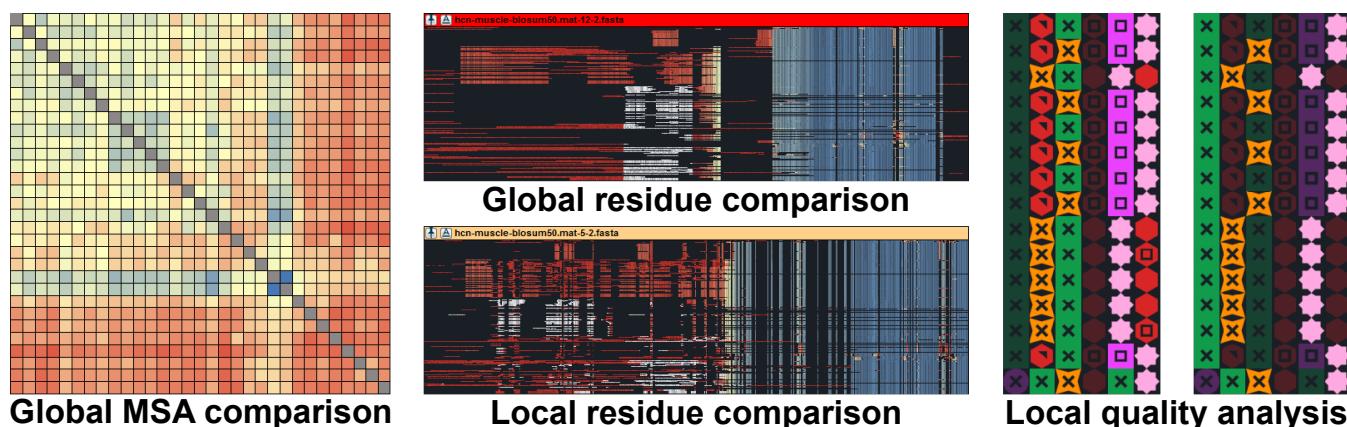


# Visual Analysis and Comparison of Multiple Sequence Alignments

M. Hess<sup>1</sup>, D. Jente<sup>1</sup>, J. Wiemeyer<sup>1</sup>, K. Hamacher<sup>1</sup> and M. Goesele<sup>1</sup>

<sup>1</sup>TU Darmstadt, Germany



**Figure 1:** Our approach allows a novel visual comparison and analysis of multiple MSAs on different levels of detail. Left: A Similarity Matrix depicting the pairwise MSA similarities of the dataset representing global comparison information. Center: Comparison of two large MSAs. For a local comparison, the MSA at the bottom highlights differently (red), slightly different (yellow) and consistently aligned residues (blue) with respect to the reference MSA shown on the top. In contrast, the reference MSA on the top shows this information globally with respect to all MSAs in the dataset. White highlights depict MSA differences for a specific selection. Right: Different highlighting modes allow the detection of potentially misaligned regions and local assessment of alignment quality.

## Abstract

Multiple Sequence Alignments (MSA) of a set of DNA, RNA or protein sequences form the fundamental basis for various biological applications such as evolutionary heritage and protein structure prediction. The quality of an MSA is crucial in order to provide useful and correct results. The analysis of MSAs is, however, still a challenging task since constructing MSAs is an NP hard problem and thus the optimal MSA is usually unknown. Additionally, MSA quality analysis is often completely ignored, especially by non-expert users, due to the lack of tools for the visual comparison and the intuitive quality assessment of alignments. In this paper, we present an interactive visual approach to simultaneously assess the quality of multiple alternative MSAs of the same set of sequences. We provide a direct assessment of the alignment quality using different highlighting techniques in combination with automatic quality measures and editing capabilities. An in-depth evaluation of our approach highlights its benefits for MSA quality assessment.

Categories and Subject Descriptors (according to ACM CCS): H.1.2 [Information Systems]: User/Machine Systems—Human information processing J.3 [Computer Applications]: Life and Medical Sciences—Biology and genetics

## 1. Introduction

One of the most fundamental tasks in computational biology is the comparison of a set of DNA, RNA or protein sequences by so called Multiple Sequence Alignments (MSA). Here, evolutionary

related (homologous) regions between the sequences are aligned to each other by relative shifts and insertions of gaps into the sequences. The resulting alignment allows to find corresponding regions and implicitly reveals similarities and differences in the underlying sequences set. These form the basis for several impor-

tant applications in biology such as evolutionary heritage or protein structure prediction. During the last years, several algorithms have been proposed that compute Multiple Sequence Alignments such as ClustalW [THG94, LBB\*07], MUSCLE [Edg04], MAFFT [KMKM02], Kalign [LFS09] and ClustalOmega [SWD\*11]. The basis for most of these algorithms (and the construction of MSAs in general) is a particular scoring model, i.e. a substitution matrix such as the BLOSUM- [HH92] or CorBLOSUM matrices [HKGH16] and an affine gap penalty model. The first models evolutionary substitution events which can be used as a similarity measure to identify homologous regions. The second penalizes the existence (Gap Opening Penalty) and the length (Gap Extension Penalty) of the aforementioned gaps. This models evolutionary insertion/deletion events which are considered to be more unlikely than substitution events and are thus penalized.

Notably, the selection of the optimal substitution matrix and the most suitable gap penalty values for a given alignment problem is a non trivial task and thus still part of current research [Edg09, KD13]. It depends on various factors such as the sequence data and the chosen MSA algorithm. There is no guarantee that a particular scoring model produces evolutionary correct alignments. In addition, the calculation of an optimal MSA for a particular scoring model is NP-hard [WJ94]. Hence, the aforementioned alignment methods rely on different heuristics to approximate the optimal but usually unknown alignment. This can result in strongly varying MSAs depending on the chosen algorithm, its parametrization such as the scoring model, and the sequence data itself.

As a result of these uncertainties, most users – in particular those without specific knowledge in the field of MSA – often generate MSAs using an arbitrary and thus potentially suboptimal alignment algorithm. Additionally, they often use default parameters which may result in MSAs with questionable quality. Since the quality of an MSA is crucial for many important applications that build upon these MSAs, the analysis of an MSA's quality is fundamental.

As mentioned above, the optimal scoring model for a given set of sequences and the corresponding optimal MSA is unknown. Hence, MSA quality is usually assessed with two different methods [KD13]: The first one is to calculate several alternative alignments for the same set of sequences using different alignment algorithms, parameter settings and scoring models. Afterwards, the generated alignments are analyzed by identifying consistently aligned regions over all alternative MSAs, which are considered to be valid [VA90]. The second method is to infer alignment quality only from the alignment structure itself by analyzing different criteria such as the number of gaps or the average symbol diversity in the MSA columns [KD13].

