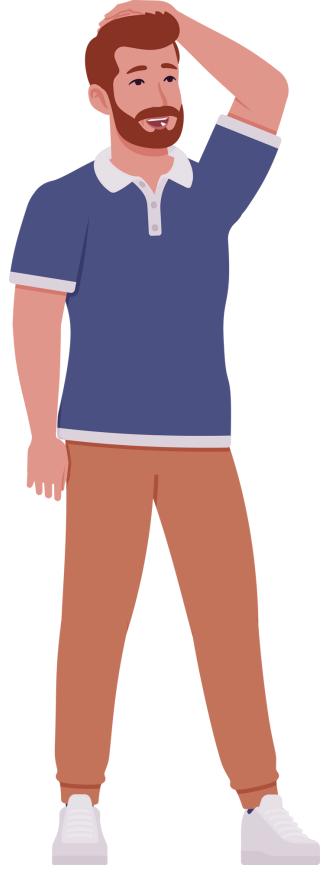
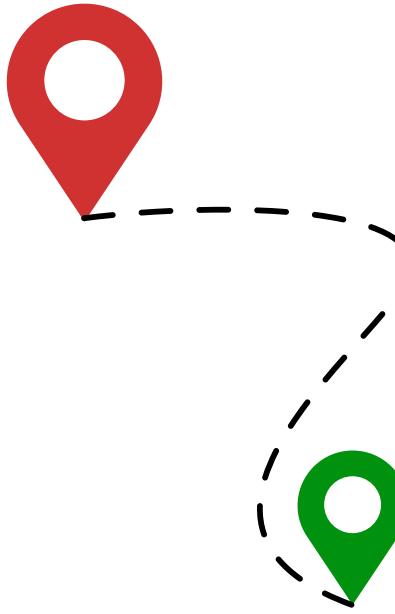


## דפוס הקפיצות (Large Genomic Gaps)

בין קבוצות החוויות יש פערים גנטיים עצומים, של מאות אלפי עד מיליון בסיסים. זהו סימן מובהק ל-*Segmental Duplications* – שכפול של בלוק DNA שלם והעתיקתו למיקום רחוק בגנים.

## דפוס החזרות המחרזיות (Periodic Repeats)

באזורים מסוימים החזרות מופיעות במרווח קבוע וסדיר מאד. דפוס מה מאפיין Tandem Arrays וمعد לרוב על אזורים בעלי תפקיד מבני ב-DNA, כמו הцентрומה, החיוני לחלוקת תא תקינה.



# מיקומים

Pos: [339850, 340480, 340982, 341427, 341846, 343177, 345810, 348443, 351187, 353931, 358045, 358779, 361577, 364210, 1550969, 1553602, 1556235, 1558979, 1561723, 1564356, 1566467, 1566912, 1567331, 1568662, 1569396, Pos: [339871, 340501, 341003, 341448, 341867, 343198, 345831, 348464, 351208, 353952, 358066, 358800, 361598, 364231, 1550990, 1553623, 1556256, 1559000, 1561744, 1564377, 1566488, 1566933, 1567352, 1568683, 1569417, Pos: [339865, 340495, 340997, 341442, 341861, 343192, 345825, 348458, 351202, 353946, 358060, 358794, 361592, 364225, 1550984, 1553617, 1556250, 1558994, 1561738, 1564371, 1566482, 1566927, 1567346, 1568677, 1569411, Pos: [339845, 340475, 340977, 341422, 341841, 343172, 345805, 348438, 351182, 353926, 358040, 358774, 361572, 364205, 1550964, 1553597, 1556230, 1558974, 1561718, 1564351, 1566462, 1566907, 1567326, 1568657, 1569411, Pos: [339878, 340508, 341010, 341455, 341874, 343205, 345838, 348471, 351215, 353959, 358073, 358807, 361605, 364238, 1550997, 1553630, 1556263, 1559007, 1561751, 1564384, 1566495, 1566940, 1567359, 1568690, 1569424, Pos: [339875, 340505, 341007, 341452, 341871, 343202, 345835, 348468, 351212, 353956, 358070, 358804, 361602, 364235, 1550994, 1553627, 1556260, 1559004, 1561748, 1564381, 1566492, 1566937, 1567356, 1568687, 1569421, Pos: [339872, 340502, 341004, 341449, 341868, 343199, 345832, 348465, 351209, 353953, 358067, 358801, 361599, 364232, 1550991, 1553624, 1556257, 1559001, 1561745, 1564378, 1566489, 1566934, 1567353, 1568684, 1569418, Pos: [339867, 340497, 340999, 341444, 341863, 343194, 345827, 348460, 351204, 353948, 358062, 358796, 361594, 364227, 1550986, 1553619, 1556252, 1558996, 1561740, 1564373, 1566484, 1566929, 1567348, 1568679, 1569413, Pos: [339849, 340479, 340981, 341426, 341845, 343176, 345809, 348442, 351186, 353930, 358044, 358778, 361576, 364209, 1550968, 1553601, 1556234, 1558978, 1561722, 1564355, 1566466, 1566911, 1567330, 1568661, 1569395, Pos: [339877, 340507, 341009, 341454, 341873, 343204, 345837, 348470, 351214, 353958, 358072, 358806, 361604, 364237, 1550996, 1553629, 1556262, 1559006, 1561750, 1564383, 1566494, 1566939, 1567358, 1568689, 1569423, Pos: [339874, 340504, 341006, 341451, 341870, 343201, 345834, 348467, 351211, 353955, 358069, 358803, 361601, 364234, 1550993, 1553626, 1556259, 1559003, 1561747, 1564380, 1566491, 1566936, 1567355, 1568686, 1569420, Pos: [339848, 340478, 340980, 341425, 341844, 343175, 345808, 348441, 351185, 353929, 358043, 358777, 361575, 364208, 1550967, 1553600, 1556233, 1558977, 1561721, 1564354, 1566465, 1566910, 1567329, 1568660, 1569394, Pos: [339847, 340477, 340979, 341424, 341843, 343174, 345807, 348440, 351184, 353928, 358042, 358776, 361574, 364207, 1550966, 1553599, 1556232, 1558976, 1561720, 1564353, 1566464, 1566909, 1567328, 1568659, 1569393, Pos: [339870, 340500, 341002, 341447, 341866, 343197, 345830, 348463, 351207, 353951, 358065, 358799, 361597, 364230, 1550989, 1553622, 1556255, 1558999, 1561743, 1564376, 1566487, 1566932, 1567351, 1568682, 1569416, Pos: [339866, 340496, 340998, 341443, 341862, 343193, 345826, 348459, 351203, 353947, 358061, 358795, 361593, 364226, 1550985, 1553618, 1556251, 1558995, 1561739, 1564372, 1566483, 1566928, 1567347, 1568678, 1569412, Pos: [339879, 340509, 341011, 341456, 341875, 343206, 345839, 348472, 351216, 353960, 358074, 358808, 361606, 364239, 1550998, 1553631, 1556264, 1559008, 1561752, 1564385, 1566496, 1566941, 1567360, 1568691, 1569425, Pos: [339876, 340506, 341008, 341453, 341872, 343203, 345836, 348469, 351213, 353957, 358071, 358805, 361603, 364236, 1550995, 1553628, 1556261, 1559005, 1561749, 1564382, 1566493, 1566938, 1567357, 1568688, 1569422, Pos: [339873, 340503, 341005, 341450, 341869, 343200, 345833, 348466, 351210, 353954, 358068, 358802, 361600, 364233, 1550992, 1553625, 1556258, 1559002, 1561746, 1564379, 1566490, 1566935, 1567354, 1568685, 1569419, Pos: [339846, 340476, 340978, 341423, 341842, 343173, 345806, 348439, 351183, 353927, 358041, 358775, 361573, 364206, 1550965, 1553598, 1556231, 1558975, 1561719, 1564352, 1566463, 1566908, 1567327, 1568658, 1569392, Pos: [339869, 340499, 341001, 341446, 341865, 343196, 345829, 348462, 351206, 353950, 358064, 358798, 361596, 364229, 1550988, 1553621, 1556254, 1558998, 1561742, 1564375, 1566486, 1566931, 1567350, 1568681, 1569415, Pos: [339868, 340498, 341000, 341445, 341864, 343195, 345828, 348461, 351205, 353949, 358063, 358797, 361595, 364228, 1550987, 1553620, 1556253, 1558997, 1561741, 1564374, 1566485, 1566930, 1567349, 1568680, 1569414, Pos: [339853, 340483, 340985, 341430, 341849, 343180, 345813, 348446, 351190, 353934, 358048, 358782, 361580, 364213, 1550972, 1553605, 1556238, 1558982, 1561726, 1564359, 1566470, 1566915, 1567334, 1568665, 1569399, Pos: [339855, 340485, 340987, 341432, 341851, 343182, 345815, 348448, 351192, 353936, 358050, 358784, 361582, 364215, 1550974, 1553607, 1556240, 1558984, 1561728, 1564361, 1566472, 1566917, 1567336, 1568667, 1569401, Pos: [339852, 340482, 340984, 341429, 341848, 343179, 345812, 348445, 351189, 353933, 358047, 358781, 361579, 364212, 1550971, 1553604, 1556237, 1558981, 1561725, 1564358, 1566469, 1566914, 1567333, 1568664, 1569398, Pos: [339864, 340494, 340996, 341441, 341860, 343191, 345824, 348457, 351201, 353945, 358059, 358793, 361591, 364224, 1550983, 1553616, 1556249, 1558993, 1561737, 1564370, 1566481, 1566926, 1567345, 1568676, 1569410, Pos: [339851, 340481, 340983, 341428, 341847, 343178, 345811, 348444, 351188, 353932, 358046, 358780, 361578, 364211, 1550970, 1553603, 1556236, 1558980, 1561724, 1564357, 1566468, 1566913, 1567332, 1568663, 1569397, Pos: [339858, 340488, 340990, 341435, 341854, 343185, 345818, 348451, 351195, 353939, 358053, 358787, 361585, 364218, 1550977, 1553610, 1556243, 1558987, 1561731, 1564364, 1566475, 1566920, 1567339, 1568670, 1569404, Pos: [339859, 340489, 340991, 341436, 341855, 343186, 345819, 348452, 351196, 353940, 358054, 358788, 361586, 364219, 1550978, 1553611, 1556244, 1558988, 1561732, 1564365, 1566476, 1566921, 1567340, 1568671, 1569405, Pos: [339860, 340490, 340992, 341437, 341856, 343187, 345820, 348453, 351197, 353941, 358055, 358789, 361587, 364220, 1550979, 1553612, 1556245, 1558989, 1561733, 1564366, 1566477, 1566922, 1567341, 1568672, 1569406, Pos: [339854, 340484, 340986, 341431, 341850, 343181, 345814, 348447, 351191, 353935, 358049, 358783, 361581, 364214, 1550973, 1553606, 1556239, 1558983, 1561727, 1564360, 1566471, 1566916, 1567335, 1568666, 1569400, Pos: [339857, 340487, 340989, 341434, 341853, 343184, 345817, 348450, 351194, 353938, 358052, 358786, 361584, 364217, 1550976, 1553609, 1556242, 1558986, 1561730, 1564363, 1566474, 1566919, 1567338, 1568669, 1569403, Pos: [339856, 340486, 340988, 341433, 341852, 343183, 345816, 348449, 351193, 353937, 358051, 358785, 361583, 364216, 1550975, 1553608, 1556241, 1558985, 1561729, 1564362, 1566473, 1566918, 1567337, 1568668, 1569402, Pos: [339862, 340492, 340994, 341439, 341858, 343189, 345822, 348455, 351199, 353943, 358057, 358791, 361589, 364222, 1550981, 1553614, 1556247, 1558991, 1561735, 1564368, 1566479, 1567343, 1568674, 1569408, Pos: [339863, 340493, 340995, 341440, 341859, 343190, 345823, 348456, 351200, 353944, 358058, 358792, 361590, 364223, 1550982, 1553615, 1556248, 1558992, 1561736, 1564369, 1566480, 1566925, 1567344, 1568675, 1569409, Pos: [339861, 340491, 340993, 341438, 341857, 343188, 345821, 348454, 351198, 353942, 358056, 358790, 361588, 364221, 1550980, 1553613, 1556246, 1558990, 1561734, 1564367, 1566478, 1566923, 1567342, 1568673, 1569407, Pos: [339881, 340511, 341013, 341458, 341877, 343208, 345841, 348474, 351218, 353962, 358076, 358810, 361608, 364241, 1551000, 1553633, 1556266, 1559010, 15617

