

Unexpected emergent properties in all discrete systems

Emergent properties are properties of the “system” as a whole.

A discrete system is one made up of distinguishable pieces, for example amino acids in proteins. Meaning is irrelevant.

Background

All discrete systems have two complementary sets of properties depending on whether the individual pieces are considered to have **specific meaning**, for example biochemical properties or **no intrinsic meaning** where they are merely distinguishable.

Background

We can identify two kinds of discrete system depending on whether the **order** of pieces is important or not.



Word counts in
a book ...

Or ...

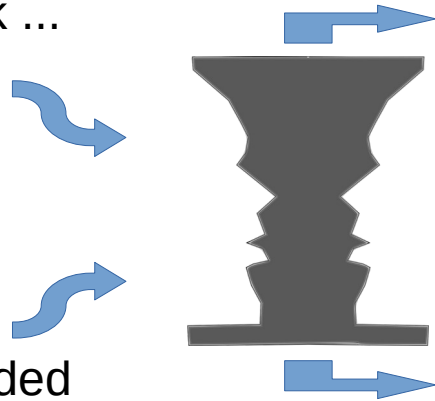
Boxes of
unordered
coloured beads



Unfolded
proteins ...

Or ...

Strings of
ordered
coloured beads



Background

Both unordered and ordered systems lead to two characteristic distributions **dominated by power-law**,

Hatton and Warr (2019), RSOS.

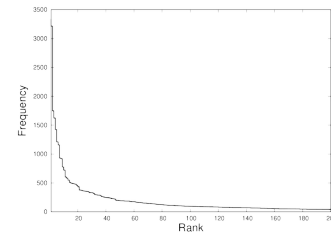
Background

Unordered systems such as word counts in books

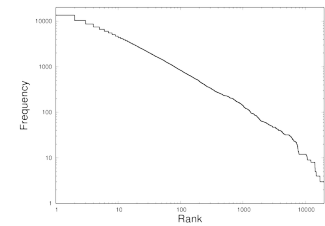
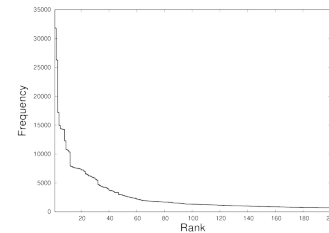
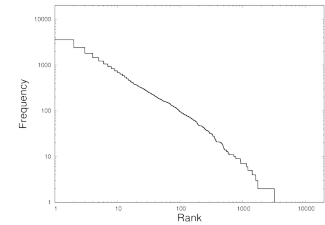


Discovered empirically by George Zipf (1935)

Three Men in a Boat



Swedish Bible

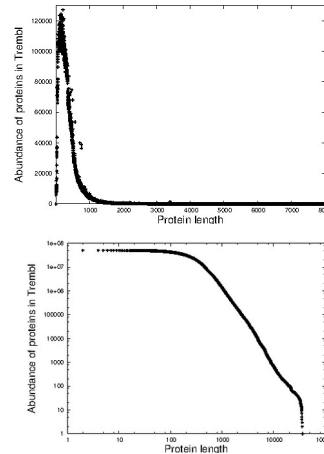


Background

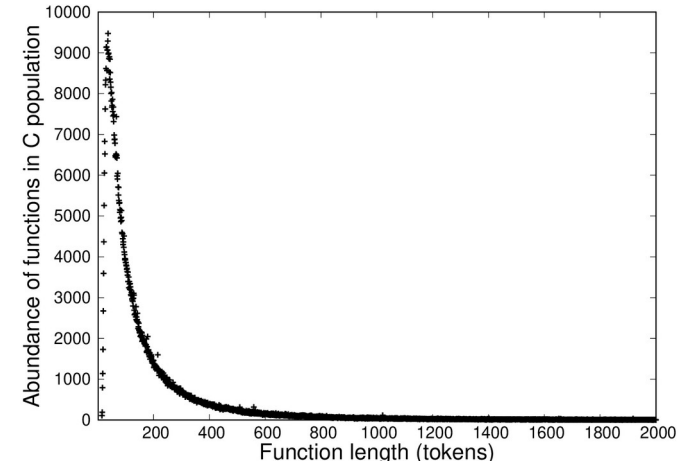
Ordered systems such as proteins, computer programs ..



