Never mind the version of record…
Which is your *format* of record?

Kaveh Bazargan
@Kaveh1000
Version of record – essential in publishing
Version of record – essential in publishing

• Responsibility of publisher
Version of record – essential in publishing

- Responsibility of publisher
- Dispute might occur decades in the future
Version of record – essential in publishing

• Responsibility of publisher
• Dispute might occur decades in the future
• Publisher has to nominate published VoR
Niso recommendation
Niso recommendation

- NISO-RP-8-2008 (30pp): Journal Article Versions
Niso recommendation

- NISO-RP-8-2008 (30pp): Journal Article Versions
  - Version of record (VoR)
Niso recommendation

- NISO-RP-8-2008 (30pp): Journal Article Versions
  - Version of record (VoR)
  - Corrected VoR
Niso recommendation

- NISO-RP-8-2008 (30pp): Journal Article Versions
  - Version of record (VoR)
  - Corrected VoR
  - Enhanced VoR
It used to be easy...
It used to be easy...

- No doubt about version of record
It used to be easy...

- No doubt about version of record
- Author and publisher signed off on print copy or galley
In the electronic age...
In the electronic age...

- DOI defines Version of Record
In the electronic age...

- DOI defines Version of Record
- Points to a web page, maintained by publisher
Disclaimer

• Use PeerJ as example because:
http://dx.doi.org/10.7717/peerj.127
A Markovian analysis of bacterial genome sequence constraints

Aaron D. Skewes¹,², Roy D. Welch³

PubMed ID: 24010012

Author and article information

Abstract

The arrangement of nucleotides within a bacterial chromosome is influenced by numerous factors. The degeneracy of the third codon within each reading frame allows some flexibility of nucleotide selection; however, the third nucleotide in the triplet of each codon is at least partly determined by the preceding two. This is most evident in organisms with a strong G + C bias, as the degenerate codon must...
A Markovian analysis of bacterial genome sequence constraints

Aaron D. Skewes, Roy D. Welch

Abstract

The arrangement of nucleotides within a bacterial chromosome is influenced by numerous factors. The degeneracy of the third codon within each reading frame allows some flexibility of nucleotide selection; however, the third nucleotide in the triplet of each codon is at least partly determined by the preceding two. This is most evident in organisms with a strong G + C bias, as the degenerate codon...
A Markovian analysis of bacterial genome sequence constraints

Aaron D. Skewes\textsuperscript{1,2}, Roy D. Welch\textsuperscript{1}

PubMed ID: 24010012

Author and article information

Abstract

The arrangement of nucleotides within a bacterial chromosome is influenced by numerous factors. The degeneracy of the third codon within each reading frame allows some flexibility of nucleotide selection; however, the third nucleotide in the triplet of each codon is at least partly determined by the preceding two. This is most
With the existence of a high-order Markov process, the number of variables (states) increases exponentially with each increase in model order. This allows a more precise determination of the probability of a particular sequence (i.e., greater resolution of transition probabilities), and thereby the identification of more sequences that are unlikely to be bacterial chromosomes. Let $X^K_L$ define a sequence of $K$ letters over an alphabet of $L$ characters, then the probability of sequence $X^K_L$ is: $P(x^K_L) = \prod_{j=1}^{K} P(X_j = x_j | X_{L}^{j-L} = x_{L}^{j-L})$, where $X_j$ represents the nucleotide at position $j$ with $x_j$ as its realization. For a DNA sequence (and assuming a 3rd-order Markov Model), $L = K = 4$. In the trivial case, where each character (nucleotide) is equally likely to occur, it can be easily shown that $P(x^K_L) = \frac{1}{L^K}$ and the expected frequency $f(x^K_L) = \frac{N-K-1}{L^K} \approx \frac{N}{L^K}$ for $K \ll N$. For any sequence that is the result of a 3rd-order Markov process and modeled as such, we get $L^K = 4^4$ times more
With the existence of a high-order Markov process, the number of variables (states) increases exponentially with each increase in model order. This allows a more precise determination of the probability of a particular sequence (i.e., greater resolution of transition probabilities), and thereby the identification of more sequences that are unlikely to be bacterial chromosomes. Let $X_L^K$ define a sequence of $K$ letters over an alphabet of $L$ characters, then the probability of sequence $X_L^K$ is: $P(x_L^K) = \prod_{j=1}^{K} P(X_j = x_j | X_L^{j-L} = x_L^{j-L})$, where $X_j$ represents the nucleotide at position $j$ with $x_j$ as its realization. For a DNA sequence (and assuming a 3rd-order Markov Model), $L = K = 4$. In the trivial case, where each character (nucleotide) is equally likely to occur, it can be easily shown that $P(x_L^K) = \frac{1}{L^K}$ and the expected frequency $f(x_L^K) = \frac{N-K-1}{L^K} \approx \frac{N}{L^K}$ for $K \ll N$. For any sequence that is the result of a 3rd-order Markov process and modeled as such, we get $L^K = 4^4$ times more
thereby the identity.