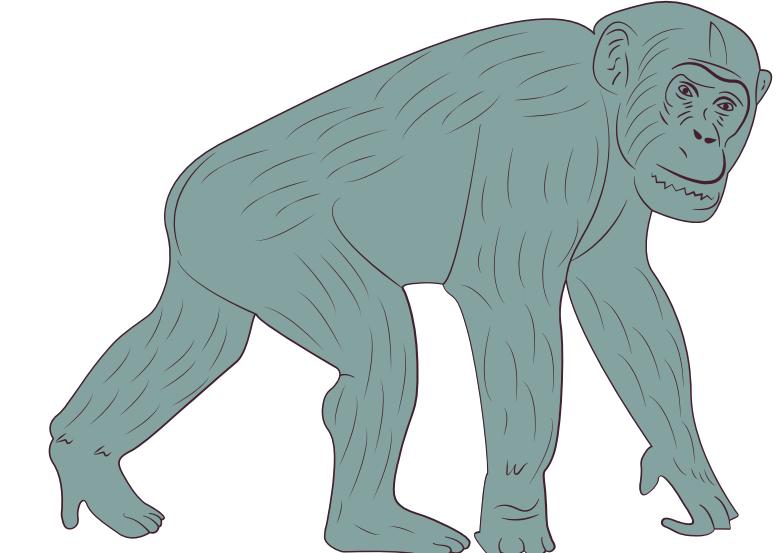
# הגנון של השימפנזה

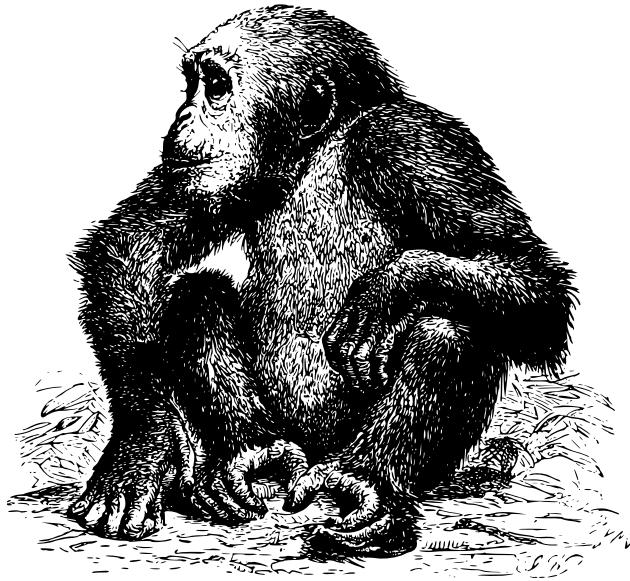
LRS

```
GGTCAGAAATCTTGGATAGAGAAAACCTTGCAACCGG  
GAATGTATCTTGTATTCTAGCACGAAAGACTCTAACAA  
GGTGGCTGGCCAGTTACCAACCAGCATATCCCC  
CCTCTGCCAAGTCAACACCCAGCAAAATGAAGAGGA  
AAACAAACAGGTGGAGACTCAGCCTGAGAAATGGCTG  
TTGCCAACACCCCAGAGCTACCCAACAGATTCTATG  
GAGTTCTGAATTCCAGGGTGGCGGATATTCCAATAAG  
CCATGTATATCCGTATCCTATGACACCAAGCCAGACTC  
ACTGCTCCATCTCATGGGAAAGATTGGCAGCTGGAACT  
CCCCAAGCTTTAATATCTGTGCATGGAGGCCTCCAGAA  
CTTGAGATGCAGCCAAAGCTGAAACAAGTCTTGGGA  
AAGGCCTGATCAAGGCTGCCATGACCACCGGGGCTG  
GATCTCACCGGGGTGTCAGCACAGGTGTTATCAGCC  
ACGTAGGGGATGCCTGAAAGACCACTCCTCCAAGTCC  
AGAGGCCGGTTGTCTAGGAATTGCTCCATGGGG  
CATCGTGGAGAATAAGGAAGACCTGGTGGAAAGGATG  
TAACAAGAGTGTACCAAGACCATGTCCAACCCCTCAAGTA  
AGCTCTGTGCTCAACAACCTCCCACACCCACTTCATCC  
TGGCTGACAATGGCACCCCTGGCAAGTATGGCGCCGAG  
GTGAAGCTGCGAAGGCTGCTGGAAAAGCACATCTCCCT  
GCAGAAGATCAACACAAGATTGGGGCAGGGCGTCCCC  
CTCGTGGTCTCGTGGAGGGGGCCCTACGTGGT  
GTCCATCGTCTTGAATACCTGCAAGAAGAGGCCTCCCGT  
CCCTGTGGTATTGTGATGGCAGCGGACGTGCCTCGG  
ACATCCTGTCTTGCAGCACAAGTACTGTGAAGAAGGCG  
GAATAATAATGAGTCCCTCAGGGAGCAGCTTAGTTA
```

