YLoc - an interpretable web server for predicting subcellular localization

Sebastian Briesemeister\textsuperscript{1,}\textsuperscript{*}, Jörg Rahnenführer\textsuperscript{2}, Oliver Kohlbacher\textsuperscript{1}

\textsuperscript{1} Div. for Simulation of Biological Systems, Universität Tübingen, Tübingen, Germany
\textsuperscript{2} Department of Statistics, TU Dortmund University, Dortmund, Germany

* E-mail: briese@informatik.uni-tuebingen.de

1 Detailed Performance Evaluation Results

Table 1, 2, 3, 4, and 5 show detailed evaluation results of the YLoc predictors. YLoc predictors that make no use of GO term-based features are marked with an asterisk. For every dataset the accuracy, the recall, the precision, and the F1 score is displayed for every location.

2 Cross-validation Evaluation

We evaluated all YLoc predictors in a 5-fold nested cross-validation evaluation scheme on the training datasets. Since MultiLoc2 was trained with the same datasets we compared both methods (see Table 6). Surprisingly, we did not observe a considerable difference between the predictors regarding prediction performance. All predictors show very high ACC and F1. The major difference between the methods lies in their complexity. MultiLoc2 is based on a complex SVM-ensemble classifier and uses hundreds of features, whereas YLoc is based on a very simple classification model.

3 Features of YLoc predictors

Table 7-24 list the features of the YLoc predictors. Every features is explained in detail and linked to a biological explanation. In cases, where we could not find a biological interpretation we transfered the biological explanation from a highly correlated feature. Predictors marked with * make no use of GO term-based features.
Table 1. Detailed performance comparison of the YLoc predictors

| Dataset           | Method        | ACC  | REC  | PRE  | F1  |
|-------------------|---------------|------|------|------|-----|
| BaCelLo Animals   | YLoc-LowRes   | 0.79 | 0.76 | 0.76 | 0.75|
|                   | YLoc-HighRes  | 0.74 | 0.70 | 0.70 | 0.69|
|                   | YLoc⁺         | 0.58 | 0.82 | 0.58 | 0.67|
|                   | YLoc-LowRes*  | 0.76 | 0.71 | 0.73 | 0.71|
|                   | YLoc-HighRes* | 0.71 | 0.70 | 0.68 | 0.69|
|                   | YLoc⁺*        | 0.58 | 0.84 | 0.56 | 0.67|
| BaCelLo Fungi     | YLoc-LowRes   | 0.56 | 0.65 | 0.66 | 0.61|
|                   | YLoc-HighRes  | 0.56 | 0.65 | 0.55 | 0.51|
|                   | YLoc⁺         | 0.48 | 0.70 | 0.42 | 0.51|
|                   | YLoc-LowRes*  | 0.53 | 0.64 | 0.60 | 0.58|
|                   | YLoc-HighRes* | 0.56 | 0.65 | 0.55 | 0.52|
|                   | YLoc⁺*        | 0.48 | 0.71 | 0.41 | 0.50|
| BaCelLo Plants    | YLoc-LowRes   | 0.71 | 0.63 | 0.56 | 0.58|
|                   | YLoc-HighRes  | 0.58 | 0.72 | 0.53 | 0.54|
|                   | YLoc⁺         | 0.53 | 0.75 | 0.45 | 0.49|
|                   | YLoc-LowRes*  | 0.58 | 0.44 | 0.44 | 0.42|
|                   | YLoc-HighRes* | 0.58 | 0.66 | 0.50 | 0.51|
|                   | YLoc⁺*        | 0.53 | 0.75 | 0.44 | 0.50|
| Höglund Animals   | YLoc-HighRes  | 0.56 | 0.31 | 0.43 | 0.34|
|                   | YLoc⁺         | 0.53 | 0.44 | 0.38 | 0.37|
|                   | YLoc-HighRes* | 0.60 | 0.37 | 0.39 | 0.36|
|                   | YLoc⁺*        | 0.56 | 0.44 | 0.42 | 0.40|

Performance of YLoc-LowRes, YLoc-LowRes*, YLoc-HighRes, YLoc-HighRes*, YLoc⁺, and YLoc⁺* on the BaCelLo IDS and the Höglund IDS animals regarding overall accuracy (ACC), average recall (REC), average precision (PRE), and average F1-score (F1). The performance of YLoc⁺ was measured using the generalized measures for multi-label classification.
### Table 2. Detailed performance comparison on the animal BaCelLo IDS

| Method        | Location | REC | PRE | F1  |
|---------------|----------|-----|-----|-----|
| YLoc-LowRes   | mi       | 0.77| 0.78| 0.78|
|               | SP       | 0.95| 0.83| 0.89|
|               | cy       | 0.43| 0.61| 0.51|
|               | nu       | 0.87| 0.82| 0.85|
| YLoc-LowRes*  | mi       | 0.70| 0.76| 0.73|
|               | SP       | 0.93| 0.80| 0.86|
|               | cy       | 0.36| 0.55| 0.44|
|               | nu       | 0.86| 0.79| 0.82|
| YLoc-HighRes  | mi       | 0.71| 0.77| 0.74|
|               | SP       | 0.91| 0.77| 0.84|
|               | cy       | 0.32| 0.50| 0.39|
|               | nu       | 0.86| 0.78| 0.82|
| YLoc-HighRes* | mi       | 0.73| 0.74| 0.70|
|               | SP       | 0.88| 0.74| 0.81|
|               | cy       | 0.45| 0.43| 0.44|
|               | nu       | 0.75| 0.82| 0.78|
| YLoc+         | mi       | 0.68| 0.66| 0.67|
|               | SP       | 0.86| 0.69| 0.77|
|               | cy       | 0.89| 0.46| 0.61|
|               | nu       | 0.85| 0.52| 0.65|
| YLoc+*        | mi       | 0.75| 0.60| 0.67|
|               | SP       | 0.91| 0.67| 0.77|
|               | cy       | 0.89| 0.45| 0.60|
|               | nu       | 0.81| 0.52| 0.64|

Detailed performance of YLoc-LowRes, YLoc-LowRes*, YLoc-HighRes, YLoc-HighRes*, YLoc+, and YLoc+* on the animal BaCelLo IDS regarding overall accuracy (ACC), average recall (REC), average precision (PRE), and average F1-score (F1). The performance of YLoc+ was measured using the generalized measures for multi-label classification.
Table 3. Detailed performance comparison on the fungi BaCelLo IDS

| Method      | Location | REC | PRE | F1  |
|-------------|----------|-----|-----|-----|
| YLoc-LowRes | mi       | 0.56| 0.66| 0.61|
|             | SP       | 0.89| 0.80| 0.84|
|             | cy       | 0.24| 0.63| 0.35|
|             | nu       | 0.91| 0.49| 0.65|
| YLoc-LowRes*| mi       | 0.62| 0.59| 0.60|
|             | SP       | 0.89| 0.76| 0.82|
|             | cy       | 0.18| 0.57| 0.27|
|             | nu       | 0.87| 0.49| 0.63|
| YLoc-HighRes| mi       | 0.42| 0.83| 0.56|
|             | SP       | 1.00| 0.22| 0.36|
|             | cy       | 0.46| 0.66| 0.54|
|             | nu       | 0.71| 0.51| 0.60|
| YLoc-HighRes*| mi      | 0.45| 0.83| 0.58|
|             | SP       | 1.00| 0.20| 0.33|
|             | cy       | 0.51| 0.64| 0.57|
|             | nu       | 0.65| 0.53| 0.58|
| YLoc⁺       | mi       | 0.44| 0.49| 0.46|
|             | SP       | 0.67| 0.29| 0.40|
|             | cy       | 0.89| 0.46| 0.61|
|             | nu       | 0.82| 0.45| 0.58|
| YLoc⁺⁺      | mi       | 0.47| 0.53| 0.50|
|             | SP       | 0.78| 0.21| 0.33|
|             | cy       | 0.79| 0.47| 0.59|
|             | nu       | 0.82| 0.44| 0.57|

