

Supplemental Figures Depicting the Comparative Analysis of Model Collections

Contents

1	Tested models	1
2	Clustering	2
3	Test Suite	6
3.1	Summary of Observations	6
3.2	Scores	6
3.3	Independent Section	8
3.3.1	Consistency	8
3.3.2	Annotation - Metabolites	14
3.3.3	Annotation - Reactions	29
3.3.4	Annotation - Genes	42
3.3.5	Annotation - SBO Terms	56
3.4	Specific Section	63
3.4.1	SBML	63
3.4.2	Basic Information	65
3.4.3	Metabolite Information	69
3.4.4	Reaction Information	72
3.4.5	Gene-Protein-Reaction (GPR) Association	77
3.4.6	Biomass	79
3.4.7	Energy Metabolism	84
3.4.8	Network Topology	86
3.4.9	Matrix Conditioning	90

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 <i>et al.</i> †	83	80	96.4

* 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.

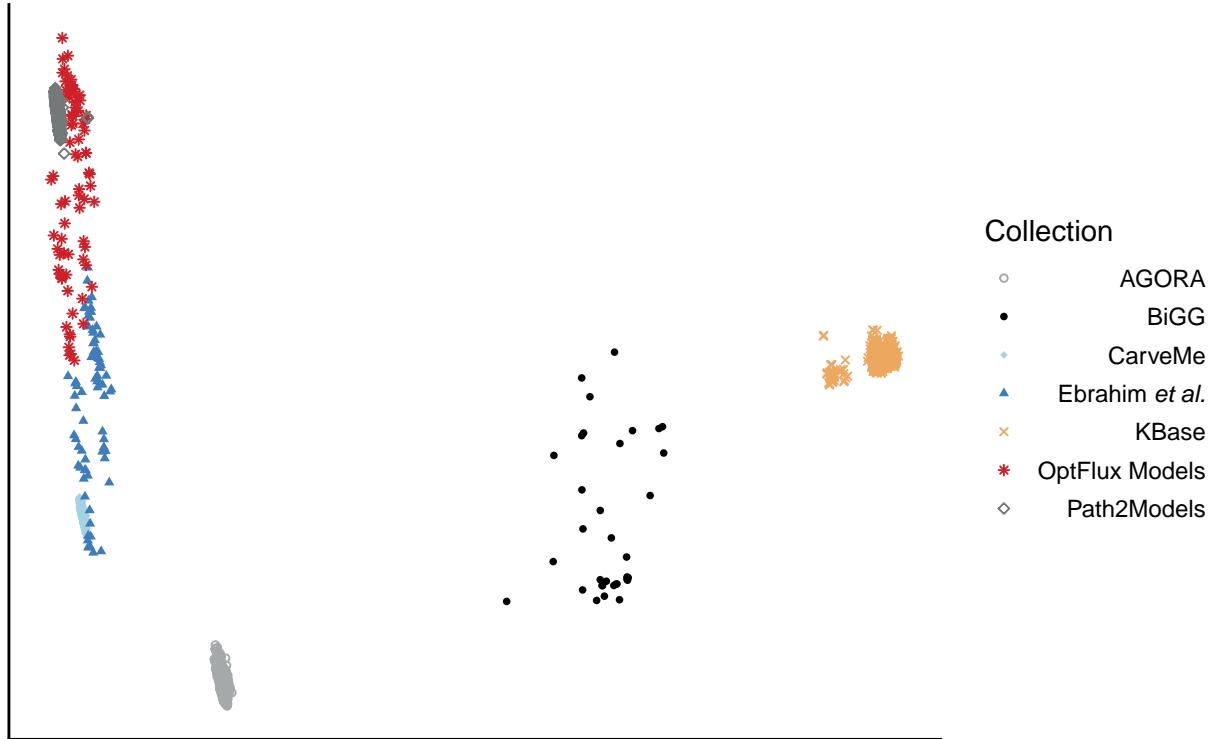


Figure S1: Depicted are the first two components of a principal components analysis of the normalized test features (metrics).

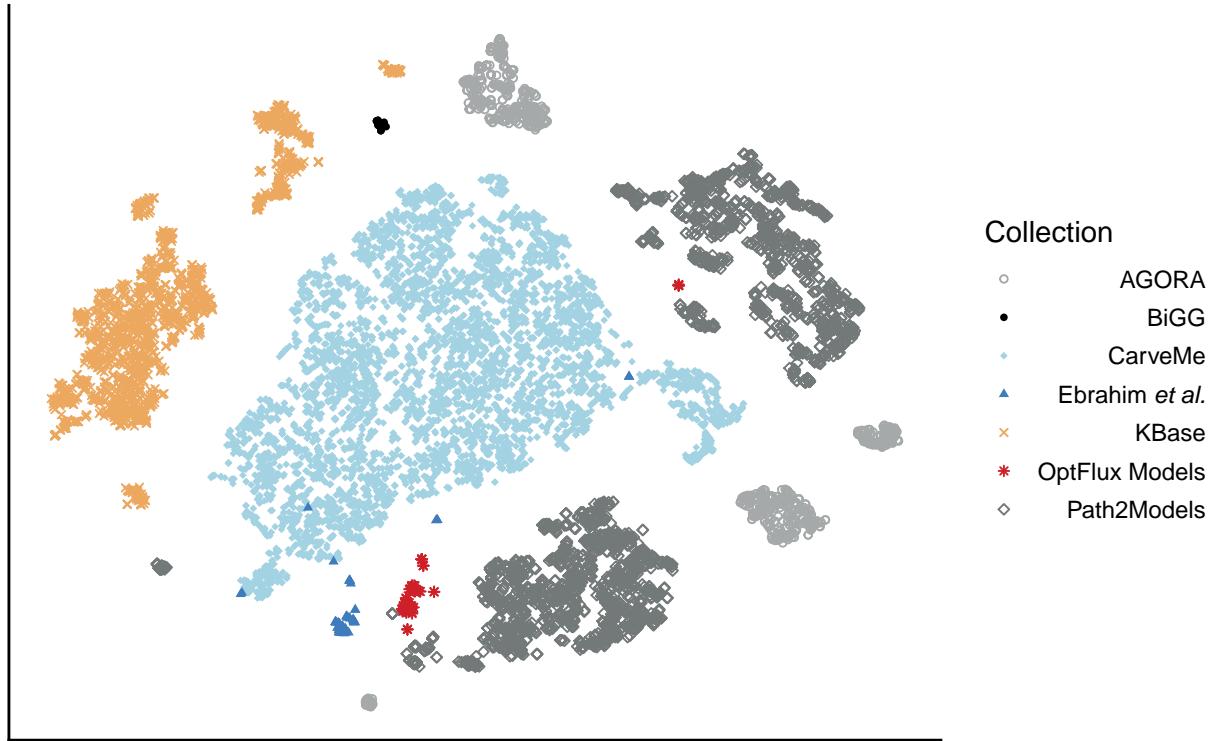


Figure S2: Depicted are the distances between models in higher order space given by the normalized test features reduced to two dimensions using t-SNE.

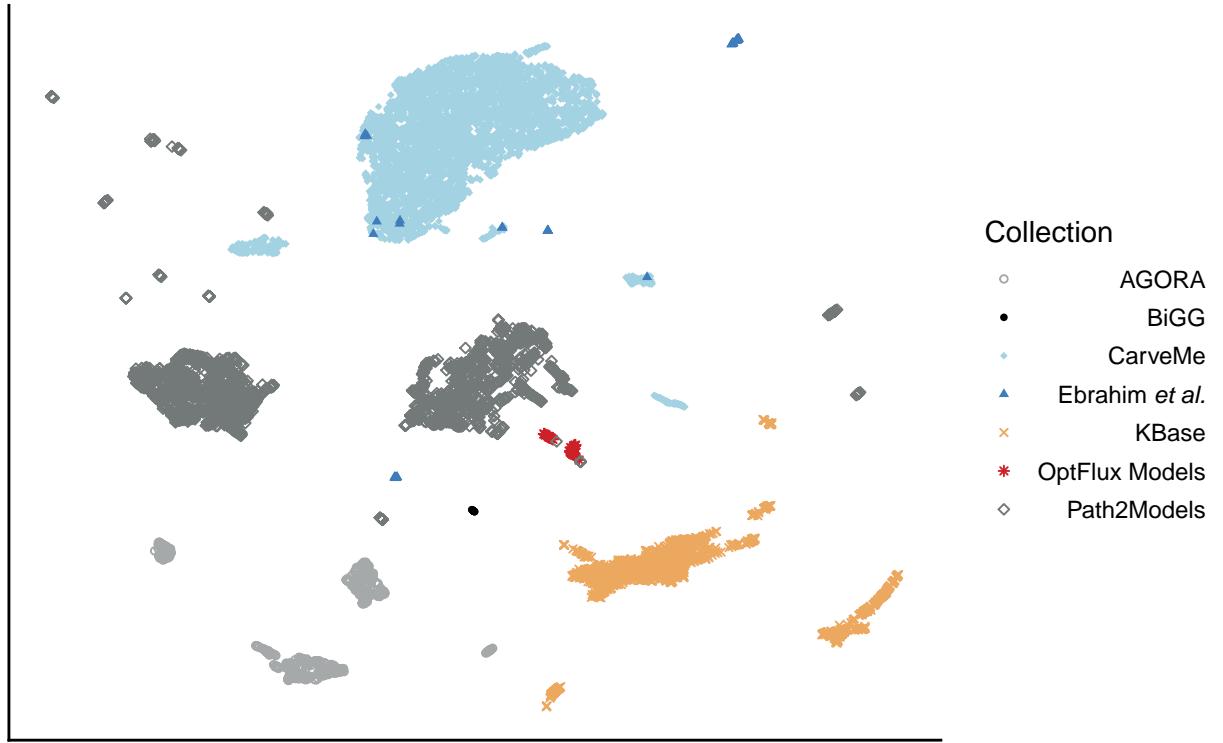


Figure S3: Depicted are the distances between models in higher order space given by the normalized test features reduced to two dimensions using t-SNE.

3 Test Suite

3.1 Summary of Observations

- SBO terms are only used consistently by models from KBase and BiGG (see Figures [S79](#)).
- Models from Path2Models and Opflux Models are formatted in legacy SBML (< Level 3, Version 1) without FBC package (Figures [S91](#) & [S92](#)).
- 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](#)).
- On biomass:
 - Only for a minority of models in BiGG, Ebrahim *et al.* and OptFlux Models memote could not identify a biomass reaction (Figure [S116](#)).
 - 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.* (Figure [S117](#)).
 - Models that cannot be simulated using the default or complete medium exist in Path2Models, BiGG, Ebrahim *et al.* and OptFlux Models (Figure [S118](#) & [S119](#)).
- Possible artifacts from automatic reconstruction with 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 [S120](#) & [S121](#) with [S118](#) & [S119](#)).
- 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 [S130](#)). 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 while that of models from Path2Models, Ebrahim *et al.* and OptFlux is relatively high (Figure [S126](#)).

3.2 Scores

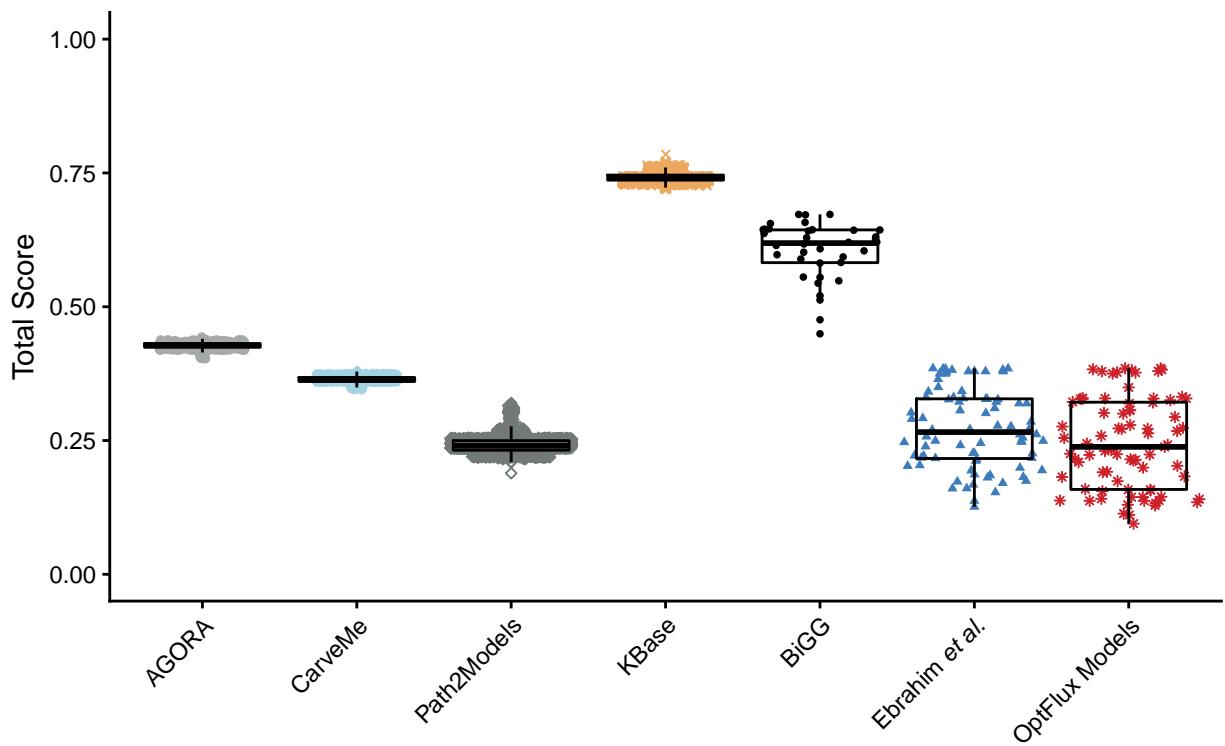


Figure S4: Total Score. Depicted are the sums of all test scores in all independent sections, applying the weights for individual test cases and sections as detailed in the snapshot report.

