Supplementary Material Depicting the Comparative Analysis of Model Collections

Contents

To simplify interpretation, the following figures are grouped by the sections of their corresponding test cases as they appear in a snapshot report. The code that was used to generate the data and figures has been deposited on GitHub https://github.com/biosustain/memote-meta-study.

1 Tested models

In order to respect the limited resources on the DTU high performance computing infrastructure, we set a maximum time limit for running the memote test suite. This introduced a bias against large models. Additionally, certain models failed the testing procedure. In the following we tabulate the total size of the collections as well as the final number of tested models. The results are shown in Table S1.

Table S1: Number of tested models.

Collection	Number of Models	Tested Models	%
AGORA	818	801	97.9
CarveMe	5587	5511	98.6
Path2Models	2641	2641	100.0
KBase	1637	1632	99.7
BiGG*	36	36	100.0
Ebrahim *et al.*†	83	80	96.4
OptFlux Models†	100	79	79.0

* Please note that we removed the large number of *Escherichia coli* strain models from the BiGG collection and only included results from the models iJR904, iAF1260, iJO1366, and iML1515.

† 39 models from these two collections are likely identical based on a filename comparison.

2 Clustering

In order to perform the clustering analyses, we used all normalized test metrics excluding some particular cases. Excluded are the Sections 3.4.2 & 3.4.6 because the basic information only contains unnormalized model dimensions and because a biomass formulation is not present in all models. We further removed individual biomass related test cases, as well as the metabolic coverage since that is not properly normalized. Additionally, test cases that contained errors were penalized with the worst metric of one.

To determine the most relevant tests to discriminate between model collections, we built a classifier using a random forest (Breiman 2001) over the collections and normalized test results (0.99 accuracy and 0.01% out-of-bag (OOB) error). Then, the importance of each variable, i.e., test case, was ranked with the Mean Decrease in Accuracy (MDA) (Louppe et al. 2013). This metric measures the total decrease in accuracy, averaged over all trees of the forest, when the value of a given variable is permuted in the OOB samples. Figure S4 represents the 15 most discriminant features on average (see last column) and their independent relevance by collection. The higher the decrease in accuracy, the higher the relative contribution of such a test to differentiate among collections. Thus, the five most discriminant tests are purely metabolic reactions, transport reactions, dead-end metabolites, orphan metabolites, and the presence of a non-growth associated maintenance reaction. Although there is a variable range of importance for each collection, e.g., for CarveMe transport reactions and orphans are more relevant; for Kbase transport reactions; for Ebrahim *et al.* purely metabolic reactions. For a detailed study of the clustering properties, please refer to the *Supplementary Clustering Analysis* notebook.

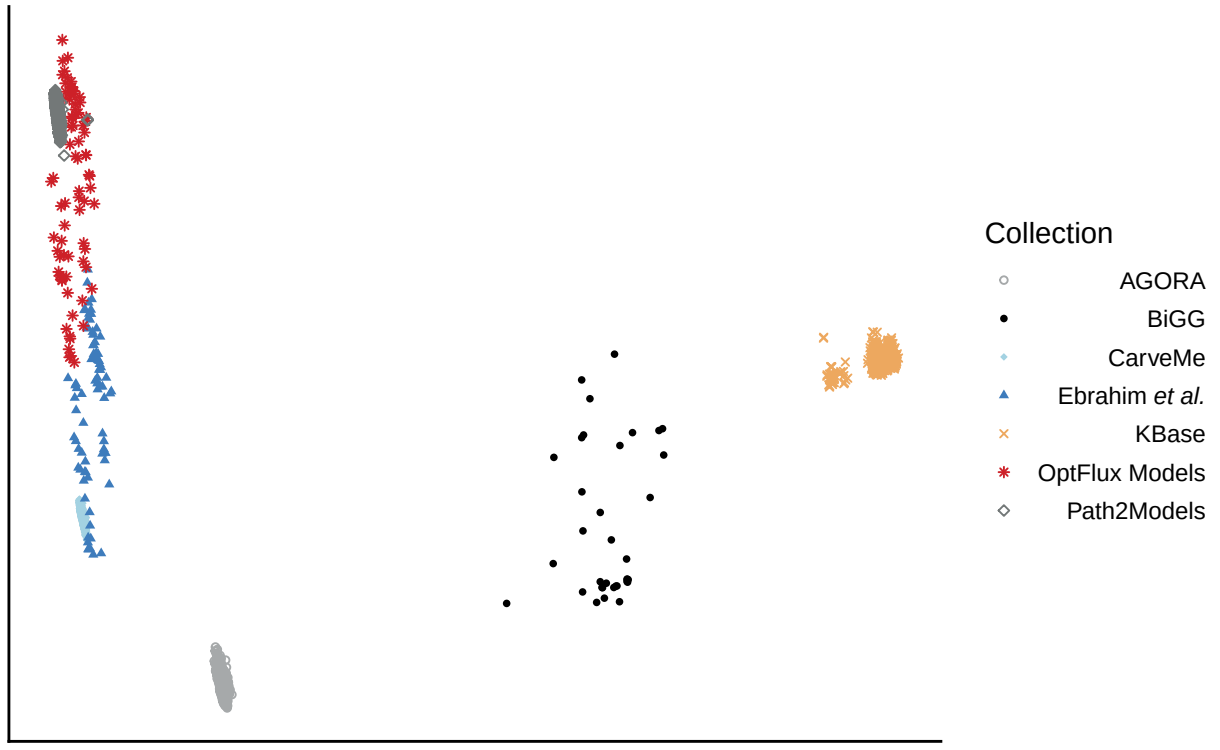

Figure S1: Depicted are the first two components of a principal components analysis of the normalized test features (metrics).

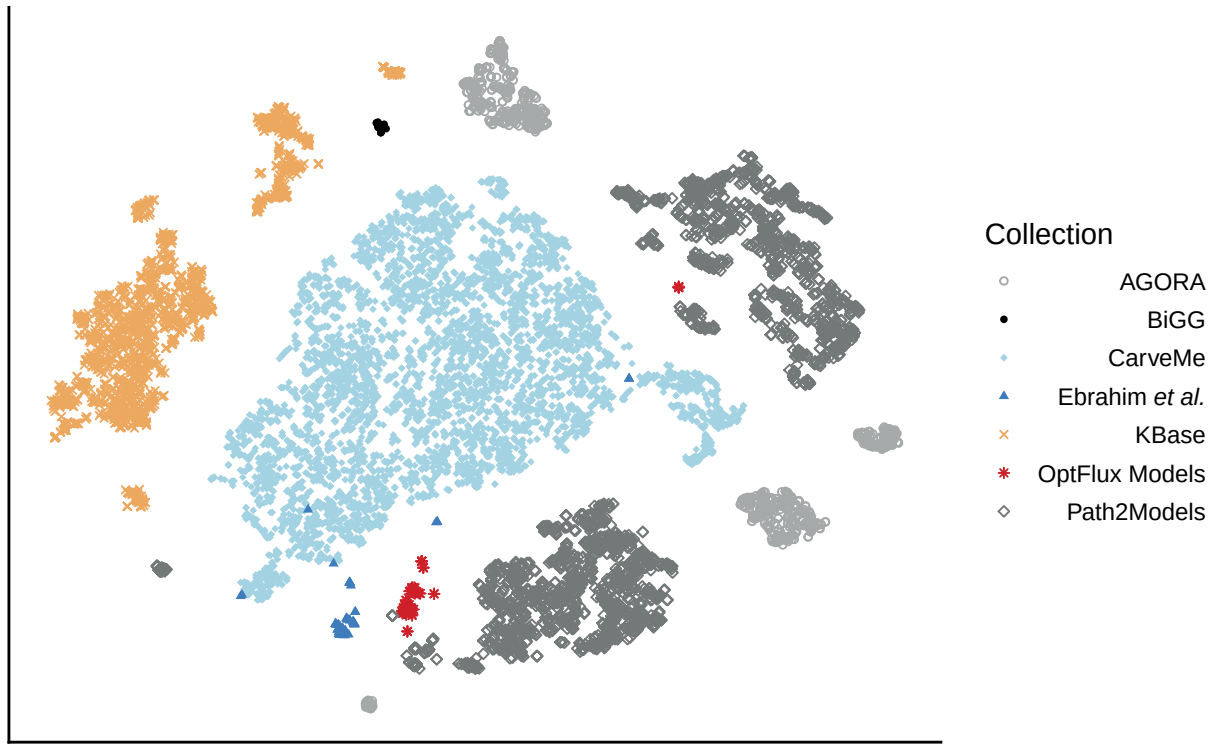

Figure S2: Depicted are the distances between models in higher order space given by the normalized test features reduced to two dimensions using t-SNE.

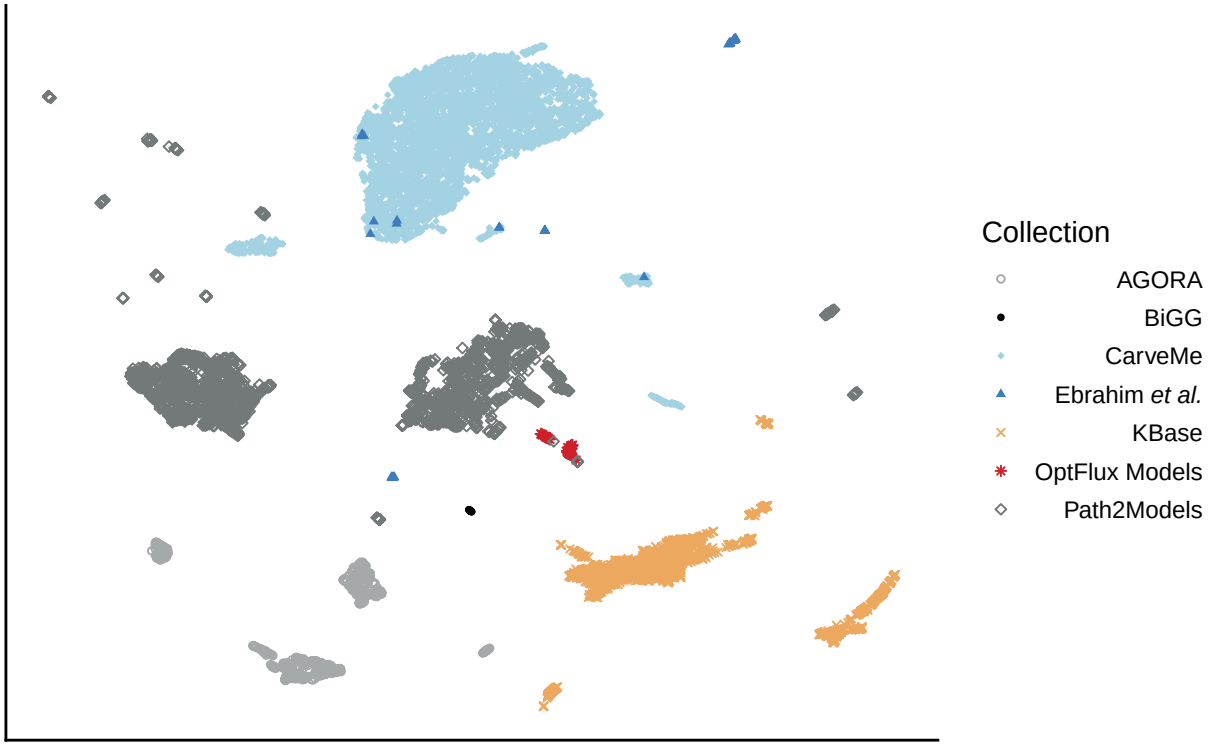

Figure S3: Depicted are the distances between models in higher order space given by the normalized test features reduced to two dimensions using UMAP.

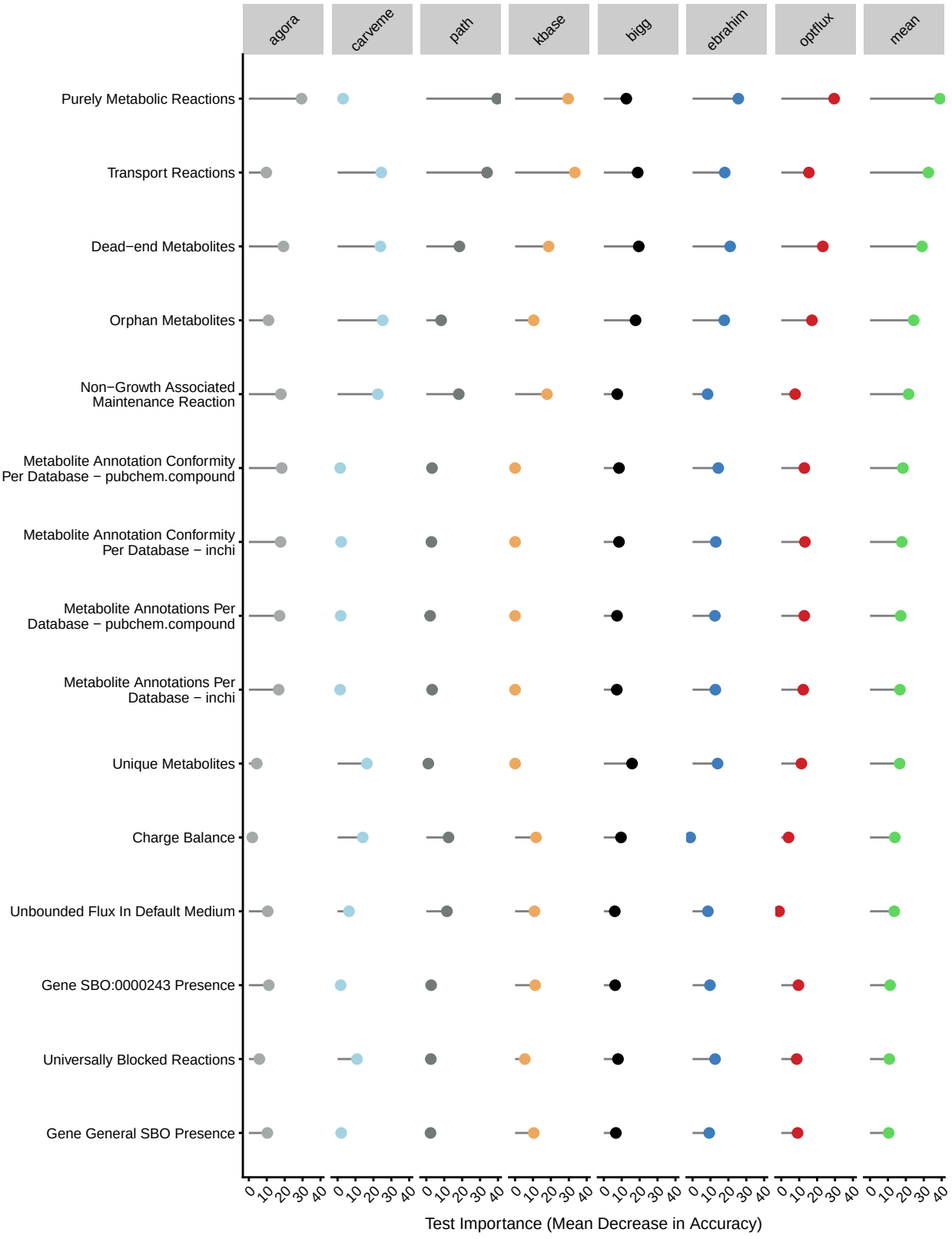

Figure S4: 15 most relevant tests to discriminate among GEM collections, for each collection and the mean. Ranked in decreasing importance according to the *mean decrease in accuracy* metric averaged over all collections (last column), computed over a random forest classification model.

3 Test Suite

3.1 Summary of Observations

- SBO terms are only used by models from KBase and BiGG (Figure S80).
- Models from Path2Models and Opflux Models are formatted in legacy SBML (< Level 3,Version 1) without FBC package (Figures S92 & S93).
- Models from the collections of Ebrahim *et al.*, and OptFlux Models are highly variable for many specific tests. Models from automatic reconstruction pipelines (AGORA, CarveMe, Path2Models, and KBase) or the controlled BiGG collection are much more similar within each collection yet still different from each other. This could be due to each collection focusing on a distinct set of taxonomies but could also be related to the algorithms and databases behind each collection (Section 3.4.8; Figures S105, S107, and S113).
- On biomass:
  - Only for a minority of models in BiGG, Ebrahim *et al.*, and OptFlux Models memote could not identify a biomass reaction (Figure S117).
  - A portion of models in the BiGG collections have inconsistent biomass equations followed by OptFlux Models and models in the collection by Ebrahim *et al.*; all models in the CarveMe and Path2Model collections have inconsistent biomass reactions (Figure S118).
  - Models that cannot be simulated using the default or complete medium exist in Path2Models, BiGG, Ebrahim *et al.*, and OptFlux Models (Figure S119 & S120).
  - Possible artifacts from automatic reconstruction are present in models from AGORA and KBase that grow despite some biomass precursors being blocked when each precursor is optimized individually in default and complete medium (compare Figures S121 & S122 with S119 & S120).
- The average fraction of reactions that participate in stoichiometrically-balanced cycles is larger for models from automatic reconstruction pipelines (AGORA, CarveMe, Path2Models, KBase) than for BiGG, Ebrahim *et al.*, and OptfluxModels (Figure S131). This could be an artifact from automatic reconstruction processes.
- Reactions that involve oxygen are integral to the energy metabolism of many organisms. Not constraining these reactions carefully can lead to predictions that deviate from the expected phenotype, i.e., allowing anaerobic growth that should not be possible. The portion of oxygen-containing reactions that are reversible varies strongly across all seven collections. Models in BiGG have the lowest variance whereas models from Path2Models, Ebrahim *et al.*, and OptFlux vary strongly (Figure S127).

3.2 Scores

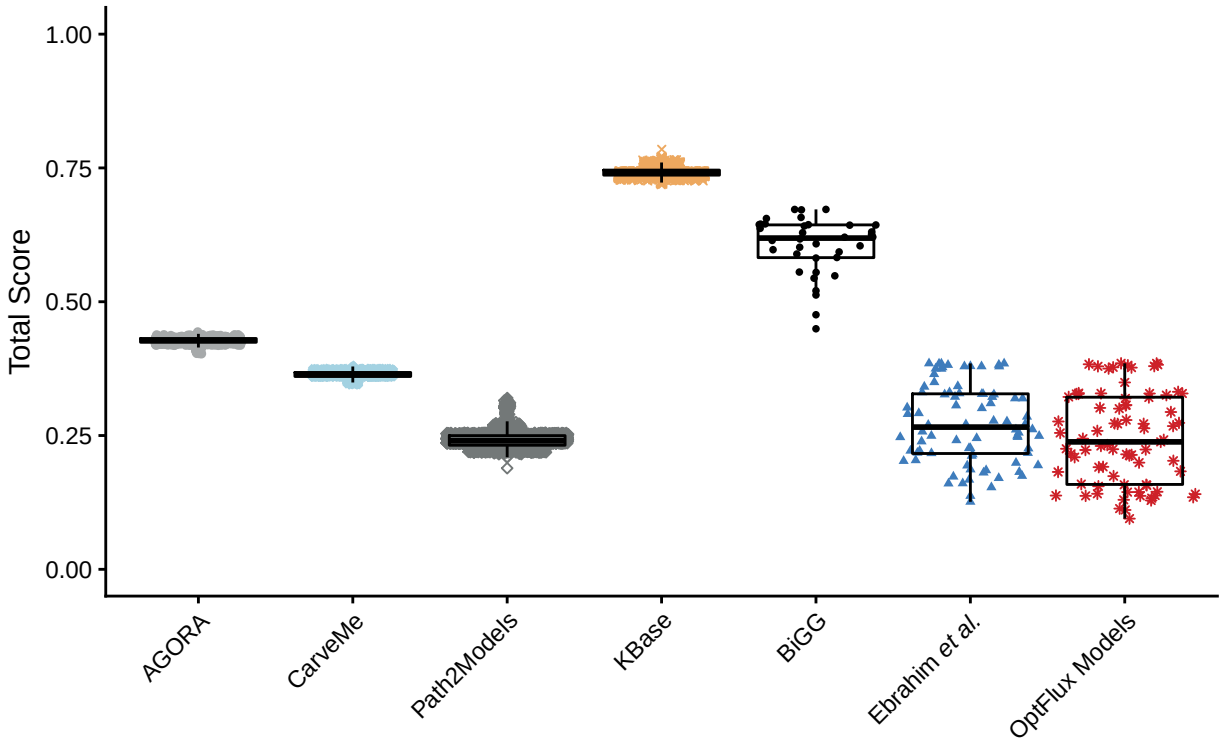

Figure S5: Total Score. Depicted are the sums of all test scores in all independent sections, applying the weights for individual test cases and sections as detailed in the snapshot report.

3.3 Independent Section

3.3.1 Consistency

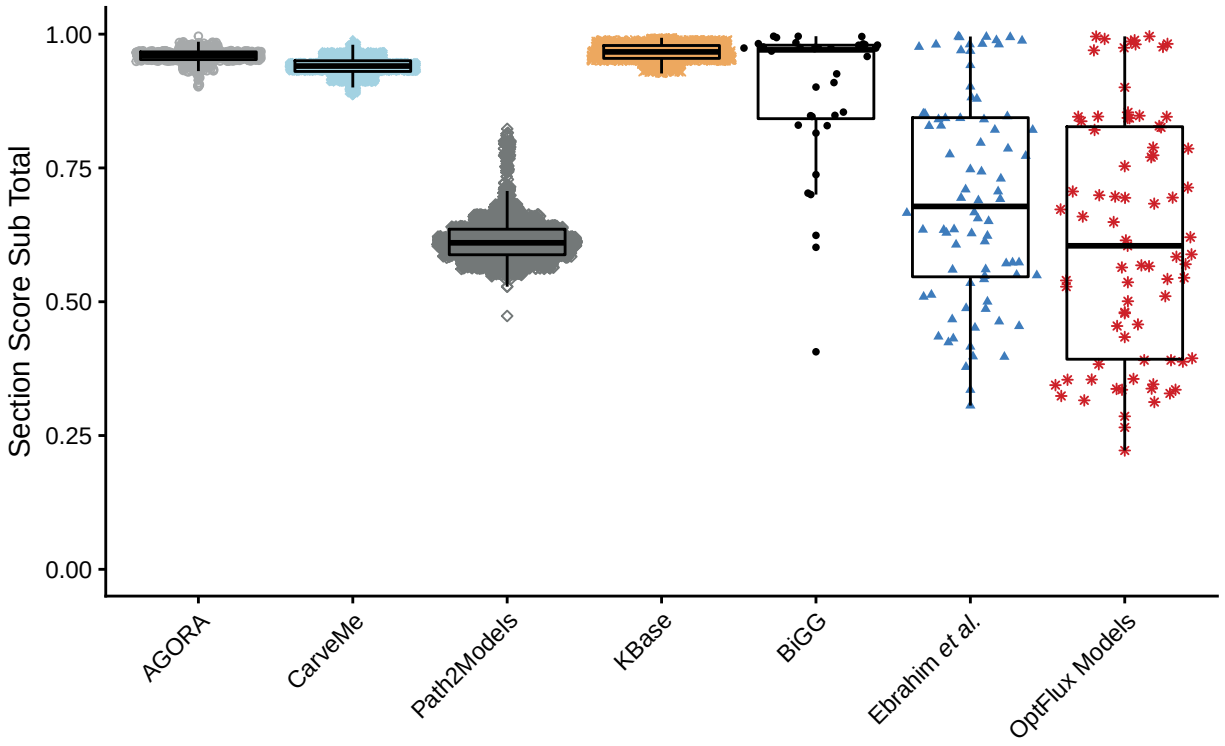

Figure S6: Consistency. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

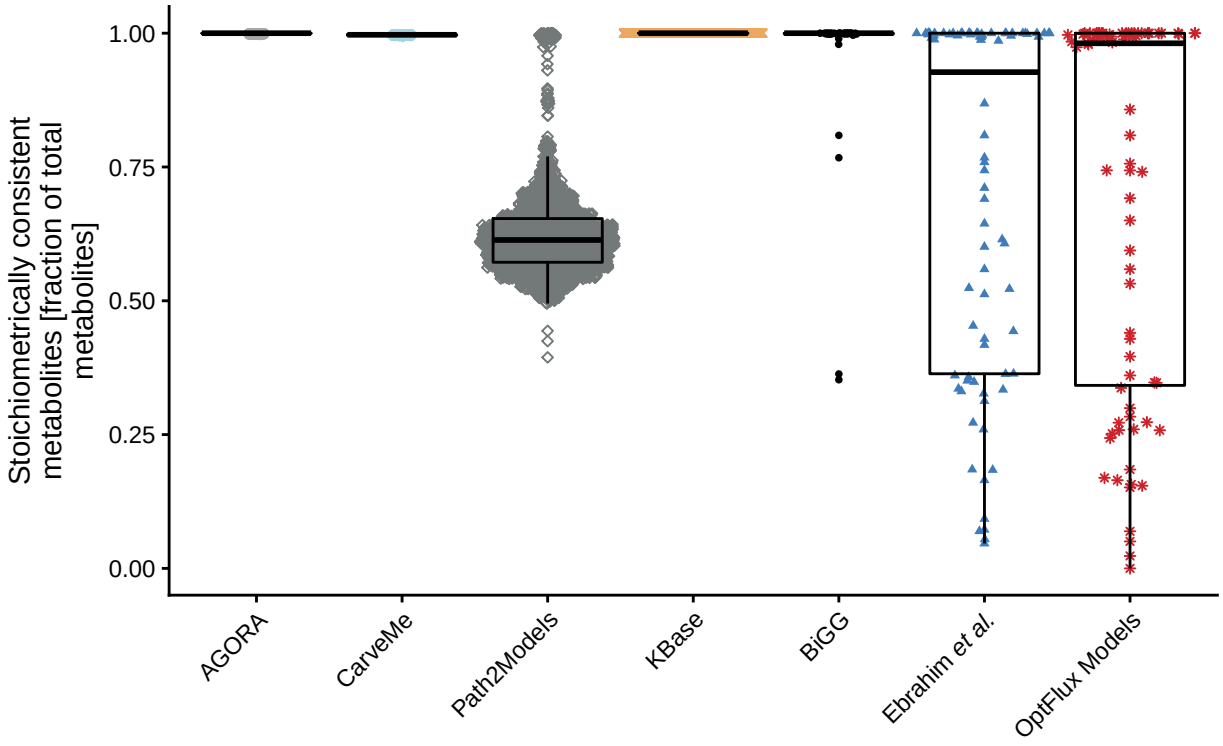

Figure S7: Stoichiometric consistency

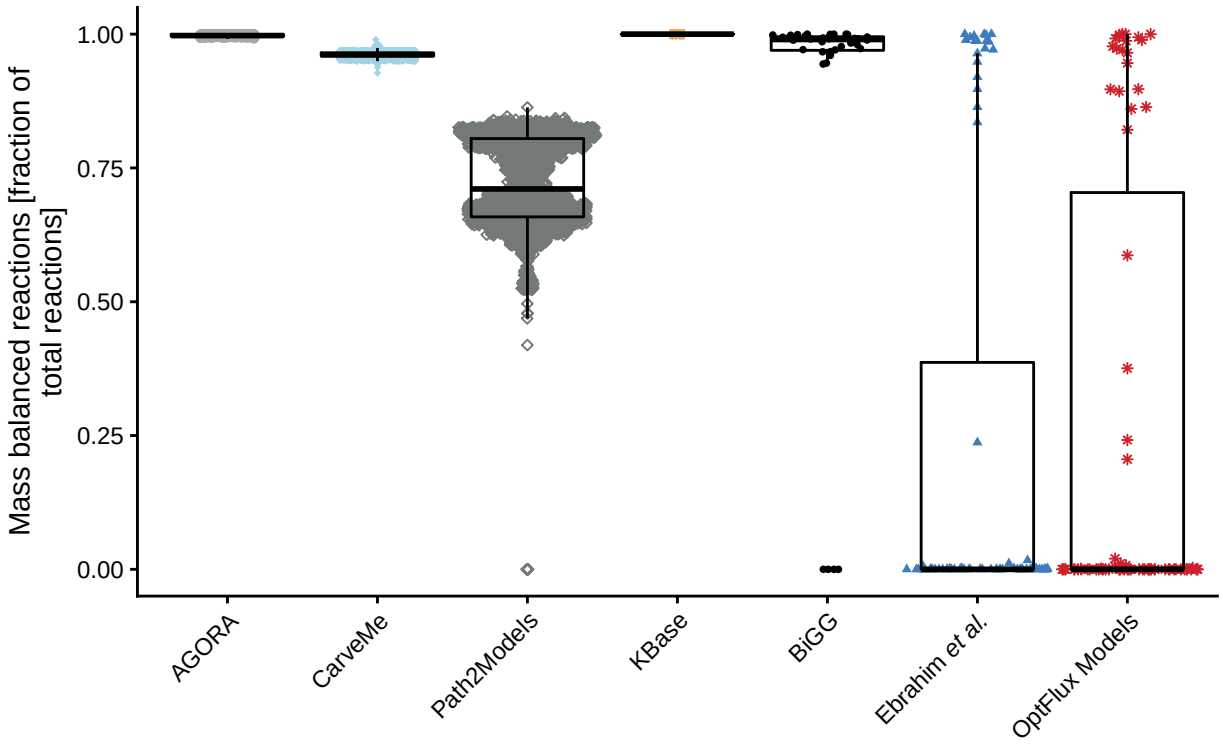

Figure S8: Mass Balance. Please note that any reaction where at least one metabolite lacks a formula annotation is considered as unbalanced for the purpose of this test.

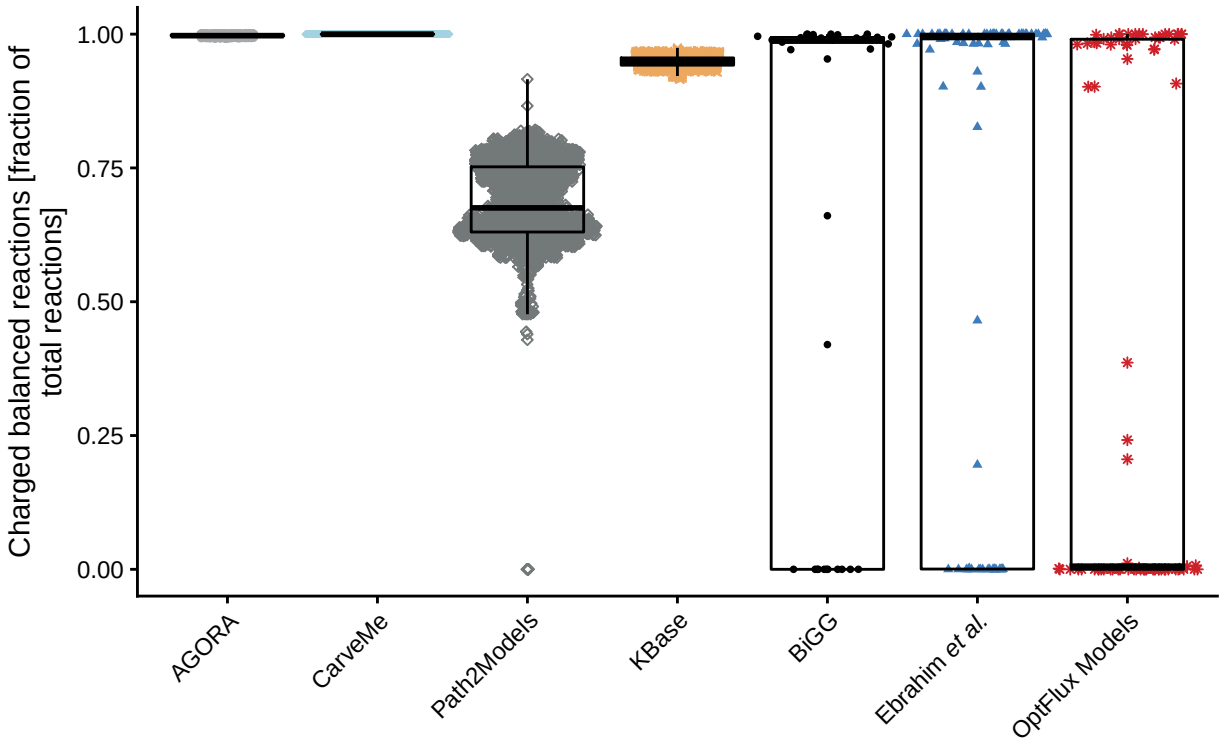

Figure S9: Charge Balance. Please note that any reaction where at least one metabolite lacks charge information is considered as unbalanced for the purpose of this test.

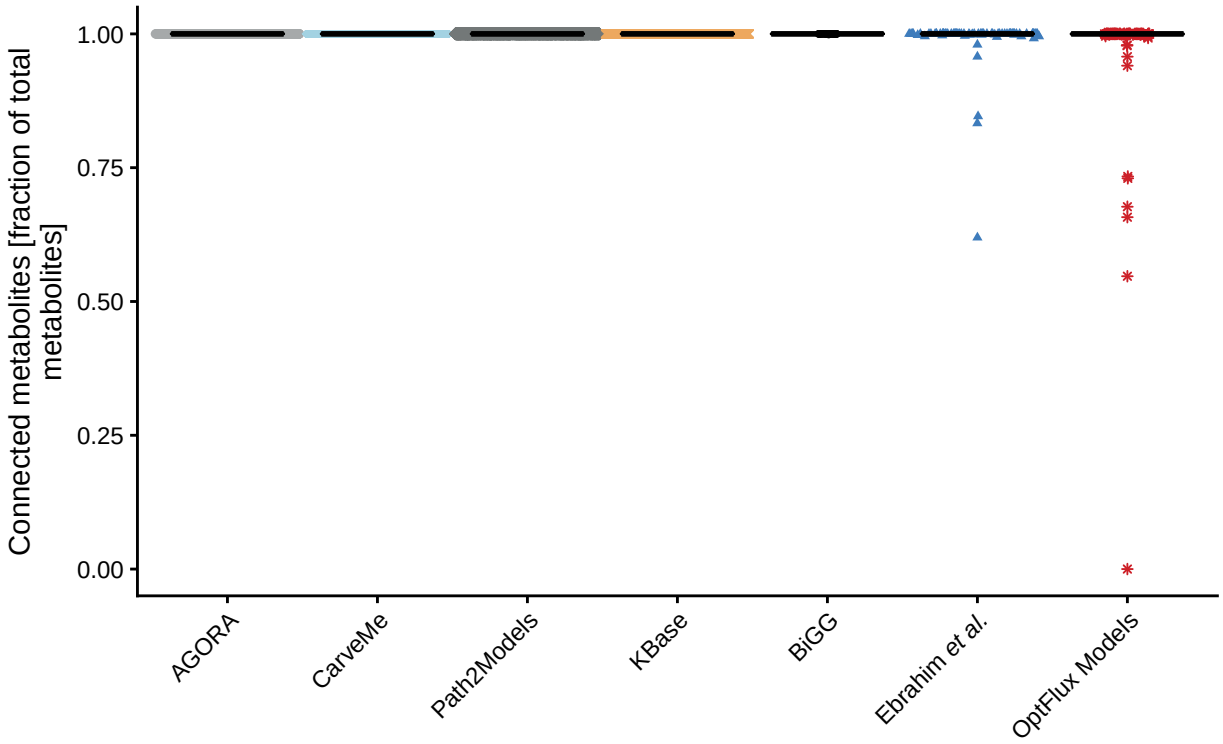

Figure S10: Metabolite Connectivity

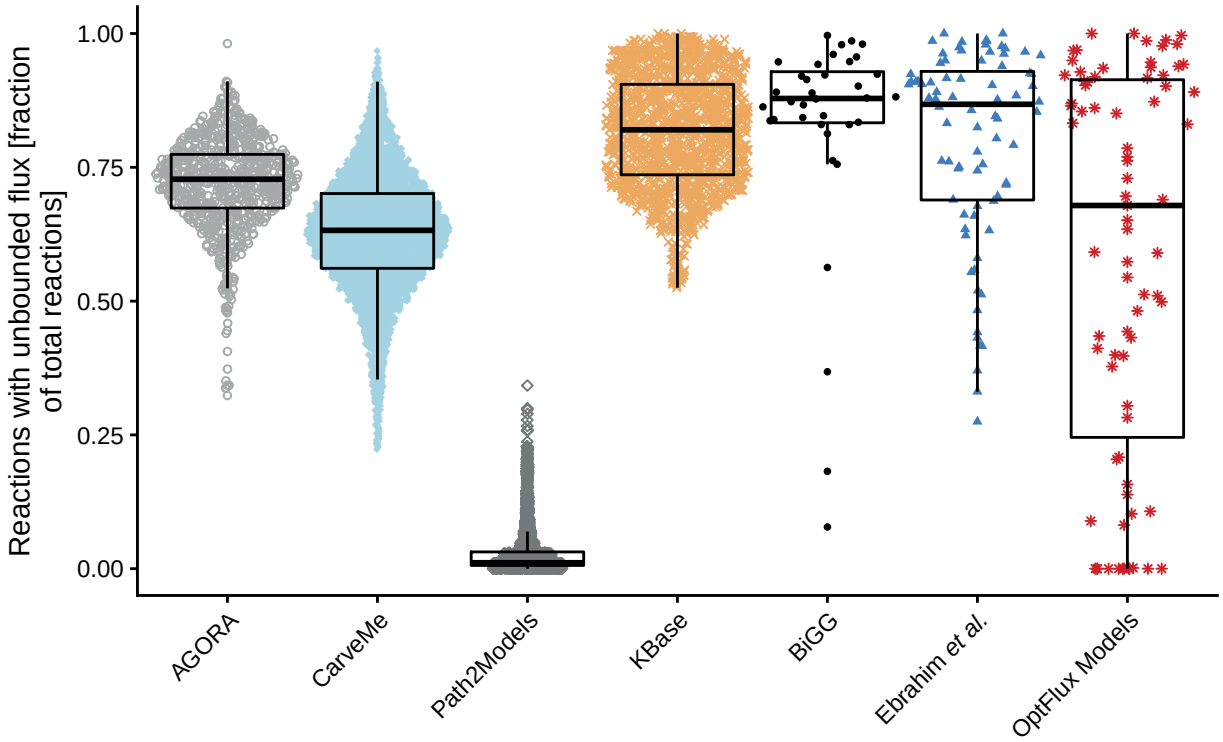

Figure S11: Unbounded Flux in Default Medium

3.3.2 Annotation - Metabolites

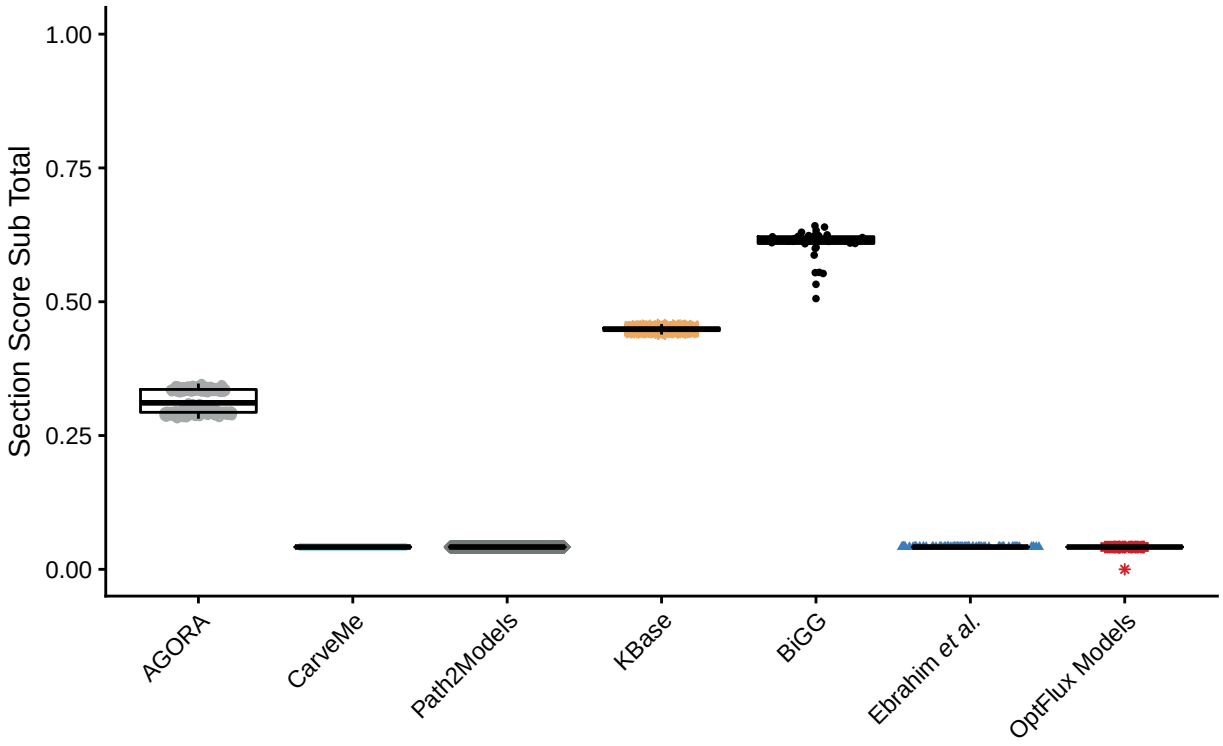

Figure S12: Annotation - Metabolites. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