Detailed performance of YLoc-LowRes, YLoc-LowRes*, YLoc-HighRes, YLoc-HighRes*, YLoc⁺, and YLoc⁺⁺ on the fungi BaCelLo IDS regarding overall accuracy (ACC), average recall (REC), average precision (PRE), and average F1-score (F1). The performance of YLoc⁺ was measured using the generalized measures for multi-label classification.
Table 4. Detailed performance comparison on the plant BaCelLo IDS

| Method          | Location | REC  | PRE  | F1   |
|-----------------|----------|------|------|------|
| YLoc-LowRes     | mi       | 0.50 | 0.31 | 0.38 |
|                 | ch       | 0.72 | 0.85 | 0.78 |
|                 | SP       | 0.67 | 0.33 | 0.44 |
|                 | cy       | 0.43 | 0.55 | 0.48 |
|                 | nu       | 0.85 | 0.76 | 0.80 |
| YLoc-LowRes*    | mi       | 0.17 | 0.06 | 0.09 |
|                 | ch       | 0.62 | 0.76 | 0.68 |
|                 | SP       | 0.33 | 0.13 | 0.19 |
|                 | cy       | 0.36 | 0.47 | 0.41 |
|                 | nu       | 0.72 | 0.76 | 0.74 |
| YLoc-HighRes    | mi       | 0.83 | 0.35 | 0.49 |
|                 | ch       | 0.42 | 0.95 | 0.59 |
|                 | SP       | 1.00 | 0.29 | 0.46 |
|                 | cy       | 0.57 | 0.28 | 0.38 |
|                 | nu       | 0.78 | 0.77 | 0.77 |
| YLoc-HighRes*   | mi       | 0.67 | 0.26 | 0.38 |
|                 | ch       | 0.46 | 0.90 | 0.61 |
|                 | SP       | 0.83 | 0.22 | 0.35 |
|                 | cy       | 0.53 | 0.34 | 0.42 |
|                 | nu       | 0.80 | 0.79 | 0.80 |
| YLoc+           | mi       | 0.83 | 0.25 | 0.39 |
|                 | ch       | 0.40 | 0.91 | 0.56 |
|                 | SP       | 0.83 | 0.24 | 0.38 |
|                 | cy       | 0.82 | 0.31 | 0.45 |
|                 | nu       | 0.88 | 0.53 | 0.66 |
| YLoc++          | mi       | 0.83 | 0.28 | 0.42 |
|                 | ch       | 0.45 | 0.87 | 0.59 |
|                 | SP       | 0.83 | 0.26 | 0.39 |
|                 | cy       | 0.82 | 0.32 | 0.46 |
|                 | nu       | 0.83 | 0.48 | 0.61 |

Detailed performance of YLoc-LowRes, YLoc-LowRes*, YLoc-HighRes, YLoc-HighRes*, YLoc+, and YLoc++ on the plants BaCelLo IDS regarding overall accuracy (ACC), average recall (REC), average precision (PRE), and average F1-score (F1). The performance of YLoc+ was measured using the generalized measures for multi-label classification.
Table 5. Detailed performance comparison on the animal Höglund IDS

| Method          | Location | REC  | PRE | F1  |
|-----------------|----------|------|-----|-----|
| YLoc-HighRes    | er       | 0.10 | 0.45| 0.16|
|                 | ex       | 0.81 | 0.86| 0.84|
|                 | go       | 0.14 | 0.33| 0.20|
|                 | ly       | 0.25 | 0.50| 0.33|
|                 | pe       | 0.00 | 0.00| 0.00|
|                 | pm       | 0.56 | 0.44| 0.49|
| YLoc-HighRes*   | er       | 0.10 | 0.29| 0.15|
|                 | ex       | 0.87 | 0.87| 0.87|
|                 | go       | 0.07 | 0.25| 0.11|
|                 | ly       | 0.25 | 0.23| 0.24|
|                 | pe       | 0.33 | 0.13| 0.18|
|                 | pm       | 0.62 | 0.60| 0.61|
| YLoc+           | er       | 0.12 | 0.42| 0.19|
|                 | ex       | 0.80 | 0.77| 0.79|
|                 | go       | 0.07 | 0.21| 0.11|
|                 | ly       | 0.25 | 0.12| 0.16|
|                 | pe       | 0.67 | 0.35| 0.46|
|                 | pm       | 0.74 | 0.43| 0.54|
| YLoc+*          | er       | 0.32 | 0.41| 0.36|
|                 | ex       | 0.84 | 0.76| 0.80|
|                 | go       | 0.07 | 0.21| 0.17|
|                 | ly       | 0.00 | 0.00| 0.00|
|                 | pe       | 0.67 | 0.44| 0.53|
|                 | pm       | 0.68 | 0.46| 0.55|

Detailed performance of YLoc-LowRes, YLoc-LowRes*, YLoc-HighRes, YLoc-HighRes*, YLoc+, and YLoc+* on the animals Höglund IDS regarding overall accuracy (ACC), average recall (REC), average precision (PRE), and average F1-score (F1). The performance of YLoc+ was measured using the generalized measures for multi-label classification.

Table 6. Cross-validation performance comparison

| Version       | MultiLoc2-LowRes | YLoc-LowRes |
|---------------|------------------|-------------|
| Animals       | 0.82 (0.84)      | 0.85 (0.87) |
| Fungi         | 0.79 (0.78)      | 0.77 (0.80) |
| Plants        | 0.79 (0.80)      | 0.77 (0.79) |

| Version       | MultiLoc2-HighRes | YLoc-HighRes | YLoc+ |
|---------------|-------------------|--------------|-------|
| Animals       | 0.84 (0.89)       | 0.85 (0.90)  | 0.77 (0.80) |
| Fungi         | 0.84 (0.89)       | 0.82 (0.90)  | 0.78 (0.81) |
| Plants        | 0.84 (0.89)       | 0.81 (0.89)  | 0.77 (0.80) |

5-fold cross-validation performance of MultiLoc2-LowRes and MultiLoc2-HighRes compared to the 5-fold nested cross-validation performance of YLoc-LowRes, YLoc-HighRes, and YLoc+ concerning the F1 and ACC (in brackets). The performance of YLoc+ was measured using the generalized measures for multi-label classification on the Höglund and the DBMLoc training dataset.
### Table 7. List of features of YLoc-LowRes Animals

| Biological interpretation                      | Detailed feature description                                                                 |
|-----------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                                  | number of amino acids                                                                      |
| mono NLS sorting signal                       | number non-hydrophobic amino acids in N-terminus and mono NLS signals                      |
| putative mitochondrial or secretory pathway sorting signal | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus |
| secretory pathway sorting signal              | autocorrelation of every third hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| putative mitochondrial sorting signal         | autocorrelation of every fifth charged amino acid within the first 30 amino acids in the N-terminus |
| charge of protein                              | overall autocorrelation of charged amino acid                                             |
| number of LxxL patterns in the N-terminus     | pseudo amino acid count of Lysine residues in a distant of three within the first 20 amino acids in the N-terminus |
| number of Cysteine in the N-terminus          | number of Cysteine within the first 30 amino acids in the N-terminus                      |
| number of Cysteine in the C-terminus          | maximal normalized pseudo amino acid count of Cysteine within the the last 30 amino acids in the C-terminus |
| hydrophobic N-terminus                        | maximal normalized pseudo amino acid count of very hydrophobic residues within the the first 90 amino acids in the N-terminus |
| negatively charged N-terminus                 | normalized number of negatively charged residues within the first 20 amino acids in the N-terminus |
| loop rich C-terminus                          | maximal pseudo amino acid count of turn favored amino acids within the the last 100 amino acids in the C-terminus |
| hydrophobic protein                           | overall maximal pseudo amino acid count of hydrophobic amino acids that often occur in beta strands CITVWY |
| typical cytoplasmic prosite pattern           | cluster of typical cytoplasmic prosite patterns                                            |
| typical plasma membrane prosite pattern       | cluster of typical plasma membrane prosite patterns                                       |
| GO:0005739 (mitochondrion)                    |                                                                                             |
| GO:0005634 (nucleus)                          |                                                                                             |
| GO:0005739 (transferase activity)             |                                                                                             |
| GO:0005576 (extracellular region)             |                                                                                             |
| GO term cluster                               |                                                                                             |

