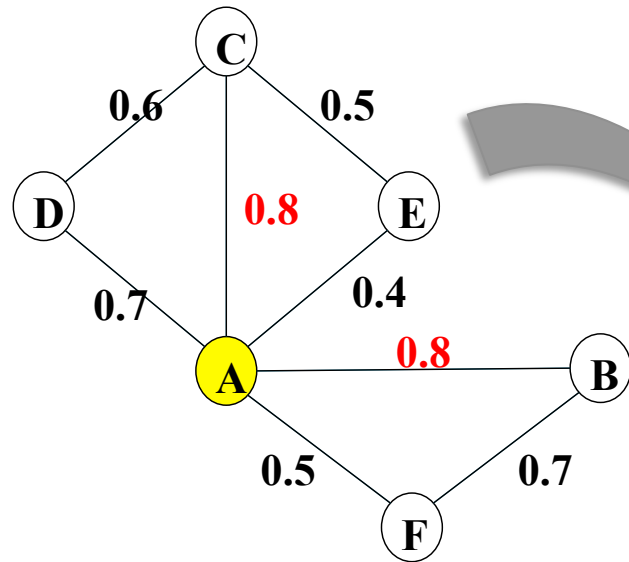


Select source organisms

Organism Name	Number of Contigs
<i>Coccomyxa subellipsoidea</i> C-169	16644
<i>Chlorella variabilis</i>	13342
<i>Volvox carteri</i>	12252
<i>Chlamydomonas reinhardtii</i>	11459
<i>Glycine max</i>	11450
<i>Populus trichocarpa</i>	11289
<i>Setaria italica</i>	11257
<i>Physcomitrella patens</i> subsp. <i>patens</i>	11253
<i>Vitis vinifera</i>	11226
<i>Brachypodium distachyon</i>	11208
<i>Solanum lycopersicum</i>	11056
<i>Arabidopsis thaliana</i>	11045
<i>Brassica rapa</i> subsp. <i>pekinensis</i>	11030
<i>Selaginella moellendorffii</i>	11017
<i>Cicer arietinum</i>	10982
<i>Fragaria vesca</i> subsp. <i>vesca</i>	10966
<i>Cucumis sativus</i>	10964
<i>Prunus persica</i>	10909
<i>Ricinus communis</i>	10800

Random Walking with Restart !

$$p^{t+1} = (1 - r)W_{p^t} + rp^0$$



$$p_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad W_{p^0} = \begin{array}{c|cccccc} & \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} \\ \hline \text{A} & 0 & 0.8 & 0.8 & 0.7 & 0 & 0.5 \\ \text{B} & 0.8 & 0 & 0 & 0 & 0 & 0.7 \\ \text{C} & 0.8 & 0 & 0 & 0.6 & 0.5 & 0 \\ \text{D} & 0.7 & 0 & 0.6 & 0 & 0 & 0 \\ \text{E} & 0.4 & 0 & 0.5 & 0 & 0 & 0 \\ \text{F} & 0.5 & 0.7 & 0 & 0 & 0 & 0 \end{array}$$

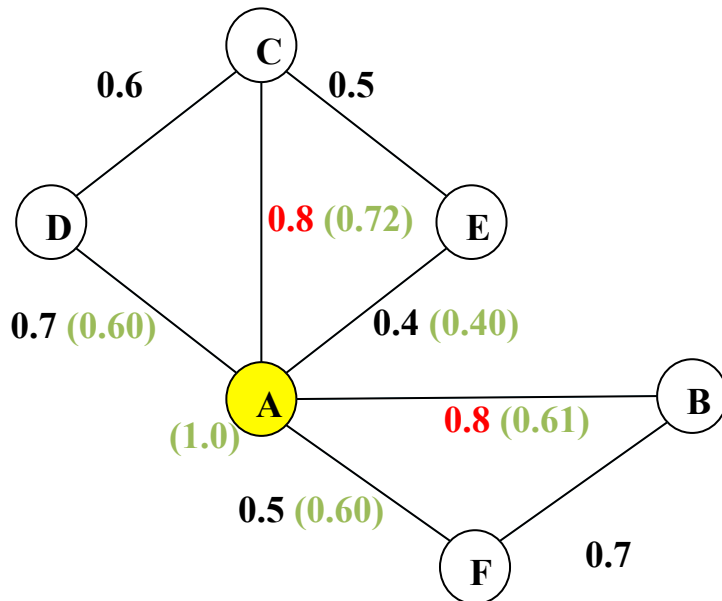
Run RWR:

tolerance = 0.06

Max Iterations = 1000

$r = 0.33$

Normalize by maximum score



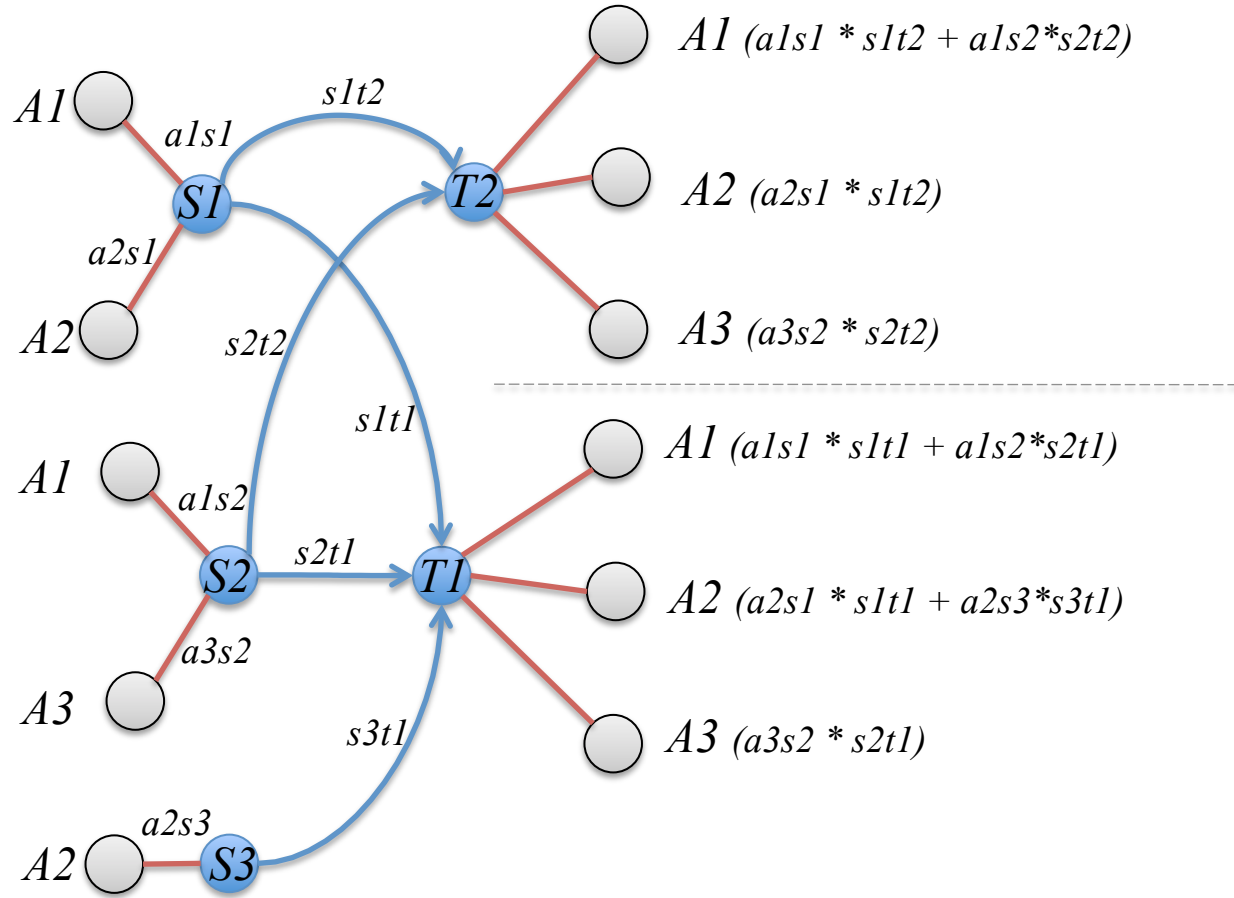
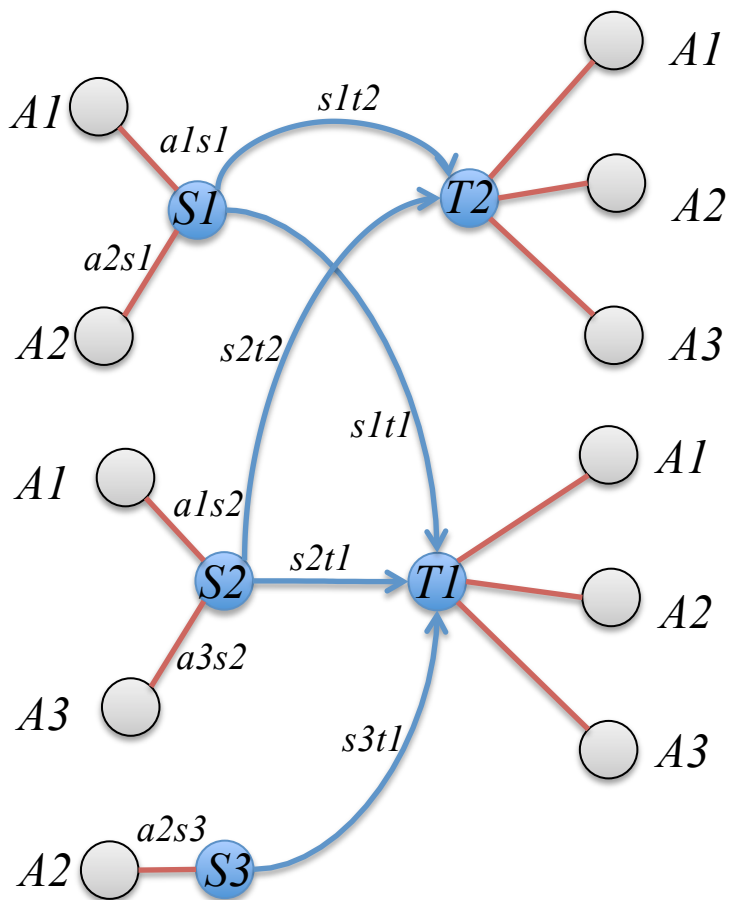


Figure X: Annotation flow network: A1-A3 are annotations of source sequences S1, S2 and S3. $axsy$ represents the user-defined confidence of the annotation ax to be associated with sequence sy . $sxty$ is the maximum score normalized orthofuzz score(sx, ty) obtained by querying the pairwise sequence similarity network using sx .