3.3 Independent Section

3.3.1 Consistency

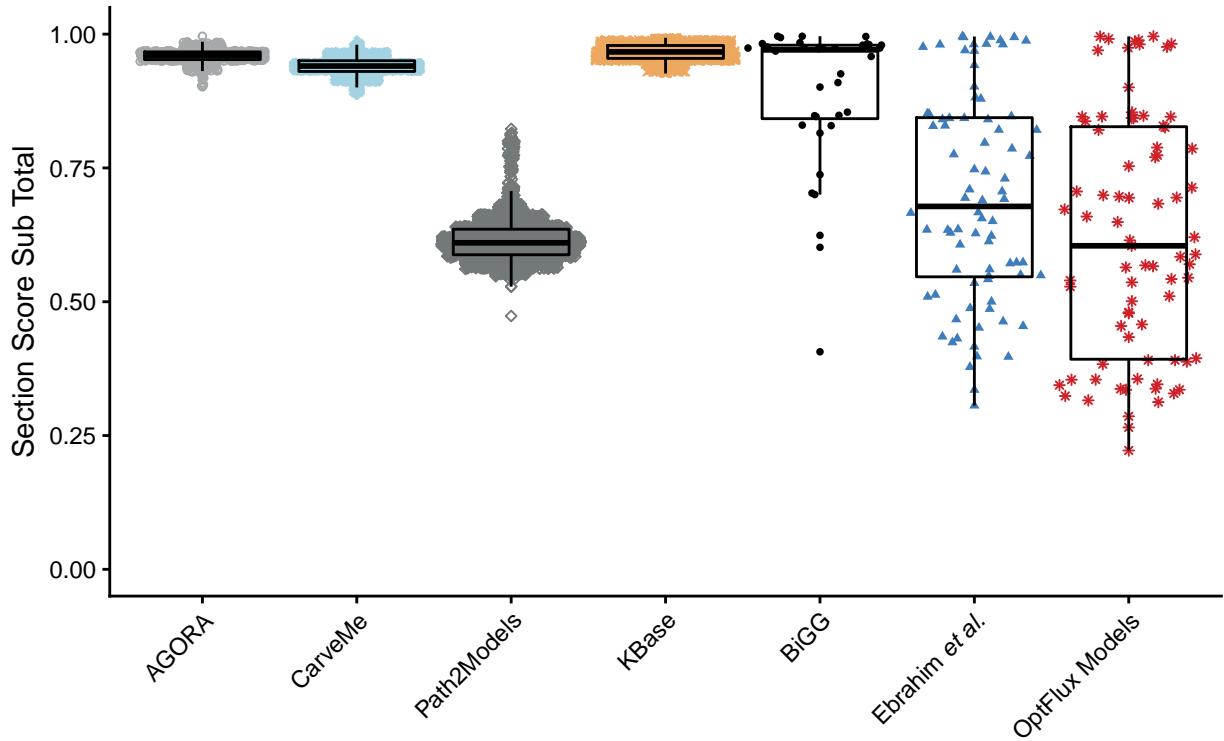


Figure S5: Consistency. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

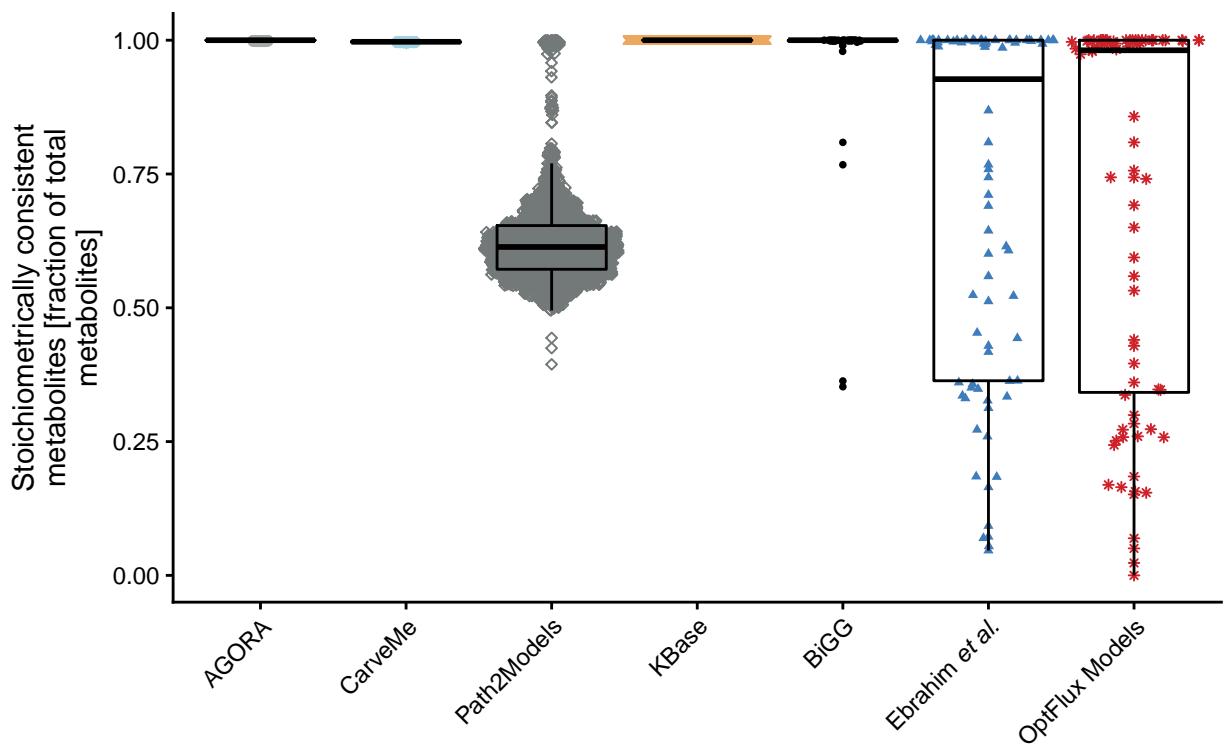


Figure S6: Stoichiometric consistency

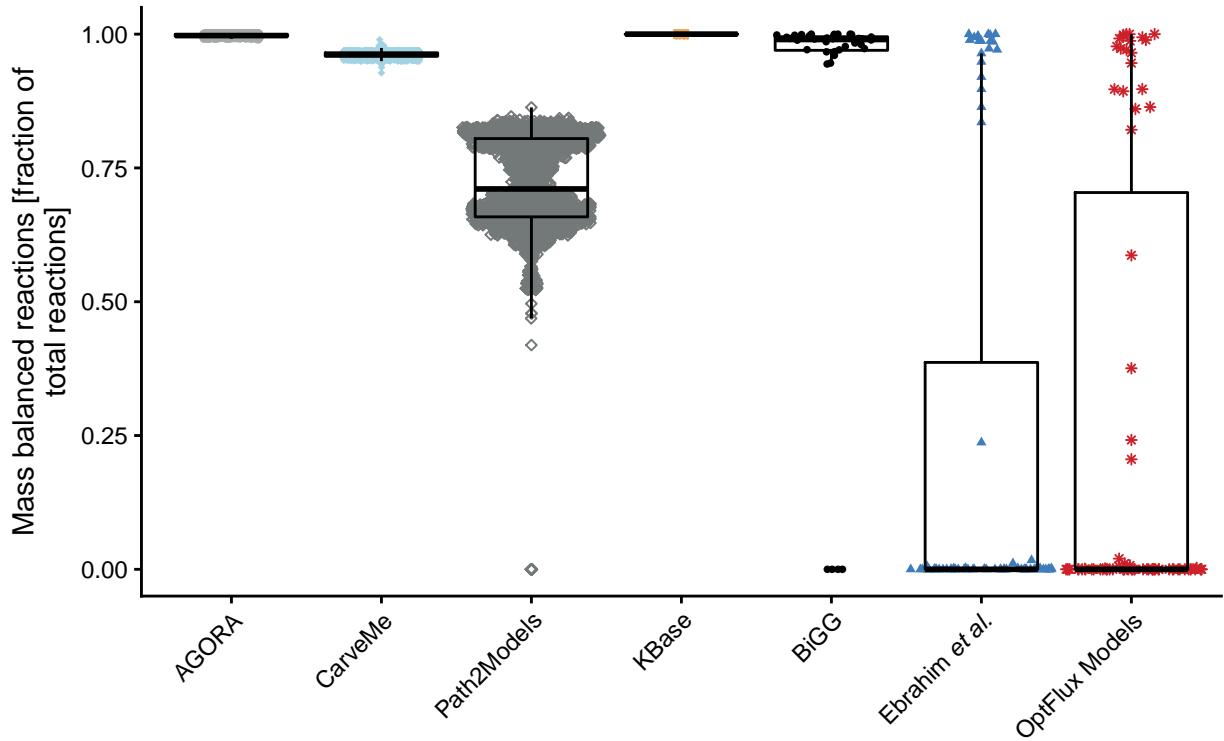


Figure S7: Mass Balance. Please note that any reaction where at least one metabolite lacks a formula annotation is considered as unbalanced for the purpose of this test.

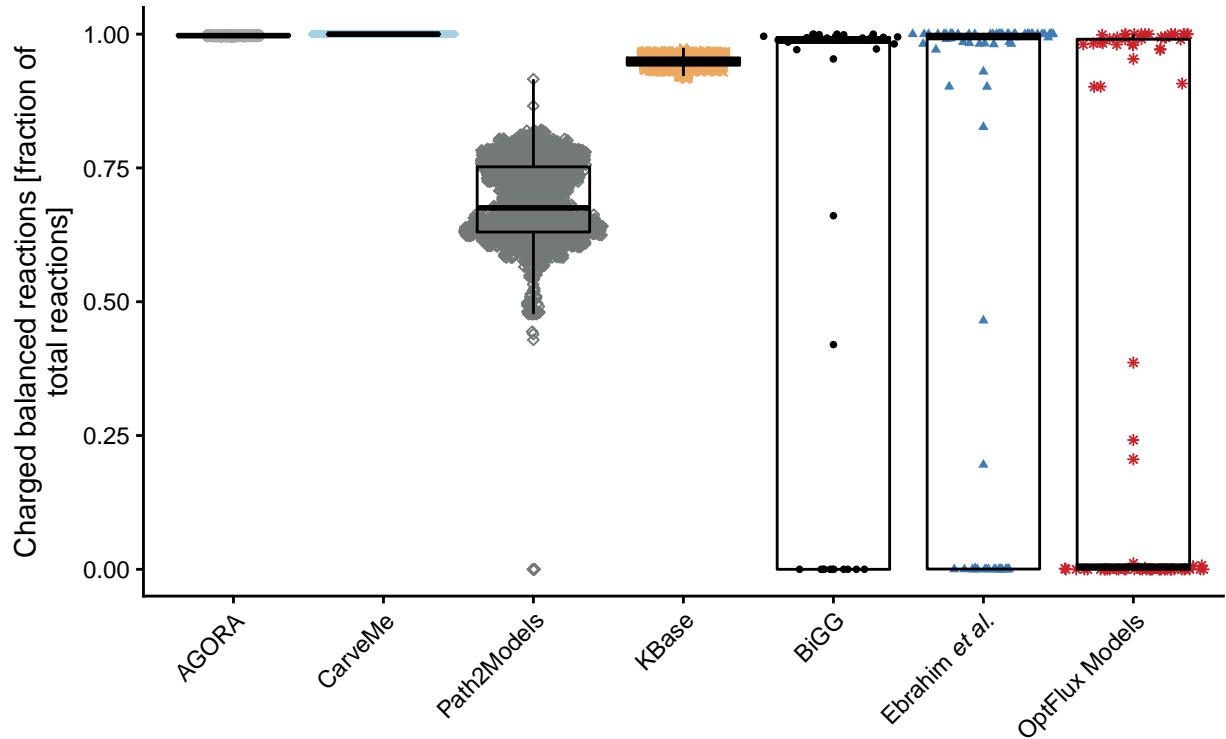


Figure S8: Charge Balance. Please note that any reaction where at least one metabolite lacks charge information is considered as unbalanced for the purpose of this test.

