Workshop on genome-scale metabolic models reconstruction with *merlin*

Files:

Workspace - https://nextcloud.bio.di.uminho.pt/workspace
merlin - https://merlin-sysbio.org/download/
PSort report - https://nextcloud.bio.di.uminho.pt/psort

hands-on:

1st Step – new workspace

1. create new workspace
   a. https://merlin-sysbio.org/documentation/#create_ws
   b. set taxonomy identifier
      i. select the strain identifier
         1. example: *Escherichia coli* – 83333
      ii. visit this site to obtain identifier
         1. https://www.ncbi.nlm.nih.gov/taxonomy
   c. select assembly record to set genome in merlin (preferably from RefSeq)
      i. information is downloaded into merlin
2. Import an existing workspace
   a. https://merlin-sysbio.org/documentation/#import_files
   b. Import the workspace (“workspace -> import -> workspace)
   c. Select the file you downloaded
   d. In advanced options you can change the name of the workspace
   e. When the workspace is created, open it (a window to open the workspace will appear)

2nd Step – load metabolic information

1. KEGG
   a. https://merlin-sysbio.org/documentation/#met_data
   b. *merlin* will download all metabolic information available in KEGG into the personal computer’s database

3rd Step – enzymes annotation

1. Perform homology search with BLAST or Diamond
   a. https://merlin-sysbio.org/documentation/#homology
   b. Set BLAST/Diamond parameters
      i. Go to “Annotation -> “Enzymes” -> BLAST -> New Blast Submission” or “Annotation -> “Enzymes” -> Diamond -> New Diamond Submission”
      ii. Select the database (example: uniprotkb_swissprot)

        iii. Change other parameters (optional)
    c. If you do not want to wait for the alignment results, you can proceed with the workspace imported in the first step
2. Perform annotation
    a. https://merlin-sysbio.org/documentation/#pipelines_gfa
    b. automatic workflow
        i. select a set of phylogenetically close organisms (preferably frequently represented in the BLAST results)
        ii. insert the previously listed organisms by order of phylogenetic proximity
        iii. all genes will be annotated
    c. SamPler (alternative to automatic workflow)
        i. select a random sample of genes with a size of your choice (as an example choose 1 or 2%)
        ii. save the sample and export the file
        iii. go to the annotation menu in the sidebar and click on "enzymes"
        iv. choose the correct EC number for each gene in the sample based on BLAST results (can be accessed by clicking on the magnifier button)
        v. go back to the SamPler menu and find the best parameters (press the "best parameters function")
        vi. apply the ones you think that fits best
        vii. go to the annotation menu in the sidebar and click on "enzymes"
        viii. annotate the ones that have scores lower than the upper threshold and greater than the lower threshold displayed in the menu in the bottom
3. Integrate annotation with metabolic data
    a. https://merlin-sysbio.org/documentation/#enzymes_integration
    b. *merlin* will add reactions encoded by enzymes identified in the genome annotation step to the model

4th Step – transporters annotation

1. Annotate genome in TranSyT's server
    a. https://merlin-sysbio.org/documentation/#transyt
    b. TranSyT will create reactions associated to genes, for metabolites available in the model
2. Add known simple diffusion transport reactions (e.g. water, oxygen)

5th Step – compartments annotation

1. Perform the compartments annotation
    a. https://merlin-sysbio.org/documentation/#compartments
    b. use the PSort report provided in *Files* (on the top of the document)
    c. import results in merlin
2. Integrate to model
    a. https://merlin-sysbio.org/documentation/#compartments_integration
    b. Choose compartments to ignore (example: "Unknown")

