‹JATS4R›

Standardising the way XML content is tagged to enhance publishing workflows and reusability

Melissa Harrison
Head of production operations, eLife

http://jats4r.org/

Kaveh Bazargan (River Valley Technologies)

Jeff Beck (PMC)

Alf Eaton (PeerJ)

Melissa Harrison (eLife)

Chris Maloney (NCBI)

Nikos Markantonatos (Atypon)

Daniel Mietchen (NIH)

Mary Seligy (Canadian Science Publishing)

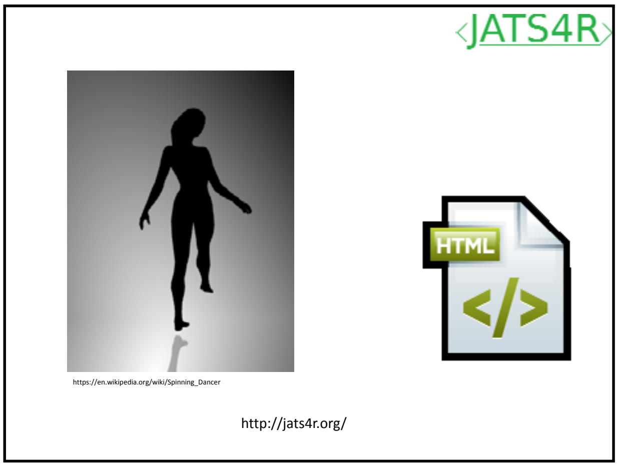
http://jats4r.org/

The XML tells the whole story

http://jats4r.org/

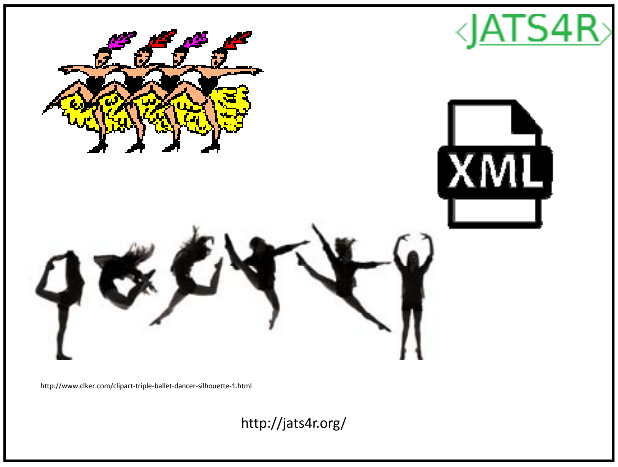
http://www.clipartsheep.com/lyrical-ballet-costumes-for-dance-recital-and-competition-clipart-92787.html

http://jats4r.org/



https://en.wikipedia.org/wiki/Spinning_Dancer

http://jats4r.org/



http://www.clker.com/clipart-triple-ballet-dancer-silhouette-1.html

http://jats4r.org/

JATS4R

JATS4R recommendations…so far

http://jats4r.org/

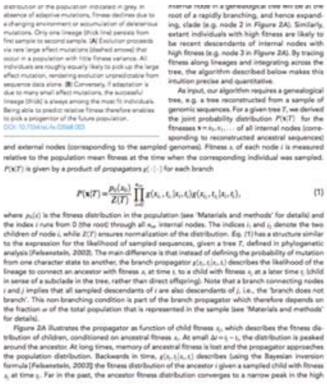
JATS4R

Math

http://jats4r.org/recommendations/math.html

http://jats4r.org/

JATS4R



distribution of the population indicated in grey. In absence of adaptive mutations, fitness declines due to a changing environment or accumulation of deleterious mutations. Only one lineage (thick line) persists from first sample to second sample. (A) Evolution proceeds via rare large effect mutations (dashed arrows) that occur in a population with little fitness variance. All individuals are roughly equally likely to pick up the large effect mutation, rendering evolution unpredictable from sequence data alone. (B) Conversely, if adaptation is due to many small effect mutations, the successful lineage (thick) is always among the most fit individuals. Being able to predict relative fitness therefore enables to pick a progenitor of the future population.
DOI: 10.7554/eLife.03568.003

internal node in a genealogical tree will be at the root of a rapidly branching, and hence expanding, clade (e.g. node 2 in Figure 2A). Similarly, extant individuals with high fitness are likely to be recent descendants of internal nodes with high fitness (e.g. node 3 in Figure 2A). By tracing fitness along lineages and integrating across the tree, the algorithm described below makes this intuition precise and quantitative.