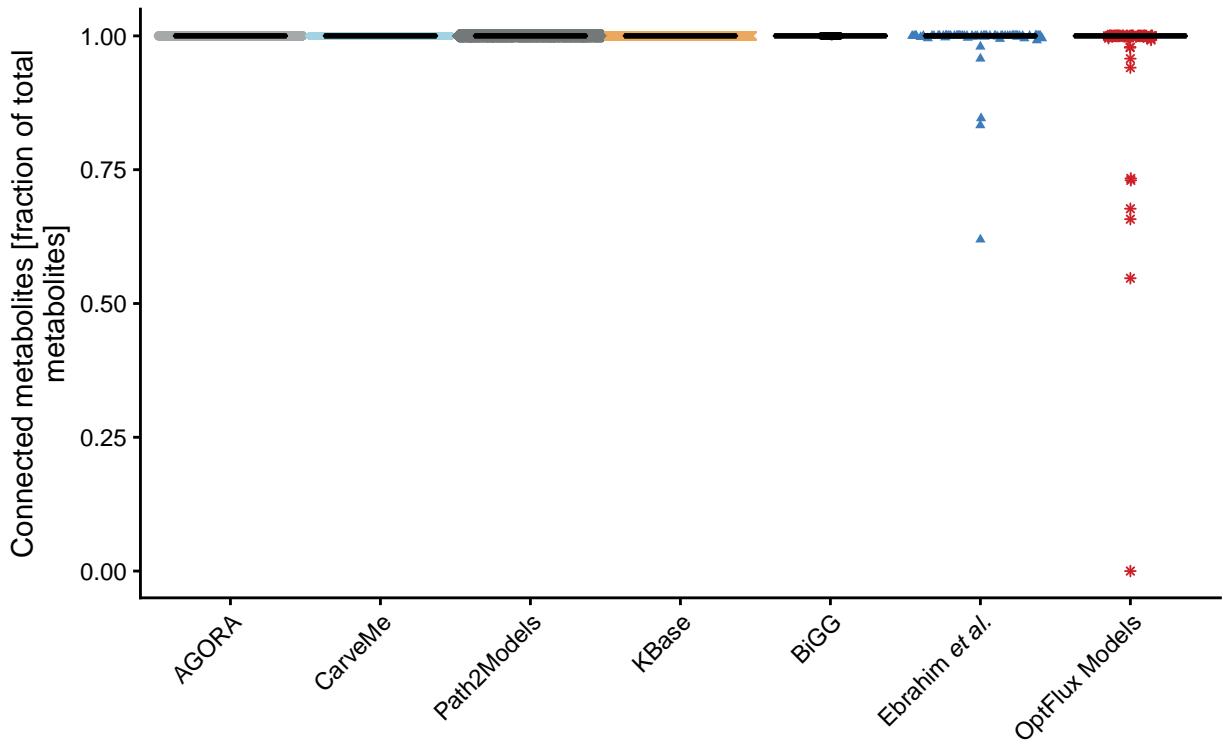


Figure S9: Metabolite Connectivity

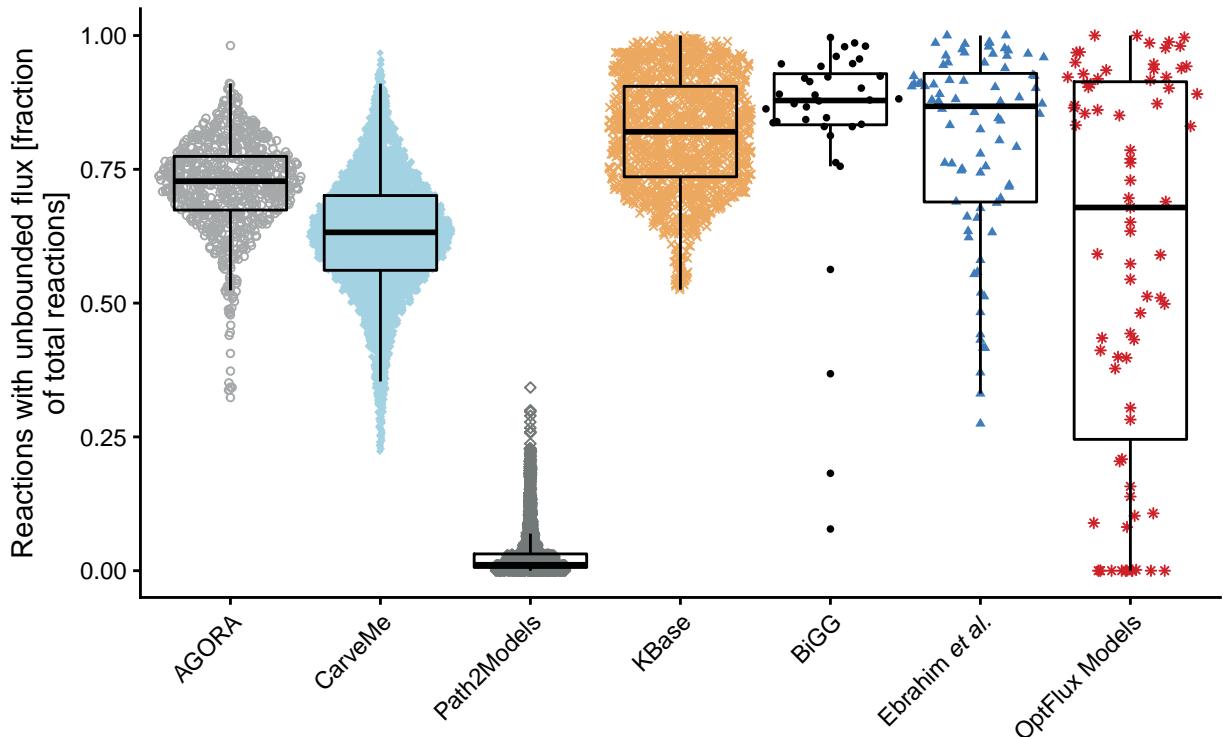


Figure S10: Unbounded Flux in Default Medium

3.3.2 Annotation - Metabolites

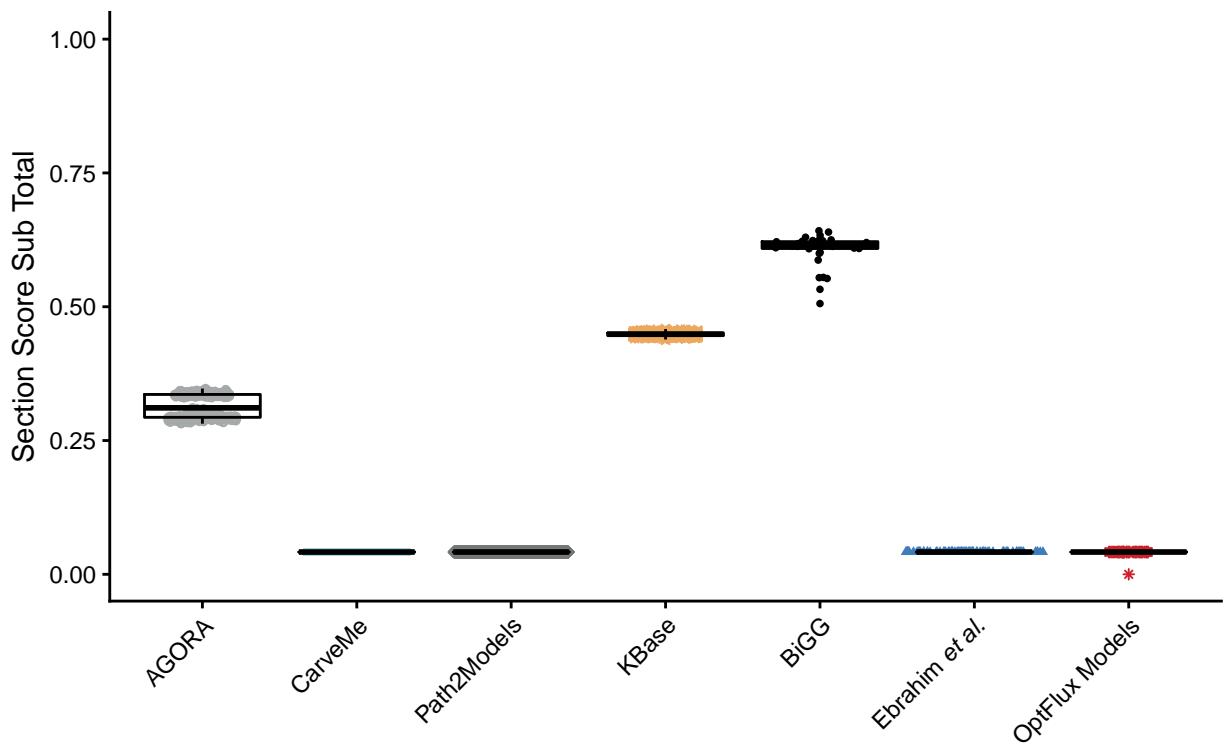


Figure S11: Annotation - Metabolites. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