**ה-LRS של השימפנזה (9,757 בסיסים)**  
אורק הרצף שנבדק הוא כ-3.2 מיליון בסיסים, וה-LRS האחרון ביותר עומד על 9,757 בסיסים.  
בשוואה לאדם, שבו ה-LRS אחר בהרבה, הדבר מצביע על קר שהabolition האנושית כליה אוירוני Segmental Duplication נרחבים יותר, שתורמים ליצירת גנים חדשים ומורכבות גבואה יותר.

```
EXECUTIVE REPORT: chimpanzee  
=====  
DATA: Total Length: 3256578  
ACTUAL LRS: GGTCAAGAAATCTTGGATAGAGA  
Length: 9757  
  
ANALYSIS RUNTIME: 2551.99 s  
PEAK MEMORY: 2915584 KB
```





# הגבום של השיפנה

# **חוֹרֶת קָצְרוֹת: המוח והמבנה של הטא**

## **חוROT קצROT (L=30) – "שפט הבקרה"**

בשימפנזה המוטיב הנפוץ הופיע 105 פעמים, לעומת 117 אצל האדם. רצפים אלה משמשים כמתגים  
והחזרה הגבוהה יותר באדם מצביעה על רשת רגולטורית מורכבת וסנכרון מתקדם של פעילות גנים,  
במיוחד במוח.

**חזרות בינויים (70=L) – “ארם הכלים המבני”**  
בשימפנזה החזרות נמוכות (עד 36), בעוד שבאדם רחלביוניים, מה שקרה שימוש חוזר אינטנסיבי באדרב תכלייתיים.

## מה זה ?N (Unknown Base)

האות N מייצגת בסיס DNA שלא זהה בזוזאות במהלך הריצוף, כאשר המכשיר מעדיף לסמן חוסר מידע במקום לנחש. הופעה חוזרת של N בשימפנזה מעידה על אזהרים קשים לרכיב, לרבות Heterochromatin.

חשיבות השוואת האדם

החוויות הרבות של א' מצביעות על כך שהגנום של השימפנזה פחות מפוענה ומכל יותר "חורים" מידעים. לעומת זאת, הגנום האנושי עבר ריצוף והשלמה מעמיקים יותר, ולכן מכל פחות אזורי א'.

# הגבום של השיפנה



## **רמת "התבניות המורכבות" (L=120) – אדם מול שיפנה**

בשימפנזה נמצאה שכיחות נמוכה מאוד של חזרות באורך זה (עד 10 חזרות) לעומת גורם אחד. מכאן ש-120 בסיסים מייצגים תתי-חידה תפקודית שלמה בחלבון, הפעם מעיד שאצל האדם מתרכז שכפול נרחב של מקטעים מורכבים, המאפשר ייצור משפחות גנים גדולות ומערכות חלבוניים מתקדמת יותר, בעוד שבשימפנזה רצפים כאלה נדרים ומשמשים בעיקר לתפקידו המרכזי כמבנה כרומוזומלי.

# הגנוּם של הכלב

LRS

```
GGTCAGAAATCTTGGATAGAGAAAACCTTGCAACGG  
GAATGTATCTTGTAAATTCTAGCACGAAAGACTCTAAC  
GGTGGCTGTGCCAGTTACCAACCAGCATATCCCC  
CCTCTGCCAAGTGCAACACCCCAGCAAAATGAAGAGGA  
AAACAAACAGGTGGAGACTCAGCCTGAGAAATGGTCTG  
TTGCCAAGCACACCCAGAGCTACCCAACAGATTCTATG  
GAGTTCTGAATTCCAGGGGGCGGATATTCCAATAAG  
CCATGTATATCCGTGTATCCTATGACACCAAGGCCAGACTC  
ACTGCTCCATCTCATGGTAAAGATTGGCAGCTGGAACT  
CCCCAAGCTTTAATATCTGTGCATGGAGGCCTCCAGAA  
CTTGAGATGCAGCCAAAGCTGAAACAAGTCTTGGGA  
AAGGCCTGATCAAGGCTGCCATGACCACCGGGGCCTG  
GATCTCACCGGGGGTGTCAAGCACAGGTGTTATCAGCC  
ACGTAGGGGATGCCTGAAAGACCACCTCCTCCAAGTCC  
AGAGGCCGGGTTGTGCTATAGGAATTGCTCCATGGGG  
CATCGTGGAGAATAAGGAAGACCTGGTGGAAAGGATG  
TAACAAAGAGTGTACCAAGACCATGTCCAACCCCTCTAAGTA  
AGCTCTCTGTGCTAACAACTCCCACACCCACTTCATCC  
TGGCTGACAATGGCACCCCTGGCAAGTATGGCGCCGAG  
GTGAAGCTGCGAAGGCTGCTGGAAAAGCACATCTCCCT  
GCAGAAGATCAACACAAGATTGGGGCAGGGCGTCCCC  
CTCGTGGGTCTCGTGGTGGAGGGGGCCCTAACGTGGT  
GTCCATCGTCTTGGAAATACCTGCAAGAAGAGGCCTCCGT  
CCCTGTGGTATTGTGATGGCAGCGGACGTGCCTCGG  
ACATCCTGTCCCTTGCGCACAGTACTGTGAAGAAGGCG  
GAATAATAATGAGTCCCTCAGGGAGCAGCTTAGTTA
```

**LRS כל הכלב – רצף חזר אחר (6,881 בסיסים)**  
בכלב, ה-LRS מראה כמעט חצי מהדימה, מה שמעיד שהחלק גדול מהגנוּם בניו מהכפלה אחת ענקית ומדויקת. בניגוד לאדם, הגנוּם של הכלב פחות מבודר וכל בלוקים גדולים שחזרים על עצמם, סימן לאבולוציה של Segmental Duplications מאסיביות.