TrEMBL protein database

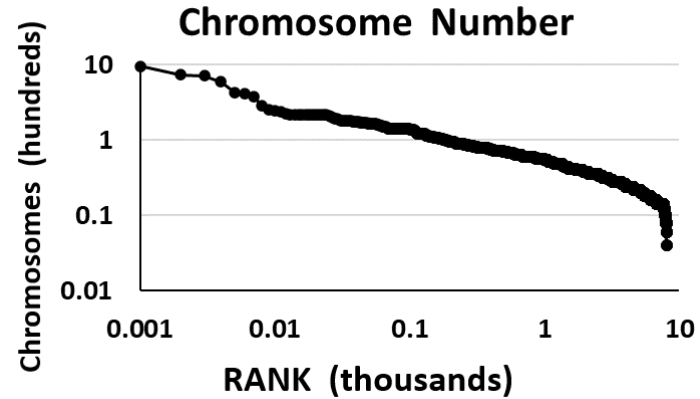
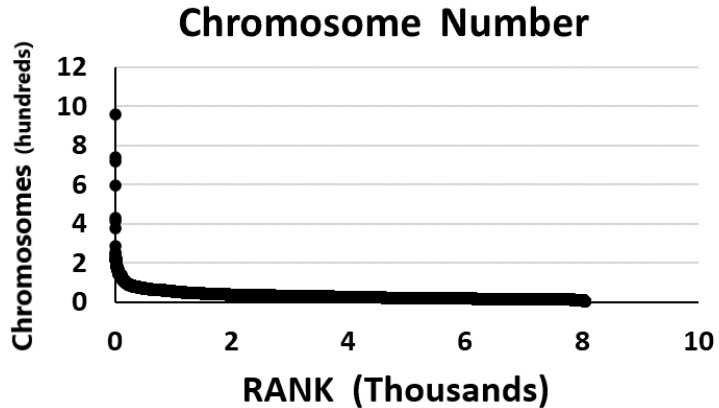