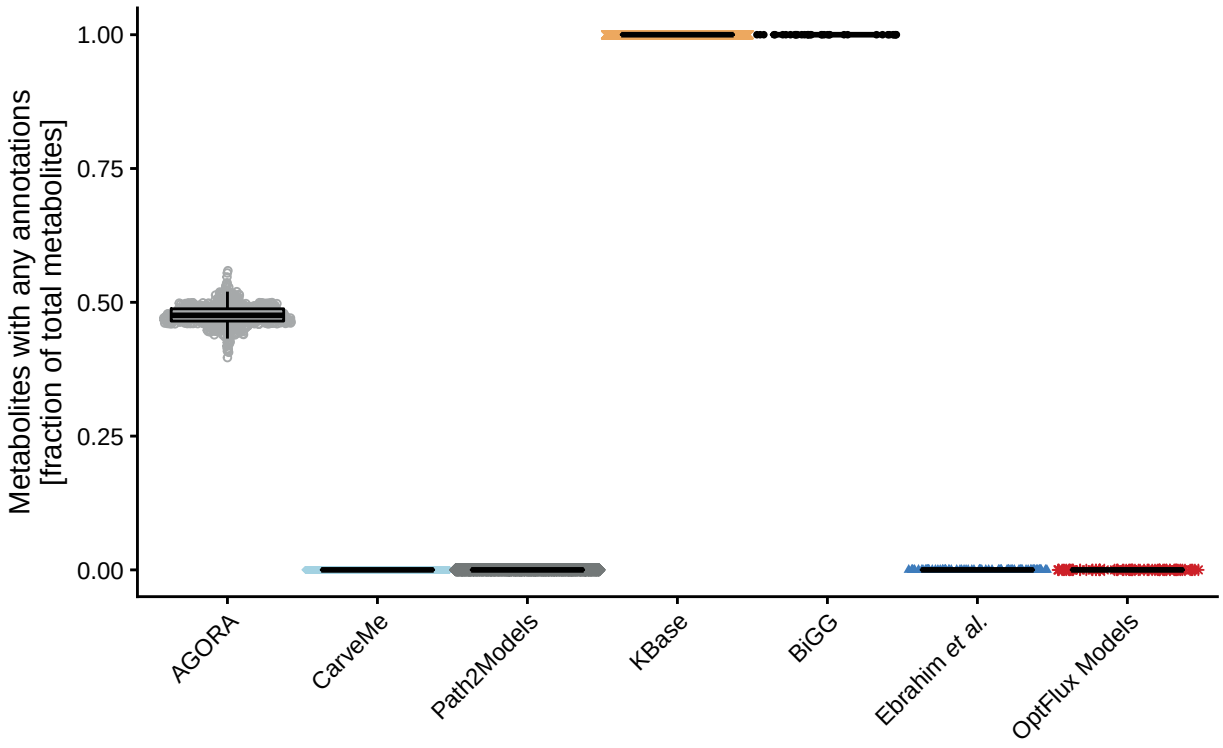

Figure S13: Presence of Metabolite Annotation

3.3.2.1 Metabolite Annotations Per Database

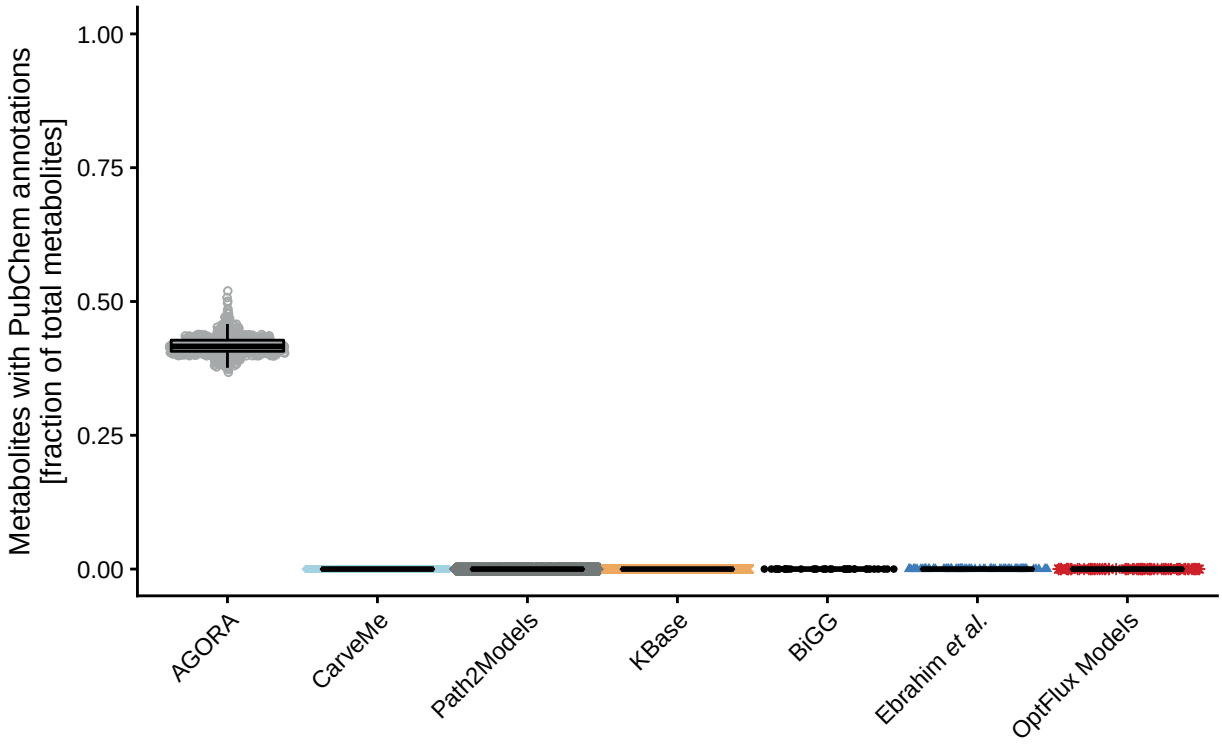

Figure S14: Metabolite Pubchem.compound Annotation

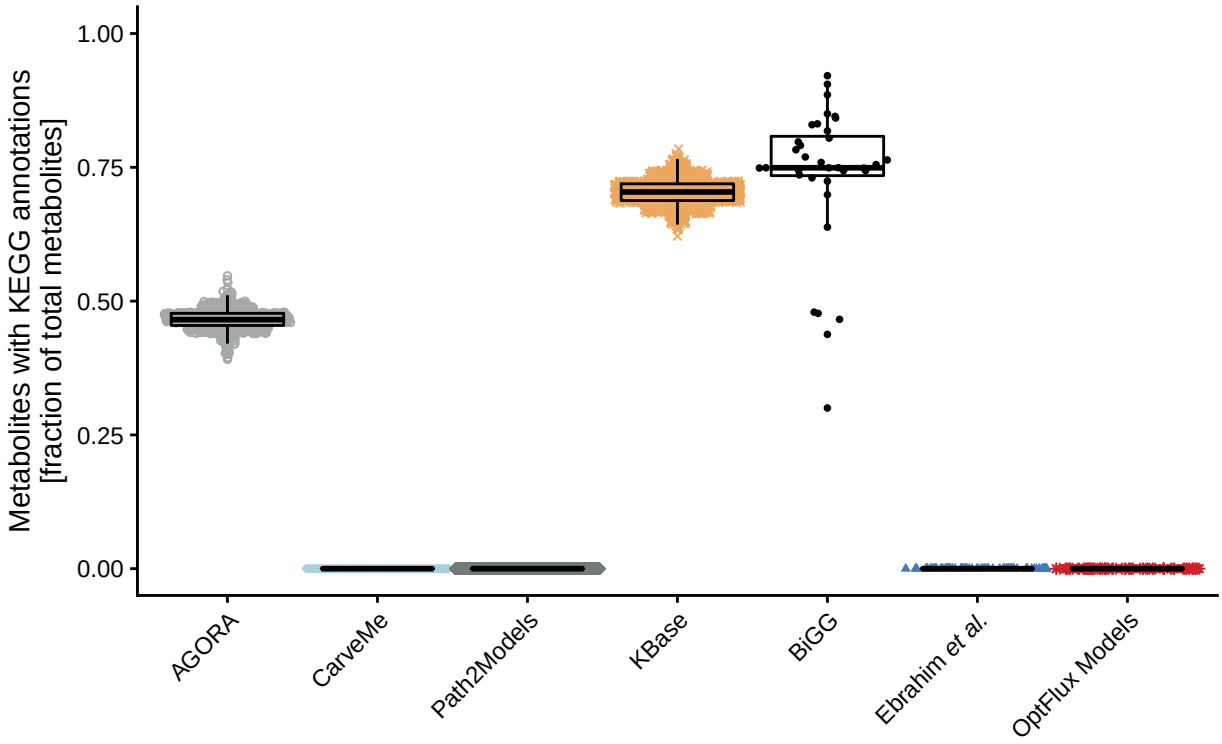

Figure S15: Metabolite KEGG.compound Annotation

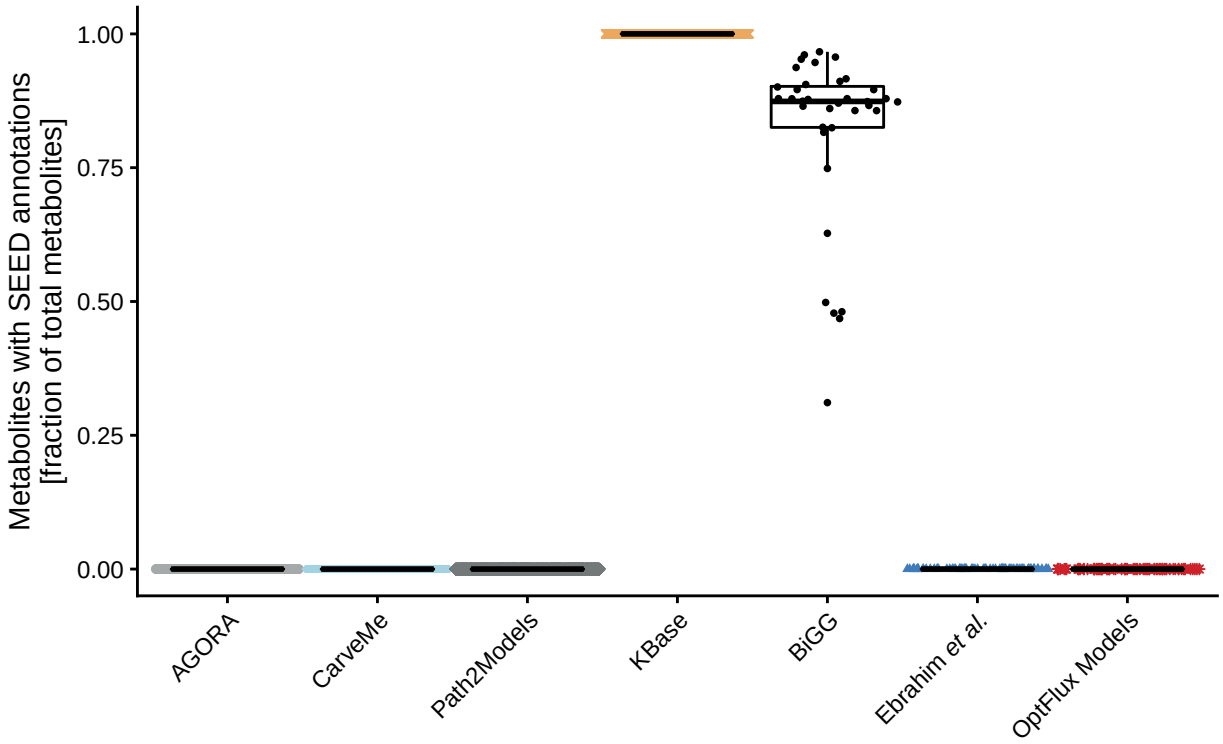

Figure S16: Metabolite SEED.compound Annotation

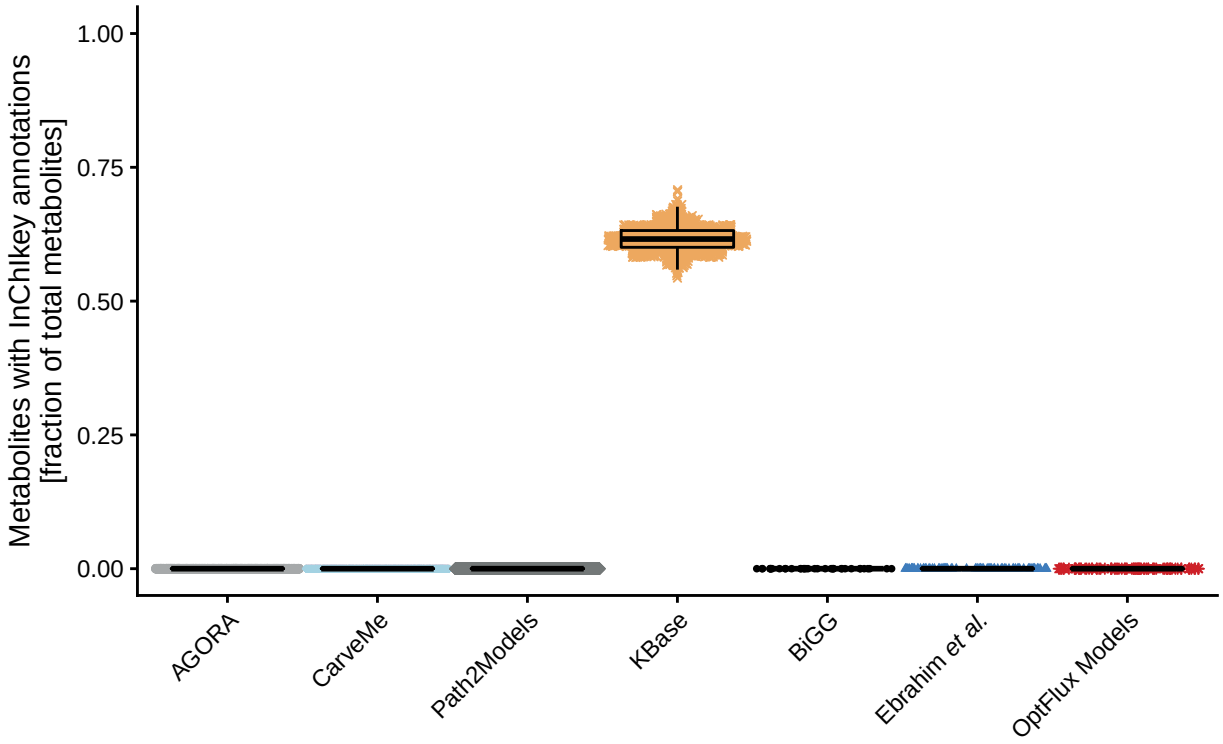

Figure S17: Metabolite InChIKey Annotation

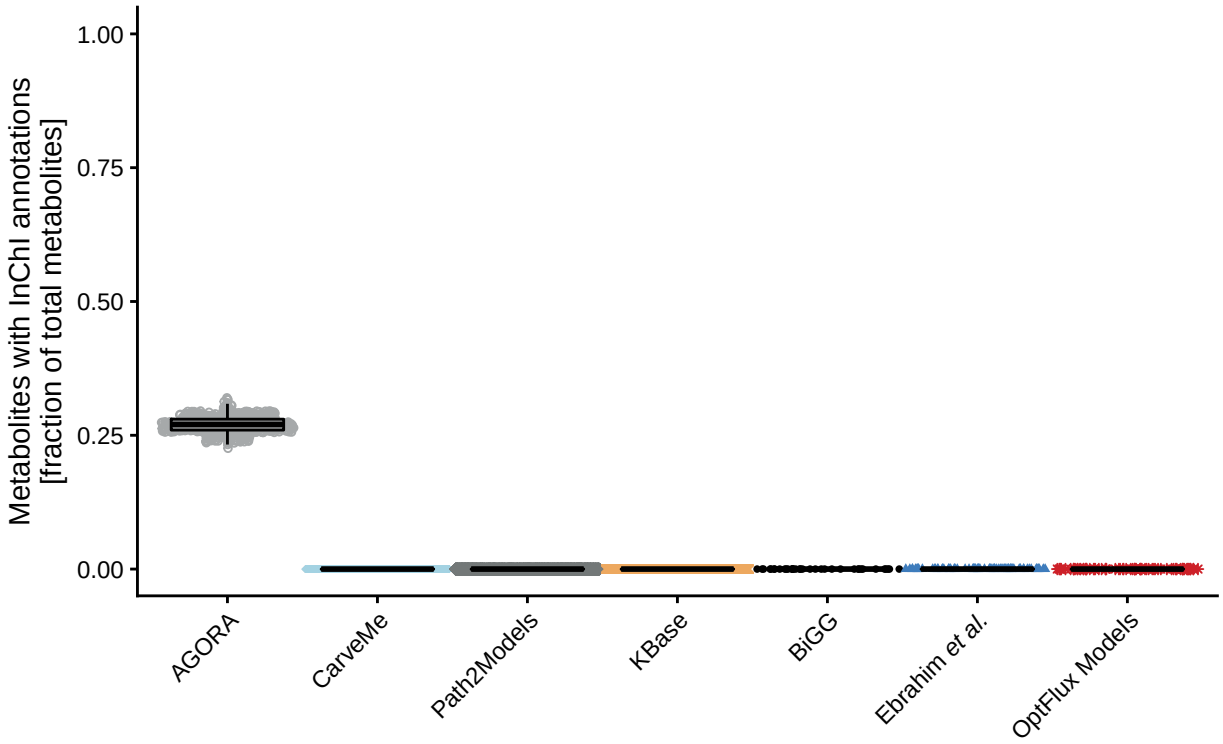

Figure S18: Metabolite InChI Annotation

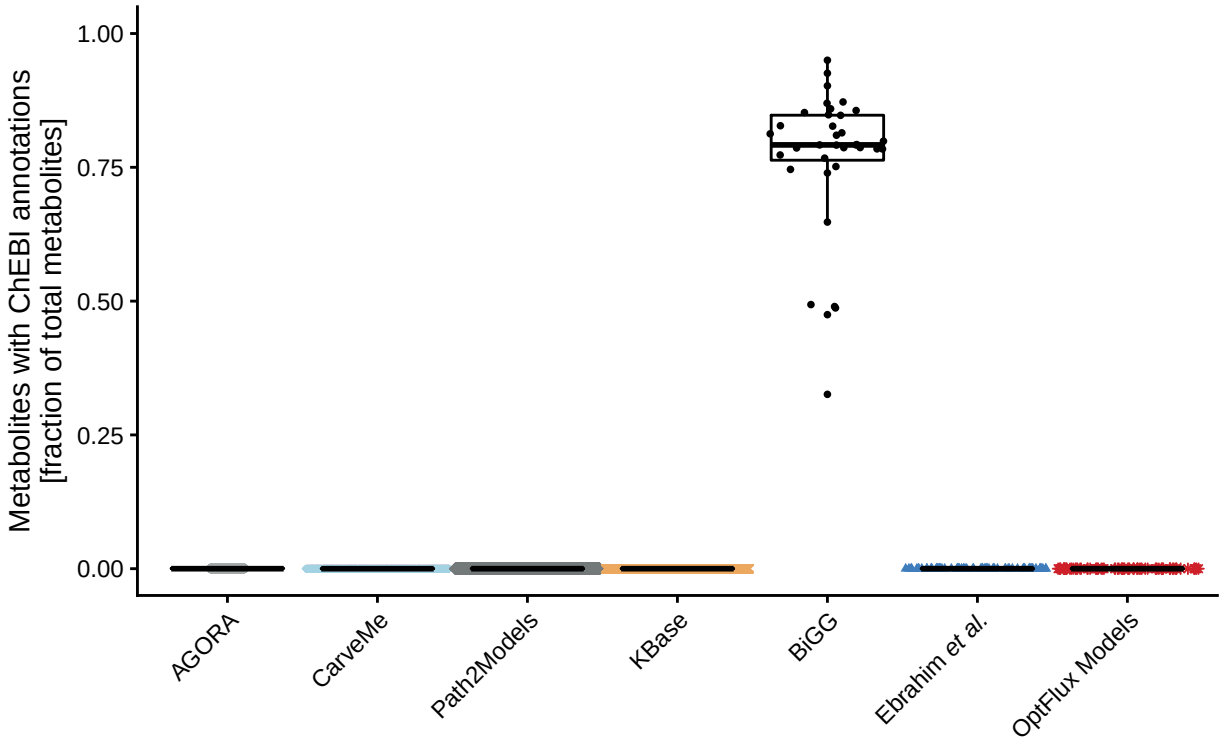

Figure S19: Metabolite ChEBI Annotation

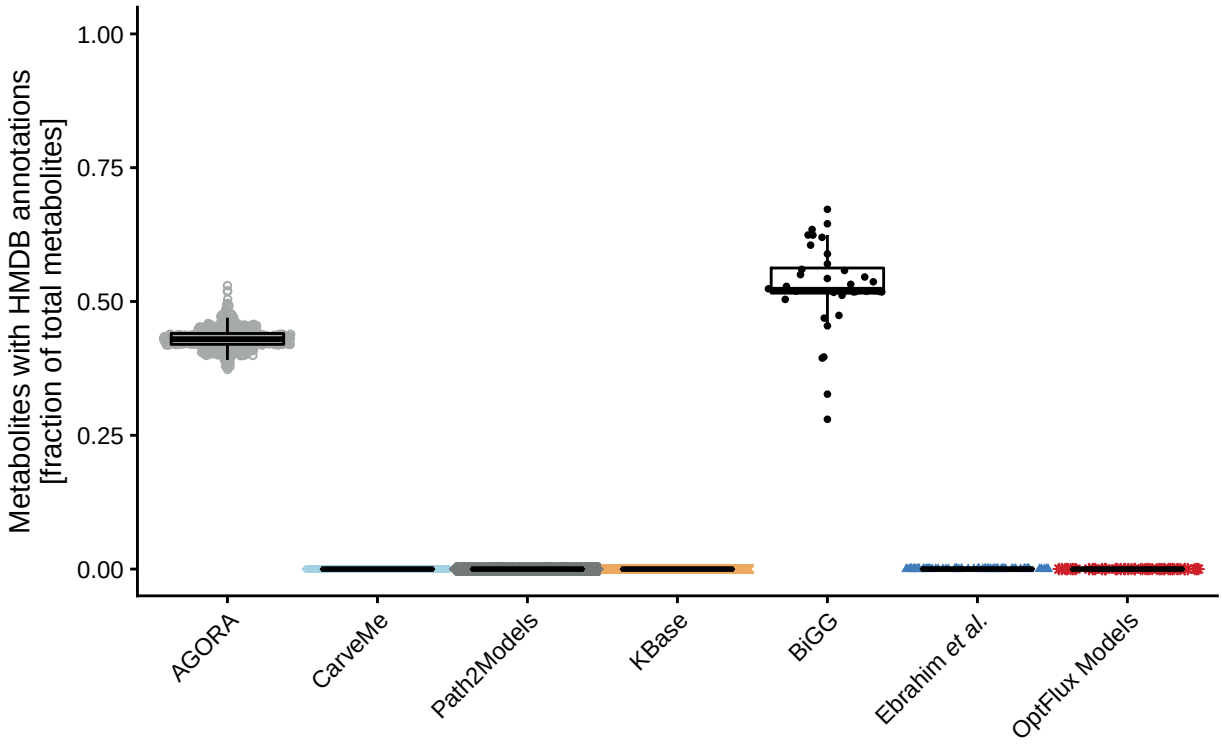

Figure S20: Metabolite HMDB Annotation

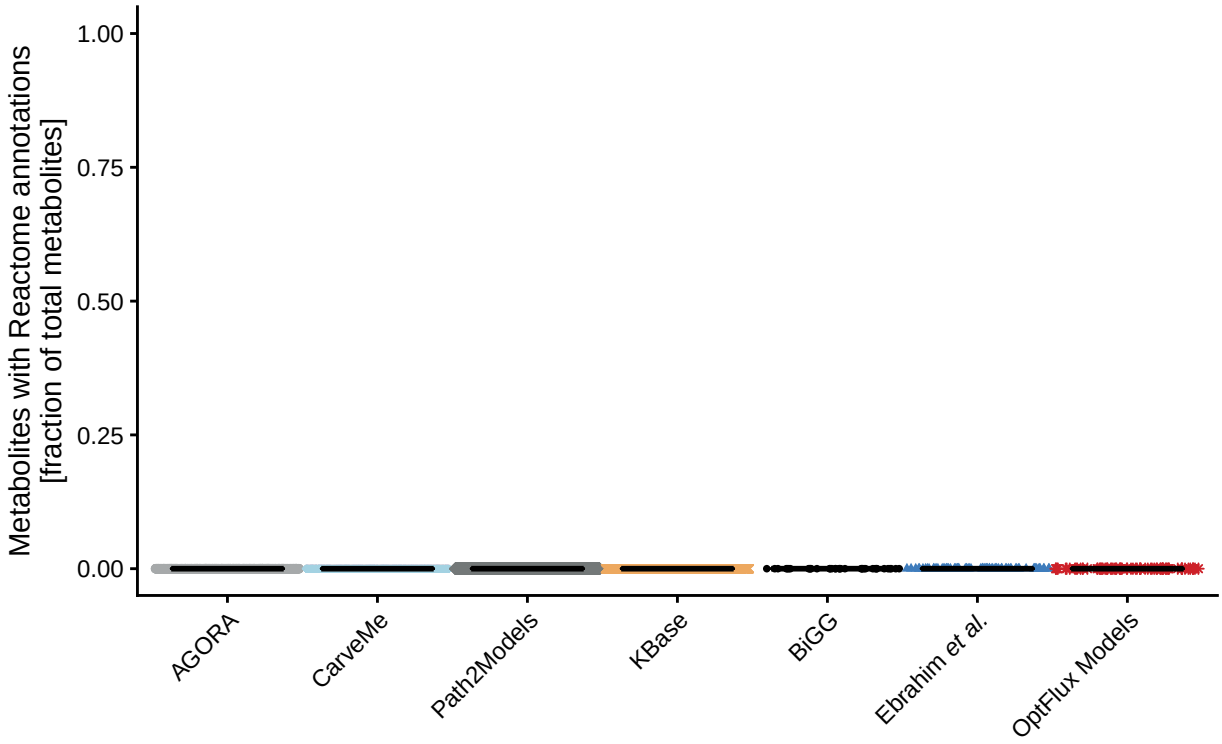

Figure S21: Metabolite Reactome Annotation

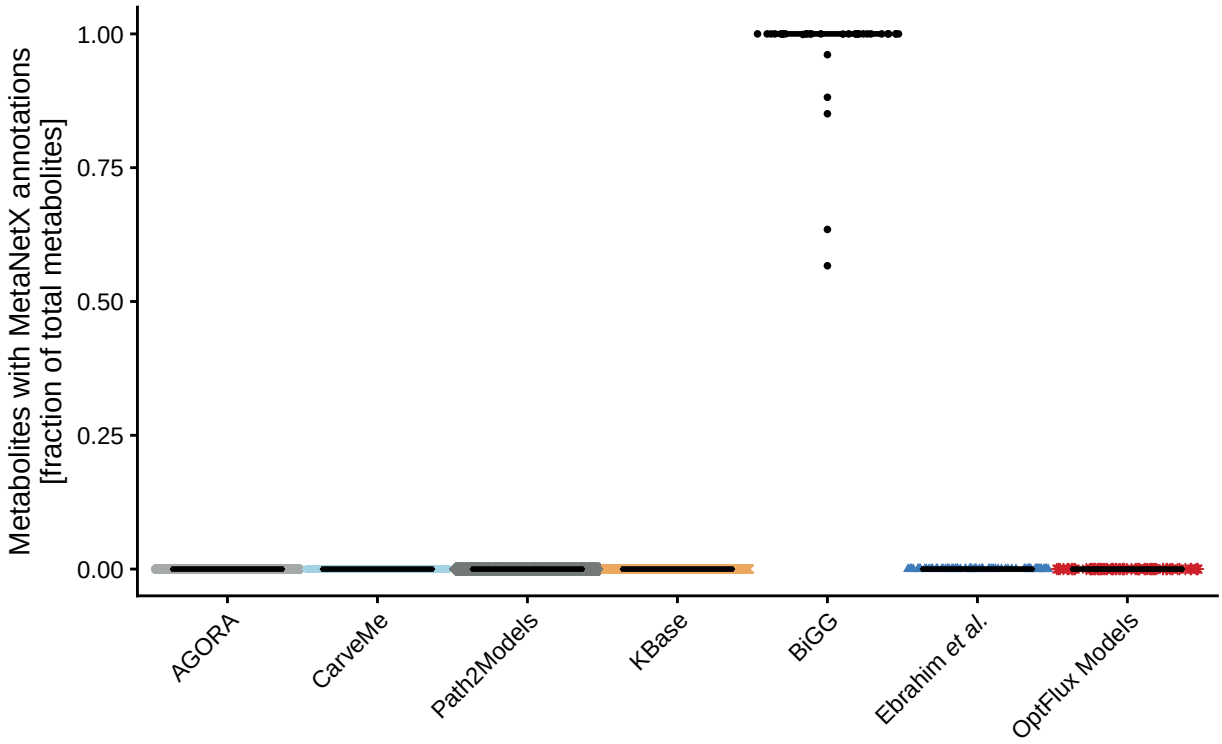

Figure S22: Metabolite MetaNetX.chemical Annotation

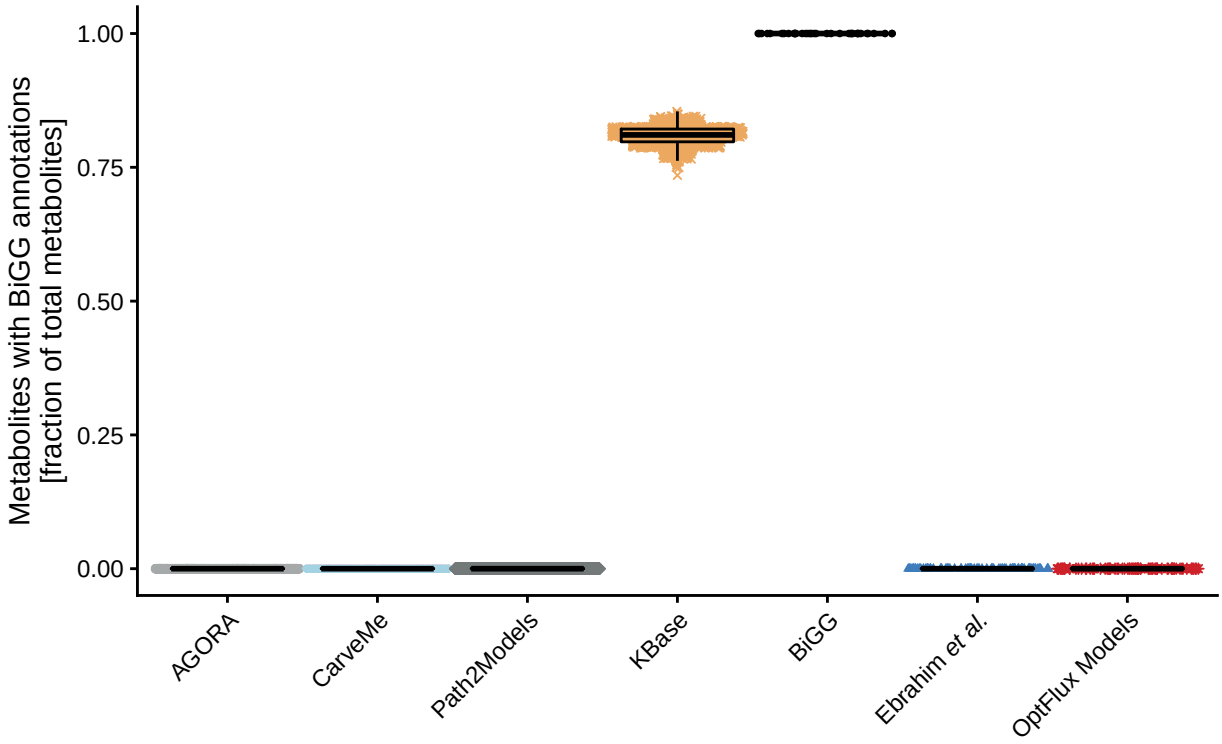

Figure S23: Metabolite BiGG.metabolite Annotation

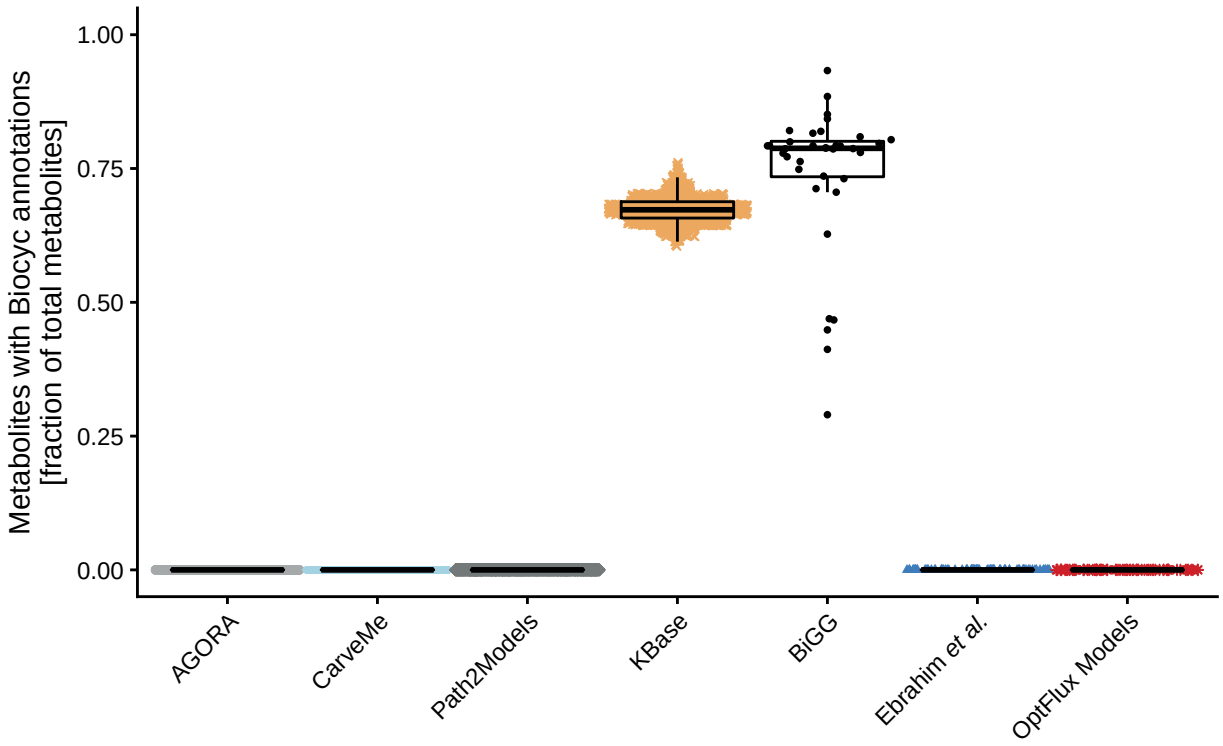

Figure S24: Metabolite BioCyc Annotation

3.3.2.2 Metabolite Annotation Conformity per Database

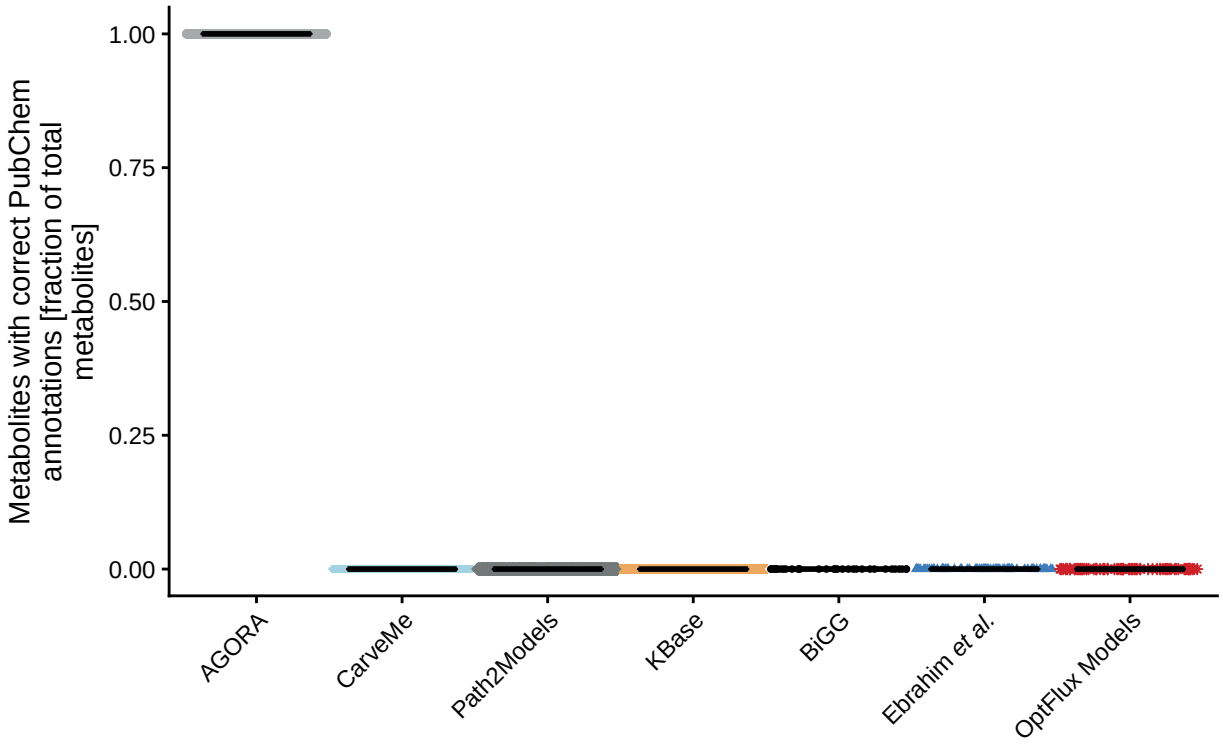

Figure S25: Correct Metabolite Pubchem.compound Annotation