To our knowledge, most users do not use these tools to assess the quality of their calculated MSAs either due to the lack of knowledge of these methods or because they blindly trust the chosen algorithms. This trend is additionally reinforced by the lack of visual analysis and comparison tools supporting this task. In this paper, we address this issue by presenting a novel interactive visual comparison and analysis approach for MSAs. It enables users to visually explore, compare and analyze multiple MSAs in order to assess their quality and the impact of different MSA algorithms

and scoring models. Our main contributions and application benefits are:

- Our approach enables the visual comparison of large sets of alternative MSAs on global and local levels in order to assess the alignment quality and the impact of different scoring models and algorithms.
- In addition, we present a novel multi-scale visual analysis approach to visually assess the quality of alignments on different levels of detail using different highlighting techniques based on automatic quality assessment.
- For this, we also provide a novel quality measure called *Cluster Shift* that allows the user to reveal potentially misaligned regions.

## 2. Related Work

Our work relates to several topics in the field of Multiple Sequence Alignment such as MSA (quality) analysis, visualization and visual comparison. In the following, we first present and discuss measures for the comparison and quality analysis of MSAs. The remaining sections present state-of-the-art visualization and visual comparison techniques for MSAs in detail.

### 2.1. MSA Comparison and Quality Analysis

In the last years, several methods have been published that address the identification of unreliable alignment regions by comparing alternative alignments [VA90, MV96, Vin96, CHK02, LS05, SAKP15]. Here, consistently aligned regions across a set of alternative alignments are considered to be reliable and thus correctly aligned. In contrast, those with large differences are assumed to be unreliably aligned.

Another approach by Cline et al. [CHK02] called *shift score* uses alignment shift information in order to predict reliable positions. Here, for each residue  $x$  – i.e. a single symbol  $x$  in a sequence – in two alternative MSAs  $A$  and  $B$ , a shift value is calculated based on the residues aligned to  $x$  in these two MSAs. If  $x$  is aligned with the same residues in both MSAs, the shift is zero. Otherwise, the shift represents the difference between their indexes (Fig. 3). The advantage of this method is that it's calculated on a residue-to-residue basis which allows the detection of similar aligned regions independent of their actual column numbers in the different MSAs. Our approach uses the *shift score* measure as basis for the visual pairwise comparison of MSAs (Sections 3.1 and 3.4).

As mentioned in the introduction, there are also approaches to assess the quality of an alignment directly by analyzing its inner structure. Two of the most commonly used measures are the column based *consensus* and *sum-of-pairs* [TPP99] scores. The first measures the overall diversity of residues per column, often based on a simple majority rule with gaps optionally taken into account. The *sum-of-pairs score* (SOP) instead, represents the quality of an alignment column by accumulating the substitution scores of its residue pairs using a particular substitution matrix such as BLOSUM62 [HH92].

Both methods have the disadvantage that they do not take neighboring columns into account and thus their scores can be misleading. For example, a diverse and thus low scoring column may be

still well aligned if this reduces the number of gaps in the alignment and thus mitigates the overall gap penalty. In addition, substitution matrices rate identical residue pairs differently depending on their type (e.g. BLOSUM62: AA = 4 vs. WW = 11). Hence, there are multiple maximum SOP scores, preventing a normalized comparison between columns. Nevertheless, we adopted both techniques into our approach in order to provide state-of-the-art MSA analysis in form of a rough assessment of alignment quality.

Kececioglu and DeBlasio [KD13] presented further alignment-only based quality measures for Protein MSAs. Here, the average *sum-of-pairs score* based on the BLOSUM62 matrix, the *Gap Open Density* (i.e. the ratio between gap occurrence and overall gap length) and three secondary structure based measures (*Secondary Structure Blockiness, Agreement and Identity*) were reported to perform best. For the latter, PSIPRED [Jon99] was used to predict the secondary structure of the sequences. From these, the first one calculates the number of non overlapping blocks of residues with identical secondary structure annotation while the second measures the fraction of identical secondary structures per column. The last measure calculates the probability that two residues belongs to the same secondary structure type by averaging the normalized confidence values reported by PSIPRED for neighboring residues. Inspired by these measures, we developed a novel quality measure called *Cluster Shift* which allows the detection of potentially misaligned regions (Section 3.2).

## 2.2. MSA Visualizations

There are many different visualization tools available which support editing and/or viewing of sequence alignments (e.g. ClustalX2 [LBB\*07], Jalview [WPM\*09], SeaView [GGG10], Webprank [LG10], SuiteMSA [ASM11], SBAL [WBW\*12], AliView [Lar14]). Some of these tools provide additional functionality such as interfaces to prominent alignment algorithms or to web services for sequence retrieval or secondary structure prediction. All these tools provide, however, only limited capabilities to visually analyze alignment quality. As mentioned above, most of them use *consensus* or *sum-of-pairs* scores, which have several limitations, in order to provide qualitative feedback. In contrast, our approach provides a broader range of quality measures that are used to directly present quality information inside the MSA visualization using several highlighting techniques. Thereby, our method enables quality assessment on global as well as on local levels and allows to focus on similarities or differences depending on the user's demand. Another important limitation of the aforementioned visualizations is that they do not provide visual comparisons between multiple alignments.

## 2.3. Visual Comparison of MSAs

To our knowledge, there are only two approaches available that allow a visual comparison of MSAs: SinicView [SLL\*06] and SuiteMSA [ASM11]. SinicView focuses on genome-scale nucleotide alignments and does not provide residue-to-residue comparisons over multiple MSAs. Hence, we compare our approach primarily with SuiteMSA which provides two independent tools for alignment comparison: the MSA Comparator and the Pixel Plot

tool. The MSA Comparator allows the visual comparison of two MSAs – shown one below the other – supported by a bar chart indicating the corresponding column *sum-of-pairs* scores. Pixel Plot allows a simple comparison of multiple MSAs shown at a very small scale. In both approaches, an interval of neighboring columns can be selected to highlight the corresponding residues in the other displayed MSAs. While the MSA Comparator highlights differently and identically aligned residues using different colors, the Pixel Plot only shows where the residues are positioned in the MSAs limiting its usefulness for alignment comparison. Also, both approaches do not provide residue type based coloring and thus the inner structure of the alignments can be barely assessed. In addition, a zooming function is missing which limits the comparison of large MSAs and alignment editing is also not supported.

## 3. Approach

Typical analysis tasks to assess the quality of an MSA involve the identification of consistently and differently aligned regions in a set of alternative MSAs and the analysis of the inherent structure of an MSA (e.g. conserved columns and local misalignments). To address these tasks and the limitations mentioned in Section 2, our approach provides several methods for the interactive visual analysis and comparison of Protein Multiple Sequence Alignments. It is designed, to support these tasks on global as well as on local levels. Our system is implemented in Java 8 in order to ensure its platform independent usage and supports the import and export of edited MSAs in the commonly used FASTA format. By using multi threaded rendering and score calculations, our implementation allows a fast visualization of multiple large MSAs at once only limited by the user's display size. To mitigate this issue, we provide continuous zooming down to a single pixel per residue and multiple window support.

An overview of this system is shown in Fig. 2. An initial visual assessment of the relative MSA similarity in the dataset is provided by a sortable Similarity Matrix View (1) showing all pairwise MSA comparisons of the dataset (Section 3.4). From there or from the Project View (2), the user can open MSA Views (Fig. 5) or select a reference MSA for the comparison with other MSAs. The MSA Views can either be shown in the MSA List (3) or in separate windows for a detailed MSA analysis (see Section 3.4 and 3.5). They support several overlay and highlighting modes using transparency or colormaps (4) to show local and global quality or similarity information directly inside the MSA Visualizations. These modes are based on state-of-the-art comparison and quality measures and the here presented approach (see Sections 3.1 and 3.2). In addition, all visualizations are interlinked in order to provide information across multiple views.