*This translates as
matrix multiplication of
Annotation Confidence Matrix
and Orthofuzz matrix*

$$\begin{array}{c}
 \begin{matrix} SS1 & SS2 & SS3 \\ A1 & A2 & A3 \end{matrix}
 \begin{pmatrix} a1s1 & a1s2 & 0 \\ a2s1 & 0 & a2s3 \\ 0 & a3s2 & 0 \end{pmatrix}
 \times
 \begin{matrix} TS1 & TS2 \\ SS1 & SS2 \\ SS3 \end{matrix}
 \begin{pmatrix} s1t1 & s1t2 \\ s2t1 & s2t2 \\ s3t1 & 0 \end{pmatrix}
 =
 \begin{matrix} TS1 & TS2 \\ A1 & A2 & A3 \end{matrix}
 \begin{pmatrix} a1s1 \times s1t1 + a1s2 \times s2t1 & a1s1 \times s1t2 + a1s2 \times s2t2 \\ a2s1 \times s1t1 + a2s3 \times s3t1 & a2s1 \times s1t2 \\ a3s2 \times s2t1 & a3s2 \times s2t2 \end{pmatrix}
 \end{array}$$

$$\begin{matrix} & SS1 & SS2 & SS3 \\ A1 & (a1s1 & a1s2 & 0 \\ A2 & (a2s1 & 0 & a2s3 \\ A3 & (0 & a3s2 & 0 \end{matrix} \times \begin{matrix} TS1 & TS2 \\ SS1 & (s1t1 & s1t2 \\ SS2 & (s2t1 & s2t2 \\ SS3 & (s3t1 & 0 \end{matrix} = \begin{matrix} TS1 & TS2 \\ A1 & (a1s1 \times s1t1 + a1s2 \times s2t1 & a1s1 \times s1t2 + a1s2 \times s2t2 \\ A2 & (a2s1 \times s1t1 + a2s3 \times s3t1 & a2s1 \times s1t2 \\ A3 & (a3s2 \times s2t1 & a3s2 \times s2t2 \end{matrix}$$

$$\mathbf{ATS}_{N_A \times N_{TS}} = \mathbf{ASS}_{N_A \times N_{SS}} \times \mathbf{SSTS}_{N_{SS} \times N_{TS}}$$

Symbol	Description
SS	<i>Set of Source Sequences</i>
N_{SS}	<i>Total Number of Source Sequences</i>
A	<i>Set of Source Annotations</i>
N_A	<i>Total Number of Annotations from Source Sequences</i>
TS	<i>Set of Target Sequences</i>
N_{TS}	<i>Total Number of Target Sequences</i>
ATS	<i>Target Annotation Weight Matrix</i>
ASS	<i>Source Annotation Confidence Matrix</i>
$SSTS$	<i>Source Target Orthofuzz Matrix</i>
$axsy$	<i>Confidence of assigning ax to sy</i>
$sytz$	<i>Max. normalized orthofuzz score of sy to tz</i>
$axtz$	<i>Annotation weight of ax to tz</i>

Cluster by Random Walking

Data:

Shared among threads:

Transferred to threads

*Random Walking with
Restart*

*Create population of
Scores*

Estimate Statistics

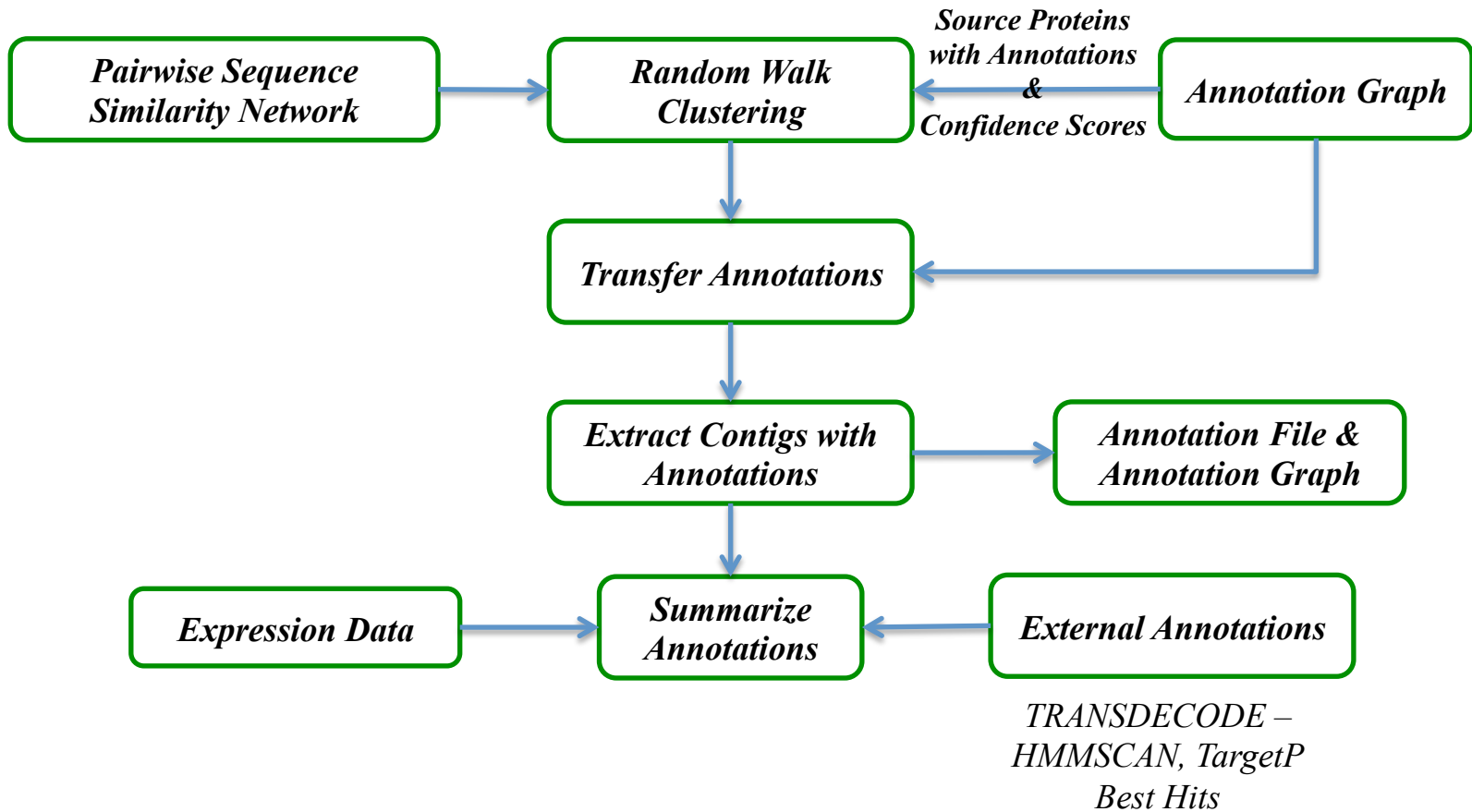


Figure 1: Protocol used for assigning functional annotations to the de-novo assembled contigs

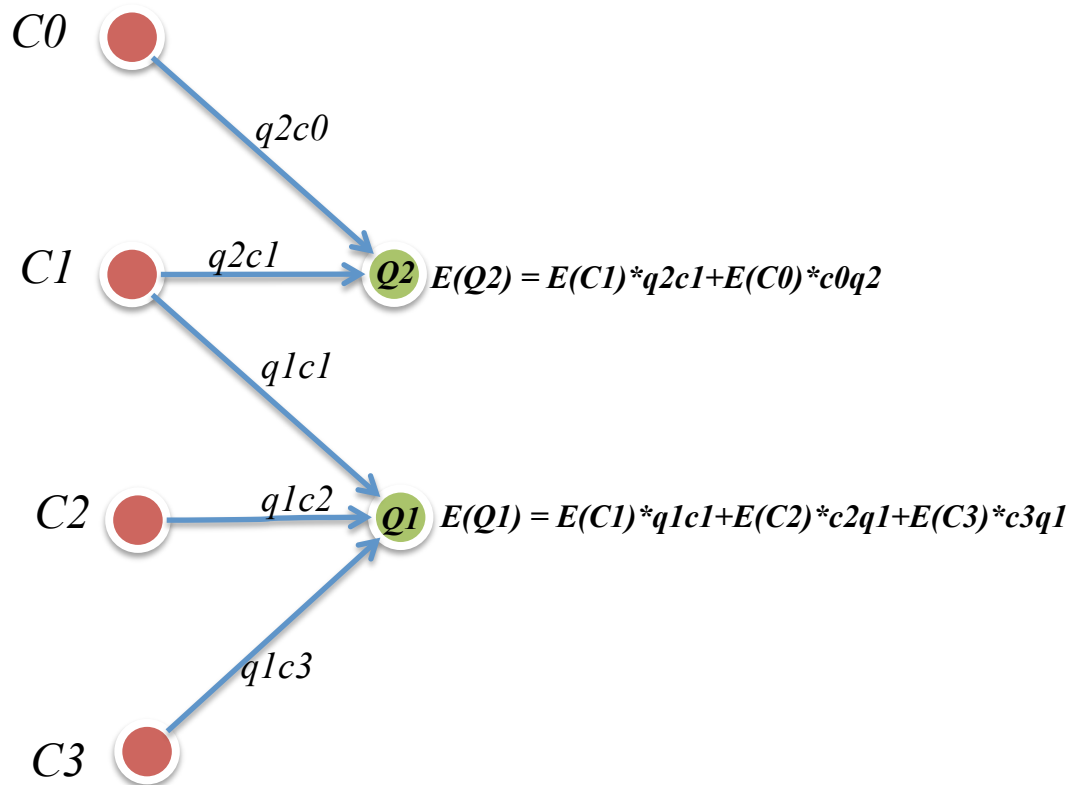
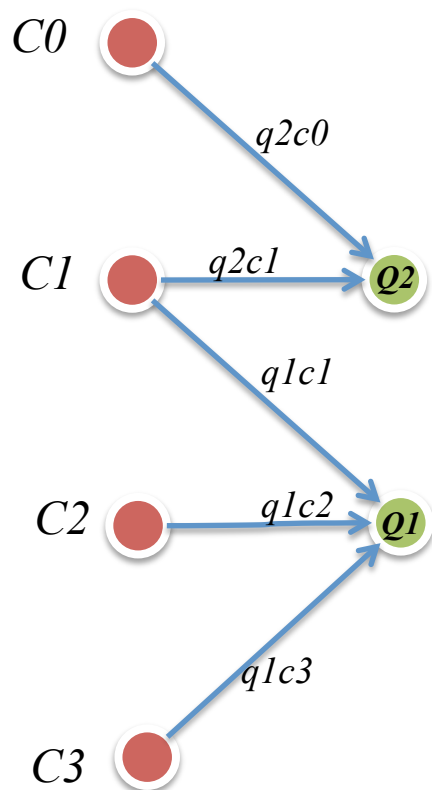


Figure X: Expression summary network: $C0$ - $C3$ are the contigs identified as homologs of Query Sets $Q1$ and $Q2$. $E(X)$ is the normalized expression level of X . q_{xyc} is the within species normalized orthofuzzscore ($q_{x,c}$) obtained by querying the network using the query set Q .