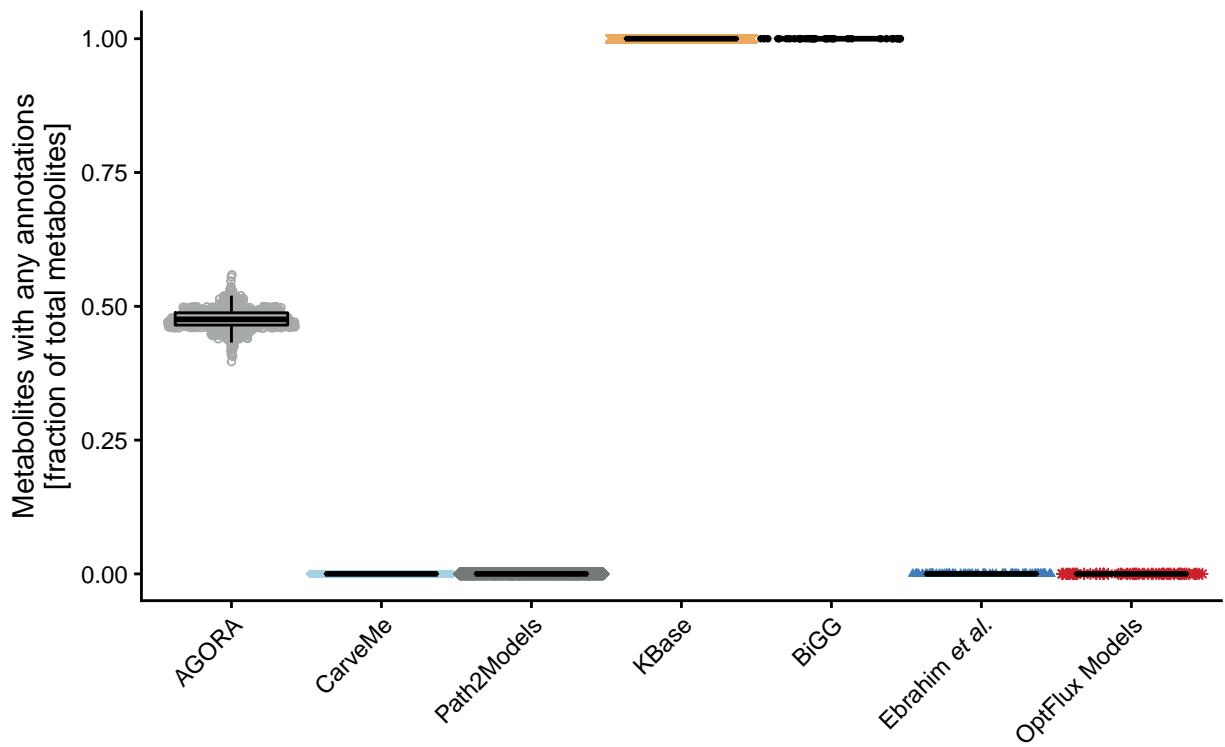


Figure S12: Presence of Metabolite Annotation

3.3.2.1 Metabolite Annotations Per Database

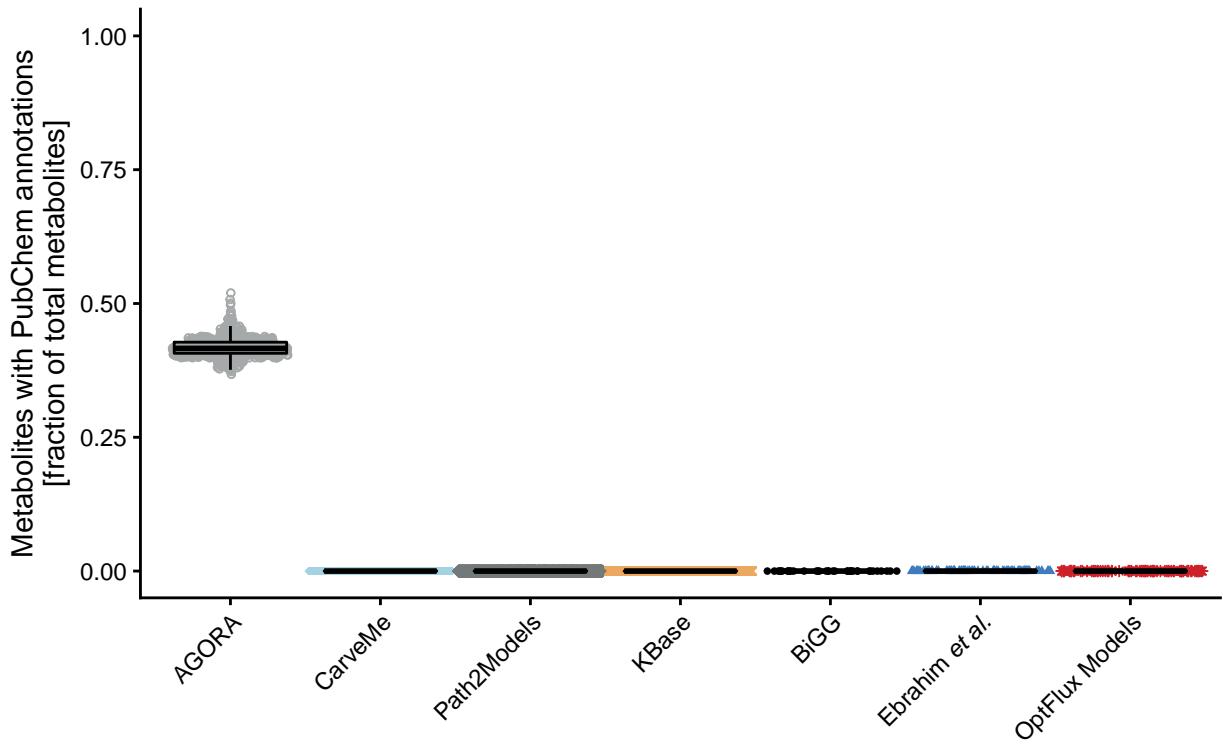


Figure S13: Metabolite Pubchem.compound Annotation

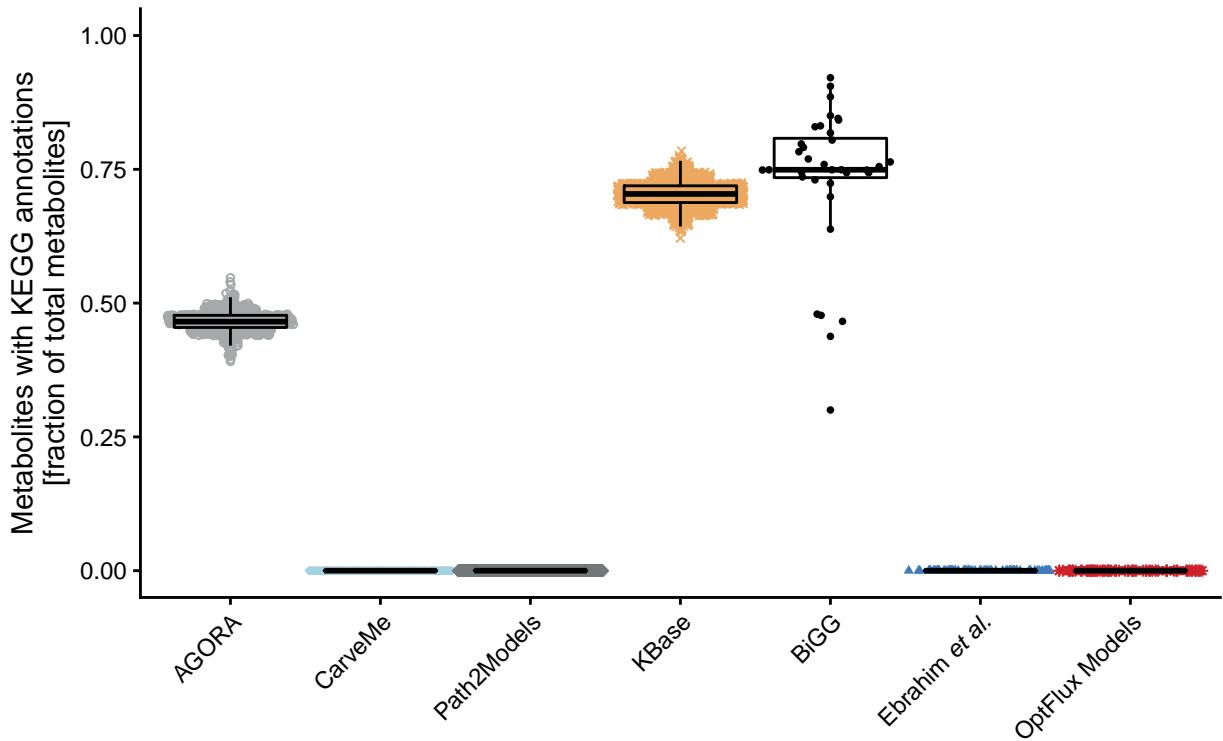


Figure S14: Metabolite KEGG.compound Annotation

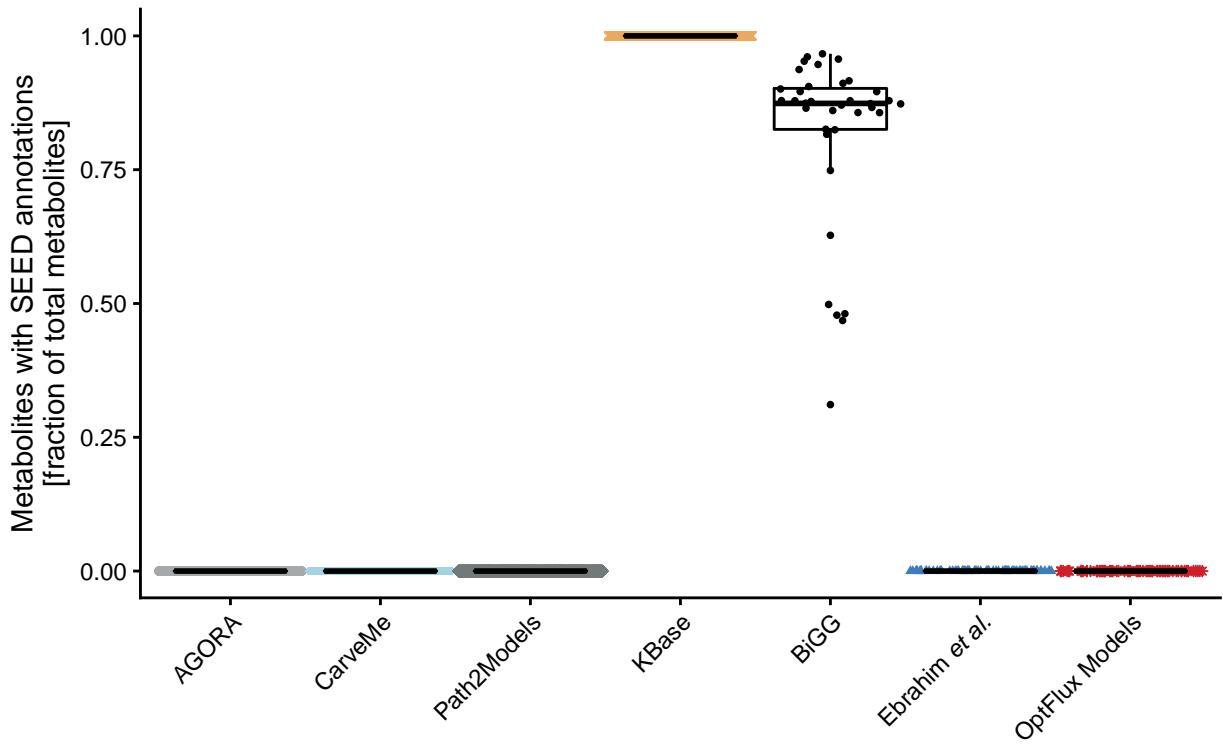


Figure S15: Metabolite SEED.compound Annotation

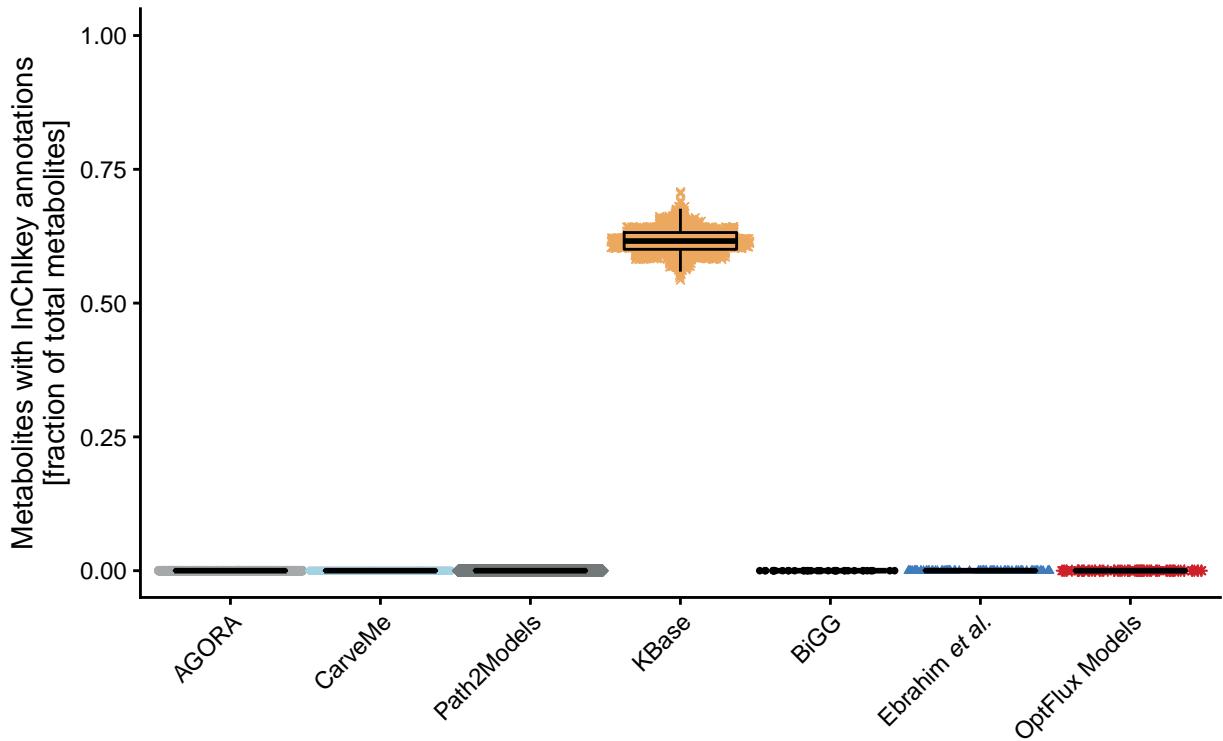


Figure S16: Metabolite InChIKey Annotation