### 3.1. MSA Comparison Measures

Our approach aims to provide automatic MSA comparison measures to support the user in assessing local and global differences and similarities across multiple MSAs. The basis for our measures is the *shift score* by Cline et al. [CHK02]. It allows an accurate assessment of local alignment differences and similarities on a per residue level. Since the *shift score* is based on pairwise alignments,



**Figure 2:** Overview of our visual comparison approach: (1) The Similarity Matrix View giving an overview over all pairwise MSA comparisons using the shift score by Cline et al. [CHK02]. The buttons below allow sorting of the matrix by several quality criteria. (2) The Project View showing all MSAs in the dataset. The sorting of the entries corresponds to the sorting of the Similarity Matrix View in order to determine high quality alignments. (3) The MSA List showing three different interlinked MSA Views. The MSA View on top of this list is chosen as reference indicated by the red title bar. This reference MSA View highlights consistently aligned regions (blue) on a global level – i.e. over the whole dataset – while the other views highlight these regions with respect to the chosen reference MSA. (4) The selection panel for switching between different quality overlays and colormaps. From here, the residues shown in the MSA Views can be filtered by setting a specific scoring threshold.

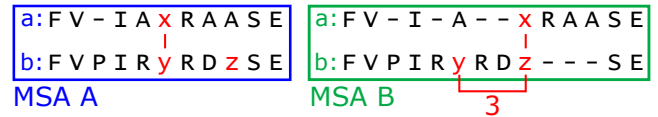
the following explanations always refer to two different sequences  $a$  and  $b$  that are both part of two alternative alignments  $A$  and  $B$ , i.e.  $a, b \in \{A, B\}$ . Thereby,  $x$  denotes a residue of sequence  $a$  while the residues  $y$  and  $z$  belong to sequence  $b$ .

For each residue  $x$  that is aligned with residue  $y$  in alignment  $A$  and residue  $z$  in alignment  $B$ , the *shift* value  $\delta(x)$  is defined as the difference of the indexes of  $y$  and  $z$  in sequence  $b$ . An example illustrating this for two pairwise alignments is shown in Fig. 3. In alignment  $A$  (left side), the red highlighted residue  $x$  in the upper sequence  $a$  is aligned with residue  $y$  of the lower sequence  $b$ . In alignment  $B$  (right side), the same residue  $x$  is aligned with residue  $z$  of the lower sequence  $b$ . The resulting *shift* value is  $\delta(x) = 3$  because the difference between the index of  $y$  and the index of  $z$  is three in sequence  $b$ .

Afterwards, the *shift* value  $\delta(x)$  is used to calculate the corresponding *shift score*  $\Delta(x)$ :

$$\Delta(x) = \frac{1 + \epsilon}{1 + |\delta(x)|} - \epsilon \quad (1)$$

Here,  $\epsilon$  is a scoring parameter that defines at which shift value the *shift score* drops into the negative range. Hence, the *shift score* ranges between  $-\epsilon$  and 1. As reported by the authors,  $\epsilon$  is normally set to 0.2. Thus accurate positions are scored with a *shift score* of 1, while small shifts are still positive reaching zero with a *shift* value of 5. Large shifts instead result in negative *shift scores* approach-

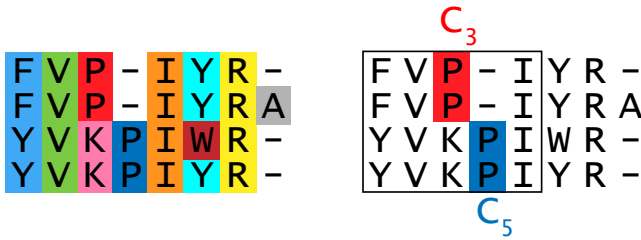


**Figure 3:** Illustration of the shift score measure. Shown are two alternative alignments of the same sequences  $A$  (left side) and  $B$  (right side). Left: The red residue  $x$  in the upper sequence  $a$  is aligned with residue  $y$  in the lower sequence  $b$ . Right: The same residue  $x$  in alignment  $B$  is aligned with residue  $z$  in sequence  $b$ . This results in a shift value of 3 because the difference between the index of  $y$  and the index of  $z$  is three in sequence  $b$ .

ing  $-\epsilon$ . Notably, the shift for residues that are aligned with gaps is undefined and thus results in a *shift score* of zero.

As mentioned above, our approach applies the *shift score* on local as well as on global levels. For the local comparison of two MSAs  $A$  and  $B$ , we calculate the average of all *shift scores*  $\Delta(x)$  obtained for a single residue  $x$  in alignment  $A$  to indicate its local difference to alignment  $B$ . We call this measure in the following *Average Residue Shift* and use it to highlight per residue differences between MSAs (see the lower MSA in Fig. 1).

In addition, we calculate another measure denoted as *Global Residue Shift* by calculating the average of all *Average Residue Shifts* obtained for a single residue in the comparison with all MSAs



**Figure 4:** Illustration of the Cluster Shift measure. In this example, a similarity level of  $\gamma = 3$  and the BLOSUM62 matrix was chosen for the clustering step. The search range for the second step was set to  $k = 2$ . Left: Shown are the obtained clusters from step 1 highlighted in different colors. Notably, the F and Y residues in Column 1 belong to the same cluster because the matrix value for the pair FY is  $S(F, Y) = 3 \geq \gamma$ . In contrast, P and K in Column 3 form separate clusters since  $S(K, P) = -1 < \gamma$ . The same occurs in Column 6 for Y and W because  $S(Y, W) = 2 < \gamma$ . Right: Here, the best matching cluster  $c_5$  (blue) to the query cluster  $c_3$  (red) is shown while the black box indicates the search window for  $c_3$ . The column offset between these clusters is 1 resulting in a Cluster Shift of  $\delta(c_3, c_5) = 1$ .

in the dataset. This allows us to assess a residue’s shift with respect to the whole dataset (see upper MSA shown in Fig. 1). The last measure denoted as *MSA Shift* is defined as the average of all *Average Residue Shift* obtained for all residues of alignment A with respect to MSA B. Thus, the *MSA Shift* represent scores for the pairwise comparison of two MSAs A and B which are used in our Similarity Matrix View (see Fig. 1, left). A detailed description of these visualization techniques is given in Section 3.4.

