

INCOMPLETE DRAFT: Smashing single cells into k -mer sketches

This manuscript ([permalink](#)) was automatically generated from [czbiohub/de-novo-orthology-paper@ab230be](#) on February 27, 2020.

Authors

- **Olga Borisovna Botvinnik**

 [0000-0003-4412-7970](#) ·  [olgabot](#) ·  [olgabot](#)

Data Sciences Platform, Chan Zuckerberg Biohub

- **Venkata Naga Pranathi Vemuri**

 [0000-0002-5748-9594](#) ·  [pranathivemuri](#) ·  [pranuvemuri](#)

Data Sciences Platform, Chan Zuckerberg Biohub

- **Phoenix Aja Logan**

 [0000-0003-4581-0552](#) ·  [phoenixAja](#) ·  [phoenixlogan](#)

Data Sciences Platform, Chan Zuckerberg Biohub

- **Saba Nafees**

 [0000-0002-3292-7703](#) ·  [snafees](#) ·  [sabanafeesTTU](#)

Data Sciences Platform, Chan Zuckerberg Biohub; Department of Biological Sciences, Texas Tech University;
Department of Mathematics & Statistics, Texas Tech University

- **Jim Karkanias**

 [0000-0002-8057-6055](#) ·  [jkensai](#) ·  [jkarkanias](#)

Data Sciences Platform, Chan Zuckerberg Biohub

Abstract

We introduce `kmermaid`, a lightweight computational method for quantifying similarity between RNA-seq transcriptomes across species. As `kmermaid` skips both alignment and gene orthology assignment, it can be applied to transcriptomes from organisms with no or poorly annotated genomes, and can perform the task of quantifying similarities much faster than current methods without loss of sensitivity. By enabling analyses across divergent species' transcriptomes in an orthology-, genome- and gene annotation-agnostic manner, `kmermaid` illustrates the potential of non-model organisms in building the cell type evolutionary tree of life.

Single-cell RNA-sequencing is a powerful technology for identifying cell types in a variety of species. However, the task of identifying even known cell types in species with poorly annotated genomes is nontrivial, as orthologous gene mapping remains an unsolved problem [1,2] and most of the 8.7 million Eukaryotic species on Earth have no sequenced genome [3]. Thus, there is an unmet need to quantitatively compare single-cell transcriptomes across species, without the need for orthologous gene mapping, gene annotations, or a reference genome. Short, k -long sequence substrings, or k -mers, have been proposed for comparing single cells [4] as they are a fast, simple way to create cell-cell similarities. Combining k -mers from putatively translated RNA-seq reads with reduced amino acid alphabets [5,5,6,7], we developed a method to identify known cell types in a non-model organism using a reference atlas from another organism, without the need for a genome or gene annotation from the non-model organism. This method relies solely on divergence time between species, which we show can be estimated from RNA-seq nucleotide k -mers (Supplemental Figure

[???]). We benchmark the genome-agnostic method on the Quest for Orthologs Opisthokonta dataset, showing that k -mers from reduced amino acid alphabets are sufficient to estimate orthology. Using human amino acid sequences, we show that one can extract putative protein-coding reads from 239 Opisthokonta species in ENSEMBL, and present the best k -mer size and alphabet for different divergence times. We apply this method on a bulk comparative transcriptomic dataset consisting of nine species and six tissues [10], showing that we can recapitulate the results as using only reads mapping to 1:1 orthologs of protein-coding genes, and we are able to resolve ... which can only be seen by using the reference-free k -mer method.

We then show that reduced amino acid protein k -mers can be used for comparing single-cell transcriptomes across mouse, human, and zebrafish hematopoiesis. We find that using k -mers has the advantage of resolving ... in comparison to using read counts from 1:1 gene orthologs. Using k -mers, we were able to resolve cell types ... , which was hidden using read counts alone. Thus, we have shown the reference-free method using the k -mers from single cells is a novel, annotation-agnostic method for comparing cells across species that is capable of identifying cell states unique to a particular organism, helping to build the cell type evolutionary tree of life.