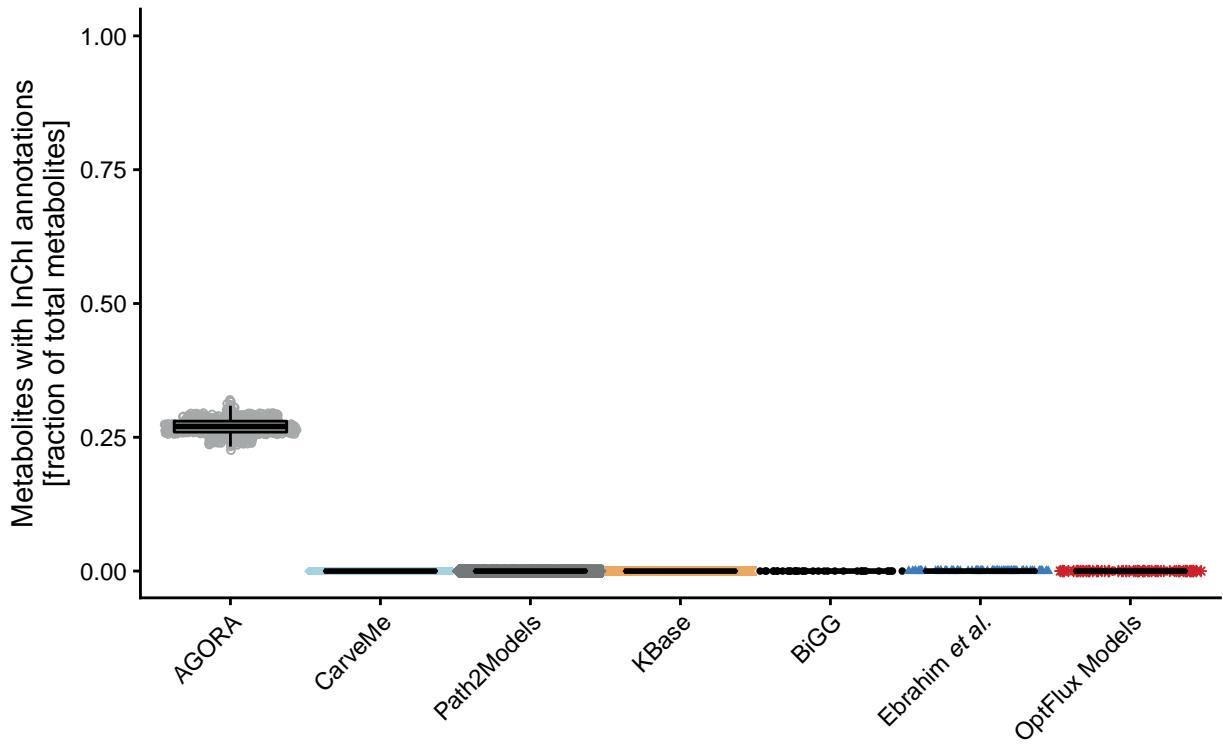


Figure S17: Metabolite InChI Annotation

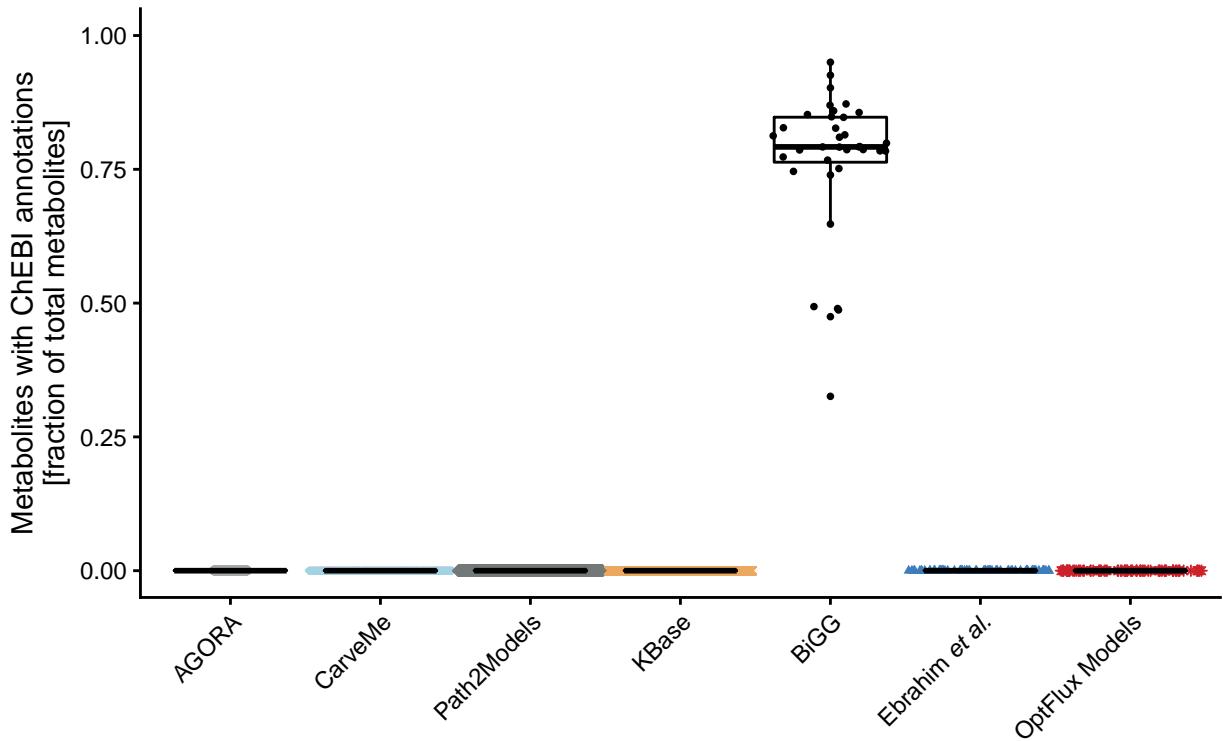


Figure S18: Metabolite ChEBI Annotation

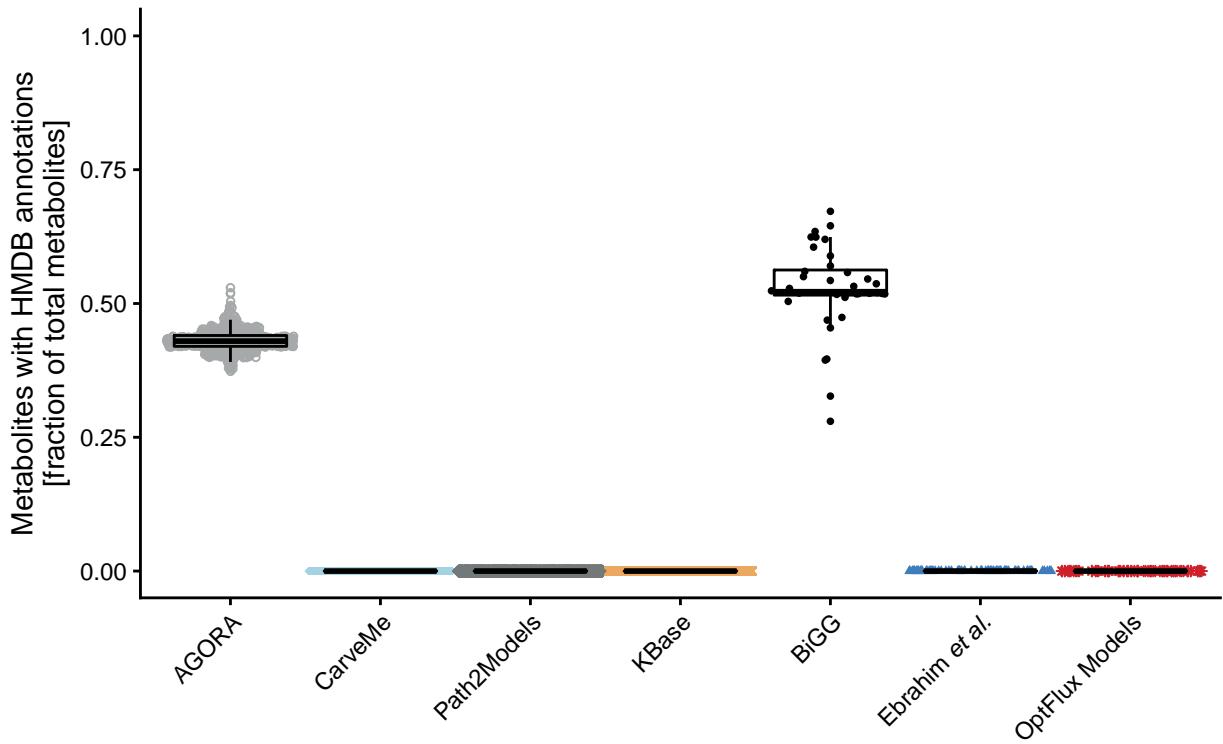


Figure S19: Metabolite HMDB Annotation

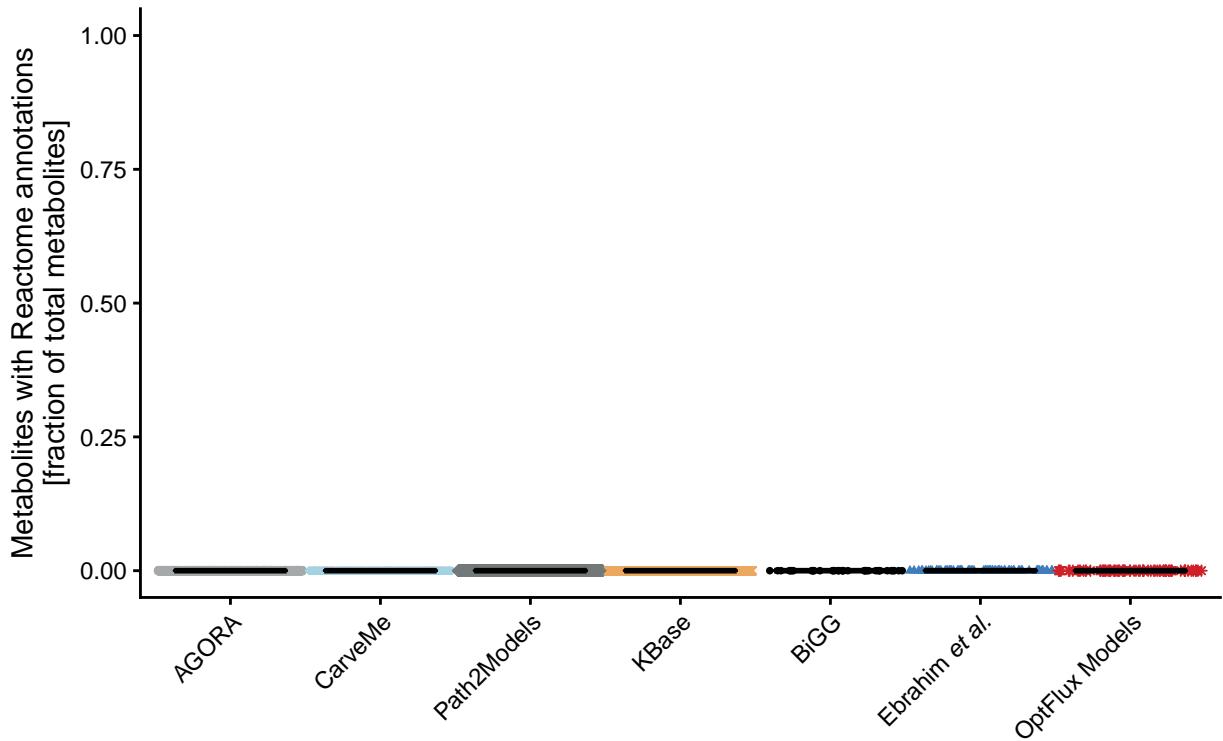


Figure S20: Metabolite Reactome Annotation

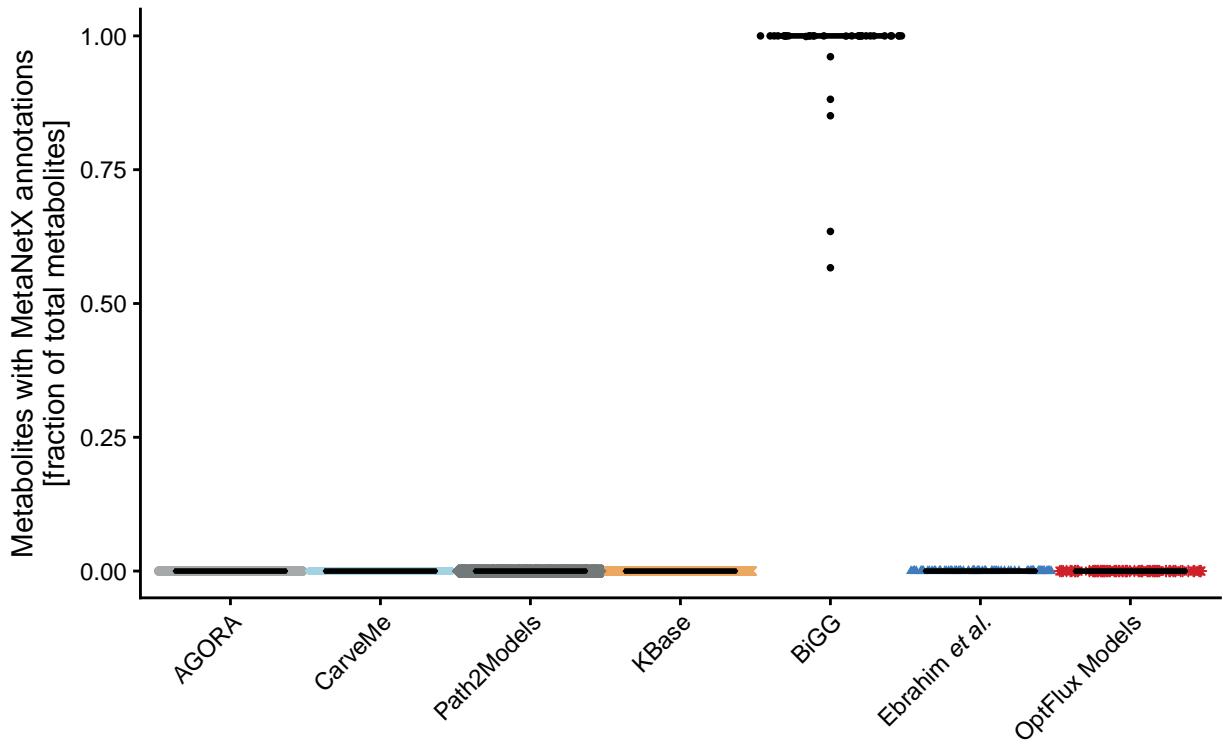


Figure S21: Metabolite MetaNetX.chemical Annotation

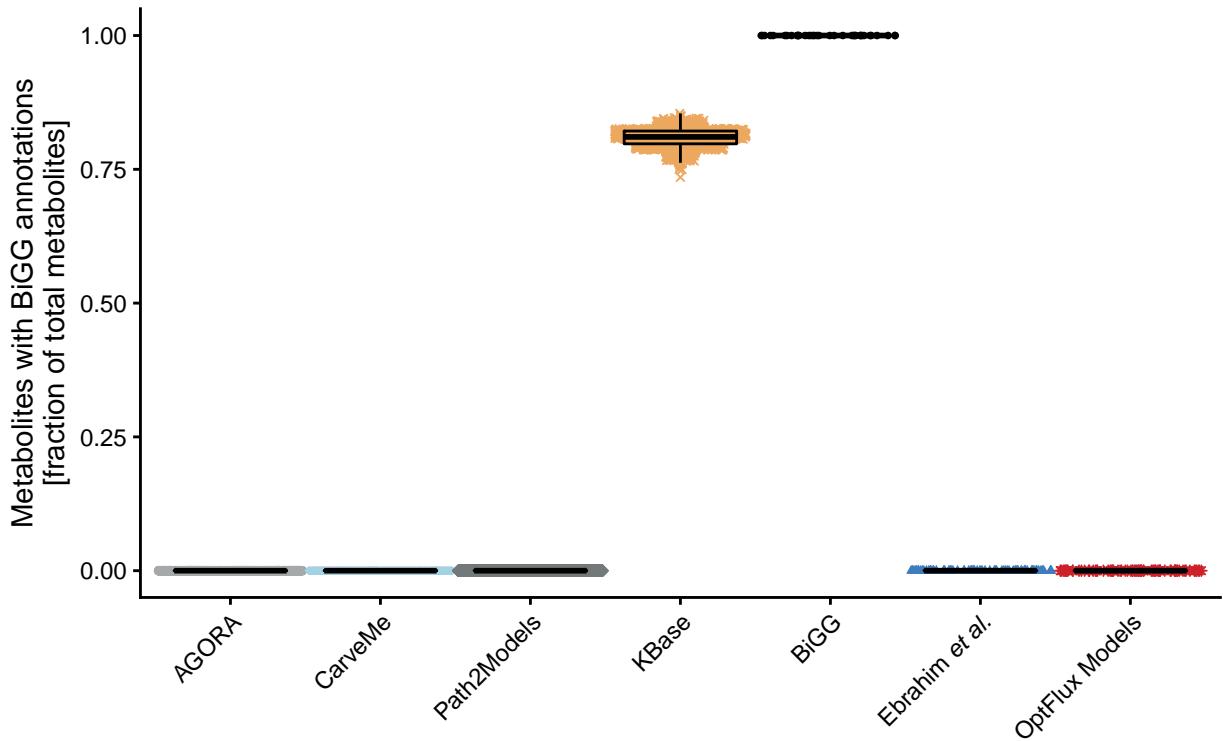