Description of the features of YLoc-LowRes Animals: All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 8. List of features of YLoc-LowRes Fungi

| Biological interpretation                  | Detailed feature description                                                                 |
|-------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                              | number of amino acids                                                                       |
| NLS sorting signal                        | number of NLS signal sequences from NLSDB                                                   |
| alternative NLS sorting signal            | autocorrelation of hydrophilic amino acids within the first 10 amino acids in the N-terminus |
| mitochondrial targeting peptide signal     | autocorrelation of every third charged amino acid within the first 20 amino acids in the N-terminus |
| positively charged N-terminus             | maximal autocorrelation of charged amino acids within the first 110 amino acids in the N-terminus |
| number of leucine in the N-terminus       | number of leucine within the first 20 amino acids in the N-terminus                         |
| number of alanine in the N-terminus       | number of alanine within the first 180 amino acids in the N-terminus                        |
| hydrophobic N-terminus                    | maximal normalized pseudo amino acid count of slightly hydrophobic residues within the first 70 amino acids in the N-terminus |
| hydrophobic C-terminus                    | pseudo amino acid count of slightly hydrophobic residues in a distant of four within the last 100 amino acids in the C-terminus |
| secretory pathway sorting signal          | pseudo amino acid count of unpolar residues in a distant of five within the first 20 amino acids in the N-terminus |
| number of polar amino acids               | maximal normalized overall pseudo amino acid count of polar residues                        |
| negatively charged N-terminus             | normalized number of negatively charged residues within the first 20 amino acids in the N-terminus |
| number of polar but not positively charged residues in N-terminus | maximal normalized pseudo amino acid count of [DEHNPQS] within the first 10 amino acids in the N-terminus |
| hydrophobic and uncharged N-terminus      | number of hydrophobic and uncharged residues within the first 50 amino acids in the N-terminus |
| polar but not positively charged protein  | Overall pseudo amino acid count of [DEHNPQS] in a distance of four                          |
| typical cytoplasmic prosite pattern       | cluster of typical cytoplasmic prosite patterns                                             |
| typical nuclear prosite                   | cluster of typical nuclear prosite patterns                                                 |
| GO:0005739 (mitochondrion)                |                                                                                             |
| GO:0005576 (extracellular region)         |                                                                                             |
| GO term cluster                           |                                                                                             |

Description of the features of YLoc-LowRes Fungi All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
### Table 9. List of features of YLoc-LowRes Plants

| Biological interpretation | Detailed feature description |
|---------------------------|------------------------------|
| protein size              | number of amino acids        |
| mono NLS sorting signal   | number non-hydrophobic amino acids in N-terminus and mono NLS signals |
| secretory pathway sorting signal | sum of hydrophobicity of the first 20 amino acids in the N-terminus |
| alternative mitochondrial targeting peptide | number of positively charged amino acid |
| putative mitochondrial or secretory pathway sorting signal | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus |
| amphiphilic helix         | maximum hydrophobic moment in a window of 18 residues |
| mitochondrial or chloroplast targeting peptide | autocorrelation of every fourth charged amino acid within the first 60 amino acids in the N-terminus |
| hydrophilic N-terminus    | normalized number of hydrophilic residues within the first 10 amino acids in the N-terminus |
| negatively charged very N-terminus | number of negatively charged residues within the first 50 amino acids in the N-terminus |
| negatively charged N-terminus | number of negatively charged residues within the first 170 amino acids in the N-terminus |
| uncharged C-terminus      | maximal normalized pseudo amino acid count of uncharged residues within the the last 100 amino acids in the C-terminus |
| positive charge of protein | overall pseudo amino acid count of positively charged residues in a distant of four |
| hydrophobic N-terminus    | maximal pseudo amino acid count of hydrophobic amino acids that often occur in beta strands CITVWY within the first 160 amino acids |
| number of potentially hydroxylated residues in the N-terminus | pseudo amino acid count of potentially hydroxylated residues in a distant of four within the first 140 amino acids in the N-terminus |
| polar but not positively charged protein | Overall pseudo amino acid count of [DEHNPQS] in a distance of four |
| typical nuclear prosite   | cluster of typical nuclear prosite patterns |
| GO:0005739 (mitochondrion) |                                |
| GO:0005576 (extracellular region) |                             |
| GO:0009507 (chloroplast)   |                                |
| GO term cluster           |                                |

Description of the features of YLoc-LowRes Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
### Table 10. List of features of YLoc-HighRes Animals

| Biological interpretation                  | Detailed feature description                                                                 |
|--------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                               | number of amino acids                                                                       |
| NLS sorting signal                         | number of NLS signal sequences from NLSDB.                                                   |
| ER retention signal                        | number of weak ER retention signals (number of K.D.E.L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRIQADNQEL, DEL, EL) at C-terminus is present it is 100. |
| alternative secretory pathway sorting signal| sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                      |
| hydrophobic and unpolar C-terminus         | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus |
| length of longest very hydrophobic region   | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| secretory pathway sorting signal           | autocorrelation of every hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| hydrophobic N-terminus                     | autocorrelation of every sixth hydrophobic amino acid within the first 70 amino acids in the N-terminus |
| mitochondrial targeting peptide            | sum of charge of first 30 amino acids in N-terminus                                           |
| charged protein                            | autocorrelation of every second charged amino acid                                           |
| weak NLS signal sequence                   | number of lysine within the first 10 amino acids in the N-terminus                           |
| number of leucine in the N-terminus        | maximal normalized pseudo amino acid count of leucine within the first 50 amino acids in the N-terminus |
| number of cysteine in the N-terminus       | maximal normalized pseudo amino acid count of cysteine within the first 170 amino acids in the N-terminus |
| alternative secretory pathway sorting signal| pseudo amino acid count of very hydrophobic residues in a distance of two within the first 50 amino acids in the N-terminus |
| hydrophobic protein                        | overall pseudo amino acid count of very hydrophobic residues in a distance of two             |
| negatively charged N-terminus              | number of negatively charged residues within the first 20 amino acids in the N-terminus       |
| uncharged protein                          | overall pseudo amino acid count of uncharged residues in a distance of two                   |
| number of small residues                   | overall pseudo amino acid count of small residues in a distance of six                       |
| number of non-aromatic residues            | maximal overall pseudo amino acid count of non-aromatic residues                              |
| number of hydrophobic uncharged residues   | minimal pseudo amino acid count of hydrophobic uncharged residues [ILVMFYWCTAG] from the amino acid at position 50 to the 20th last amino acid |
| the middle part of the protein             |                                                                                             |
| typical cytoplasmic prosite pattern        | cluster of typical cytoplasmic prosite patterns                                              |
| typical nuclear prosite pattern            | cluster of typical nuclear prosite patterns                                                  |
| typical plasma membrane prosite p.         | cluster of typical plasma membrane prosite patterns                                          |
| GO:0005737 (cytoplasm)                     |                                                                                             |
| GO:0005783 (endoplasmic reticulum)         |                                                                                             |
| GO:0005739 (mitochondrion)                 |                                                                                             |
| GO:0005576 (extracellular)                 |                                                                                             |
| GO:0005777 (peroxisome)                    |                                                                                             |
| GO:0005764 (lysosome)                      |                                                                                             |
| GO term cluster                            |                                                                                             |

Description of the features of YLoc-HighRes Animals All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 11. List of features of YLoc-HighRes Fungi

| Biological interpretation                                      | Detailed feature description                                                                 |
|---------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                                                 | number of amino acids                                                                       |
| NLS sorting signal                                           | number of NLS signal sequences from NLSDB.                                                  |
| ER retention signal                                          | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| alternative secretory pathway sorting signal                 | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                       |
| hydrophobic and unpolar C-terminus                            | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus |
| length of longest very hydrophobic region                     | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| putative mitochondrial or secretory pathway sorting signal     | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus    |
| secretory pathway sorting                                    | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| alternative secretory pathway sorting signal                 | autocorrelation of every hydrophobic amino acid within the first 90 amino acids in the N-terminus |
| mitochondrial targeting peptide                               | sum of charge of first 30 amino acids in N-terminus                                         |
| charged protein                                               | autocorrelation of every second charged amino acid                                         |
| number of leucine in the N-terminus                           | maximal normalized pseudo amino acid count of leucine within the first 40 amino acids in the N-terminus |
| number of lysine in the N-terminus                            | maximal pseudo amino acid count of lysine within the first 40 amino acids in the N-terminus |
| number of cysteine in the N-terminus                          | maximal pseudo amino acid count of cysteine within the first 170 amino acids in the N-terminus |
| overall number of serine                                      | overall maximal pseudo amino acid count of serine                                           |
| alternative secretory pathway sorting signal                 | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 60 amino acids in the N-terminus |
| hydrophobic protein                                           | maximal overall pseudo amino acid count of very hydrophobic residues                        |
| negatively charged N-terminus                                | number of negatively charged residues within the first 20 amino acids in the N-terminus     |
| uncharged protein                                             | overall pseudo amino acid count of uncharged residues in a distance of two                  |
| number of hydrophobic residues                                | minimal overall pseudo amino acid count of hydrophobic residues often occur in beta strands CITVWY |
| hydrophilic and non-positively charged N-terminus             | maximal normalized pseudo amino acid count of hydrophilic, non-positively charged residues DEHNPQS within the first 170 amino acids in the N-terminus |
| typical cytoplasmic prosite pattern                           | cluster of typical cytoplasmic prosite patterns                                              |
| typical nuclear prosite                                       | cluster of typical nuclear prosite patterns                                                  |
| typical plasma membrane prosite p.                           | cluster of typical plasma membrane prosite patterns                                          |
| GO:0005737 (cytoplasm)                                       |                                                                                             |
| GO:0005783 (endoplasmic reticulum)                           |                                                                                             |
| GO:0005739 (mitochondrion)                                   |                                                                                             |
| GO:0005576 (extracellular region)                            |                                                                                             |
| GO:0005777 (peroxisome)                                      |                                                                                             |
| GO term cluster                                               |                                                                                             |

