Citizen science and Phylo

Silja Laht

Bioinformatics journal club 26.03.2012

Citizen science

Scientific research conducted, in whole or in part, by amateur or nonprofessional scientists.

spring alive

- gather data that will be analyzed by professional researchers (bird counting)
- analyze data (SETI Live, Galaxy Zoo, Ancient Lives)
- share computer resources (SETI@home)
- FoldIt and Phylo - computer games that solve scientific problems

STARDUST @ HOME

Field Expedition: Mongolia

ANCIENT LIVES

GALAXY ZOO

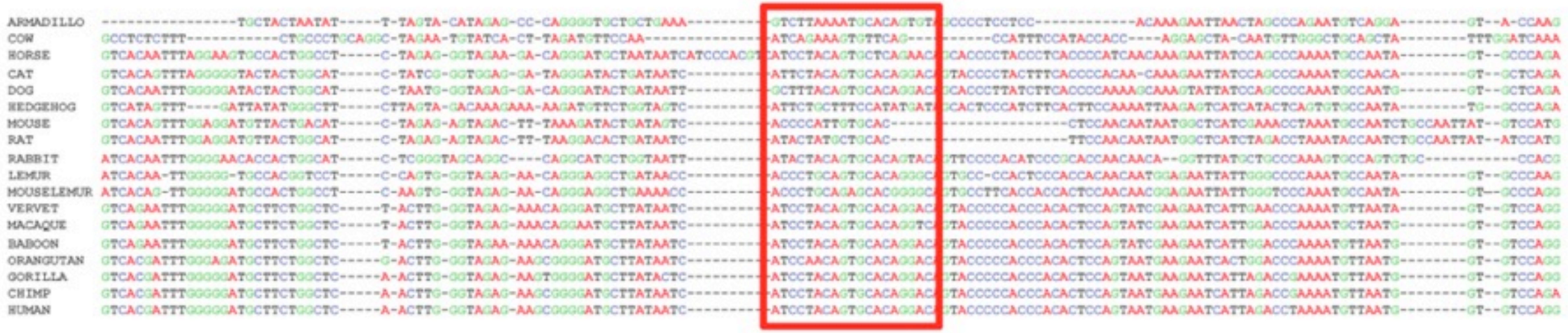
foldit BETA
Solve Puzzles for Science

Phylo: A Citizen Science Approach for Improving Multiple Sequence Alignments

Kawrykow et al. PLoS ONE, March 2012

- Aligning multiple sequences is a fundamental question in bioinformatics, but a difficult computational task
- Humans excel at visual pattern recognition
- NP-hard computational problem embedded in casual game and played by people without scientific training.

Whole-genome multiple alignment (UCSC 44-way Multiz MSA)