EXECUTIVE REPORT: dog  
=====

DATA: Total Length: 1666555

ACTUAL LRS: CAGGTGATTTCAAATCAGAG  
Length: 6881

ANALYSIS RUNTIME: 389.24 s

PEAK MEMORY: 635315 KB





# הגןום של הכלב

## חוות קצחות: המוח והמבנה של התא

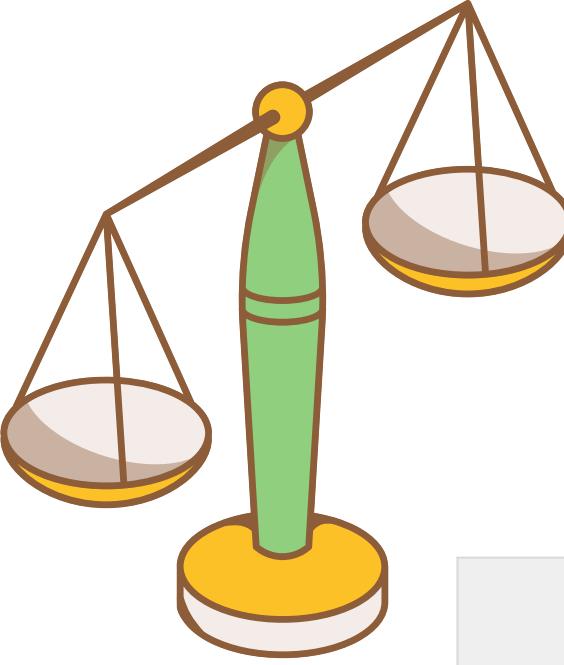
```
--- L = 70 ---
Total Unique Motifs: 30171
Avg Frequency: 3.02
Median Frequency: 3
Most Frequent Motif: AGATAGGCAGAACCAATGAGACTCGCTATGAGGTACAGGCCCTGGAAACCCGGGACCCGTACAACCTCAG (6 times)
Top-5 motifs (motif | count):
AGATAGGCAGAACCAATGAGACTCGCTATGAGGTACAGGCCCTGGAAACCCGGGACCCGTACAACCTCAG | 6
CAGATAGGCAGAACCAATGAGACTCGCTATGAGGTACAGGCCCTGGAAACCCGGGACCCGTACAACCTCA | 6
CCAGATAGGCAGAACCAATGAGACTCGCTATGAGGTACAGGCCCTGGAAACCCGGGACCCGTACAACCTC | 6
GATAGGCAGAACCAATGAGACTCGCTATGAGGTACAGGCCCTGGAAACCCGGGACCCGTACAACCTCAGC | 6
AAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTGAGAGGAAATGAAGGGCAAGGTGCTGGCGAGTGCACGTTGAAAGAAAGTC
Rarest repeating motif: AAAAAAAAGTGCCTCGCTACATACAAGGTGAAGGCAGTCGCTCACGTTGAGATGGAAGC (3 times)
Cartesian variants: 30171
Parameterized variants: 30171
-----
--- L = 120 ---
Total Unique Motifs: 27872
Avg Frequency: 3.02
Median Frequency: 3
Most Frequent Motif: AAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTGAGAGGAAATGAAGGGCAAGGTGCTGGCGAGTGCACGTTGAAAGAAAGTC
Top-5 motifs (motif | count):
AAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTGAGAGGAAATGAAGGGCAAGGTGCTGGCGAGTGCACGTTGAAAGAAAGTCGGCC | 4
AAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTGAGAGGAAATGAAGGGCAAGGTGCTGGCGAGTGCACGTTGAAAGAAAGTCGGCCATCAACCAGATCC | 4
AAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTGAGAGGAAATGAAGGGCAAGGTGCTGGCGAGTGCACGTTGAAAGAAAGTCGGCC | 4
AAACAGGAATCTCCAGAGTGACGTCGGCTACTCCACAGCTCAAAGCATCAGGACTCCAAAAGGAAGAAGAAAAGAAGAAGCCCCACATAAAGAACCTCTTAACGCATTCTATGTTG | 4
AAACCCACCTCCACACTTACAGCTACGTTAGACCCCCAAACAGGAATCCCACGGCTCCGACCCCTCCAGATATCTCGTATTACCCGCTATCGCCTGGCACCGTAGGACAAATCCC | 4
Rarest repeating motif: AAAAAAAAGTGCCTCGCTACATACAAGGTGAAGGCAGTCGCTCACGTTGAGATGGAAGCTTACTAGACTCGCCTCCCTCCCTCCCCAACCTGCTAGGCTCC
Cartesian variants: 27872
Parameterized variants: 27872
```

### השוואת ה"מטרגים" (L=30) – היכן הבקרה?

בכלב השכיחות המקסימלית של מטרגים היא 44 חזרות בלבד, לעומת 105 בשימפנזה ו-111 באדם. משמעות הדבר היא שלכלב יש פחות נזקודות בקרה משותפות, והוא מפעיל שתות גנים קטנות וממוקדות יותר, עם חזרות כמו AGCAGCAGC קטנה משמעותית.

### רמת ה"מודלים" (L=70 ו-L=120) – עילות מול מורכבות

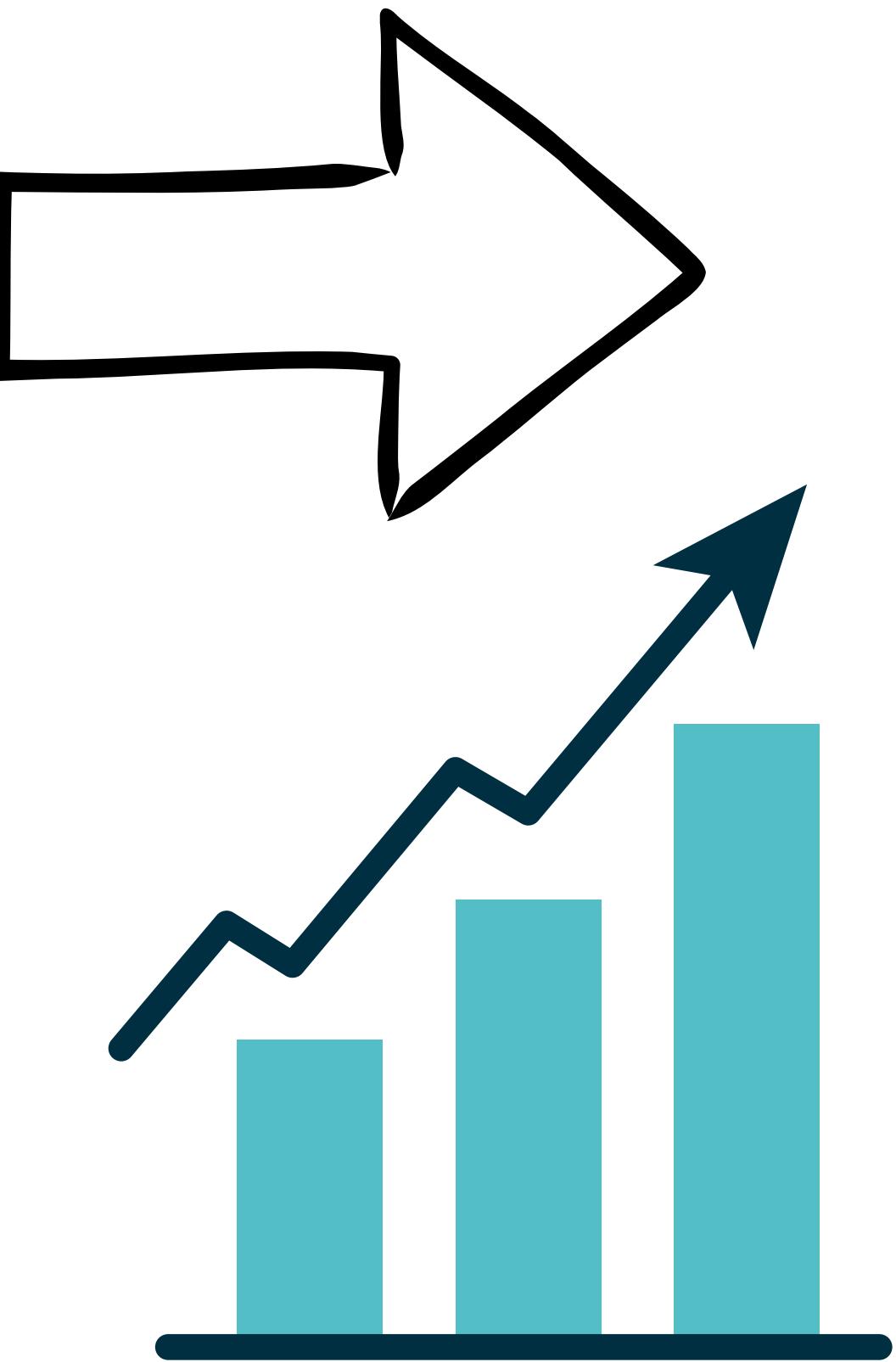
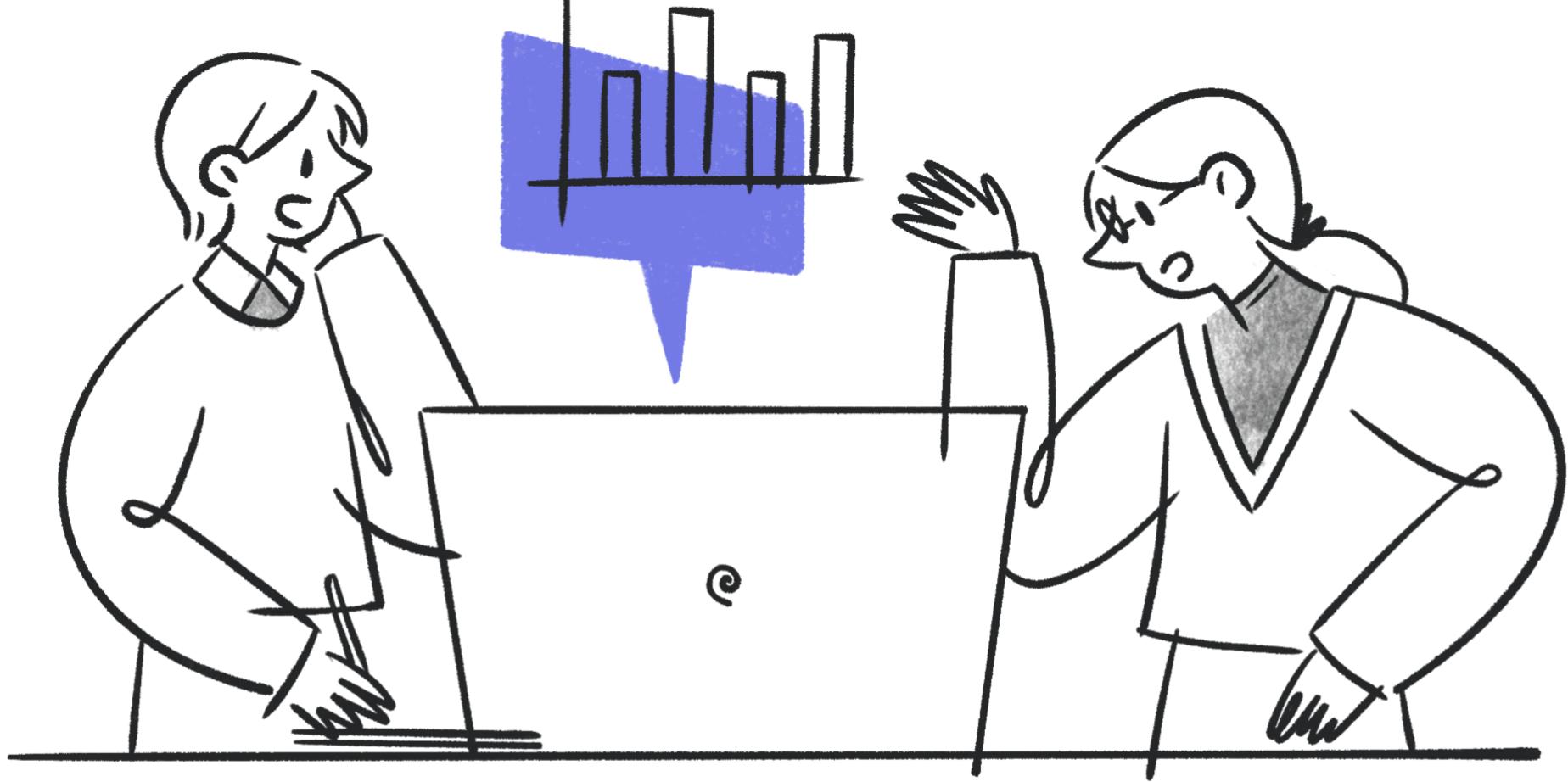
בשכיחות רצפי 70–120 בסיסים הכלב מגיע ל-4–6 חזרות בלבד, בעוד שבאדם השכיחות נשארת גבוהה (114–116). זה מעיד שלכלב יש "ארם כלים" חלבוני ייחודי, בעוד האדם משכפל יחידות מוצלחות שוב ושוב לצירת חלבונים מורכבים ורבתתכליתיים.