Figure 1 – Overview of `kmermaid` pipeline

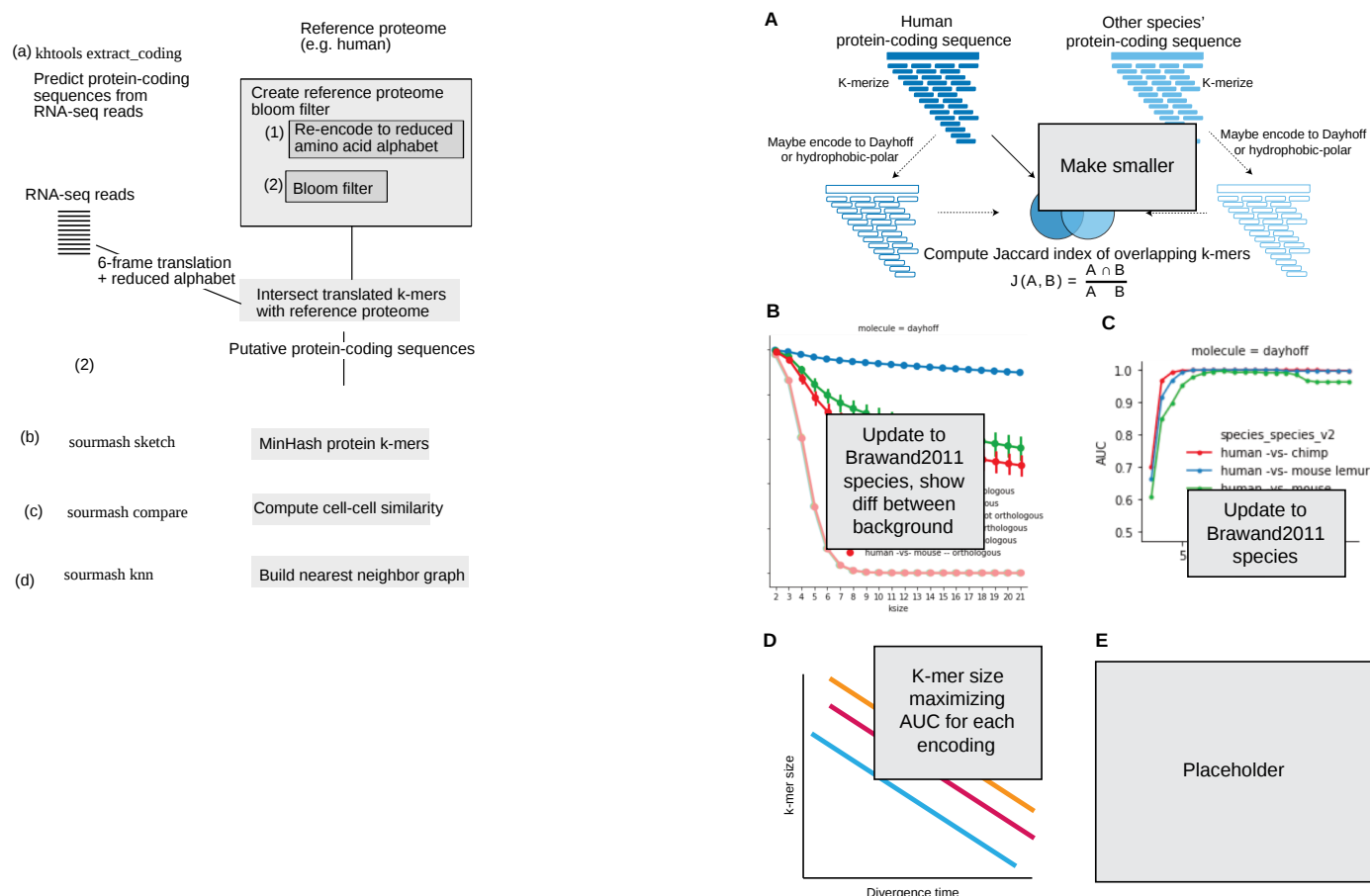


Figure 1: A. Overview of the `kmermaid` pipeline. **(a, b, c)** `kmermaid` consists of a protein-coding prediction phase **(a)** that is invoked by the command `khtools extract_coding`, a k -mer sketch computation phase **(b)** invoked by the command `sourmash sketch`, a signature similarity comparison phase **(c)** invoked by the command `sourmash compare`, and an optional database-creation phase **(d)** invoked by the command `sourmash index`. The coding prediction phase has three components: (1) six-frame translation, removal of stop-codon frames, and subsequent k -merization of RNA-sequencing reads; (2) a degenerate protein alphabet which allows for protein-coding detection from a wide variety of species; (3) a bloom filter containing known protein-coding sequences from a well annotated organism; and (4) computation of the Jaccard index of translated RNA-seq reading frames. The sketch computation phase involves randomly subsetting the degenerate peptide k -mers using a MinHash algorithm. The sketch comparison phase consists of computing the Jaccard intersection of MinHashed degenerate peptide k -mers between all pairs of samples.

To determine whether short segments of sequences could detect gene orthologues, we k -merized orthologous genes derived from the ENSEMBL version 97 [11] COMPARA database [12] (Figure [1]). We compared human protein sequences to orthologous chimpanzee, mouse, (orangutan, bonobo, gorilla, macaque, opossum, platypus, chicken) protein sequences, as these are species used in [10]. As a background, we randomly chose 10 non-orthologous genes relative to the human gene. In addition to k -merizing the protein-coding sequence, we also re-encoded the protein-coding sequence into a six-letter Dayhoff alphabet [13], a nine-letter encoding [8], and a two-letter hydrophobic-polar encodings [14,15], show in Table [1].