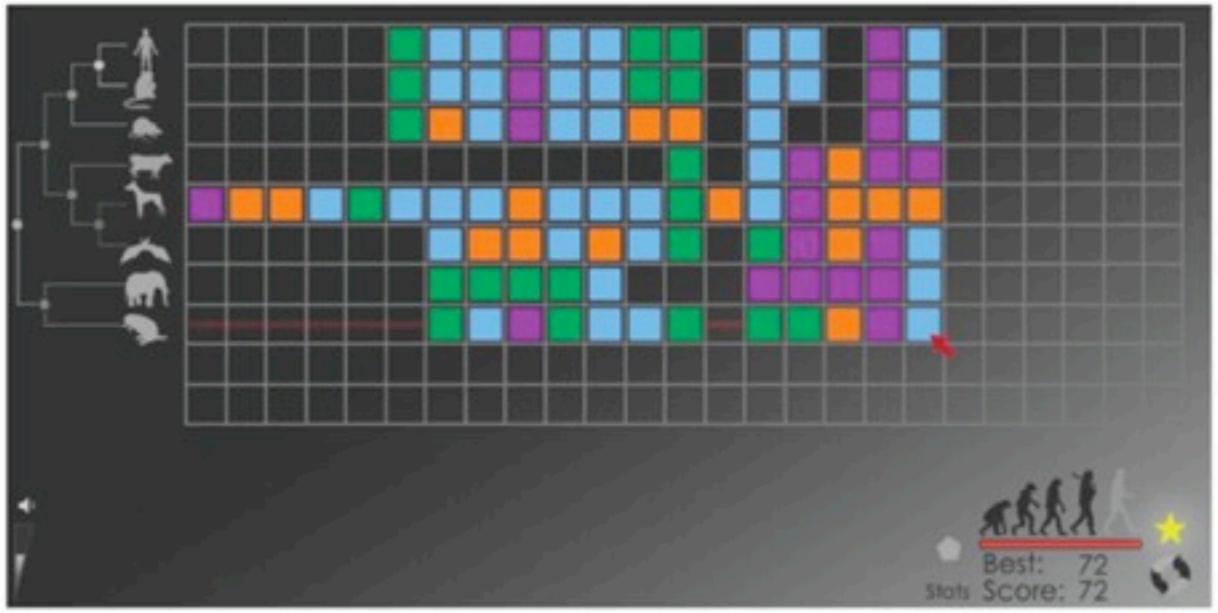
Extract dubious alignment region

Reinsertion into original alignment + Evaluation

Database of interesting puzzles

Video game:

- Computers
- Tablets
