

فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن فاعلاتن

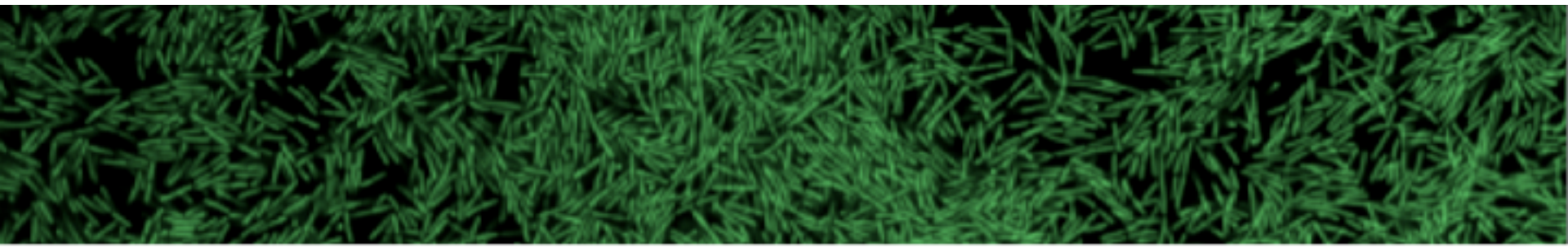
Bioinformatic Approaches to the Computation of Poetic Meter

A. Sean Pue (@seanpue)

Tracy K. Teal (@tracykteal)

C. Titus Brown (@ctitusbrown)

Michigan State University



Problem

- How to computationally scan Urdu poetry in a scalable and effective way

رنج سے خوگر ہو انساناں تو مٹ جاتا ہے رنج

- Focusing (for now) on “classical” poetry
- Uses a metrical system derived from Persian (Farsi) and Arabic
- Quantitative, not based on stress but on combinations of syllables in particular patterns
- Word-final “long” vowels are flexible (can be short), and word-final consonants can graft with following word-initial long vowels or short vowels with consonants

Challenges

- Need to know the unwritten short vowels, e.g. how the text is actually pronounced.

رنج سے خوگر ہو انساں تو مٹ جاتا ہے رنج

- Classical prosody usually describes verse in relation to a pattern, e.g. : fā'ilātun fā'ilātun fā'ilātun fā'ilun

فاعلاتن فاعلاتن فاعلاتن فاعلن

- Can be described as a series of “long” (=) and “short” (-) syllables, e.g. =-==/=--==/=--==/=--==