We found that, consistent with previous knowledge, that 1:1 orthologues had higher k -mer similarities as determined by the Jaccard Index. This approach is similar to SwiftOrtho [8], a k -mer based orthology relationship finder.

Additionally, more recently diverged genes had higher k -mer similarity as well.

Outline

- Kmers can approximate orthologies
 - Jaccard similarity of orthologues is higher than non-orthologues
 - Benchmarking using <https://orthology.benchmarkservice.org/cgi-bin/gateway.pl>
 - Finding orthologues
 - Gold standard
 - ENSEMBL COMPARA
 - Quest for Orthologs consortium, Altenhoff, A. M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D. A., DeLuca, T., et al. (2016). Standardized benchmarking in the quest for orthologs. Nature Methods, 13(5), 425–430. <http://doi.org/10.1038/nmeth.3830> [16]
 - Orthologous groups/Conserved Domain Database [17]

Figure 2 – k -mers from lossily-encoded putative protein-coding reads faithfully pull out reads from protein-coding genes within amniotes

A

B

C

Use all
Brawand2011
species

Figure 2: Figure 2.

- Overview of kmermaid pipeline
 - Comparison of tissue across species
 - Partition reads to coding/noncoding bins
 - MinHash the Dayhoff-encoded coding sequences
 - Jaccard similarity on the MinHashes
- Which reads are found to have coding features but didn't map to the genome?
- Do these features map to novel genes or gene fusions?
- Kmers can find correct reading from of RNA-seq reads

- Human peptides → human, chimp, bonobo, orangutan, gorilla, macaque, mouse, opossum, platypus, chicken RNAseq from Brawand2011 data
- Comparison to other methods: RNASamba [[18](#)]

Figure 3 – k -mers can pull out only reads from transcription factors and Amniotes can be compared on the MinHashes of their protein-coding sequences