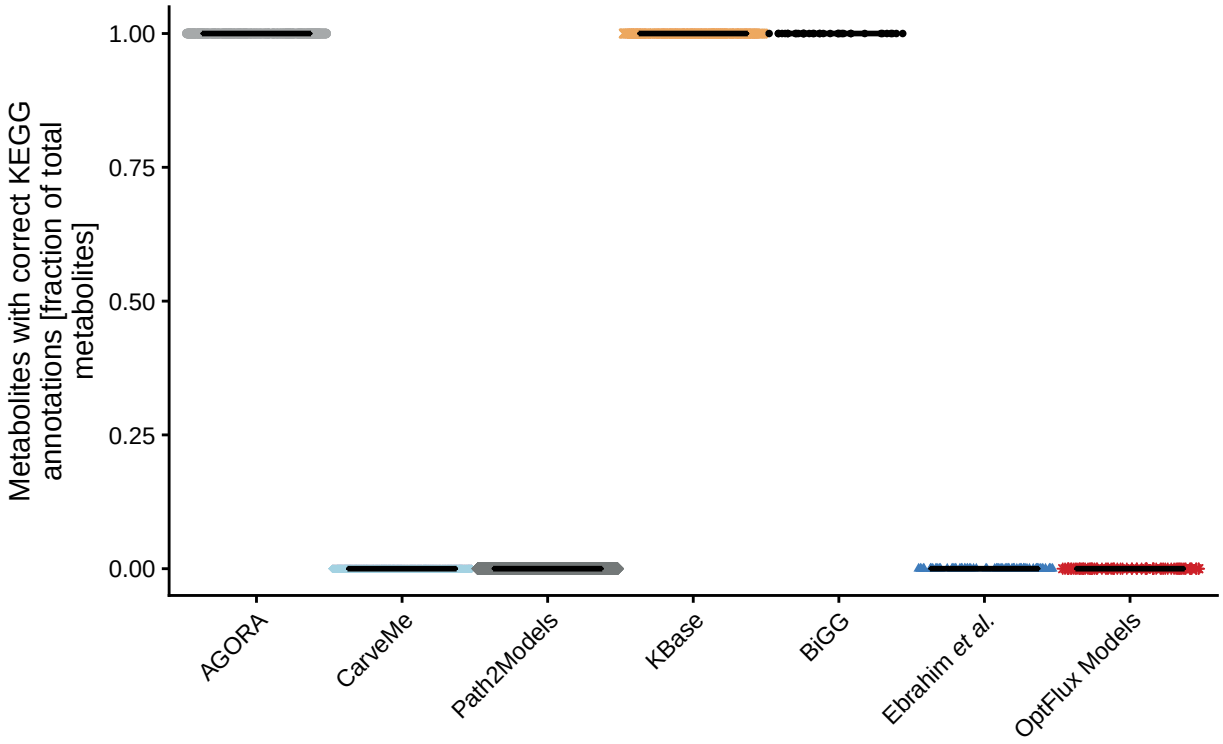

Figure S26: Correct Metabolite KEGG.compound Annotation

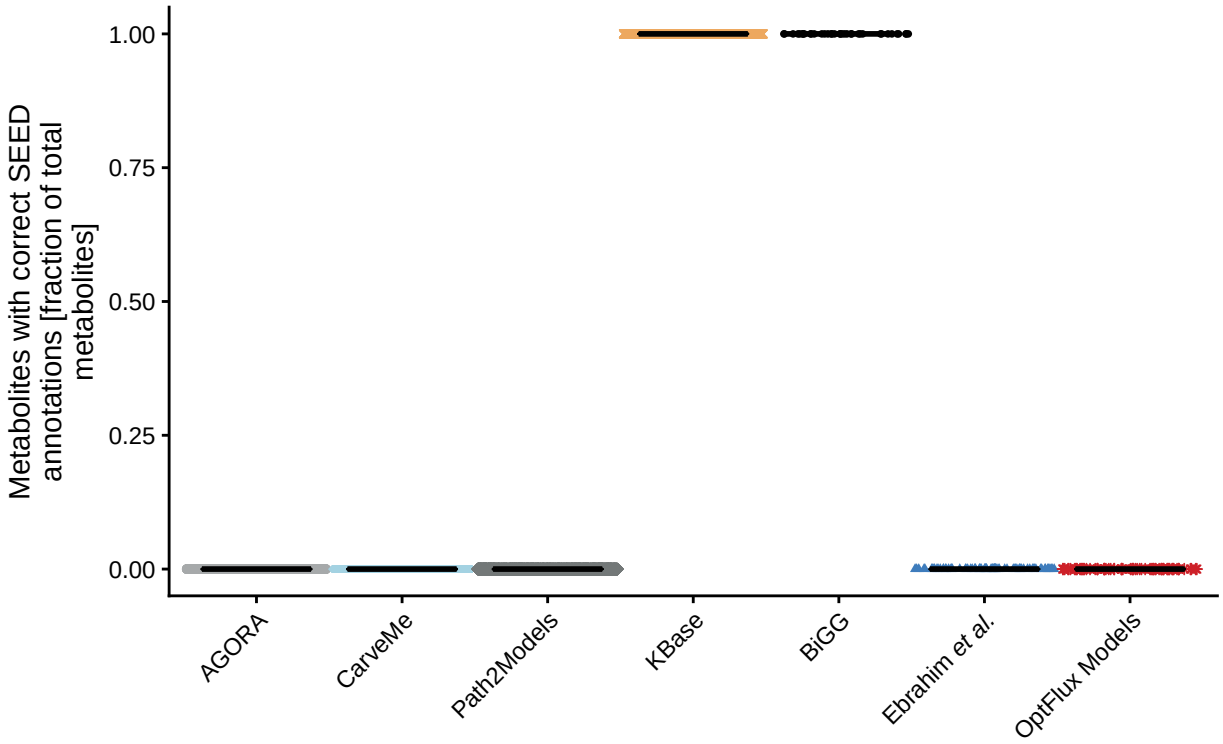

Figure S27: Correct Metabolite SEED.compound Annotation

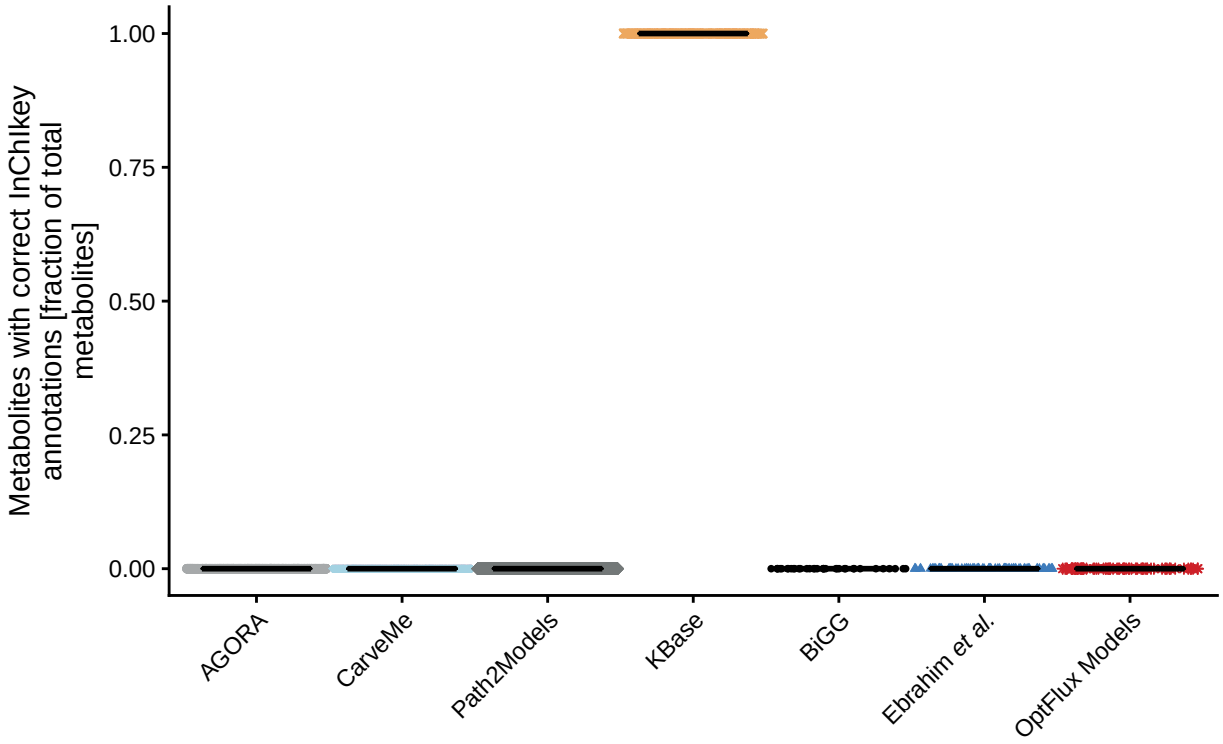

Figure S28: Correct Metabolite InChIKey Annotation

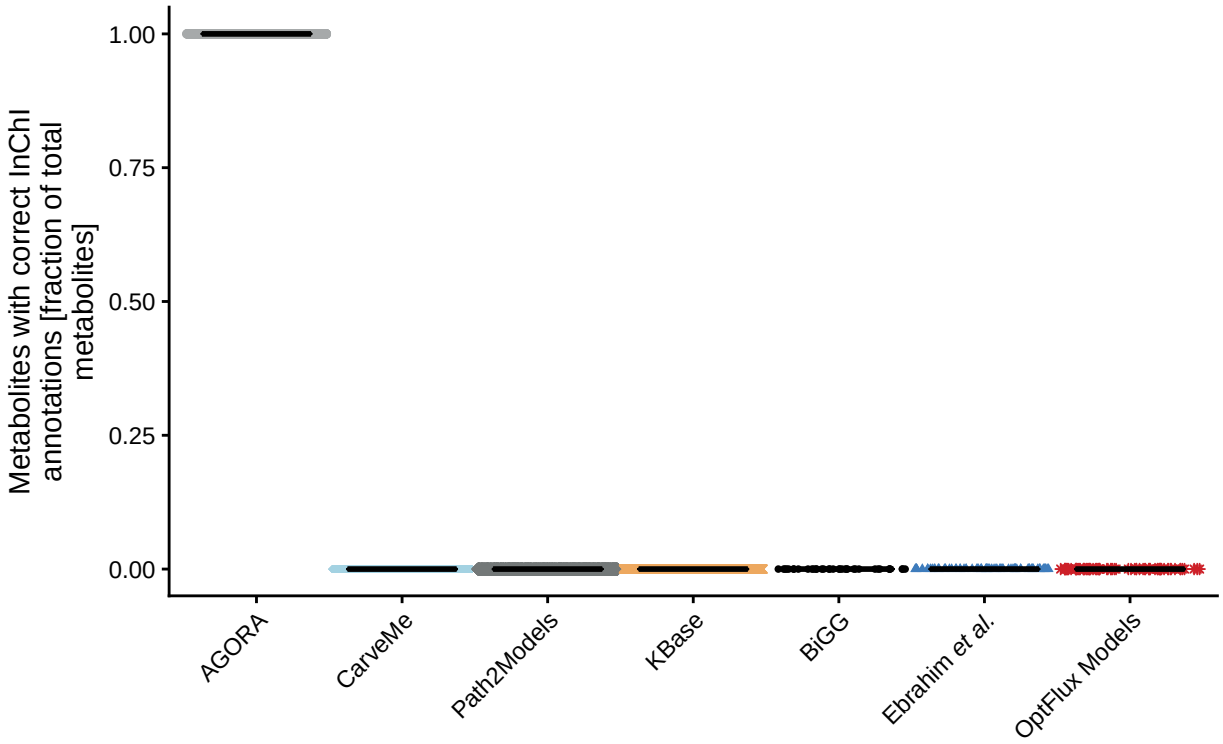

Figure S29: Correct Metabolite InChI Annotation

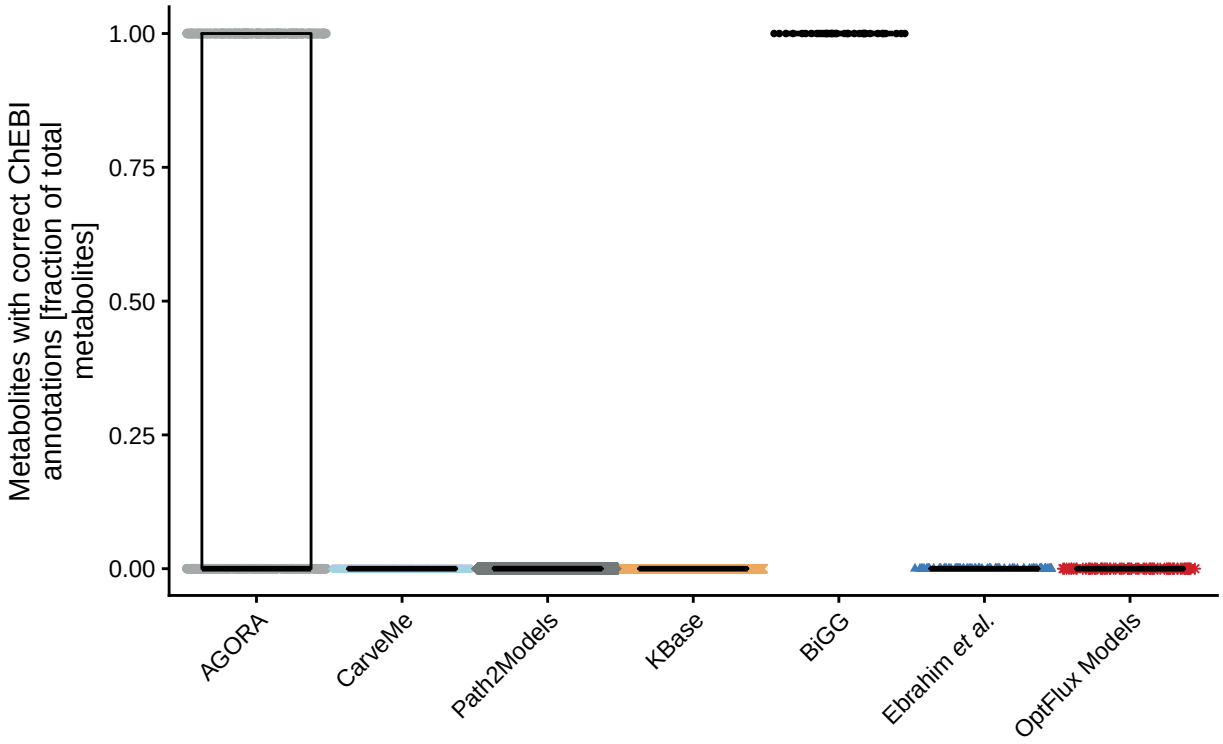

Figure S30: Correct Metabolite ChEBI Annotation

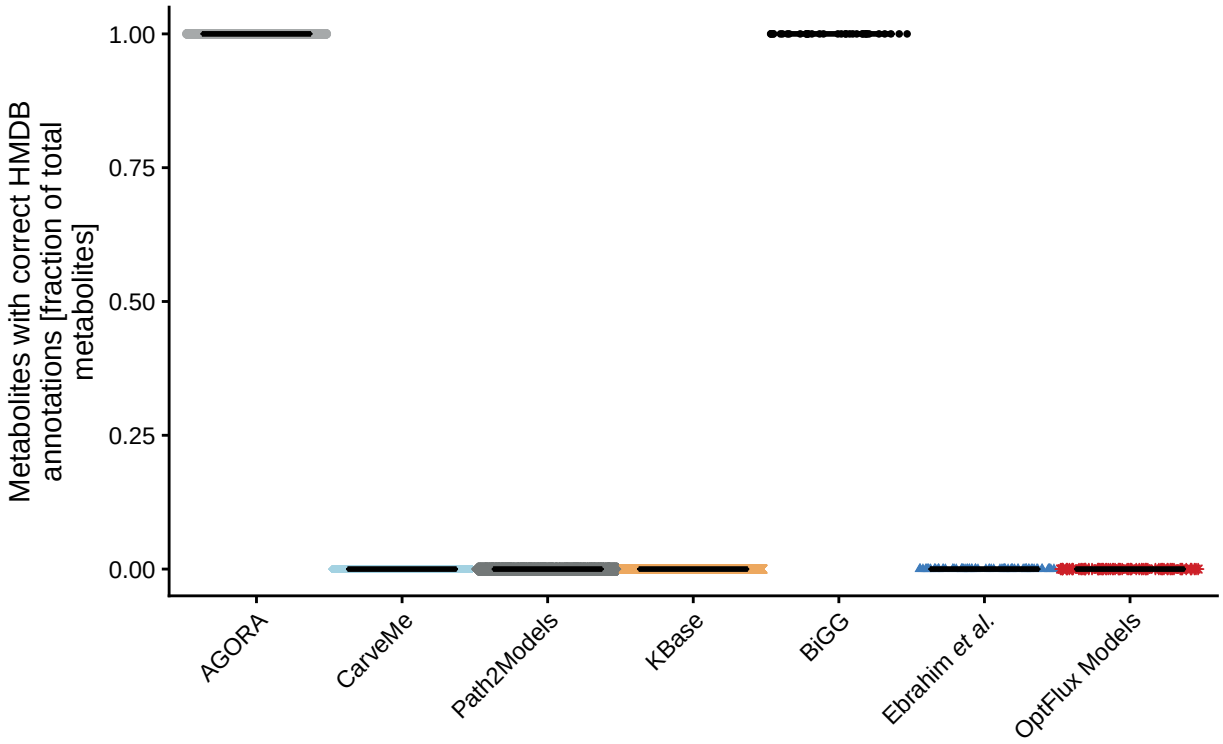

Figure S31: Correct Metabolite HMDB Annotation

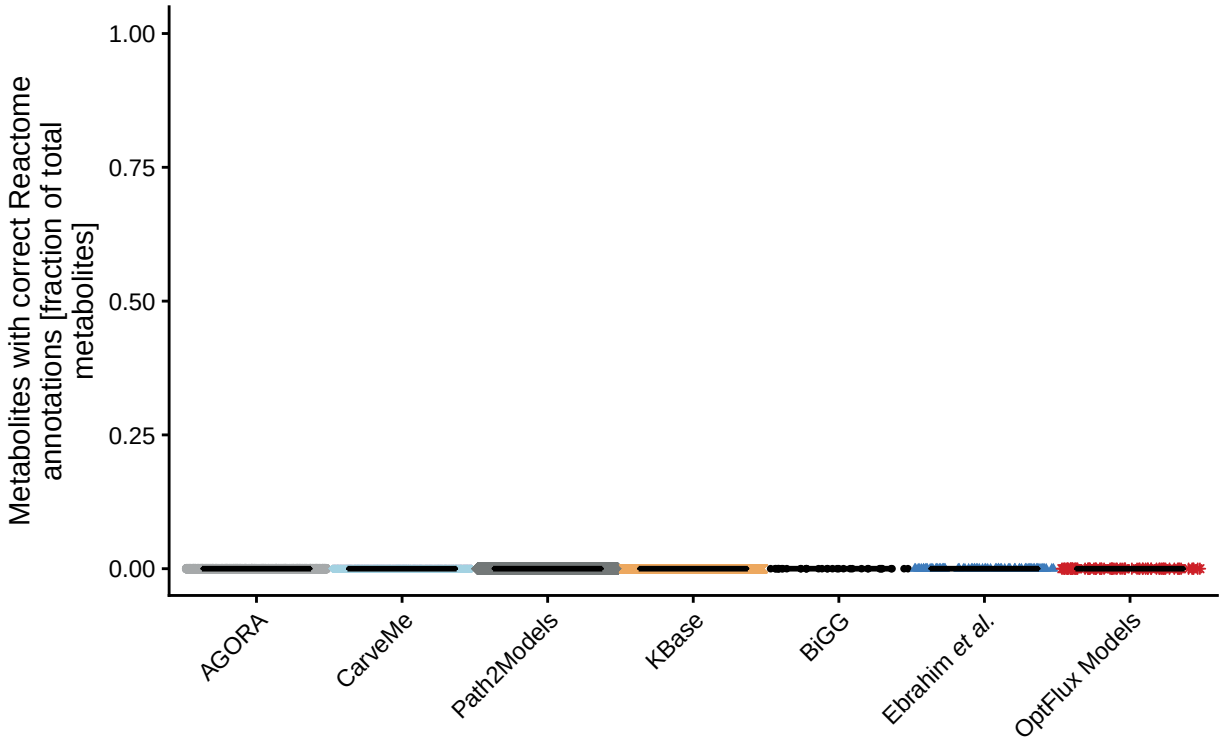

Figure S32: Correct Metabolite Reactome Annotation

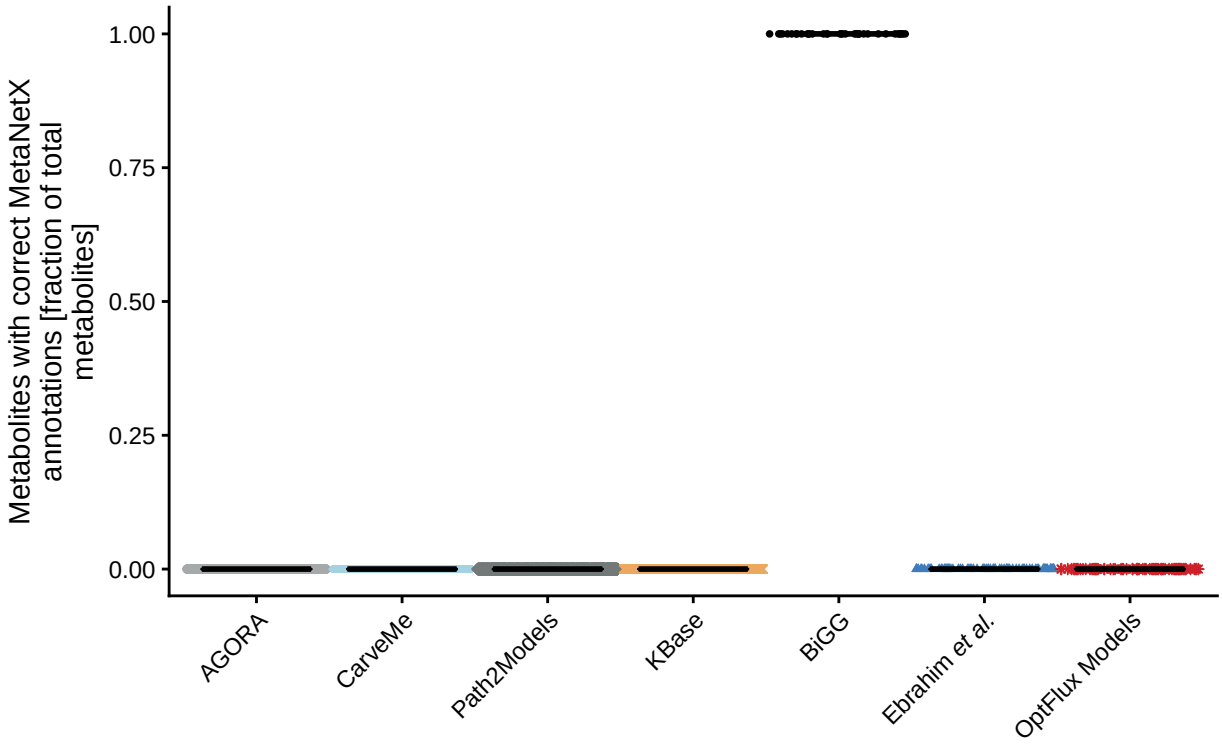

Figure S33: Correct Metabolite MetaNetX.chemical Annotation

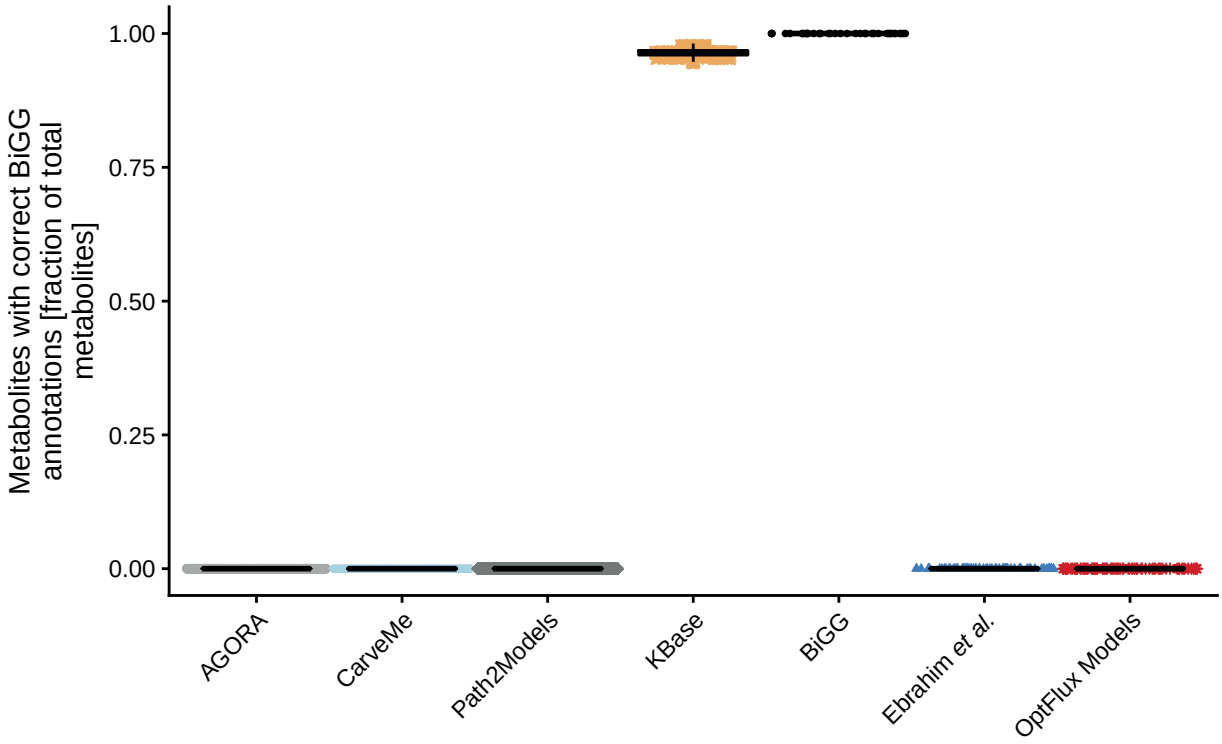

Figure S34: Correct Metabolite BiGG.metabolite Annotation

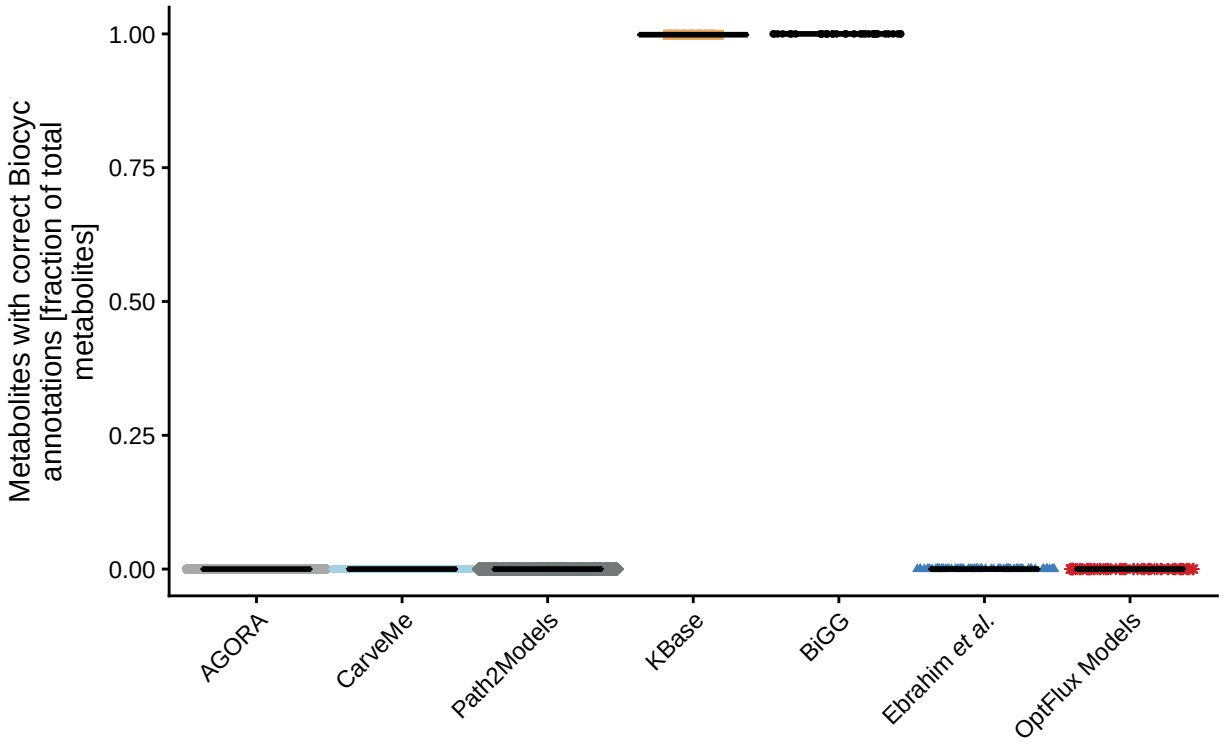

Figure S35: Correct Metabolite BioCyc Annotation

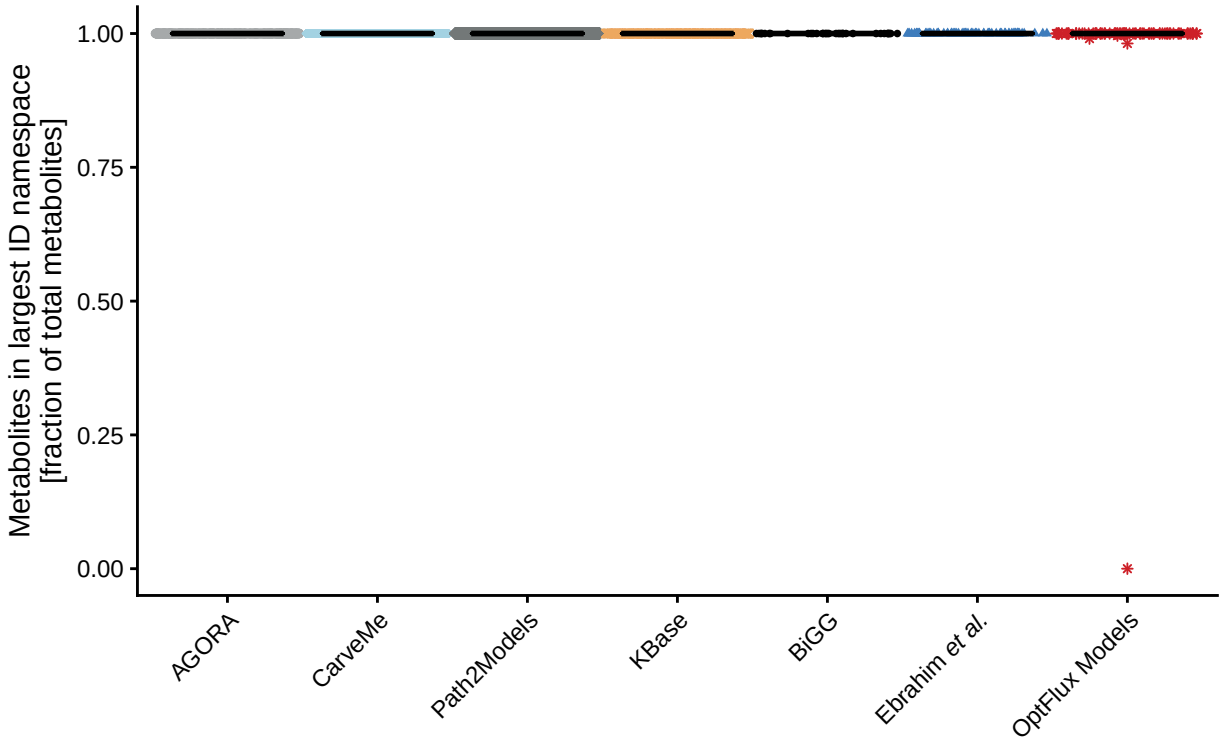

Figure S36: Uniform Metabolite Identifier Namespace

3.3.3 Annotation - Reactions

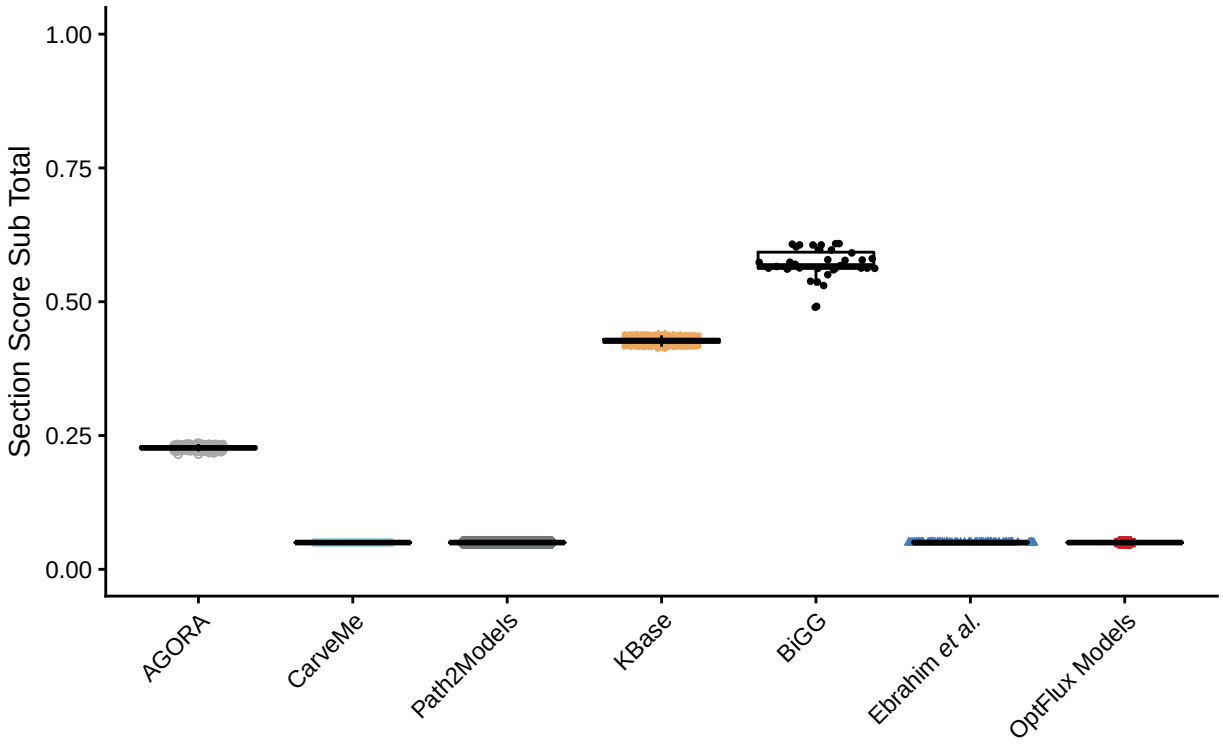

Figure S37: Annotation - Reactions. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