Description of the features of YLoc-HighRes Fungi. All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 12. List of features of YLoc-HighRes Plants

| Biological interpretation                                      | Detailed feature description                                                                 |
|---------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                                                 | number of amino acids                                                                        |
| NLS sorting signal                                           | number of NLS signal sequences from NLSDB.                                                   |
| ER retention signal                                          | number of weak ER retention signals (number of K.D.E.L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| alternative secretory pathway sorting signal                 | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage.                       |
| hydrophobic and unpolar C-terminus                           | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus |
| unpolar cluster in the N-terminus                            | maximal normalized number of unpolar residues in a window of size 6.8,10 or 12 within the first 50 residues in the N-terminus |
| length of longest very hydrophobic region                    | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| alternative mitochondrial targeting peptide                  | number of positively charged amino acid                                                     |
| number of large amino acids in the N-terminus                | sum of volume of first 10 amino acids in the N-terminus                                      |
| secretary pathway sorting signal                              | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| hydrophobic N-terminus                                       | autocorrelation of every sixth hydrophobic amino acid within the first 70 amino acids in the N-terminus |
| charged N-terminus                                           | autocorrelation of every sixth charged amino acid within the first 60 amino acids in the N-terminus |
| number of leucine in the N-terminus                          | maximal normalized pseudo amino acid count of leucine within the first 40 amino acids in the N-terminus |
| number of cysteine in the N-terminus                         | maximal pseudo amino acid count of cysteine within the first 180 amino acids in the N-terminus |
| alternative chloroplast targeting signal                     | maximal pseudo amino acid count of weakly hydrophobic [AGHPSTY] within the first 40 amino acids in the N-terminus |
| alternative secretory pathway sorting signal                 | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 50 amino acids in the N-terminus |
| hydrophobic protein                                          | overall pseudo amino acid count of very hydrophobic residues in a distance of two            |
| negatively charged N-terminus                                | number of negatively charged residues within the first 20 amino acids in the N-terminus     |
| uncharged protein                                            | overall pseudo amino acid count of uncharged residues in a distance of two                  |
| hydrophilic and non-positively charged N-terminus            | number of hydrophilic, non-positively charged residues [DEHPNQS] within the first 30 amino acids in the N-terminus |
| number of hydrophilic, non-positively charged residues DEHPNQS| maximal overall normalized pseudo amino acid count of hydrophilic, non-positively charged residues DEHPNQS |
| typical cytoplasmic prosite pattern                          | cluster of typical cytoplasmic prosite patterns                                             |
| typical nuclear prosite                                       | cluster of typical nuclear prosite patterns                                                  |
| typical plasma membrane prosite p.                           | cluster of typical plasma membrane prosite patterns                                          |
| GO:0005783 (endoplasmic reticulum)                           |                                                                                             |
| GO:0005739 (mitochondrion)                                   |                                                                                             |
| GO:0005576 (extracellular region)                            |                                                                                             |
| GO:0005777 (peroxisome)                                      |                                                                                             |
| GO:0009507 (chloroplast)                                     |                                                                                             |
| GO term cluster                                              |                                                                                             |

Description of the features of YLoc-HighRes Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 13. List of features of YLoc+ Animals

| Biological interpretation | Detailed feature description |
|---------------------------|------------------------------|
| protein size              | number of amino acids        |
| ER retention signal       | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting signal (PTS) | weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.) |
| number of leucine and leucine clusters in the N-terminus | product of scores for leucine clusters (1.5 for LL, 2.5 for LLL, 5 for LLLL, 12 for LLLLL, 20 for LLLLLL) in the first 50 amino acids in the N-terminus |
| alternative secretory pathway sorting signal | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage |
| length of longest very hydrophobic region | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| putative mitochondrial sorting signal | weighted sum of typical amino acids for mitochondrial proteins in N-terminus |
| secretory pathway sorting peptide | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| mitochondrial targeting peptide | maximal autocorrelation of every sixth charged amino acid within the first 30 amino acids in the N-terminus |
| number of methionine in the N-terminus | number of methionine within the first 70 amino acids in the N-terminus |
| number of asparagine in the N-terminus | number of asparagine within the first 70 amino acids in the N-terminus |
| number of cysteine in the N-terminus | maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus |
| number of lysine in the N-terminus | maximal normalized pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus |
| number of tryptophane in the N-terminus | number of tryptophane within the first 120 amino acids in the N-terminus |
| hydrophobic N-terminus | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 130 amino acids in the N-terminus |
| hydrophobic C-terminus | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus |
| negatively charged N-terminus | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus |
| number of alpha-helix preferred residues AFGHKLMNR | minimal overall pseudo amino acid count of alpha-helix preferred residues |
| number of small amino acids in the N-terminus | number of small amino acids within the first 20 amino acids in the N-terminus |
| number of aromatic residues | maximal normalized overall pseudo amino acid count of aromatic residues |
| number of potentially hydroxylated residues in the C-terminus | pseudo amino acid count of potentially hydroxylated residues in a distance of three within the last 100 amino acids in the C-terminus |
| uncharged and hydrophobic protein | overall pseudo amino acid count of uncharged hydrophobic residues (I,L,V,M,F,Y,W,C,T,A,G) in a distant of two |
| typical plasma membrane prosite p. | cluster of typical plasma membrane prosite patterns |
| typical nuclear prosite pattern | cluster of typical nuclear prosite patterns |
| GO:0005783 (endoplasmic reticulum) | |
| GO:0005739 (mitochondrion) | |
| GO:0005576 (extracellular region) | |
| GO:0042025 (host cell nucleus) | |
| GO:0005778 (peroxisomal membrane) | |
| GO term cluster | |

Description of the features of YLoc+ Animals All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 14. List of features of YLoc+ Fungi

| Biological interpretation | Detailed feature description |
|---------------------------|-------------------------------|
| protein size              | number of amino acids         |
| ER retention signal       | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting signal (PTS) | weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.) |
| number of leucine and leucine clusters in the N-terminus | product of scores for leucine clusters (1.5 for LL, 2.5 for LLL, 5 for LLLL, 12 for LLLLL, 20 for LLLLLL) in the first 50 amino acids in the N-terminus |
| alternative secretory pathway sorting signal | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage |
| length of longest very hydrophobic region | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| putative mitochondrial sorting signal | weighted sum of typical amino acids for mitochondrial proteins in N-terminus |
| secretary pathway sorting signal | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| mitochondrial targeting peptide | maximal autocorrelation of every sixth charged amino acid within the first 30 amino acids in the N-terminus |
| number of asparagine in the N-terminus | number of asparagine within the first 70 amino acids in the N-terminus |
| number of Lysine in the N-terminus | maximal normalized pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus |
| number of tryptophane in the N-terminus | number of tryptophanee within the first 120 amino acids in the N-terminus |
| number of phenylalanine in the N-terminus | maximal normalized pseudo amino acid count of phenylalanine within the first 130 amino acids in the N-terminus |
| number of cysteine in the N-terminus | maximal pseudo amino acid count of cysteine within the first 130 amino acids in the N-terminus |
| hydrophobic N-terminus | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 130 amino acids in the N-terminus |
| hydrophobic C-terminus | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus |
| negatively charged N-terminus | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus |
| number of alpha-helix preferred residues AFGHKLMNR | minimal overall pseudo amino acid count of alpha-helix preferred residues |
| number of small amino acids in the N-terminus | number of small amino acids within the first 20 amino acids in the N-terminus |
| number of potentially hydroxylated residues in the C-terminus | pseudo amino acid count of potentially hydroxylated residues in a distance of three within the last 100 amino acids in the C-terminus |
| uncharged and hydrophobic protein | overall pseudo amino acid count of uncharged hydrophobic residues (I,L,V,M,F,Y,W,C,T,A,G) in a distant of two |
| typical plasma membrane prosite p. | cluster of typical plasma membrane prosite patterns |
| typical nuclear prosite | cluster typical nuclear prosite patterns |
| GO:0005783 (endoplasmic reticulum) | |
| GO:0005739 (mitochondrion) | |
| GO:0005576 (extracellular region) | |
| GO:0042025 (host cell nucleus) | |
| GO:0005773 (vacuole) | |
| GO:0005778 (peroxisomal membrane) | |
| GO term cluster | |