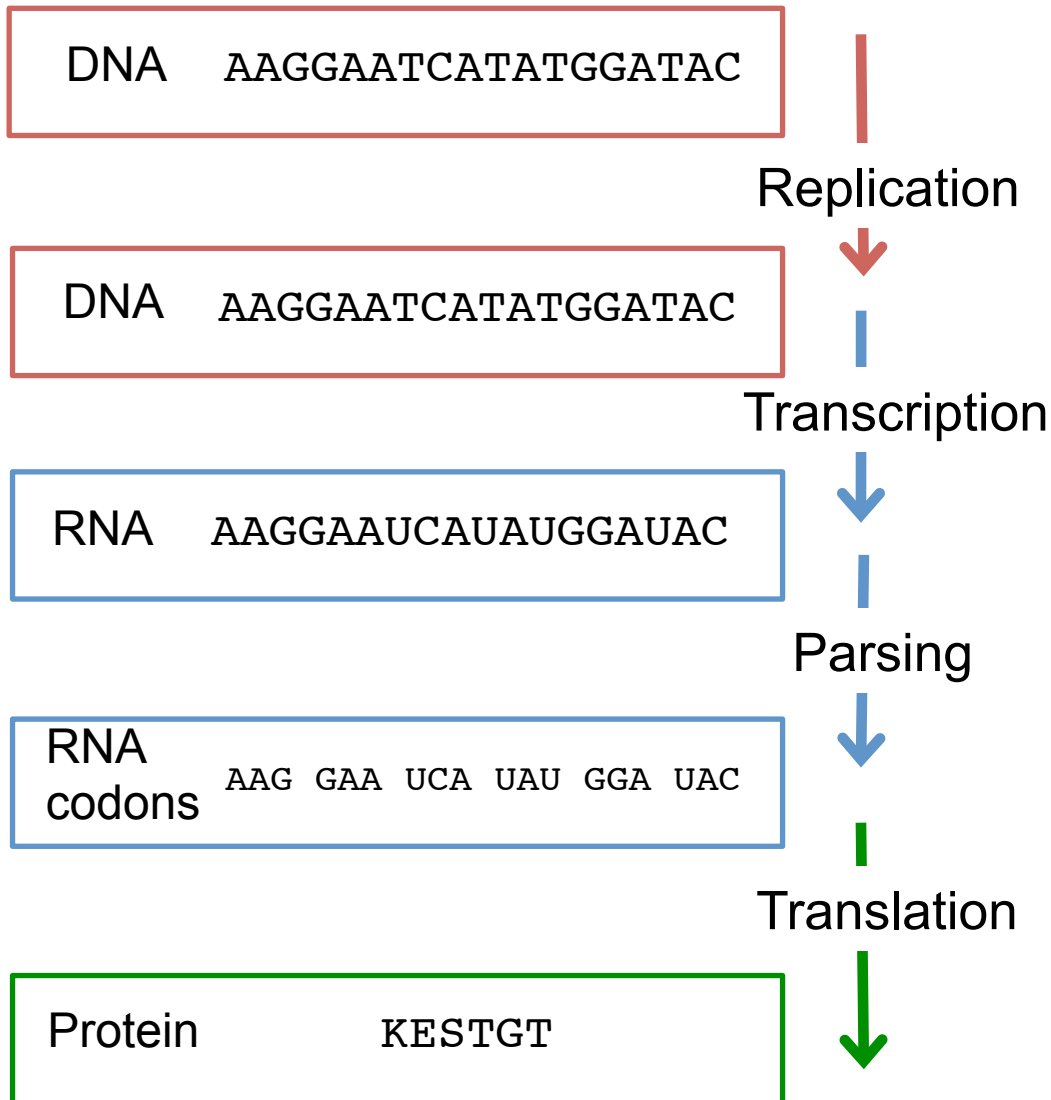
Why?

- Accurate textual encoding of literary corpus, i.e. error checking, metrical encoding (in [TEI](#))
- With proper meter, the possibility of extracting the phonetic pronunciation of text over time (marking long and short vowels with IPA length markers)
- Possibility of correlating text with performance
- Opens up possibilities of algorithmic criticism, stylometry, evolutionary studies of meter/poetry, misattribution detection
- Can be used for teaching as well, since all rules are encoded
- Techniques expandable to other languages (Hindi, Bengali, Punjabi, Sanskrit, Persian, Arabic, etc.)