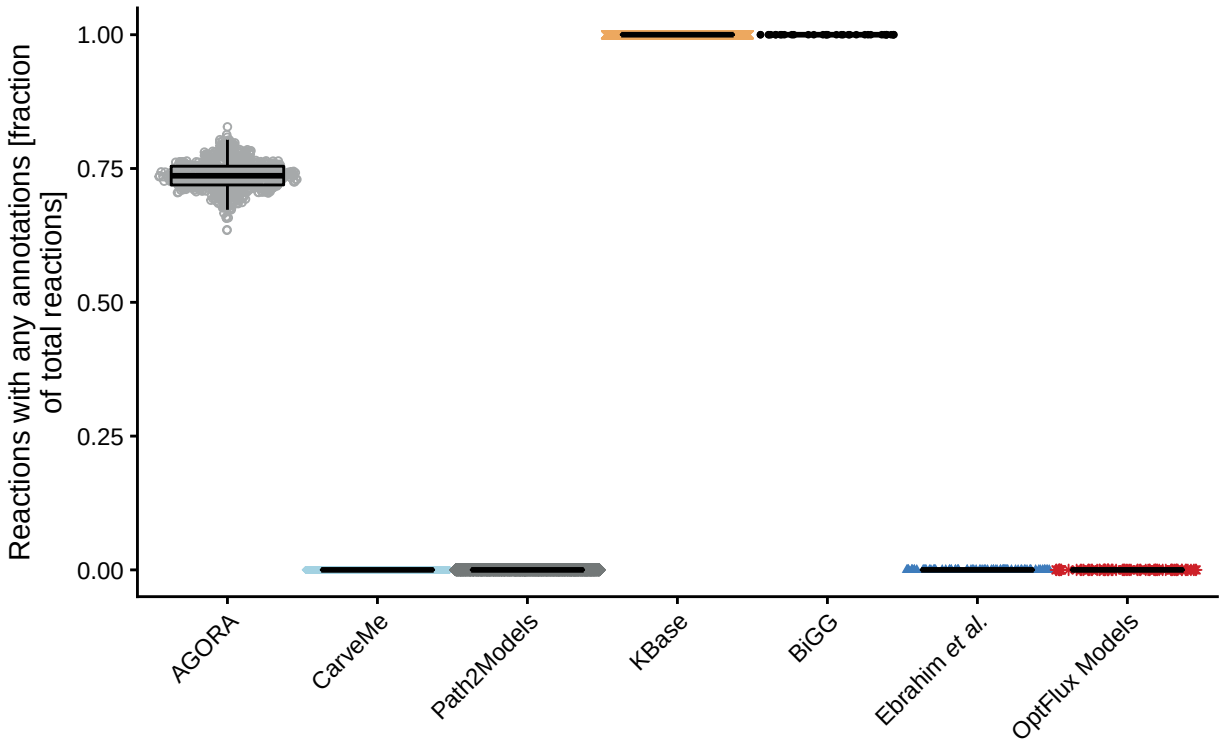

Figure S38: Presence of Reaction Annotation

3.3.3.1 Reaction Annotations Per Database

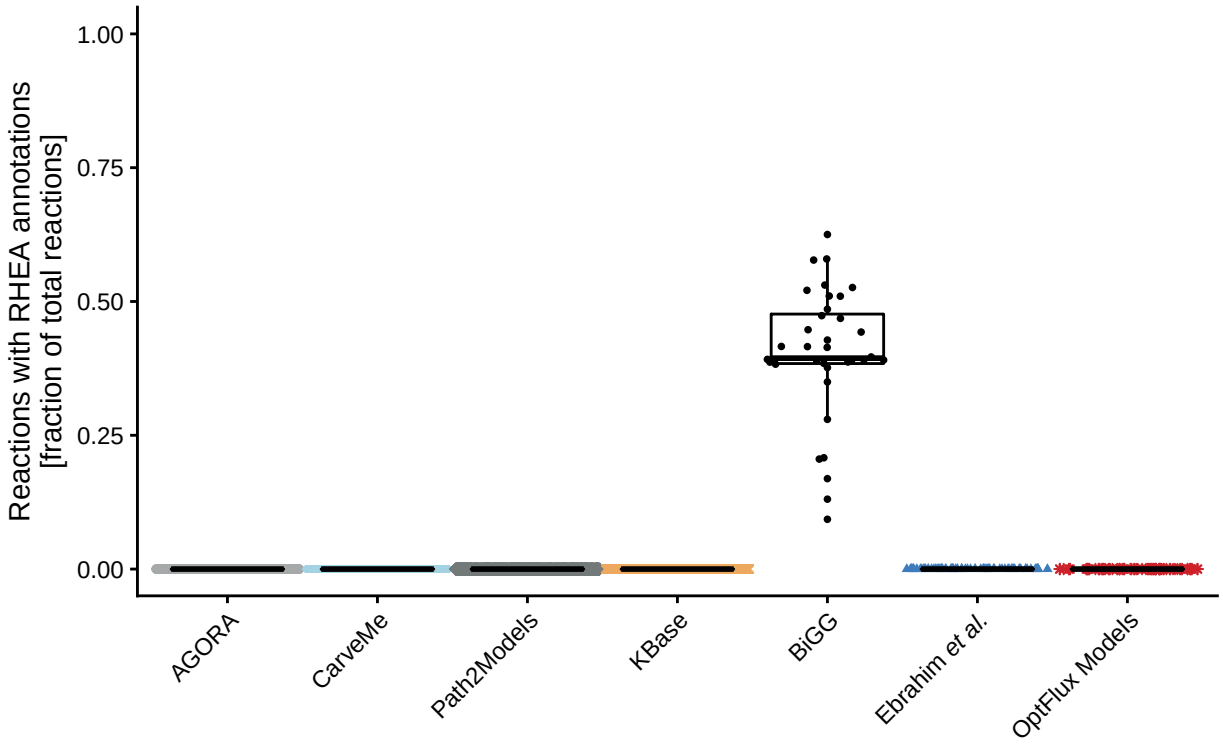

Figure S39: Reaction Rhea Annotation

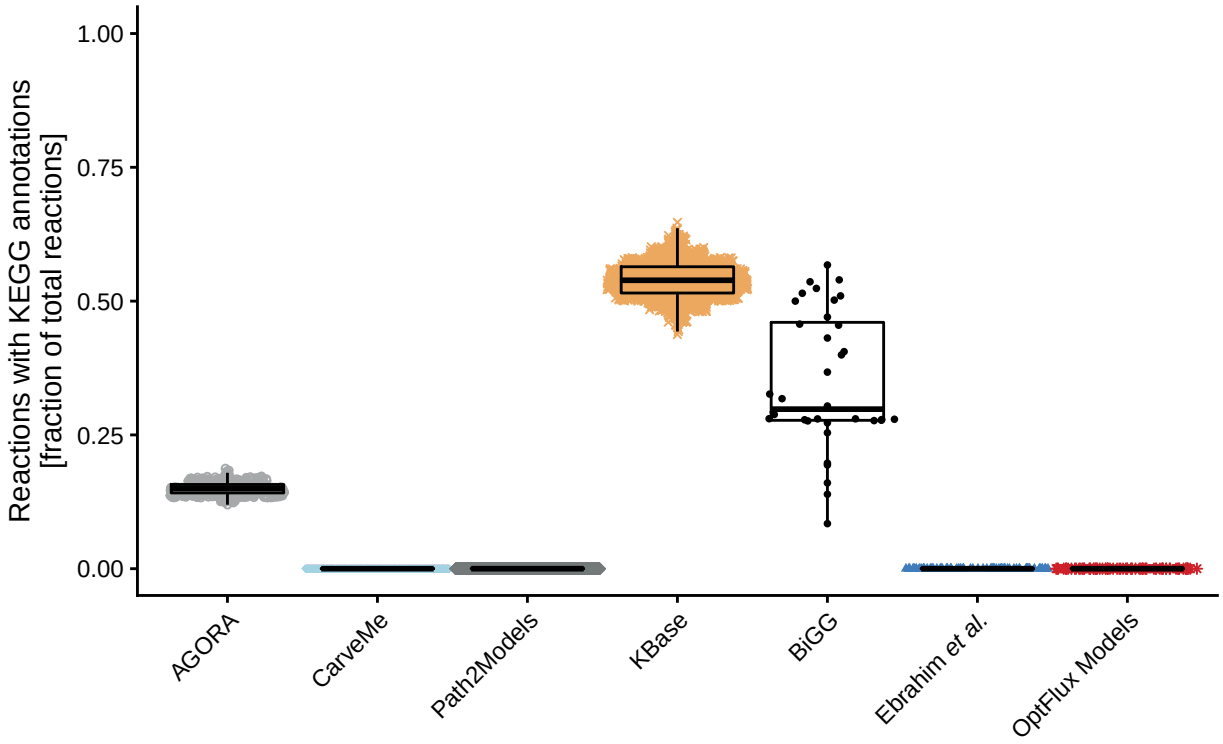

Figure S40: Reaction KEGG.reaction Annotation

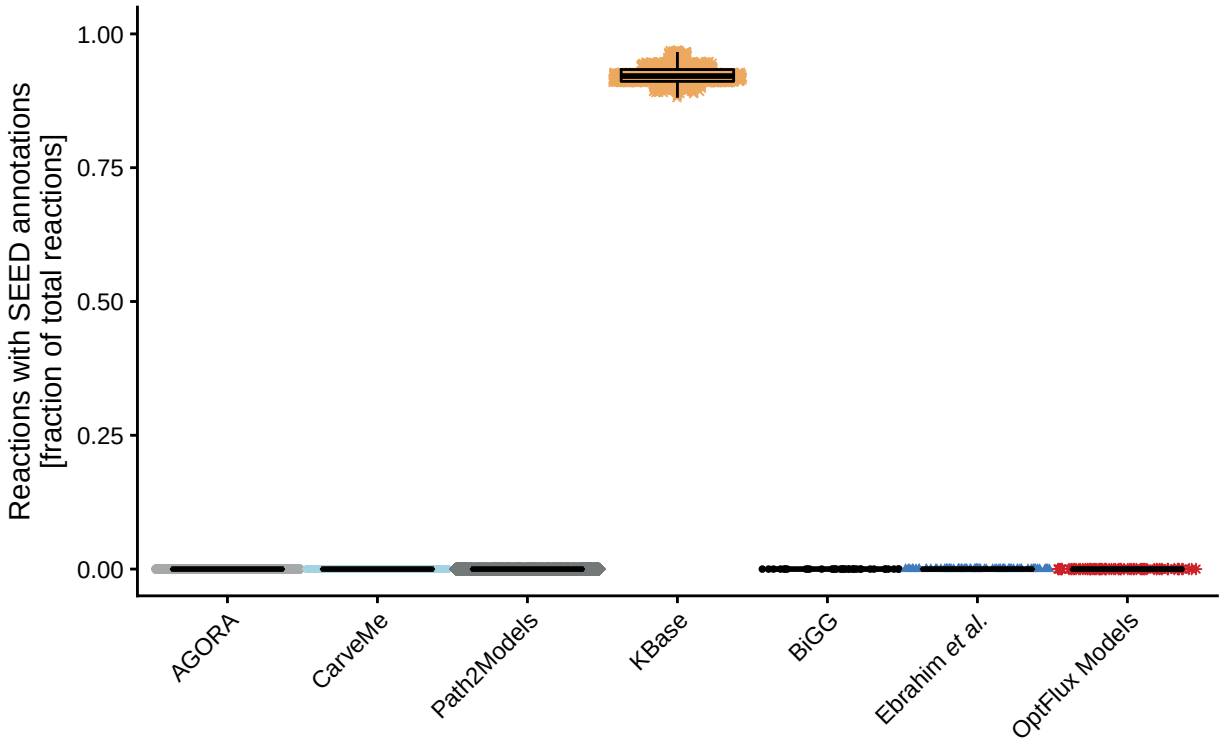

Figure S41: Reaction SEED.reaction Annotation

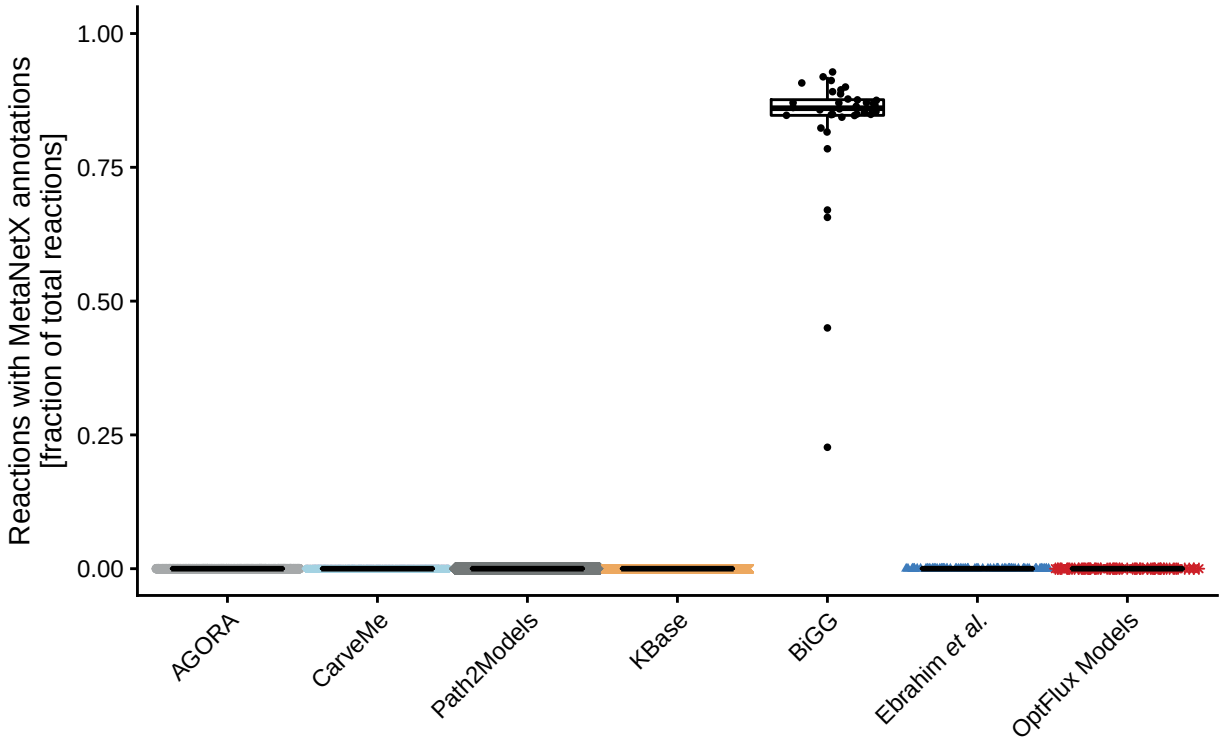

Figure S42: Reaction MetaNetX.reaction Annotation

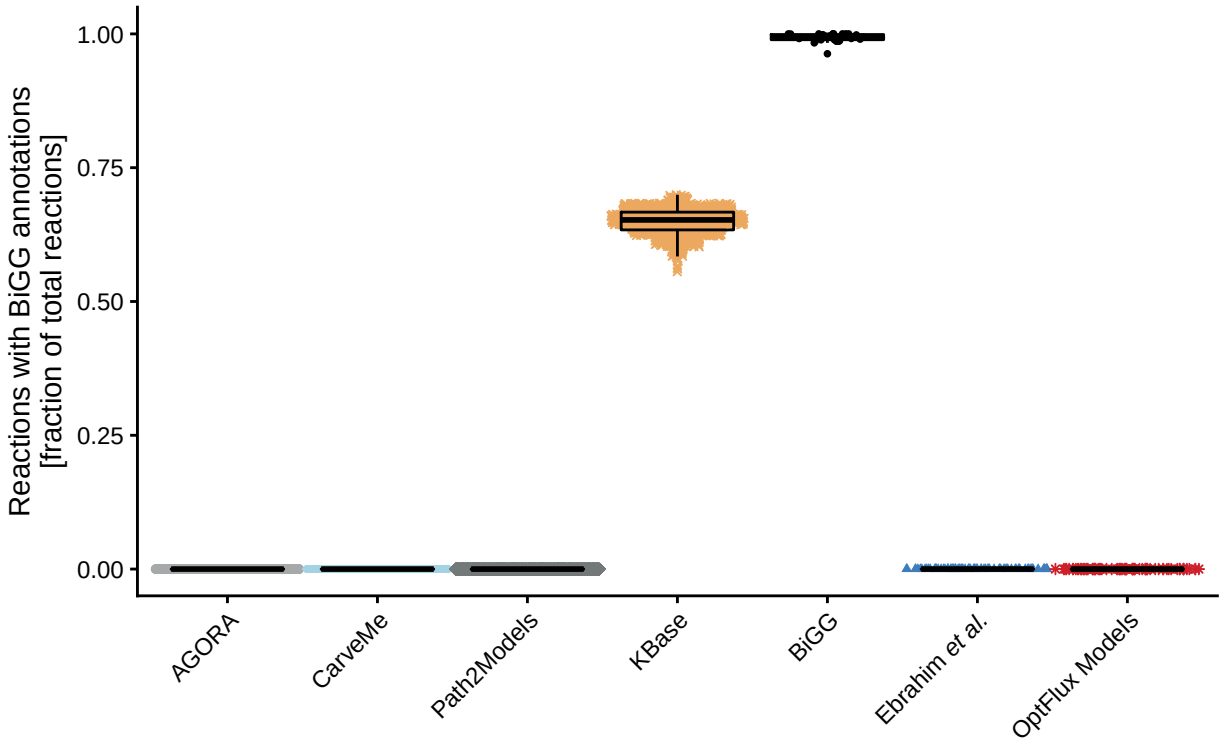

Figure S43: Reaction BiGG.reaction Annotation

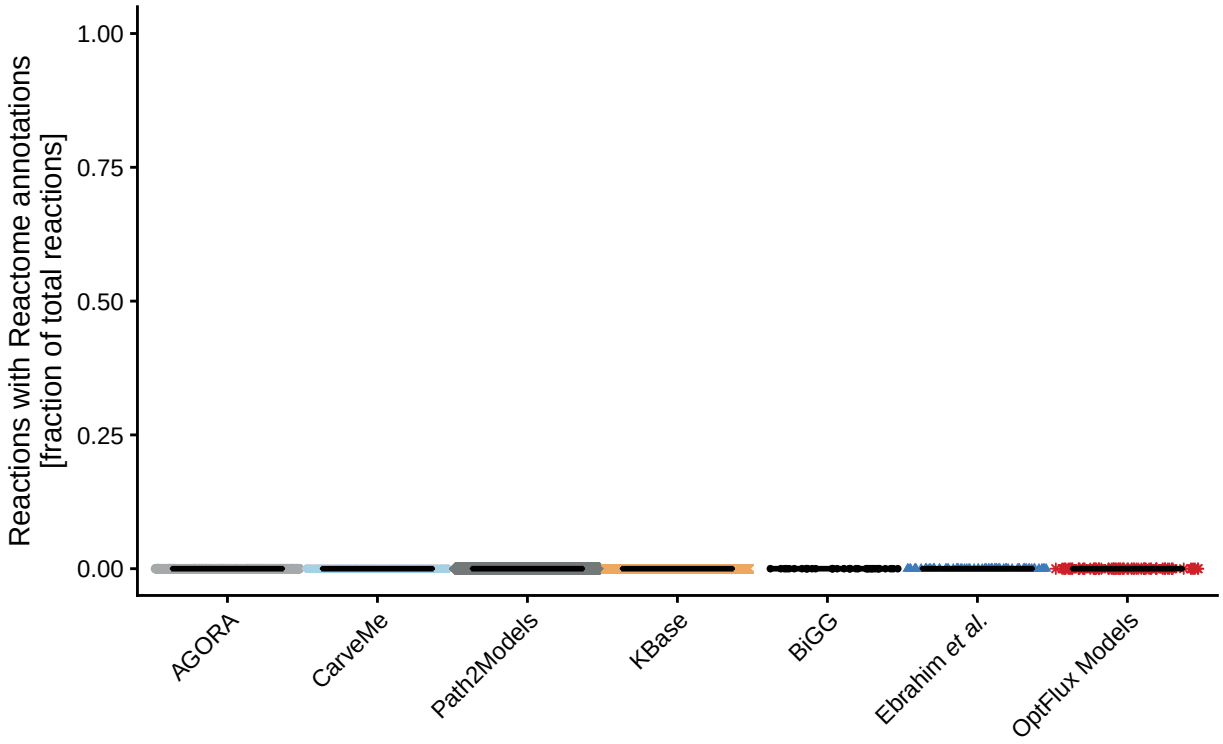

Figure S44: Reaction Reactome Annotation

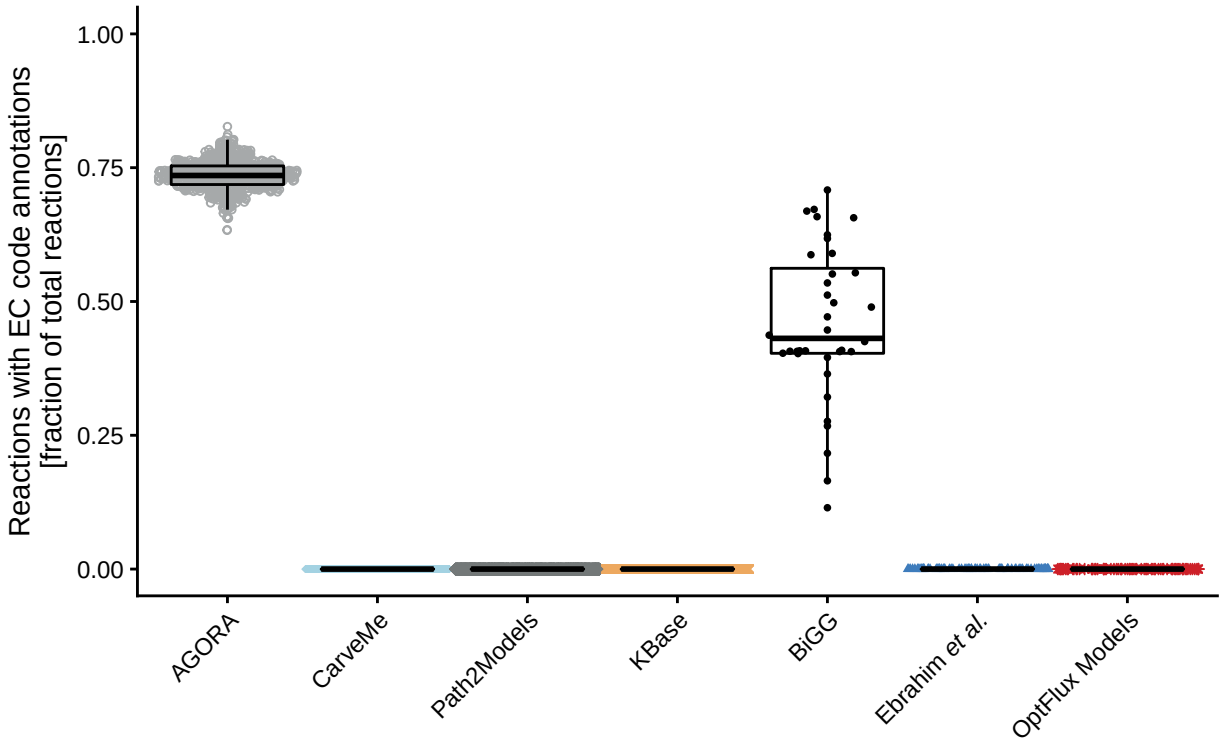

Figure S45: Reaction Enzyme Classification Annotation

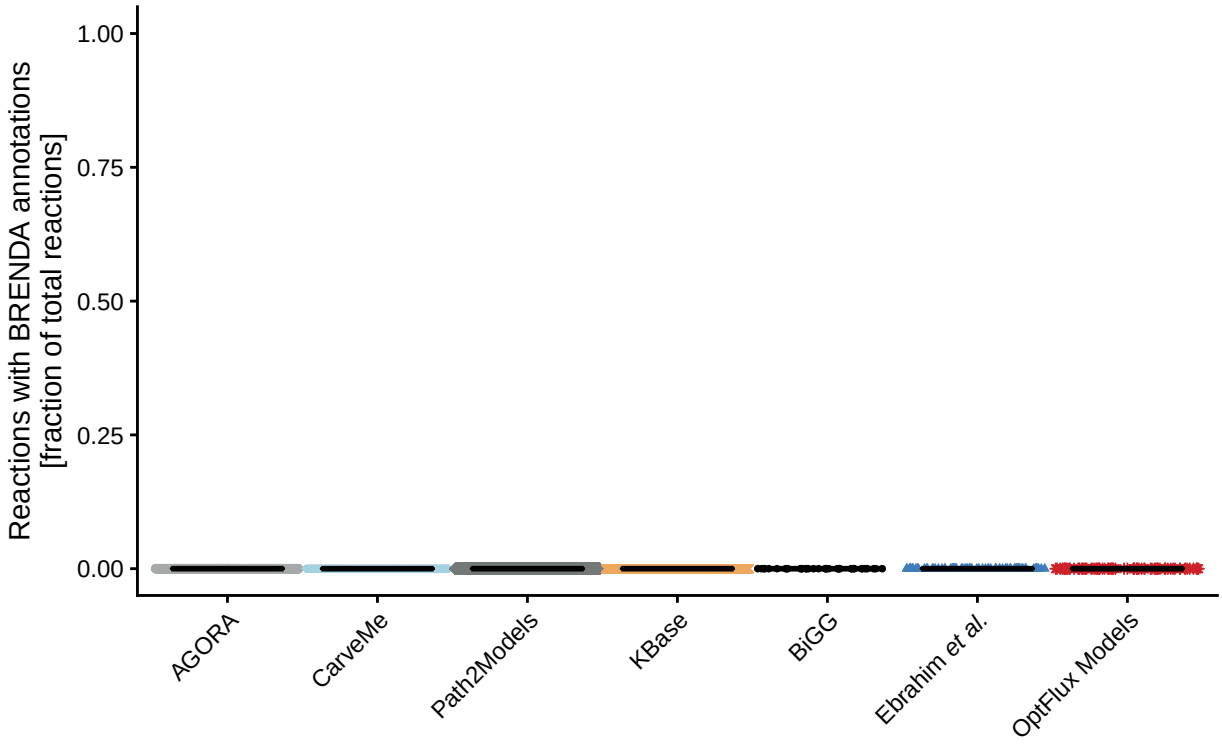

Figure S46: Reaction BRENDA Annotation

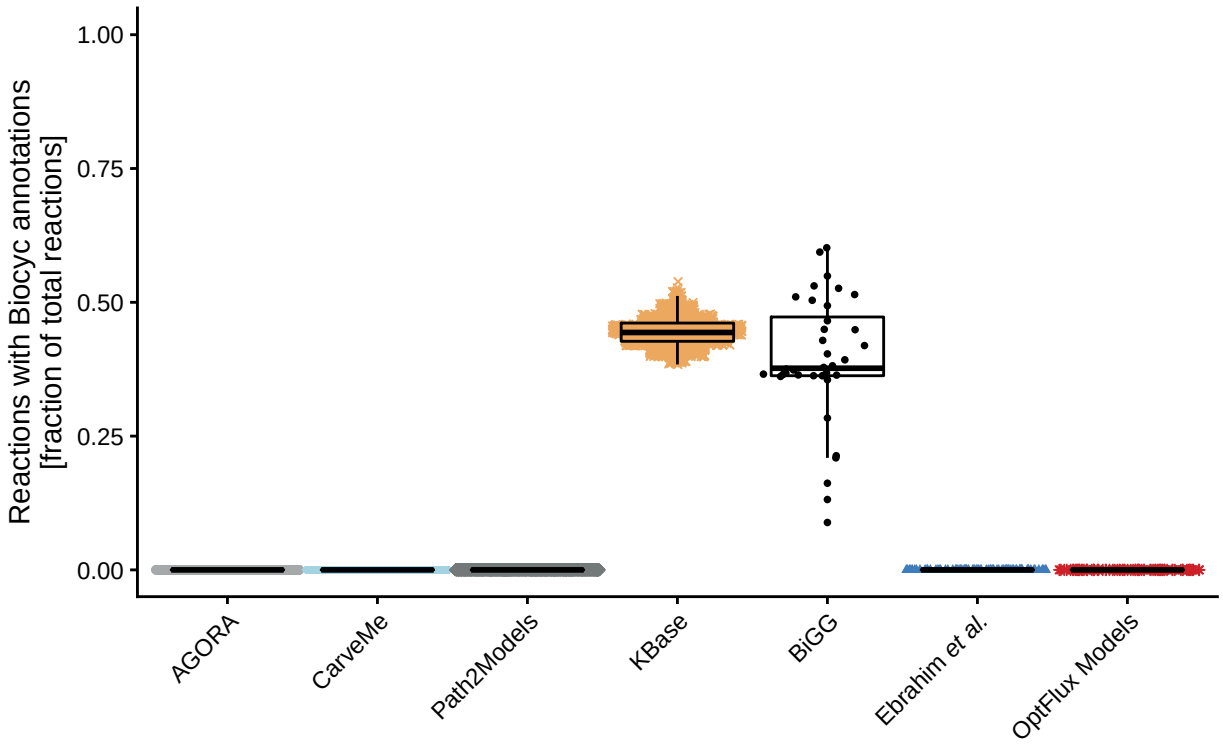

Figure S47: Reaction BioCyc Annotation

3.3.3.2 Reaction Annotation Conformity Per Database

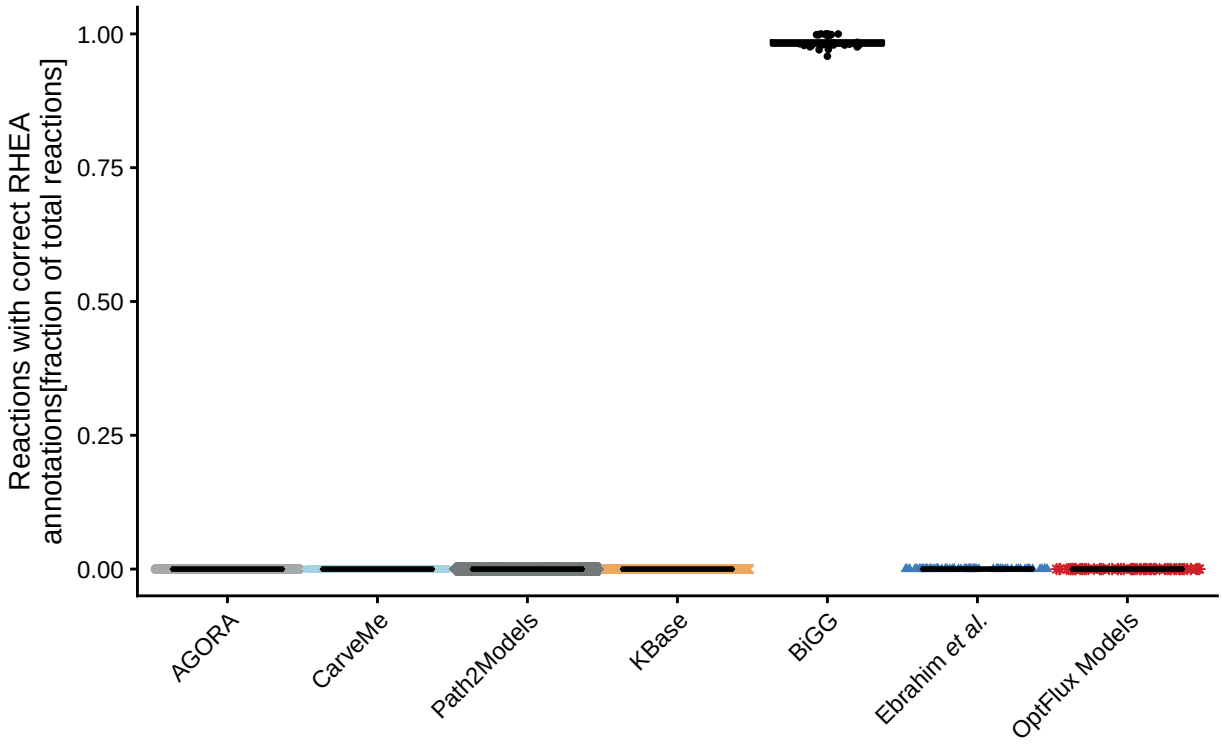

Figure S48: Correct Reaction Rhea Annotation

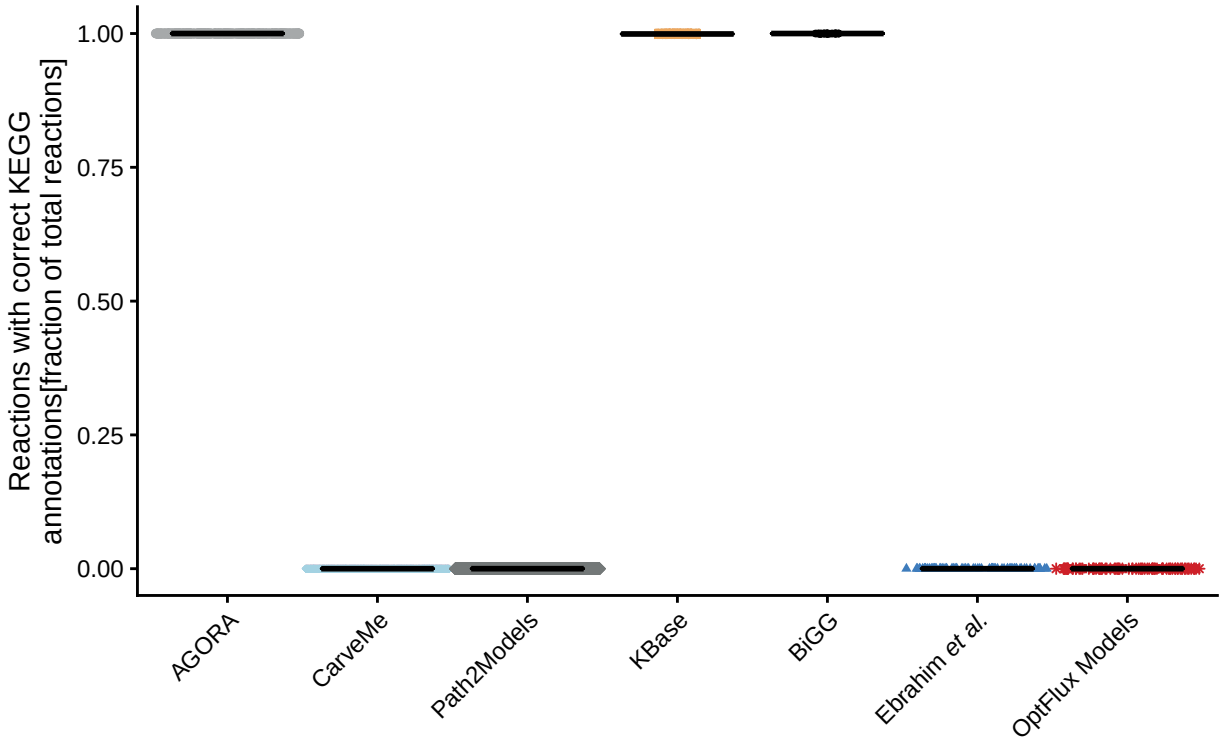

Figure S49: Correct Reaction KEGG.reaction Annotation

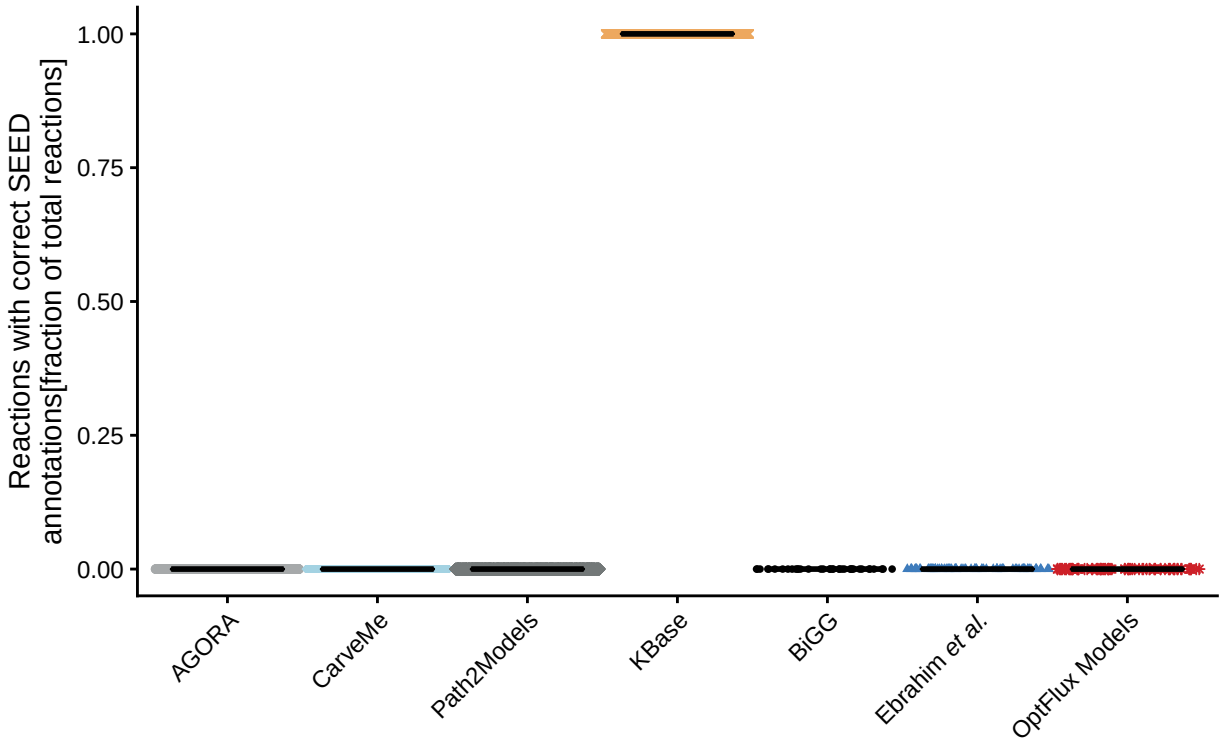

Figure S50: Correct Reaction SEED.reaction Annotation

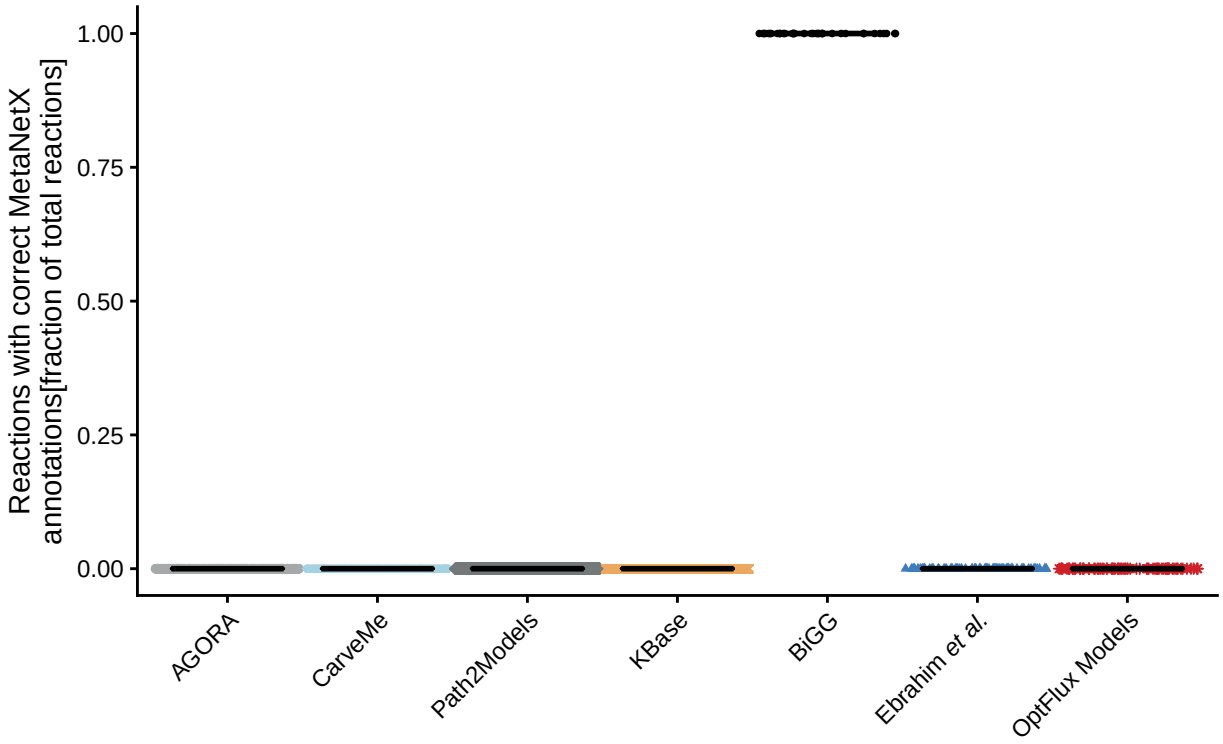

Figure S51: Correct Reaction MetaNetX.reaction Annotation

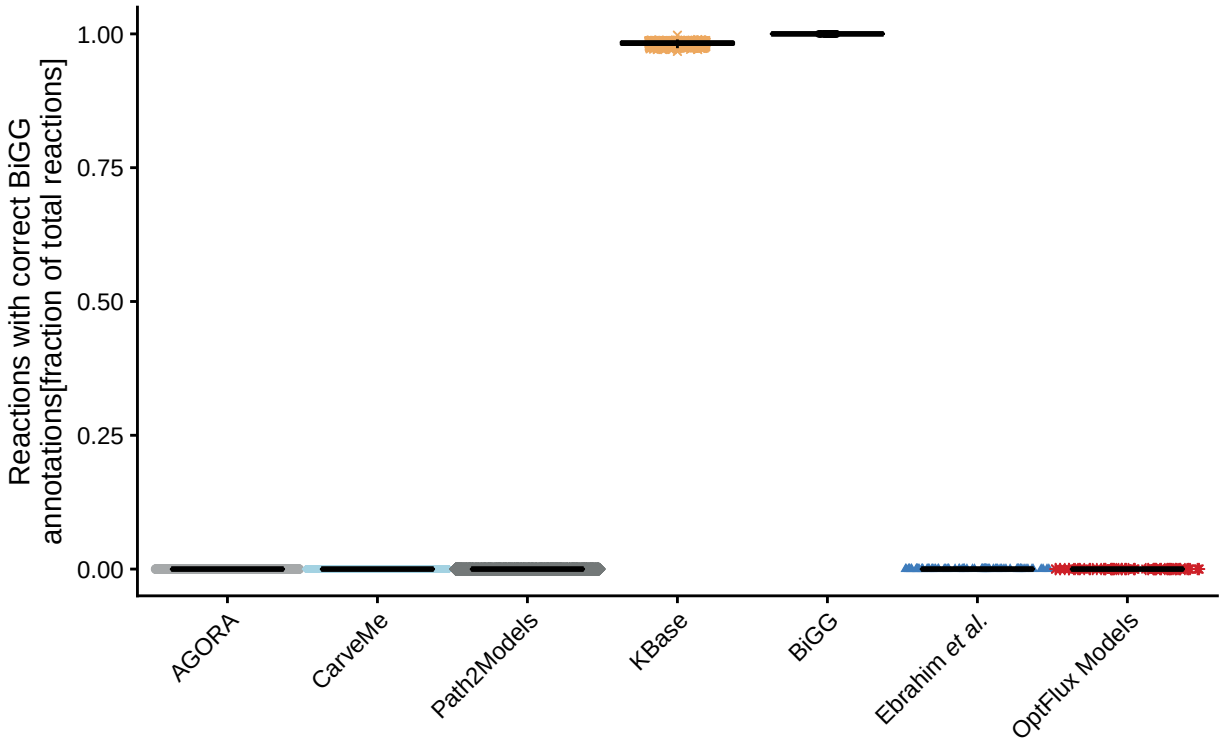

Figure S52: Correct Reaction BiGG.reaction Annotation

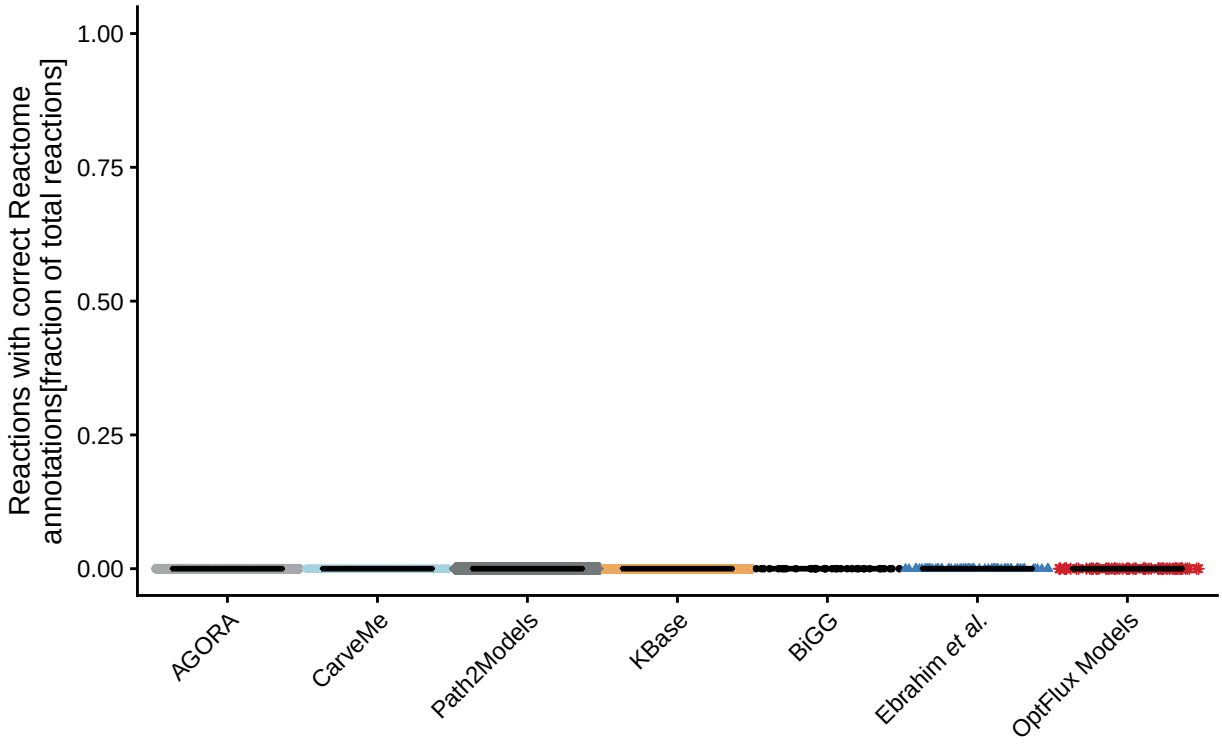

Figure S53: Correct Reaction Reactome Annotation

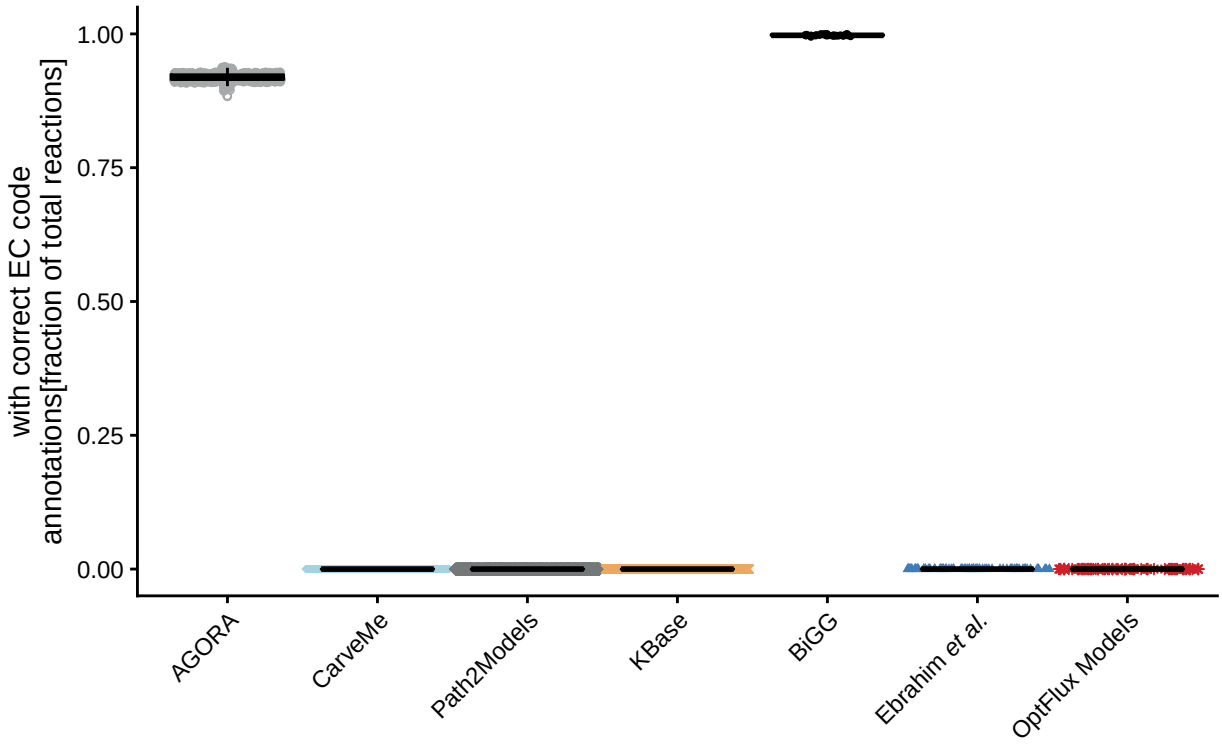

Figure S54: Correct Reaction Enzyme Classification Annotation

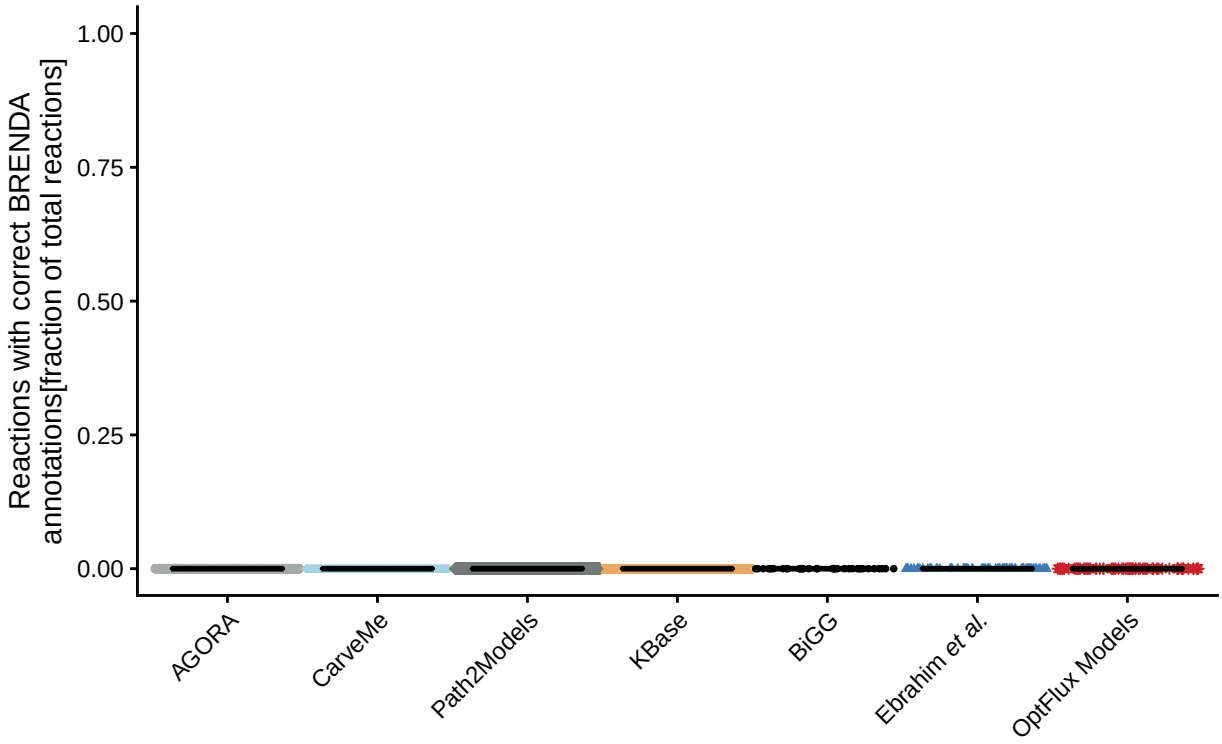

Figure S55: Correct Reaction BRENDA Annotation

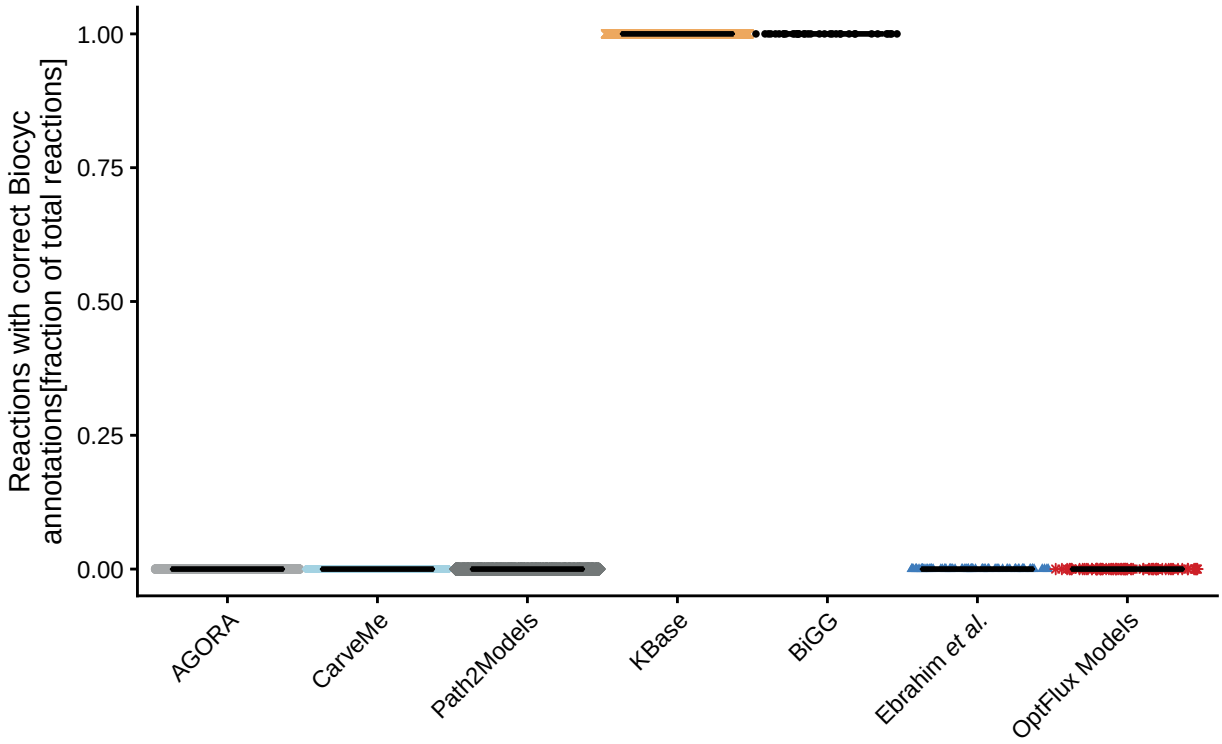

Figure S56: Correct Reaction BioCyc Annotation

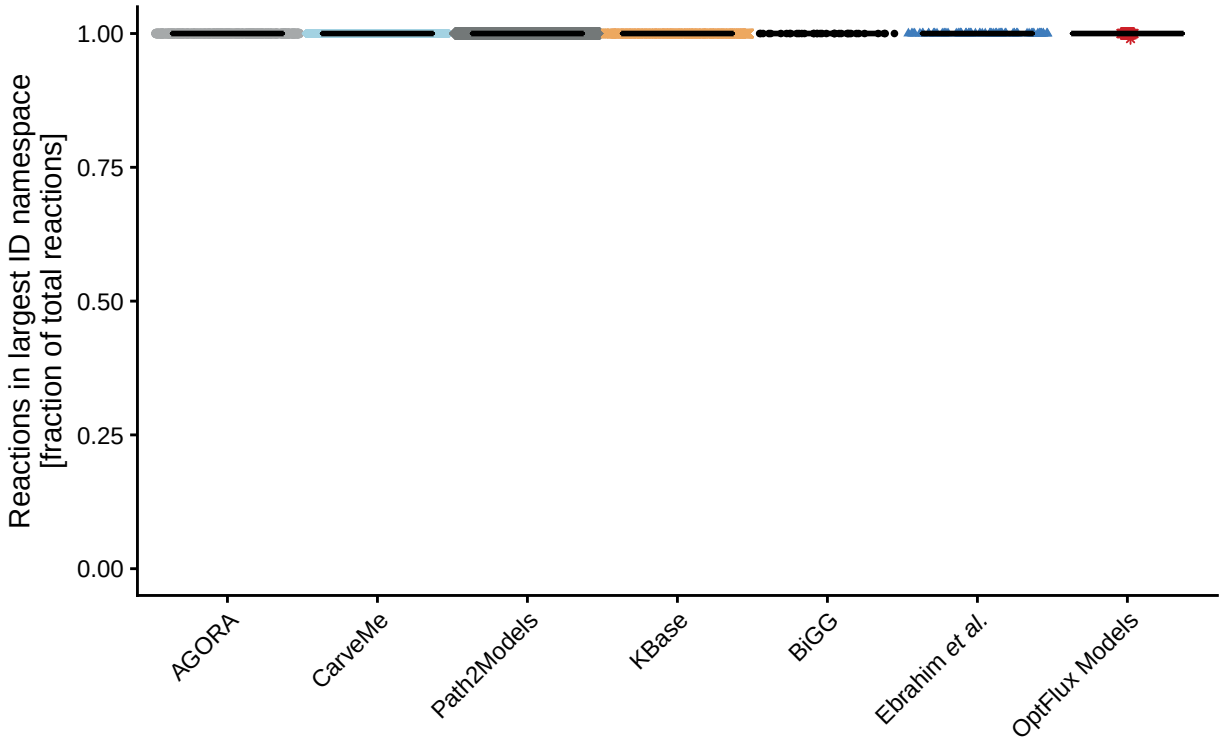

Figure S57: Uniform Reaction Identifier Namespace

3.3.4 Annotation - Genes

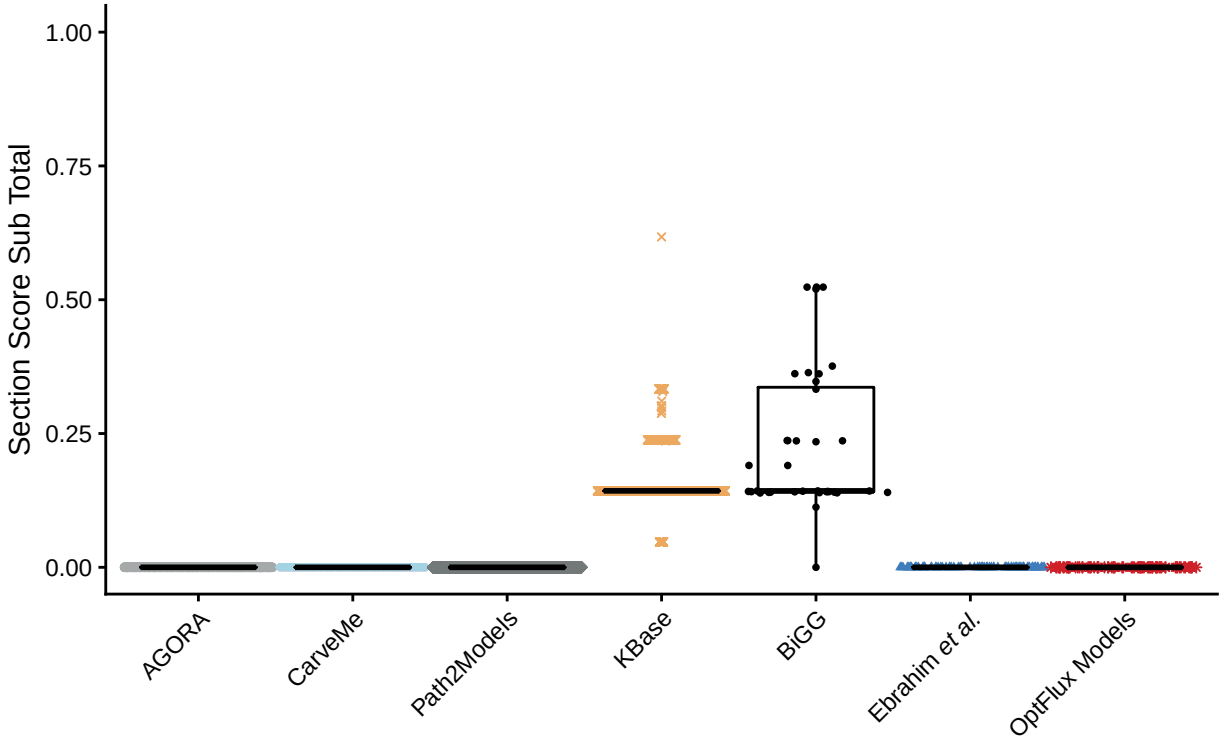

Figure S58: Annotation - Genes. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