100 million lines of C programs



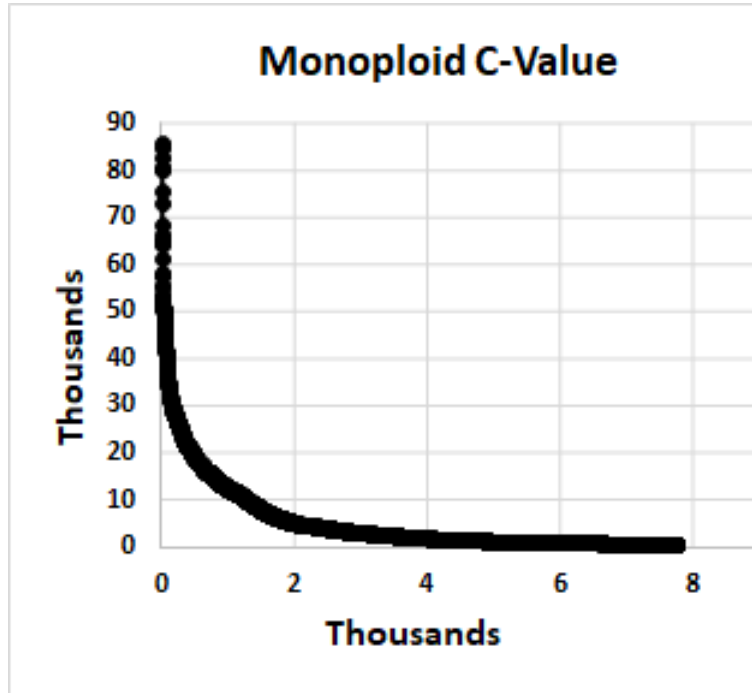
Examples from Kew plant data

Chromosome number



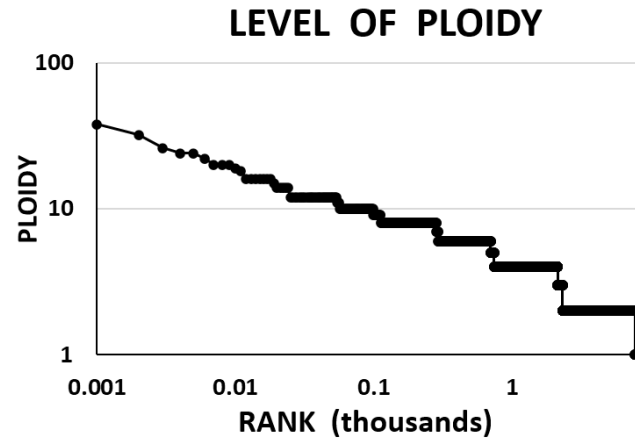
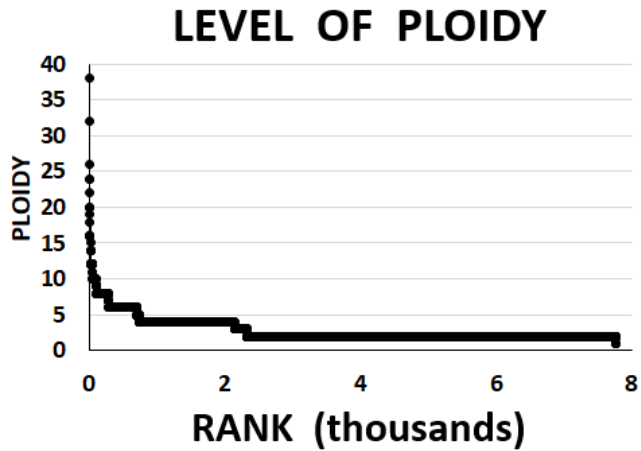
Examples from Kew plant data

C-value



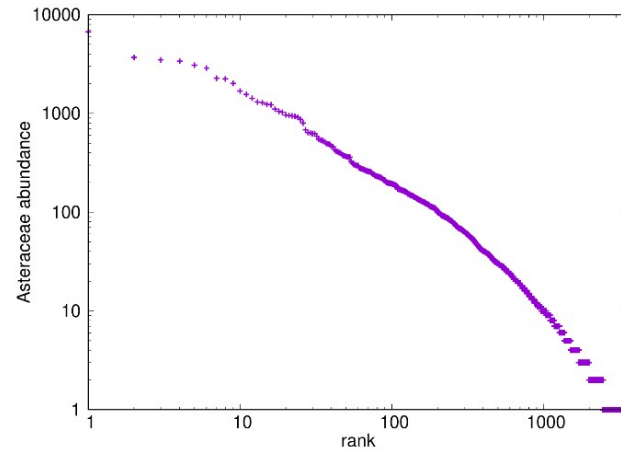
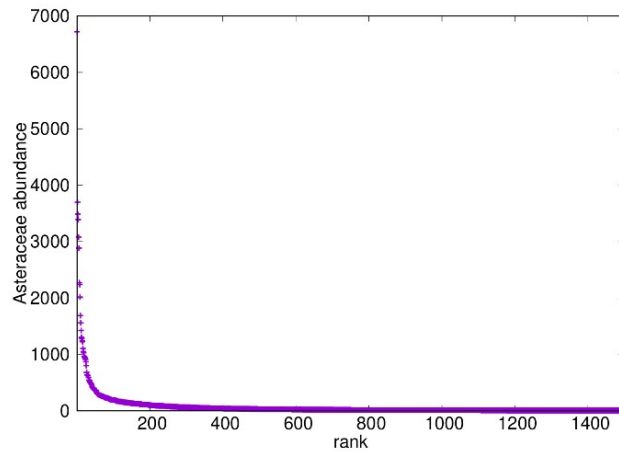
Examples from Kew plant data