- Cell phones



http://phylo.cs.mcgill.ca

Multiz

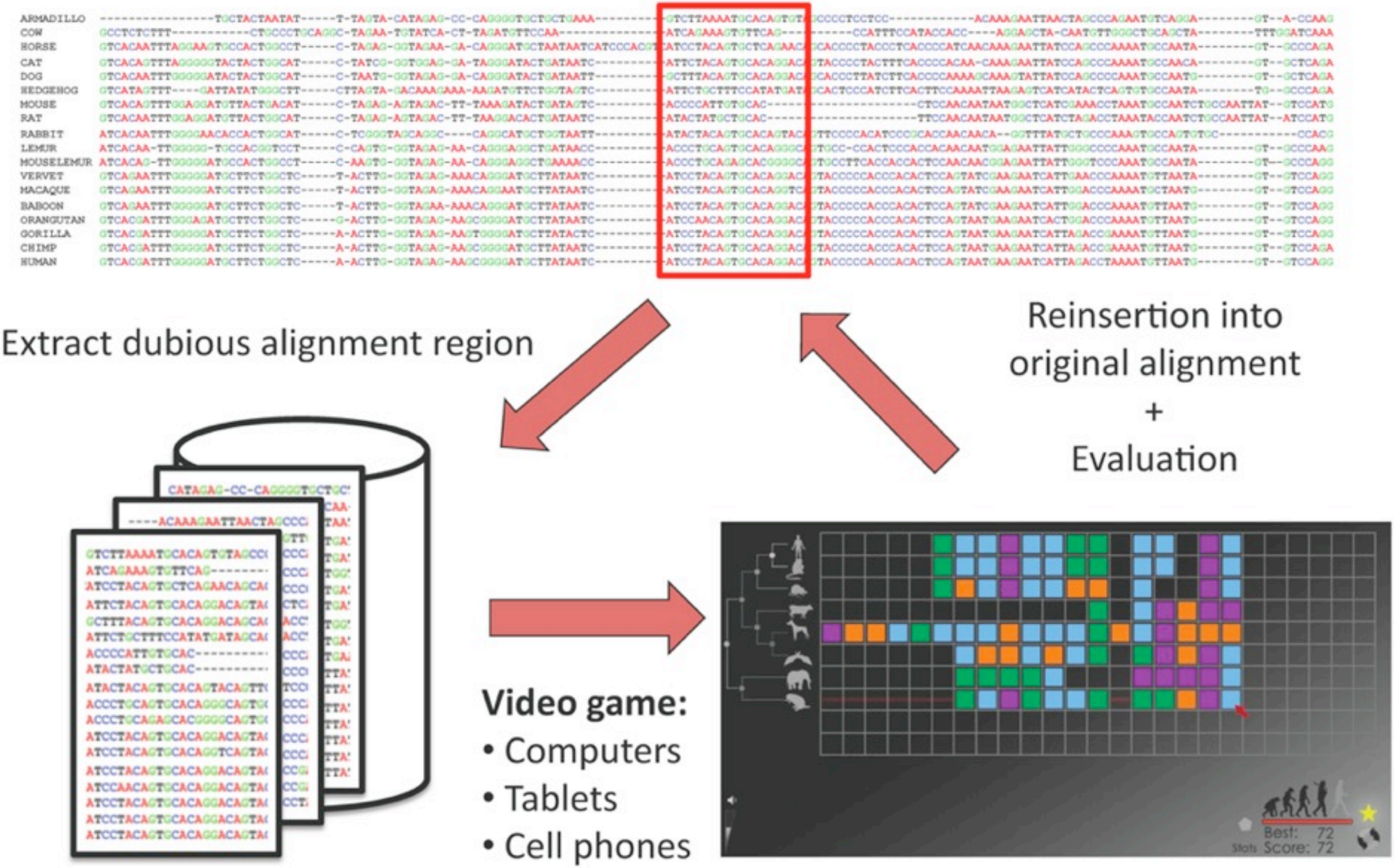
Blanchett et al. , Genome Research, 2004
“Aligning multiple genomic sequences with the threaded blockset aligner”