### 3.2. MSA Quality Measures

Analogous to the comparison measures used in our approach, our goal is to support the user’s analysis task by several MSA quality measures to select from. We provide state-of-the-art measures such as *sum-of-pairs score* [TPP99] based on a user selectable substitution matrix as well as a consensus style measure we call *Symbol Identity*. The latter calculates the fraction of amino acids per column that have the same equivalence class in spirit in of the *Amino Acid Identity* used by Kececioğlu and DeBlasio [KD13] (Fig. 1, right). These equivalence classes represent groups of amino acids that share similar chemical attributes (e.g. amino acids F, Y and W that have aromatic side chains). The user can choose from various reduced alphabets to specify these equivalence classes. These measures allow an easy but rough assessment of alignment quality and can be used to identify potentially well aligned columns. In addition, we use the aforementioned *Global Residue Shift* score per residue (Section 3.1) to measure the quality of a single residue with respect to the whole dataset (see upper MSA shown in Fig. 1). Here, residues with small *Global Residue Shift* values indicate consistently aligned regions in the whole dataset which are assumed to be reliably and thus correctly aligned (see Section 2).

Most MSA analysis approaches use column based quality measures which only allow the detection of potential misalignments

in single columns. Instead, our goal is to provide a measure that is able to detect potential misalignments across multiple columns. Hence, we developed a new quality criteria called *Cluster Shift* which searches for misalignments in neighboring columns in spirit of the *Secondary Structure Blockiness* measure introduced by Kececioğlu and DeBlasio [KD13].

The *Cluster Shift* measure is calculated in three steps: In the first step, we cluster the residues of each column based on their relative similarity which depends on two parameters, a substitution matrix  $S$  of the user’s choice and a user selectable similarity level  $\gamma$ . The parameter  $\gamma$  defines at which minimum similarity score two residues  $x$  and  $y$  are allowed to belong to the same cluster. Notably, acids of the same type always form a cluster and gap characters are ignored. An example clustering result (colored boxes) is illustrated on the left side of Fig. 4 for  $\gamma = 3$  and the BLOSUM62 matrix. All columns with the exception of Column 1, 3 and 6 only consist of identical acids and thus directly form clusters. The acids in Column 1 form a single cluster (light blue) since F and Y have a BLOSUM62 score of  $S(F, Y) = 3 \geq \gamma$ . In contrast, the P and K acids in Column 3 form two different clusters because the BLOSUM62 score for P and K is only  $S(K, P) = -1 < \gamma$ . The same occurs in Column 6 since the BLOSUM62 score of Y and W is  $S(Y, W) = 2$  which is still smaller than the required  $\gamma$ .

In the second step, we search the best matching cluster  $b_j$  for each identified cluster  $c_i$  in a range of  $k$  neighboring columns. Here,  $i$  and  $j$  denote the id of the corresponding cluster. The range parameter  $k$  can be selected by the user and defines the window in which locally misalignments should be detected. The best matching cluster  $b_j$  to a cluster  $c_i$  corresponds to that neighboring cluster  $c_j$  which have the highest similarity to cluster  $c_i$ . In addition, the set of sequences covered by  $c_i$  and  $c_j$  must be fully disjoint, i.e. the clusters are not allowed to overlap. The similarity corresponds to the average *sum-of-pairs* score of all residues in  $c_i$  to those in  $c_j$ . Fig. 4 (right side) shows an example for this step. Here, the best matching cluster of the red highlighted cluster  $c_3$  is cluster  $c_5$  colored in blue.

In the third and last step, we filter the obtained cluster-to-best-match assignments. We check for each identified best matching cluster  $b_j$  of cluster  $c_i$ , if  $c_i$  is also the best matching cluster for  $b_j$ . In other words, we test if both clusters refer to each other as the best matching cluster. If this is true, we keep these best match assignments. Otherwise, we drop these assignments. For the remaining clusters, we calculate the column offset between them in order to determine the corresponding *Cluster Shift Score*  $\delta(c_i, c_j)$ , i.e. the relative shift of the clusters. For example, the red and blue highlighted clusters on the right side of Fig. 4 have a column difference of 1 resulting in a *Cluster Shift* of  $\delta(c_3, c_5) = 1$ .

By highlighting clusters with a *Cluster Shift* below or above a specific threshold (Fig. 1, right), our approach allows to reveal potentially misaligned regions in an alignment based on relative amino acid similarity. Since the search window for this measure can be adaptively adjusted, the user can focus on local misalignments only or on those within a larger range. The visual concepts using the aforementioned measures are presented in detail in Section 3.5.

### 3.3. MSA Visualization and Editing

Since our approach focuses on assessing an MSA's quality, amino acid similarities play a central role. In order to visually emphasize these similarities and to assess various analysis aspects, our visualization allows to individually adapt the appearance of the shown residues, i.e. shape, color and texture. As shown on the left of Fig. 5, the color of a residue can be used, e.g. to indicate whether its corresponding amino acid type is hydrophobic or hydrophilic. Analogous, another amino acid attribute could be mapped to the residue's shape while the texture can be used for a third attribute. These three degrees of freedom can be freely assigned by the user depending on his or her demands, i.e. the user is not only restricted to potentially suboptimal visualization presets.

Nonetheless, we also offer different predefined style sets that are designed to emphasize residue similarities implicitly encoded by different substitution matrices such as BLOSUM62. By assigning different colors and shapes to groups of acids that share high acid-to-acid matrix values, the user can detect whether the acids in a column are similar and thus provide high scores to the MSA scoring function or not (Fig. 5, right). For example, the group of aliphatic amino acids I, L, M, and V are colored in red using a hexagonal shape. Since this combination of color and shape is unique, only these four acids appear as similar in the visualization. The texture of these acids is used to distinguish the different acid types. Several evaluations of this style set based on our previous work highlighted its usefulness for the visual assessment of local MSA quality especially by non-expert users, e.g., casual gamers [HWHG14].

Notably, the user can switch between custom shape or default shape rendering (colored squares) mode at any time. Analogous, the texture mode can be switched between custom texture, one letter code or disabled. The grid spacing of the alignment visualization can also be freely adjusted. In summary, the user can fully alter the MSA visualization on demand to focus on various interesting analysis aspects.

Our approach also features basic editing of MSAs. The user can easily select residues, e.g., by dragging a selection rectangle or by double clicking on a residue, which selects all residues of the same type in the corresponding column. The selected residues can be moved left or right using the arrow keys on the keyboard. Thereby, we offer different movements modes which can be triggered by holding a modifier key such as shift or alt. For example, the user can shift whole rows or multiple residues, either until neighboring residues are "touched" or by pushing those residues in the movement direction.

### 3.4. Visual Comparison of Multiple MSAs

As mentioned above, our approach aims to enable the visual comparison of multiple alternative MSAs on global as well as on local levels. This involves several challenges that need to be addressed. In the following, we describe these challenges and how we address them in detail.