Ploidy



Other plant examples

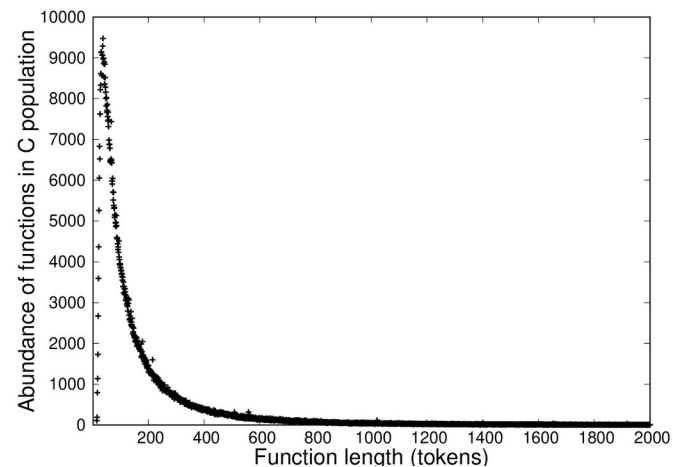
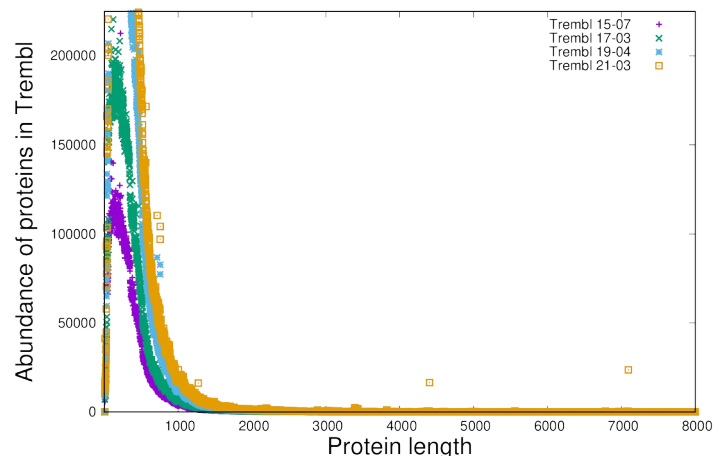
Asteraceae



Lengths of proteins and maximum length



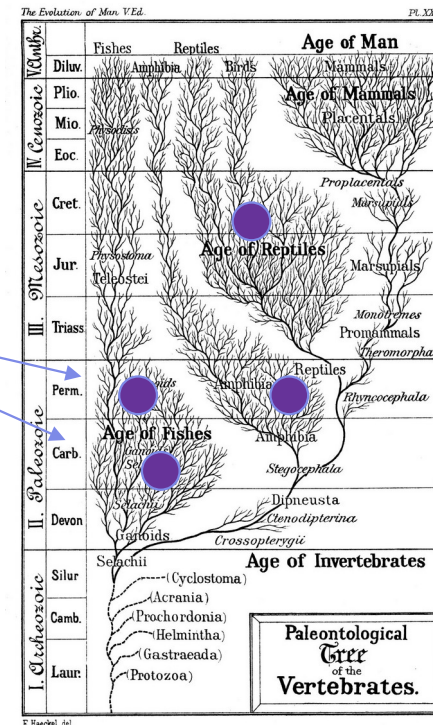
Distribution of protein lengths in Trembl



HGT: Protein copying round the tree of life and viruses

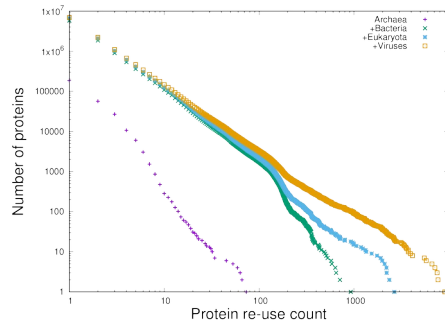
The only requirements for the homogeneous model is that bins do not overlap (each bead is distinguished only by its colour) and that the whole system is represented (1 bead for every discrete piece).

Each bead is a protein and a unique colour is used if a protein appears N times in the trees of life and viruses, N=1,...,10000 (or so in TrEMBL 21-03).

