



Background and Rationale

Since computational modelling has been effective in metabolic engineering for microorganisms, it has been recently extended to the realm of plant metabolism engineering. However, the higher complexity of the metabolic pathways in plants have complicated the task of creating metabolic models of plants which are accurate in its simulation of metabolism.

Yet, the rewards from successfully creating accurate plant metabolic models are huge: we will be able use these models to identify the key metabolic pathways in plants that produce desired metabolites/biomasses for industry usage, and even analyze how some plant species are able to survive in certain climates more successfully than others. In this project, we modelled the plant species *Neuontobotrys tarapacana* (NTARA) in order to understand how the metabolism of such desert plants allows them to effectively grow in dry conditions.

Methods

1) Creation of basic metabolic model from source file containing Gene-Reaction Pathways.

ID	Name	Pathways	Reactions
All-Genes All-Genes			
NTARA_0	Ntara_0		RXN-13684, AMINOACYLASE-RXN
NTARA_1	Ntara_1		RXN-13684, AMINOACYLASE-RXN
NTARA_1000	Ntara_1000		
NTARA_1001	Ntara_1001		
NTARA_1002	Ntara_1002		
NTARA_10023	Ntara_10023		RXN-9935
NTARA_10024	Ntara_10024	FWY-6442	RXN-11259, RXN-11258,
NTARA_10025	Ntara_10025	FWY-6442	RXN-11259, RXN-11258,
NTARA_10026	Ntara_10026	FWY-6442	RXN-11259, RXN-11258,
NTARA_10027	Ntara_10027	FWY-6442	RXN-11259, RXN-11258,
NTARA_1003	Ntara_1003		
NTARA_1004	Ntara_1004		
NTARA_10048	Ntara_10048		2.4.99.3-RXN
NTARA_10069	Ntara_10069		PYRIDOXINE-4-DEHYDROGENASE-RXN
NTARA_10070	Ntara_10070		PYRIDOXINE-4-DEHYDROGENASE-RXN
NTARA_10071	Ntara_10071		PYRIDOXINE-4-DEHYDROGENASE-RXN
NTARA_10072	Ntara_10072		PYRIDOXINE-4-DEHYDROGENASE-RXN
NTARA_10080	Ntara_10080		ALDEHYDE-OXIDASE-RXN, RXN-8090