One challenge is to provide an useful overview of the overall MSA similarities that scales well with the number of MSAs in the dataset. For this, we provide a Similarity Matrix View (see

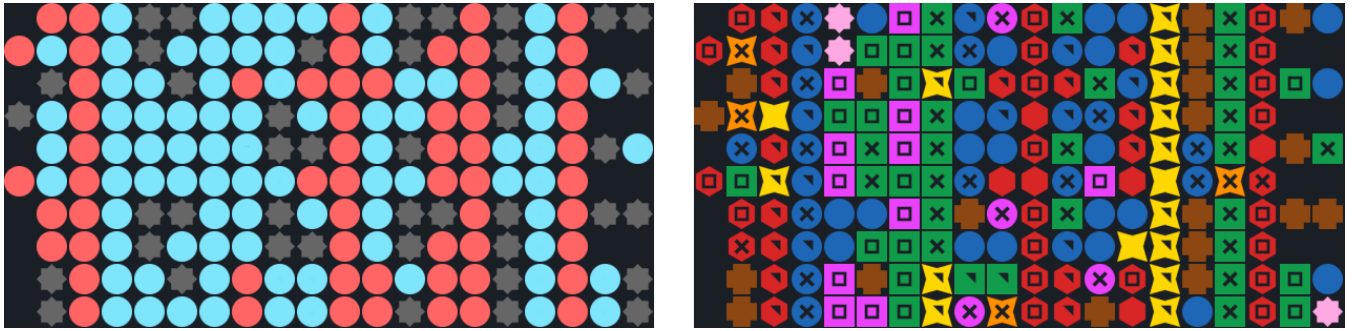
Fig. 1, left) showing the similarities of the MSAs in the dataset using different colors. Each cell in the matrix represents the similarity of two specific MSAs by mapping their corresponding *MSA Shift* score (Section 3.1) to a specific color retrieved from a user selectable colormap. This score can also be normalized in order to investigate small value differences. In Fig. 1, e.g., the "Red Yellow Blue" colormap is selected and thus cells representing the pairwise comparison of similar MSAs – i.e. those with shift scores near 1.0 – are highlighted in a bluish tone. In contrast, yellow colored cells indicate slightly different MSAs while those highlighted in red represent MSAs with different structure. Further details about single comparisons such as the corresponding MSAs and scores are shown in a tooltip by moving the mouse cursor over specific cells. Additionally, the user can sort the matrix by one of the implemented quality measures (see Section 3.2). This allows the user to get a fast overview of the overall quality of the dataset and the selection of interesting or conspicuous MSAs. For example, it enables a global scale analysis of the impact of different MSA algorithms or scoring models such as different substitution matrices and gap penalties.

Other challenges arise for the visual comparison of multiple alternative MSAs on local levels by comparing their inner structure. For example, one challenge is to visualize local differences in the pairwise comparison of two alignments. Another is the simultaneous assessment of local differences of a single MSA to all other MSAs in order to detect consistently aligned regions over the whole dataset. Since these regions are assumed to be correctly aligned (Section 2.1), this comparison can be especially helpful for non-expert users because it enables an easy assessment of alignment quality.

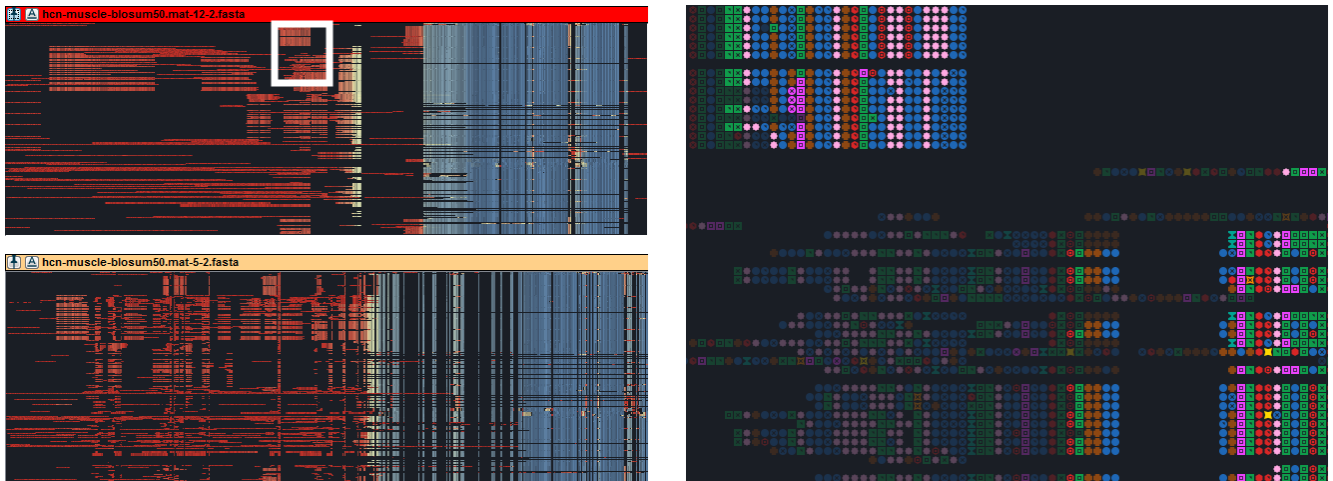
As mentioned in Section 3.1, we first address these challenges in an automatic fashion: We calculate the aforementioned *Average Residue Shift* and *Global Residue Shift* scores for each residue in an alignment with respect to all other alignments in the dataset. Afterwards, we use these scores to visually highlight differently and consistently aligned residues using transparency or different colors based on user selectable colormaps (Fig. 6). We offer two comparison modes: The first mode emphasizes this information with respect to a user selectable reference alignment and allows the pairwise visual comparison of MSAs. The second mode shows this information with respect to all MSAs which allows the detection of high quality regions in the whole dataset at once.

An example for these comparison view modes is shown on the left side in Fig. 6. Two MSAs are compared in detail with the reference MSA shown on top as indicated by the red title bar. Local per residue differences in the lower MSA with respect to the reference MSA – i.e. *Average Residue Shift* values (Section 3.1) – are highlighted using the "Red Yellow Blue" colormap. Consistently aligned regions are highlighted in blue while slightly and completely different regions are indicated by yellow and red colors, respectively. This allows the direct visual assessment of local alignment differences and similarities. In contrast, the highlights shown in the reference alignment are based on the *Global Residue Shift* (see Section 3.1) indicating per residue alignment differences with respect to the whole dataset. This enables the visual comparison of each residue in a global context.

One interesting aspect in the analysis of the per residue align-



**Figure 5:** Illustration of our MSA Visualization approach. Left: An example MSA shown with a custom appearance mapping focusing on two chemical properties. While cyan circles depict hydrophilic amino acids, red circles represent hydrophobic amino acids. Other amino acids are shown as dark gray stars in order to prevent visual distraction. Right: The same MSA using the BLOSUM62 style set which visually emphasizes the amino acid similarity information encoded in the BLOSUM62 substitution matrix. A group of acids with high matrix values to each other is highlighted with a specific color, shape and texture. For example, the aliphatic amino acids I, L, M and V are colored in red using hexagonal shapes. The texture of these acids is used to distinguish their different acid types inside a group. While the color of a group is unique and thus represents the most important visual similarity feature, the shape is also used to indicate weaker similarities across groups. For example, the pink and green squares represent acids that belong to different groups but still have similarities to each other.