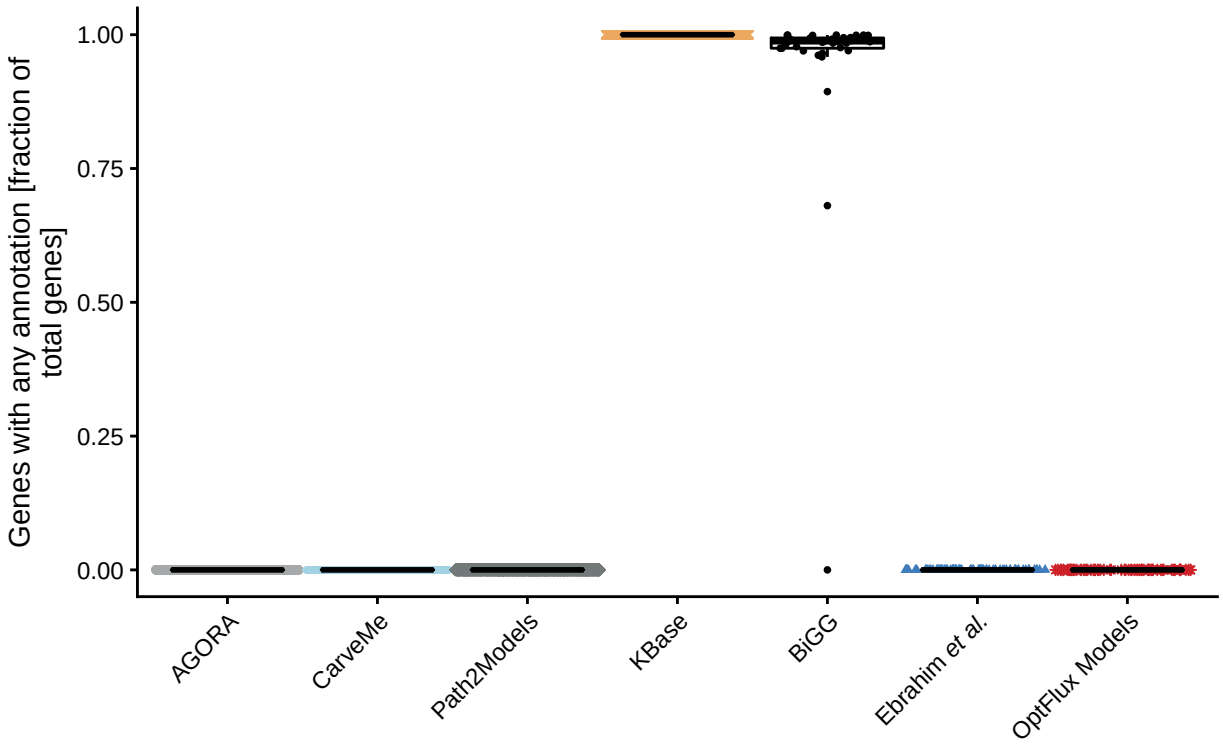

Figure S59: Presence of Gene Annotation

3.3.4.1 Gene Annotations Per Database

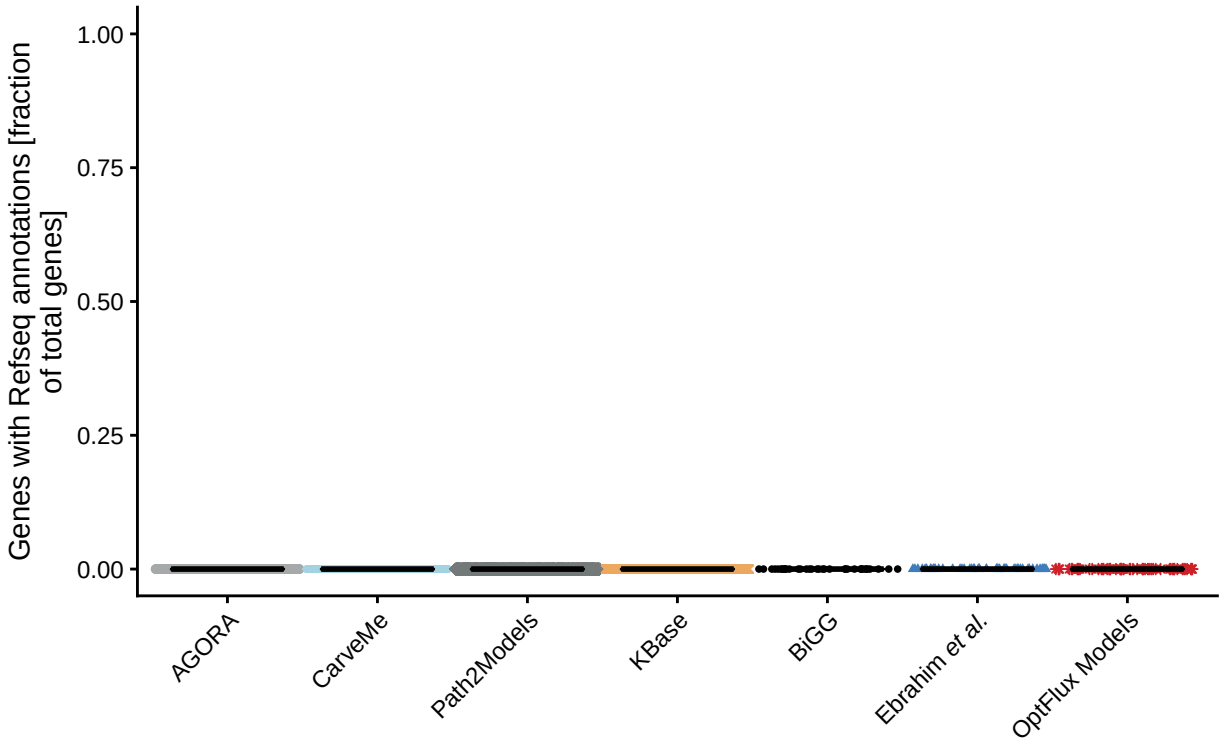

Figure S60: Gene RefSeq Annotation

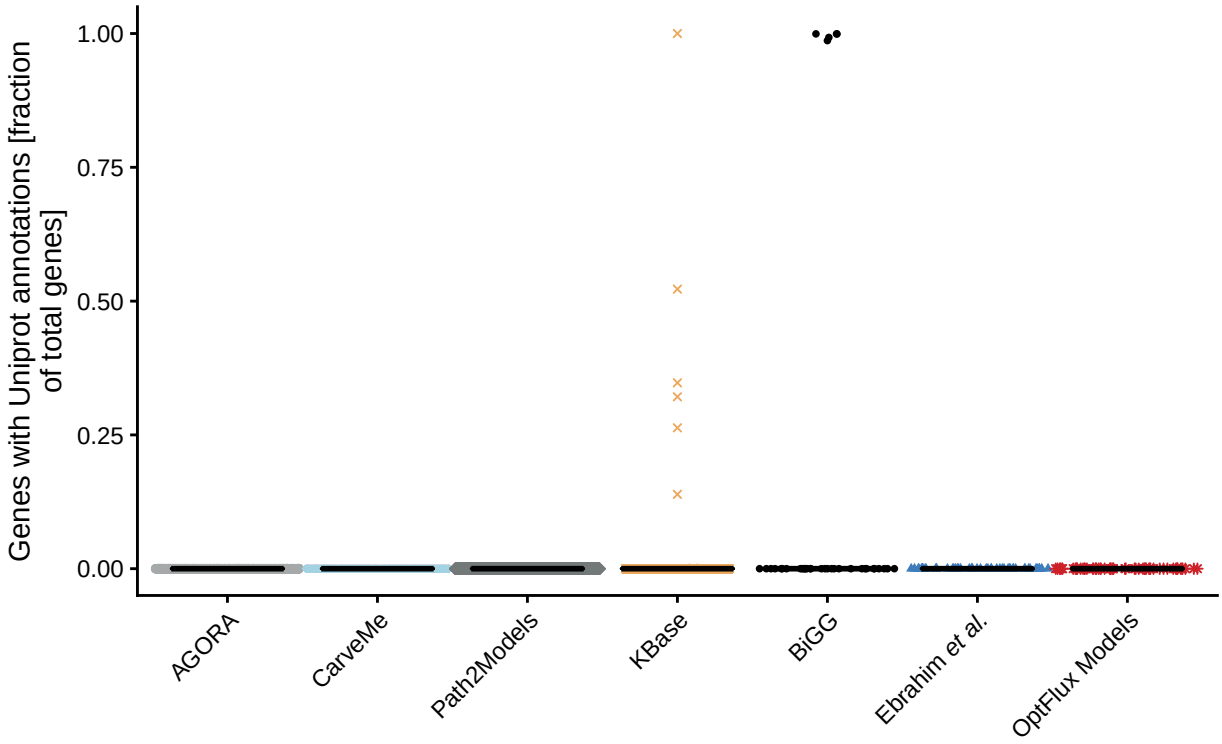

Figure S61: Gene UniProt Annotation

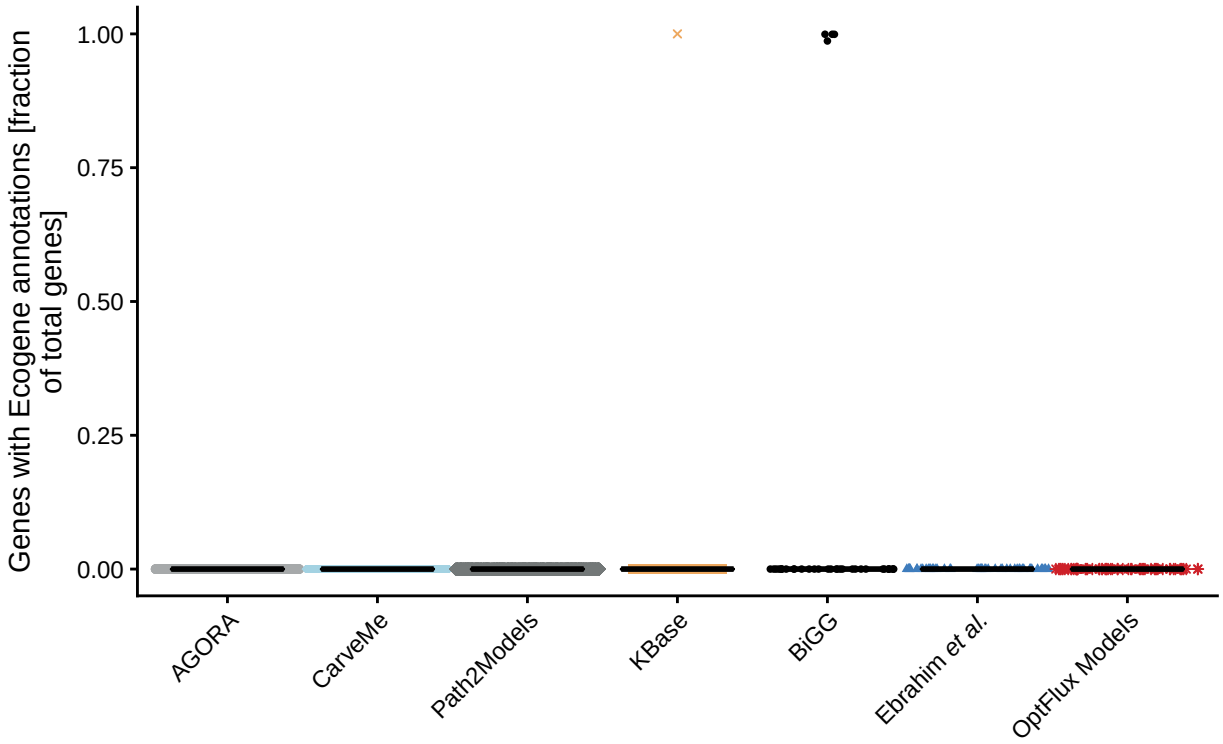

Figure S62: Gene EcoGene Annotation

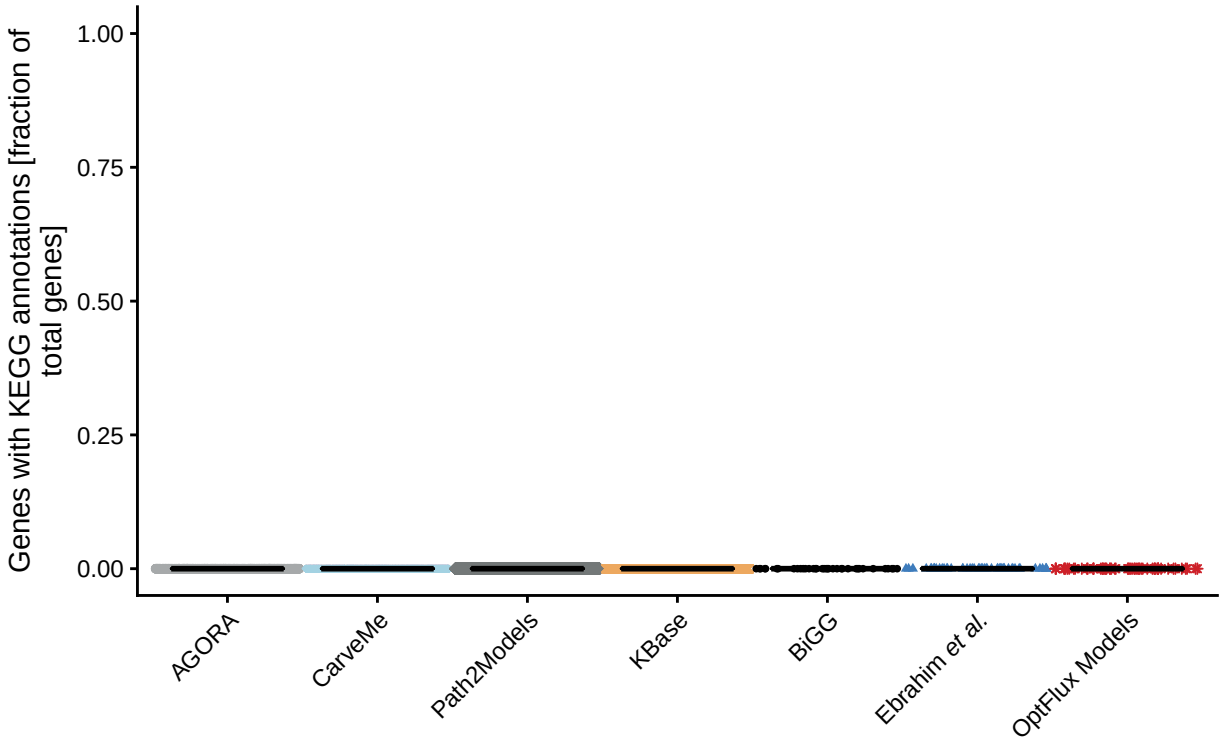

Figure S63: Gene KEGG.genes Annotation

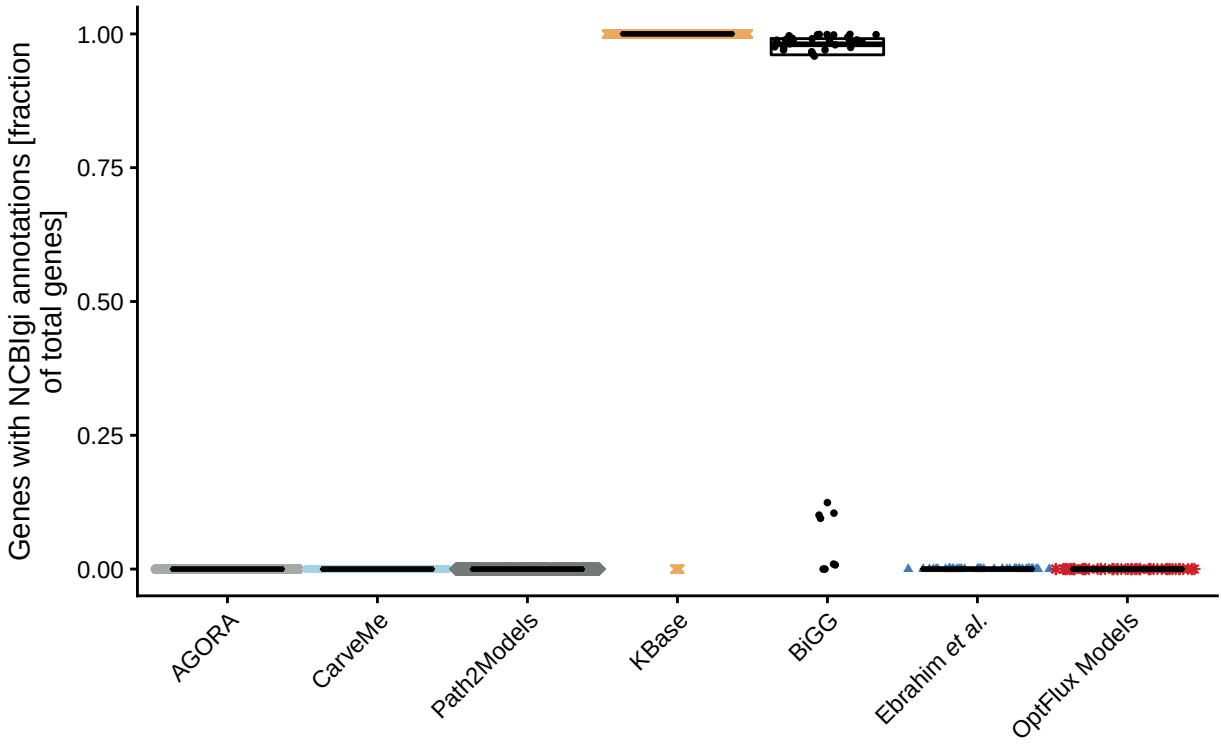

Figure S64: Gene NCBIgi Annotation

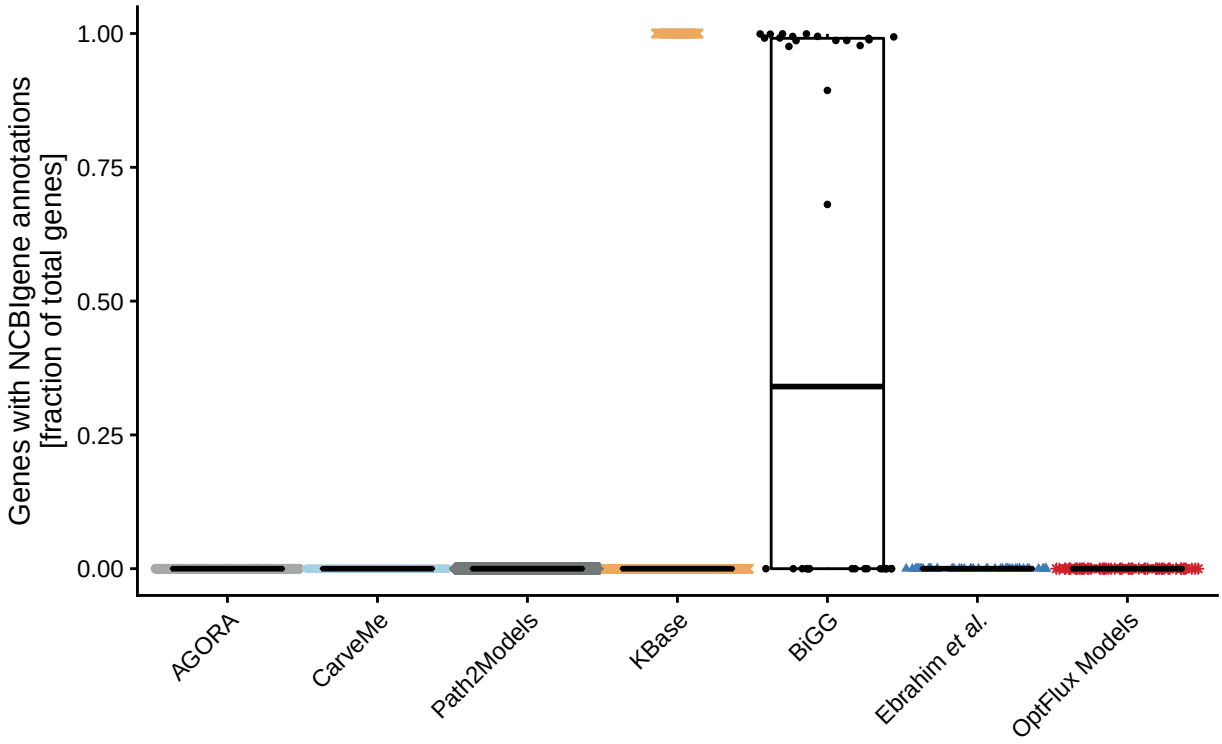

Figure S65: Gene NCBIgene Annotation

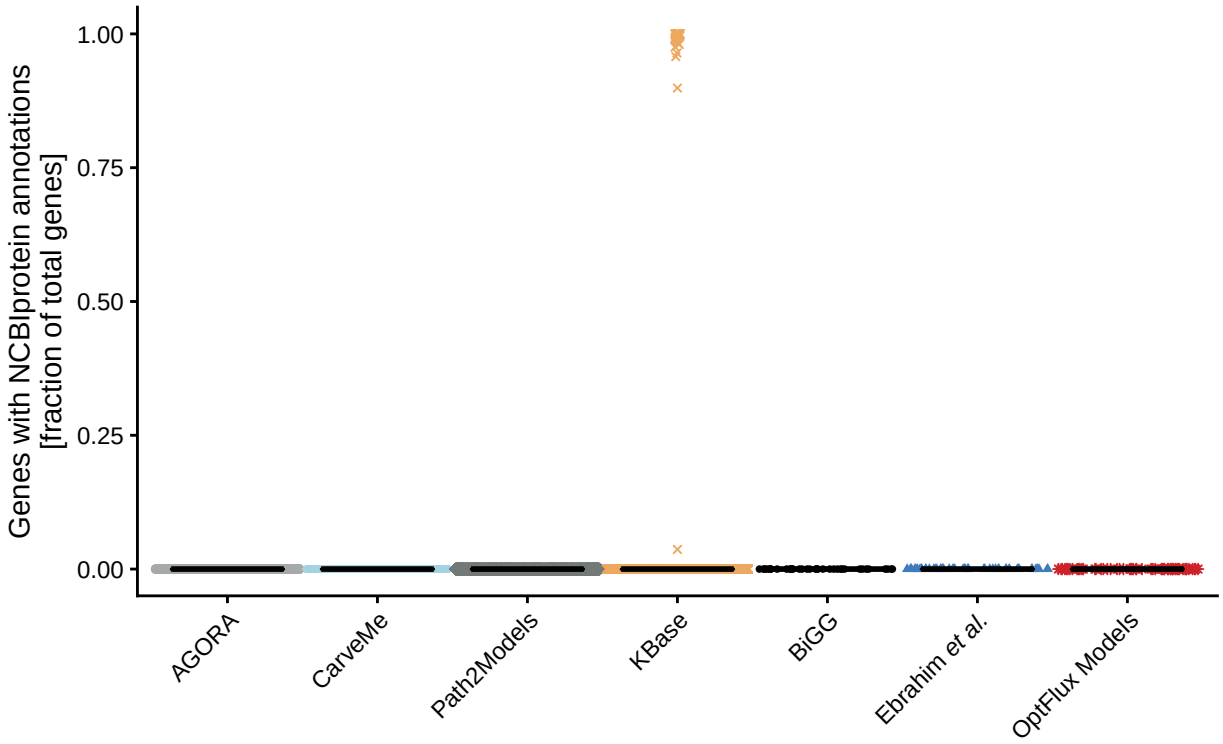

Figure S66: Gene NCBIprotein Annotation

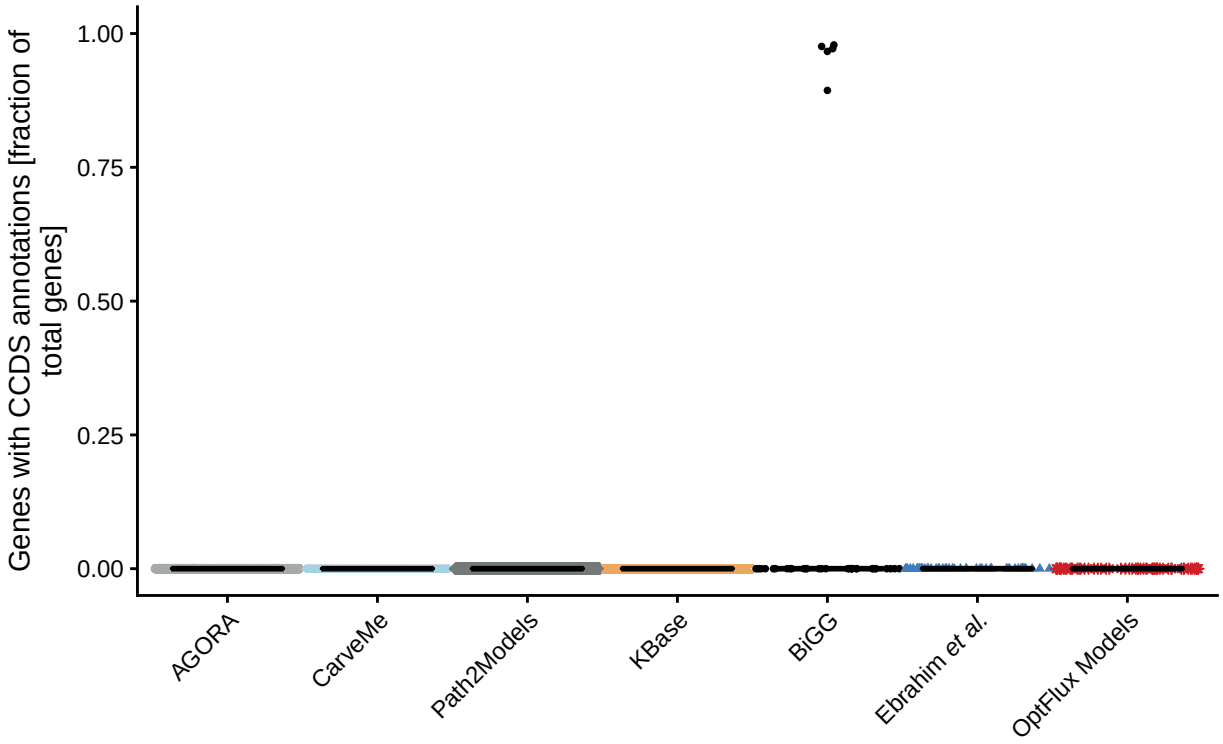

Figure S67: Gene CCDS Annotation

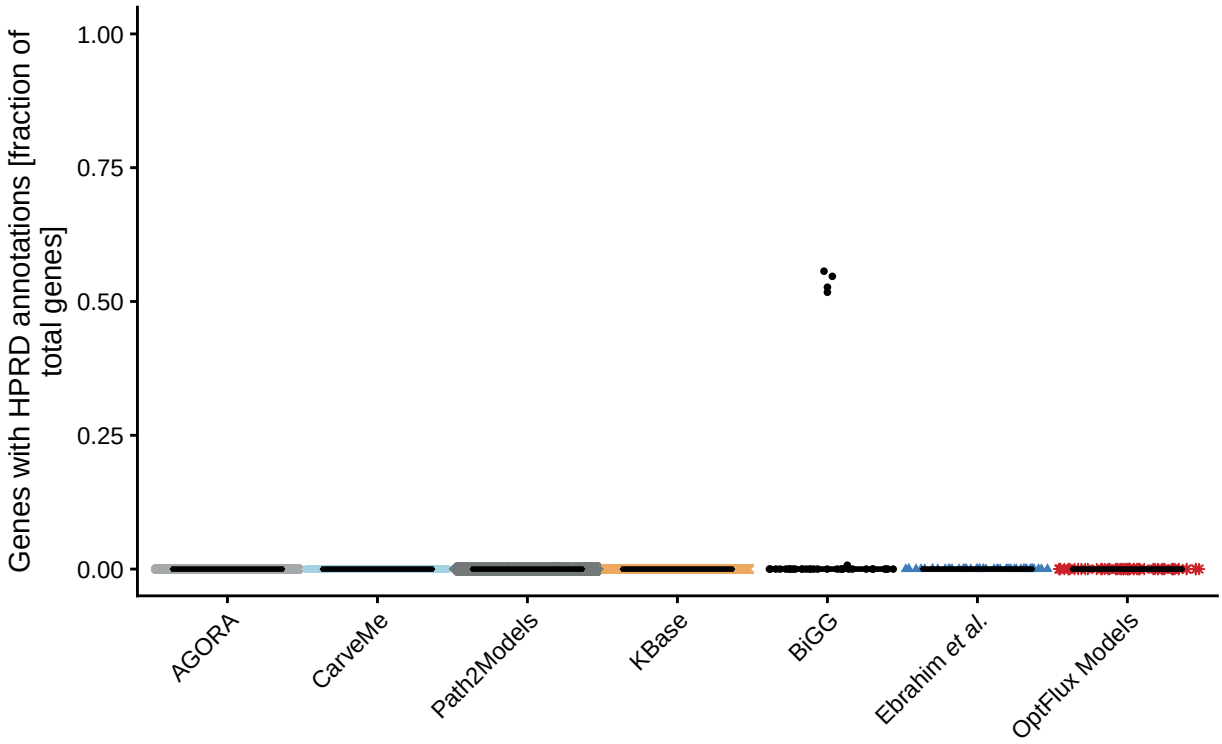

Figure S68: Gene HPRD Annotation

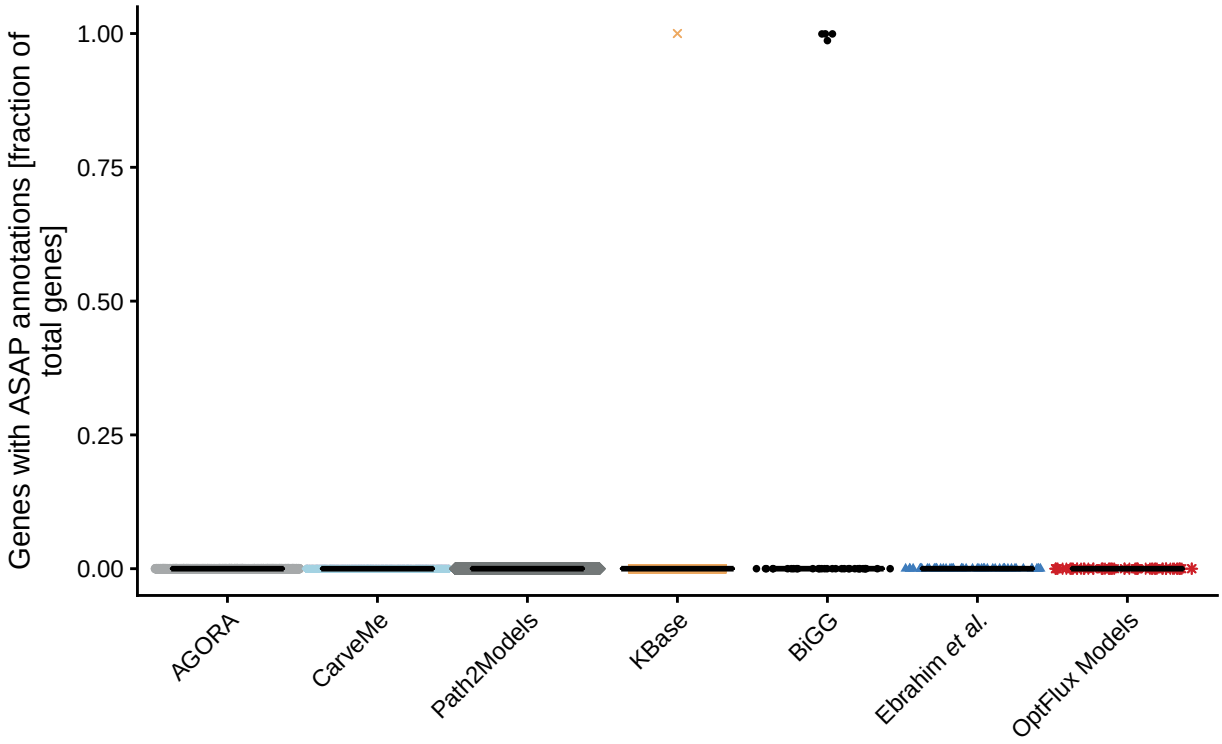

Figure S69: Gene ASAP Annotation

3.3.4.2 Gene Annotation Conformity Per Database

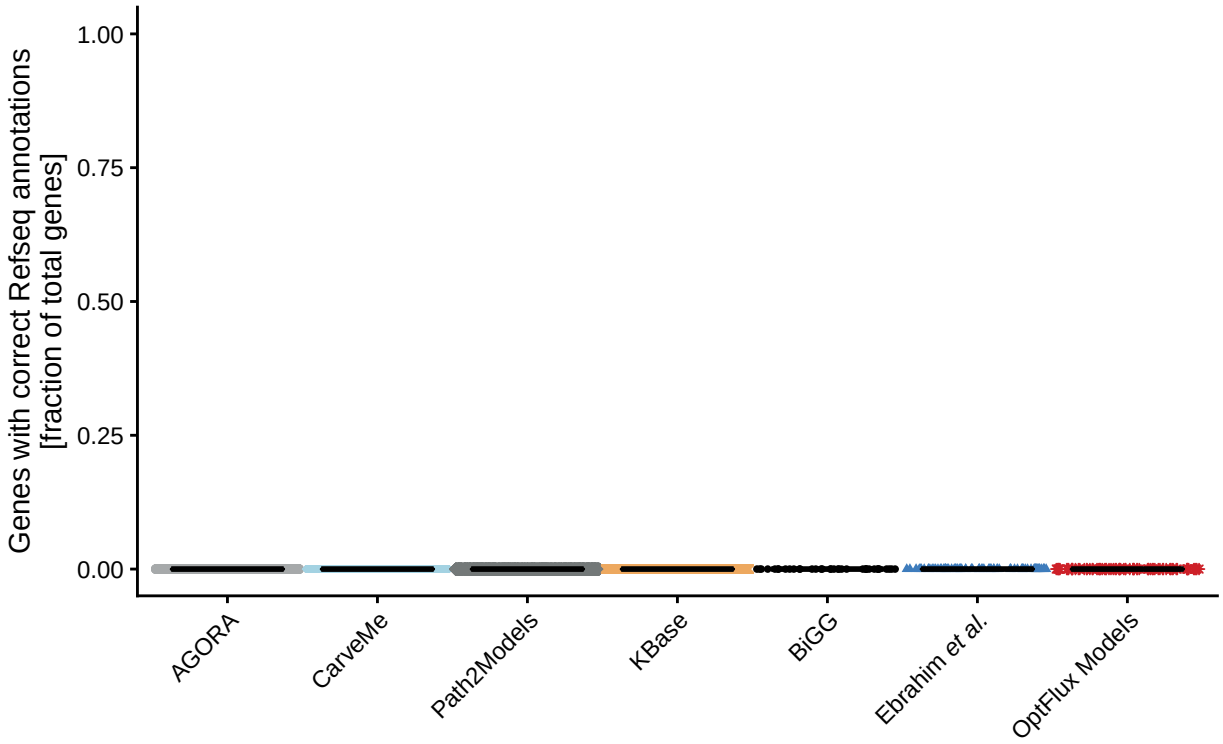

Figure S70: Correct Gene RefSeq Annotation

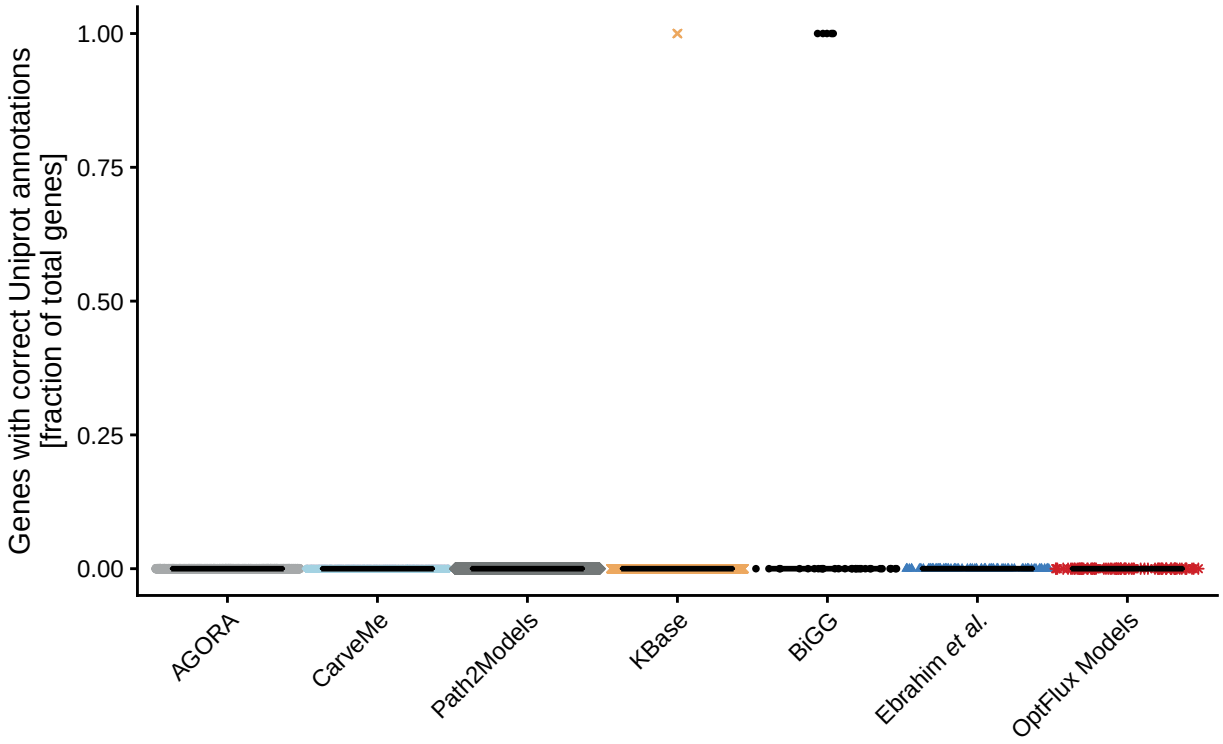

Figure S71: Correct Gene UniProt Annotation

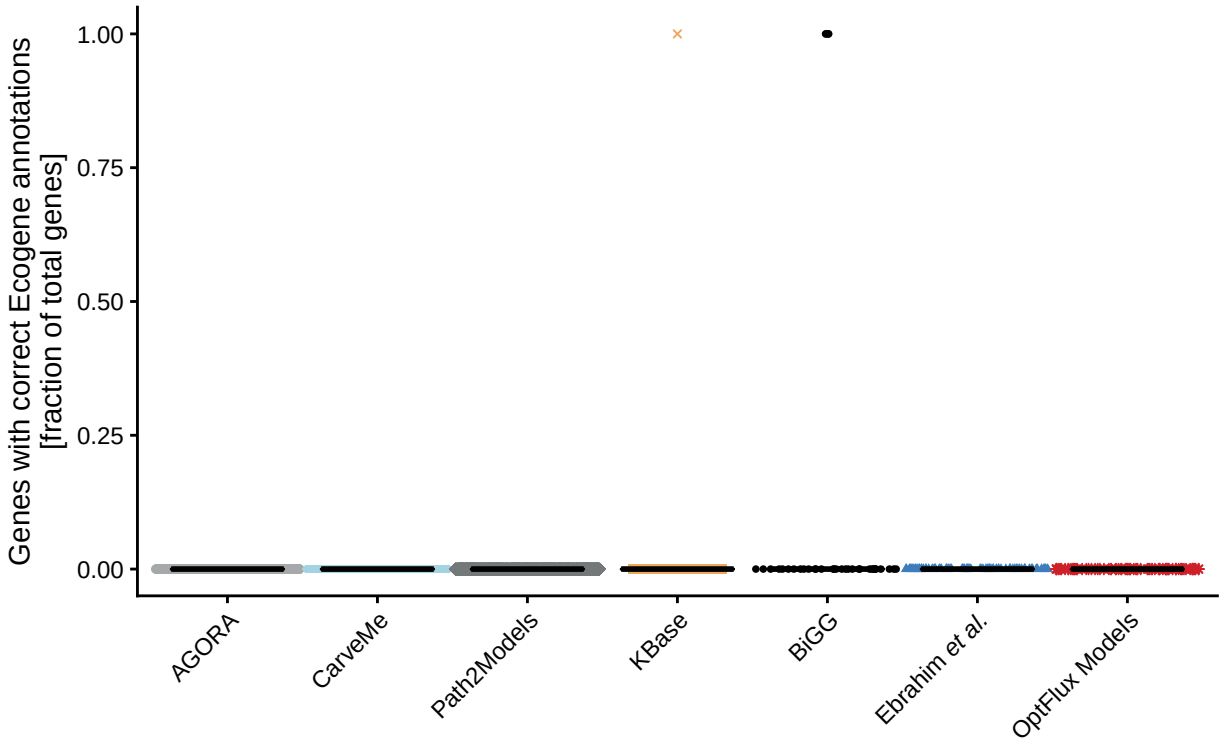

Figure S72: Correct Gene EcoGene Annotation

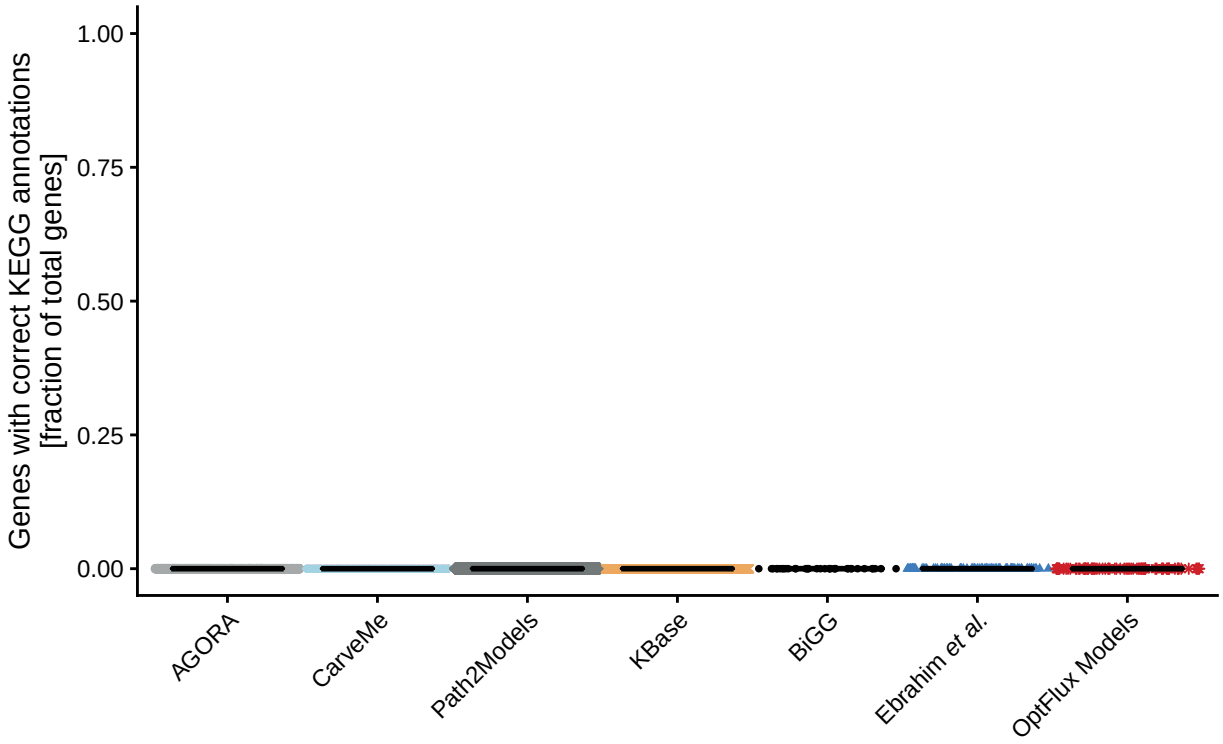

Figure S73: Correct Gene KEGG.genes Annotation

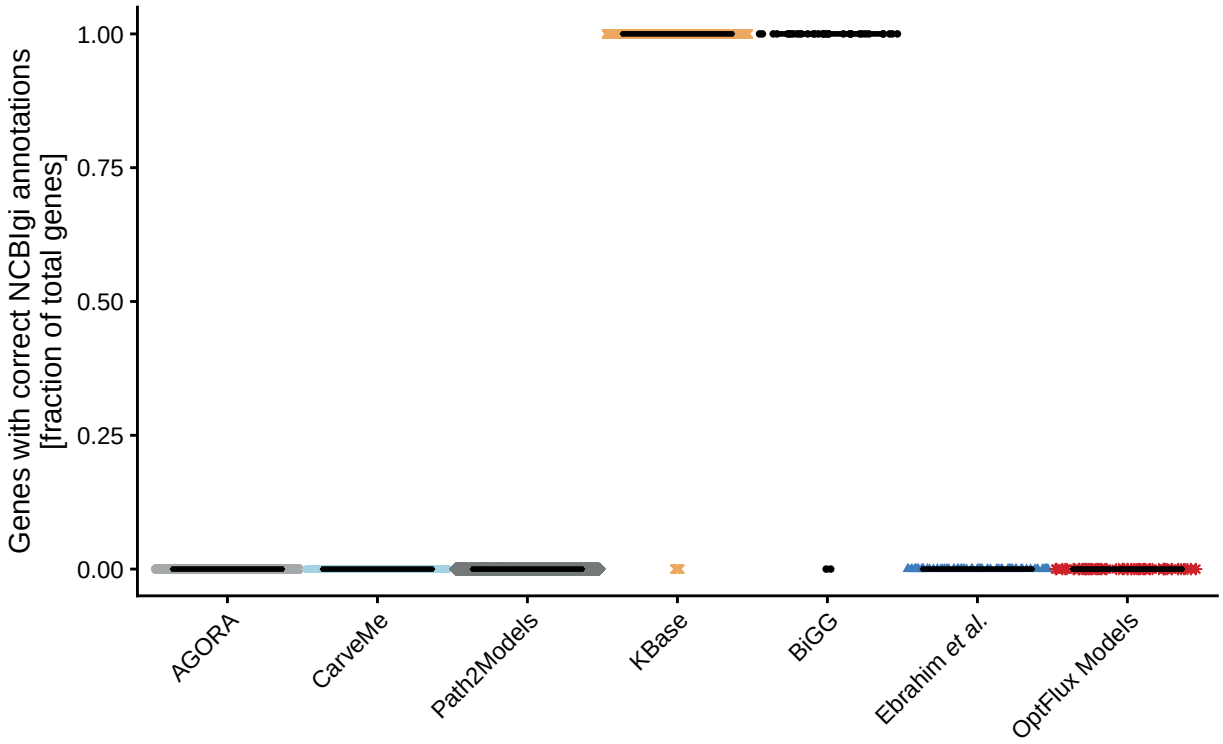

Figure S74: Correct Gene NCBIgi Annotation

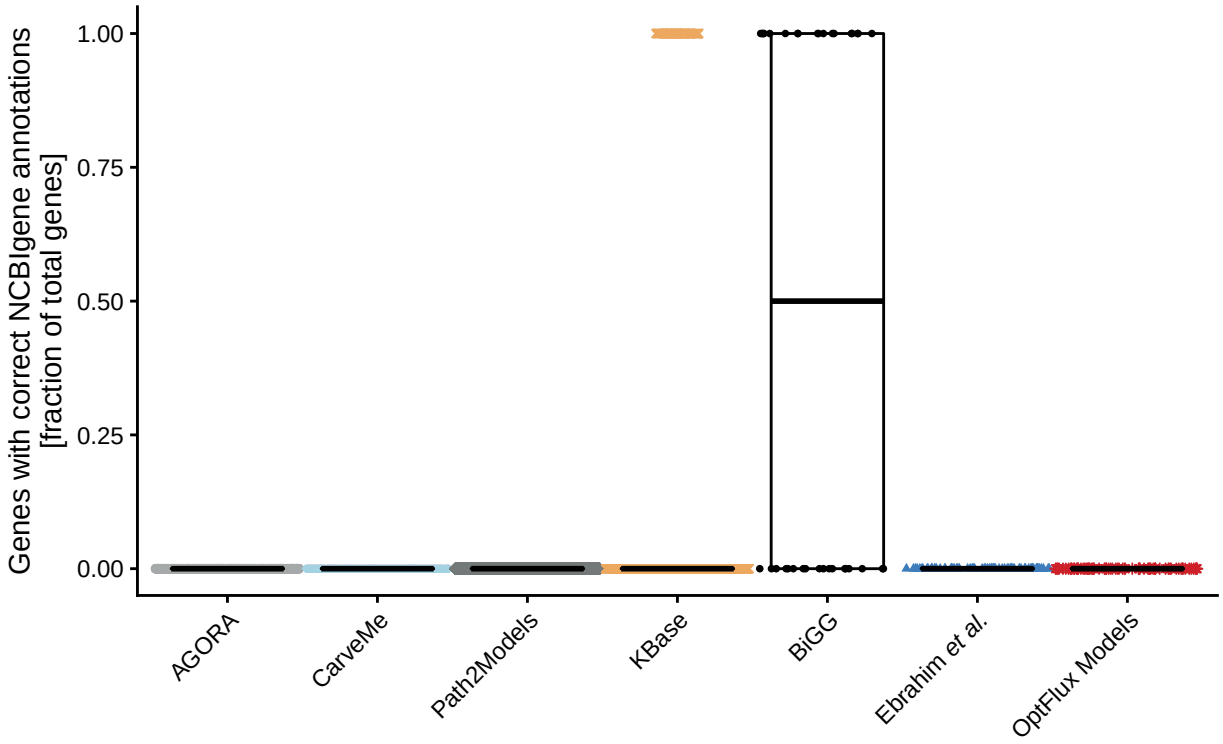

Figure S75: Correct Gene NCBIgene Annotation

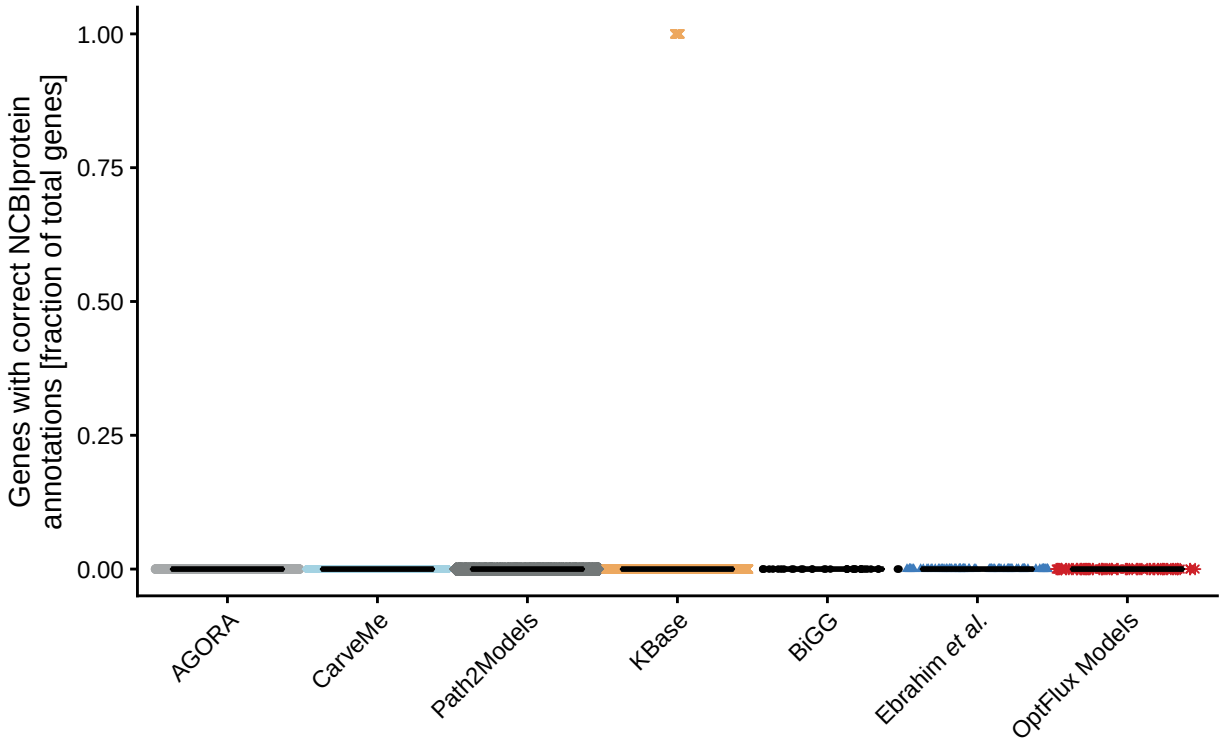

Figure S76: Correct Gene NCBIprotein Annotation

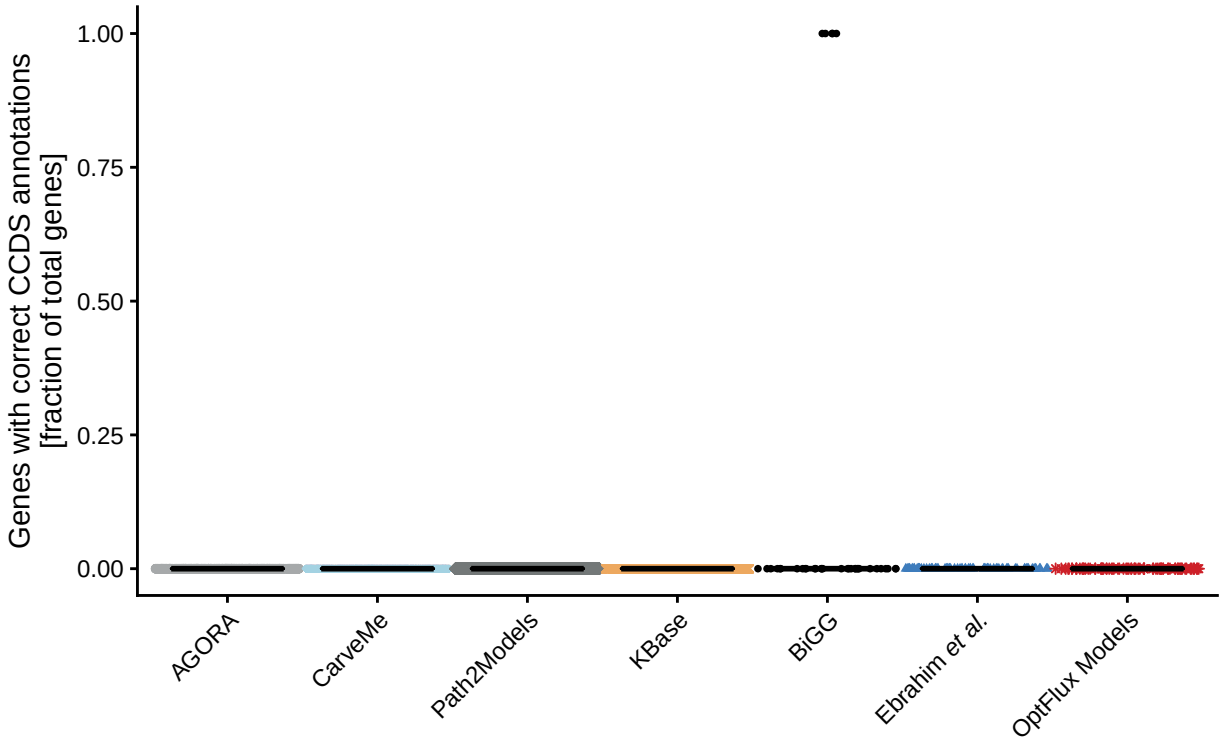

Figure S77: Correct Gene CCDS Annotation

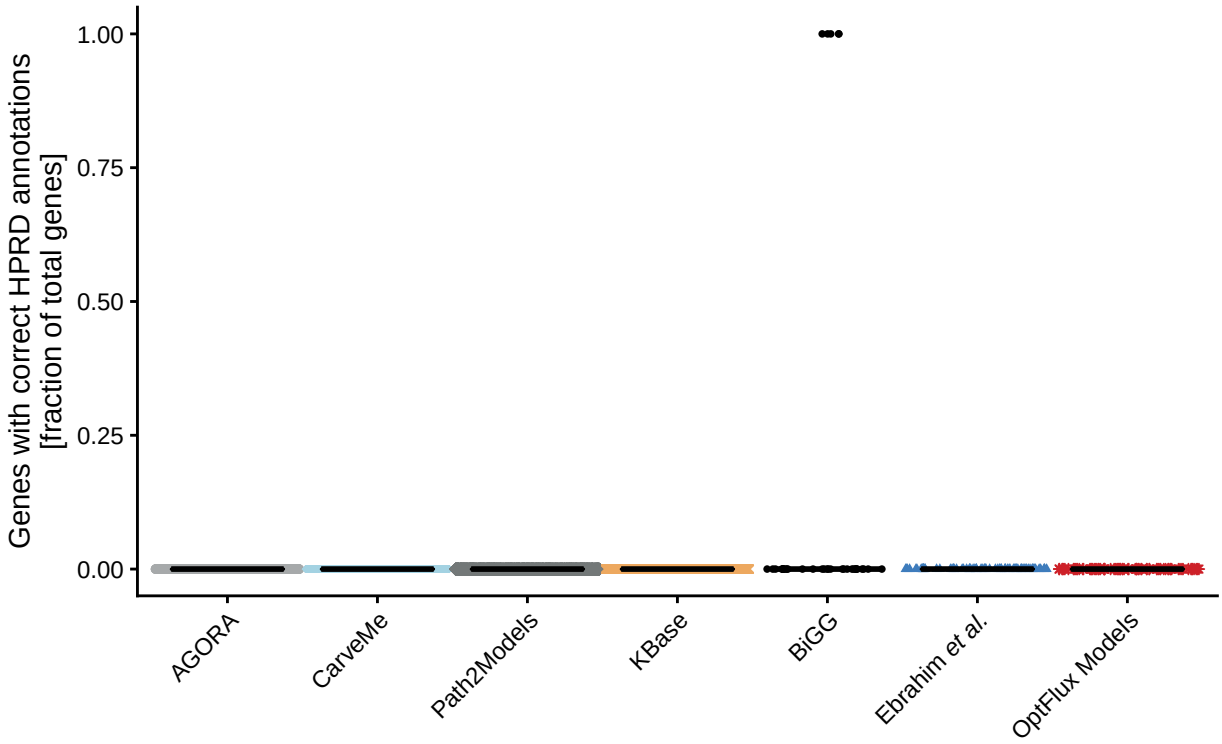

Figure S78: Correct Gene HPRD Annotation

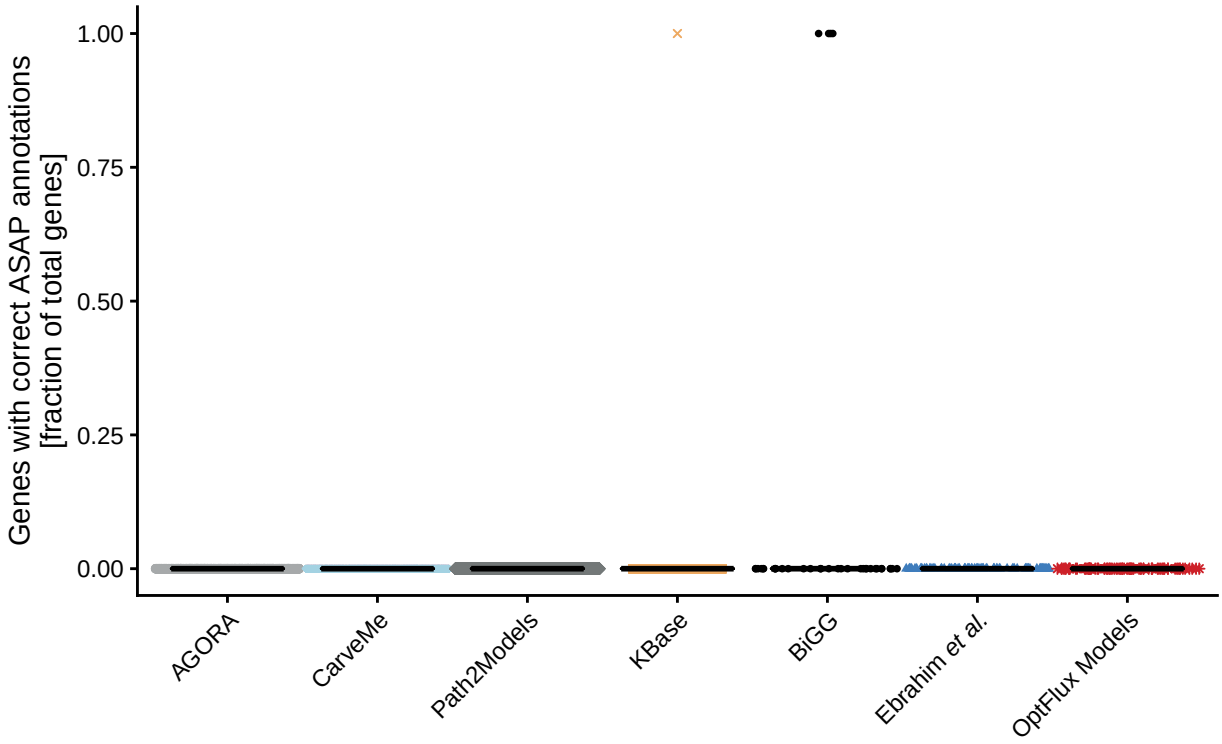

Figure S79: Correct Gene ASAP Annotation

3.3.5 Annotation - SBO Terms

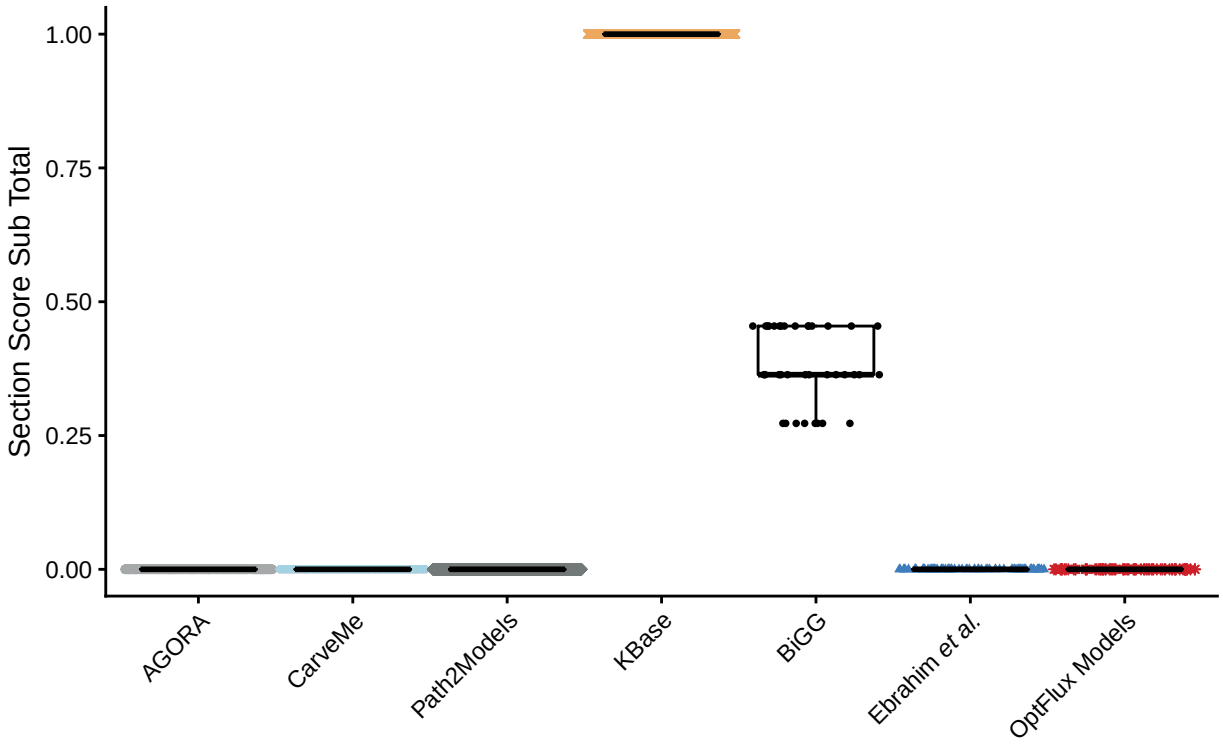

Figure S80: Annotation - SBO Terms. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