- Threaded blockset aligner (TBA)
- Designed for aligning genome sequences of many species as a set of local MSAs
- Multiz is a phylogenetic tree directed multiple alignment.
- Multiz can be used to align highly rearranged or incompletely sequenced genomes
- http://www.bx.psu.edu/miller_lab/

Phylo

- 44-vertebrate whole-genome alignment
- short alignment regions with signs of misalignment of disease associated promoter regions
- one of these likely misalignments is represented to player like a puzzle-game
- upon completion the player’s solution is sent to server, reinserted into alignment, evaluated and if better than original, retained

Whole-genome multiple alignment (UCSC 44-way Multiz MSA)



Extract dubious alignment region

Reinsertion into original alignment + Evaluation

Video game:

- Computers
- Tablets
- Cell phones

Database of interesting puzzles

http://phylo.cs.mcgill.ca

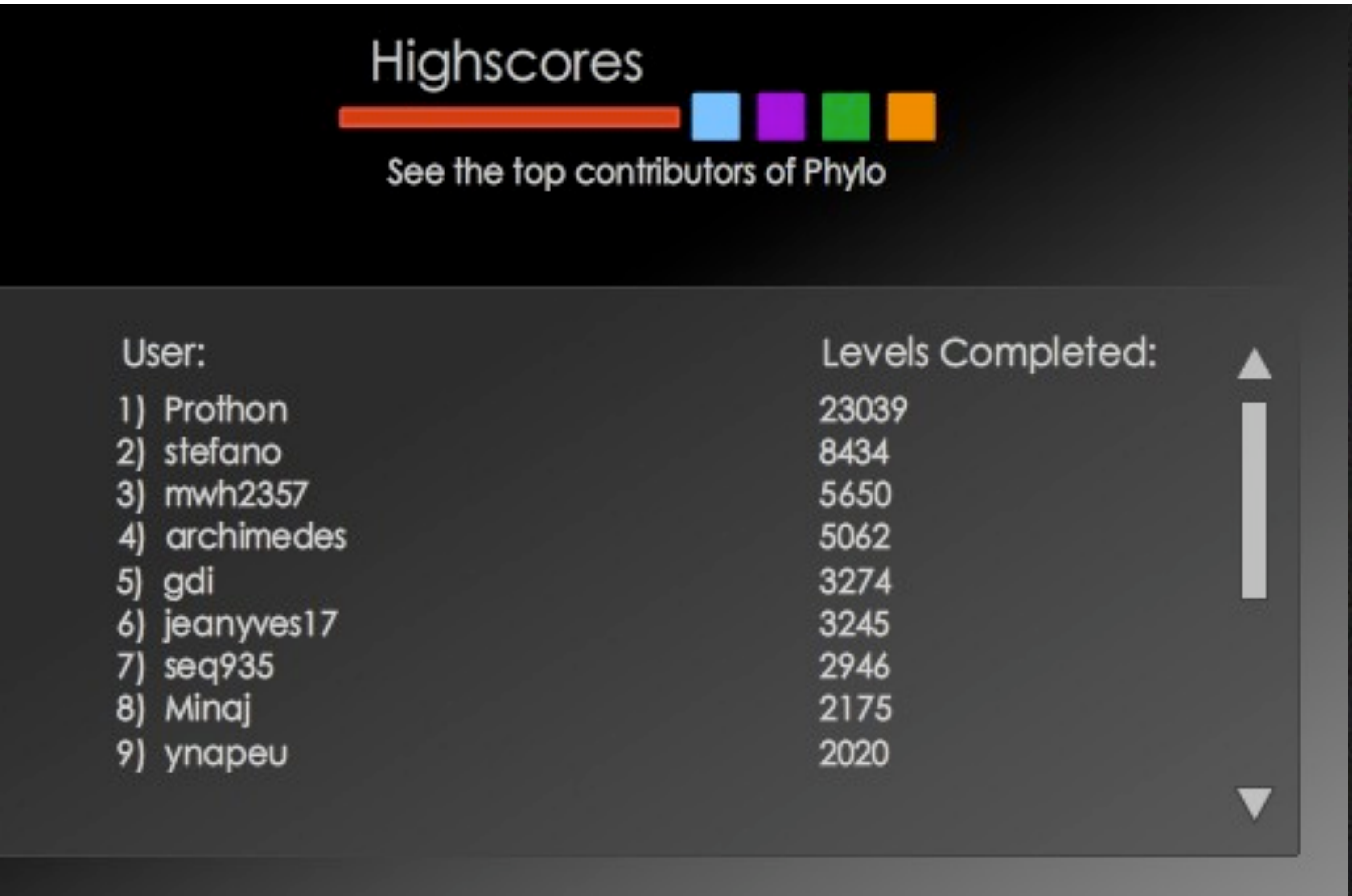
Data selection

- Human promoters associated to genes with known implications in various diseases in OMIM database
- 1 kb region upstream of annotated transcription start site
- 24-column regions likely to contain alignment errors and suitable to make interesting puzzles
- At most 8 species (distant) is kept in each puzzle
- 739 puzzles were created
- It's possible to send your own puzzle to be solved at Phylo

Phylo- game

- Sequences are progressively added starting with 2
- Scoring is based on ancestral sequences predicted from alignment
- You have to at least as good as the Multiz alignment score to proceed
- You can revise any part of the alignment at any time
- Each stage must be completed within a certain time limit
- The top 20 users with most puzzles solved are shown in highscores