Figure S22: Metabolite BiGG.metabolite Annotation

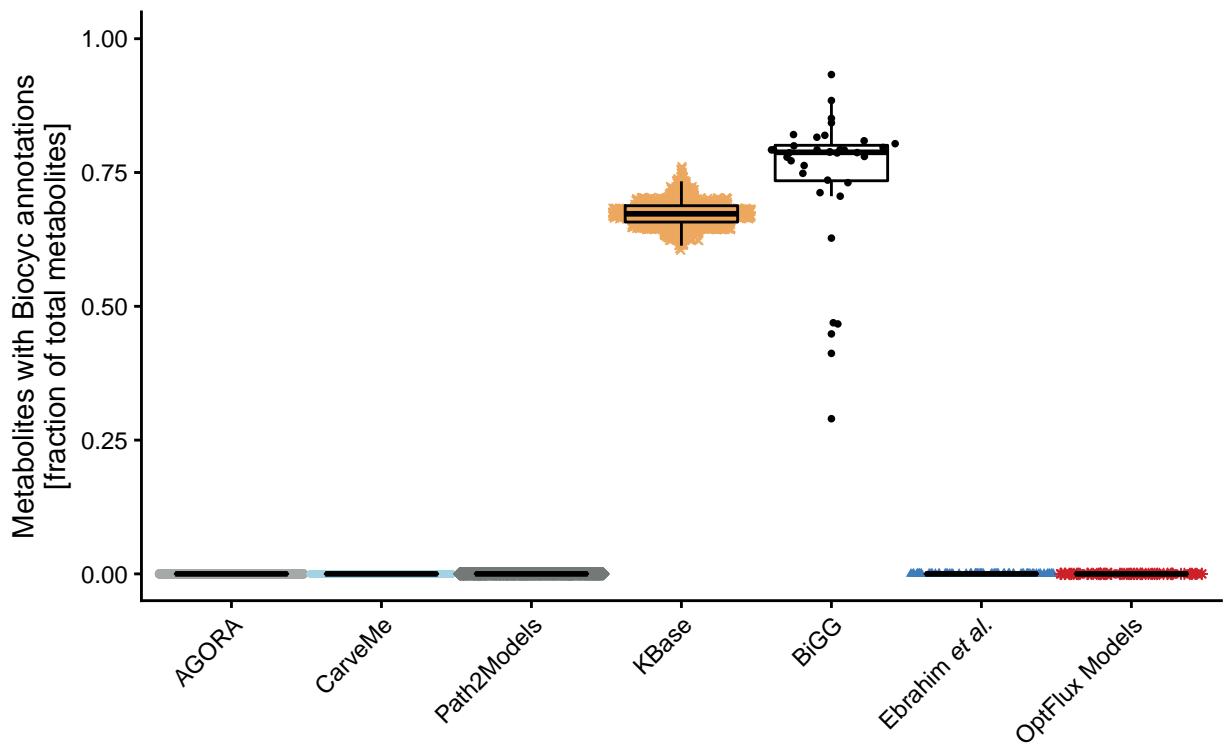


Figure S23: Metabolite BioCyc Annotation

3.3.2.2 Metabolite Annotation Conformity per Database

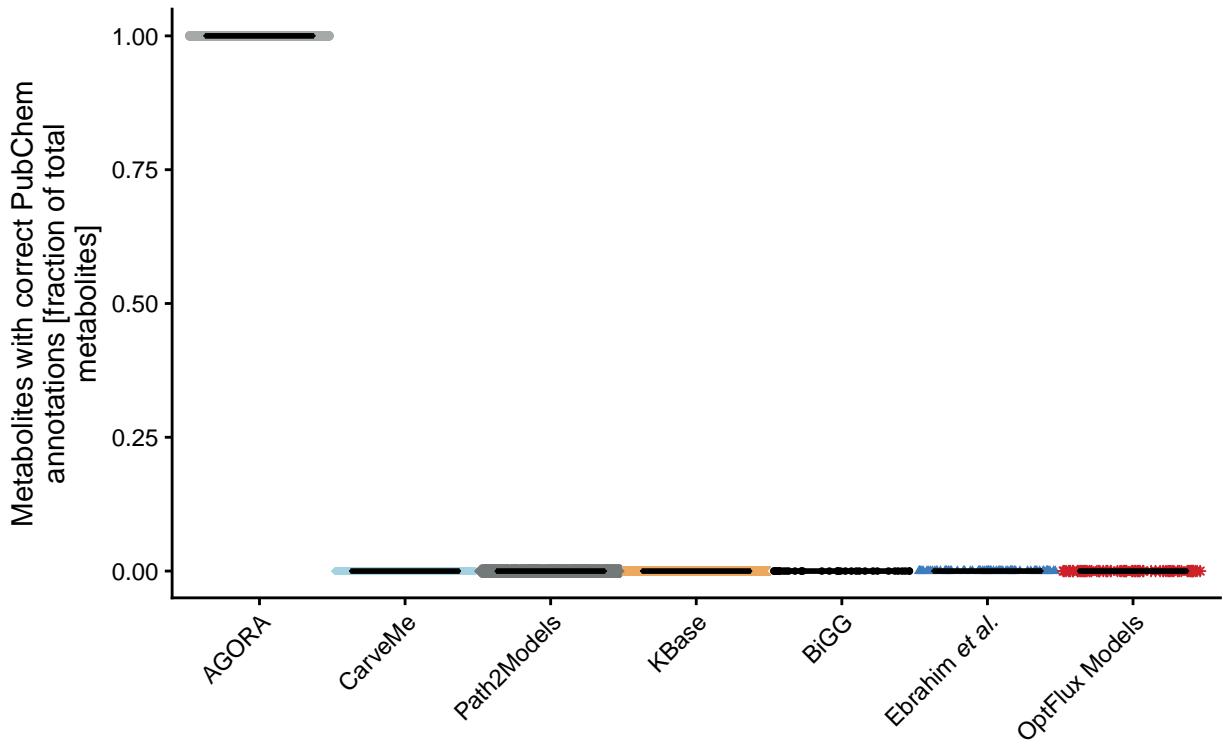


Figure S24: Correct Metabolite Pubchem.compound Annotation

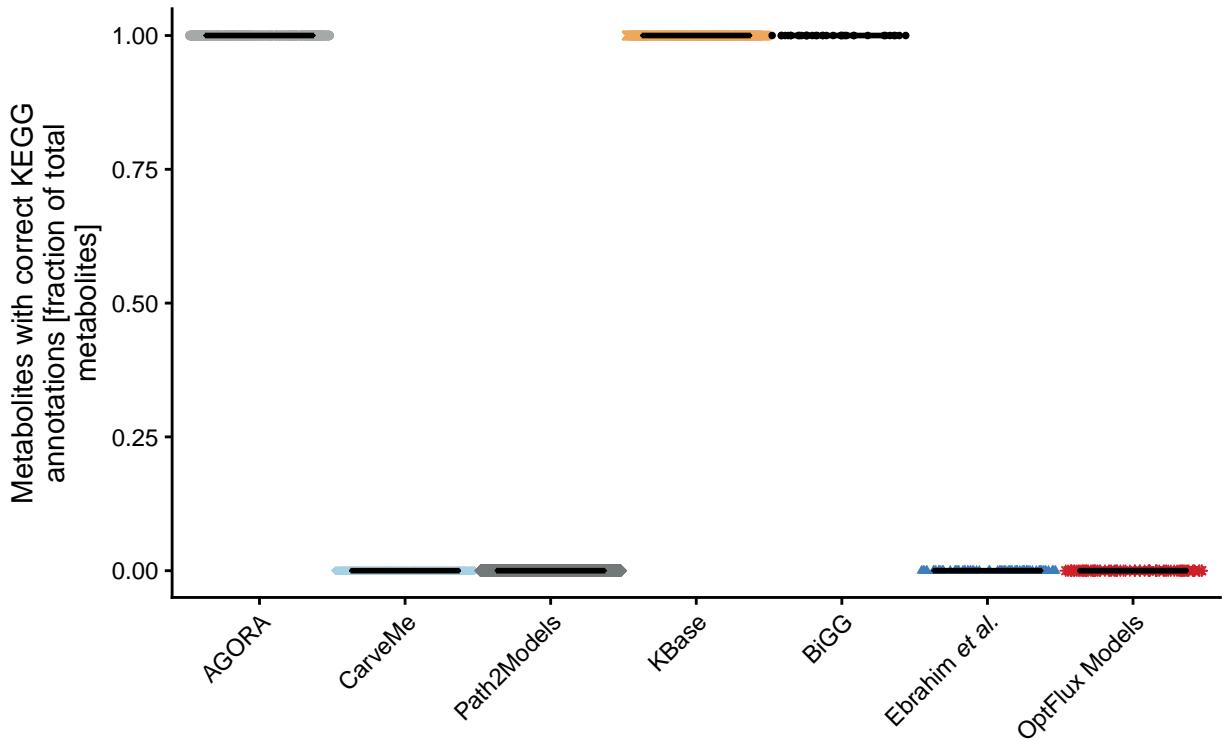


Figure S25: Correct Metabolite KEGG.compound Annotation

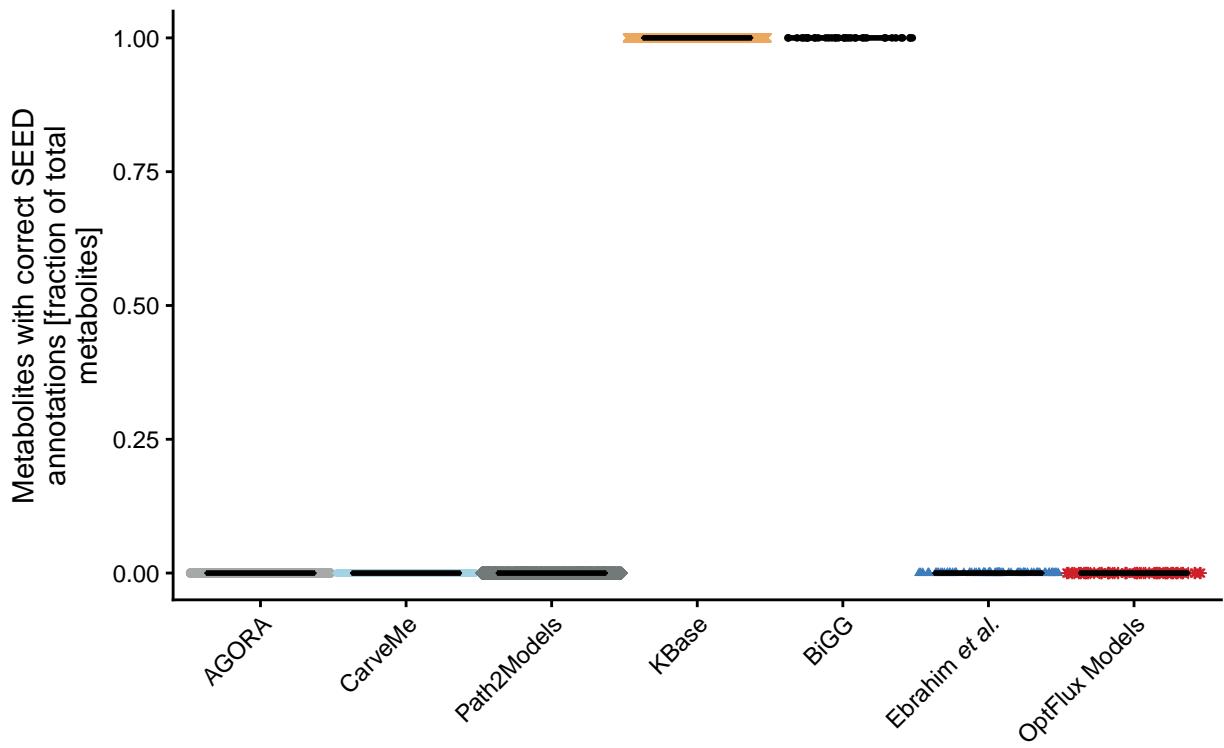


Figure S26: Correct Metabolite SEED.compound Annotation

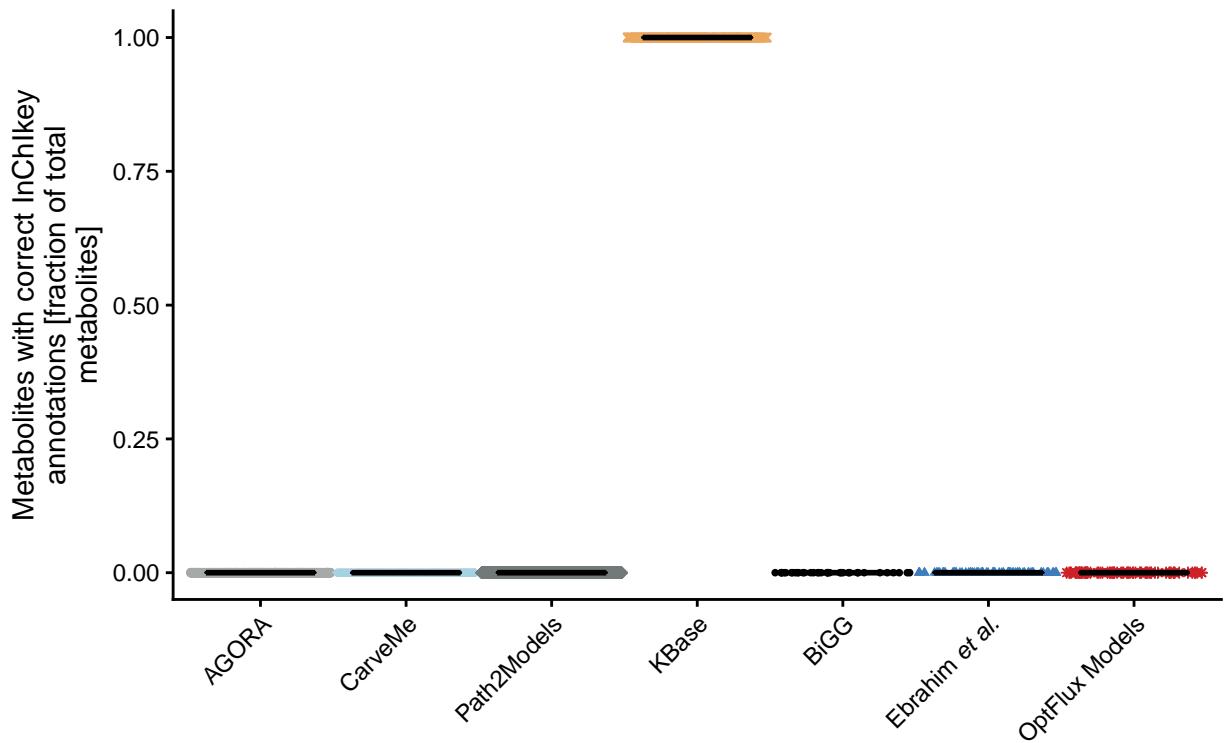