**Figure 6:** Illustration of our visual MSA comparison approach. Left: A pairwise comparison of two large MSAs with the reference alignment shown on the top. The MSA shown at the bottom highlights local alignment differences and similarities with respect to the reference alignment. Consistently aligned regions are colored in blue while slightly or completely different regions are colored in yellow and red respectively. In contrast, the same colors in the reference MSA indicate this information on a global level, i.e. with respect to all other MSAs in the dataset. Right: A zoomed part of the reference MSA (white box) using the transparency mode. The view is filtered using a Global Residue Shift score threshold of  $\geq 0.12$  in order to focus only those residues that are consistently or slightly differently aligned in the whole dataset.

ment differences is the magnitude of the individual shifts. To support this analysis on a more detailed level, the user can set a threshold to filter residues with shift values smaller or greater and equal a specific threshold (see Fig. 6, right). Here, a small part of the reference alignment is shown in more detail (indicated by the white box on the left). This zoomed part of the reference alignment uses the transparency view mode and a *Global Residue Shift* threshold of  $\geq 0.12$ . Hence, only residues above this threshold are shown fully opaque to emphasize those regions that are consistently or only slightly different aligned with respect to whole dataset. Furthermore, the user can select specific or multiple residues in any

alignment to highlight their corresponding counterparts in all other MSA Views (see center of Fig. 1). This allows an on demand detail analysis of local alignment differences and similarities.

In summary, the comparison view mode allows a fast detection of consistently aligned regions on a global level – i.e. over the whole dataset – and locally with respect to the selected reference MSA. Since, consistently aligned regions over a large set of alignments are usually assumed to be correctly aligned [VA90], this mode allows to effectively judge alignment quality on a global and local level.

### 3.5. Visual Quality Analysis of MSAs

The second focus of our approach is the visual quality analysis of single MSAs solely based on their inherent inner structure. The challenge here is to provide reasonable quality measures and to visualize them in a comprehensible way. Similar to the comparison modes, we address this by offering the user several quality overlays which highlight well and potentially badly aligned residues using color and transparency. Again, this information is based on different automatic quality criteria (Section 3.2).

The visual quality assessment of an example MSA using different overlays is shown on the right side of Fig. 1. The quality overlay on the left highlights columns with different conservation based on the *Symbol Identity* measure and using the transparency mode. Analogous to the comparison mode, the user can also switch to a colormap and can filter the shown results. For example, the user can toggle the filter mode to focus on highly conserved columns instead of those with different conservation. This overlay provides a first but rough overview about locally good or potentially bad aligned MSA regions.

On the right side, the overlay uses our novel quality measure *Cluster Shift* for indicating potential misalignments. Here three different residue types are visually emphasized which could eventually be better aligned. The user can now easily correct these misalignments by selecting and adjusting the corresponding residues. Afterwards, these conspicuous regions are no longer highlighted indicating a local quality improvement of the alignment.

## 4. Results and Discussion

For the evaluation of our approach, we created a set of 104 alternative alignments of 211 HCN (Hyperpolarizationactivated cyclic nucleotide-gated) ion channel proteins using different algorithms and parameter settings (Sections 4.1 and 4.2). HCN ion channels are important for the functionality of cells in the heart and we thus address current research interests in biomedicine. To obtain our 211 sequences we run a BLAST search using as a query the known sequence `gi355749904`. As a threshold in the *E*-value we used 0.00001. In addition, we kept only those sequences annotated as "hyperpolarization" and "cyclic". In a third step, a molecular biologist and bioinformatician deleted thirteen additional sequences as non-hits. In total, we obtained a set of 211 valid HCN sequences [HBW\* 14].

Our evaluation covers two common analysis tasks that may occur when creating MSAs. Our first goal is to investigate alignment differences and similarities that are induced by the usage of different MSA algorithms with their default parameters. This covers the scenario that a user – and in particular a non-expert user – wants to create an MSA but does not know which MSA algorithm and which parameter setting to choose. Our second goal is to analyze the impact of different scoring models on the resulting MSAs when using a specific MSA algorithm. This analysis task represents the scenario that a user has chosen a favorite algorithm but is unsure which parameter settings work best for the given task. In the following sections, each scenario is discussed in detail.

### 4.1. Analysis of the impact of different MSA algorithms

One goal of our evaluation is to analyze the impact of different MSA algorithms on the resulting MSAs. We evaluated 8 MSAs generated with ClustalW2, ClustalOmega, MUSCLE, Kalign and MAFFT using their default settings. For the latter, we additionally used its three different high accuracy modes. In order to detect differences between the generated MSAs on a global level, we compared them using our Similarity Matrix View (Fig. 7). Here, the overall similarity between the MSAs was relatively high with *MSA Shift Scores* of  $\geq 0.9$ . In order to visually emphasize the differences between these scores, we switched the Similarity Matrix View to show the min/max normalized scores. Not surprisingly, the four different MAFFT parameterizations (columns 1 to 4) produced similar alignments with the Default MAFFT MSA being the most different one. In contrast, larger differences between the MSAs generated by MUSCLE, ClustalW2 and Kalign can be observed with normalized *MSA Shift* scores of  $< 0.12$  as indicated by the red cells. These alignments are also quite different to the four MAFFT MSAs with maximum *MSA Shift* scores of  $\sim 0.32$  for MUSCLE and ClustalW2 and  $\sim 0.52$  for Kalign, respectively. Interestingly, the MSA calculated using ClustalOmega (most right Column) is very consistent with all other MSAs.

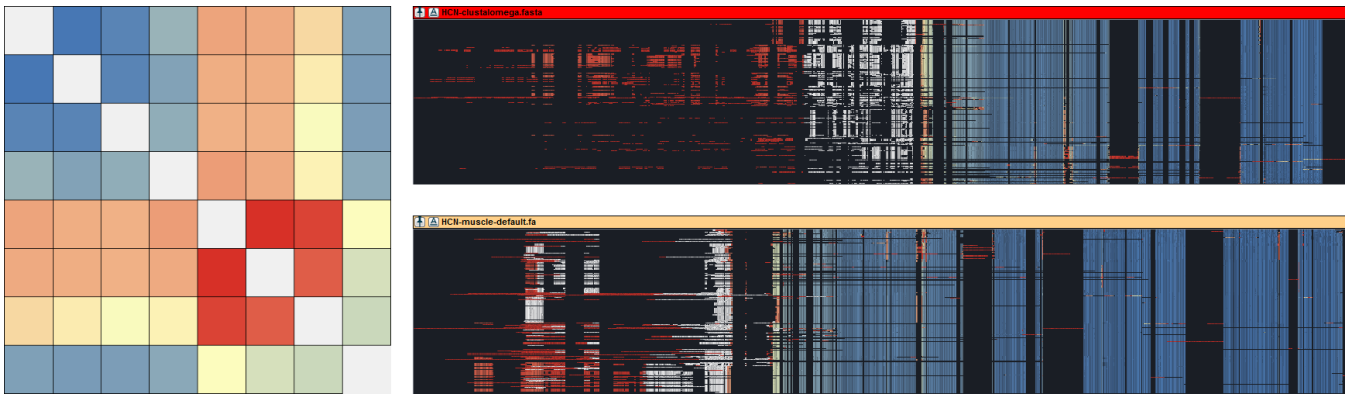
For a detailed analysis, we compared the MUSCLE MSA with the ClustalOmega MSA selected as reference (Fig. 7, right). Here, the leading parts of these large alignments are shown with the ClustalOmega MSA depicted at the top. Interestingly, the residues on the right side of the ClustalOmega MSA – i.e. the center of the MSAs – are highlighted in blue. This indicates that the blue regions are consistently aligned in the whole dataset independent of the chosen MSA algorithm and thus can be assumed to be correctly aligned. In contrast, the residues at the beginning of the sequences differ to a greater extent. This is visually emphasized by the white highlights showing selected residues in both MSAs. In MUSCLE these regions are more compactly aligned, while ClustalOmega tends to generate more spread out alignment blocks.