Transfer of sequences that encode information

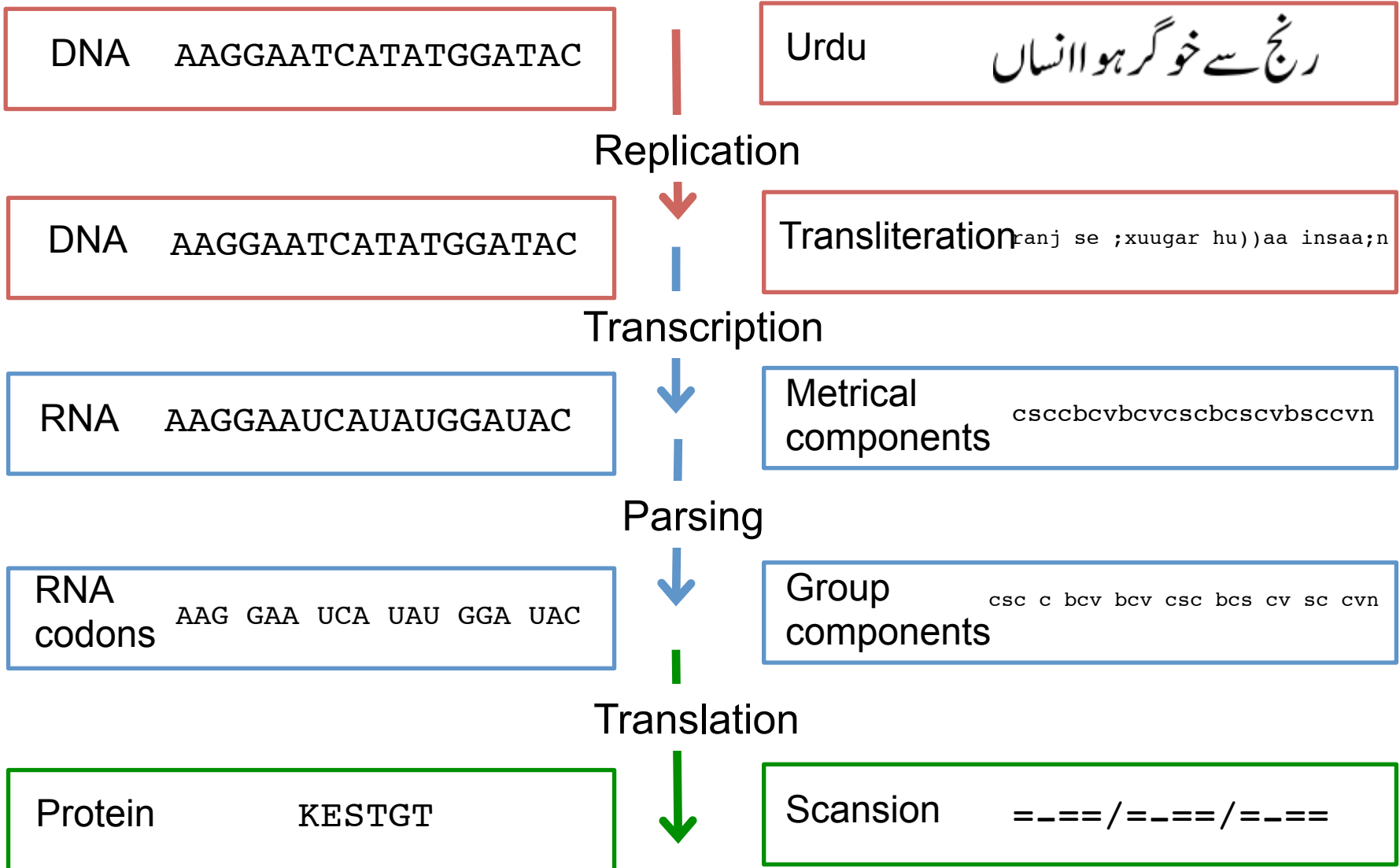
Transfer of sequences that encode information

The Central Dogma of Molecular Biology



Transfer of sequences that encode information

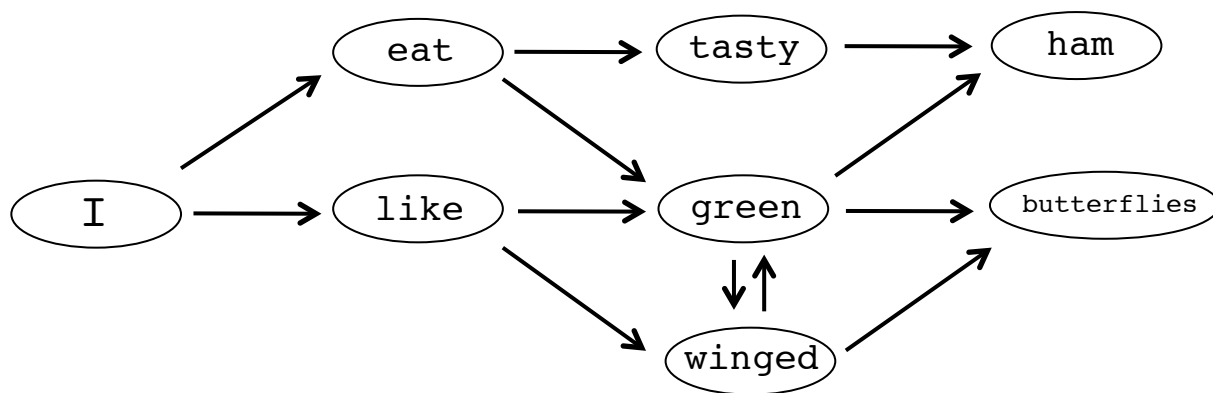
The Central Dogma of Molecular Biology



Representations of sequences

Markov chain

A mathematical system that undergoes transitions from one state to another, between a finite or countable number of possible states