6th Step – model curation

1. Run the "correct reversibility" tool
    a. https://merlin-sysbio.org/documentation/#reversibility
    b. Select the Zeng or ModelSeed source, and an adequate template for your organism
2. Analyze pathways
    a. Draw the pathway in Browser
        i. https://merlin-sysbio.org/documentation/#reactions_board
            1. Green enzymes: the enzyme was identified in the genome annotation
            2. Dark Blue enzymes: the reaction is present in the model, but it is being promoted by an enzyme other than the one available for this pathway
    b. Identify missing reactions
        i. Check the genome annotation: search in the enzymes annotation view for the EC number (in the search bar, change "name" to "all" before typing the EC number). Search also for incomplete EC numbers.
        ii. Search in biological databases (examples: KEGG, BRENDA, MetaCyc, BiGG Models)

3. Identify blocked reactions
    a. https://merlin-sysbio.org/documentation/#find_blocked
    b. Blocked reactions will be highlighted in red
    c. Check the dead-end metabolites in the reaction
    d. Identify candidate reactions to consume/produce these metabolites
        i. Check the reversibility of reactions associated to each dead-end metabolite (the reversibility of reactions in MetaCyc is usually manually curated)
        ii. Search in KEGG for reactions that can consume/produce the metabolite
        iii. Search in other biological databases for reactions not available in KEGG, spontaneous/non enzymatic reactions, and transport reactions
4. Balance validation
    a. https://merlin-sysbio.org/documentation/#find_unbalanced
    b. Unbalanced reactions will be highlighted in bold
    c. Metabolites without formula
        i. Search the molecular formula in biological databases (BiGG Models, BRENDA, MetaCyC)
    d. Missing protons or water
        i. Unbalanced protons are usually caused by the protonation state of the metabolites in the reaction.
    e. Polymerization/depolymerization reactions
        i. Remove the polymer from the reactants (polymerization reactions) or from the products (depolymerization reactions)
        ii. Adjust the stoichiometry of the monomers appropriately

7th step – Create the biomass formulation

1. Creating the biomass equation
    a. https://merlin-sysbio.org/documentation/#e_biomass
    b. Select an appropriate template
    c. e-Protein, e-DNA and e-RNA are determined automatically from the genome information
    d. The others must be adjusted manually

8ß step – model curation with BioISO

1. Assess biomass formulation
    a. https://merlin-sysbio.org/documentation/#bioiso
2. Find gaps and errors in the model
3. BioISO will highlight all biomass precursors having zero flux, thereby speeding up the curation process

9th step – network visualisation

1. Visualise small networks with Escher maps:
    a. https://merlin-sysbio.org/documentation/#escher_maps_draw
    b. https://merlin-sysbio.org/documentation/#escher_maps_templates
    c. https://merlin-sysbio.org/documentation/#escher_maps_draw_path
    d. Validation -> network visualization
        i. Choose *Escher;*
        ii. Press the "+" button to choose the pathways to visualise;
        iii. Select "all reactions";
        iv. Press "ok";
        v. Wait;
        vi. A new separator will be opened in your default browser;
        vii. Then you can see that the dots are metabolites whereas the arrows correspond to the reactions;
        viii. The red arrows are reactions present in the KEGG database but not in our model, whereas the blue arrows are the reactions in our model.
2. Visualise larger networks with MetExploreViz:
    a. https://merlin-sysbio.org/documentation/#metexploreviz
    b. Validation -> network visualization
        i. Choose *MetExplore*
        ii. Press the "+" button to choose the pathways to visualise (one with reactions);
        iii. Select "all reactions";
        iv. Press "ok";
        v. Wait;
        vi. A new separator will be opened in your default browser;
        vii. Wait a few seconds, even if the page is not rendering anything;

viii. Then you can see that the dots are metabolites whereas the arrows correspond to the reactions;tm

9th step – run MEMOTE

1. Test the quality of your model with MEMOTE:
   a. https://merlin-sysbio.org/documentation/#memote_overview
   b. *Validation -> memote*
   c. Select the workspace;
   d. Press “ok” and wait for the results;
   e. A comprehensive table will docker below the dashboards’ item *validation*;
   f. Navigate through the table’s magnifier button to check the state of you model;

10th step – export to SBML

1. Export model to in the standard format
   a. https://merlin-sysbio.org/documentation/#export_model