Source File Formatting

	Abbreviation	Description	Reaction	GPR	Genes	Proteins	Subsystem
2	+	NEOMENTHOL		CPD-1905, NTARA_23799 or NTARA_42192	NTARA_42192	AT3G6122	PWY-3061
3	-	LIMONENE		CPD-4886, NTARA_16201	NTARA_16201	MONOME	PWY-5928
4	-	MENTHOL		--MENTHOL, NTARA_23799 or NTARA_42192	NTARA_42192	AT3G6122	PWY-3061
5	1-	ACYLGLYCEROL		ACYL-ACP, NTARA_3908	NTARA_3908	1-ACYLGLY	PWY0-1319
6	1-	PHOSPHATIC		ATP_c + L- NTARA_29247 or NTARA_32545	NTARA_32545	CPLX-7926	PWY-6352
7	1-	PHOSPHATIC		ATP_c + L- NTARA_10712 or NTARA_27261	NTARA_27261	HS00529-1	PWY-6352
8	1.1.1.141-	RXN		SZ13E-155, NTARA_24964	NTARA_24964		
9	1.1.1.145-	RXN		3b-hydrox, NTARA_1420 or NTARA_1420	NTARA_1420	N CPLX66-220	
10	1.1.1.170-	RXN		CPD-5846, NTARA_26835 or NTARA_26836	NTARA_26836	HS07423-1	
11	1.1.1.197-	RXN		ALPROSTA, NTARA_23800 or NTARA_23802	NTARA_23802		
12	1.1.1.210-	RXN		ALPHA-AN, NTARA_46344 or NTARA_9160	NTARA_9160	N CPLX66-220	
13	1.1.1.247-	RXN		CODEINE, NTARA_2042	NTARA_2042	MONOME	PWY-5270
14	1.1.1.248-	RXN		NADP_c + NTARA_45632	NTARA_45632	MONOME	PWY-5270
15	1.1.1.271-	RXN		CPD-13111, NTARA_5396 or NTARA_5396	NTARA_5396	N CPLX0-388	PWY-66
16	1.1.1.285-	RXN		CPD-7227, NTARA_2037 or NTARA_2039	NTARA_2039	N MONOME	PWY-5912
17	1.1.1.288-	RXN		CPD-7279, NTARA_3287 or NTARA_3287	NTARA_3287	N AT1G5234	PWY-695
18	1.1.1.34-	RXN		CO-A_c + NTARA_21945 or NTARA_35751	NTARA_35751	IMONOME	PWY-7524
19	1.1.1.39-	RXN		MAL_c + NTARA_13712 or NTARA_13712	NTARA_13712	IMALIC-NA	GLUCONEC
20	1.1.1.51-	RXN		NAD-P-OR, NTARA_1420 or NTARA_1420	NTARA_1420	N CPLX-8373	PWY-6943
21	1.1.1.64-	RXN		NADP_c + NTARA_1420 or NTARA_1420	NTARA_1420	N HS03054-1	PWY66-378
22	1.1.1.8-	RXN_c		GLYCEROL, NTARA_2282 or NTARA_27092	NTARA_27092	CPLX-8755	PWY-6118
23	1.1.3.37-	RXN		CPD-356_c, NTARA_19472 or NTARA_40931	NTARA_40931	IYML086C-	PWY30-6
24	1.1.99.28-	RXN		CPD-1538, NTARA_11818 or NTARA_11819	NTARA_11819	CPLX-7444	PWY-5530
25	1.10.2.2-	RXN		2.0 Cytoch, NTARA_13003 or NTARA_21445	NTARA_21445	G30-3124	PWY-7082
26	1.11.1.12-	RXN		A-LIPID-H, NTARA_13024 or NTARA_7267	NTARA_7267	N HS09562-1	PWY-4081
27	1.11.1.15-	RXN		Alkyl-Hyd, NTARA_10388 or NTARA_10389	NTARA_10389		
28	1.14.11.18-	RXN		2-KETOGL, NTARA_50999	NTARA_50999		

Model File

- Filter out all unique reactions from the source file.
- Create an initial model of the plant using Gene information from the Metacyc Database, which might contain incorrect Genes/ GPR information.
- Fill up the Genes and GPR column with the correct information from the source file. Done via siphoning out all the genes related to each unique reaction in the source file.
- Identify metabolites that have missing information (related to its Charge, Neutral Formula etc) and source as much of these information from other completed models.
- Then, fill up these missing entries for the metabolites which we have information of, while leaving the rest blank.

2) Compartmentalize the reactions in the model.

- Using a reference Python dictionary which contains the name of the compartments and its constituent reactions, remove all of the reactions that are not in the model from this reference dictionary.
- Then, use this updated dictionary to compartmentalize the model. Compartmentalization entails the addition of proper suffixes at the end of the reaction (Data Tagging) to allow the user and program to identify which compartment the reaction belongs to.

How the Compartmentalization code works

Step 1: Add all reactions in model that are not in 'default' compartment of input dictionary to a new list, reac_not_default.

Step 2: Add all reactions in model that are not in both reac_not_default and default compartment of input dictionary to the default compartment.