Find the best transition state trajectory

Data Set

- *Dīvān-e Ghālib*
Mirzā Asadullah Ḳhān ‘Ghālib’
1797-1869
- Text taken from Frances Pritchett’s
A Desertful of Roses
- <http://www.columbia.edu/itc/mealac/pritchett/00ghalib/>
- 3314 lines of transcribed *ghazal* (poetic genre) poetry, with correct meter indicated



- *Look at IPython Notebook*
- *Visible via iPython Notebook Viewer:*
- [http://nbviewer.ipython.org/urls/
raw.githubusercontent.com/asp49/meter/graph/Shared
%20Horizons%20Presentation.ipynb](http://nbviewer.ipython.org/urls/raw.githubusercontent.com/asp49/meter/graph/Shared%20Horizons%20Presentation.ipynb)

Future Directions

- Sequence Alignment
- Assembly
- DeBruijn Graphs
- Free Verse
- Other languages

Sequence Alignment and Free Verse

AC-TGAT-CCA
|| || | | |
AC-TGATACTA
|| || | | |
ACGTGTTACCA

Multiple sequence alignment

== - = - == = - -
| | | | | | | |
== - - - = - = *

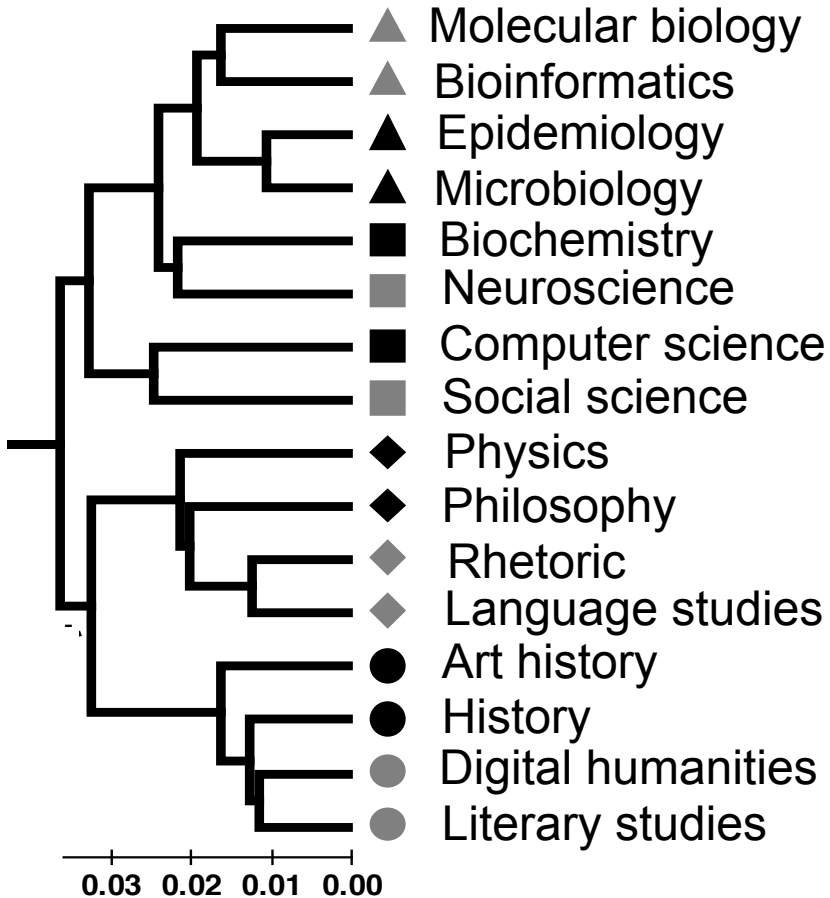
Free verse poetry scansion

Sequence Alignment and Free Verse

AC-TGAT-CCA
|| || | | |
AC-TGATACTA
|| || | | |
ACGTGTACCA

Multiple sequence alignment

Area of study taxonomy



NTS

Functional taxonomy

