BMC Bioinformatics

Correction

Open Access

Algorithm of OMA for large-scale orthology inference
Alexander CJ Roth*, Gaston H Gonnet* and Christophe Dessimoz*

Address: ETH Zurich, and Swiss Institute of Bioinformatics, 8092 Zurich, Switzerland

E-mail: Alexander CJ Roth* - alexande@inf.ethz.ch; Gaston H Gonnet* - gonnet@inf.ethz.ch; Christophe Dessimoz* - cdessimoz@inf.ethz.ch
*Corresponding author

Abstract

Since the publication of our article (Roth, Gonnet, and Dessimoz: BMC Bioinformatics 2008 9: 518), we have noticed several errors, which we correct in the following.

Correction

We lately identified inadvertent errors in our publication [1]. We regret these errors, and offer our sincere apologizes for the confusion and inconvenience. The corrections are described in detail in what follows.

Formation of stable pairs

In the main text of section "Formation of stable pairs", the formula for stable pair formation is missing a minus sign after the "greater than" symbol. The sentence should read:

"Formally, a pair of sequences (x, y) from genomes X and Y is considered a stable pair if and only if, for all $x_i \in X$, $x_i \neq x$, and for all $y_j \in Y$, $y_j \neq y$:

$$d_{xy_j} - d_{xy} > -k\sqrt{\sigma^2(d_{xy_j} - d_{xy})}$$

and

$$d_{x_i y} - d_{xy} > -k\sqrt{\sigma^2(d_{x_i y} - d_{xy})}$$

where d is a pairwise maximum likelihood distance estimate and k, the tolerance parameter of the standard deviation between the two distances, where $\sigma^2\left(d_{xy_j} - d_{xy}\right) = \sigma^2\left(d_{xy_j}\right) + \sigma^2\left(d_{xy}\right) - 2 \cdot \text{ccv}\left(d_{xy_j}, d_{xy}\right)$."

Ortholog clustering

In the subsection "Ortholog clustering", the example describing the clustering algorithm suggests that our algorithm could find best global maximum edge weight clique partition. This is incorrect, as our current implementation consists of a k-greedy approximation algorithm which is not guaranteed to find the best global maximum edge weight clique. To avoid any confusion, figure Eight has been updated (Fig. 1 here). The figure caption remains unchanged, but the numbers reported in the main text should now read:

"Figure 1A shows a graph with edges between all vertices except (z_1, z_2) and (z_1, y_2), which are paralogous relations. The highest scoring partition is $\{w_1, x_1, z_1\}$, $\{y_2, z_2\}$, with the total sum of edge weights of $4 \cdot 900 = 3600$. The score is higher than the
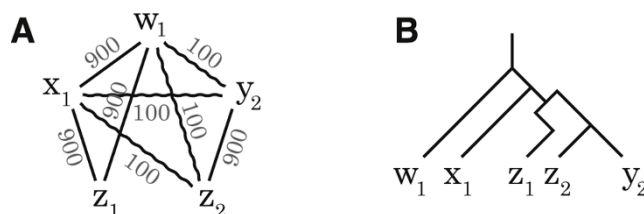

Figure 1
Corrected Figure Eight – Maximum edge weight cliques for inference of orthologous groups. **A** An example graph containing one 4-clique, four 3-cliques, and eight 2-cliques is provided. The highest edge scoring partition of the graph is $\{w_1, x_1, z_1\}$, $\{y_2, z_2\}$. **B** A possible evolutionary scenario corresponding to the graph.
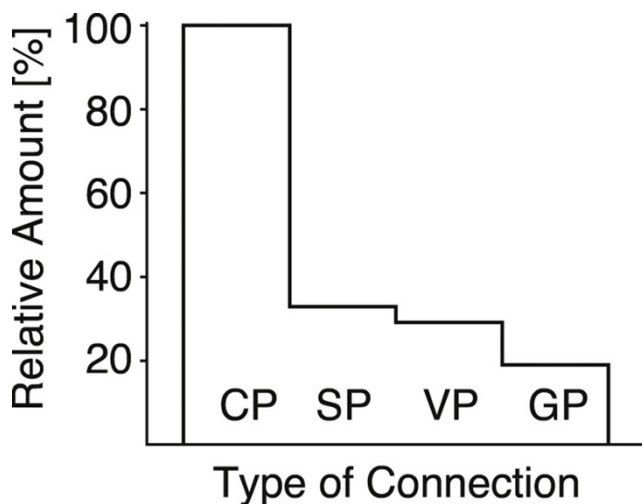
Figure 2
Corrected Figure Ten – Number of pairs reported after each step. Each step of the algorithm reduces the number of pairs, and the largest reduction is observed with the formation of stable pairs.

highest scoring maximum size clique $\{w_1, x_1, y_2, z_2\}$, $\{z_1\}$, where the sum of the scores is $2 \cdot 900 + 4 \cdot 100 = 2200$."

Figure Ten
In section "Results and discussion", figure Ten shows the decrease of the relative number of pairs after each step. The y-axis in the figure was scaled incorrectly. A corrected version is provided here (Fig. 2). The caption remains unchanged.

Bibliography Update
Finally, this correction gives us the opportunity to update a bibliographic reference: since publication of our original manuscript, Reference 33 [2], "in press" then, has been now published. The full reference is provided below.

References
1. Roth ACJ, Gonnet GH and Dessimoz C: **Algorithm of OMA for large-scale orthology inference.** *BMC Bioinformatics* 2008, **9:**518.
2. Altenhoff AM and Dessimoz C: **Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods.** *PLoS Comput Biol* 2009, **5:**e1000262.