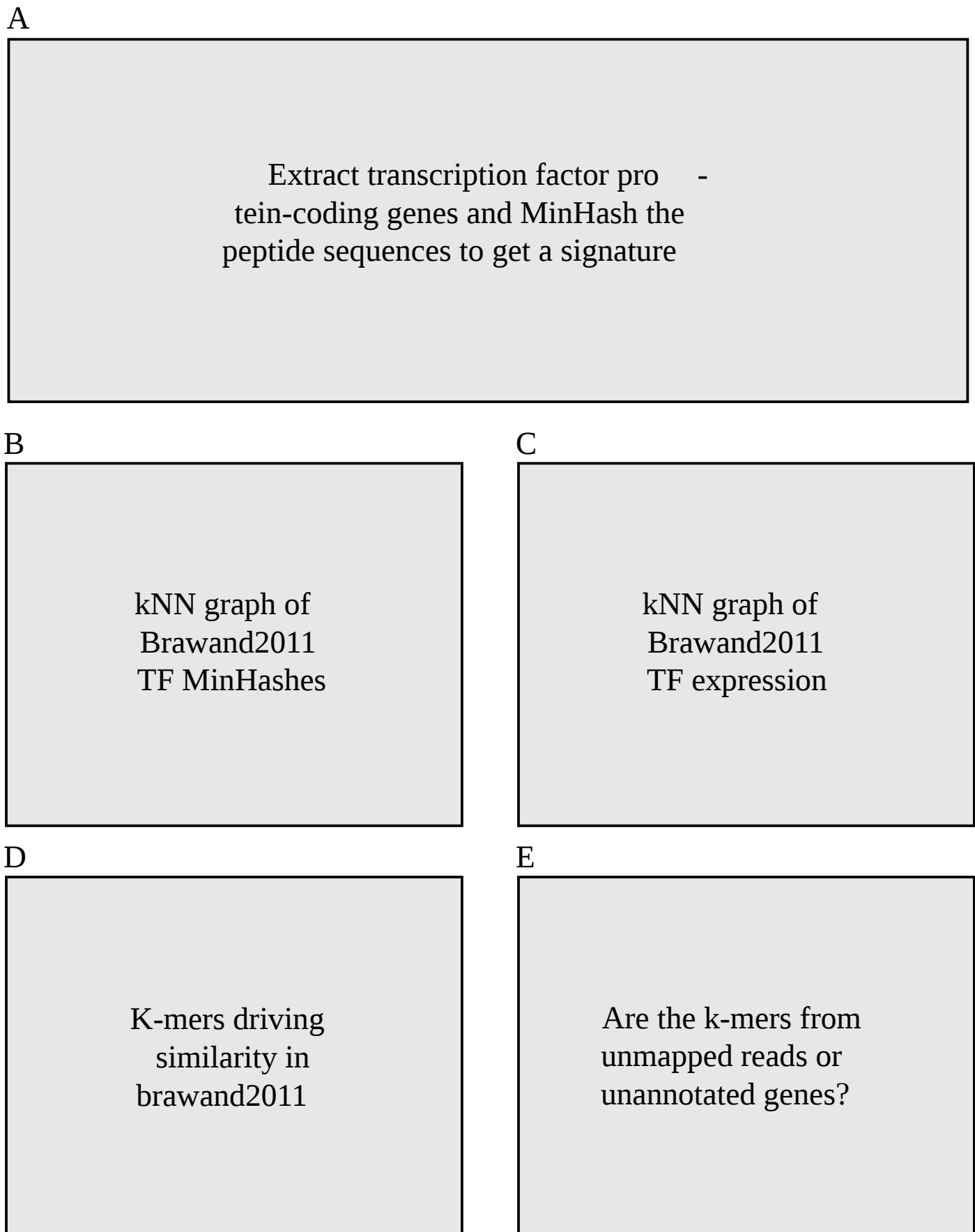


Figure 3: Figure 3.

- Kmers can find only transcription factor reads of TFs from RNA-seq reads
 - Human peptides → human, chimp, bonobo, orangutan, gorilla, macaque, mouse, opossum, platypus, chicken RNAseq from Brawand2011 data

kmermaid implements the concept of lightweight orthology assignment using k-mers to the problem of cross-species RNA-seq analyses and achieves unprecedented speed of analysis. By removing the orthology inference step, **kmermaid** opens up the possibility of finding shared and divergent tissue and cell types across a broad range of species, paving the way for evolutionary analyses of cell types across species. **kmermaid** can be used in *de novo* setting for non-model organisms, finding similar cell types within an organism, or finding similar cell types relative to a reference organism, without the need for a reference genome or transcriptome. The memory usage of **kmermaid** is quite low, using only 50MB for extracting coding sequences and 50MB for assigning protein k-mer signatures. As the number of RNA-seq datasets, especially single-cell RNA-seq datasets continues to grow, we expect **kmermaid** to be widely used for identifying cell types in non-model organisms.

kmermaid is free and open-source software and is available as Supplementary Data and at <http://github.com/czbiohub/kmermaid> and as a scalable Nextflow workflow at <http://github.com/nf-core/nf-kmermaid>.