*This translates as
matrix multiplication of
OrthoFuzz Matrix and
Contig Expression Vector*

$$\begin{matrix} & C0 & C1 & C2 & C3 \\ Q1 & \begin{pmatrix} 0 & q1c1 & q1c2 & q1c3 \end{pmatrix} \\ Q2 & \begin{pmatrix} q2c0 & q2c1 & 0 & 0 \end{pmatrix} \end{matrix} \times \begin{matrix} E \\ C0 \begin{pmatrix} e_{c0} \\ e_{c1} \\ e_{c2} \\ e_{c3} \end{pmatrix} \end{matrix} = \begin{matrix} E \\ Q1 \begin{pmatrix} e_{c1} \times q1c1 + e_{c2} \times q1c2 + e_{c3} \times q1c3 \\ e_{c0} \times q2c0 + e_{c1} \times q2c1 \end{pmatrix} \end{matrix}$$

$$\begin{matrix} & C0 & C1 & C2 & C3 \\ Q1 \left(\begin{matrix} 0 & q1c1 & q1c2 & q1c3 \end{matrix} \right) \\ Q2 \left(\begin{matrix} q2c0 & q2c1 & 0 & 0 \end{matrix} \right) \end{matrix} \times \begin{matrix} E \\ C0 \left(\begin{matrix} e_{c0} \\ e_{c1} \\ e_{c2} \\ e_{c3} \end{matrix} \right) \end{matrix} = \begin{matrix} E \\ Q1 \left(\begin{matrix} e_{c1} \times q1c1 + e_{c2} \times q1c2 + e_{c3} \times q1c3 \\ e_{c0} \times q2c0 + e_{c1} \times q2c1 \end{matrix} \right) \end{matrix}$$

$$\mathbf{QC}_{N_Q \times N_C} \times \mathbf{Ce}_{N_C \times 1} = \mathbf{Qe}_{N_Q \times 1}$$

Symbol	Description
Q	<i>Set of Query Sets</i>
N_Q	<i>Total Number of Query Sets</i>
N_C	<i>Total Number of Expressed Contigs</i>
QC	<i>Query Contig Orthofuzz Matrix</i>
$qxcy$	<i>Species.normalized orthofuzz score of qx to cy</i>
Ce	<i>Vector containing expression values of Contigs</i>
e_{cy}	<i>Expression value of contig y</i>
Qe	<i>Vector containing expression values of Query</i>

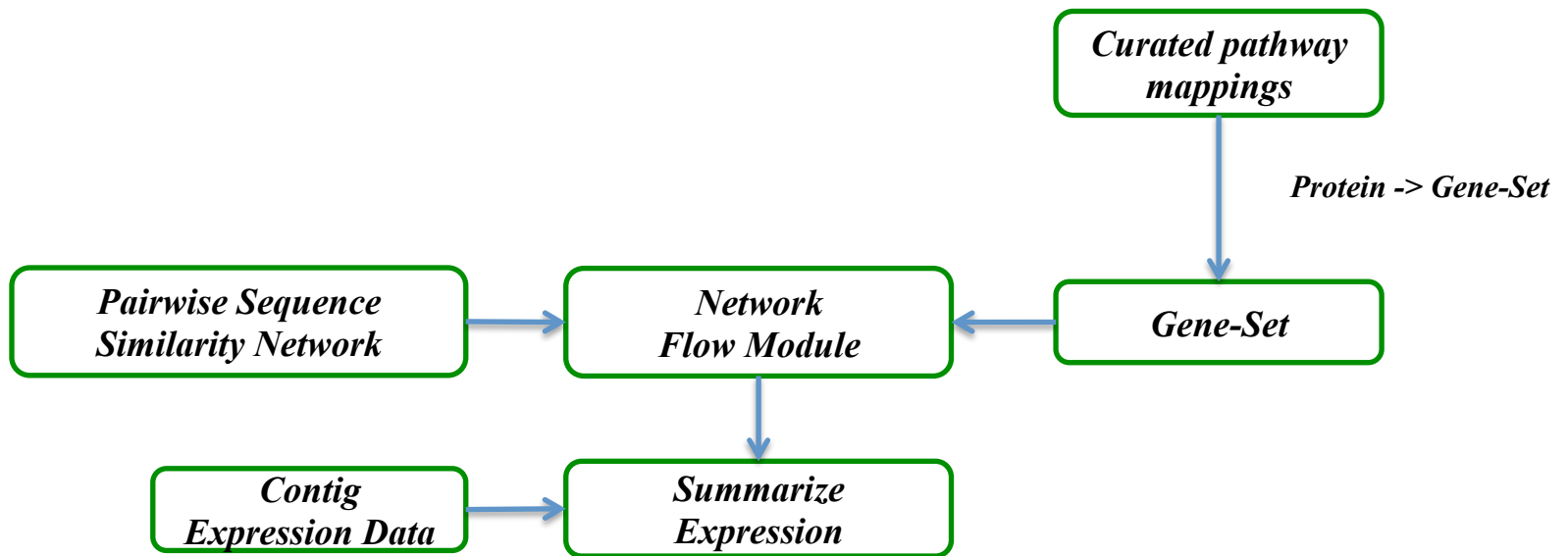
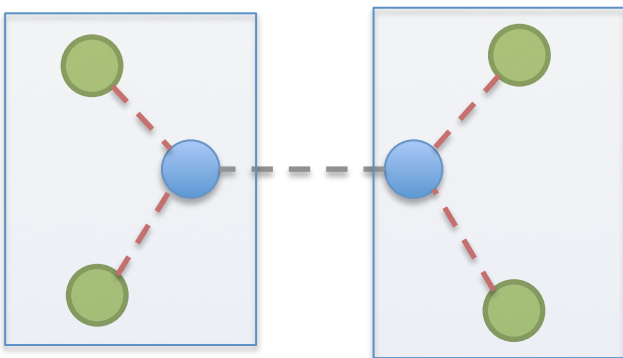
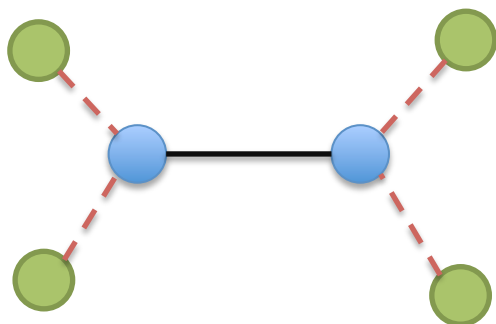
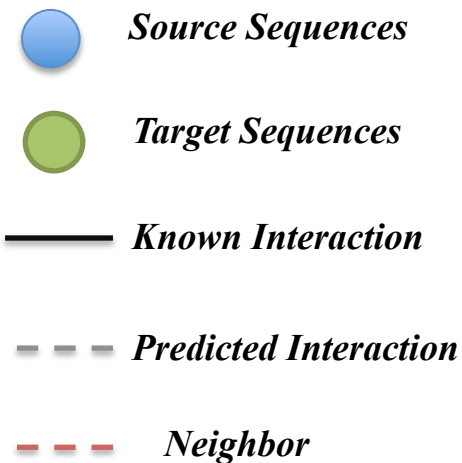