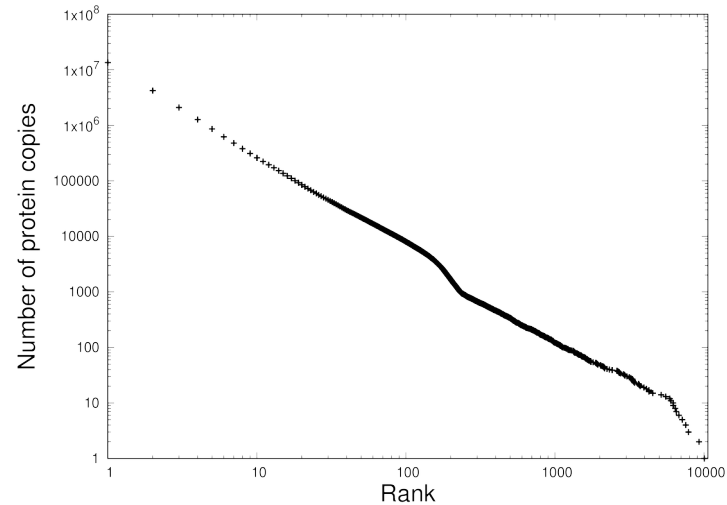


HGT: Protein copying round the tree of life and viruses

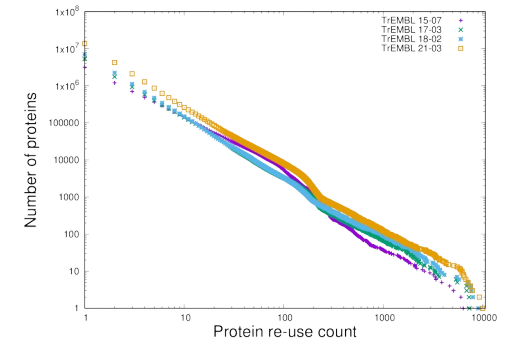
Adding domains



21-03 combined

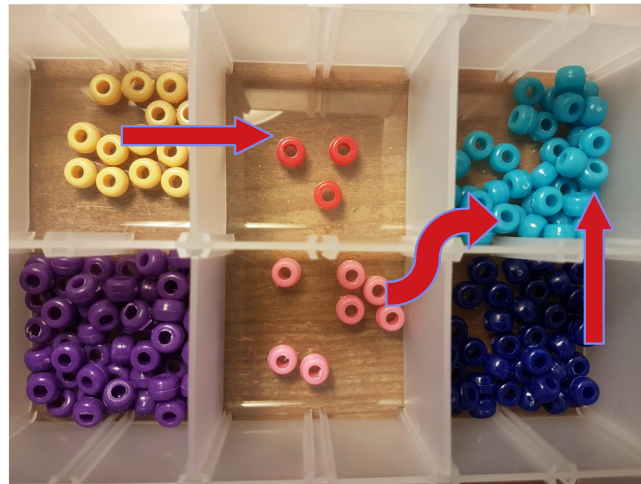


15-07 to 21-03



HGT: Protein copying round the tree of life and viruses

Natural Selection could not handle this as *beads move between bins as the size of the proteome increases*



Universal Inverse Symmetry



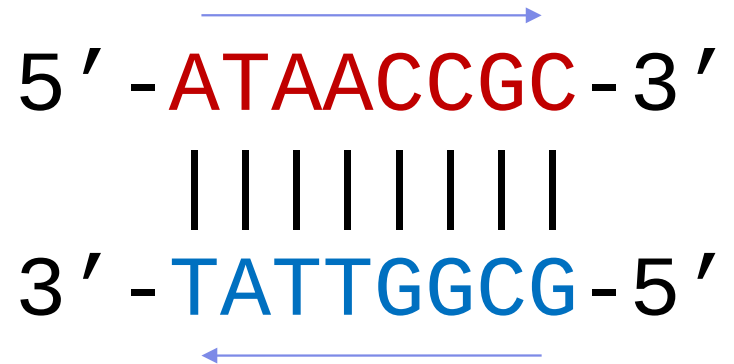
We know in a diploid (two-stranded) DNA molecule,
 $\%(A) = \%(T)$ and $\%(C) = \%(G)$