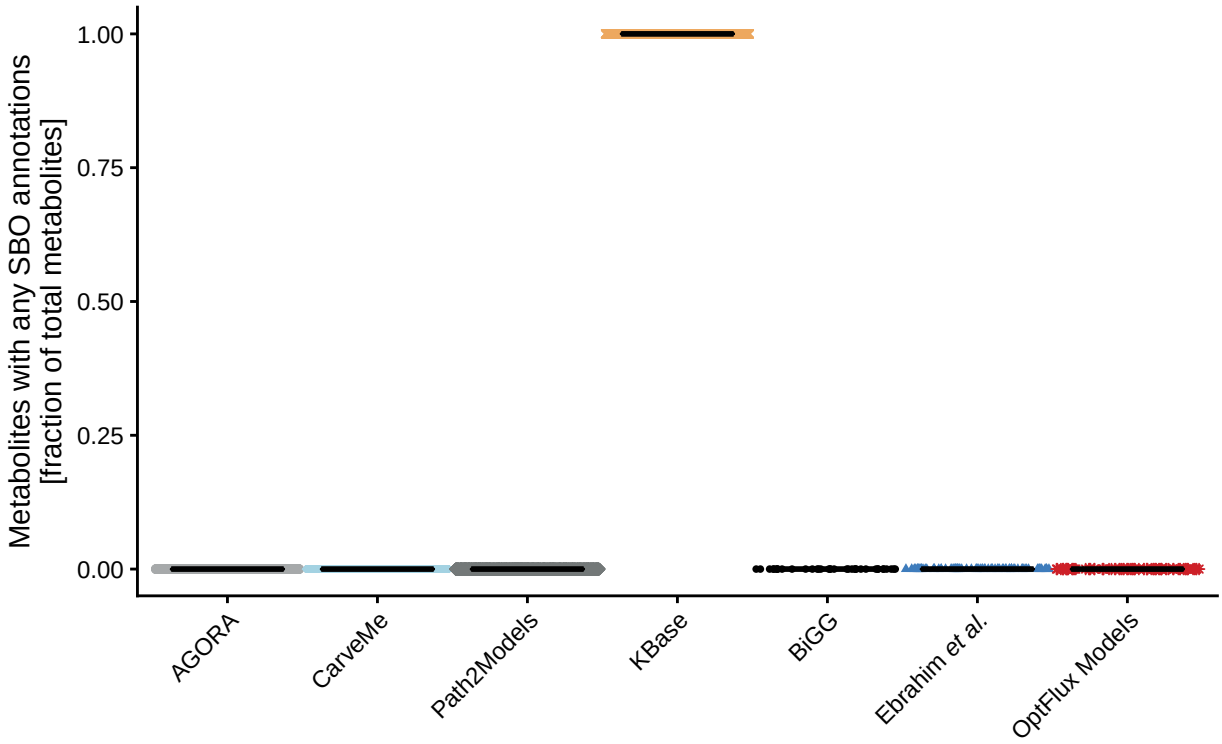

Figure S81: Metabolite General SBO Presence

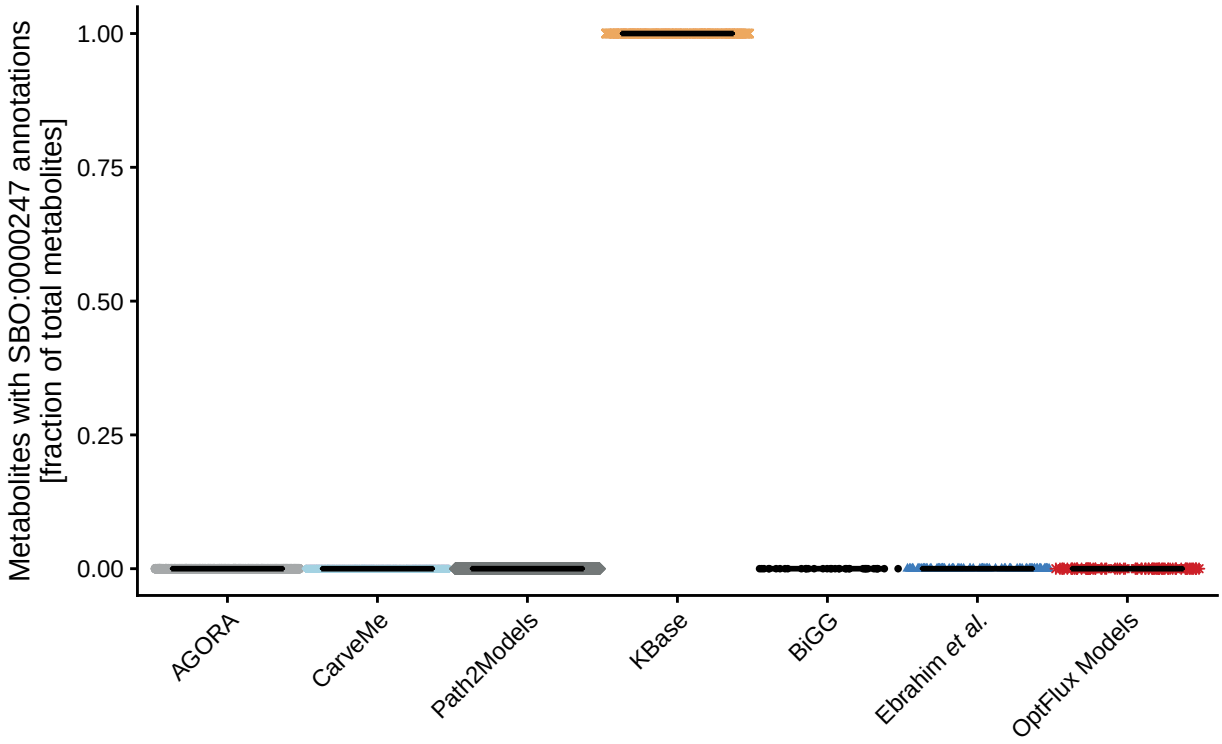

Figure S82: Metabolite SBO:0000247 Presence

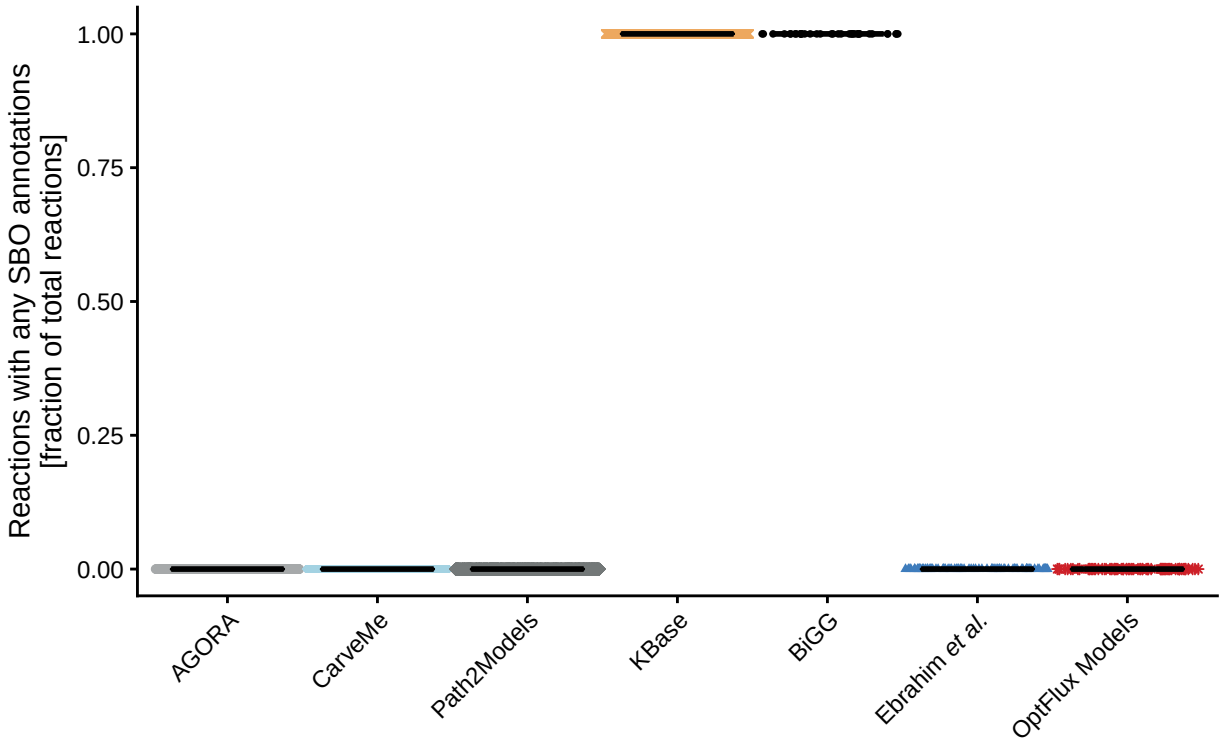

Figure S83: Reaction General SBO Presence

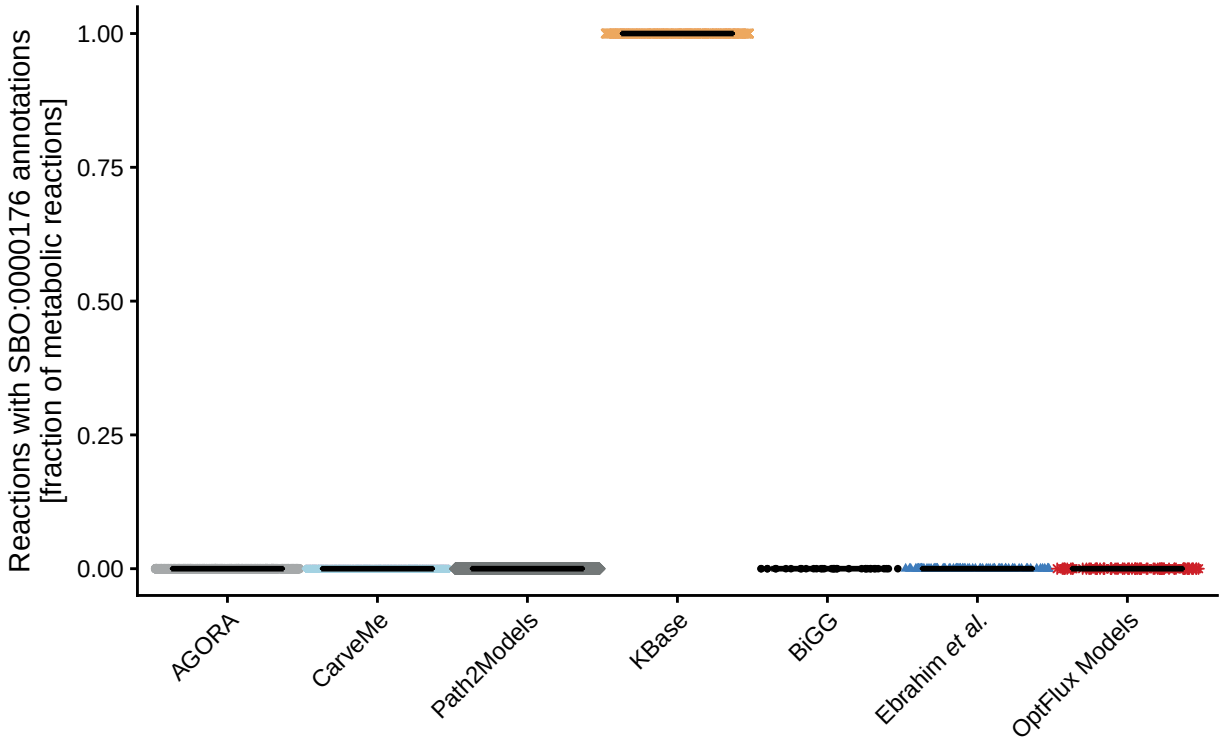

Figure S84: Metabolic Reaction SBO:0000176 Presence

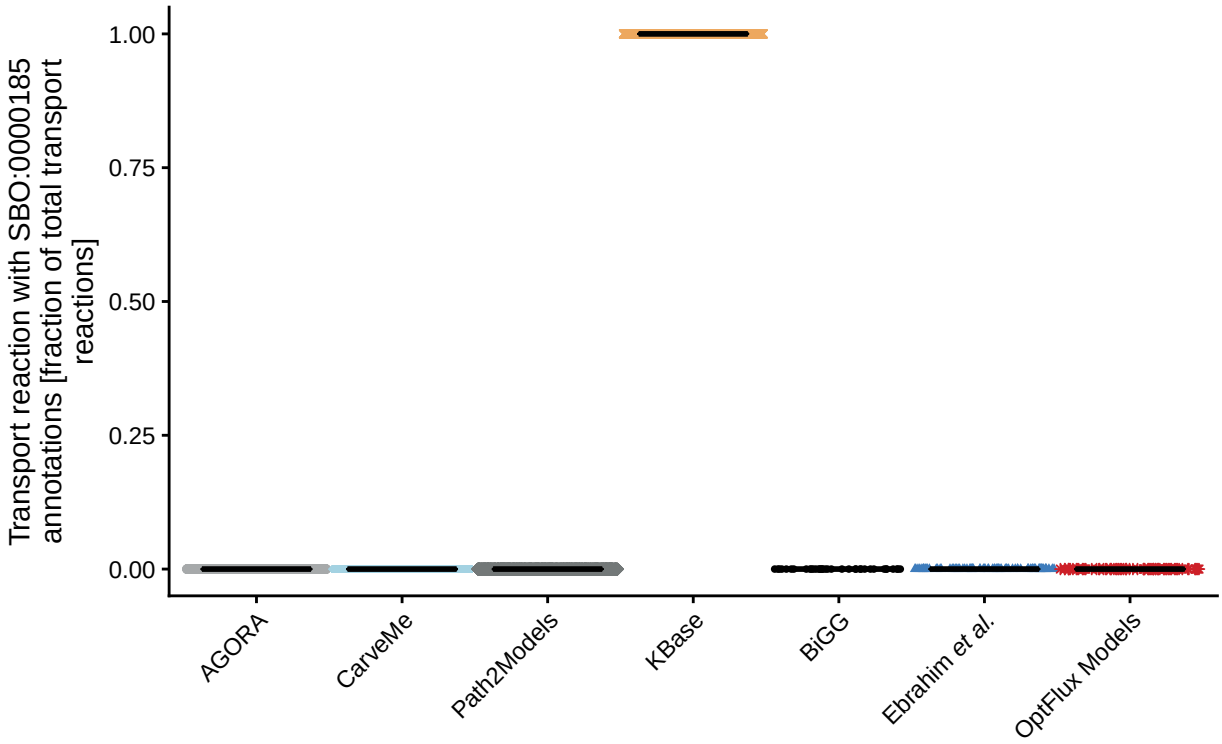

Figure S85: Transport Reaction SBO:0000185 Presence

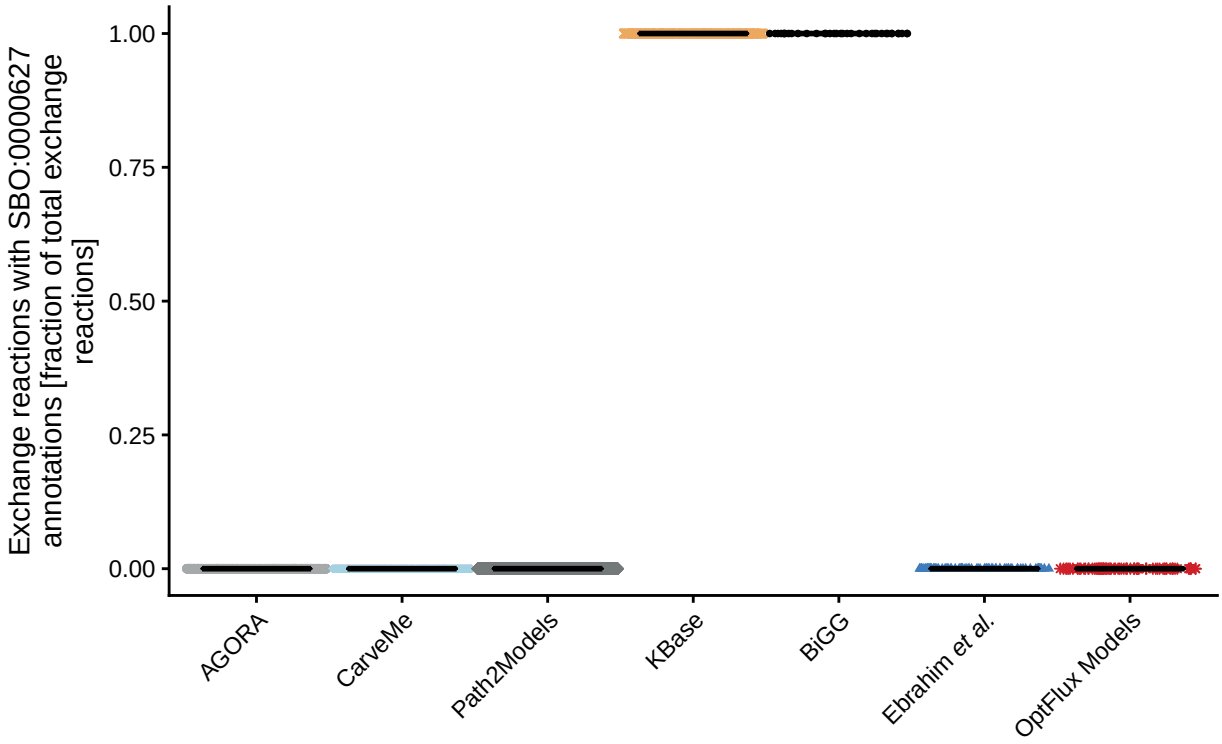

Figure S86: Exchange Reaction SBO:0000627 Presence

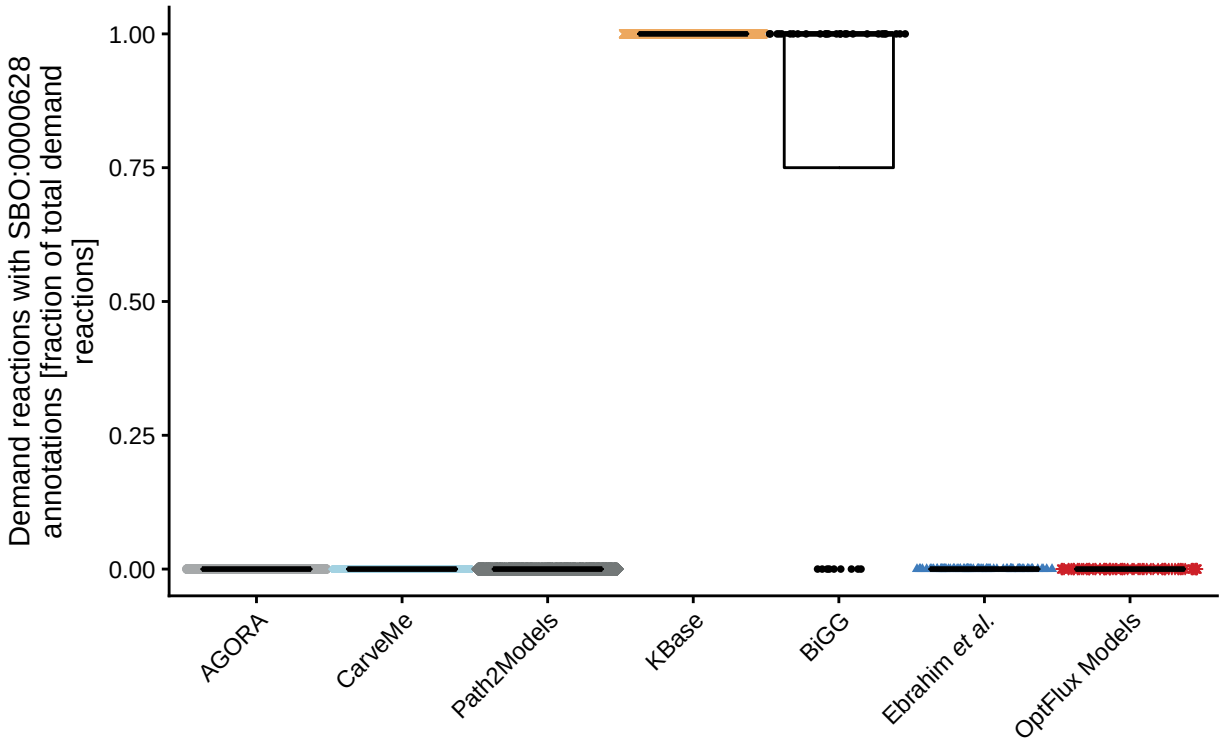

Figure S87: Demand Reaction SBO:0000628 Presence

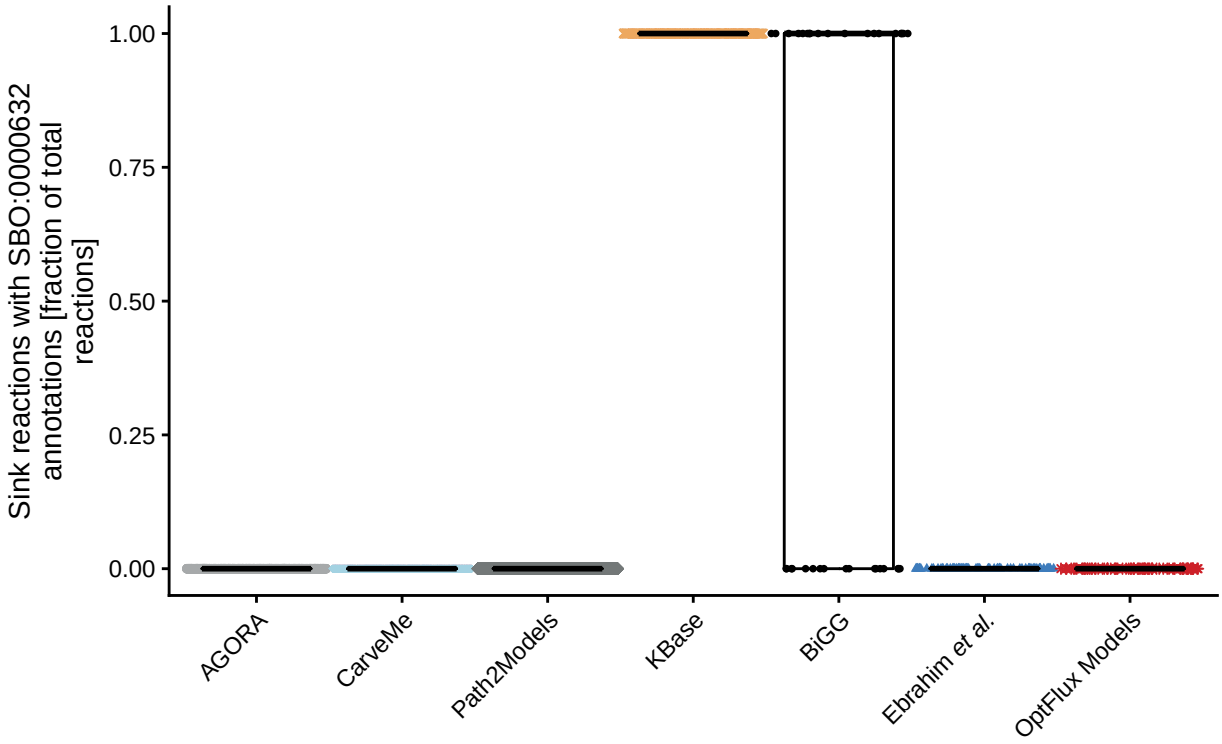

Figure S88: Sink Reaction SBO:0000632 Presence

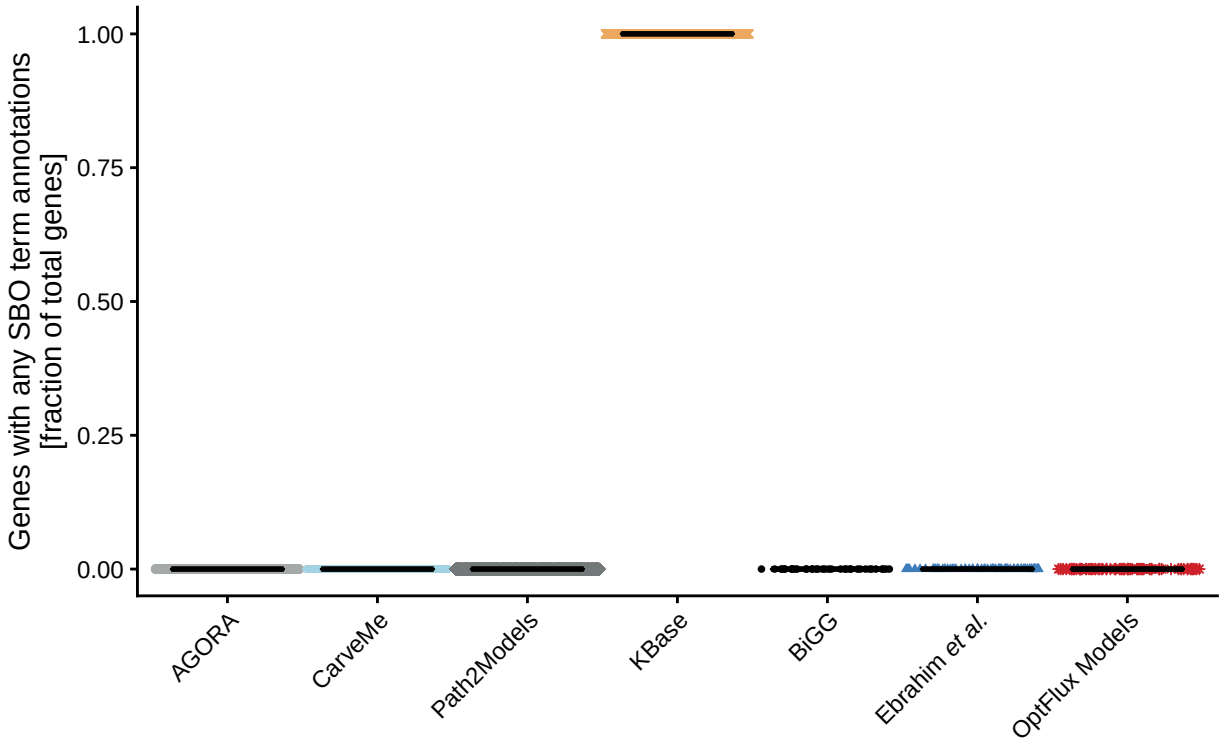

Figure S89: Gene General SBO Presence

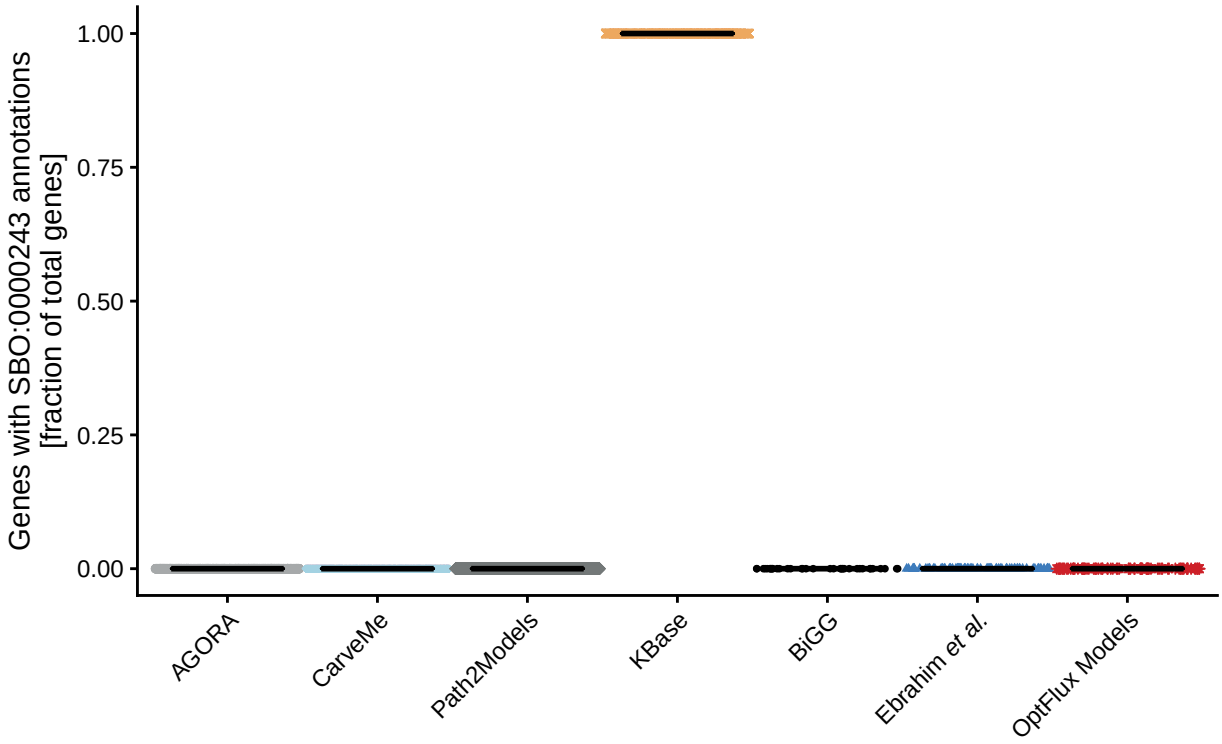

Figure S90: Gene SBO:0000243 Presence

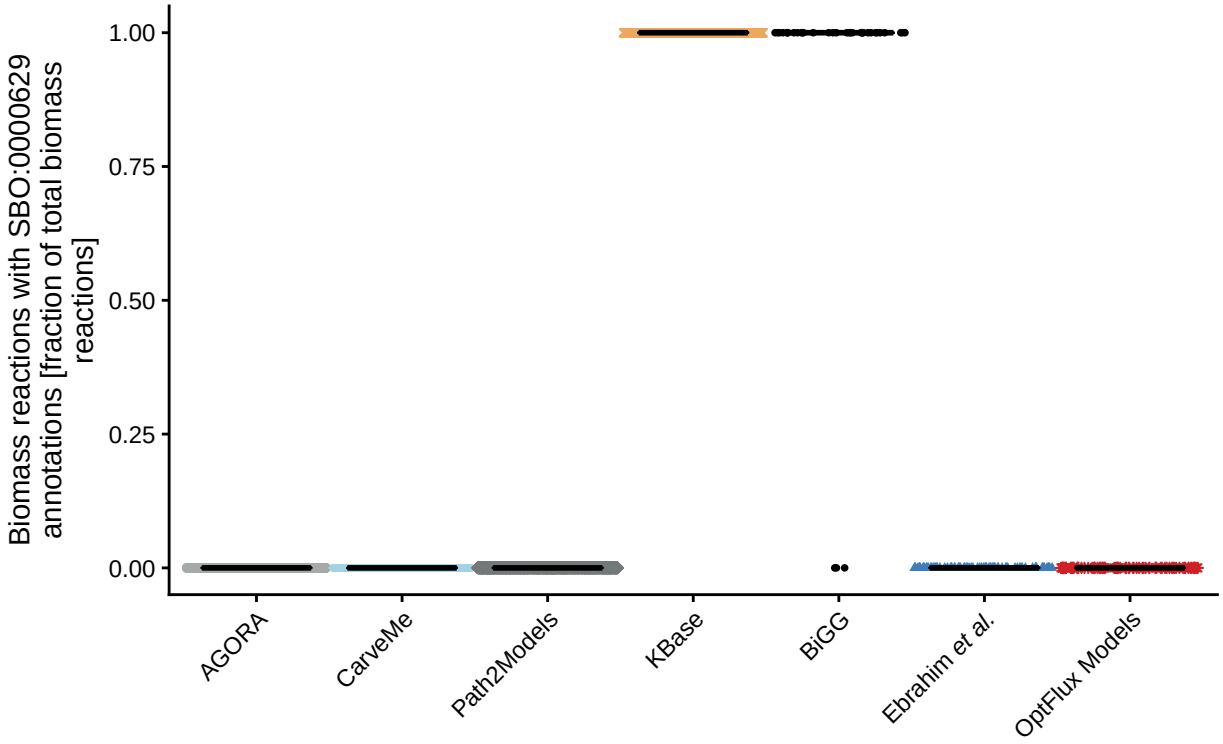

Figure S91: Biomass Reaction SBO:0000629 Presence

3.4 Specific Section

3.4.1 SBML

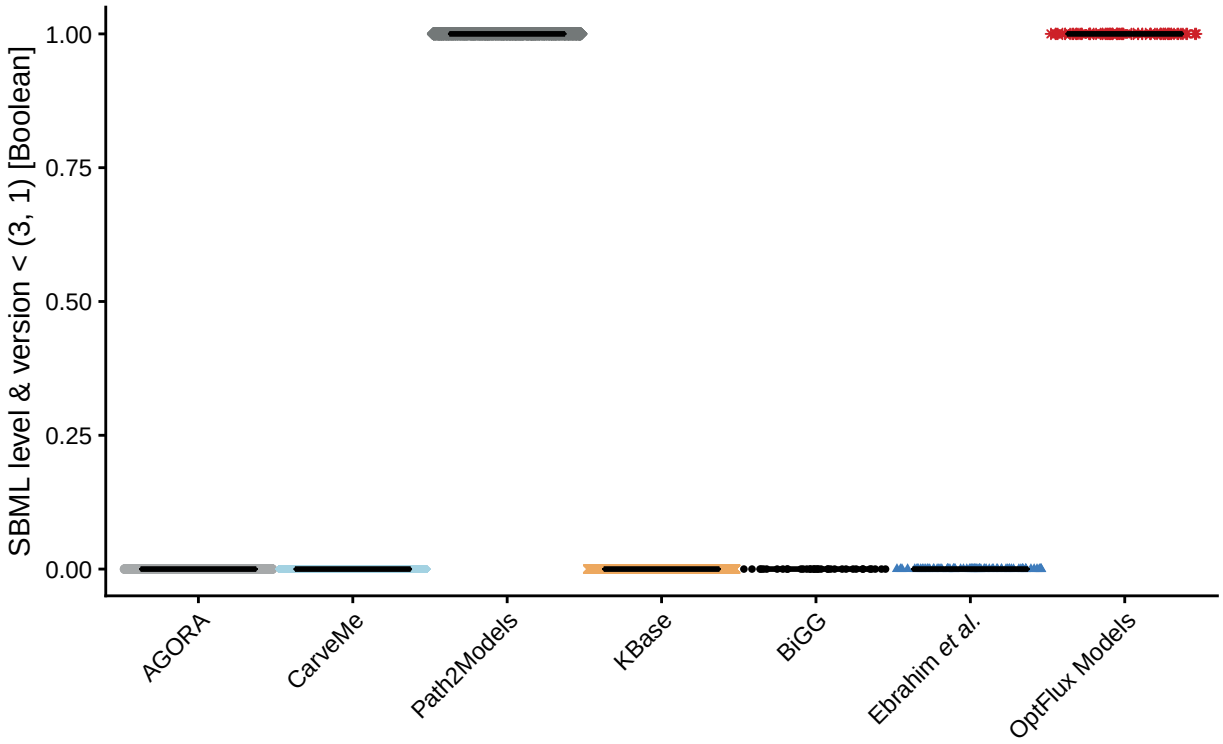

Figure S92: SBML Level and Version

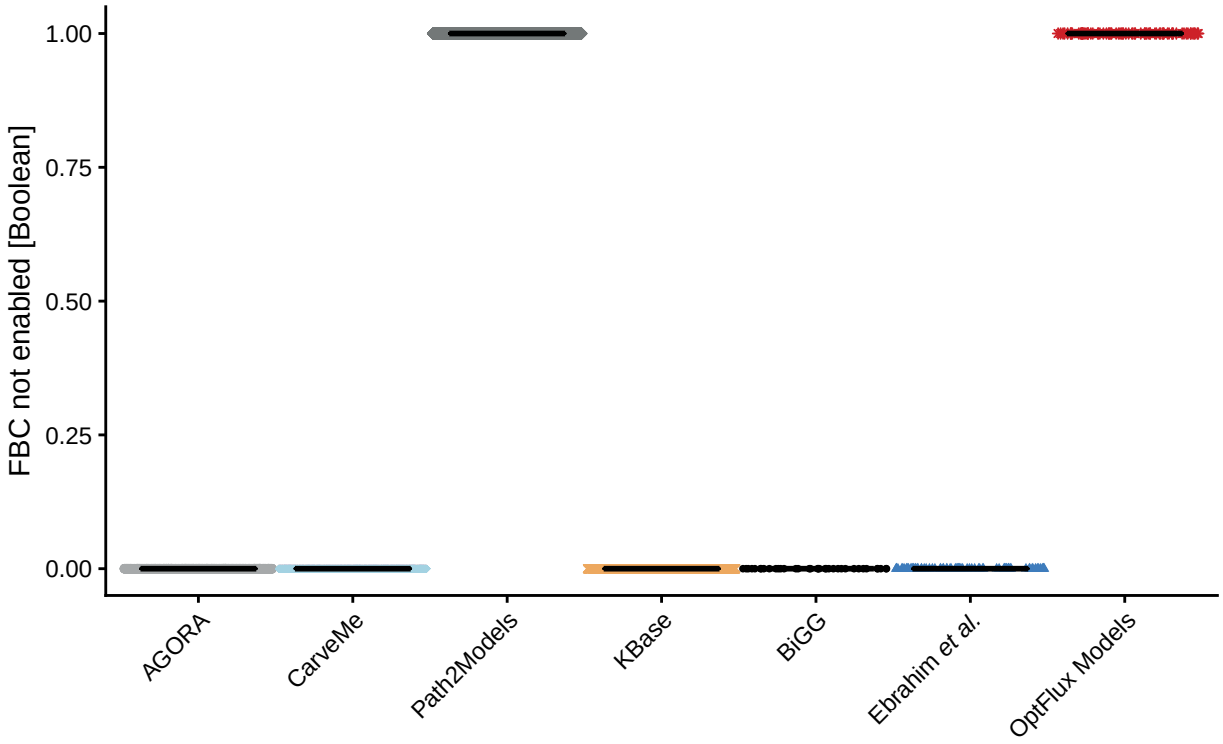

Figure S93: FBC not Enabled

3.4.2 Basic Information

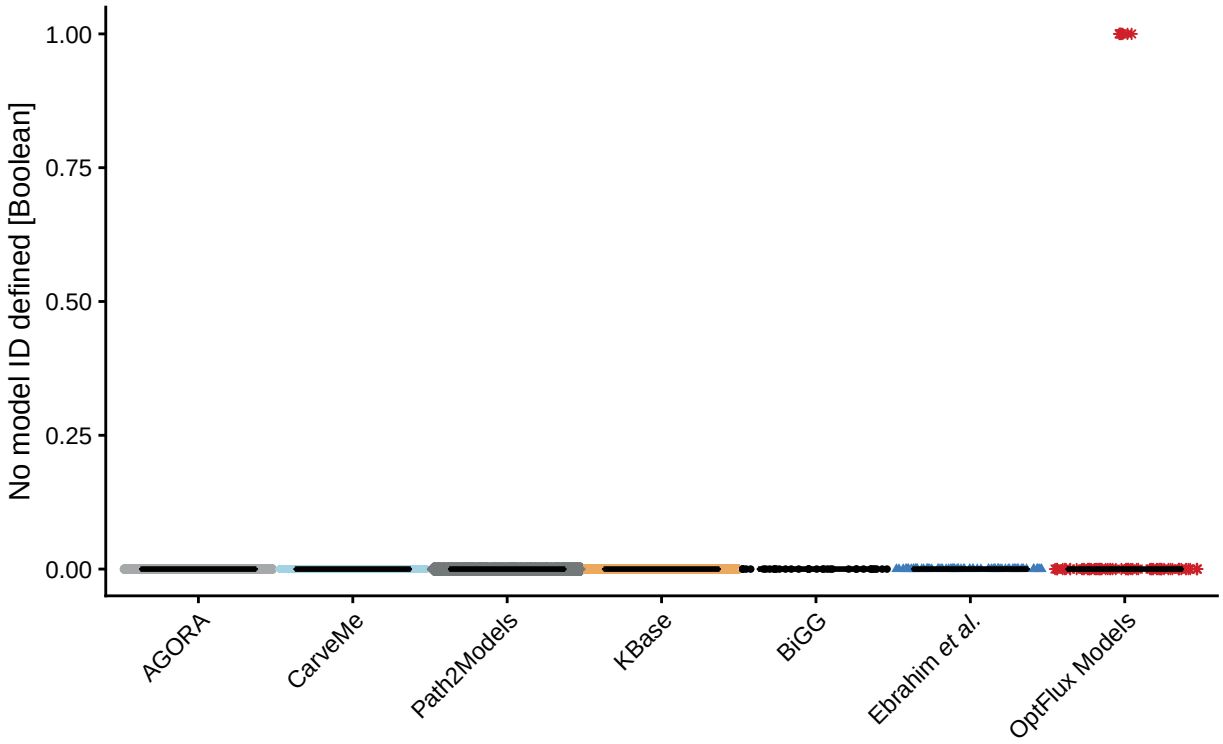

Figure S94: Model Identifier Presence

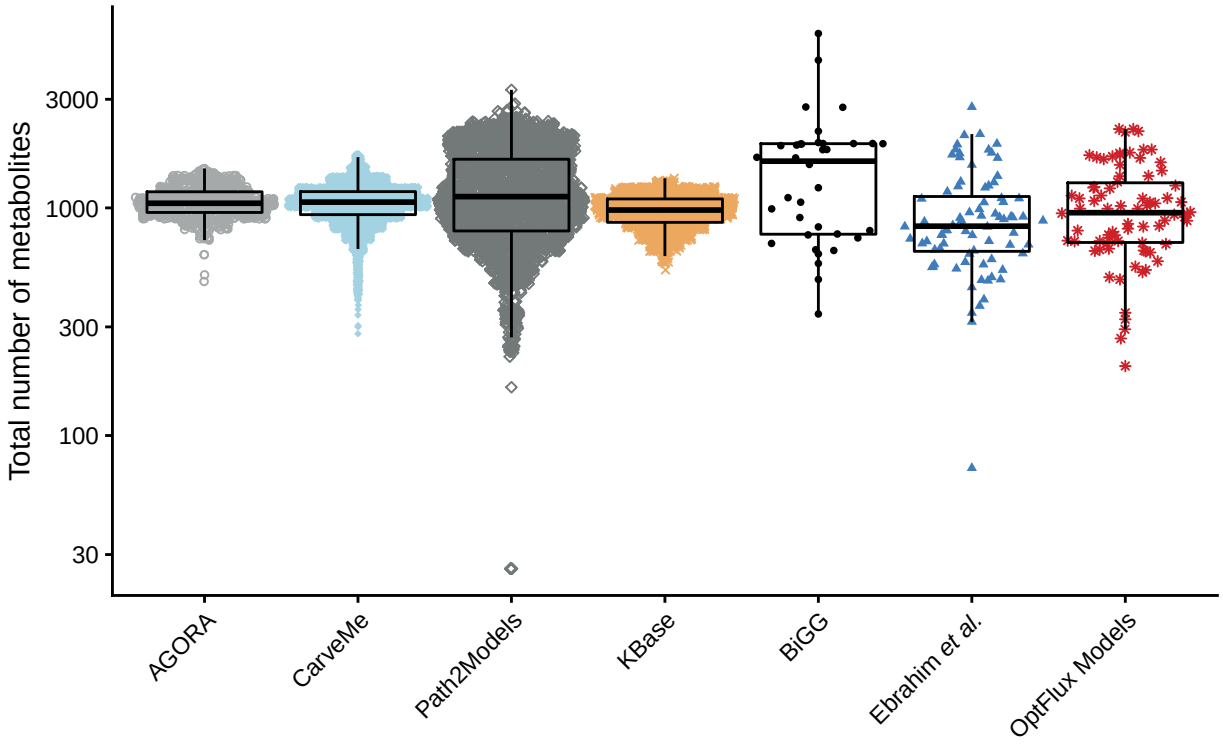

Figure S95: Number of Metabolites

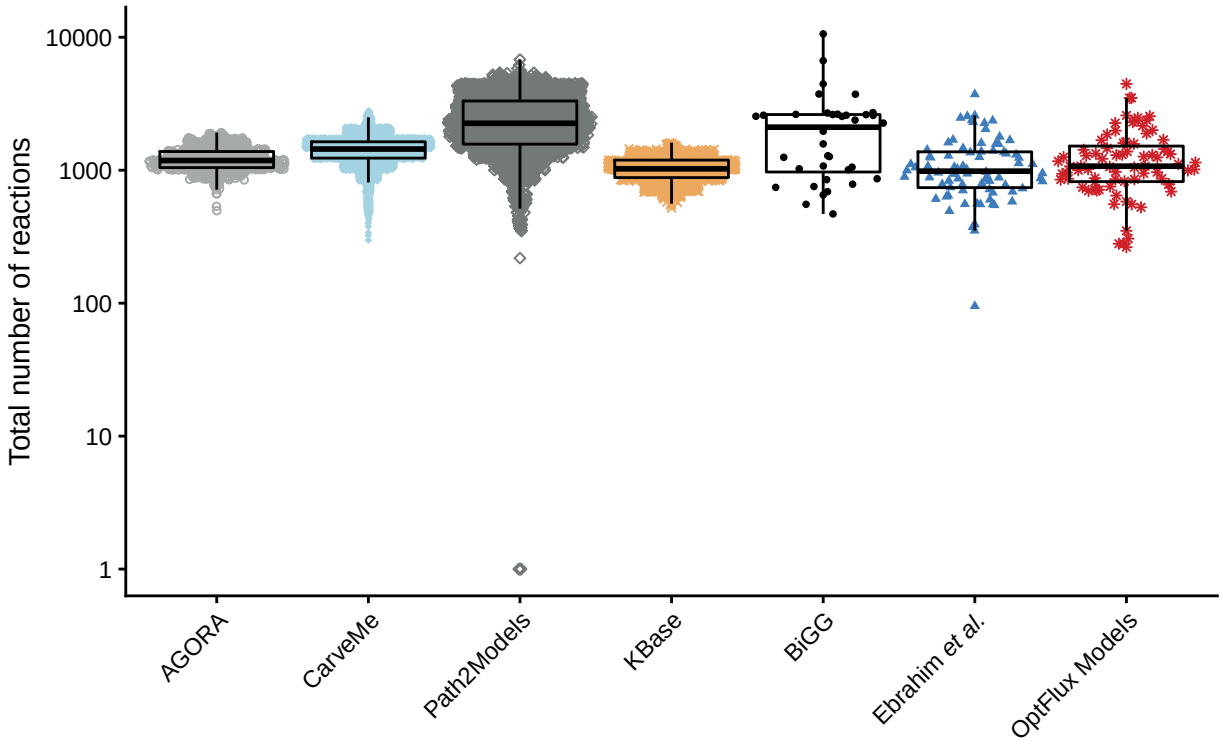

Figure S96: Number of Reactions

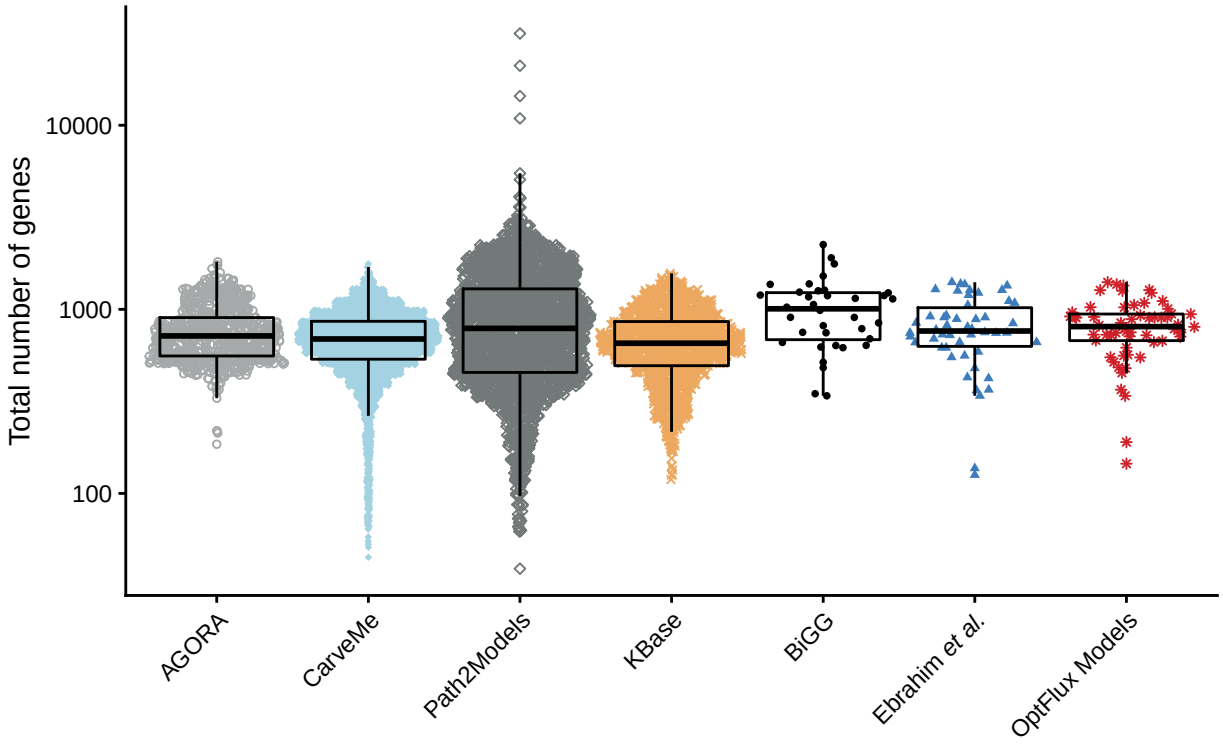

Figure S97: Number of Genes

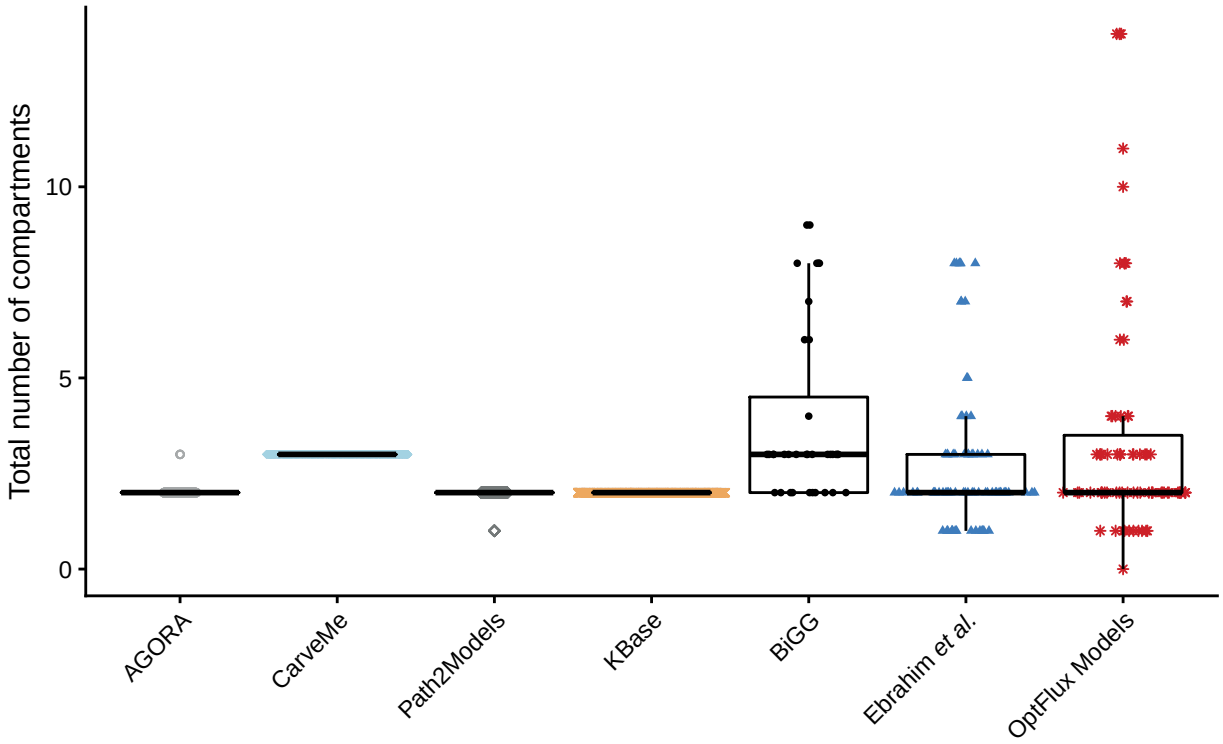

Figure S98: Number of Compartments

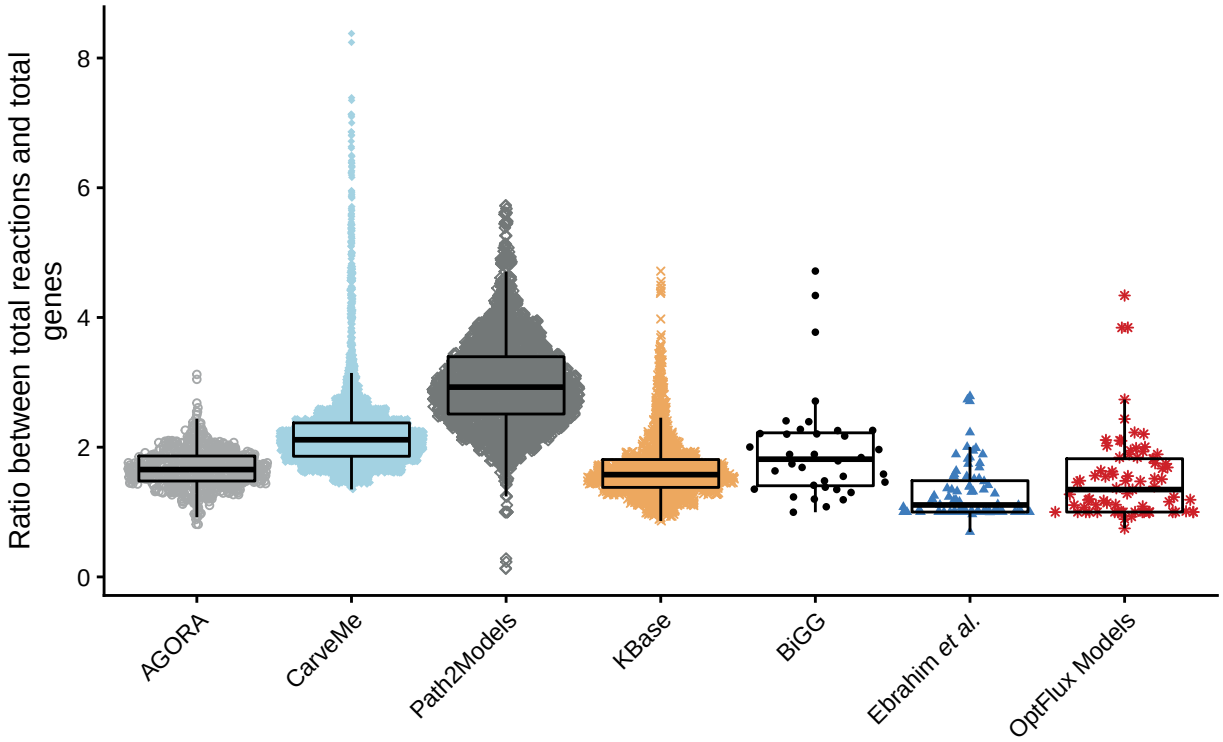

Figure S99: Metabolic Coverage

3.4.3 Metabolite Information

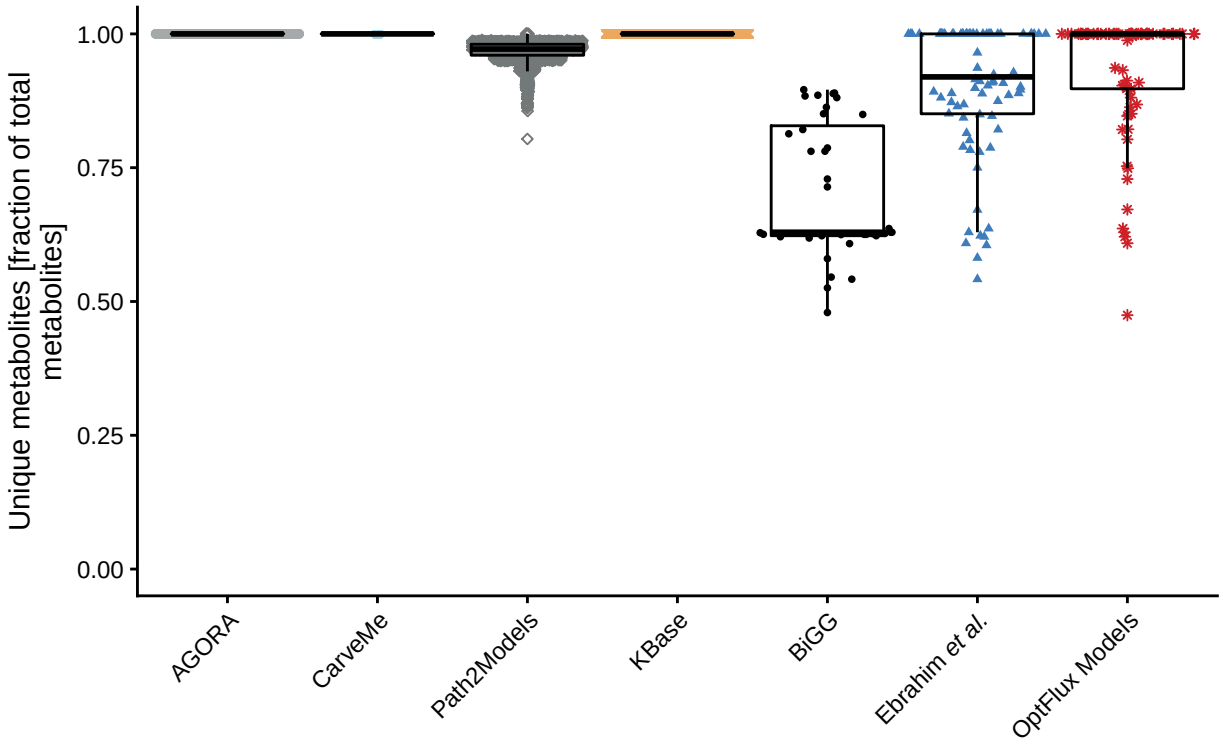

Figure S100: Unique Metabolites

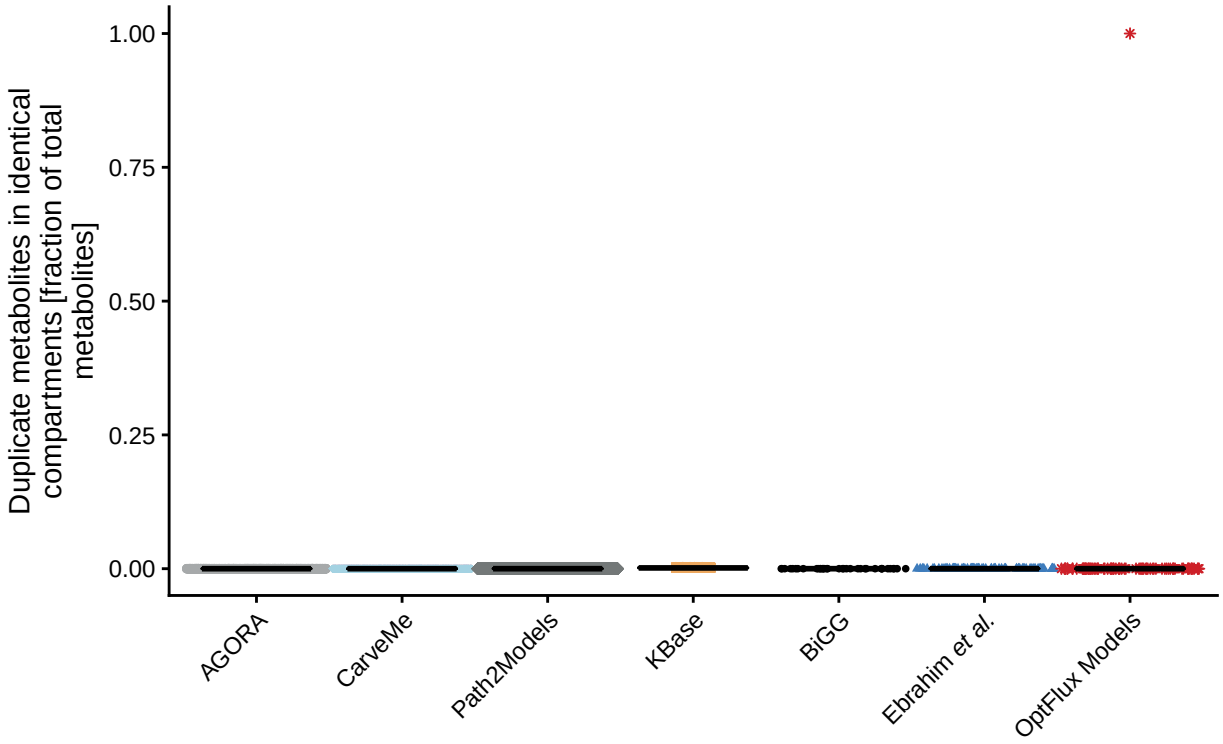

Figure S101: Duplicate Metabolites in Identical Compartments

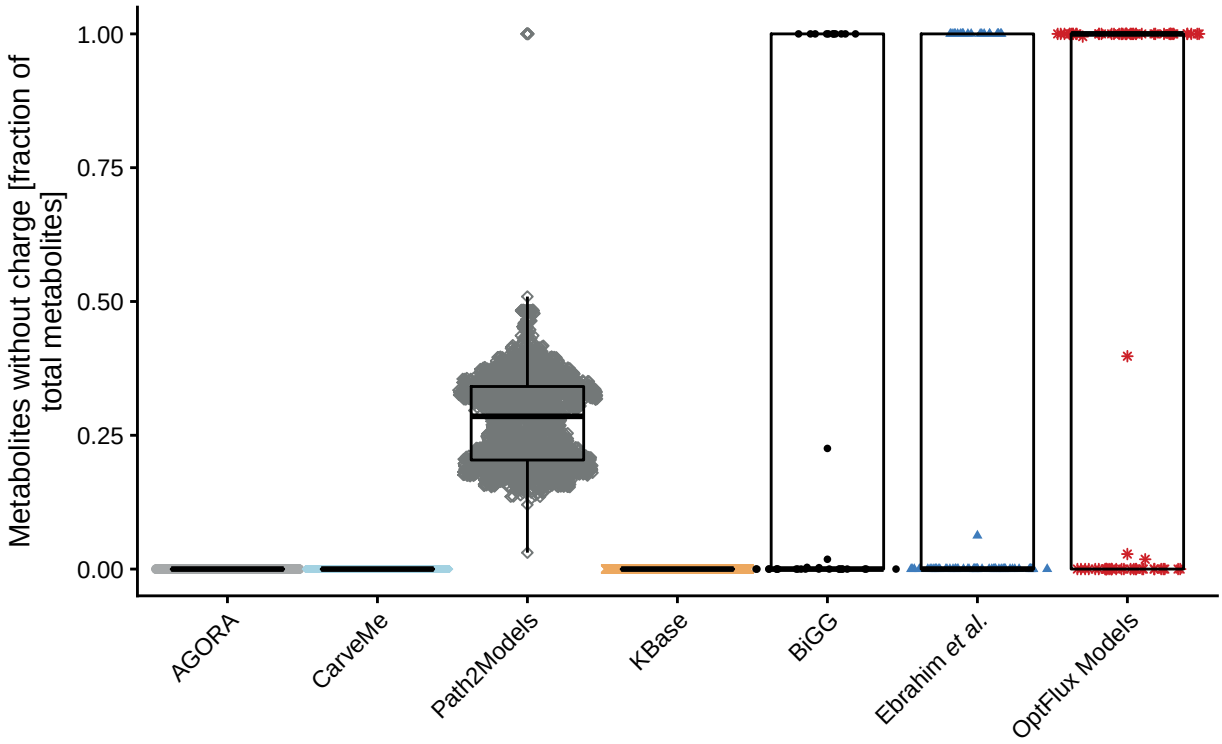