Figure: Protocol used for estimating expression levels of gene-sets



Is there evidence for protein complexes ?



Expression Levels

Interaction Unit 1

Interaction Unit 2

<i>Conditions</i>		

Are there interaction units whose expression levels are changing ?

Are they co-expressed ?



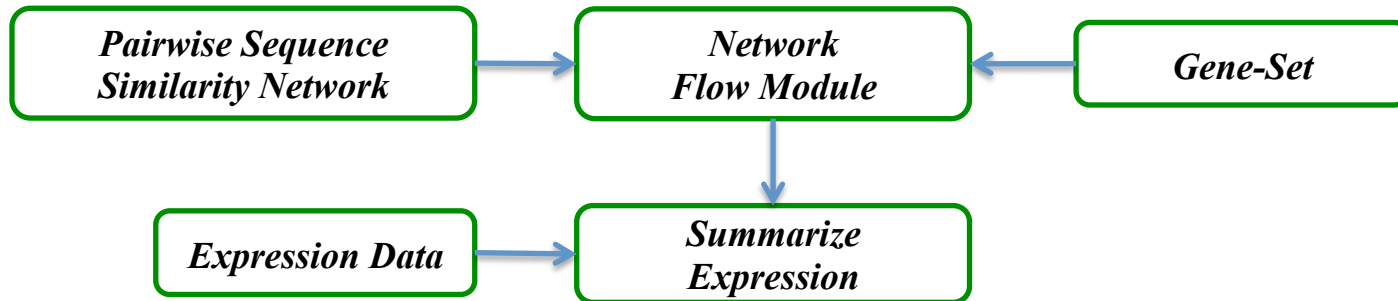


Figure: Protocol used for estimating expression levels of gene-sets

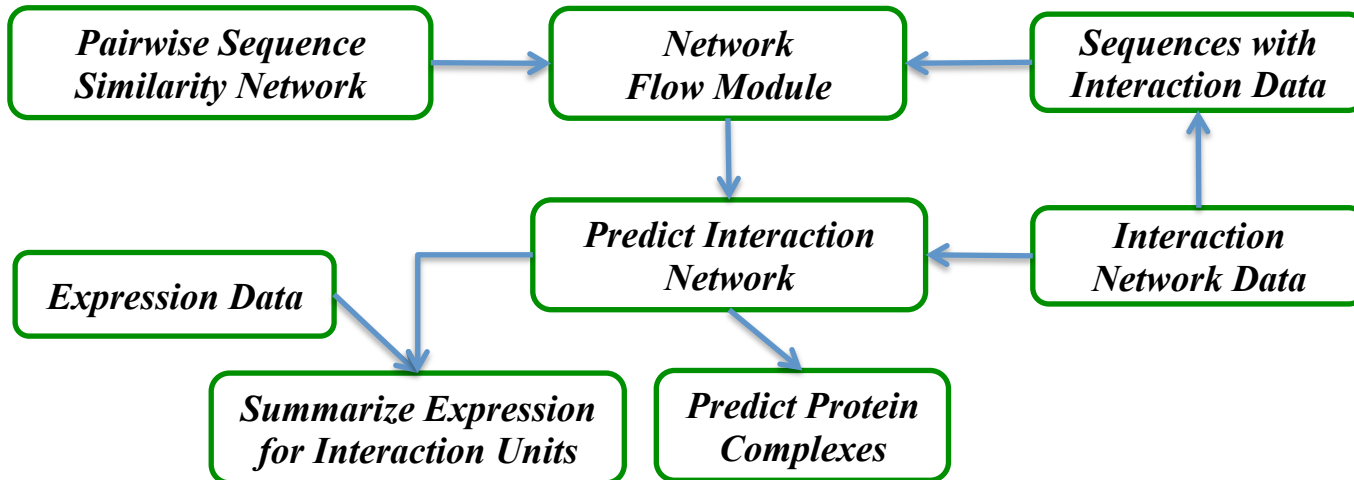
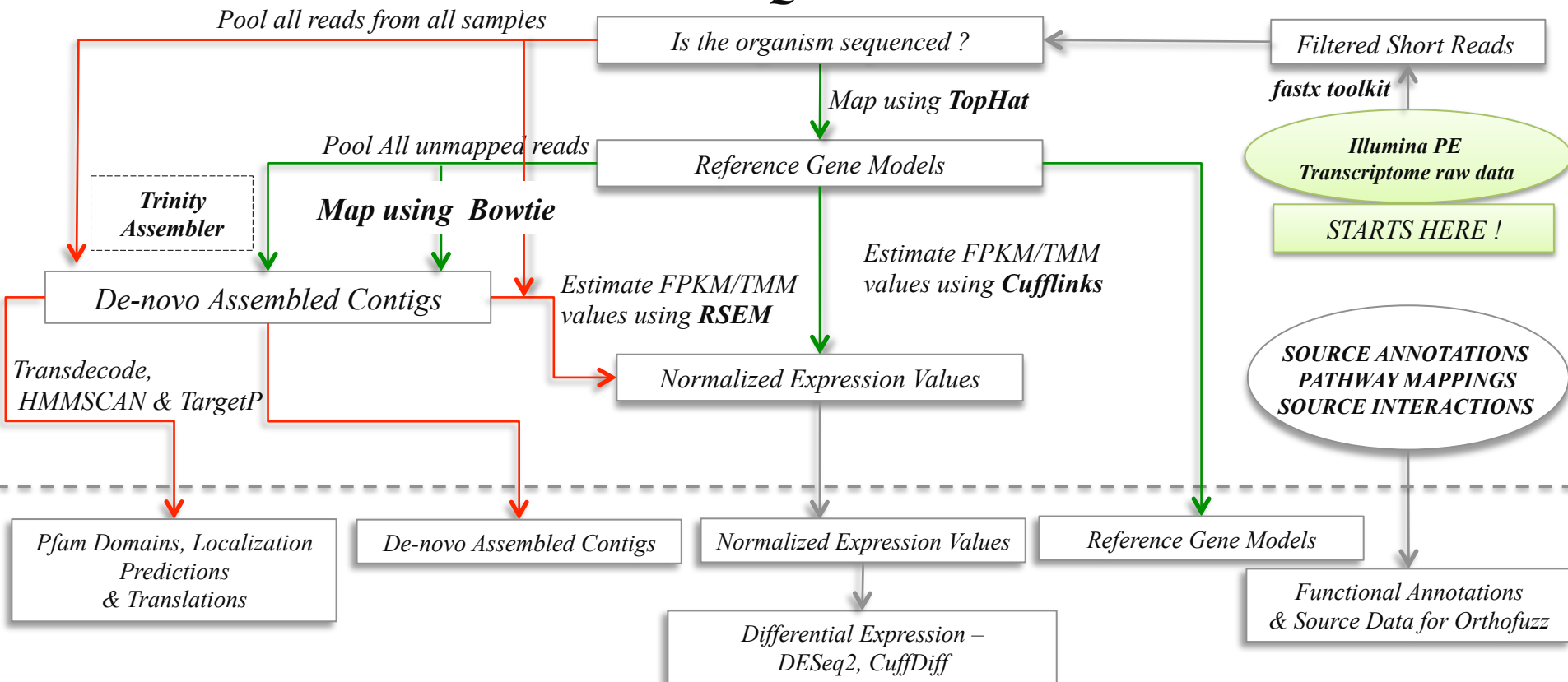