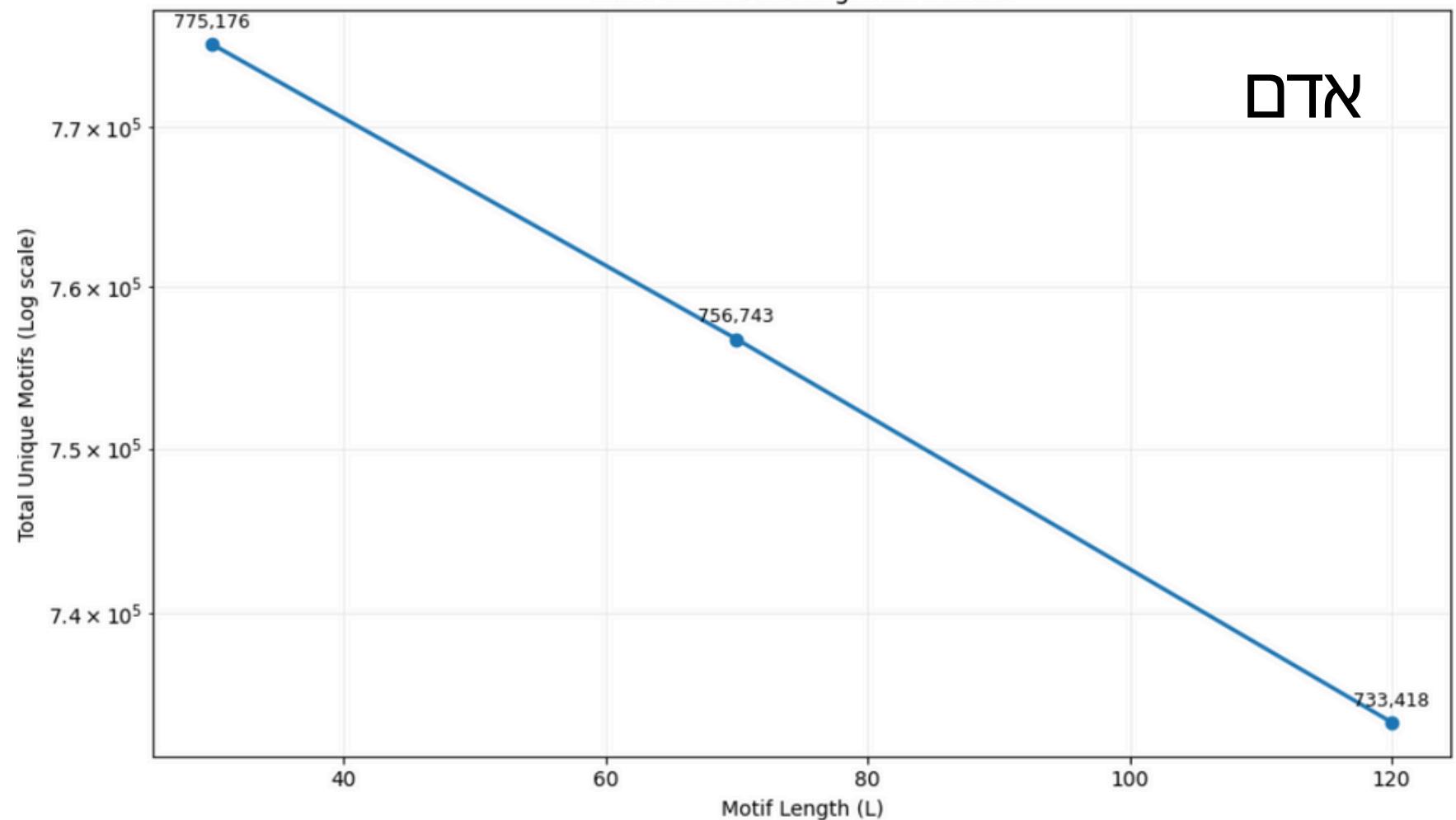
# טבלת השוואות

מדד	אדם	שימפנזה	כלב
LRS אורך	29,500~ (שיא המורכבות)	9,700~	6,800~
L=30 שכיחות ב-	117 (בקלה מסועפת)	105	44 (בקלה פשוטה)
L=120 שכיחות ב-	114 (שכפול תבניות)	10	4 (יחידות גבואה)
שימוש ב- N (חוריים)	נמוך מאוד	גבוה (36)	נמוך