As input, our algorithm requires a genealogical tree, e.g. a tree reconstructed from a sample of genomic sequences. For a given tree T, we derived the joint probability distribution $P(\mathbf{x}|T)$ for the fitnesses $\mathbf{x} = x_0, x_1, \ldots$ of all internal nodes (corresponding to reconstructed ancestral sequences) and external nodes (corresponding to the sampled genomes). Fitness x_i of each node i is measured relative to the population mean fitness at the time when the corresponding individual was sampled. $P(\mathbf{x}|T)$ is given by a product of *propagators* $g(\cdot|\cdot)$ for each branch

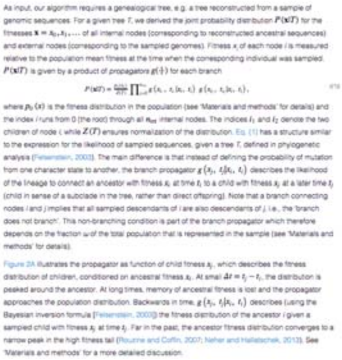
$$P(\mathbf{x}|T) = \frac{p_0(x_0)}{Z(T)} \prod_{i=0}^{n_{int}} g(x_{i_1}, t_{i_1}|x_i, t_i) g(x_{i_2}, t_{i_2}|x_i, t_i), \tag{1}$$

where $p_0(x)$ is the fitness distribution in the population (see 'Materials and methods' for details) and the index i runs from 0 (the root) through all n_{int} internal nodes. The indices i_1 and i_2 denote the two children of node i, while $Z(T)$ ensures normalization of the distribution. Eq. (1) has a structure similar to the expression for the likelihood of sampled sequences, given a tree T, defined in phylogenetic analysis (Felsenstein, 2003). The main difference is that instead of defining the probability of mutation from one character state to another, the branch propagator $g(x_j, t_j|x_i, t_i)$ describes the likelihood of the lineage to connect an ancestor with fitness x_i at time t_i to a child with fitness x_j at a later time t_j (child in sense of a subclade in the tree, rather than direct offspring). Note that a branch connecting nodes i and j implies that all sampled descendants of i are also descendants of j, i.e., the 'branch does not branch'. This non-branching condition is part of the branch propagator which therefore depends on the fraction ω of the total population that is represented in the sample (see 'Materials and methods' for details).

Figure 2A illustrates the propagator as function of child fitness x_j, which describes the fitness distribution of children, conditioned on ancestral fitness x_i. At small $\Delta t = t_j - t_i$ the distribution is peaked around the ancestor. At long times, memory of ancestral fitness is lost and the propagator approaches the population distribution. Backwards in time, $g(x_j, t_j|x_i, t_i)$ describes (using the Bayesian inversion formula [Felsenstein, 2003]) the fitness distribution of the ancestor i given a sampled child with fitness x_j at time t_j. Far in the past, the ancestor fitness distribution converges to a narrow peak in the high

http://jats4r.org/

JATS4R

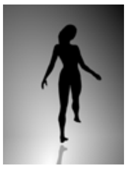


As input, our algorithm requires a genealogical tree, e.g. a tree reconstructed from a sample of genomic sequences. For a given tree T, we derived the joint probability distribution $P(\mathbf{x}|T)$ for the fitnesses $\mathbf{x} = x_0, x_1, \ldots$ of all internal nodes (corresponding to reconstructed ancestral sequences) and external nodes (corresponding to the sampled genomes). Fitness x_i of each node i is measured relative to the population mean fitness at the time when the corresponding individual was sampled. $P(\mathbf{x}|T)$ is given by a product of *propagators* $g(\cdot|\cdot)$ for each branch

$$P(\mathbf{x}|T) = \frac{p_0(x_0)}{Z(T)} \prod_{i=0}^{n_{int}} g(x_{i_1}, t_{i_1}|x_i, t_i)\; g(x_{i_2}, t_{i_2}|x_i, t_i), \tag{1}$$