Figure S102: Metabolites Without Charge

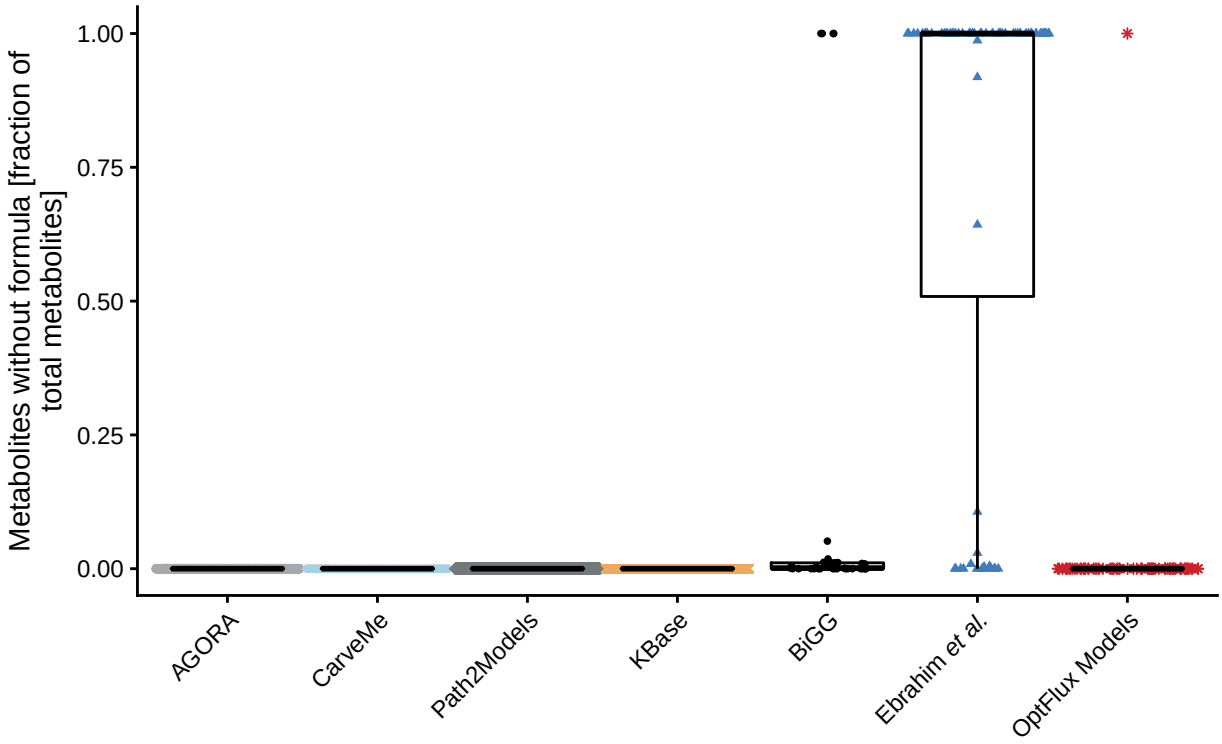

Figure S103: Metabolites Without Formula

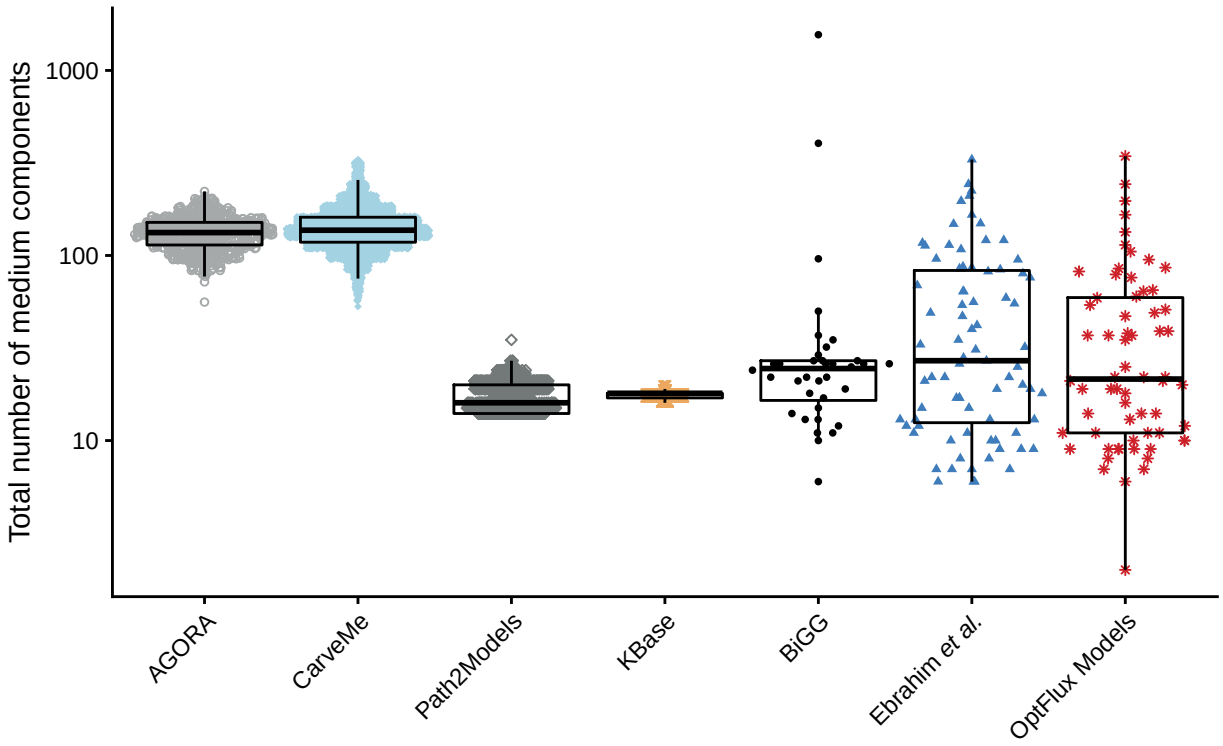

Figure S104: Number of Medium Components

3.4.4 Reaction Information

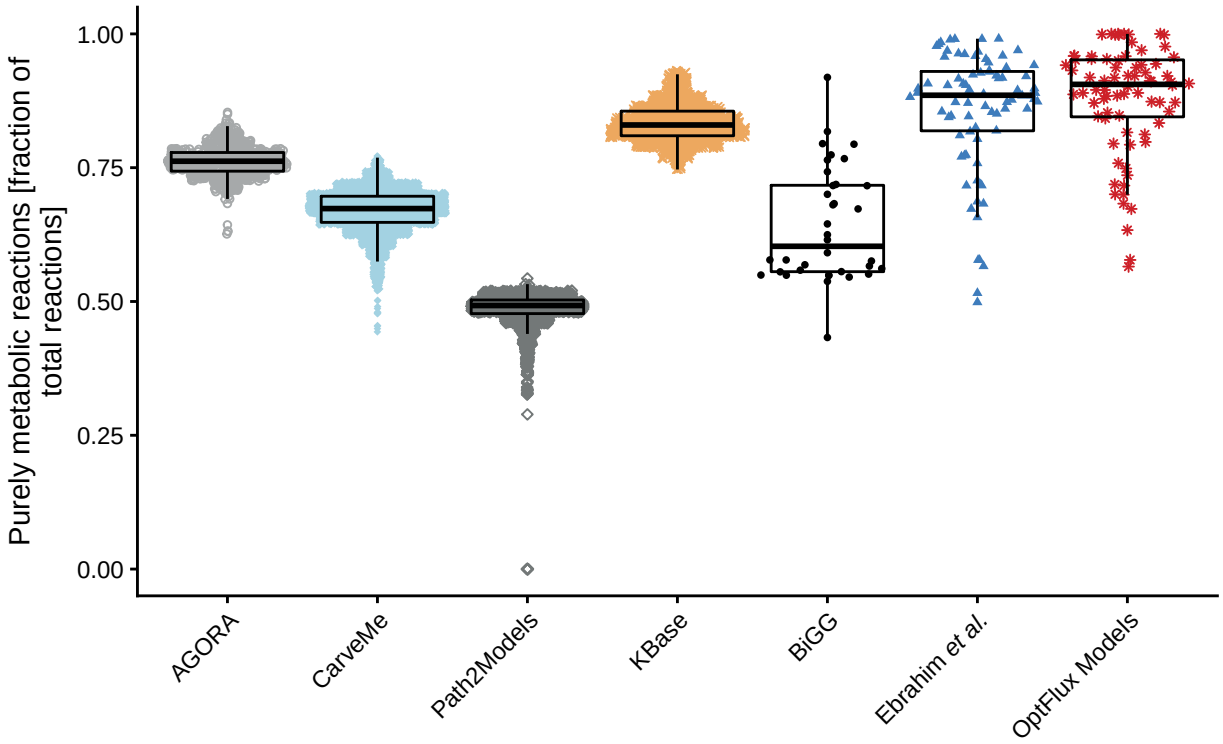

Figure S105: Purely Metabolic Reactions

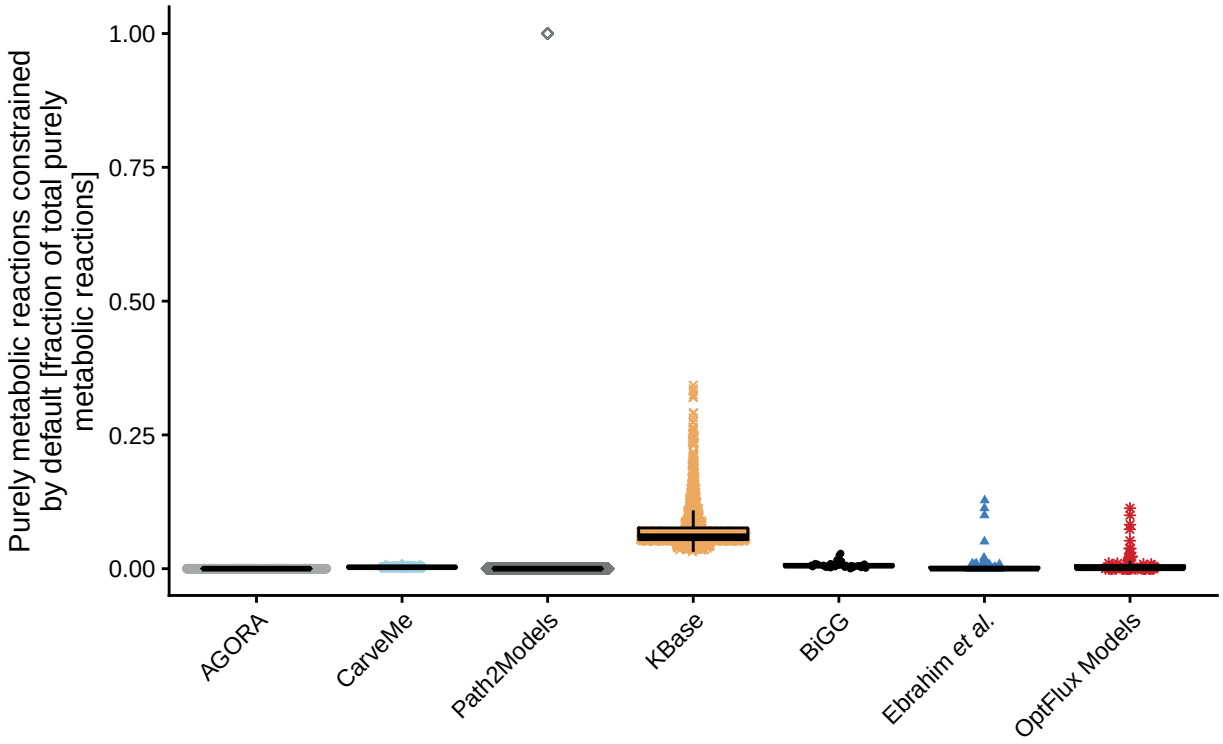

Figure S106: Purely Metabolic Reactions with Constraints

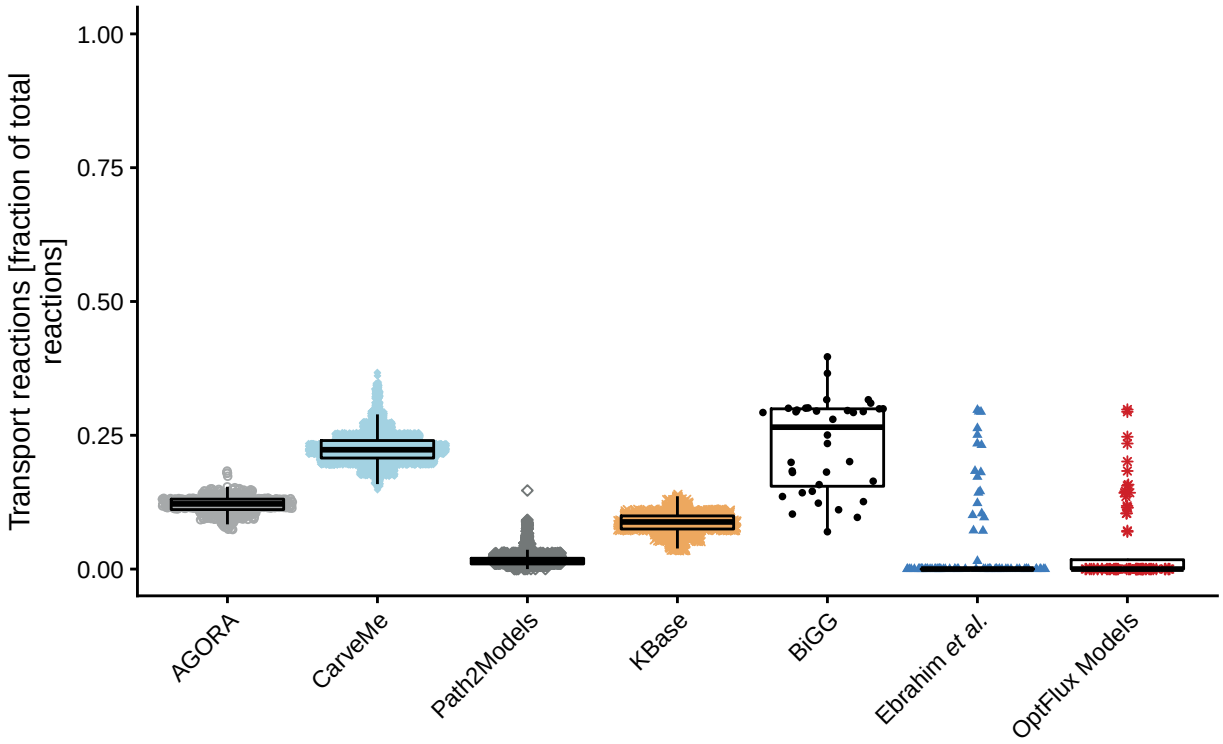

Figure S107: Transport Reactions

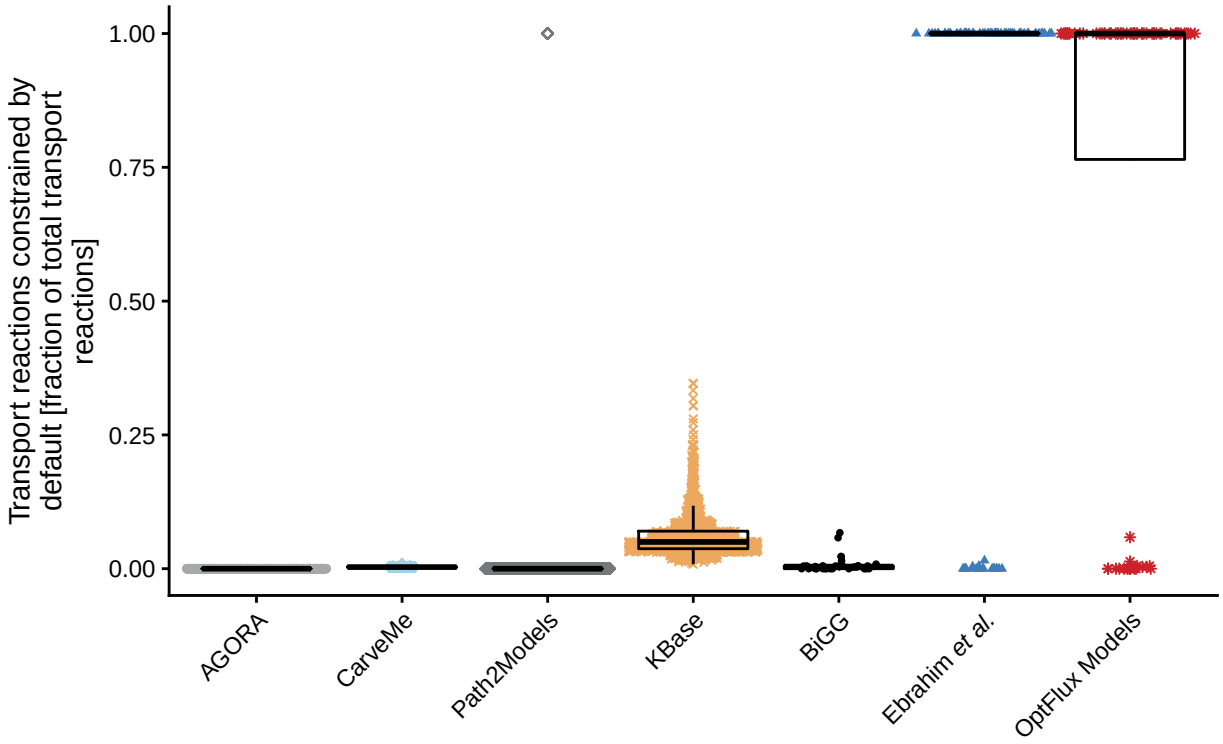

Figure S108: Transport Reactions with Constraints

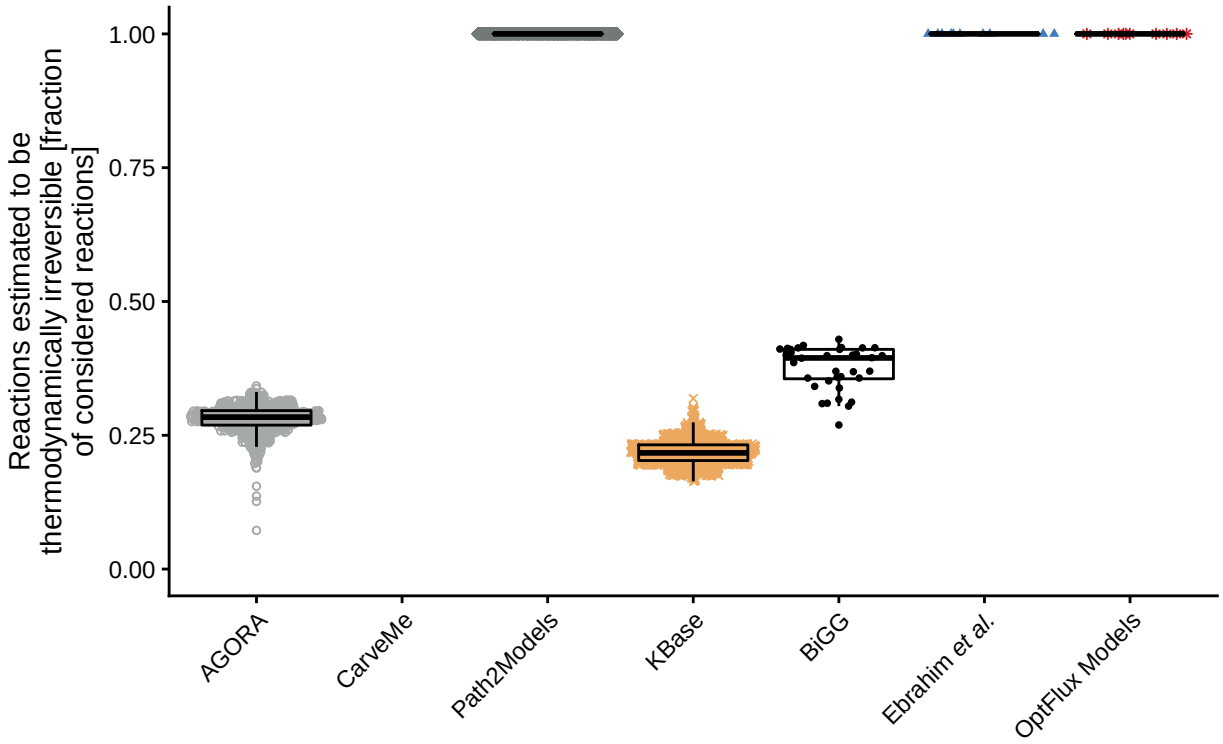

Figure S109: Thermodynamic Reversibility of Purely Metabolic Reactions

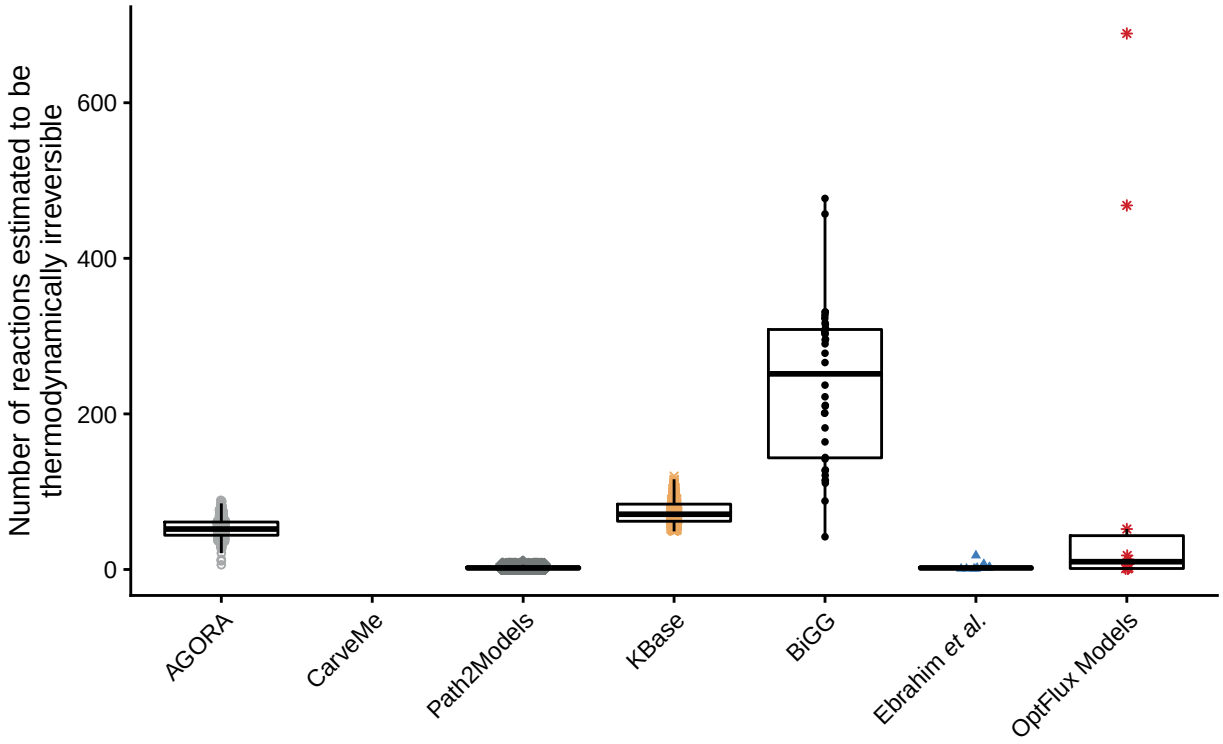

Figure S110: Thermodynamic Reversibility of Purely Metabolic Reactions

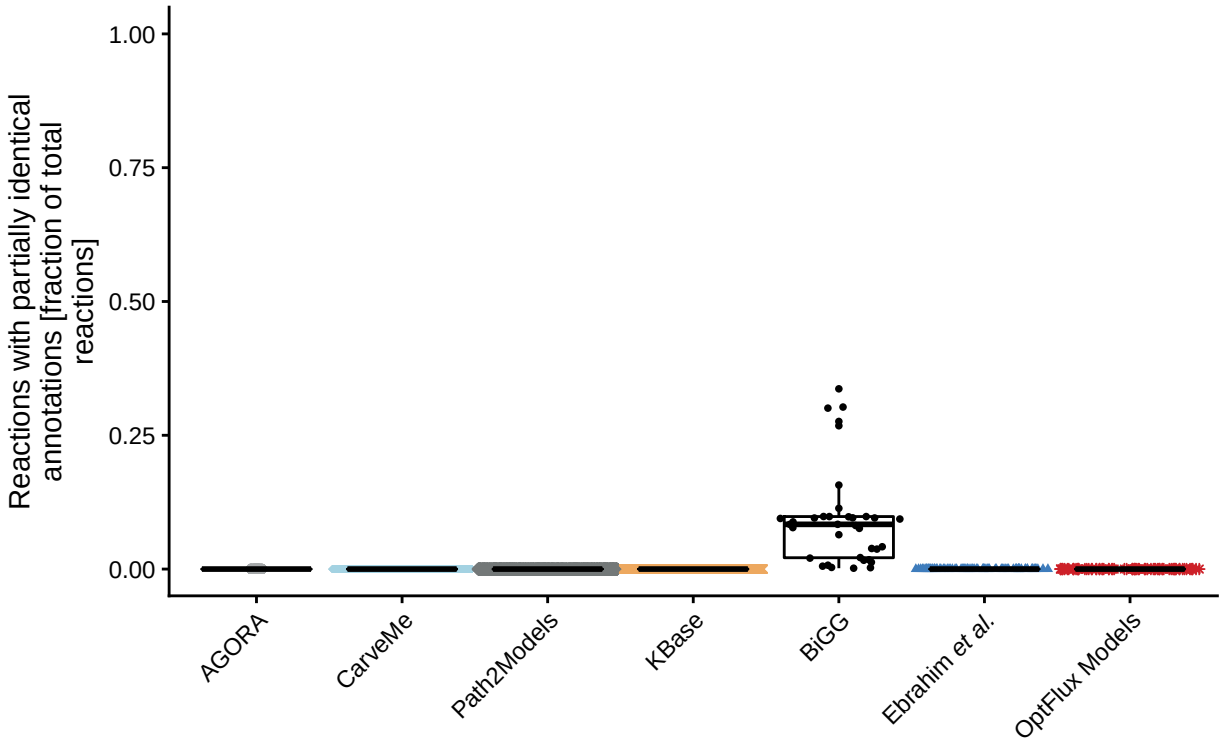

Figure S111: Reactions with Partially Identical Annotations

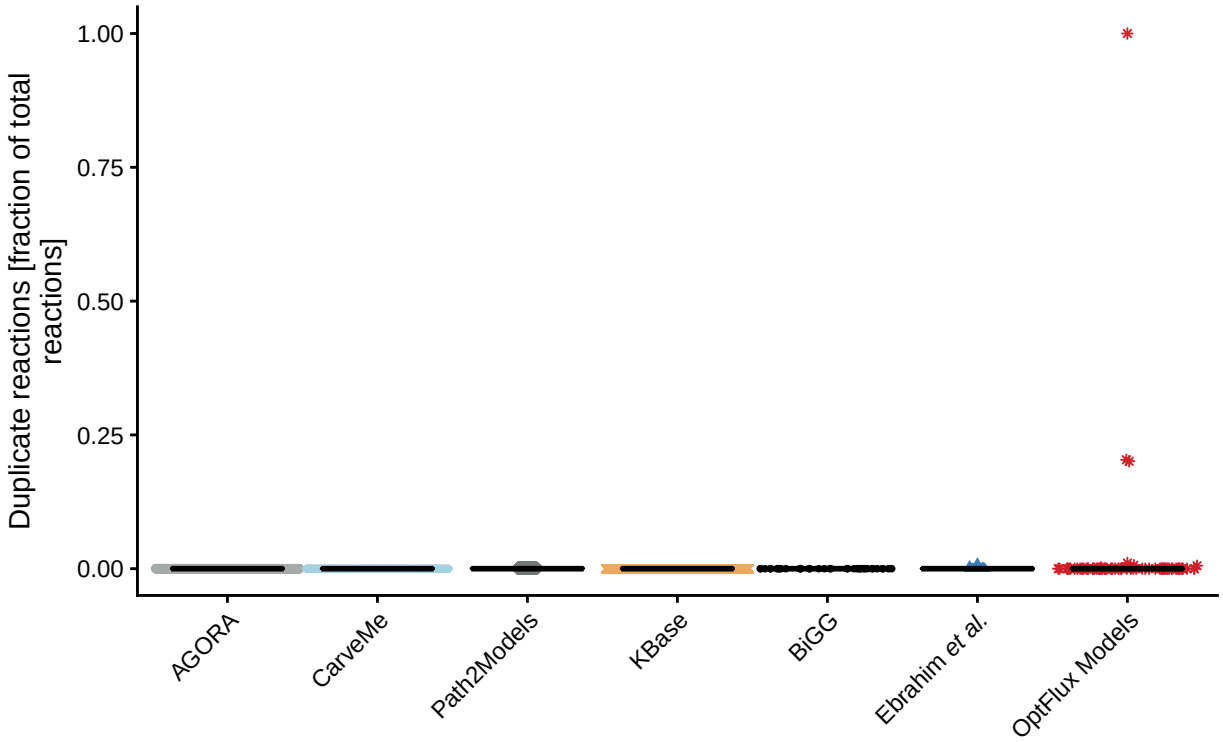

Figure S112: Duplicate Reactions

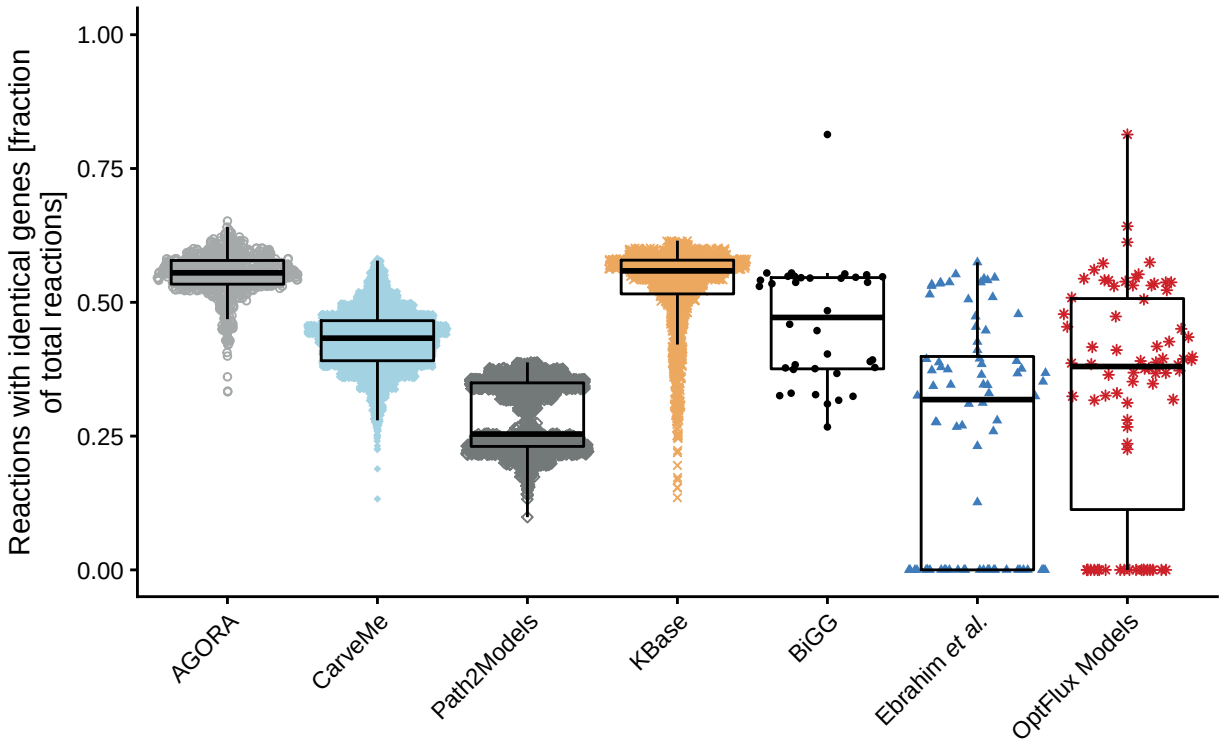

Figure S113: Reactions with Identical Genes

3.4.5 Gene-Protein-Reaction (GPR) Association

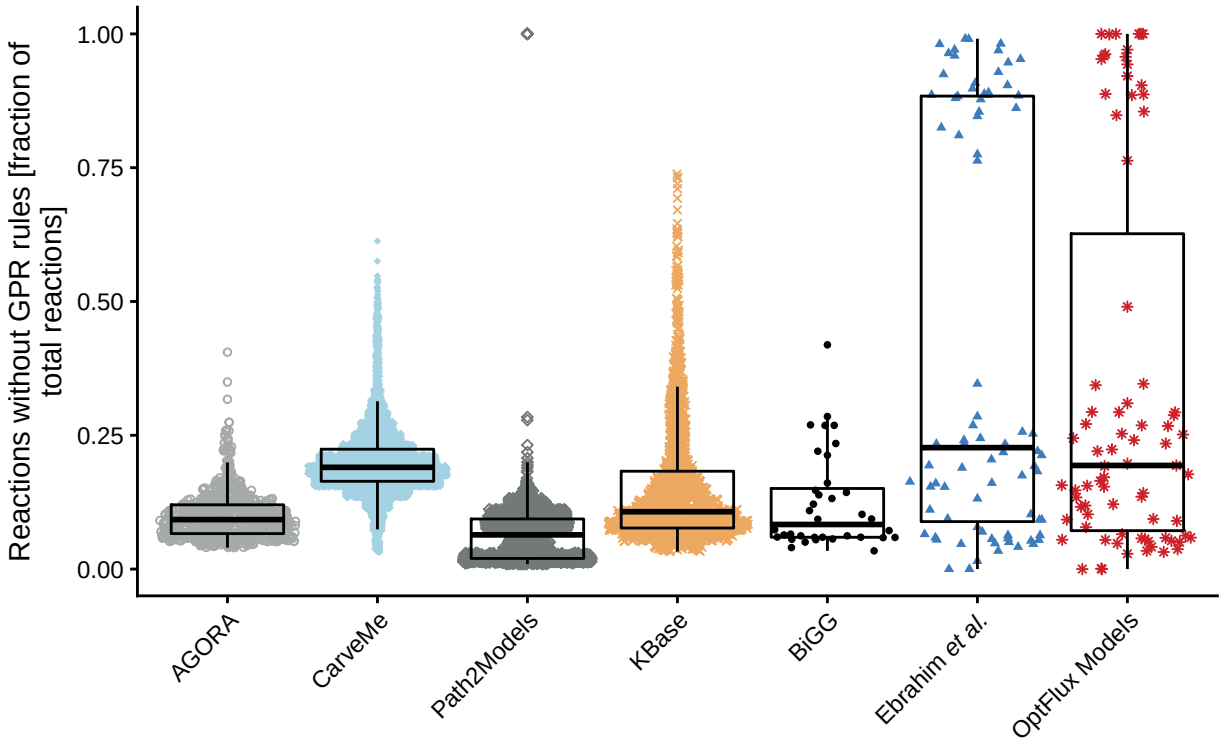

Figure S114: Reactions without GPR

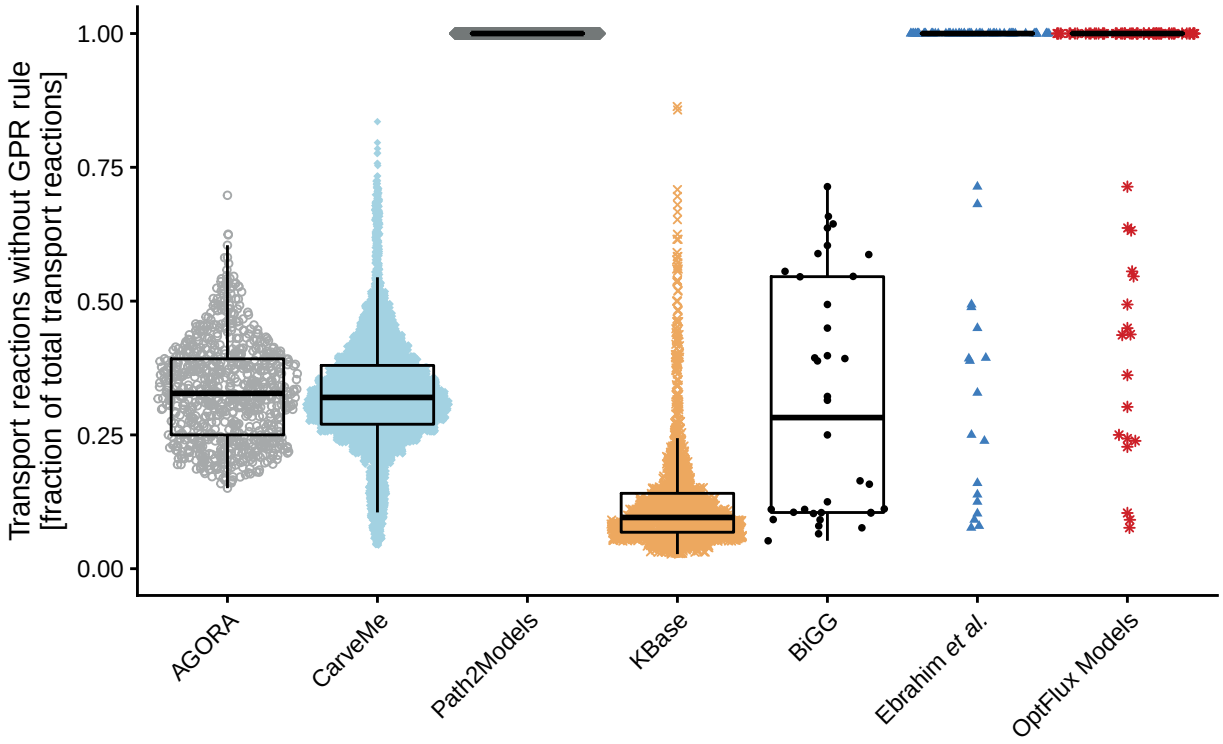

Figure S115: Fraction of Transport Reactions without GPR

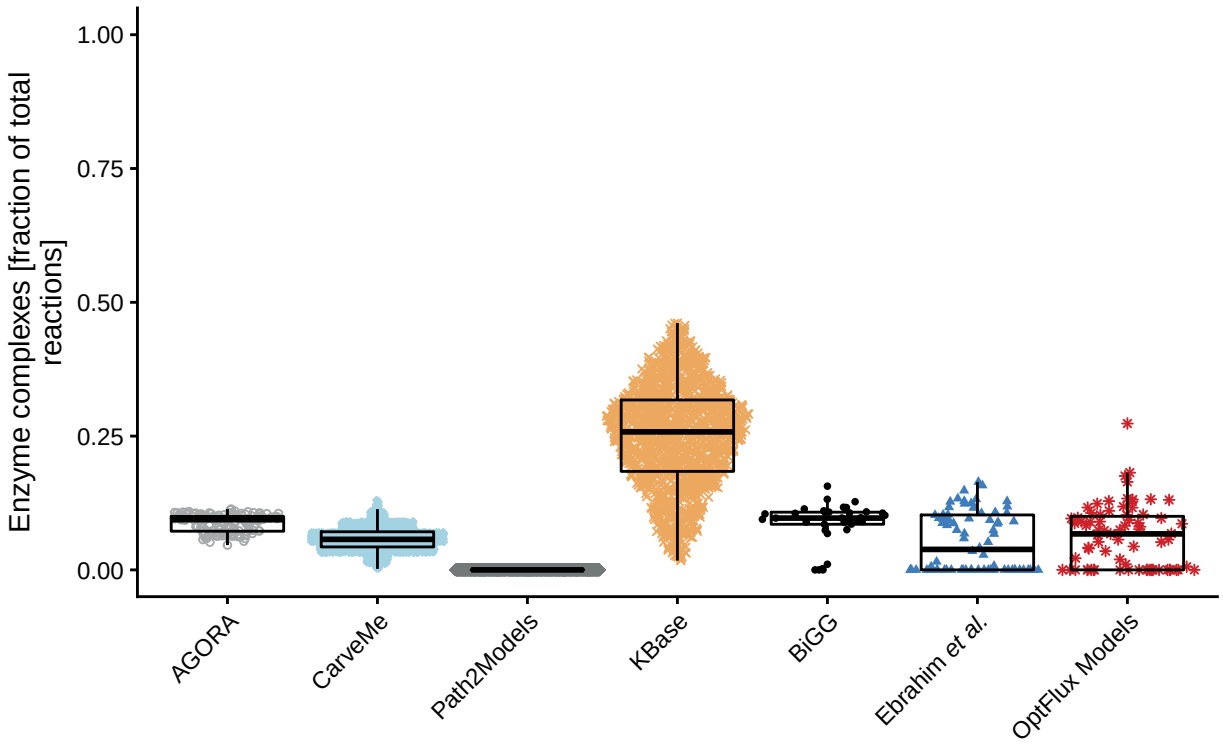

Figure S116: Enzyme Complexes

3.4.6 Biomass

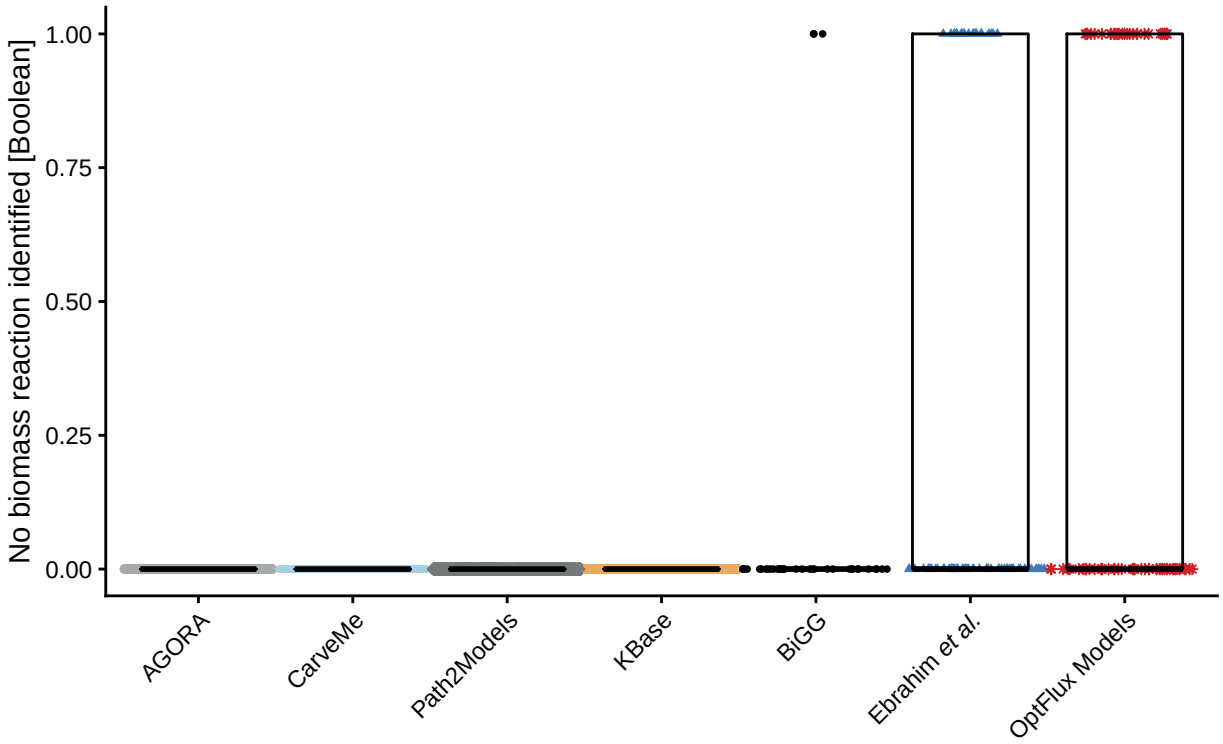

Figure S117: Biomass Reactions Identified

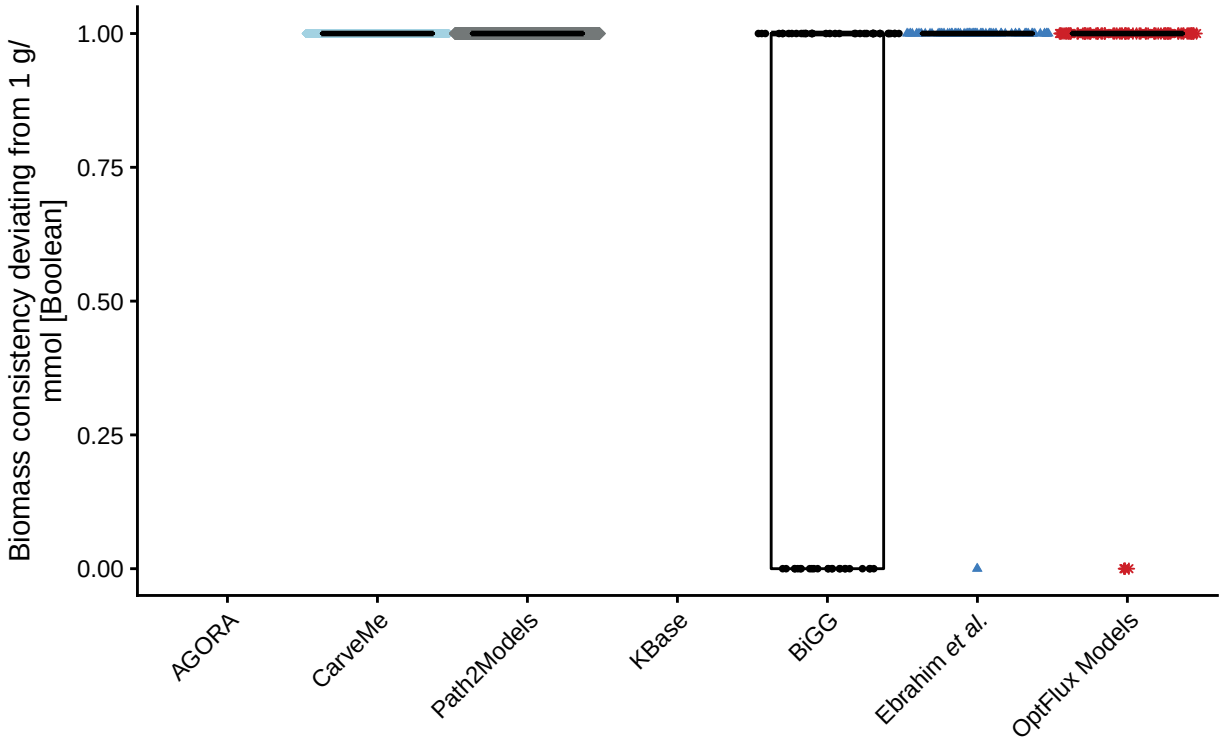

Figure S118: Biomass Consistency

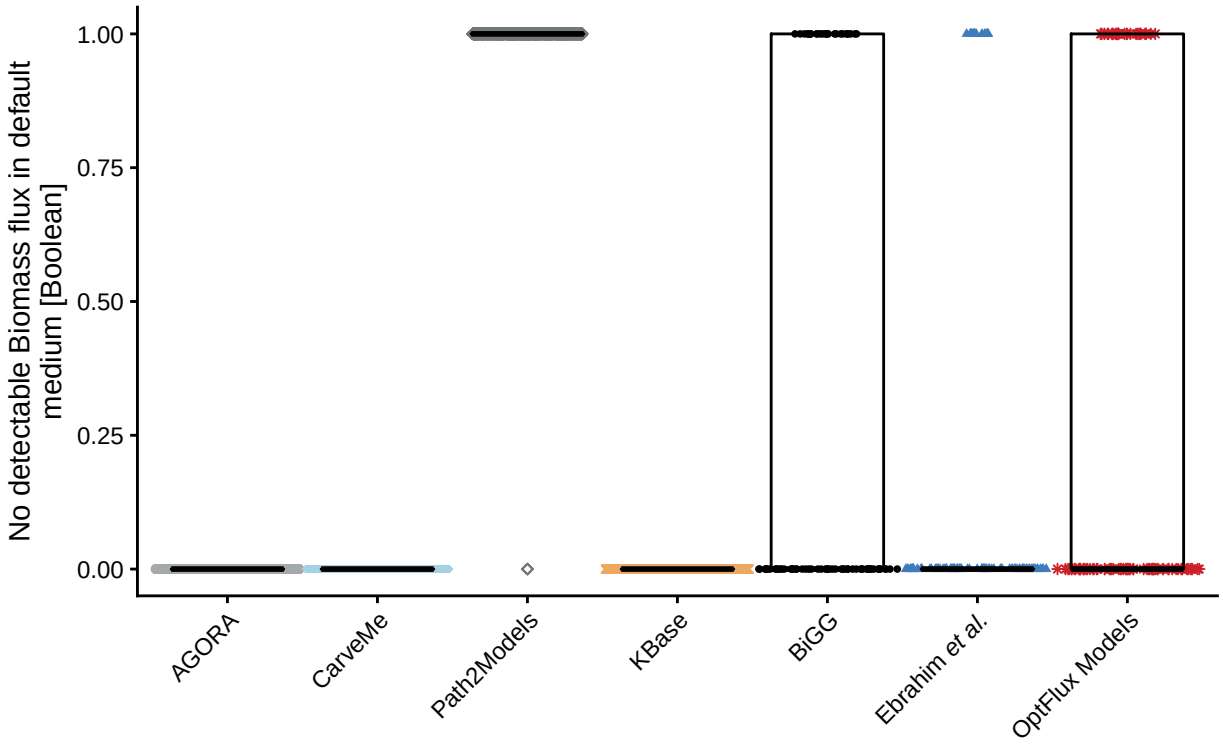

Figure S119: Biomass Production in Default Medium

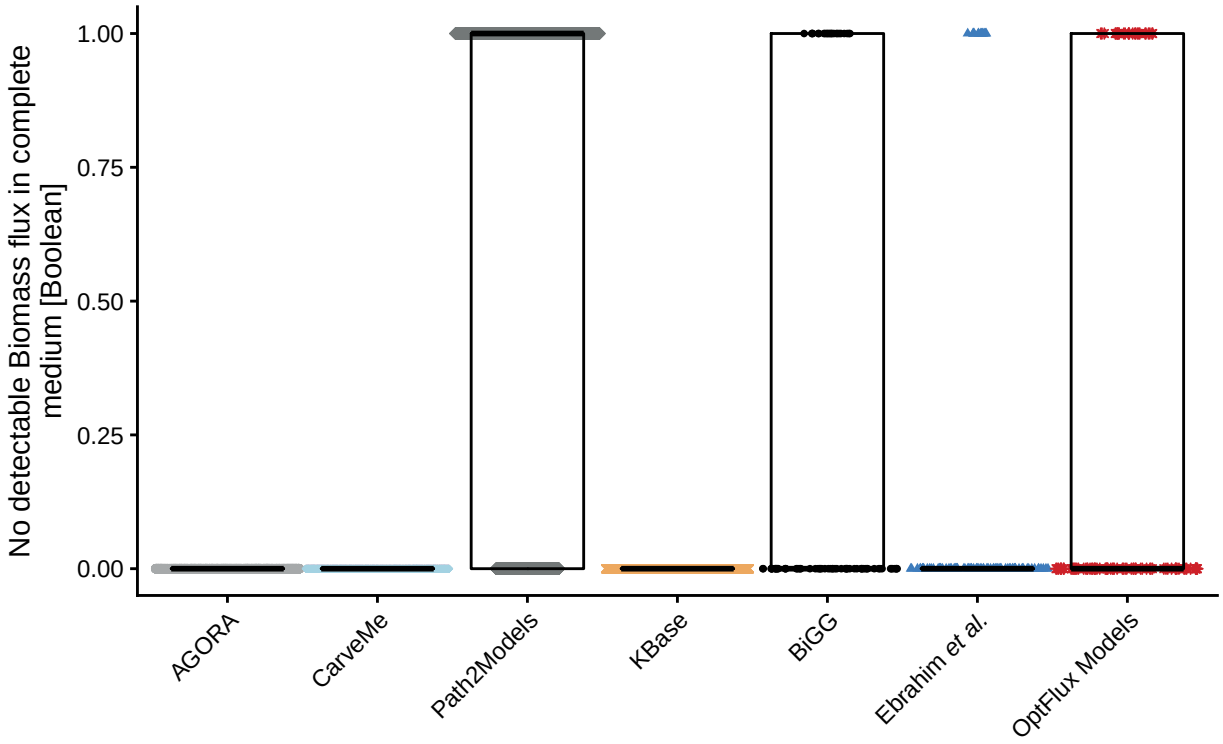

Figure S120: Biomass Production in Complete Medium

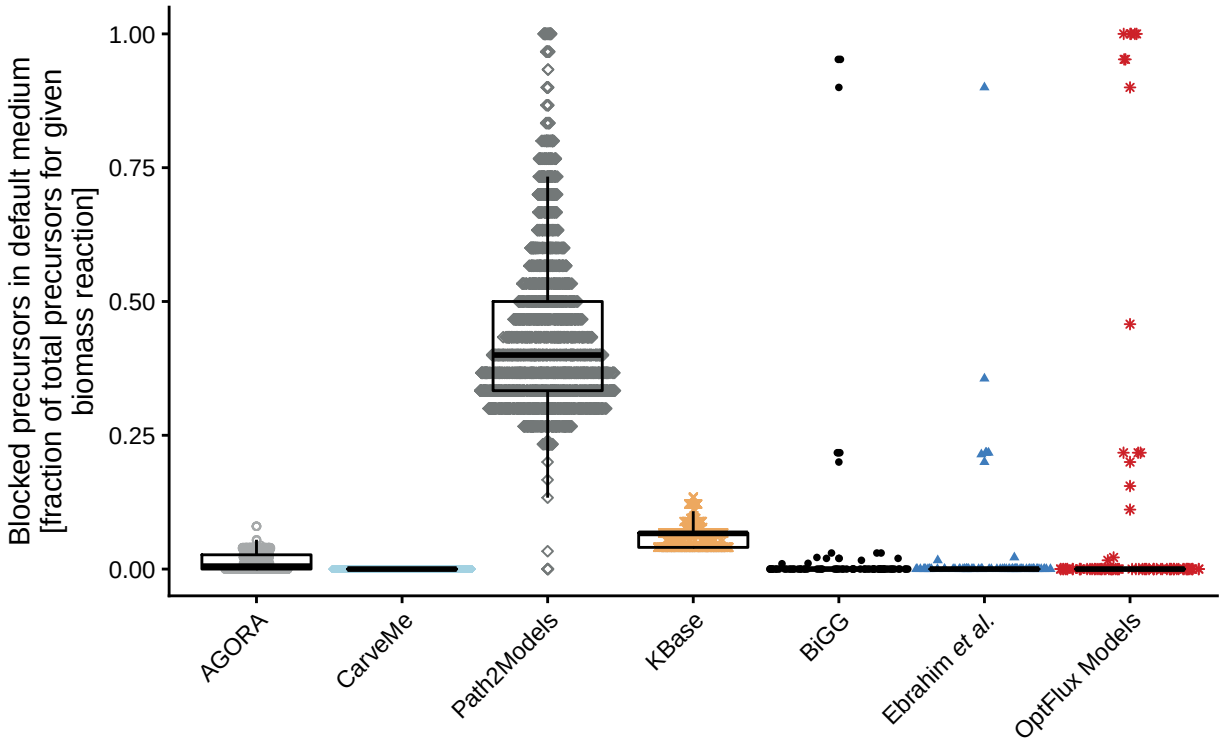

Figure S121: Blocked Biomass Precursors in Default Medium