Description of the features of YLoc+ Fungi All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 15. List of features of YLoc+ Plants

| Biological interpretation | Detailed feature description |
|---------------------------|-----------------------------|
| protein size              | number of amino acids       |
| ER retention signal       | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| alternative secretory     | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage |
| pathway sorting signal    | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| length of longest very    | maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus |
| hydrophobic region        | maximal pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus |
| putative mitochondrial    | maximal pseudo amino acid count of phenylalanine within the first 120 amino acids in the N-terminus |
| sorting signal            | number of lysine in the N-terminus |
| putative mitochondrial or chloroplast targeting peptide | number of cysteine in the N-terminus |
| number of Serine in the N-terminus | number of Serine within the first 50 amino acids in the N-terminus |
| number of leucine in the N-terminus | maximal normalized pseudo amino acid count of leucine within the first 80 amino acids in the N-terminus |
| number of cysteine in the N-terminus | maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus |
| number of lysine in the N-terminus | maximal pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus |
| number of phenylalanine in the N-terminus | number of phenylalanine within the first 120 amino acids in the N-terminus |
| number of tryptophane in the N-terminus | number of tryptophane within the first 120 amino acids in the N-terminus |
| alternative secretory     | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 100 amino acids in the N-terminus |
| pathway sorting signal    | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus |
| hydrophobic C-terminus    | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus |
| negatively charged N-terminus | pseudo amino acid count of alpha-helix preferred residues [AFGHKLMNR] in a distance of three in the first 20 amino acids |
| number of alpha-helix preferred residues in the N-terminus | minimal overall pseudo amino acid count of alpha-helix preferred residues |
| number of alpha-helix preferred residues | number of large amino acids within the first 20 amino acids in the N-terminus |
| number of large amino acids in the N-terminus | number of potentially hydroxylated residues |
| number of potentially hydroxylated residues | maximal overall pseudo amino acid count of potentially hydroxylated residues |
| uncharged and hydrophobic protein | overall pseudo amino acid count of uncharged hydrophobic residues (L,L,V,M,F,Y,W,C,T,A,G) in a distant of two |
| typical plasma membrane prosite p. | cluster of typical plasma membrane prosite patterns |
| typical nuclear prosite    | cluster typical nuclear prosite patterns |
| GO:0005783 (endoplasmic reticulum) | |
| GO:0005739 (mitochondrion) | |
| GO:0005576 (extracellular region) | |
| GO:0042025 (host cell nucleus) | |
| GO:0009507 (chloroplast) | |
| GO term cluster | |

Description of the features of YLoc+ Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 16. List of features of YLoc-LowRes* Animals

| Biological interpretation | Detailed feature description |
|---------------------------|-----------------------------|
| protein size              | number of amino acids       |
| mono NLS sorting signal   | number non-hydrophobic amino acids in N-terminus and mono NLS signals |
| putative mitochondrial or secretory pathway sorting signal | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus |
| secretory pathway sorting signal | autocorrelation of every third hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| putative mitochondrial sorting signal | autocorrelation of every fifth charged amino acid within the first 30 amino acids in the N-terminus |
| charge of protein         | overall autocorrelation of charged amino acid |
| number of tryptophane in the N-terminus | number of tryptophane within the first 20 amino acids in the N-terminus |
| number of LxxL patterns in the N-terminus | pseudo amino acid count of leucine residues in a distant of three within the first 30 amino acids in the N-terminus |
| number of cysteine in the N-terminus | number of cysteine within the first 30 amino acids in the N-terminus |
| number of alanine in the N-terminus | maximal normalized pseudo amino acid count of alanine within the first 60 amino acids in the N-terminus |
| number of CxC patterns | overall pseudo amino acid count of cysteine residues in a distance of |
| hydrophobic N-terminus | maximal normalized pseudo amino acid count of very hydrophobic residues within the the first 90 amino acids in the N-terminus |
| negatively charged N-terminus | normalized number of negatively charged residues within the first 20 amino acids in the N-terminus |
| uncharged C-terminus | minimal pseudo amino acid count of uncharged residues within the the last 80 amino acids in the C-terminus |
| loop rich C-terminus | maximal pseudo amino acid count of turn favored amino acids within the the last 100 amino acids in the C-terminus |
| hydrophobic protein | overall maximal pseudo amino acid count of hydrophobic amino acids that often occur in beta strands CITVWY |
| typical cytoplasmic prosite pattern | cluster of typical cytoplasmic prosite patterns |
| typical secreted pathway prosite pattern | cluster of typical secreted pathway prosite patterns |
| typical mitochondrial prosite pattern | cluster of typical mitochondrial prosite patterns |
| typical nuclear prosite pattern | cluster of typical nuclear prosite patterns |

Description of the features of YLoc-LowRes* Animals All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
**Table 17. List of features of YLoc-LowRes* Fungi**

| Biological interpretation                  | Detailed feature description                                                                 |
|-------------------------------------------|-----------------------------------------------------------------------------------------------|
| protein size                              | number of amino acids                                                                        |
| NLS sorting signal                        | number of NLS signal sequences from NLSDB.                                                    |
| alternative NLS sorting signal            | autocorrelation of hydrophilic amino acids within the first 10 amino acids in the N-terminus |
| alternative secretory pathway sorting     | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                        |
| mitochondrial targeting peptide           | autocorrelation of every third charged amino acid within the first 20 amino acids in the N-terminus |
| positively charged N-terminus             | maximal autocorrelation of charged amino acids within the first 110 amino acids in the N-terminus |
|                                           | number of leucine within the first 20 amino acids in the N-terminus                          |
|                                           | number of alanine within the first 40 amino acids in the N-terminus                          |
|                                           | number of glycine within the first 180 amino acids in the N-terminus                         |
|                                           | number of glycine within the last 90 amino acids in the C-terminus                          |
| hydrophobic N-terminus                    | maximal normalized pseudo amino acid count of slightly hydrophobic residues within the first 70 amino acids in the N-terminus |
| secretory pathway sorting signal          | pseudo amino acid count of unpolar residues in a distance of five within the first 20 amino acids in the N-terminus |
|                                           | maximal normalized overall pseudo amino acid count of polar residues                         |
| number of polar amino residues            | normalized number of negatively charged residues within the first 20 amino acids in the N-terminus |
| negatively charged N-terminus             | overall pseudo amino acid count of large residues in a distance of four                      |
|                                           | maximal normalized pseudo amino acid count of [DEHNPQS] within the first 10 amino acids in the N-terminus |
| number of large amino acids               | number of hydrophobic and uncharged residues within the first 50 amino acids in the N-terminus |
| number of polar but not positively charged residues in N-terminus | overall pseudo amino acid count of [DEHNPQS] in a distance of four |
| hydrophobic and uncharged residues        | cluster of typical cytoplasmic prosite patterns                                              |
| polar but not positively charged protein  | cluster of typical nuclear prosite patterns                                                  |
| typical cytoplasmic prosite pattern       |                                                                                              |
| typical nuclear prosite pattern           |                                                                                              |