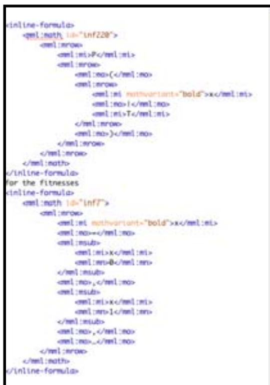
where $p_0(x)$ is the fitness distribution in the population (see 'Materials and methods' for details) and the index i runs from 0 (the root) through all n_{int} internal nodes. The indices i_1 and i_2 denote the two children of node i, while $Z(T)$ ensures normalization of the distribution. Eq. (1) has a structure similar to the expression for the likelihood of sampled sequences, given a tree T, defined in phylogenetic analysis (Felsenstein, 2003). The main difference is that instead of defining the probability of mutation from one character state to another, the branch propagator $g(x_j, t_j|x_i, t_i)$ describes the likelihood of the lineage to connect an ancestor with fitness x_i at time t_i to a child with fitness x_j at a later time t_j (child in sense of a subclade in the tree, rather than direct offspring). Note that a branch connecting nodes i and j implies that all sampled descendants of i are also descendants of j, i.e., the 'branch does not branch'. This non-branching condition is part of the branch propagator which therefore depends on the fraction ω of the total population that is represented in the sample (see 'Materials and methods' for details).

Figure 2A illustrates the propagator as function of child fitness x_j, which describes the fitness distribution of children, conditioned on ancestral fitness x_i. At small $\Delta t = t_j - t_i$, the distribution is peaked around the ancestor. At long times, memory of ancestral fitness is lost and the propagator approaches the population distribution. Backwards in time, $g(x_j, t_j|x_i, t_i)$ describes (using the Bayesian inversion formula [Felsenstein, 2003]) the fitness distribution of the ancestor i given a sampled child with fitness x_j at time t_j. Far in the past, the ancestor fitness distribution converges to a narrow peak in the high fitness tail (Rouzine and Coffin, 2007; Neher and Hallatschek, 2013). See 'Materials and methods' for a more detailed discussion.

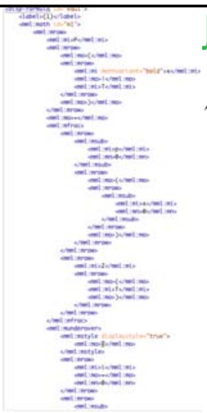


http://jats4r.org/

JATS4R



```
<inline-formula>
    <mml:math id="inf220">
        <mml:mrow>
            <mml:mi>P</mml:mi>
            <mml:mrow>
                <mml:mo>(</mml:mo>
                <mml:mrow>
                    <mml:mi mathvariant="bold">x</mml:mi>
                    <mml:mo>|</mml:mo>
                    <mml:mi>T</mml:mi>
                </mml:mrow>
                <mml:mo>)</mml:mo>
            </mml:mrow>
        </mml:mrow>
    </mml:math>
</inline-formula>
for the fitnesses
<inline-formula>
    <mml:math id="inf7">
        <mml:mrow>
            <mml:mi mathvariant="bold">x</mml:mi>
            <mml:mo>=</mml:mo>
            <mml:msub>
                <mml:mi>x</mml:mi>
                <mml:mn>0</mml:mn>
            </mml:msub>
            <mml:mo>,</mml:mo>
            <mml:msub>
                <mml:mi>x</mml:mi>
                <mml:mn>1</mml:mn>
            </mml:msub>
            <mml:mo>,</mml:mo>
            <mml:mo>…</mml:mo>
        </mml:mrow>
    </mml:math>
</inline-formula>
```



```
<disp-formula id="equ1">
    <label>(1)</label>
    <mml:math id="m1">
        <mml:mrow>
            <mml:mi>P</mml:mi>
            <mml:mrow>
                <mml:mo>(</mml:mo>
                <mml:mrow>
                    <mml:mi mathvariant="bold">x</mml:mi>
                    <mml:mo>|</mml:mo>
                    <mml:mi>T</mml:mi>
                </mml:mrow>
                <mml:mo>)</mml:mo>
            </mml:mrow>
            <mml:mo>=</mml:mo>
            <mml:mfrac>
                <mml:mrow>
                    <mml:msub>
                        <mml:mi>p</mml:mi>
                        <mml:mn>0</mml:mn>
                    </mml:msub>
                    <mml:mrow>
                        <mml:mo>(</mml:mo>
                        <mml:mrow>
                            <mml:msub>
                                <mml:mi>x</mml:mi>
                                <mml:mn>0</mml:mn>
                            </mml:msub>
                        </mml:mrow>
                        <mml:mo>)</mml:mo>
                    </mml:mrow>
                </mml:mrow>
                <mml:mrow>
                    <mml:mi>Z</mml:mi>
                    <mml:mrow>
                        <mml:mo>(</mml:mo>
                        <mml:mi>T</mml:mi>
                        <mml:mo>)</mml:mo>
                    </mml:mrow>
                </mml:mrow>
            </mml:mfrac>
            <mml:munderover>
                <mml:mstyle displaystyle="true">
                    <mml:mo>∏</mml:mo>
                </mml:mstyle>
                <mml:mrow>
                    <mml:mi>i</mml:mi>
                    <mml:mo>=</mml:mo>
                    <mml:mn>0</mml:mn>
                </mml:mrow>
                <mml:mrow>
                    <mml:msub>
```