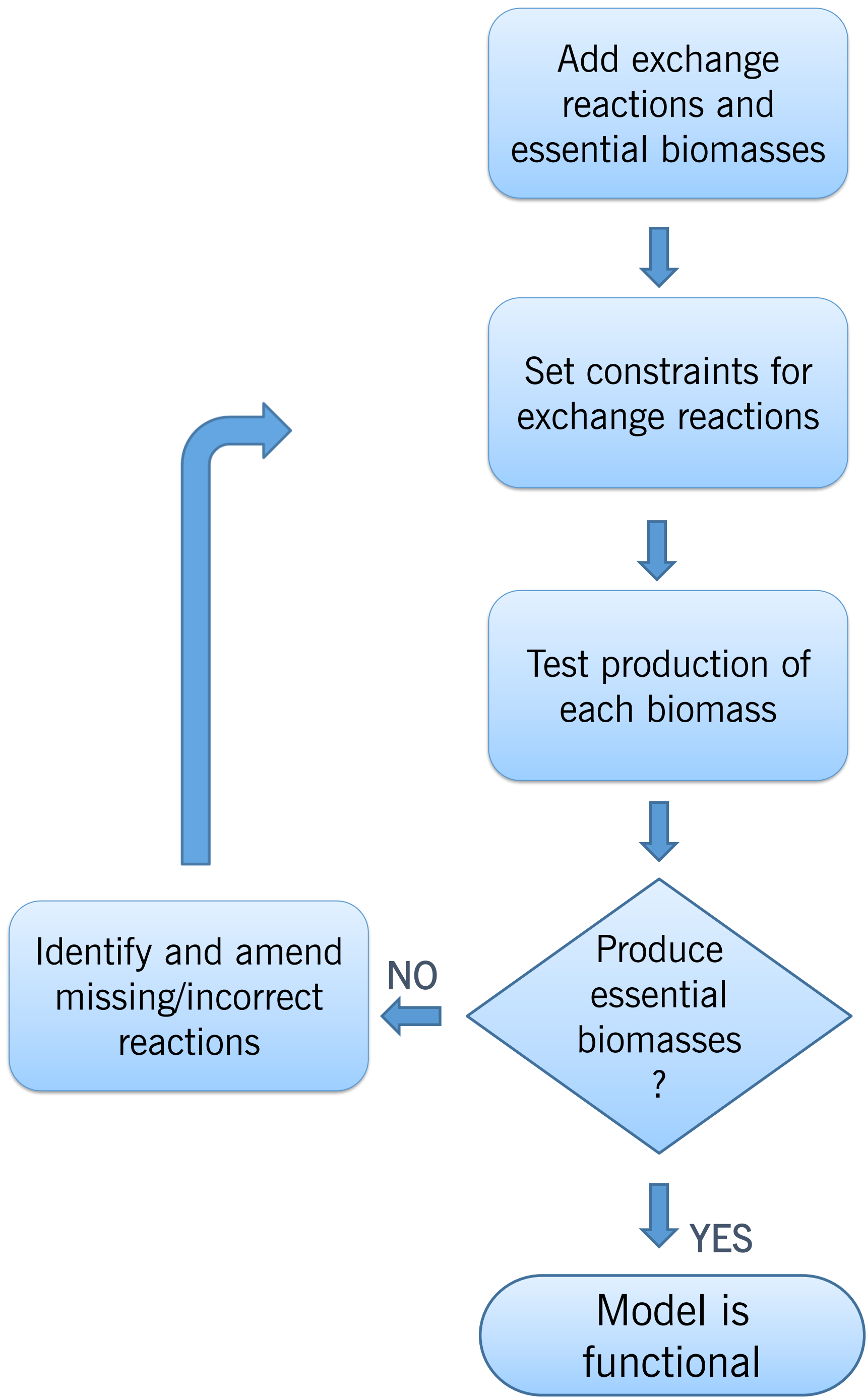
Step 3: Create an empty model, compartmented_model.

Step 4: Add all reactions in the model to the compartmented_model with their suffixes and IDs.

Step 5: Add all metabolites of the aforementioned reactions into the model too with their respective suffixes.

Step 6: Return this model, compartmented_model, to the user.

3) Test the model's ability to produce essential biomasses and check for unbalanced or missing reactions..



Results and Future Directions

- The final draft model contains a sizeable amount of information, namely: 3073 Reactions, 3209 Metabolites, 9020 Genes.
- Upon testing the model, essential reactions containing Pumped-Proton have either incorrectly named metabolites or were missing from the model. Some essential transporters were missing from the model as well.
- After amending the above issues, the number of biomasses produced for the NTARA model rose to 11. The biomasses are: GLC, Glycerol, NO_3^- , Pi, SO_4^{2-} , Xylan, sETOH, sFUM, sGLC, sMAL and sSucrose.
- Some biomasses, however, remain unproduced, and work is needed to identify the remaining missing/incorrect reactions. Once we have a functional model, we can use FBA to simulate the metabolic processes of the plant.

Acknowledgements

I would like to give my sincerest thanks to Professor Maurice Cheung for giving me the valuable opportunity to participate in this exciting research project, the JY Pillay Global-Asia Programme for the generous funding which enabled me to complete my research project in Singapore without any inconvenience, and of course the Dean of Faculty Office at Yale-NUS College for these valuable opportunities. I also would like to acknowledge the data provided by Dr. Nilo-Poyanco (Project Fondecyt Iniciacion 11150107).