Description of the features of YLoc-LowRes* Fungi. All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
### Table 18. List of features of YLoc-LowRes* Plants

| Biological interpretation                  | Detailed feature description                                                                 |
|-------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                              | number of amino acids                                                                      |
| mono NLS sorting signal                   | number non-hydrophobic amino acids in N-terminus and mono NLS signals                      |
| secretory pathway sorting signal          | sum of hydrophobicity of the first 20 amino acids in the N-terminus                         |
| alternative mitochondrial targeting peptide| number of positively charged amino acid                                                    |
| putative mitochondrial or secretory pathway sorting signal | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus     |
| amphiphilic helix                         | maximum hydrophobic moment in a window of 18 residues                                       |
| mitochondrial or chloroplast targeting peptide| autocorrelation of every fourth charged amino acid within the first 60 amino acids in the N-terminus |
| number of cysteine in the C-terminus       | number of cysteine within the last 70 amino acids in the C-terminus                         |
| hydrophilic N-terminus                    | normalized number of hydrophilic residues within the first 10 amino acids in the N-terminus |
| negatively charged very N-terminus        | number of negatively charged residues within the first 50 amino acids in the N-terminus     |
| negatively charged N-terminus             | number of negatively charged residues within the first 170 amino acids in the N-terminus    |
| uncharged C-terminus                      | maximal normalized pseudo amino acid count of uncharged residues within the last 100 amino acids in the C-terminus |
| positive charge of protein                | overall pseudo amino acid count of positively charged residues in a distant of four        |
| hydrophobic N-terminus                    | maximal pseudo amino acid count of hydrophobic amino acids that often occur in beta strands CITVWY within the first 160 amino acids |
| number of large amino acids               | overall pseudo amino acid count of large residues in a distance of six                     |
| number of potentially hydroxylated residues in the N-terminus | pseudo amino acid count of potentially hydroxylated residues in a distance of four within the first 140 amino acids in the N-terminus |
| number of basic residues in the C-terminus | maximal pseudo amino acid count of basic residues within the last 90 amino acids in the C-terminus |
| polar but not positively charged protein  | Overall pseudo amino acid count of [DEHNPQS] in a distance of four                         |
| typical cytoplasmic prosite pattern       | cluster of typical cytoplasmic prosite patterns                                            |
| typical nuclear prosite                    | cluster typical nuclear prosite patterns                                                   |

Description of the features of YLoc-LowRes* Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 19. List of features of YLoc-HighRes* Animals

| Biological interpretation                  | Detailed feature description                                                                                                                                                 |
|--------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| protein size                               | number of amino acids                                                                                                                                                        |
| NLS sorting signal                         | number of NLS signal sequences from NLSDB.                                                                                                                                 |
| ER retention signal                        | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQ8ADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting sequence (PTS)       | C-terminal PTS1 signal sequence or N-terminal PTS2 signal sequence                                                                                                        |
| alternative secretory pathway sorting signal| sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                                                                                                      |
| hydrophobic and unpolar C-terminus         | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus                                                        |
| length of longest very hydrophobic region   | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases                                       |
| secretary pathway sorting signal           | autocorrelation of every hydrophobic amino acid within the first 20 amino acids in the N-terminus                                                                       |
| hydrophobic N-terminus                     | autocorrelation of every sixth hydrophobic amino acid within the first 70 amino acids in the N-terminus                                                                  |
| mitochondrial targeting peptide            | sum of charge of first 30 amino acids in N-terminus                                                                                                                        |
| charged protein                            | autocorrelation of every second charged amino acid                                                                                                                         |
| weak NLS signal sequence                   | number of lysine within the first 20 amino acids in the N-terminus                                                                                                |
| number of asparagine in the N-terminus     | maximal pseudo amino acid count of asparagine within the first 30 amino acids in the N-terminus                                                                         |
| number of leucine in the N-terminus        | maximal normalized pseudo amino acid count of leucine within the first 50 amino acids in the N-terminus                                                                   |
| number of tryptophane in the N-terminus    | number of tryptophane within the first 90 amino acids in the N-terminus                                                                                            |
| number of cysteine in the N-terminus       | maximal normalized pseudo amino acid count of cysteine within the first 170 amino acids in the N-terminus                                                                  |
| alternative secretory pathway sorting signal| pseudo amino acid count of very hydrophobic residues in a distance of two within the first 50 amino acids in the N-terminus                                                       |
| hydrophobic C-terminus                     | maximal pseudo amino acid count of very hydrophobic residues within the last 40 amino acids in the C-terminus                                                             |
| hydrophobic protein                        | overall pseudo amino acid count of very hydrophobic residues in a distance of two                                                                                         |
| negatively charged N-terminus              | number of negatively charged residues within the first 20 amino acids in the N-terminus                                                                                     |
| uncharged protein                          | overall pseudo amino acid count of uncharged residues in a distance of two                                                                                                 |
| number of small residues                   | overall pseudo amino acid count of small residues in a distance of six                                                                                                |
| number of non-aromatic hydroxylated residues| maximal overall pseudo amino acid count of non-aromatic residues                                                                                                           |
| number of potentially hydroxylated residues| maximal normalized overall pseudo amino acid count of potentially hydroxylated                                                                                           |
| number of hydrophobic residues in the middle part of the protein | minimal pseudo amino acid count of hydrophobic uncharged residues [ILVMFYWCTAG] from the amino acid at position 50 to the 20th last amino acid |
| typical cytoplasmic prosite pattern        | cluster of typical cytoplasmic prosite patterns                                                                                                                           |
| typical extracellular prosite pattern      | cluster of typical extracellular prosite patterns                                                                                                                         |
| typical mitochondrial prosite pattern      | cluster of typical mitochondrial prosite patterns                                                                                                                          |
| typical nuclear prosite pattern            | cluster of typical nuclear prosite patterns                                                                                                                                |
| typical plasma membrane prosite p.         | cluster of typical plasma membrane prosite patterns                                                                                                                        |

Description of the features of YLoc-HighRes* Animals All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
## Table 20. List of features of YLoc-HighRes* Fungi

| Biological interpretation                  | Detailed feature description                                                                 |
|-------------------------------------------|---------------------------------------------------------------------------------------------|
| protein size                              | number of amino acids                                                                       |
| NLS sorting signal                        | number of NLS signal sequences from NLSDB                                                     |
| ER retention signal                       | number of weak ER retention signals (number of K.D.E.L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQ8ADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting sequence (PTS)      | C-terminal PTS1 signal sequence or N-terminal PTS2 signal sequence                           |
| alternative secretory pathway sorting signal | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                      |
| hydrophobic and unpolar C-terminus        | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus |
| length of longest very hydrophobic region  | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to zero mean drops below zero in at most two cases |
| putative mitochondrial or secretory pathway sorting signal | weighted sum of typical amino acids for mitochondrial and secreted proteins in N-terminus |
| secretory pathway sorting signal          | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| hydrophobic C-terminus                    | autocorrelation of every third hydrophobic amino acid within the last 30 amino acids in the C-terminus |
| mitochondrial targeting                   | sum of charge of first 30 amino acids in N-terminus                                          |
| charged protein                           | autocorrelation of every second charged amino acid                                           |
| number of leucine in the N-terminus       | maximal normalized pseudo amino acid count of leucine within the first 40 amino acids in the N-terminus |
| number of lysine in the N-terminus        | maximal pseudo amino acid count of lysine within the first 40 amino acids in the N-terminus |
| number of cysteine in the N-terminus      | maximal pseudo amino acid count of cysteine within the first 170 amino acids in the N-terminus |
| number of serines                         | maximal overall pseudo amino acid count of serine                                            |
| alternative secretory pathway sorting signal | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 60 amino acids in the N-terminus |
| hydrophobic protein                       | maximal overall pseudo amino acid count of very hydrophobic residues                         |
| negatively charged N-terminus             | number of negatively charged residues within the first 20 amino acids in the N-terminus      |
| uncharged protein                         | overall pseudo amino acid count of uncharged residues in a distance of                       |
| number of beta-sheet preferred residues CITVWWY | minimal overall pseudo amino acid count of very hydrophobic residues [CITVWWY]          |
| number of aromatic residues               | maximal normalized overall pseudo amino acid count of aromatic residues                     |
| number of potentially hydroxylated residues in the N-terminus | number of potentially hydroxylated residues within the first 110 amino acids in the N-terminus |
| hydrophilic and non-positively charged N-terminus | maximal normalized pseudo amino acid count of hydrophilic, non-positively charged residues DEHNPQS within the first 170 amino acids in the N-terminus |
| typical cytoplasmic prosite pattern       | cluster of typical cytoplasmic prosite patterns                                             |
| typical extracellular prosite pattern      | cluster of typical extracellular prosite patterns                                            |
| typical mitochondrial prosite pattern      | cluster of typical mitochondrial prosite patterns                                           |
| typical nuclear prosite pattern           | cluster of typical nuclear prosite patterns                                                  |
| typical plasma membrane prosite p.        | cluster of typical plasma membrane prosite patterns                                         |