http://jats4r.org/

JATS4R

Permissions and Licenses

http://jats4r.org/recommendations/permissions.html

http://jats4r.org/

⟨JATS4R⟩

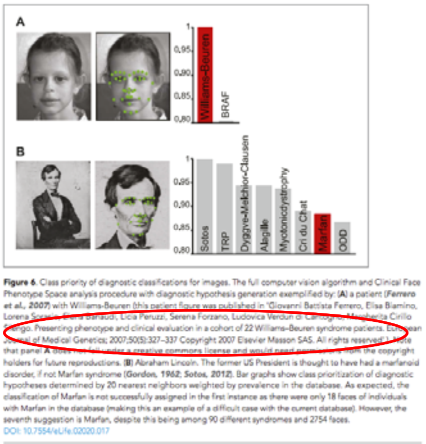


Figure 6. Class priority of diagnostic classifications for images. The full computer vision algorithm and Clinical Face Phenotype Space analysis procedure with diagnostic hypothesis generation exemplified by: (**A**) a patient (***Ferrero et al., 2007***) with Williams-Beuren (this patient figure was published in "Giovanni Battista Ferrero, Elisa Biamino, Lorena Sorasio, Elena Banaudi, Licia Peruzzi, Serena Forzano, Ludovica Verdun di Cantogno, Margherita Cirillo Silengo. Presenting phenotype and clinical evaluation in a cohort of 22 Williams–Beuren syndrome patients. European Journal of Medical Genetics; 2007;50(5):327–337 Copyright 2007 Elsevier Masson SAS. All rights reserved".). Note that panel **A** does not fall under a creative commons license and would need permissions from the copyright holders for future reproductions. (**B**) Abraham Lincoln. The former US President is thought to have had a marfanoid disorder, if not Marfan syndrome (***Gordon, 1962; Sotos, 2012***). Bar graphs show class prioritization of diagnostic hypotheses determined by 20 nearest neighbors weighted by prevalence in the database. As expected, the classification of Marfan is not successfully assigned in the first instance as there were only 18 faces of individuals with Marfan in the database (making this an example of a difficult case with the current database). However, the seventh suggestion is Marfan, despite this being among 90 different syndromes and 2754 faces.
DOI: 10.7554/eLife.02020.017

⟨JATS4R⟩

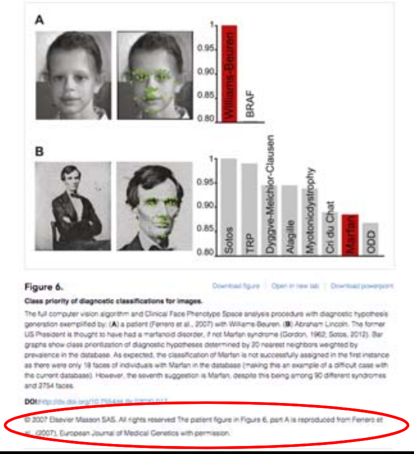


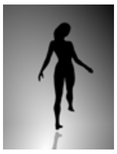
Figure 6.

Class priority of diagnostic classifications for images.

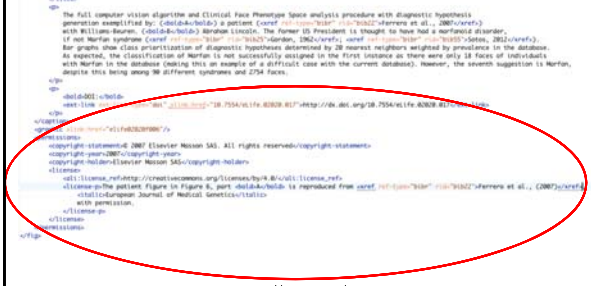
The full computer vision algorithm and Clinical Face Phenotype Space analysis procedure with diagnostic hypothesis generation exemplified by: (**A**) a patient (Ferrero et al., 2007) with Williams-Beuren. (**B**) Abraham Lincoln. The former US President is thought to have had a marfanoid disorder, if not Marfan syndrome (Gordon, 1962; Sotos, 2012). Bar graphs show class prioritization of diagnostic hypotheses determined by 20 nearest neighbors weighted by prevalence in the database. As expected, the classification of Marfan is not successfully assigned in the first instance as there were only 18 faces of individuals with Marfan in the database (making this an example of a difficult case with the current database). However, the seventh suggestion is Marfan, despite this being among 90 different syndromes and 2754 faces.