Some potential references

Gene expression evolution through duplications

- Farre, D., & Alba, M. M. (2010). Heterogeneous Patterns of Gene-Expression Diversification in Mammalian Gene Duplicates. *Molecular Biology and Evolution*, 27(2), 325–335. <http://doi.org/10.1093/molbev/msp242> [19]
- Thornton, J. W., & DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annual Review of Genomics and Human Genetics*, 1(1), 41–73. <http://doi.org/10.1146/annurev.genom.1.1.41> [20]
- Farre, D., & Alba, M. M. (2010). Heterogeneous Patterns of Gene-Expression Diversification in Mammalian Gene Duplicates. *Molecular Biology and Evolution*, 27(2), 325–335. <http://doi.org/10.1093/molbev/msp242> [19]

Taxa-restricted genes

- Human-specific genes in fetal neocortex Florio, M., Heide, M., Pinson, A., Brandl, H., Albert, M., Winkler, S., et al. (2018). Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife*, 7, D635. <http://doi.org/10.7554/eLife.32332> [21]
- Insects – Santos, M. E., Le Bouquin, A., Crumière, A. J. J., & Khila, A. (2017). Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science*, 358(6361), 386–390. <http://doi.org/10.1126/science.aan2748> [22]

Correlated evolution of celltypes?

- Liang, C., Musser, J. M., Cloutier, A., Prum, R. O., & Wagner, G. P. (2018). Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes. *Genome Biology and Evolution*, 10(2), 538–552. <http://doi.org/10.1093/gbe/evy016> [23]

Cell type homology

- Thornton, J. W., & DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annual Review of Genomics and Human Genetics*, 1(1), 41–73. <http://doi.org/10.1146/annurev.genom.1.1.41> [20]
- Tschopp, P., & Tabin, C. J. (2017). Deep homology in the age of next-generation sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1713), 20150475–8. <http://doi.org/10.1098/rstb.2015.0475> [24]
- Hejnal, A., & Lowe, C. J. (2015). Embracing the comparative approach: how robust phylogenies and broader developmental sampling impacts the understanding of nervous system evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684), 20150045–16. <http://doi.org/10.1098/rstb.2015.0045> [25]
- Santos, M. E., Le Bouquin, A., Crumière, A. J. J., & Khila, A. (2017). Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. *Science*, 358(6361), 386–390. <http://doi.org/10.1126/science.aan2748> [22]
- Mammalian decidual cell

Cell type evolution

- Erkenbrack, E. M., Maziarz, J. D., Griffith, O. W., Liang, C., Chavan, A. R., Nnamani, M. C., & Wagner, G. P. (2018). The mammalian decidual cell evolved from a cellular stress response. *PLOS Biology*, 16(8), e2005594–27. <http://doi.org/10.1371/journal.pbio.2005594> [26]

Methods

Methods go here.

Experimental

Primate brain organoid protocols

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Single-cell capture of primate brain organoids

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Long read library prep

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Short read library prep

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Sequencing

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Computational

Caption for Supplemental figure 1

k-mer comparison of orthologous genes

We used ENSEMBL version 97. We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Extraction of putative coding reads from RNA-seq

We did things. One sentence per line. Prefer DOI for references, but for Biorxiv use the URL. DOI example: [16]. Biorxiv example: [27]. Multiple citations per line example: [16,27].

Supplemental Methods

Table 1: Dayhoff and hydrophobic-polar encodings are a reduced amino acid alphabet allowing for permissive cross-species sequence comparisons. For example, the amino acid sequence SASHAFIERCE would be Dayhoff-encoded to bbbdbfecdac , and HP-encoded to phpphhpppp .