### 4.2. Analysis of the impact of different scoring models

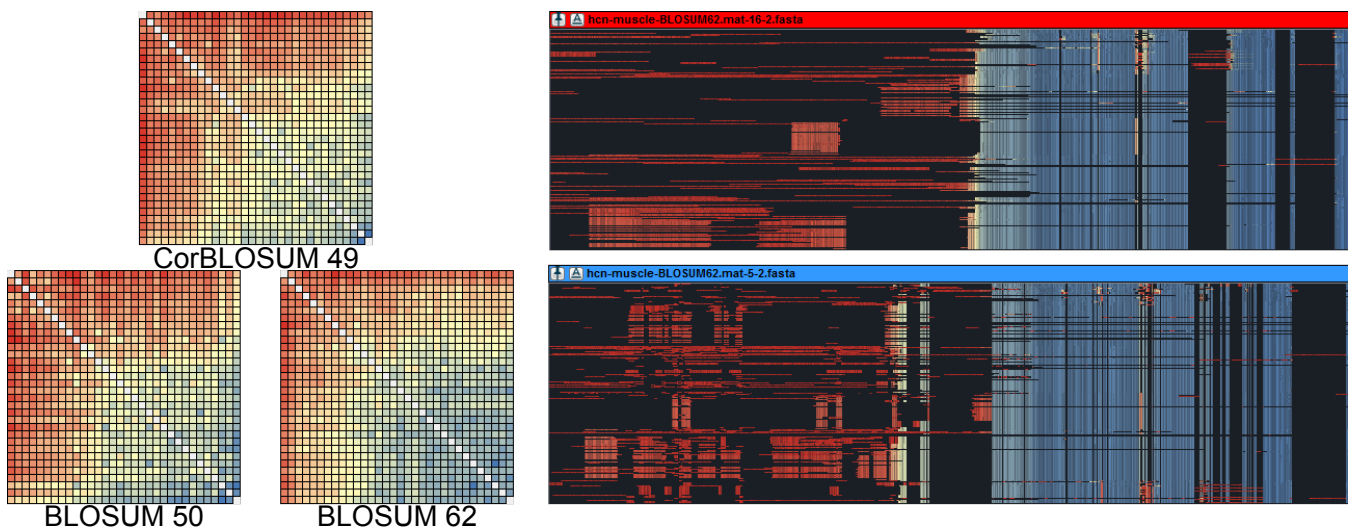
In the second part of our evaluation, we analyzed the effect of using different parameter settings with the same algorithm. We selected the MUSCLE algorithm as basis for our evaluation because MUSCLE is reported to be one the most accurate MSA algorithms. Using MUSCLE, we constructed three different MSA subsets based on the BLOSUM50, BLOSUM62 and CorBLOSUM49 matrix and different gap penalties. For the latter, we varied the gap open penalties between 5 and 20 and the gap extension penalties between 1 and 2, respectively. Similar to the first evaluation, we first investigated the MSA differences on a global level. Again, we analyzed the pairwise MSA similarities using normalized *MSA Shift* scores since the overall MSA similarity is rather high.

As shown in Fig. 8, the similarity of the created MSAs varied depending on the chosen substitution matrix. The MSAs based on the CorBLOSUM49 matrix showed mixed results. Here, the most consistent MSAs are those generated with relatively high gap open penalties. While the overall similarity of MSAs generated using the BLOSUM50 matrix is rather small, BLOSUM62 produced more





**Figure 7:** Left: Shown is the Similarity Matrix View depicting the pairwise comparison of the eight MSAs calculated by the different algorithms using their default settings. Right: The detailed comparison of the MUSCLE MSA (bottom) with the ClustalOmega MSA (top) selected as reference for the comparison. While the residues on the right side of the MSAs are consistently aligned (indicated by the blue colored residues), those red colored residues at the beginning of the sequences differ to a greater extent. To analyze these differences in more detail, a set of residues is selected in the upper MSA (white colored residues) to highlight their exact alignment differences in the lower MSA (white).



**Figure 8:** Assessment of the impact of different scoring models on the MSAs using MUSCLE. Left: The Similarity Matrix Views showing the overall MSA similarity for each of three MSA datasets. The usage of the BLOSUM62 produced the most stable MSAs. Right: Comparison between two BLOSUM62 MSAs generated with different gap open penalties. The reference MSA shown at the top was constructed using a gap opening penalty of 16, while the MSA on the bottom was calculated with a gap opening penalty of 5. The gap extension was set to 2 for both MSAs. As expected, the lower MSA shows more gaps in total on the left side, while the MSA on top is aligned more compactly. Independent of the chosen penalties, both MSAs consistently aligned the highly conserved regions on the right (blue).

consistent alignments. In summary, the most consistent MSAs were constructed using high gap open penalties of  $\geq 12$ . The gap extension penalty showed no apparent impact on the consistency of the generated MSAs.

In order to analyze the influence of the chosen gap penalties on a local level, we compared two different BLOSUM62 based MSAs in detail (see Fig. 8). Here, the BLOSUM62 MSA based on a gap opening penalty of 16 was chosen as reference shown on top. At the bottom, the comparison MSA based on a gap opening penalty of 5 is shown. As expected, the lower MSA shows much more gaps

in total on the left side, while the MSA on top is aligned in a more compact way. The highly conserved regions on the right (blue) are aligned consistently independent of the chosen penalties.

## 5. Limitations and Future Work

Even though our approach provides several benefits for the visual comparison and analysis of MSAs, there are still some limitations. Our approach only supports basic MSA editing in comparison to other tools, e.g. we do not provide auto-alignment or masking

mechanisms to hide specific parts of large alignments. Additionally, our evaluation is focused on the analysis task only and does not directly investigate the chosen visualizations. We plan to address these issues in the future. Additionally, we would like to integrate further quality criteria and information such as protein secondary structure to provide an even better MSA analysis approach.

## 6. Conclusion

We presented a novel interactive visual comparison and analysis approach to assess the quality of a set of alternative Multiple Sequence Alignments. Our approach combines automatic comparison and quality measures with different highlighting techniques to reveal differently and consistently alignment regions on global and local levels. This enables the users to visually explore, compare and analyze multiple alternative MSAs in order to assess their quality. We evaluated our approach on a large set of alternative MSAs based on real biological data (211 HCN ion channel proteins) that have current research interests in biomedicine. Here, we were able to assess the impact of different MSA algorithms and scoring models on the resulting alignments. In particular, our approach revealed consistently (and thus probably correct) aligned regions and was able to highlight alignment differences on a per residue level. This information led to a better understanding of different alignment parameters and enables a more reliable selection of a representative alignment for the use in further applications such as protein structure prediction.