# גראף

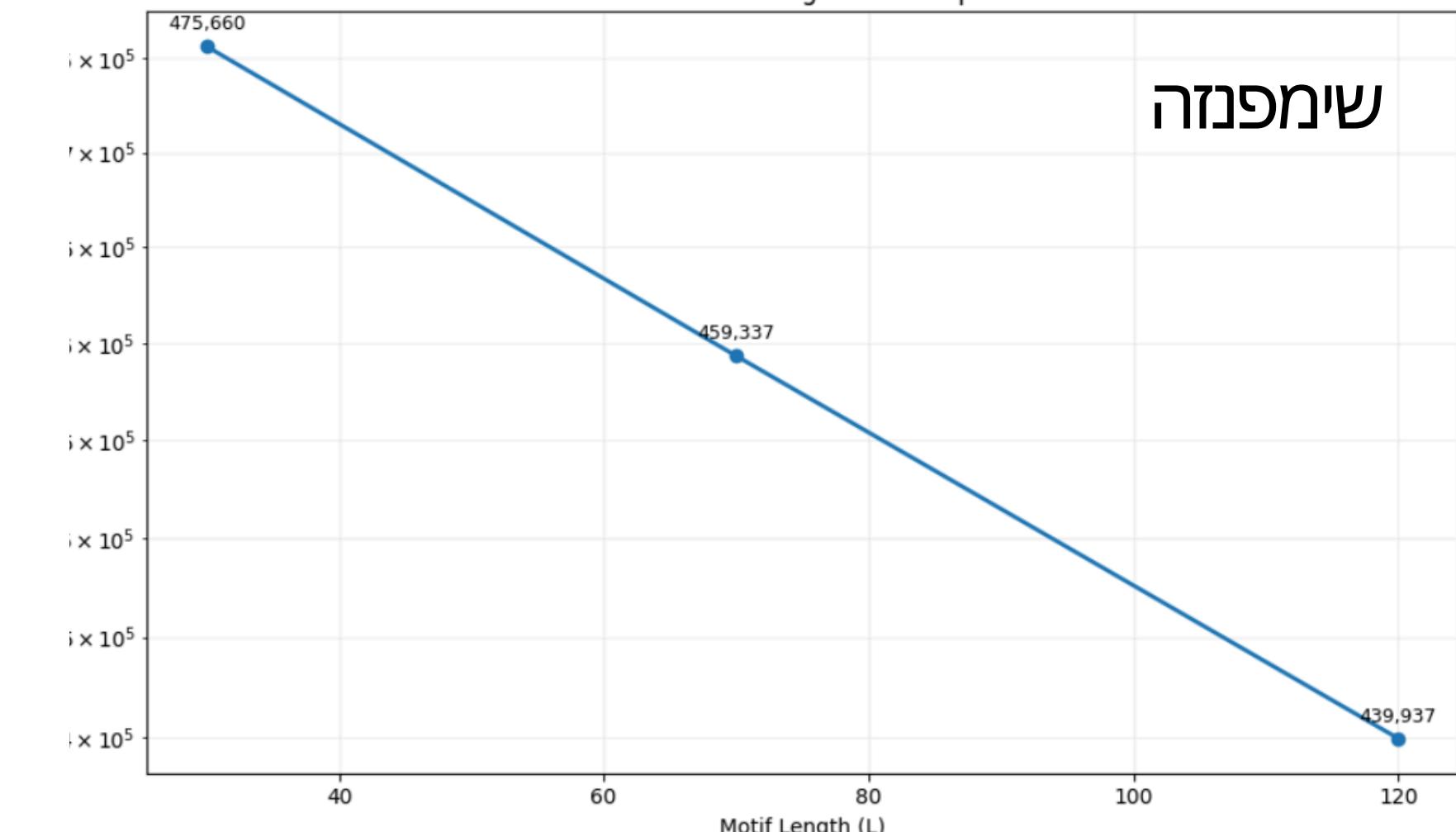


Motif Counts vs Length for Human



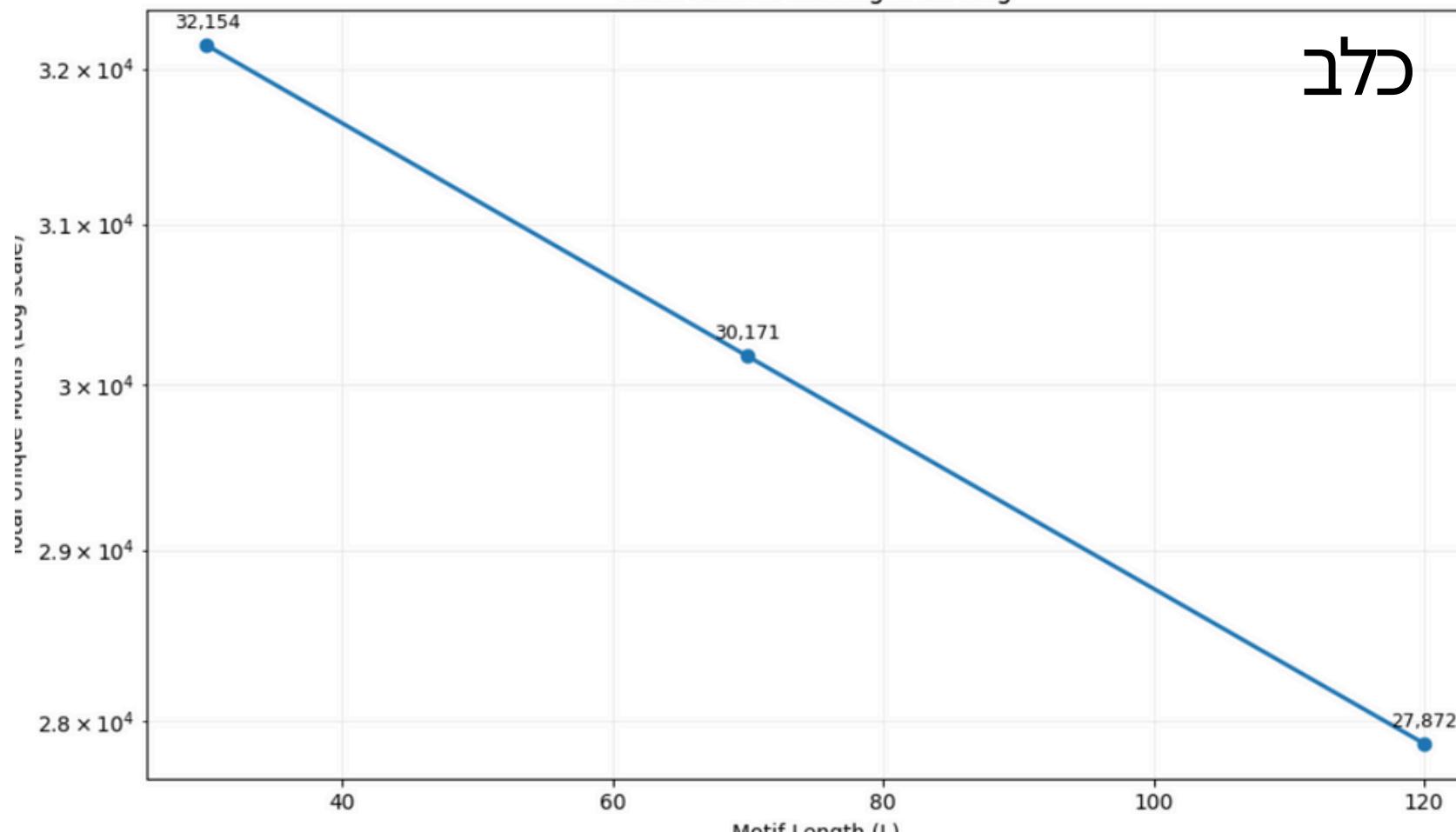
אדם

Motif Counts vs Length for Chimpanzee



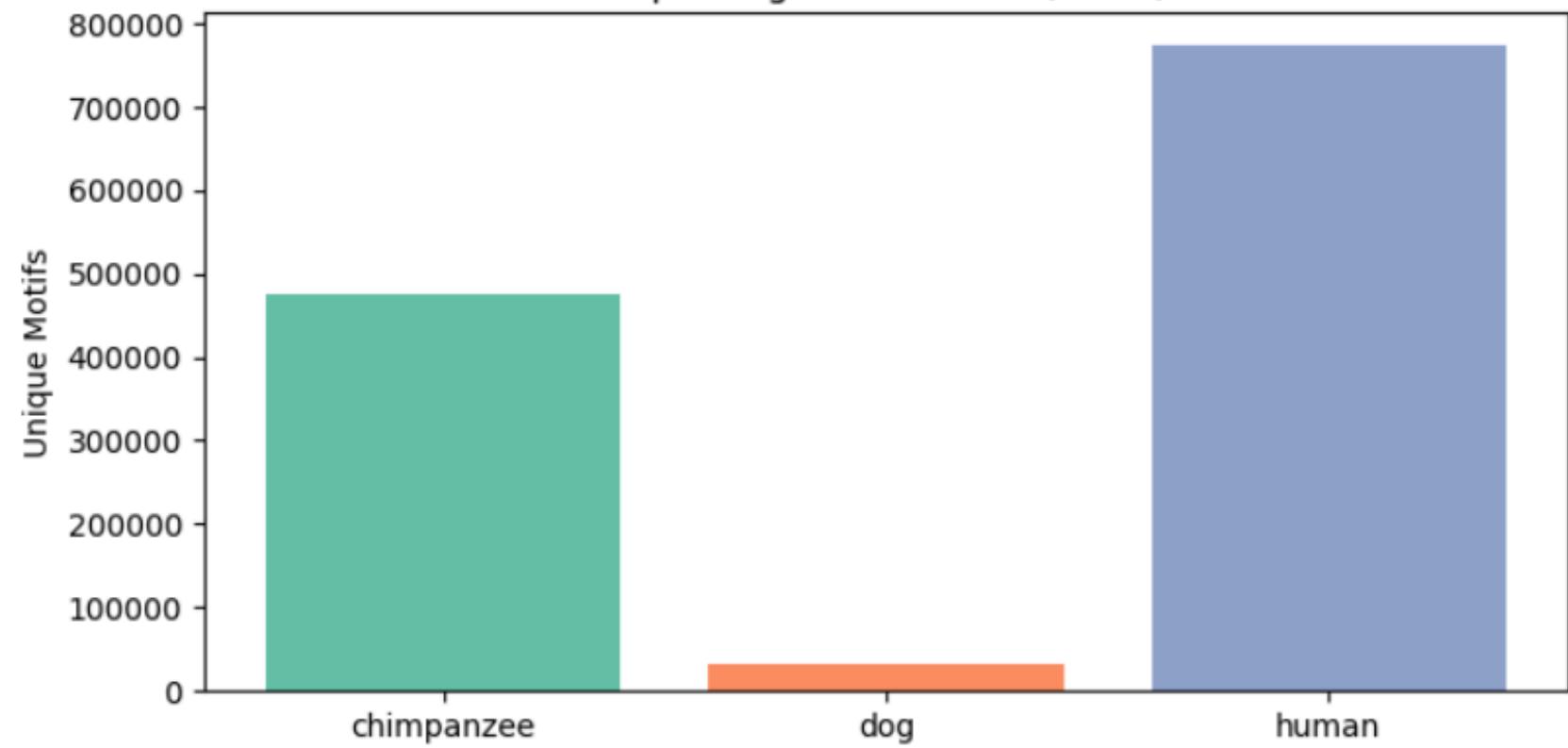
שימפנזה

Motif Counts vs Length for Dog

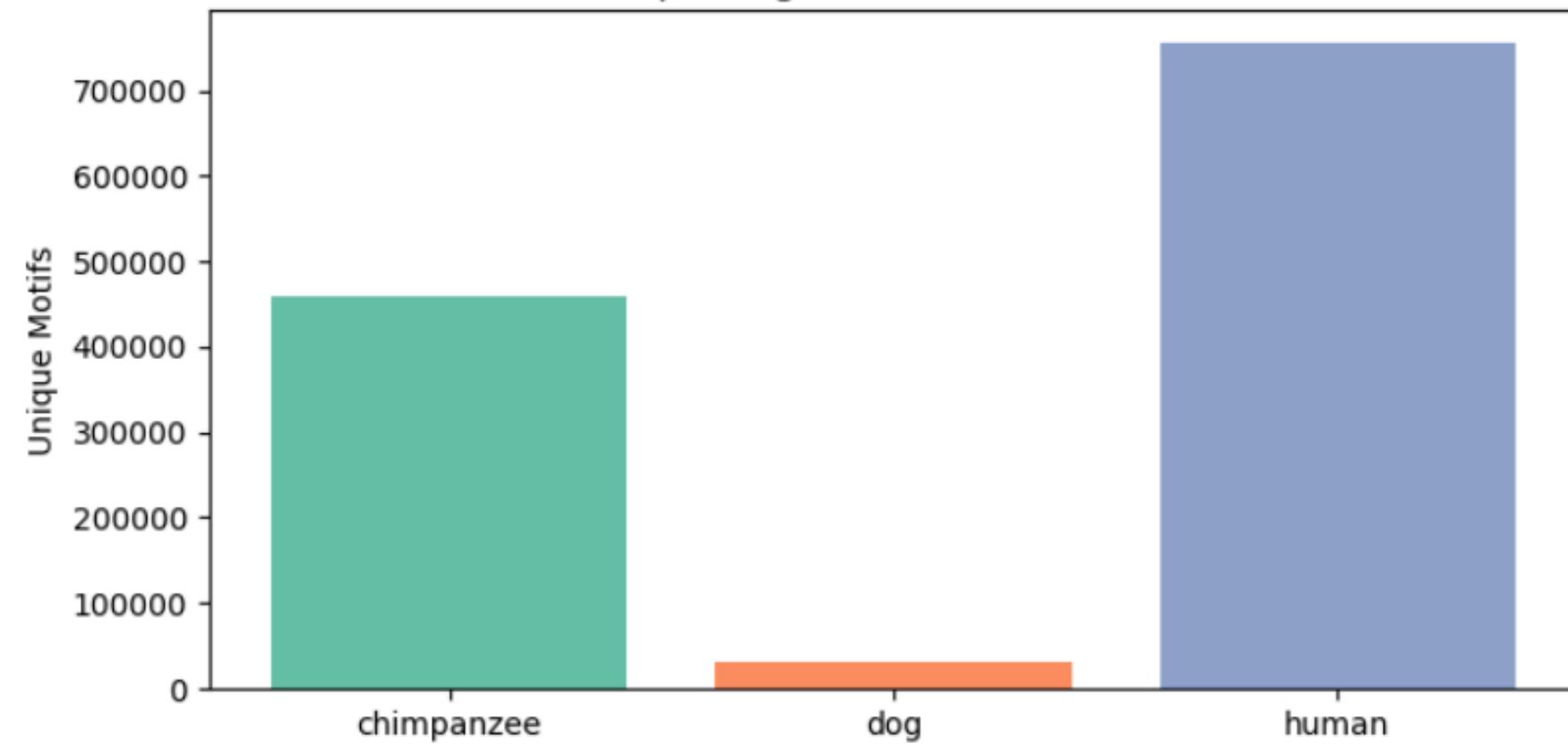


כלב

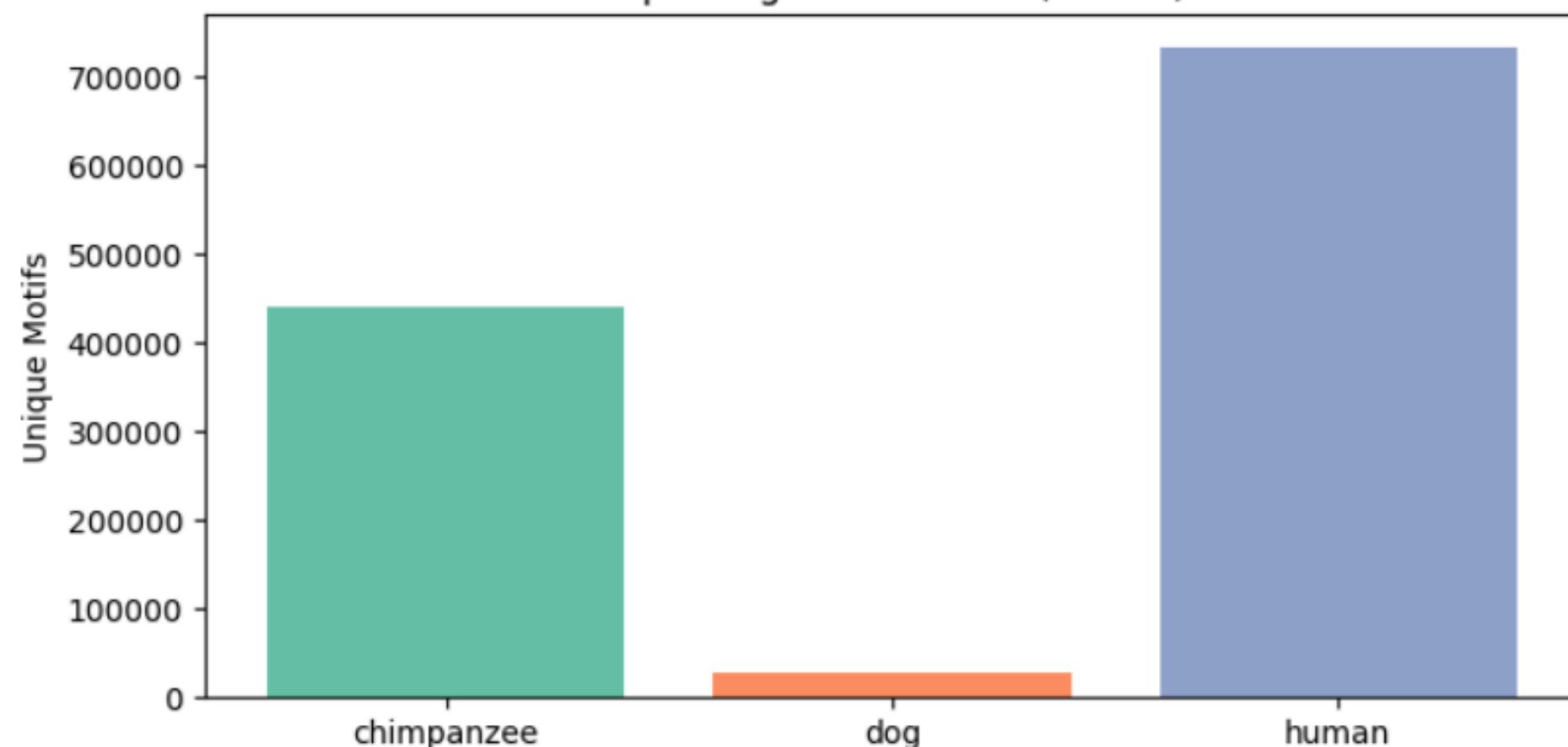
Repeating Motifs Count (L=30)



Repeating Motifs Count (L=70)



Repeating Motifs Count (L=120)



[ AT LENGTH L = 30 ]

Number of motifs shared by ALL 3 species: 759

Pairwise Shared Counts:

- chimpanzee vs dog: 1071 motifs
- chimpanzee vs human: 333742 motifs
- dog vs human: 1664 motifs

[ AT LENGTH L = 70 ]