Amino acid	Property	Dayhoff	Hydrophobic-polar (HP)
C	Sulfur polymerization	a	p
A, G, P, S, T	Small	b	A, G, P: h S,T: p
D, E, N, Q	Acid and amide	c	p
H, K, R	Basic	d	p
I, L, M, V	Hydrophobic	e	h
F, W, Y	Aromatic	f	h

References

1. The origin and evolution of cell types

Detlev Arendt, Jacob M. Musser, Clare V. H. Baker, Aviv Bergman, Connie Cepko, Douglas H. Erwin, Mihaela Pavlicev, Gerhard Schlosser, Stefanie Widder, Manfred D. Laubichler, Günter P. Wagner
Nature Reviews Genetics (2016-11-07) <https://doi.org/f9b62x>
DOI: [10.1038/nrg.2016.127](https://doi.org/10.1038/nrg.2016.127) · PMID: [27818507](https://pubmed.ncbi.nlm.nih.gov/27818507/)

2. How Single-Cell Genomics Is Changing Evolutionary and Developmental Biology

John C. Marioni, Detlev Arendt
Annual Review of Cell and Developmental Biology (2017-10-06) <https://doi.org/ggb632>
DOI: [10.1146/annurev-cellbio-100616-060818](https://doi.org/10.1146/annurev-cellbio-100616-060818) · PMID: [28813177](https://pubmed.ncbi.nlm.nih.gov/28813177/)

3. How Many Species Are There on Earth and in the Ocean?

Camilo Mora, Derek P. Tittensor, Sina Adl, Alastair G. B. Simpson, Boris Worm
PLoS Biology (2011-08-23) <https://doi.org/fpr4z8>
DOI: [10.1371/journal.pbio.1001127](https://doi.org/10.1371/journal.pbio.1001127) · PMID: [21886479](https://pubmed.ncbi.nlm.nih.gov/21886479/) · PMCID: [PMC3160336](https://pubmed.ncbi.nlm.nih.gov/PMC3160336/)

4. K-mer counting with low memory consumption enables fast clustering of single-cell sequencing data without read alignment

Christina Huan Shi, Kevin Y. Yip
bioRxiv (2019-08-02) <https://www.biorxiv.org/content/10.1101/723833v1>
DOI: [10.1101/723833](https://doi.org/10.1101/723833)

5. Reduced amino acid alphabets exhibit an improved sensitivity and selectivity in fold assignment

Eric L. Peterson, Jané Kondev, Julie A. Theriot, Rob Phillips
Bioinformatics (2009-04-07) <https://doi.org/btqmpnp>
DOI: [10.1093/bioinformatics/btp164](https://doi.org/10.1093/bioinformatics/btp164) · PMID: [19351620](https://pubmed.ncbi.nlm.nih.gov/19351620/) · PMCID: [PMC2732308](https://pubmed.ncbi.nlm.nih.gov/PMC2732308/)

6. Simplified amino acid alphabets for protein fold recognition and implications for folding

Lynne Reed Murphy, Anders Wallqvist, Ronald M. Levy
Protein Engineering, Design and Selection (2000-03) <https://doi.org/bdtngnh>
DOI: [10.1093/protein/13.3.149](https://doi.org/10.1093/protein/13.3.149) · PMID: [10775656](https://pubmed.ncbi.nlm.nih.gov/10775656/)

7. Local homology recognition and distance measures in linear time using compressed amino acid alphabets

R. C. Edgar
Nucleic Acids Research (2004-01-02) <https://doi.org/ckg5d4>
DOI: [10.1093/nar/gkh180](https://doi.org/10.1093/nar/gkh180) · PMID: [14729922](https://pubmed.ncbi.nlm.nih.gov/14729922/) · PMCID: [PMC373290](https://pubmed.ncbi.nlm.nih.gov/PMC373290/)

8. SwiftOrtho: A fast, memory-efficient, multiple genome orthology classifier

Xiao Hu, Iddo Friedberg
GigaScience (2019-10-01) <https://doi.org/ggcr5x>
DOI: [10.1093/gigascience/giz118](https://doi.org/10.1093/gigascience/giz118) · PMID: [31648300](https://pubmed.ncbi.nlm.nih.gov/31648300/) · PMCID: [PMC6812468](https://pubmed.ncbi.nlm.nih.gov/PMC6812468/)