Figure S27: Correct Metabolite InChIKey Annotation

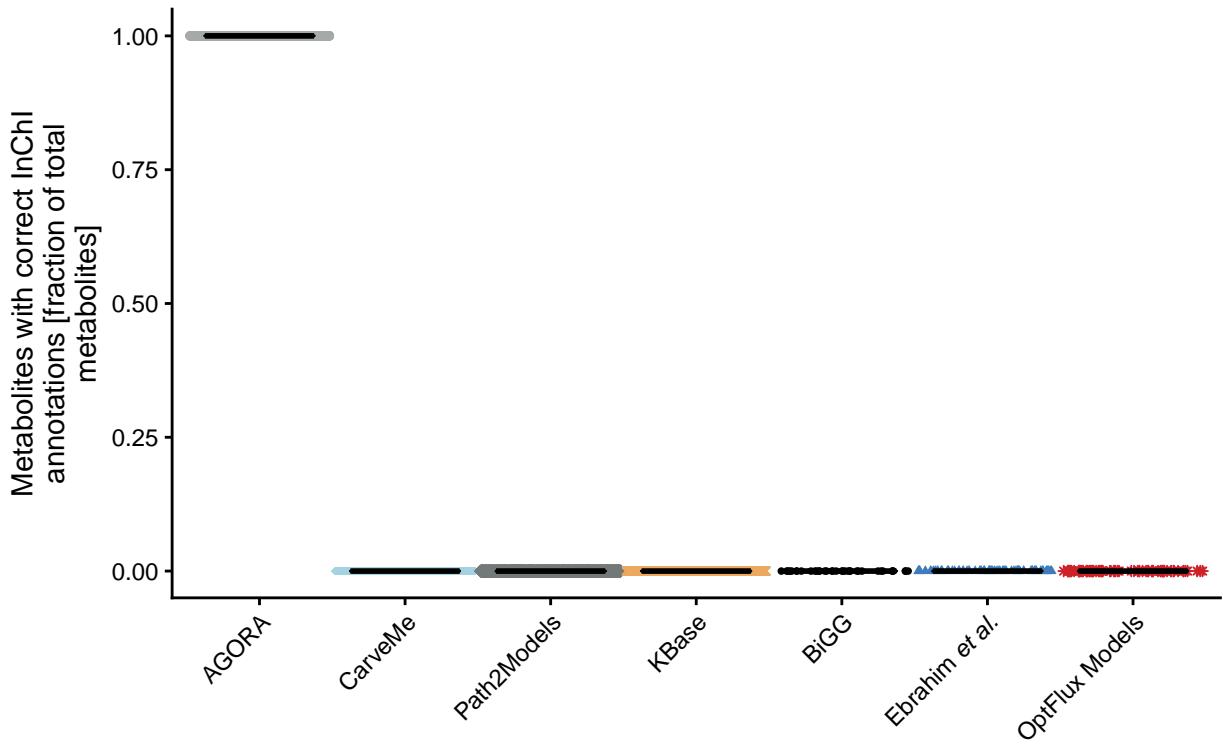


Figure S28: Correct Metabolite InChI Annotation

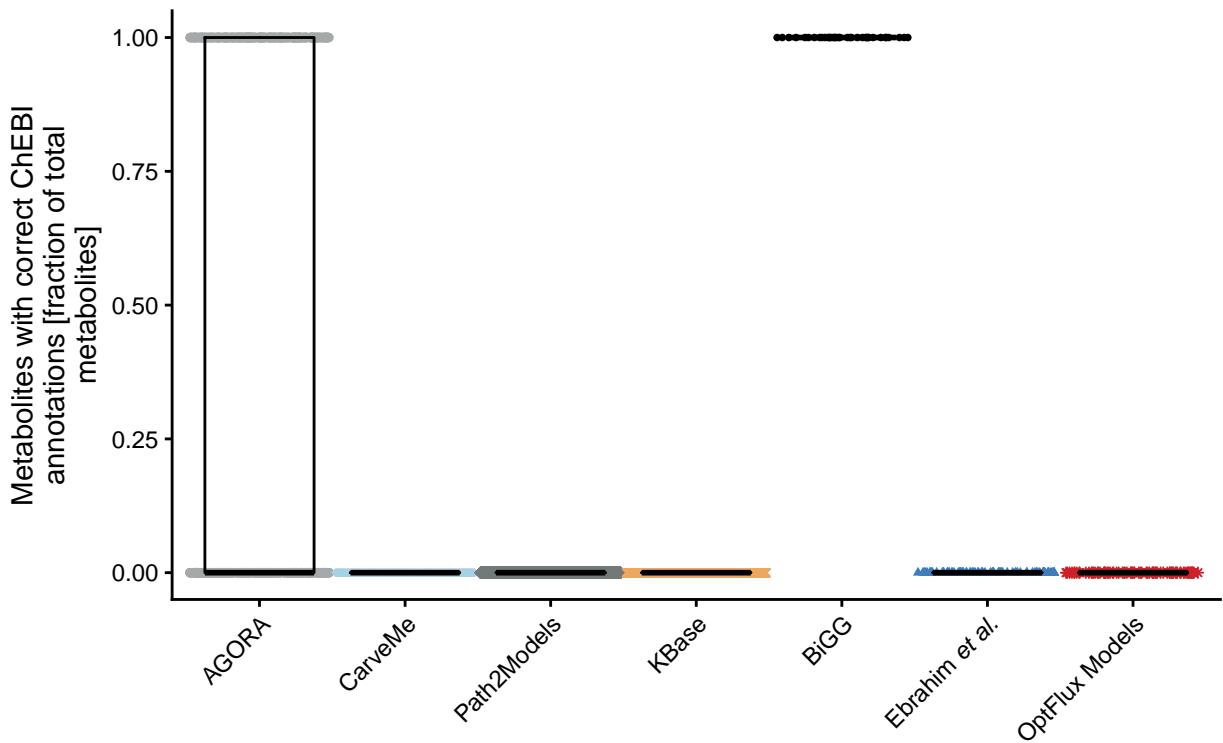


Figure S29: Correct Metabolite ChEBI Annotation

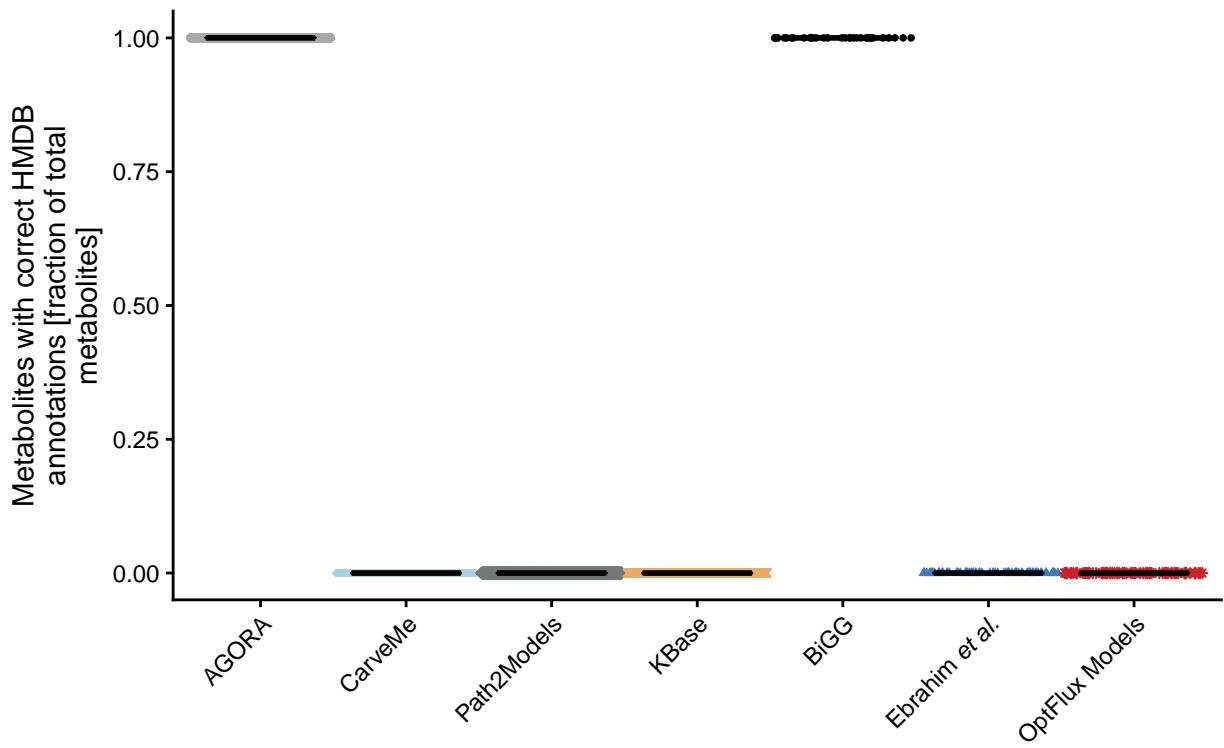


Figure S30: Correct Metabolite HMDB Annotation

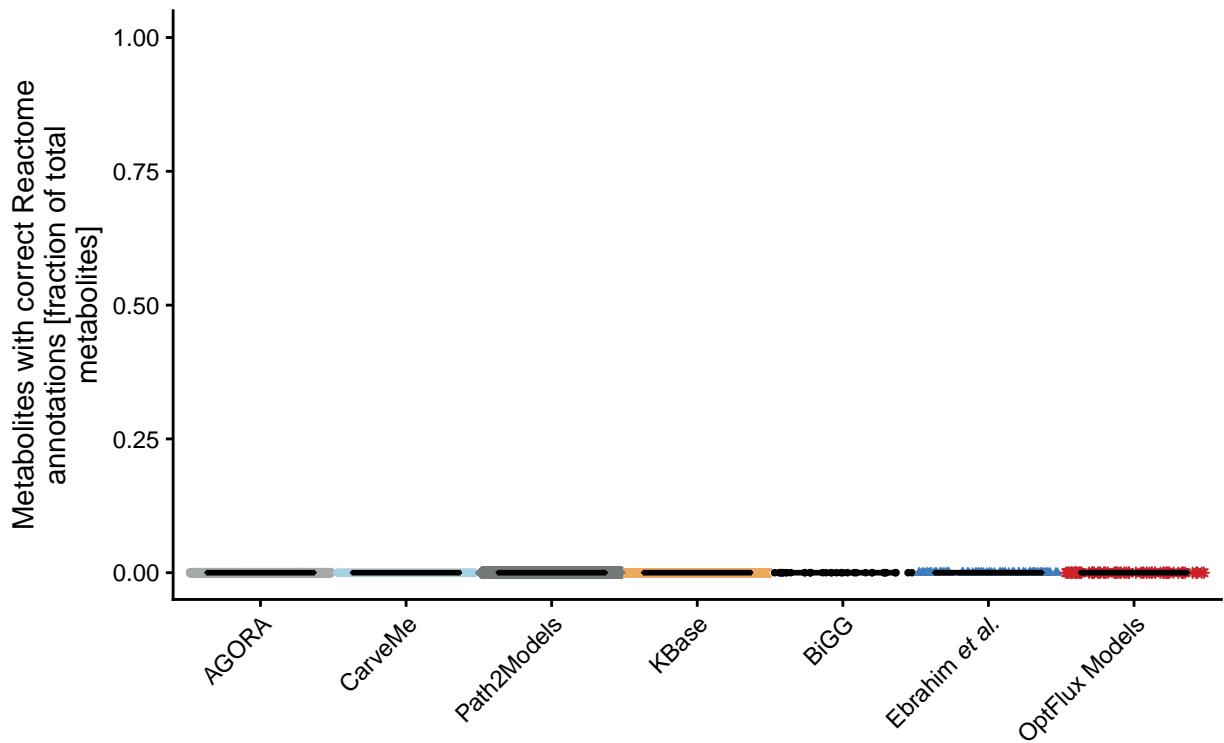


Figure S31: Correct Metabolite Reactome Annotation

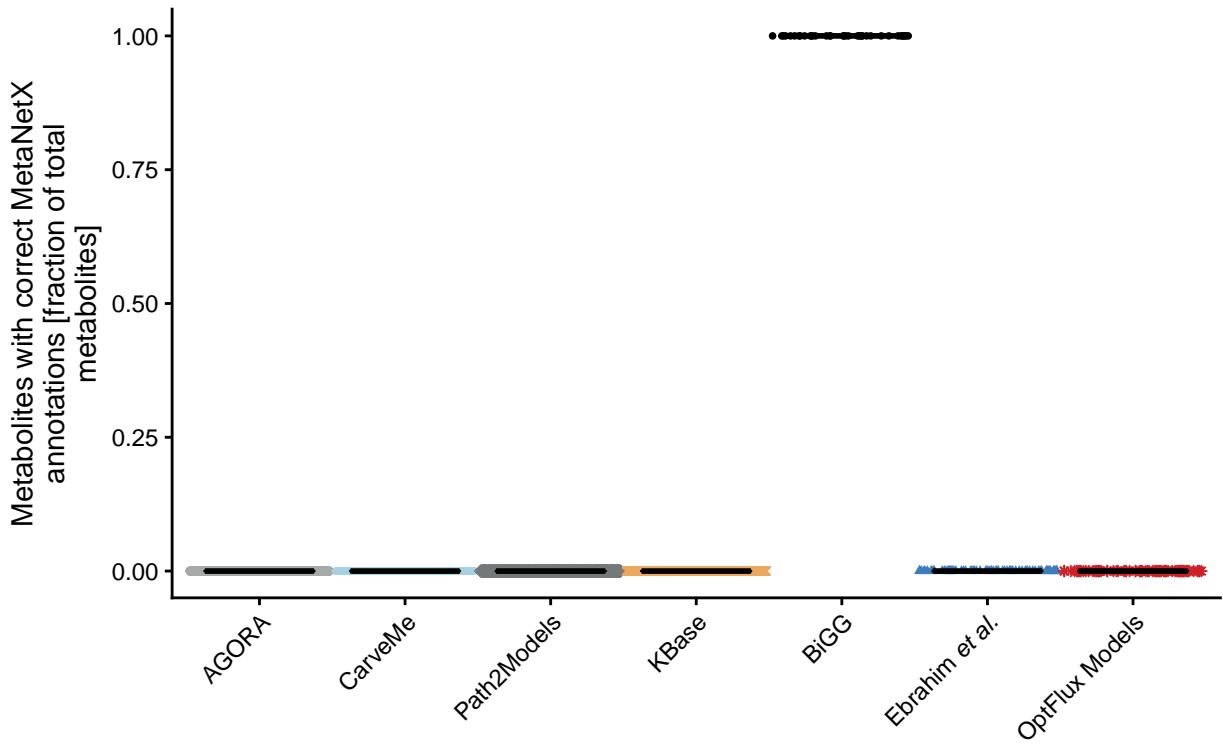