Description of the features of YLoc-HighRes* Fungi All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
### Table 21. List of features of YLoc-HighRes* Plants

| Biological interpretation | Detailed feature description |
|---------------------------|------------------------------|
| protein size              | number of amino acids        |
| NLS sorting signal        | number of NLS signal sequences from NLSDB. |
| ER retention signal       | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| alternative secretory     | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage |
| pathway sorting signal    | sum of volume of first 10 amino acids in N-terminus |
| hydrophobic and unpolar   | maximal number of hydrophobic unpolar residues in a window of size 24 within the last 150 residues in the C-terminus |
| C-terminus                | maximal normalized number of unpolar residues in a window of size 6,8,10 or 12 within the first 50 residues in the N-terminus |
| unpolar cluster in the    | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| N-terminus                | number of positively charged amino acid |
| length of longest very    | number of lysine within the first 20 amino acids in the N-terminus |
| hydrophobic region        | maximal normalized pseudo amino acid count of leucine within the first 40 amino acids in the N-terminus |
| alternative mitochondrial | maximal pseudo amino acid count of cysteine within the first 180 amino acids in the N-terminus |
| targeting peptide         | maximal pseudo amino acid count of slightly hydrophobic residues within the first 40 amino acids in the N-terminus |
| number of large amino     | number of positively charged residues within the first 20 amino acids in the N-terminus |
| acids in the N-terminus   | overall pseudo amino acid count of charged residues [DEHNPQS] within the first 30 amino acids in the N-terminus |
| secretory pathway sorting | overall pseudo amino acid count of non-aromatic residues DEHNPQS |
| signal                    | number of very hydrophobic residues [CITVWY] within the first 160 amino acids in the N-terminus |
| hydrophobic N-terminus    | number of very hydrophobic residues [CITVWY] within the first 160 amino acids in the N-terminus |
| charged N-terminus        | number of very hydrophobic residues [CITVWY] within the first 160 amino acids in the N-terminus |
| weak NLS signal sequence  | number of positively charged residues within the first 20 amino acids in the N-terminus |
| number of leucine in the  | maximal overall pseudo amino acid count of non-aromatic residues DEHNPQS |
| N-terminus                | number of hydrophilic, non-positively charged residues DEHNPQS |
| number of cysteine in the | number of hydrophilic, non-positively charged residues DEHNPQS |
| N-terminus                | number of hydrophilic, non-positively charged residues DEHNPQS |
| number of slightly        | number of hydrophilic, non-positively charged residues DEHNPQS |
| hydrophobic residues in    | number of hydrophilic, non-positively charged residues DEHNPQS |
| the N-terminus            | number of hydrophilic, non-positively charged residues DEHNPQS |
| alternative secretory     | number of hydrophilic, non-positively charged residues DEHNPQS |
| pathway sorting signal    | number of hydrophilic, non-positively charged residues DEHNPQS |
| hydrophobic protein       | number of hydrophilic, non-positively charged residues DEHNPQS |
| negatively charged        | number of hydrophilic, non-positively charged residues DEHNPQS |
| N-terminus                | number of hydrophilic, non-positively charged residues DEHNPQS |
| uncharged protein         | number of hydrophilic, non-positively charged residues DEHNPQS |
| number of beta-sheet      | number of hydrophilic, non-positively charged residues DEHNPQS |
| preferred residues CITVWY | number of hydrophilic, non-positively charged residues DEHNPQS |
| in the N-terminus         | number of hydrophilic, non-positively charged residues DEHNPQS |
| number of non-aromatic    | number of hydrophilic, non-positively charged residues DEHNPQS |
| residues in the N-terminus| number of hydrophilic, non-positively charged residues DEHNPQS |
| number of potentially     | number of hydrophilic, non-positively charged residues DEHNPQS |
| hydroxylated residues in  | number of hydrophilic, non-positively charged residues DEHNPQS |
| the N-terminus            | number of hydrophilic, non-positively charged residues DEHNPQS |
| hydrophilic and non-positively charged N-terminus | number of hydrophilic, non-positively charged residues DEHNPQS |
| number of hydrophilic,    | number of hydrophilic, non-positively charged residues DEHNPQS |
| non-positively charged    | number of hydrophilic, non-positively charged residues DEHNPQS |
| residues                  | number of hydrophilic, non-positively charged residues DEHNPQS |
| typical cytoplasmic prosite pattern | number of hydrophilic, non-positively charged residues DEHNPQS |
| typical extracellular prosite pattern | number of hydrophilic, non-positively charged residues DEHNPQS |
| typical mitochondrial prosite pattern | number of hydrophilic, non-positively charged residues DEHNPQS |
| typical nuclear prosite pattern | number of hydrophilic, non-positively charged residues DEHNPQS |
| typical plasma membrane prosite pattern | number of hydrophilic, non-positively charged residues DEHNPQS |

Description of the features of YLoc-HighRes* Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
Table 22. List of features of YLoc+* Animals

| Biological interpretation                        | Detailed feature description                                                                 |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------|
| protein size                                     | number of amino acids                                                                         |
| ER retention signal                              | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting signal (PTS)               | weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.)     |
| number of leucine and                            | product of scores for leucine clusters (1.5 for LL, 2.5 for LLL, 5 for LLLL, 12 for LLLLL, 20 for LLLLLL) in the first 50 amino acids in the N-terminus |
| leucine clusters in the N-terminus               |                                                                                               |
| alternative secretory                            | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                        |
| pathway sorting signal                           |                                                                                               |
| length of longest very                            | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| hydrophobic region                               |                                                                                               |
| putative mitochondrial                           | weighted sum of typical amino acids for mitochondrial proteins in N-terminus                  |
| sorting signal                                   |                                                                                               |
| secretory pathway sorting signal                 | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| mitochondrial targeting peptid                   | maximal autocorrelation of every sixth charged amino acid within the first 30 amino acids in the N-terminus |
| number of asparagine in the N-terminus           | number of asparagine within the first 70 amino acids in the N-terminus                         |
| number of serines in the N-terminus              | pseudo amino acid count of serines in a distance of two within the first 90 amino acids in the N-terminus |
| number of cysteine in the N-terminus             | maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus |
| number of lysine in the N-terminus               | maximal normalized pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus |
| number of methionine in the N-terminus           | number of methionine within the first 120 amino acids in the N-terminus                        |
| number of tryptophane in the N-terminus          | number of tryptophane within the first 120 amino acids in the N-terminus                      |
| hydrophobic N-terminus                           | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 130 amino acids in the N-terminus |
| hydrophobic C-terminus                           | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus |
| negatively charged N-terminus                    | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus |
| number of alpha-helix preferred residues AFGHKLMNR| minimal overall pseudo amino acid count of alpha-helix preferred residues                     |
| number of small amino acids in the N-terminus    | number of small amino acids within the first 20 amino acids in the N-terminus                 |
| number of aromatic residues                      | maximal normalized overall pseudo amino acid count of aromatic residues                       |
| number of potentially hydroxylated residues in the C-terminus | pseudo amino acid count of potentially hydroxylated residues in a distance of three within the last 100 amino acids in the C-terminus |
| uncharged and hydrophobic protein                | overall pseudo amino acid count of uncharged hydrophobic residues (I,L,V,M,F,Y,W,C,T,A,G) in a distant of two |
| PS00639 (cysteine proteases)                     | Prosite pattern cysteine proteases often in lysosomal proteins                                 |
| PS50041 (C-type lectin domain)                   | Prosite pattern C-type lectin domain present in membrane and extracellular                     |
| typical extracellular prosite pattern            | cluster of typical extracellular prosite patterns                                             |
| typical plasma membrane prosite p.               | cluster of typical plasma membrane prosite patterns                                         |
| typical nuclear prosite pattern                  | cluster of typical nuclear prosite patterns                                                  |
| typical mitochondrial prosite pattern            | cluster of typical mitochondrial prosite patterns                                          |
| typical cytoplasmic prosite pattern              | cluster of typical cytoplasmic prosite patterns                                               |