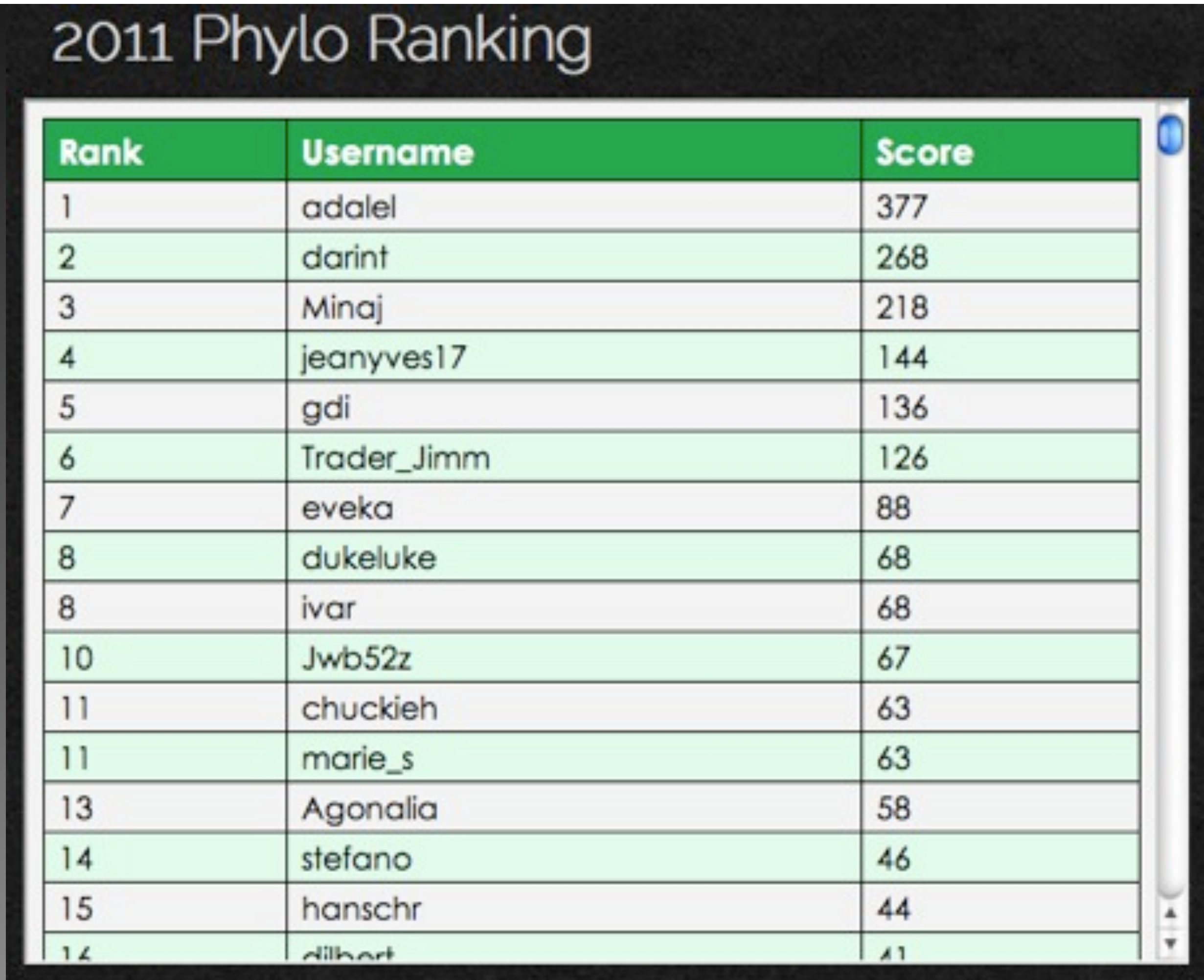
- Released online November 29th 2010.
- About 300 puzzle solutions per day
- Top player has completed 23 000 levels
- 821 different users obtained the best score for at least one puzzle