Extending to tuples (8 here),
 $\%(ATAACCGC)$ on strand 1
 $= \%(GCGGTTAT)$ on strand 2

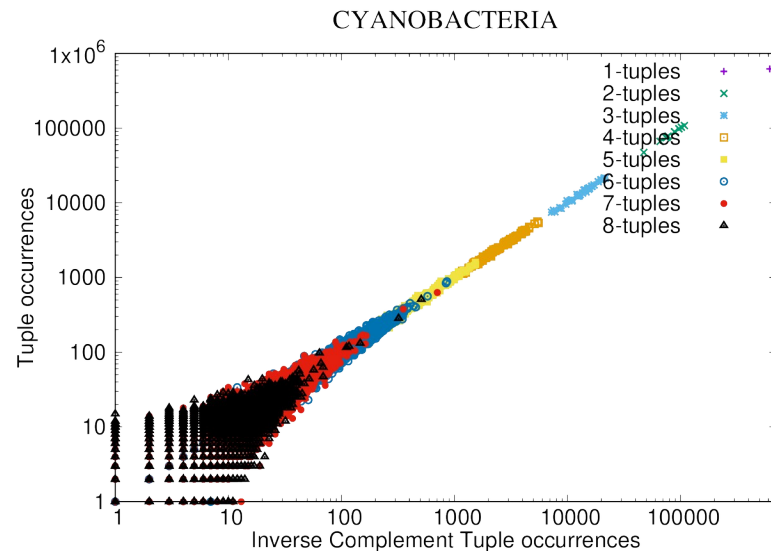
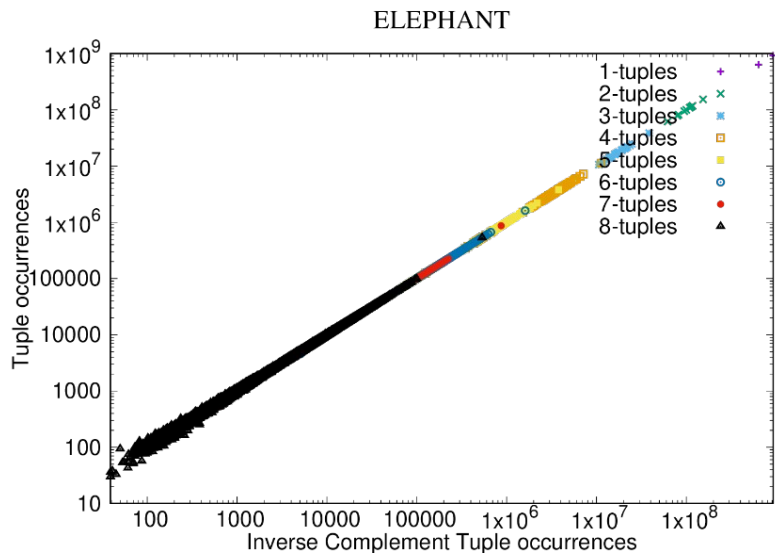
The FLI predicts that

**$\%(ATAACCGC)$ on strand 1
 $= \%(GCGGTTAT)$ on strand 1 !**

**... and all n-tuples, degrading stochastically with
genome size**



Universal Inverse Symmetry on same strand



Universal Inverse Symmetry on same strand

<https://www.leshatton.org/Documents/178-lifes-eternal-symmetry.mp4>

Universal Inverse Symmetry on same strand

This offers an explanation for ...

- The astonishing stability of the genome over > 3.6 billion years
- Why prokaryotes are much more agile than eukaryotes, (they are not so constrained).

References

Introduction ...

- Les Hatton and Gregory Warr (2019), “Strong evidence of an information-theoretical conservation principle linking all discrete systems”, Royal Society Open Science, <https://doi.org/10.1098/rsos.191101>

General Background ...

- Les Hatton and Gregory Warr (2022), “Exposing Nature’s Bias: The Hidden Clockwork behind Society, Life and the Universe”, Bluespear Publishing, ISBN 978-1908422040