9. Fast databank searching with a reduced amino-acid alphabet

Claudine Landès, Jean-Loup Risler
Bioinformatics (1994) <https://doi.org/cvrjmw>
DOI: [10.1093/bioinformatics/10.4.453](https://doi.org/10.1093/bioinformatics/10.4.453) · PMID: [7804879](https://pubmed.ncbi.nlm.nih.gov/7804879/)

10. The evolution of gene expression levels in mammalian organs

David Brawand, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, Angélica Liechti, Ayinuer Aximu-Petri, Martin Kircher, ... Henrik Kaessmann

Nature (2011-10) <https://doi.org/fcvk54>

DOI: [10.1038/nature10532](https://doi.org/10.1038/nature10532) · PMID: [22012392](https://pubmed.ncbi.nlm.nih.gov/22012392/)

11. Ensembl 2018

Daniel R Zerbino, Premanand Achuthan, Wasiiu Akanni, M Ridwan Amode, Daniel Barrell, Jyothish Bhai, Konstantinos Billis, Carla Cummins, Astrid Gall, Carlos García Girón, ... Paul Flicek

Nucleic Acids Research (2017-11-16) <https://doi.org/gcwg6r>

DOI: [10.1093/nar/gkx1098](https://doi.org/10.1093/nar/gkx1098) · PMID: [29155950](https://pubmed.ncbi.nlm.nih.gov/29155950/) · PMCID: [PMC5753206](https://pubmed.ncbi.nlm.nih.gov/PMC5753206/)

12. Ensembl comparative genomics resources

Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M. J. Searle, Ridwan Amode, Simon Brent, ... Paul Flicek

Database (2016) <https://doi.org/ggb9tv>

DOI: [10.1093/database/bav096](https://doi.org/10.1093/database/bav096) · PMID: [26896847](https://pubmed.ncbi.nlm.nih.gov/26896847/) · PMCID: [PMC4761110](https://pubmed.ncbi.nlm.nih.gov/PMC4761110/)

13. Atlas of protein sequence and structure

Margaret O Dayhoff

National Biomedical Research Foundation. (1969)

14. Physical biology of the cell

Rob Phillips, Julie Theriot, Jane Kondev, Hernan Garcia

Garland Science (2012)

15. Theory for the folding and stability of globular proteins

Ken A. Dill

Biochemistry (1985-03-12) <https://doi.org/fnj5k7>

DOI: [10.1021/bi00327a032](https://doi.org/10.1021/bi00327a032) · PMID: [3986190](https://pubmed.ncbi.nlm.nih.gov/3986190/)

16. Standardized benchmarking in the quest for orthologs

Adrian M Altenhoff, Brigitte Boeckmann, Salvador Capella-Gutierrez, Daniel A Dalquen, Todd DeLuca, Kristoffer Forslund, Jaime Huerta-Cepas, Benjamin Linard, Cécile Pereira, ... Christophe Dessimoz

Nature Methods (2016-04-04) <https://doi.org/f3rpzx>

DOI: [10.1038/nmeth.3830](https://doi.org/10.1038/nmeth.3830) · PMID: [27043882](https://pubmed.ncbi.nlm.nih.gov/27043882/) · PMCID: [PMC4827703](https://pubmed.ncbi.nlm.nih.gov/PMC4827703/)

17. <https://www.ebi.ac.uk/miriam/main/collections/MIR:00000119>

18. RNAsamba: coding potential assessment using ORF and whole transcript sequence information

Antonio P. Camargo, Vsevolod Sourkov, Marcelo F. Carazzolle

Cold Spring Harbor Laboratory (2019-04-28) <https://doi.org/ggdtxk>

DOI: [10.1101/620880](https://doi.org/10.1101/620880)

19. Heterogeneous Patterns of Gene-Expression Diversification in Mammalian Gene Duplicates

D. Farre, M. M. Alba

Molecular Biology and Evolution (2009-10-12) <https://doi.org/dxrtmd>

DOI: [10.1093/molbev/msp242](https://doi.org/10.1093/molbev/msp242) · PMID: [19822635](https://pubmed.ncbi.nlm.nih.gov/19822635/)