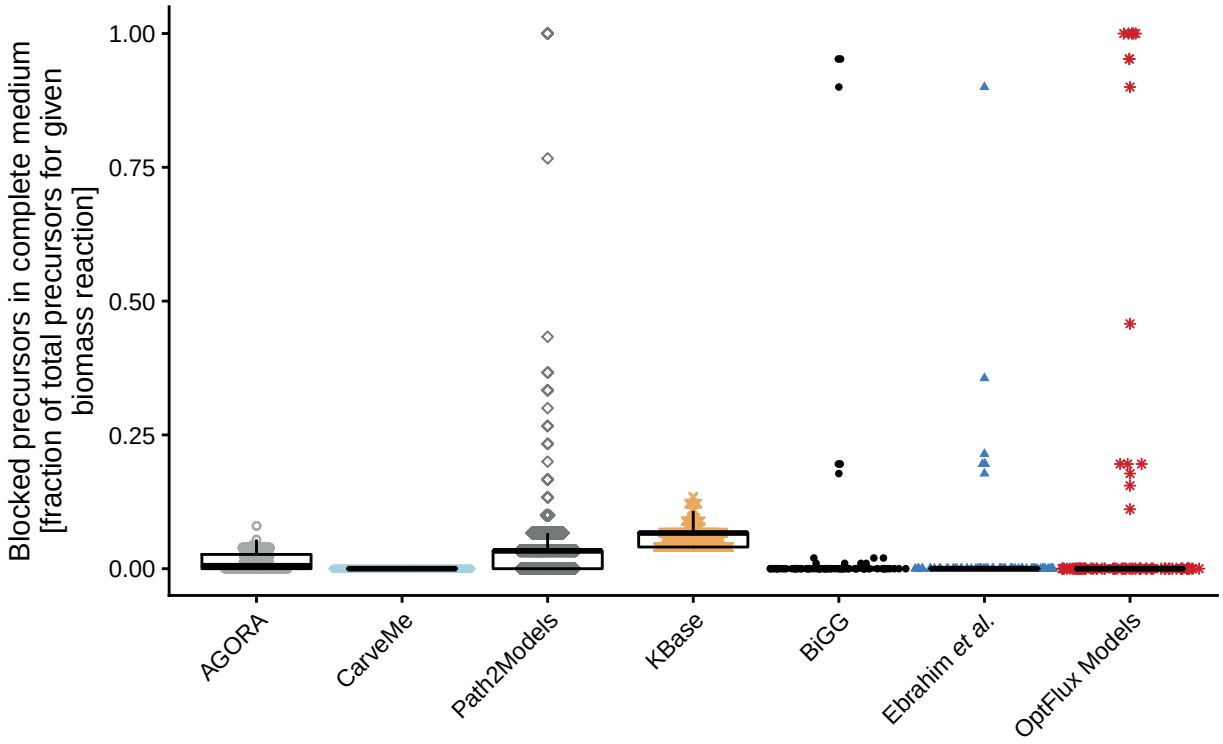

Figure S122: Blocked Biomass Precursors in Complete Medium

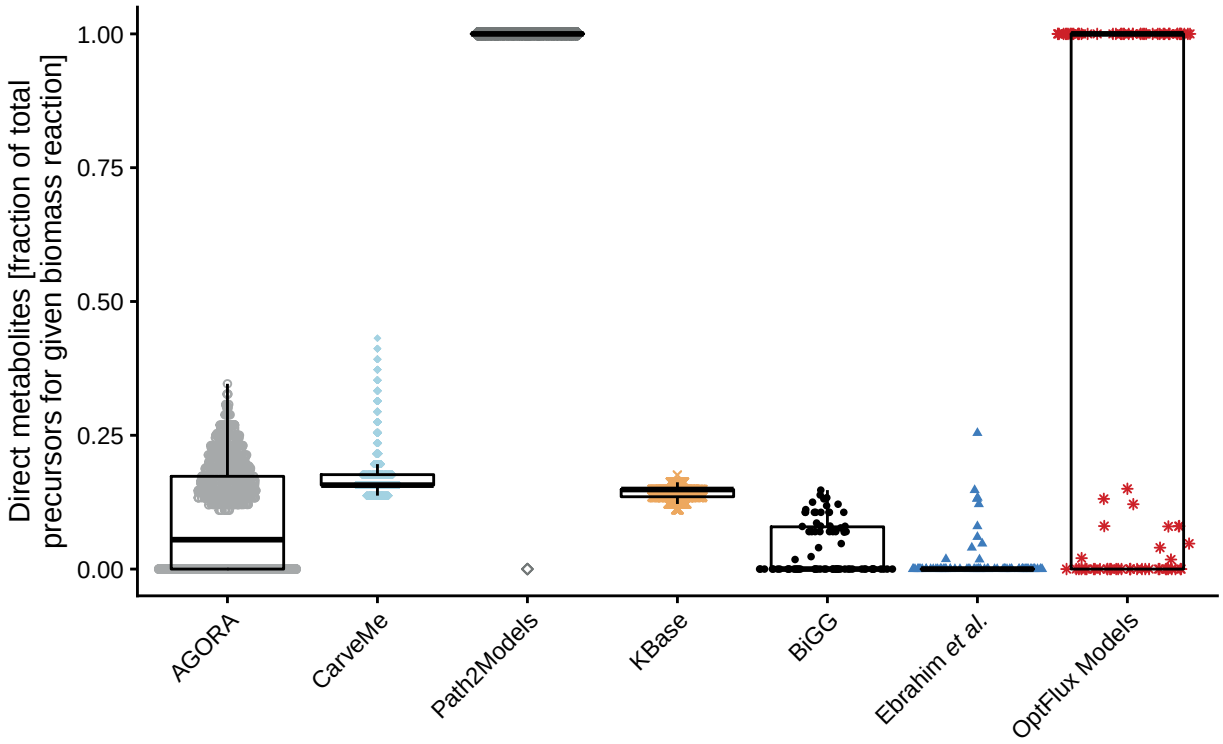

Figure S123: Ratio of Direct Metabolites in Biomass Reaction

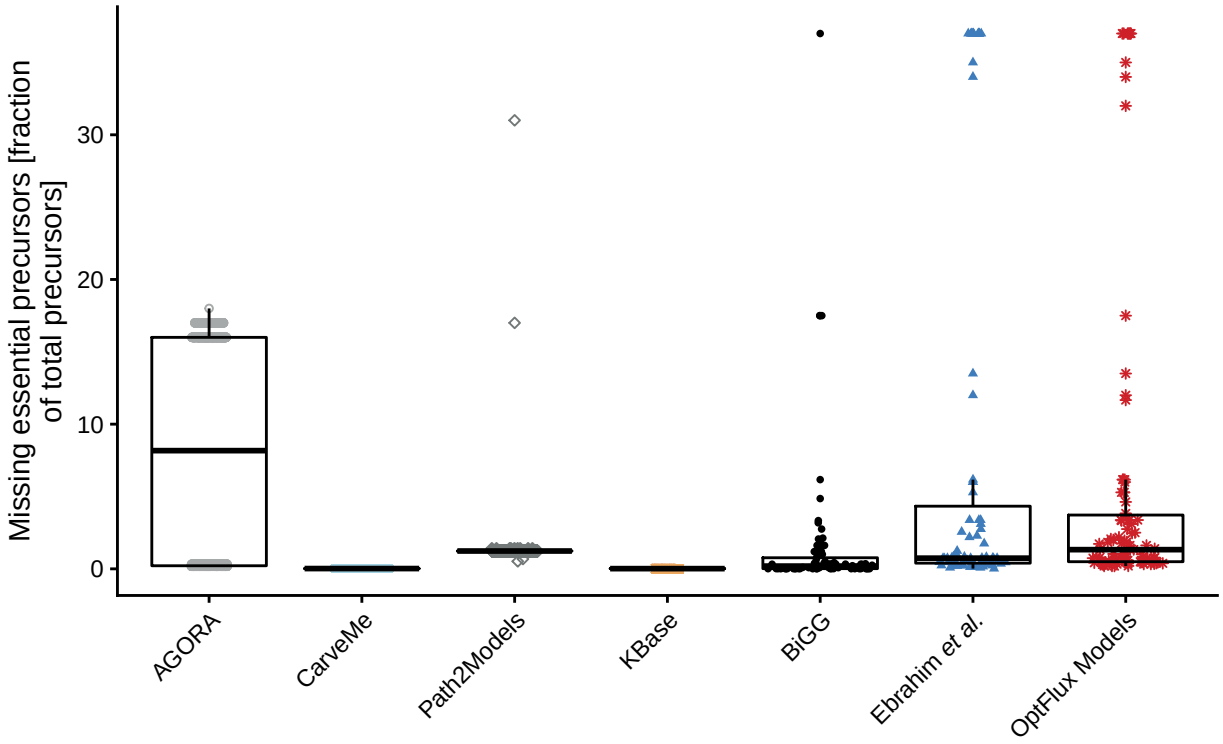

Figure S124: Number of Missing Essential Biomass Precursors

3.4.7 Energy Metabolism

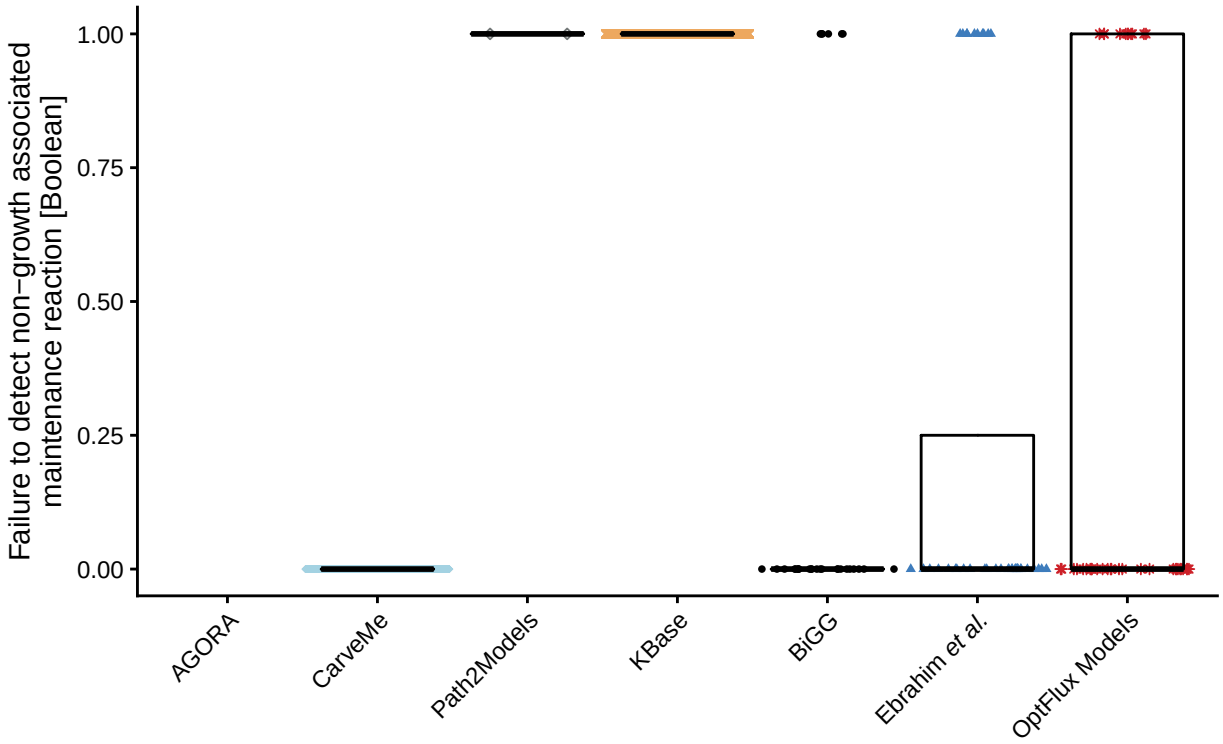

Figure S125: Non-Growth Associated Maintenance Reaction

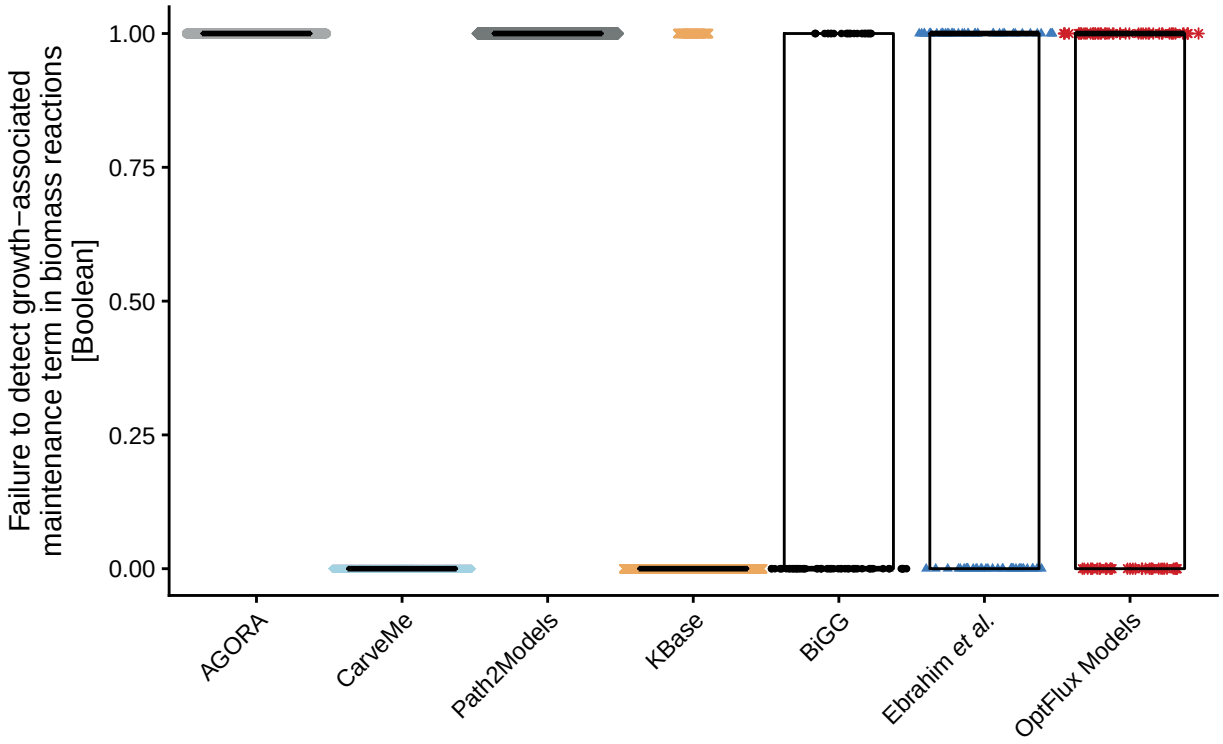

Figure S126: Growth-associated Maintenance in Biomass Reaction

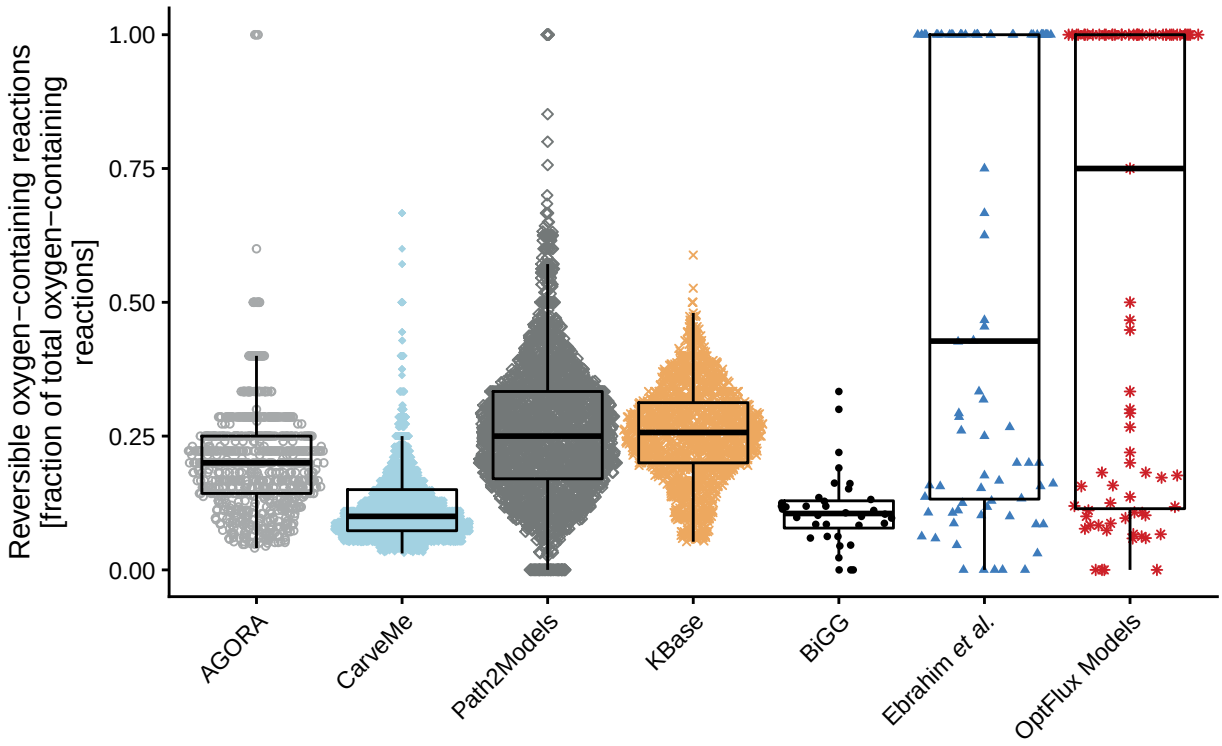

Figure S127: Number of Reversible Oxygen-Containing Reactions

3.4.8 Network Topology

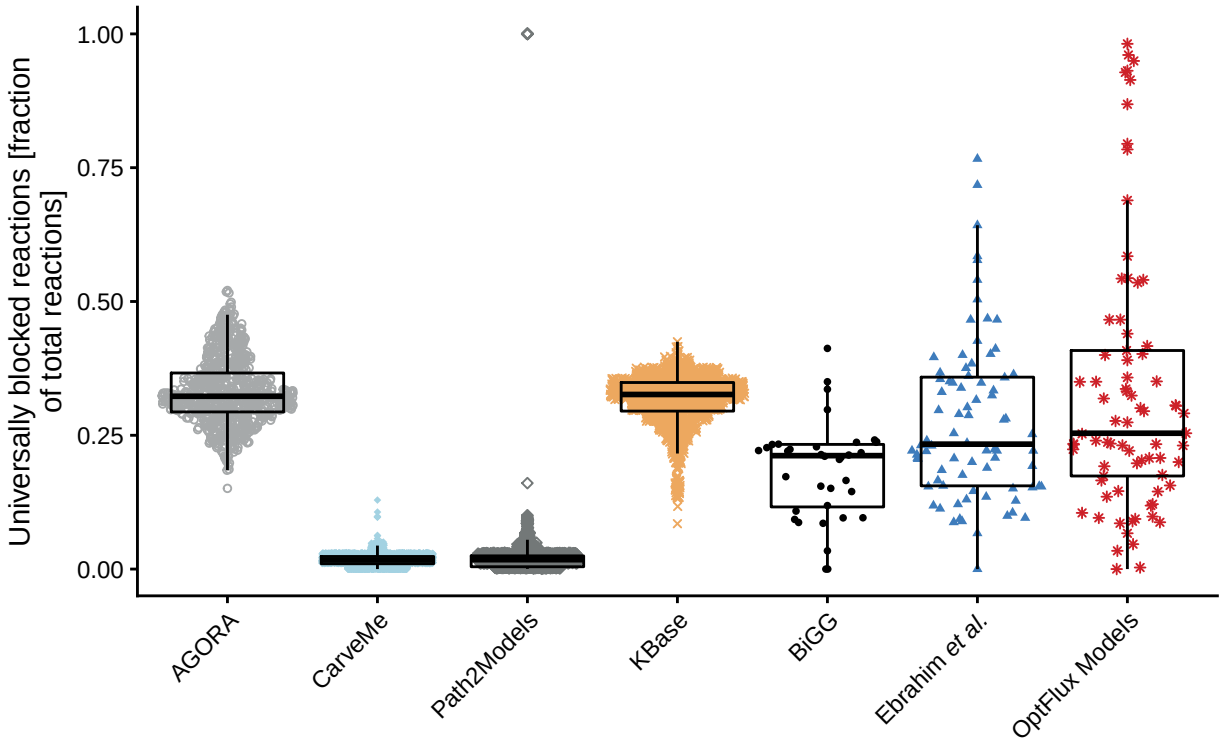

Figure S128: Universally Blocked Reactions

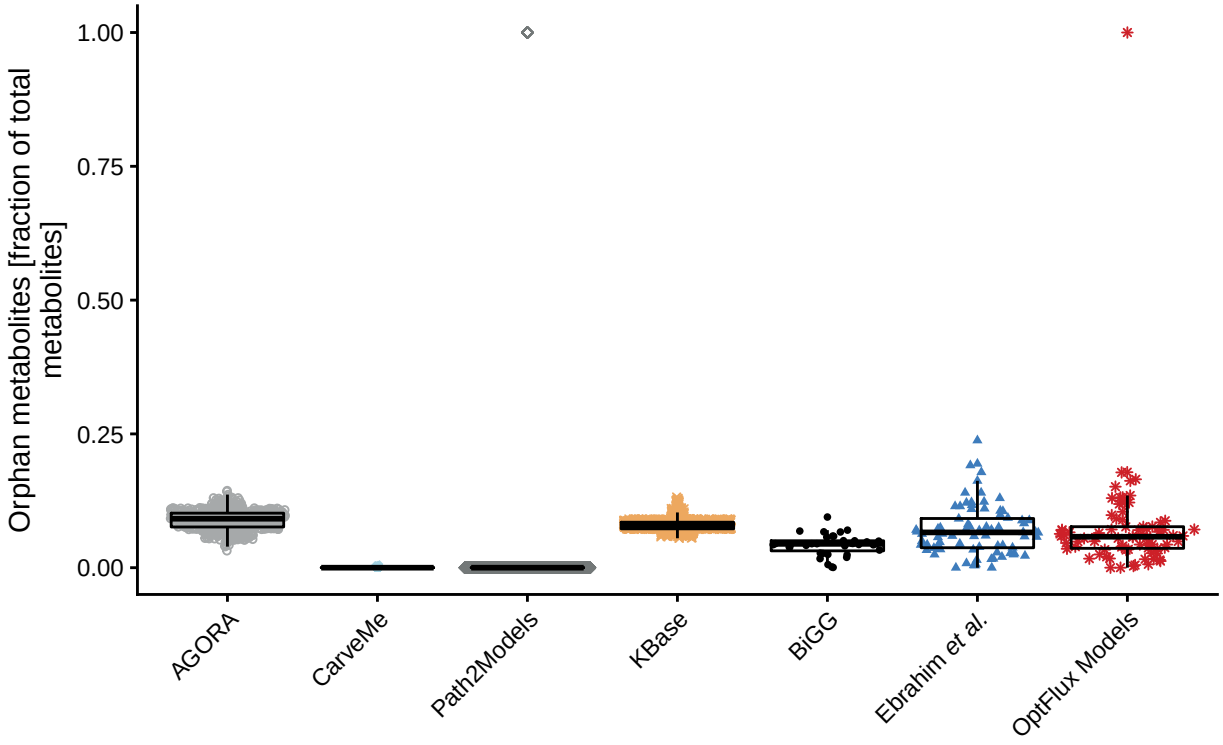

Figure S129: Orphan Metabolites

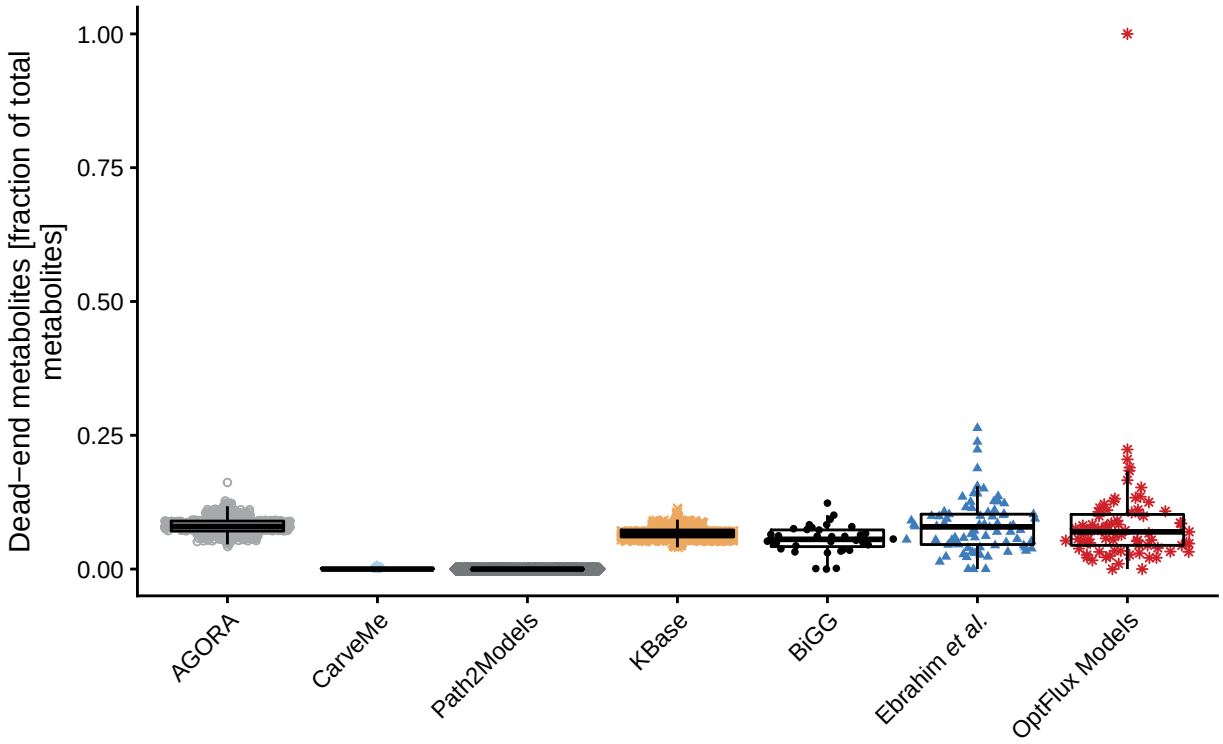

Figure S130: Dead-end Metabolites

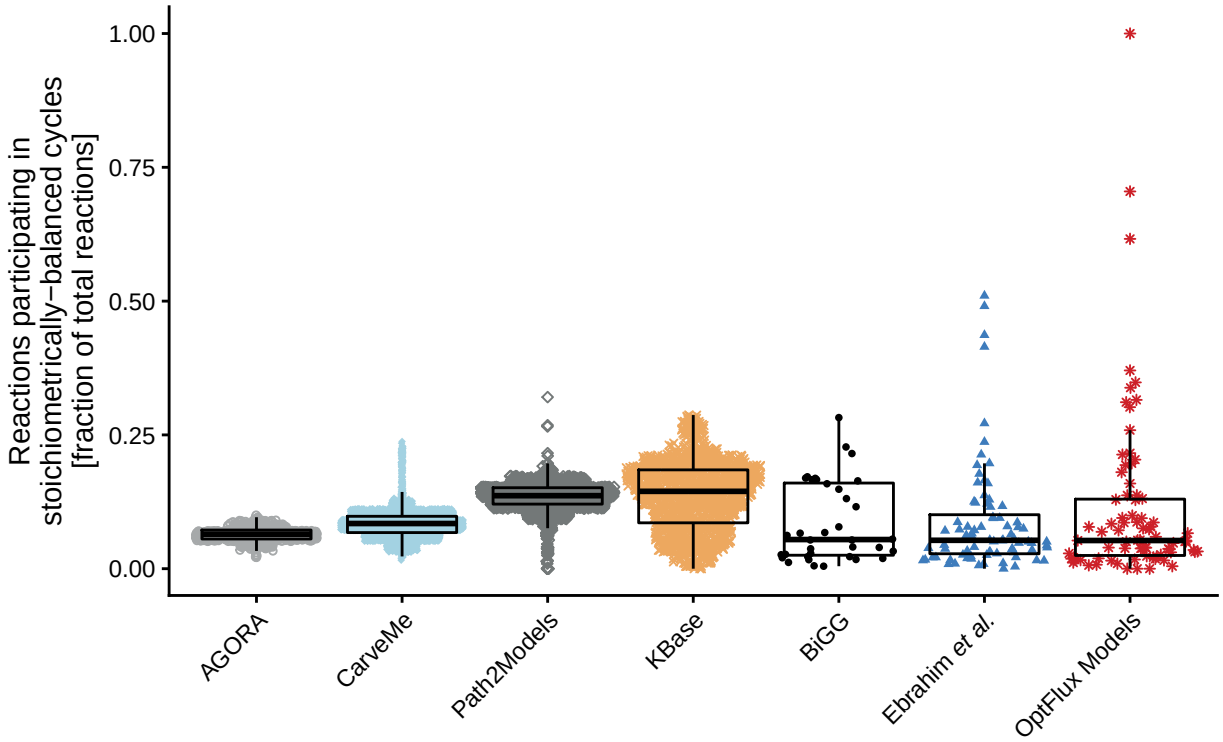

Figure S131: Stoichiometrically Balanced Cycles

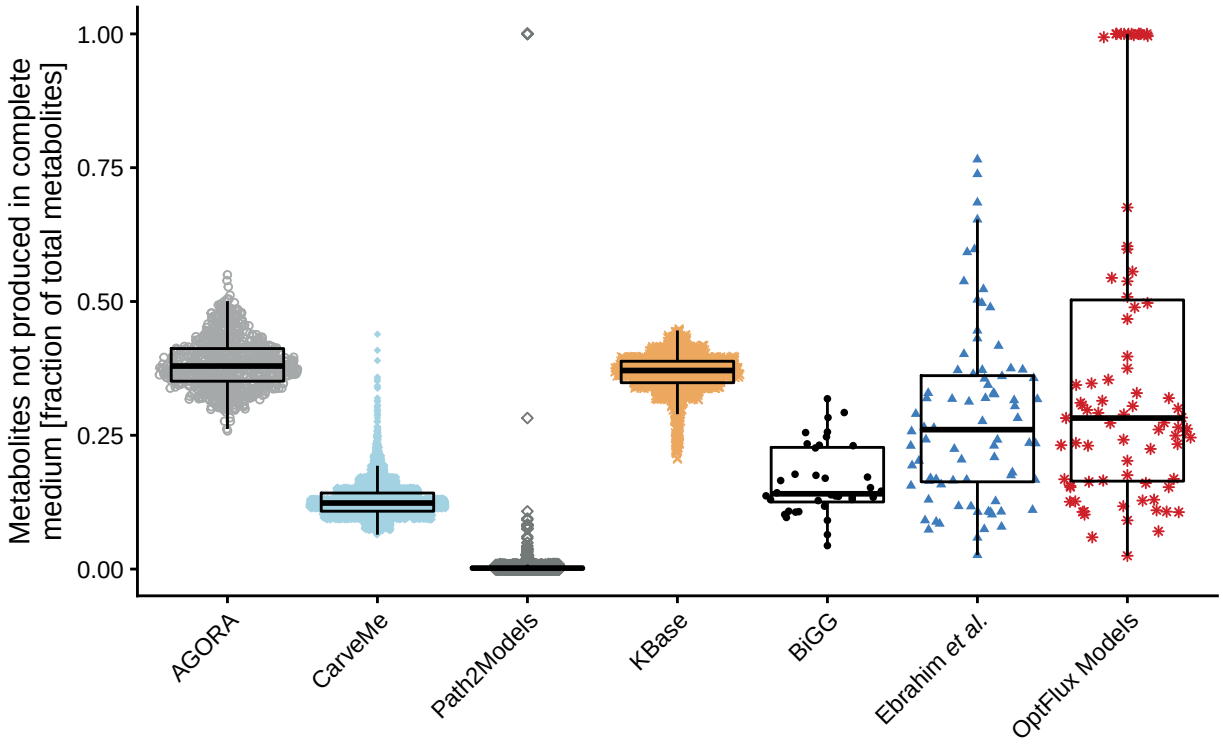

Figure S132: Metabolite Production in Complete Medium

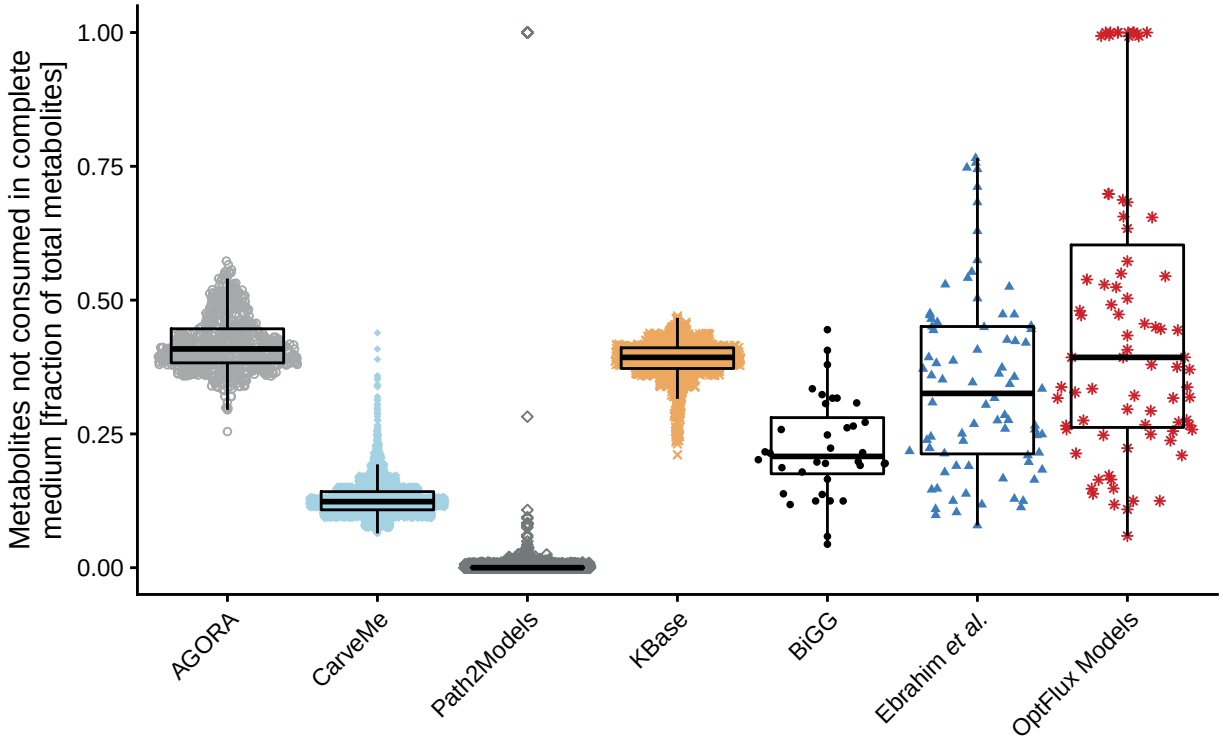

Figure S133: Metabolite Consumption in Complete Medium

3.4.9 Matrix Conditioning

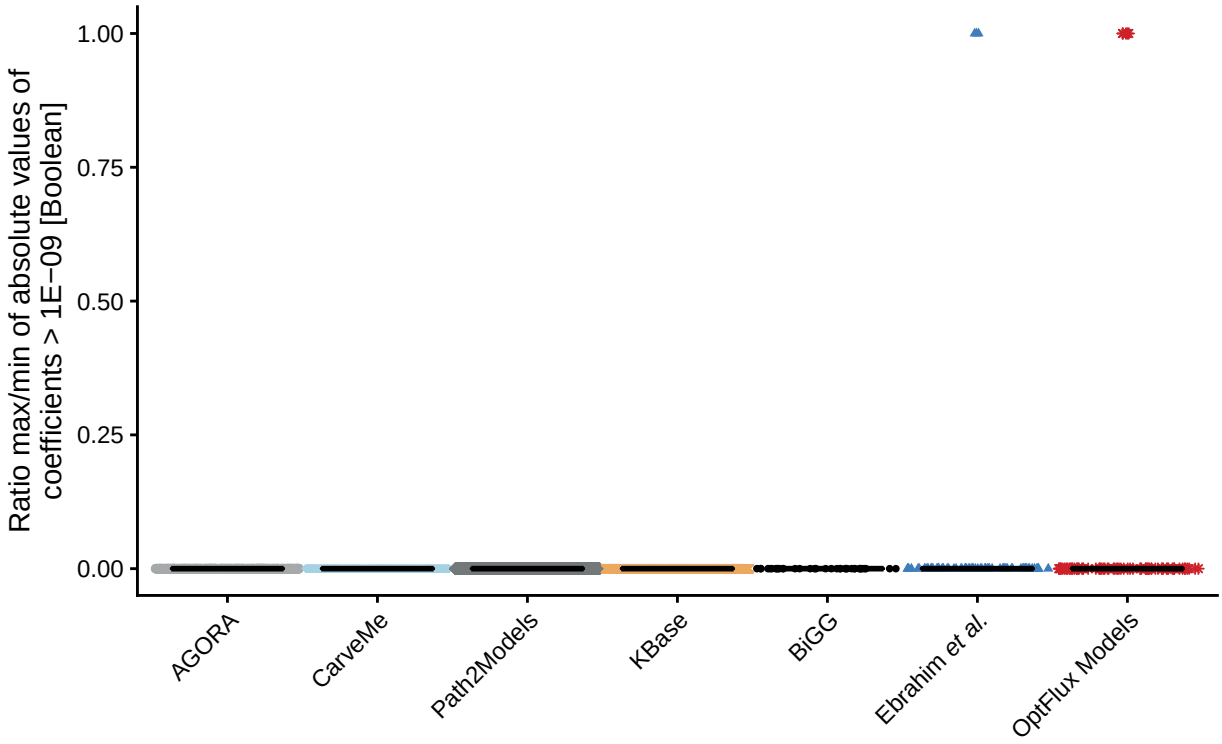

Figure S134: Ratio Min/Max Non-Zero Coefficients

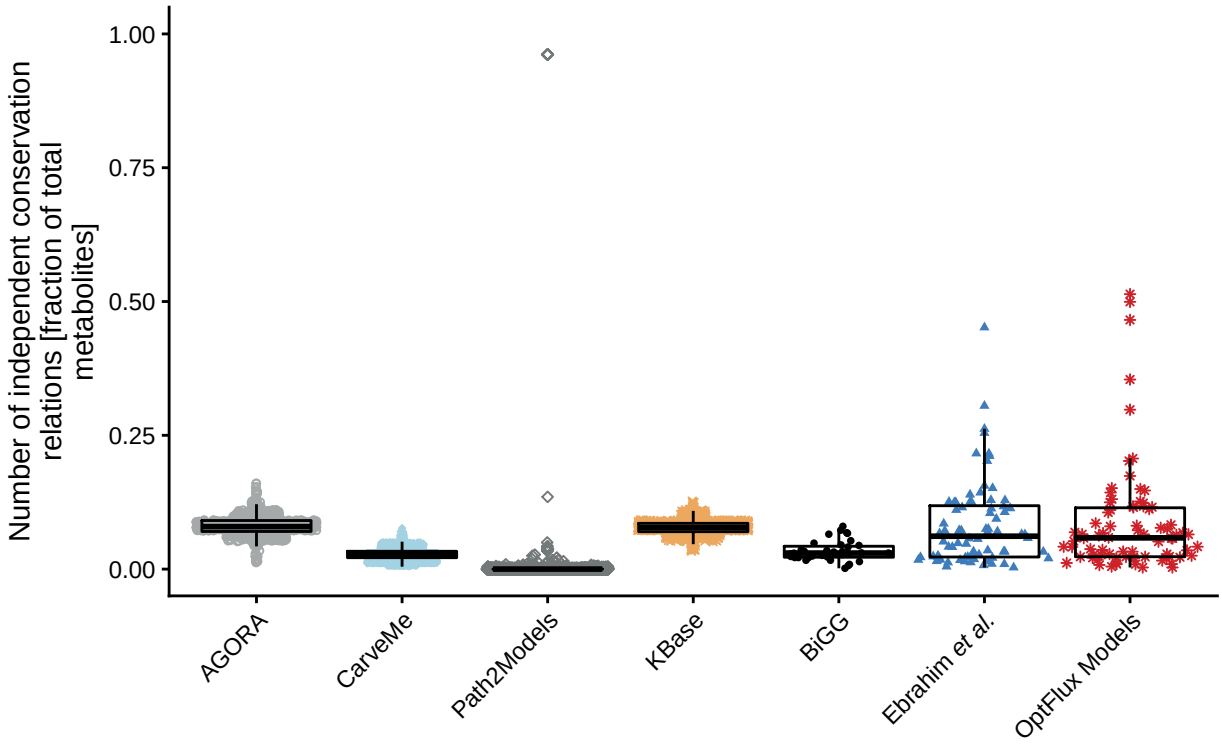

Figure S135: Independent Conservation Relations

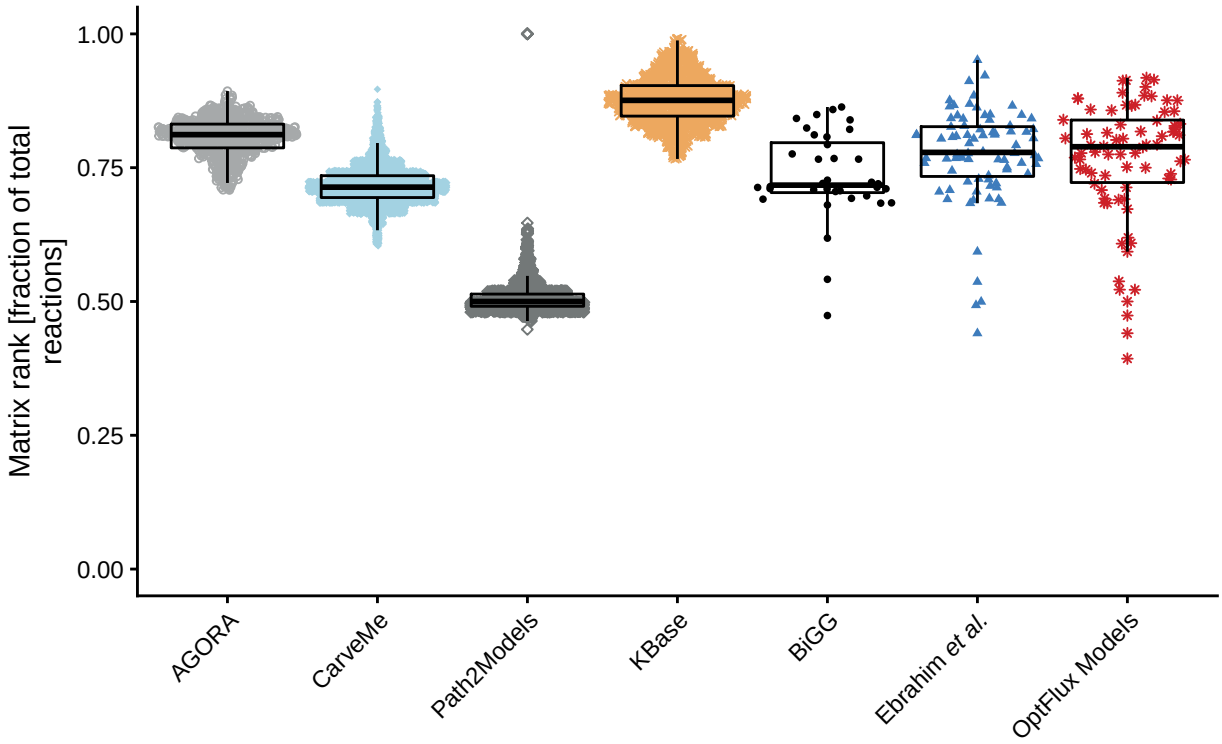

Figure S136: Rank

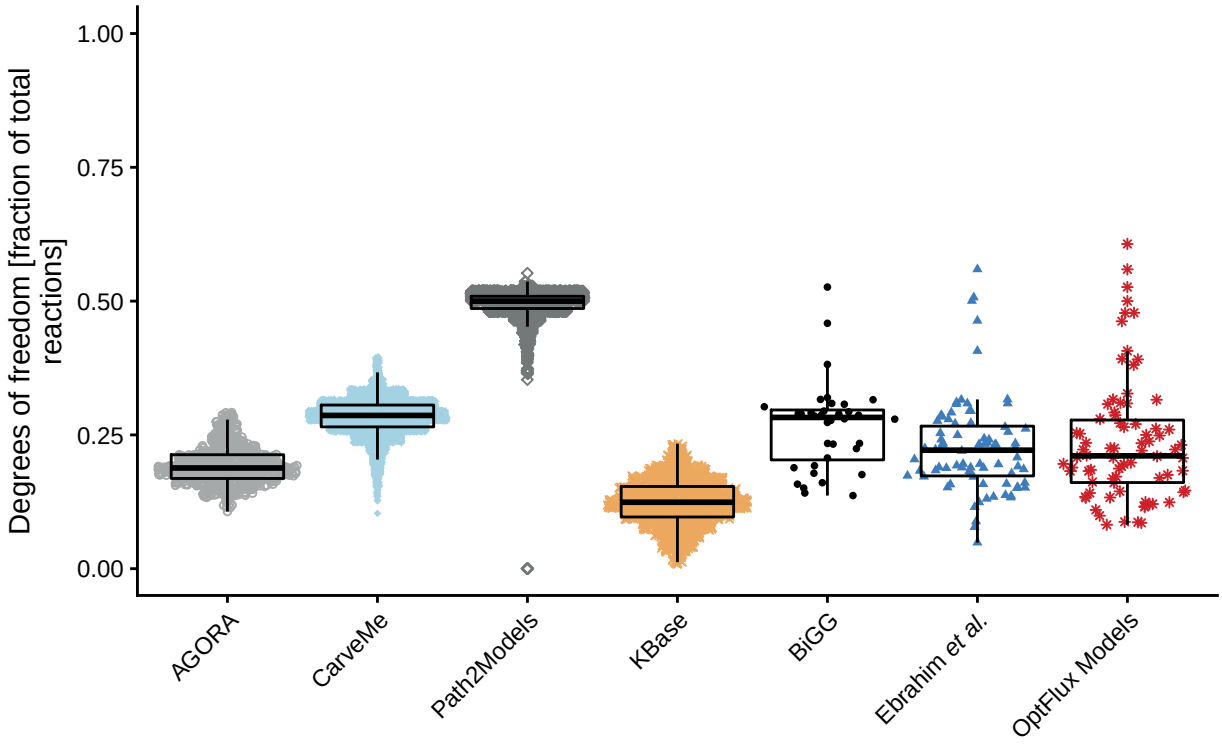

Figure S137: Degrees of Freedom

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Louppe, Gilles, Louis Wehenkel, Antonio Sutera, and Pierre Geurts. 2013. "Understanding Variable Importances in Forests of Randomized Trees." In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, 431–39. NIPS'13. USA: Curran Associates Inc. http://dl.acm.org/citation.cfm?id=2999611.2999660.