Figure: Protocol used for estimating expression levels of gene-sets

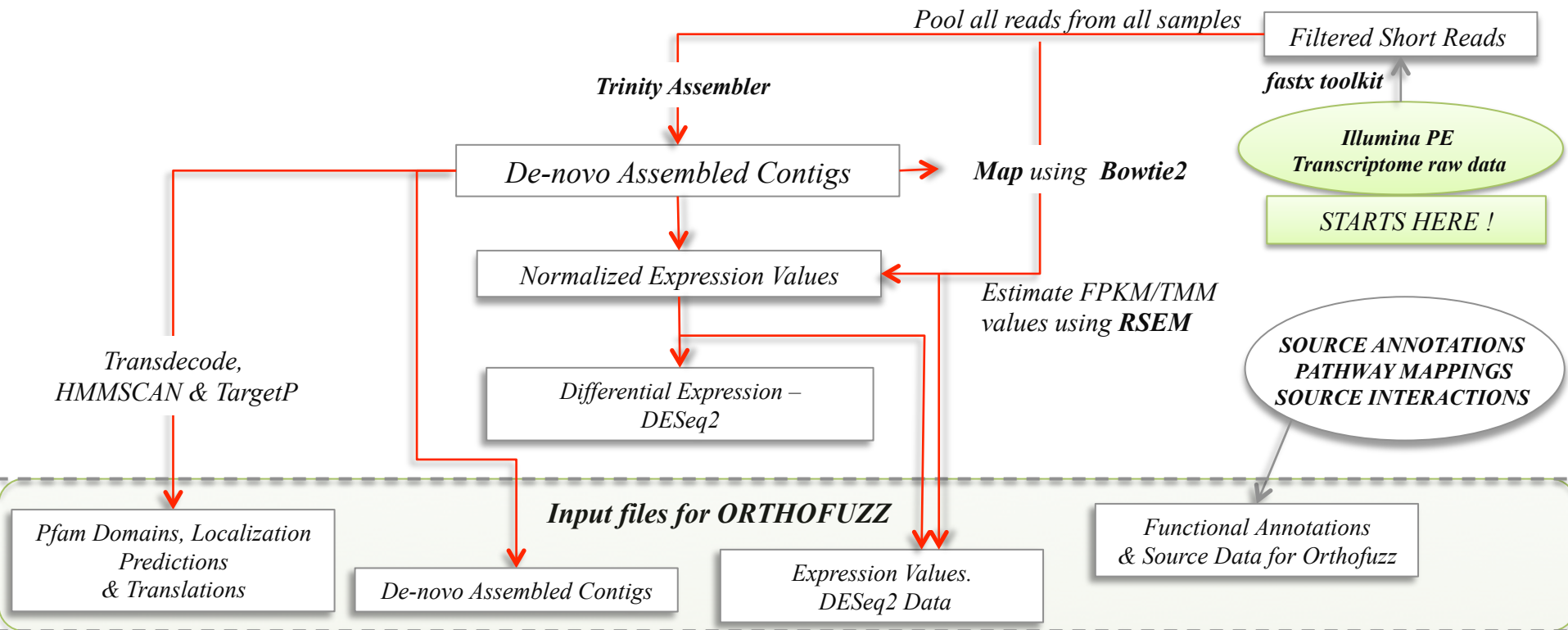
RNA-SEQ PIPELINE



ORTHOFUZZ package

Downstream analysis step	Orthofuzz Program
Functional Annotation	Orthofuzz–TransferAnnotation tool
List Enrichment Statistics	Orthofuzz–GeneListEnrichment tool
Summarize Gene-Sets Expression	Orthofuzz–TransferExpression tool
Gene-Sets Enrichment Statistics	Orthofuzz–GeneSetEnrichment tool
Predict Interactions	Orthofuzz–PredictInteractions tool
Co-expression	Orthofuzz–ParallelCorrelation tool

De-novo RNA-SEQ PIPELINE



ORTHOFUZZ package

Downstream analysis step	Orthofuzz Program
Functional Annotation	Orthofuzz-TransferAnnotation tool
List Enrichment Statistics	Orthofuzz-GeneListEnrichment tool
Summarize Gene-Sets Expression	Orthofuzz-TransferExpression tool
Gene-Sets Enrichment Statistics	Orthofuzz-GeneSetEnrichment tool
Predict Interactions	Orthofuzz-PredictInteractions tool
Co-expression	Orthofuzz-ParallelCorrelation tool

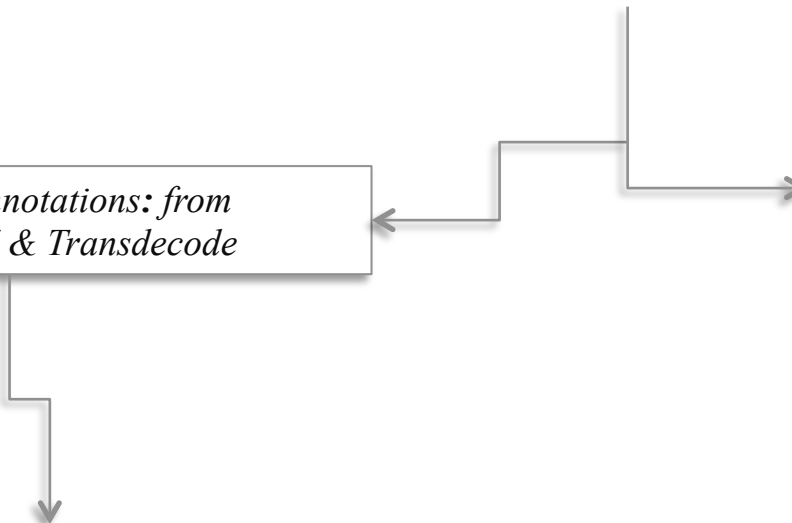
<i>Downstream analysis step</i>	<i>Software Used</i>
<i>Differential Expression</i>	<i>DESeq2, CuffDiff</i>
<i>Functional Annotation</i>	<i>Orthofuzz – TransferAnnotation tool</i>
<i>List Enrichment Statistics</i>	<i>TopGO for Gene Ontology Enrichment Statistics</i> <i>Orthofuzz – GeneListEnrichment tool</i>
<i>Gene Set Enrichment Statistics</i>	<i>GSA method for gene-set statistics</i> <i>Orthofuzz GeneSetEnrichment tool</i>
<i>Co-expression & Clustering</i>	<i>Orthofuzz – ParallelCorrelation tool.</i> <i>Cytoscape – Visualize co-expression network</i> <i>GenePattern for heatmap</i> <i>WGCNA package for clustering</i>
<i>Pathway Visualization</i>	<i>Pathvisio for drawing and mapping expression data on to pathways</i>

<i>Downstream analysis step</i>	<i>Software Used</i>
<i>Summarize Expressions of Gene Sets</i>	<i>Orthofuzz – TransferExpression tool</i>

*Functional Annotations: from
ORTHOFUZZ & Transdecode*

*Interaction Data:
Predicted Co-expressed Interaction
Complexes with Enrichment Statistics*

*Pathway Data:
Pathway Enrichment Statistics
Expression of Enzymes in Pathways*



<i>Downstream analysis step</i>	<i>Software Used</i>
<i>Differential Expression</i>	<i>DESeq2, CuffDiff</i>
<i>Functional Annotation</i>	<i>Orthofuzz–TransferAnnotation tool</i>
<i>List Enrichment Statistics</i>	<i>TopGO for Gene Ontology Enrichment Statistics Orthofuzz–GeneListEnrichment tool</i>
<i>Gene-Sets Summarize Expression</i>	<i>Orthofuzz–TransferExpression tool</i>
<i>Gene-Sets Enrichment Statistics</i>	<i>GSA method for gene-set statistics Orthofuzz-GeneSetEnrichment tool</i>
<i>Predict Interactions</i>	<i>Orthofuzz-PredictInteractions</i>

<i>Downstream analysis step</i>	<i>Software Used</i>
<i>Co-expression & Clustering</i>	<i>Orthofuzz–ParallelCorrelation tool. Cytoscape – Visualize co-expression network GenePattern for heatmap WGCNA package for clustering</i>
<i>Pathway Visualization</i>	<i>Pathvisio for drawing and mapping expression data on to pathways</i>