Highscores

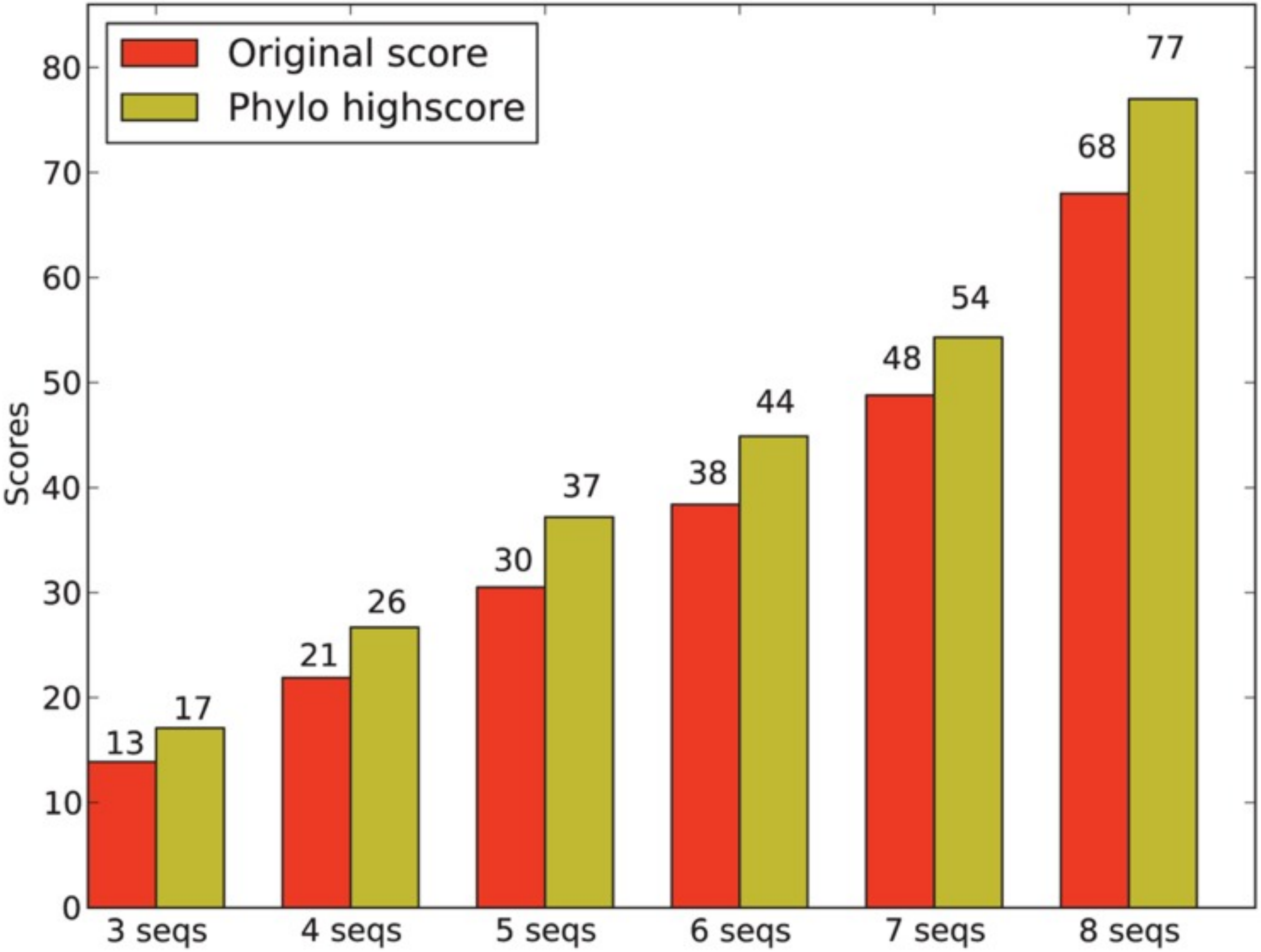
See the top contributors of Phylo

User:	Levels Completed:
1) Prothon	23039
2) stefano	8434
3) mwh2357	5650
4) archimedes	5062
5) gdi	3274
6) jeanyves17	3245
7) seq935	2946
8) Minaj	2175
9) ynapeu	2020



2011 Phylo Ranking

Rank	Username	Score
1	adalel	377
2	darint	268
3	Minaj	218
4	jeanyves17	144
5	gdi	136
6	Trader_Jimm	126
7	eveka	88
8	dukeluke	68
8	ivar	68
10	Jwb52z	67
11	chuckieh	63
11	marie_s	63
13	Agonalia	58
14	stefano	46
15	hanschr	44
16	dilbert	41