Number of motifs shared by ALL 3 species: 62

Pairwise Shared Counts:

- chimpanzee vs dog: 62 motifs
- chimpanzee vs human: 265054 motifs
- dog vs human: 151 motifs

[ AT LENGTH L = 120 ]

Number of motifs shared by ALL 3 species: 0

Pairwise Shared Counts:

- chimpanzee vs dog: 0 motifs
- chimpanzee vs human: 202243 motifs
- dog vs human: 25 motifs

## 1. האדם והשימפנזה – "אחים גנטיים"

ב- $L=30$  יש 333,742 מוטיבים משותפים; גם ב- $L=120$  יש 202,243 מוטיבים זהים.

מסקנה: רוב מנגנוני הבקרה והחלבונים זהים לחלוtein, מאושך קרבה אבולוציונית מיידית.

## 2. הכלב – "הזר המוכר"

מול האדם יש רק 1,664 מוטיבים משותפים ב- $L=30$ .

מסקנה: מערכות הבקרה של הכלב נבנו באופן עצמאי ושוונה מהפרימטים, והוא חולק איתנו רק את ה"לביה" של החסרים.

## 3. הלביה המשותפת לכלם (CORE)

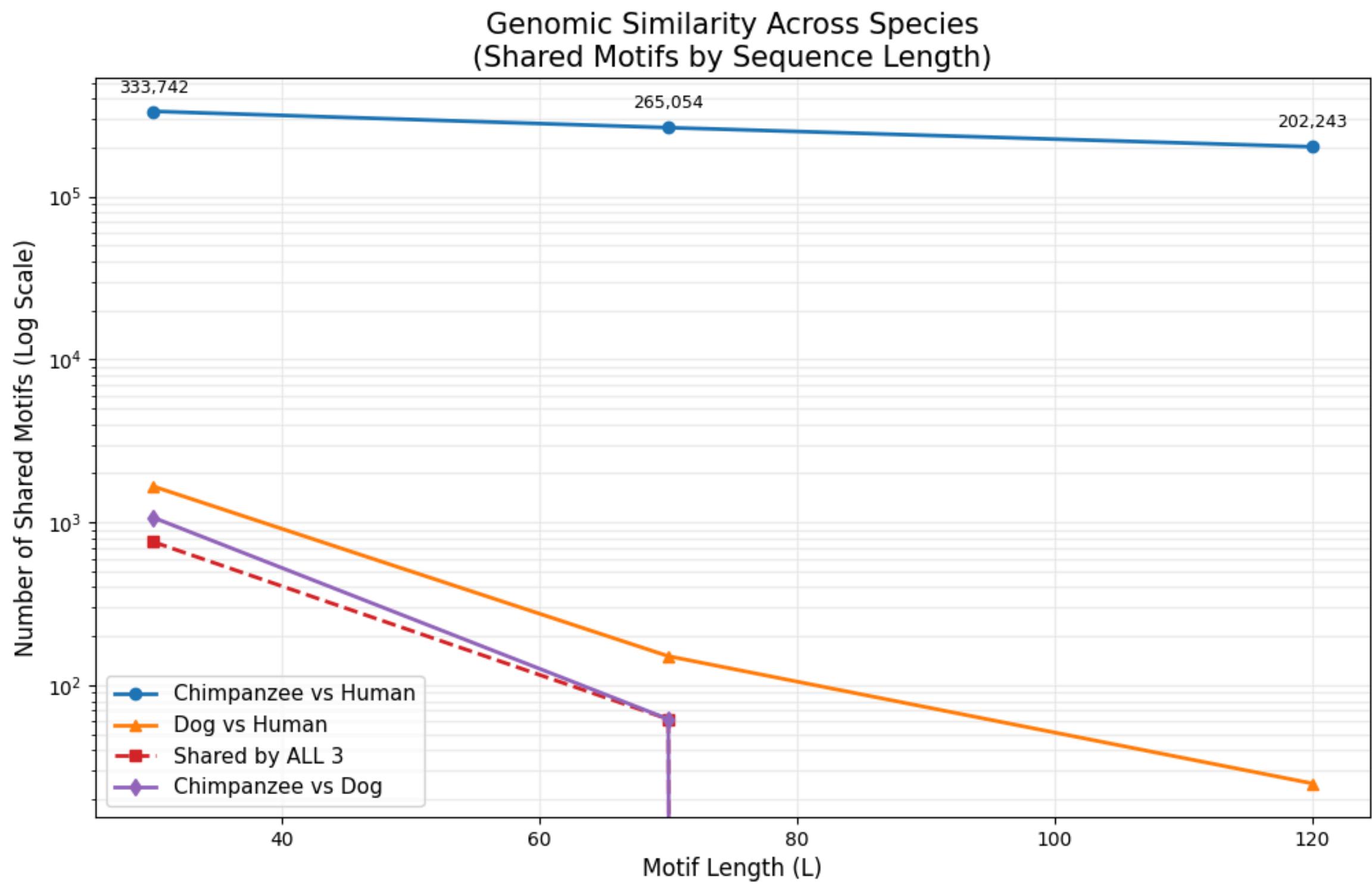
759 מוטיבים משותפים ב- $L=30$ , אך ב- $L=120$ .

מסקנה: אלו רצפים בסיסיים של יונקים – נשימה תאית, חלוקת תאים וSCPOL DNA. רצפים ארוכים שונים לכל מין.

## 4. חוק ה-120 – הגבול של הדמיון

בין השימפנזה לכלב אין אפילו רצף אחר זהה ( $L=120$ ).

מסקנה: פרימטים יוצאים דופן מבחינה גנטית, בעוד כלב מייצג סדרה אבולוציונית נפרדת.



**אדם ושימפנזה:**  
דמיוון מוחלט כמעט בכל אורך.  
**הכלב:**  
דמיוון נמוך כבר ברצפים  
קצרים, מוחלט ברצפים  
ארוכים.  
**הLIBRARY של 759 מוטיבים:**  
פרוטוקול בסיסי שמאפיין את  
כל היעונקים.

# תודה על ההקשבה!