## Acknowledgements

The authors are grateful for a TU Darmstadt FiF grant supporting this study. KH thanks the LOEWE project iNAPO funded by the Ministry of Higher Education, Research and the Arts (HMWK) of the Hessian state.

## References

- [ASM11] ANDERSON C. L., STROPE C. L., MORIYAMA E. N.: Suit-  
emsa: visual tools for multiple sequence alignment comparison and  
molecular sequence simulation. *BMC Bioinf.* 12, 1 (2011). 3
- [CHK02] CLINE M., HUGHEY R., KARPLUS K.: Predicting reliable  
regions in protein sequence alignments. *Bioinformatics* 18, 2 (2002). 2,  
3, 4
- [Edg04] EDGAR R. C.: Muscle: a multiple sequence alignment method  
with reduced time and space complexity. *BMC Bioinf.* 5, 1 (2004). 2
- [Edg09] EDGAR R. C.: Optimizing substitution matrix choice and gap  
parameters for sequence alignment. *BMC Bioinf.* 10, 1 (2009). 2
- [GGG10] GOUY M., GUINDON S., GASCUEL O.: Seaview version 4: A  
multiplatform graphical user interface for sequence alignment and phy-  
logenetic tree building. *Mol. Biol. Evol.* 27, 2 (2010). 3
- [HBW\*14] HESS M., BREMM S., WEISSGRAEBER S., HAMACHER K.,  
GOESELE M., WIEMEYER J., VON LANDESBERGER T.: Visual ex-  
ploration of parameter influence on phylogenetic trees. *IEEE Comput.  
Graphics Appl.* 34, 2 (2014). 8
- [HH92] HENIKOFF S., HENIKOFF J. G.: Amino acid substitution ma-  
trices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* 89, 22 (1992). 2
- [HKG16] HESS M., KEUL F., GOESELE M., HAMACHER K.: Ad-  
dressing inaccuracies in blosum computation improves homology search  
performance. *BMC Bioinf.* 17, 1 (2016). 2
- [HWHG14] HESS M., WIEMEYER J., HAMACHER K., GOESELE M.:  
Serious games for solving protein sequence alignments - combining citi-  
zen science and gaming. In *Games for Training, Education, Health and  
Sports*. Springer International Publishing, 2014. 6
- [Jon99] JONES D. T.: Protein secondary structure prediction based on  
position-specific scoring matrices. *J. Mol. Biol.* 292, 2 (1999). 3
- [KD13] KECECIOGLU J., DEBLASIO D.: Accuracy estimation and pa-  
rameter advising for protein multiple sequence alignment. *J. Comput.  
Biol.* 20, 4 (2013). 2, 3, 5
- [KMKM02] KATOH K., MISAWA K., KUMA K., MIYATA T.: Mafft: a  
novel method for rapid multiple sequence alignment based on fast fourier  
transform. *Nucleic Acids Res.* 30, 14 (2002). 2
- [Lar14] LARSSON A.: Aliview: a fast and lightweight alignment viewer  
and editor for large datasets. *Bioinformatics* 30, 22 (2014). 3
- [LBB\*07] LARKIN M., BLACKSHIELDS G., BROWN N., CHENNA R.,  
MCGETTIGAN P., MCWILLIAM H., VALENTIN F., WALLACE I.,  
WILM A., LOPEZ R., THOMPSON J., GIBSON T., HIGGINS D.: Clustal  
w and clustal x version 2.0. *Bioinformatics* 23, 21 (2007). 2, 3
- [LFS09] LASSMANN T., FRINGS O., SONNHAMMER E. L. L.: Kalign2:  
high-performance multiple alignment of protein and nucleotide se-  
quences allowing external features. *Nucleic Acids Res.* 37, 3 (2009). 2
- [LG10] LÖYTYNOJA A., GOLDMAN N.: webprank: a phylogeny-aware  
multiple sequence aligner with interactive alignment browser. *BMC  
Bioinf.* 11, 1 (2010). 3
- [LS05] LASSMANN T., SONNHAMMER E. L. L.: Automatic assessment  
of alignment quality. *Nucleic Acids Res.* 33, 22 (2005). 2
- [MV96] MEVISSSEN H. T., VINGRON M.: Quantifying the local reliabil-  
ity of a sequence alignment. *Protein Eng.* 9, 2 (1996). 2
- [SAKP15] SELA I., ASHKENAZY H., KATOH K., PUPKO T.: Guid-  
ance2: accurate detection of unreliable alignment regions accounting  
for the uncertainty of multiple parameters. *Nucleic Acids Res.* 43, W1  
(2015). 2
- [SLL\*06] SHIH A., LEE D., LIN L., PENG C.-L., CHEN S.-H., WU  
Y.-W., WONG C.-Y., CHOU M.-Y., SHIAO T.-C., HSIEH M.-F.:  
Sinicview: a visualization environment for comparisons of multiple nu-  
cleotide sequence alignment tools. *BMC Bioinf.* 7, 1 (2006). 3
- [SWD\*11] SIEVERS F., WILM A., DINEEN D., GIBSON T. J.,  
KARPLUS K., LI W., LOPEZ R., MCWILLIAM H., REMMERT M.,  
SÖDING J., THOMPSON J. D., HIGGINS D. G.: Fast, scalable genera-  
tion of high-quality protein multiple sequence alignments using clustal  
omega. *Mol. Syst. Biol.* 7, 1 (2011). 2
- [THG94] THOMPSON J. D., HIGGINS D. G., GIBSON T. J.: Clustal  
w: improving the sensitivity of progressive multiple sequence alignment  
through sequence weighting, position-specific gap penalties and weight  
matrix choice. *Nucleic Acids Res.* 22, 22 (1994). 2
- [TPP99] THOMPSON J. D., PLEWNIAK F., POCH O.: A comprehensive  
comparison of multiple sequence alignment programs. *Nucleic Acids  
Res.* 27, 13 (1999). 2, 5
- [VA90] VINGRON M., ARGOS P.: Determination of reliable regions in  
protein sequence alignments. *Protein Eng.* 3, 7 (1990). 2, 7
- [Vin96] VINGRON M.: Near-optimal sequence alignment. *Curr. Opin.  
Struct. Biol.* 6, 3 (1996). 2
- [WBW\*12] WANG C. K., BRODER U., WEERATUNGA S. K., GASSER  
R. B., LOUKAS A., HOFMANN A.: Sbal: a practical tool to generate and  
edit structure-based amino acid sequence alignments. *Bioinformatics* 28,  
7 (2012). 3
- [WJ94] WANG L., JIANG T.: On the complexity of multiple sequence  
alignment. *J. Comput. Biol.* 1, 4 (1994). 2
- [WPM\*09] WATERHOUSE A. M., PROCTER J. B., MARTIN D. M.,  
CLAMP M., BARTON G. J.: Jalview version 2.10: a multiple sequence  
alignment editor and analysis workbench. *Bioinformatics* 25, 9 (2009). 3