DOI:http://dx.doi.org/10.7554/eLife.02020.017

© 2007 Elsevier Masson SAS. All rights reserved The patient figure in Figure 6, part A is reproduced from Ferrero et al., (2007), European Journal of Medical Genetics with permission.



⟨JATS4R⟩

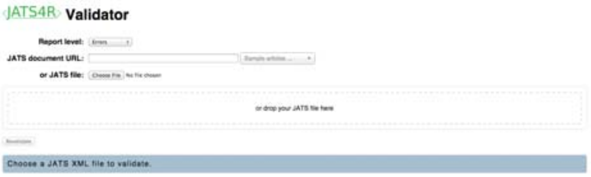


```
<fig id="fig6" position="float">
    <object-id pub-id-type="doi">10.7554/eLife.02020.017</object-id>
    <label>Figure 6.</label>
    <caption>
        <title>
            Class priority of diagnostic classifications for images.
        </title>
        <p>
            The full computer vision algorithm and Clinical Face Phenotype Space analysis procedure with diagnostic hypothesis
            generation exemplified by: (<bold>A</bold>) a patient (<xref ref-type="bibr" rid="bib22">Ferrero et al., 2007</xref>)
            with Williams-Beuren. (<bold>B</bold>) Abraham Lincoln. The former US President is thought to have had a marfanoid disorder,
            if not Marfan syndrome (<xref ref-type="bibr" rid="bib25">Gordon, 1962</xref>; <xref ref-type="bibr" rid="bib55">Sotos, 2012</xref>).
            Bar graphs show class prioritization of diagnostic hypotheses determined by 20 nearest neighbors weighted by prevalence in the database.
            As expected, the classification of Marfan is not successfully assigned in the first instance as there were only 18 faces of individuals
            with Marfan in the database (making this an example of a difficult case with the current database). However, the seventh suggestion is Marfan,
            despite this being among 90 different syndromes and 2754 faces.
        </p>
        <p>
            <bold>DOI:</bold>
            <ext-link ext-link-type="doi" xlink:href="10.7554/eLife.02020.017">http://dx.doi.org/10.7554/eLife.02020.017</ext-link>
        </p>
    </caption>
    <graphic xlink:href="elife02020f006"/>
    <permissions>
        <copyright-statement>© 2007 Elsevier Masson SAS. All rights reserved</copyright-statement>
        <copyright-year>2007</copyright-year>
        <copyright-holder>Elsevier Masson SAS</copyright-holder>
        <license>
            <ali:license_ref>http://creativecommons.org/licenses/by/4.0/</ali:license_ref>
            <license-p>The patient figure in Figure 6, part <bold>A</bold> is reproduced from <xref ref-type="bibr" rid="bib22">Ferrero et al., (2007)</xref>
                <italic>European Journal of Medical Genetics</italic>
                with permission.
            </license-p>
        </license>
    </permissions>
</fig>
```

http://jats4r.org/

⟨JATS4R⟩

Online validation tool



Report level: Errors

JATS document URL:

or JATS file: No file chosen

or drop your JATS file here

Choose a JATS XML file to validate.

http://jats4r.org/validator/

http://jats4r.org/

〈JATS4R〉

Website redesign and additional resources

- Website
- XML learning centre
- Further recommendations

http://jats4r.org/

〈JATS4R〉

Getting involved

- Biweekly call
- Google drive
- Listserv

JATS4R / JATS4R Participant Hub

- Github repos
- Twitter: @jats4r

http://jats4r.org/

〈JATS4R〉

RIVER VALLEY TECHNOLOGIES

HighWire

NCBI National Center for Biotechnology Information

BIR The British Institute of Radiology

eLIFE

CAMBRIDGE UNIVERSITY PRESS

PLOS

Atypon

Thank you

m.harrison@elifesciences.org

F1000Research Open for Science

Canadian Science Publishing

PeerJ

frontiers

Europe PMC

http://jats4r.org/

OpenBook Publishers Knowledge is for sharing