Let $X^K_L$ define a subfamily of sequence $X$. 
thereby the identity

Let $X^K_L$ define a probability of sequence $X$. 

Identification of more problems. Let $X^K_L$ define a probability of sequence $X$.
Let $X^K_L$ define a property of sequence $X$. Let $X^K_L$ define another property of sequence $X$. Therefore, the identity...
\[ X \subseteq L \}

\[ K \]
\[ X^L \]
$x^k L^k$
\[Gr\]
$X_L^K$
\[ G^{X_L}_K \]
So far...
So far...

- PDF

\[ X^K_L \]
So far...

- PDF
- html

\[ X^K_L \]
So far...

- PDF
- html
- PNG
So far...

- PDF
- html
- PNG
- MathML
So far…

- PDF
- html
- PNG
- MathML
- TeX
So far...

- PDF
- html
- PNG
- MathML
- TeX
- Epub 2, Epub 3, + future formats
Nightmare scenario
Nightmare scenario

10 years after publication, a mismatch is reported
Nightmare scenario

10 years after publication, a mismatch is reported
e.g. PDF and html don’t match…
Nightmare scenario

10 years after publication, a mismatch is reported
e.g. PDF and html don’t match…

Worse, in XML, TeX and MathML differ
Nightmare scenario

10 years after publication, a mismatch is reported e.g. PDF and html don’t match…

Worse, in XML, TeX and MathML differ

Author has passed away
Nightmare scenario

10 years after publication, a mismatch is reported
e.g. PDF and html don’t match…

Worse, in XML, TeX and MathML differ

Author has passed away

Publisher must nominate format of record
The solution
The solution

- XML should have content in one form only
The solution

- XML should have content in one form only
- e.g. MathML or TeX for mathematics
The solution

• XML should have content in one form only

• e.g. MathML or TeX for mathematics

• All formats should be derived from XML with 100% automation
The solution

- XML should have content in one form only
- e.g. MathML or TeX for mathematics
- All formats should be derived from XML with 100% automation
- XML should be format of record
Let’s go further...
Let’s go further...

- Publishers publish good XML, and only the XML
Let’s go further...

• Publishers publish good XML, and only the XML

• Let’s throw away idea of “house style”
Let’s go further…

- Publishers publish good XML, and *only* the XML
- Let’s throw away idea of “house style”
- 3rd parties specialize in “rendering” the XML to reader preferences, using automated systems (free or commercial)
Let’s go further…

• Publishers publish good XML, and only the XML

• Let’s throw away idea of “house style”

• 3rd parties specialize in “rendering” the XML to reader preferences, using automated systems (free or commercial)

• …e.g. single column PDF, for iPad mini, Garamond font, Harvard references
Let’s go further...

- Publishers publish good XML, and *only* the XML
- Let’s throw away idea of “house style”
- 3rd parties specialize in “rendering” the XML to reader preferences, using automated systems (free or commercial)
- ...e.g. single column PDF, for iPad mini, Garamond font, Harvard references
- I want to read *your* content, but with *my* style
• Publishers don’t worry about formats but concentrate on content, peer review, copy editing
• Publishers don’t worry about formats but concentrate on content, peer review, copy editing

• 3rd parties can cater for accessibility too, e.g. blind mathematicians; a reader for the blind is just another “rendering”