20. **GENEFAMILYEVOLUTION ANDHOMOLOGY: Genomics Meets Phylogenetics**
Joseph W. Thornton, Rob DeSalle
Annual Review of Genomics and Human Genetics (2000-09) <https://doi.org/bjp5pm>
DOI: [10.1146/annurev.genom.1.1.41](https://doi.org/10.1146/annurev.genom.1.1.41) · PMID: [11701624](https://pubmed.ncbi.nlm.nih.gov/11701624/)
21. **Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex**
Marta Florio, Michael Heide, Anneline Pinson, Holger Brandl, Mareike Albert, Sylke Winkler, Pauline Wimberger, Wieland B Huttner, Michael Hiller
eLife (2018-03-21) <https://doi.org/gc678k>
DOI: [10.7554/elife.32332](https://doi.org/10.7554/elife.32332) · PMID: [29561261](https://pubmed.ncbi.nlm.nih.gov/29561261/) · PMCID: [PMC5898914](https://pubmed.ncbi.nlm.nih.gov/PMC5898914/)
22. **Taxon-restricted genes at the origin of a novel trait allowing access to a new environment**
M. Emília Santos, Augustin Le Bouquin, Antonin J. J. Crumière, Abderrahman Khila
Science (2017-10-19) <https://doi.org/gcgjbs>
DOI: [10.1126/science.aan2748](https://doi.org/10.1126/science.aan2748) · PMID: [29051384](https://pubmed.ncbi.nlm.nih.gov/29051384/)
23. **Pervasive Correlated Evolution in Gene Expression Shapes Cell and Tissue Type Transcriptomes**
Cong Liang, Jacob M Musser, Alison Cloutier, Richard O Prum, Günter P Wagner
Genome Biology and Evolution (2018-01-23) <https://doi.org/gc69v9>
DOI: [10.1093/gbe/evy016](https://doi.org/10.1093/gbe/evy016) · PMID: [29373668](https://pubmed.ncbi.nlm.nih.gov/29373668/) · PMCID: [PMC5800078](https://pubmed.ncbi.nlm.nih.gov/PMC5800078/)
24. **Deep homology in the age of next-generation sequencing**
Patrick Tschopp, Clifford J. Tabin
Philosophical Transactions of the Royal Society B: Biological Sciences (2017-02-05)
<https://doi.org/gfzpbq>
DOI: [10.1098/rstb.2015.0475](https://doi.org/10.1098/rstb.2015.0475) · PMID: [27994118](https://pubmed.ncbi.nlm.nih.gov/27994118/) · PMCID: [PMC5182409](https://pubmed.ncbi.nlm.nih.gov/PMC5182409/)
25. **Embracing the comparative approach: how robust phylogenies and broader developmental sampling impacts the understanding of nervous system evolution**
Andreas Hejnol, Christopher J. Lowe
Philosophical Transactions of the Royal Society B: Biological Sciences (2015-12-19)
<https://doi.org/ggcd2m>
DOI: [10.1098/rstb.2015.0045](https://doi.org/10.1098/rstb.2015.0045) · PMID: [26554039](https://pubmed.ncbi.nlm.nih.gov/26554039/) · PMCID: [PMC4650123](https://pubmed.ncbi.nlm.nih.gov/PMC4650123/)
26. **The mammalian decidual cell evolved from a cellular stress response**
Eric M. Erkenbrack, Jamie D. Maziarz, Oliver W. Griffith, Cong Liang, Arun R. Chavan, Mauris C. Nnamani, Günter P. Wagner
PLOS Biology (2018-08-24) <https://doi.org/gd5b9s>
DOI: [10.1371/journal.pbio.2005594](https://doi.org/10.1371/journal.pbio.2005594) · PMID: [30142145](https://pubmed.ncbi.nlm.nih.gov/30142145/) · PMCID: [PMC6108454](https://pubmed.ncbi.nlm.nih.gov/PMC6108454/)
27. **OrthoFinder: phylogenetic orthology inference for comparative genomics**
David M. Emms, Steven Kelly
bioRxiv (2019-04-24) <https://www.biorxiv.org/content/10.1101/466201v2>
DOI: [10.1101/466201](https://doi.org/10.1101/466201)