Description of the features of YLoc+* Animals All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.
### Table 23. List of features of YLoc+* Fungi

| Biological interpretation                                      | Detailed feature description                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| protein size                                                  | number of amino acids                                                                                                                                                                                                                                                                                                                                        |
| ER retention signal                                           | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting signal (PTS)                           | weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.)                                                                                                                                                                                                           |
| number of leucine and leucine clusters in the N-terminus      | product of scores for leucine clusters (1.5 for LL, 2.5 for LLL, 5 for LLLL, 12 for LLLLL, 20 for LLLLLL) in the first 50 amino acids in the N-terminus                                                                                                                                                      |
| alternative secretory pathway sorting signal                 | sum of hydrophobicity of 10 amino acids before the N-terminal cleavage                                                                                                                                                                                                                             |
| length of longest very hydrophobic region                    | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases                                                                                                                                                        |
| putative mitochondrial sorting signal                         | weighted sum of typical amino acids for mitochondrial proteins in N-terminus                                                                                                                                                                                                                     |
| secretory pathway sorting signal                              | autocorrelation of every second hydrophobic amino acid within the first 20 amino acids in the N-terminus                                                                                                                                                                                      |
| mitochondrial targeting peptide                               | maximal autocorrelation of every sixth charged amino acid within the first 30 amino acids in the N-terminus                                                                                                                                                                                        |
| number of asparagine in the N-terminus                       | number of asparagine within the first 70 amino acids in the N-terminus                                                                                                                                                                                                                               |
| number of tryptophane in the N-terminus                      | maximal normalized pseudo amino acid count of tryptophane within the first 120 amino acids in the N-terminus                                                                                                                                                                                        |
| number of lysine in the N-terminus                            | maximal pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus                                                                                                                                                                                                        |
| number of methionine in the N-terminus                       | number of methionine within the first 120 amino acids in the N-terminus                                                                                                                                                                                                                             |
| number of tryptophane in the N-terminus                      | number of tryptophane within the first 120 amino acids in the N-terminus                                                                                                                                                                                                                           |
| number of phenylalanine in the N-terminus                    | maximal normalized pseudo amino acid count of phenylalanine within the first 130 amino acids in the N-terminus                                                                                                                                                                                      |
| number of cysteine in the N-terminus                          | maximal pseudo amino acid count of cysteine within the first 130 amino acids in the N-terminus                                                                                                                                                                                                       |
| hydrophobic N-terminus                                        | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 130 amino acids in the N-terminus                                                                                                                                                                           |
| hydrophobic C-terminus                                        | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus                                                                                                                                                                                    |
| negatively charged N-terminus                                 | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus                                                                                                                                                                                     |
| number of alpha-helix preferred residues AFGHKLMNR            | minimal overall pseudo amino acid count of alpha-helix preferred residues                                                                                                                                                                                                                           |
| number of small amino acids in the N-terminus                | number of small amino acids within the first 20 amino acids in the N-terminus                                                                                                                                                                                                                      |
| number of aromatic residues in the C-terminus                | maximal pseudo amino acid count of aromatic residues within the last 100 residues in the C-terminus                                                                                                                                                                                               |
| number of potentially hydroxylated residues in the C-terminus | pseudo amino acid count of potentially hydroxylated residues in a distance of three within the last 100 amino acids in the C-terminus                                                                                                                                                                  |
| uncharged and hydrophobic protein                             | overall pseudo amino acid count of uncharged hydrophobic residues (I,L,V,M,F,Y,W,C,T,A,G) in a distance of two                                                                                                                                                                                  |
| PS50041 (C-type lectin domain)                                | Prosite pattern C-type lectin domain present in membrane and extracellular membrane and secreted proteins                                                                                                                      |
| typical extracellular prosite pattern                         | cluster of typical extracellular prosite patterns                                                                                                                                                                                                                                               |
| typical plasma membrane prosite p.                           | cluster of typical plasma membrane prosite patterns                                                                                                                                                                                                                                              |
| typical nuclear prosite pattern                               | cluster of typical nuclear prosite patterns                                                                                                                                                                                                                                                           |
| typical mitochondrial prosite pattern                         | cluster of typical mitochondrial prosite patterns                                                                                                                                                                                                                                                  |
| typical cytoplasmic prosite pattern                           | cluster of typical cytoplasmic prosite patterns                                                                                                                                                                                                                                                     |

Description of the features of YLoc+* Fungi. All features are explained in detail and are linked to a biological interpretation. In cases where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated feature.
Table 24. List of features of YLoc+* Plants

| Biological interpretation                        | Detailed feature description                                                                                                                                 |
|-------------------------------------------------|-----------------------------------------------------------------------------------------------------|
| protein size                                     | number of amino acids                                                                             |
| ER retention signal                              | number of weak ER retention signals (number of K,D,E,L in the very C-terminus) times a factor that depends on the fact that no transmembrane helix and no glycolization signal is present. If a strong ER retention signal (KDEL, KRHQSADENQEL, DEL, EL) at C-terminus is present it is 100. |
| peroxisomal targeting signal (PTS)               | weighted PTS sum = SKL*0.83+SKF*0.5+[SAGCN][RKH][LIVMAF]*0.25 (presented by Nakai et al.)             |
| alternative secretory pathway sorting signal     | sum of hydrophobicity of 10 amino acids before N-terminal cleavage                                 |
| length of longest very hydrophobic region        | maximal length of very hydrophobic region where the iterative sum of hydrophobicity normed to mean zero drops below zero in at most two cases |
| putative mitochondrial sorting signal            | weighted sum of typical amino acids for mitochondrial proteins in N-terminus                       |
| typical chloroplast protein N-terminus [MA]      | chloroplast motif [MA] at the beginning of the protein                                            |
| secretory pathway sorting signal                 | autocorrelation of every hydrophobic amino acid within the first 20 amino acids in the N-terminus |
| mitochondrial targeting peptide                  | sum of charge of first 10 amino acids in N-terminus                                                |
| putative mitochondrial or chloroplast targeting peptide | autocorrelation of every sixth charged amino acid within the first 50 amino acids in the N-terminus |
| number of Serine in the N-terminus               | number of Serine within the first 50 amino acids in the N-terminus                                 |
| number of leucine in the N-terminus              | maximal normalized pseudo amino acid count of leucine within the first 80 amino acids in the N-terminus |
| number of cysteine in the N-terminus             | maximal pseudo amino acid count of cysteine within the first 120 amino acids in the N-terminus     |
| number of lysine in the N-terminus               | maximal pseudo amino acid count of lysine within the first 120 amino acids in the N-terminus      |
| number of phenylalanine in the N-terminus        | number of phenylalanine within the first 120 amino acids in the N-terminus                          |
| number of tryptophane in the N-terminus          | number of tryptophane within the first 120 amino acids in the N-terminus                           |
| alternative secretory pathway sorting signal     | pseudo amino acid count of very hydrophobic residues in a distance of two within the first 100 amino acids in the N-terminus |
| hydrophobic C-terminus                           | maximal normalized pseudo amino acid count of very hydrophobic residues in the last 40 amino acids in the C-terminus |
| negatively charged N-terminus                    | maximal pseudo amino acid count of negatively charged residues within the first 30 amino acids in the N-terminus |
| number of alpha-helix preferred residues in the N-terminus | pseudo amino acid count of alpha-helix preferred residues [AFGHKLMNR] in a distance of three in the first 20 amino acids |
| number of alpha-helix preferred residues         | minimal overall pseudo amino acid count of alpha-helix preferred residues                          |
| number of large amino acids in the N-terminus    | number of large amino acids within the first 20 amino acids in the N-terminus                      |
| number of potentially hydroxylated residues      | maximal overall pseudo amino acid count of potentially hydroxylated residues                     |
| uncharged and hydrophobic protein                | overall pseudo amino acid count of uncharged hydrophobic residues (L,L,V,M,F,Y,W,C,T,A,G) in a distance of two |
| typical extracellular prosite pattern            | cluster of typical extracellular prosite patterns                                                  |
| typical plasma membrane prosite p.               | cluster of typical plasma membrane prosite patterns                                               |
| typical nuclear prosite pattern                  | cluster of typical nuclear prosite patterns                                                       |
| typical mitochondrial prosite pattern            | cluster of typical mitochondrial prosite patterns                                                 |
| typical chloroplast prosite pattern              | cluster of typical chloroplast prosite patterns                                                   |
| typical cytoplasmic prosite pattern              | cluster of typical cytoplasmic prosite patterns                                                   |

Description of the features of YLoc+* Plants All features are explained in detail and are linked to a biological interpretation. In cases, where a biological interpretation is not straightforward, the explanation was transferred from a highly correlated features.