(a) Average improvement of puzzle scores per level

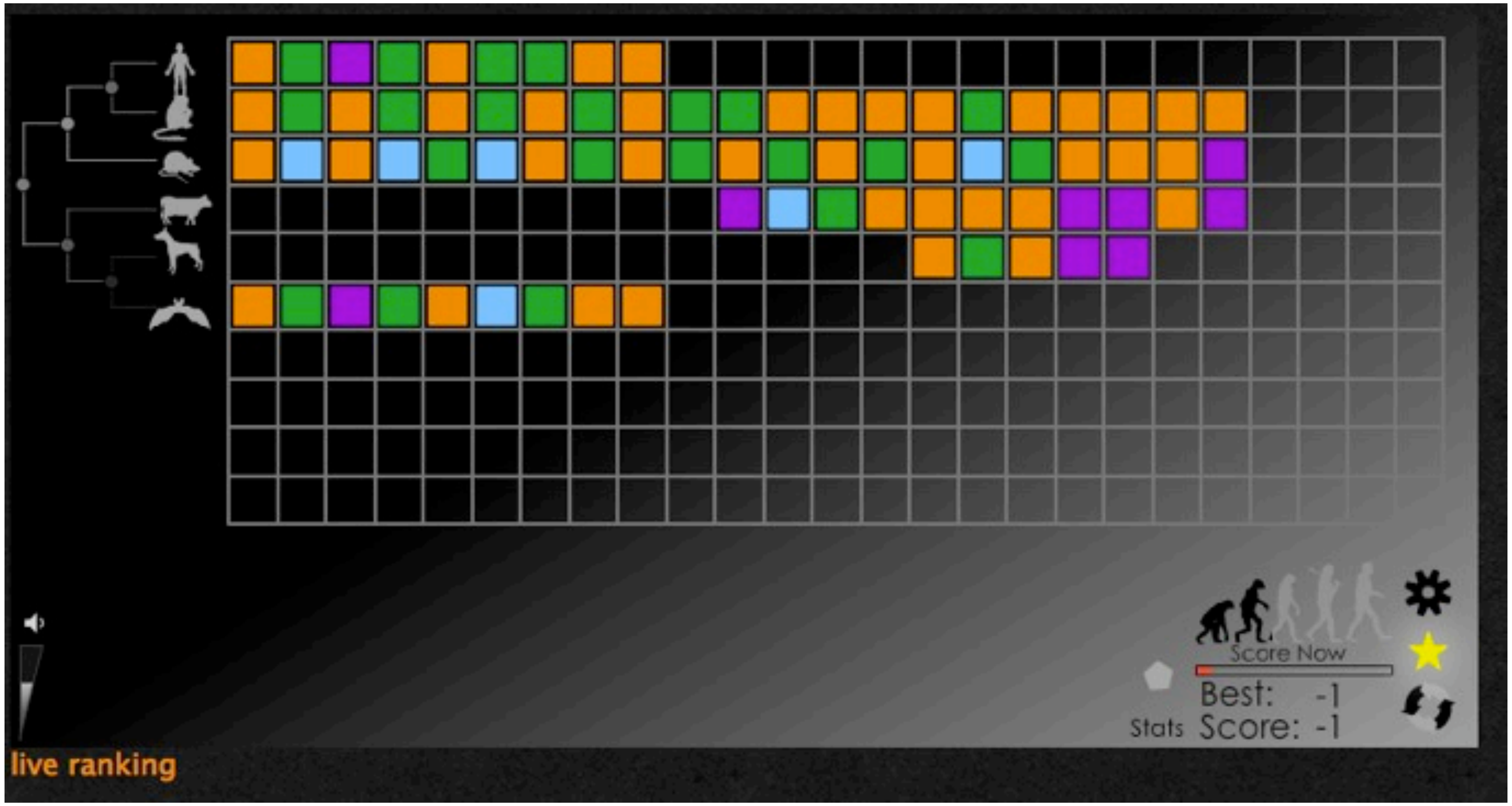
Improvements in alignments

- The best Phylo alignment outscored Multiz for 70% of the blocks
- Phylo alignments were better than original Multiz or local de novo alignment in 36% of puzzles
- De novo alignments were best in 46% of cases
- Relative score increase >10% in 78% of cases
- The full block alignment score is improved in 55% of cases. The more sequences used for puzzle and the more solutions, the better results

Future plans

- Increase the size of the puzzles
- Addition of flanking regions to increase the correlation between Phylo puzzle scores and final alignment scores
- Improved system for selecting puzzles for players

Billions of human-brain peta-flops of computation are waisted daily playing games that do not contribute to advancing knowledge.



🟧	🟩	🟪	🟩	🟧	🟩	🟩	🟧	🟧																
🟧	🟩	🟧	🟩	🟧	🟩	🟧	🟩	🟧	🟩	🟩	🟧	🟧	🟧	🟧	🟩	🟧	🟧	🟧	🟧	🟧				
🟧	🟦	🟧	🟦	🟩	🟦	🟧	🟩	🟧	🟩	🟧	🟩	🟧	🟩	🟧	🟦	🟩	🟧	🟧	🟧	🟪				
										🟪	🟦	🟩	🟧	🟧	🟧	🟧	🟪	🟪	🟧	🟪				
														🟧	🟩	🟧	🟪	🟪						
🟧	🟩	🟪	🟩	🟧	🟦	🟩	🟧	🟧																

Best: -1

Stats Score: -1