Figure S32: Correct Metabolite MetaNetX.chemical Annotation

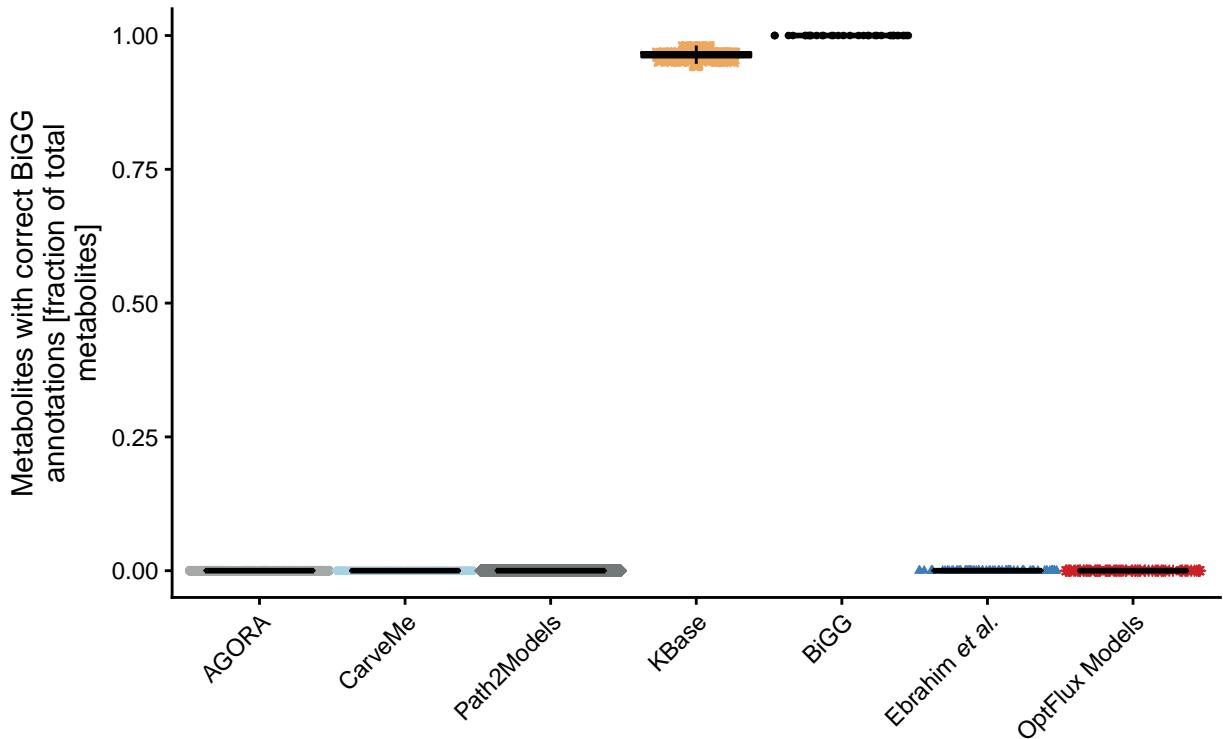


Figure S33: Correct Metabolite BiGG.metabolite Annotation

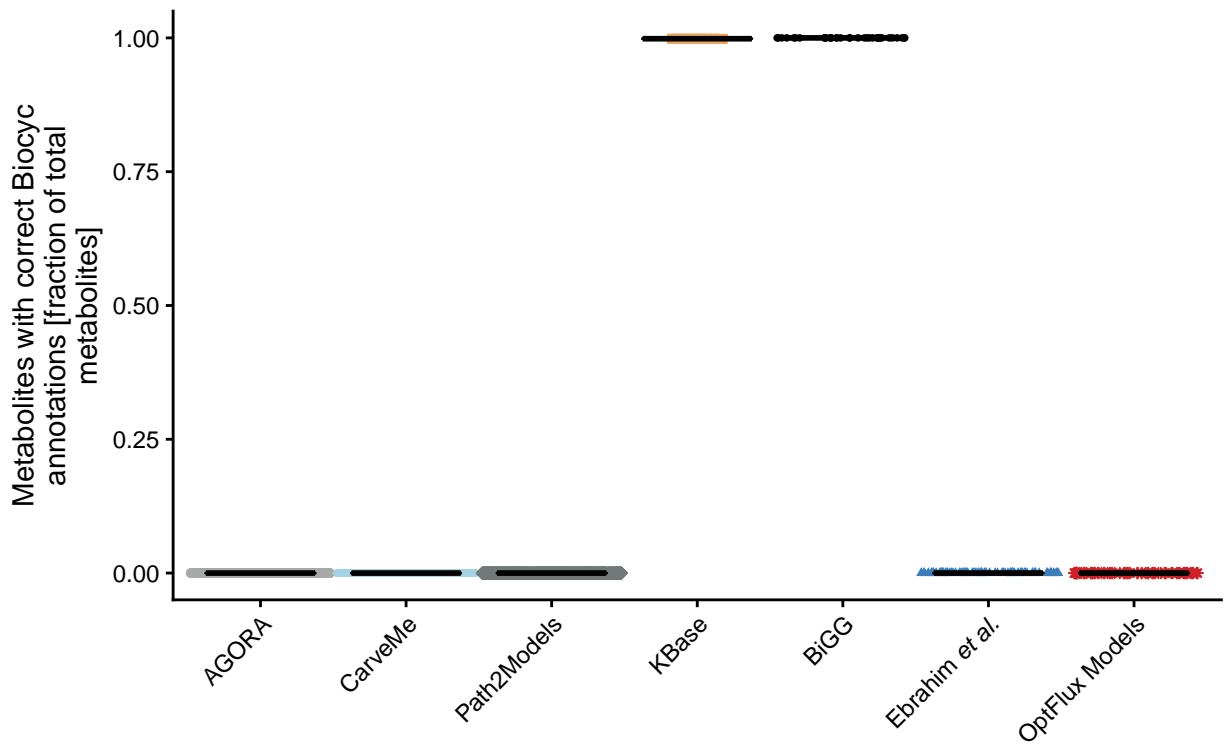


Figure S34: Correct Metabolite BioCyc Annotation

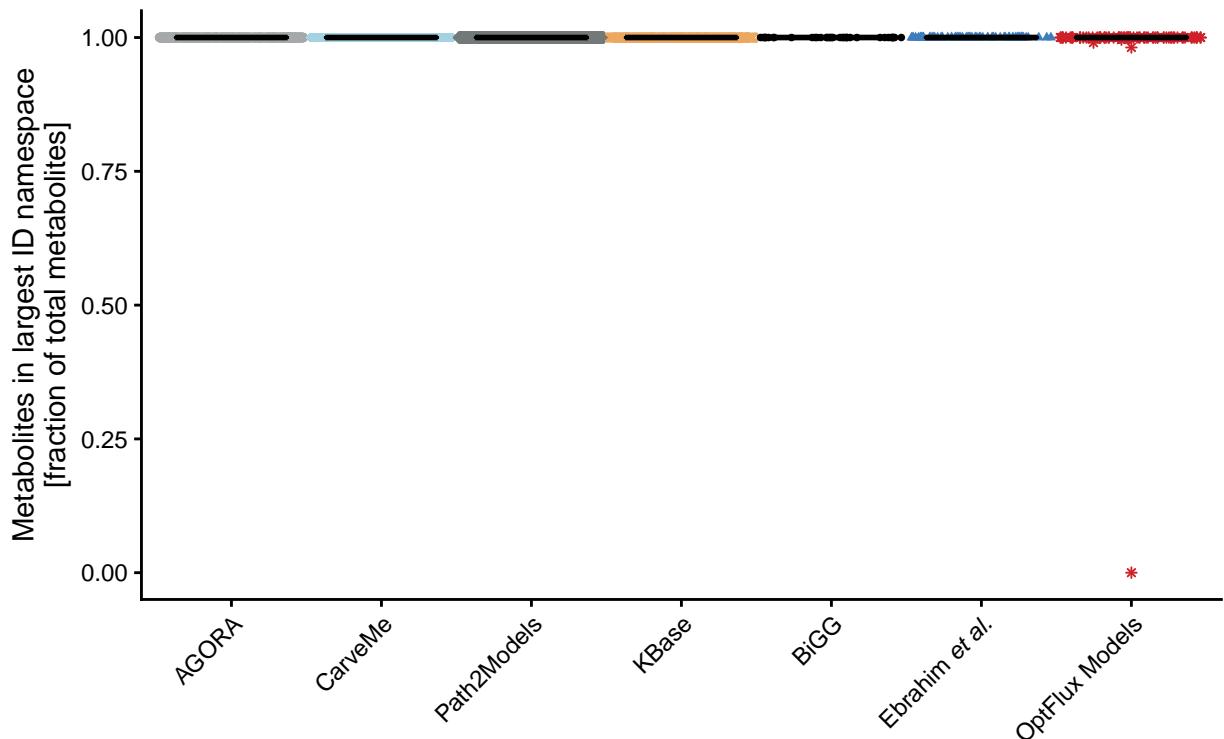


Figure S35: Uniform Metabolite Identifier Namespace

3.3.3 Annotation - Reactions

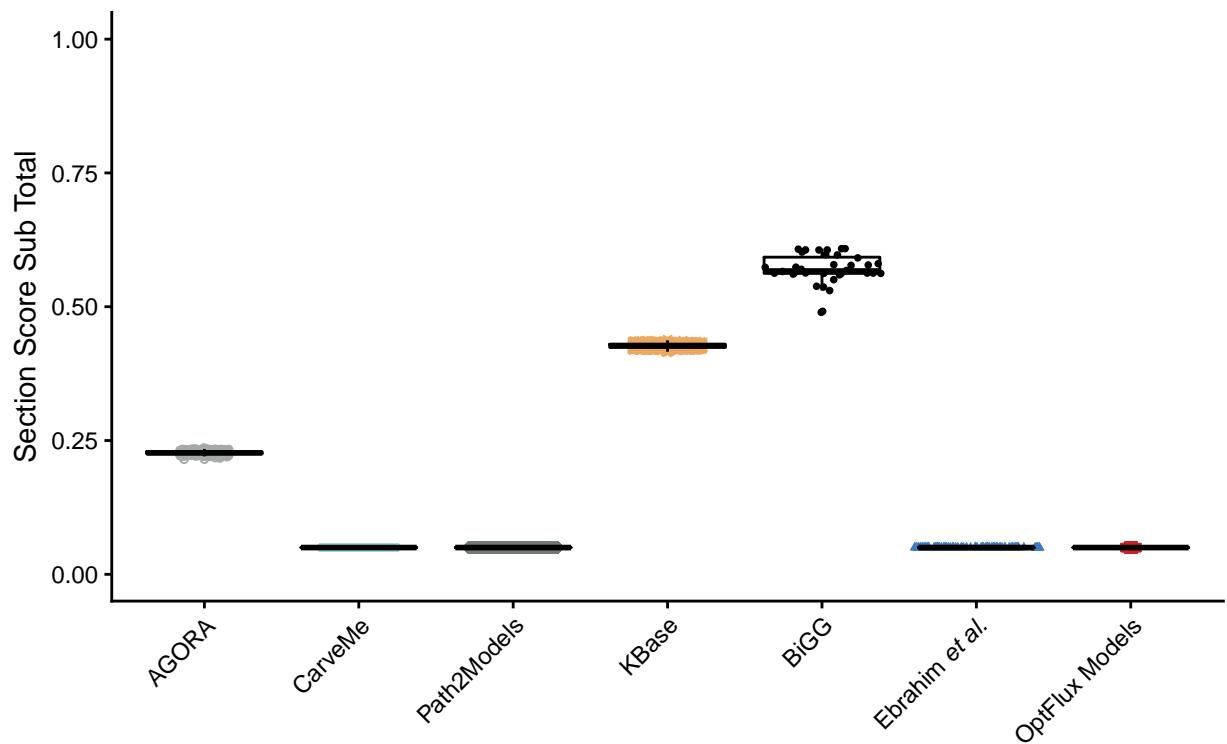


Figure S36: Annotation - Reactions. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

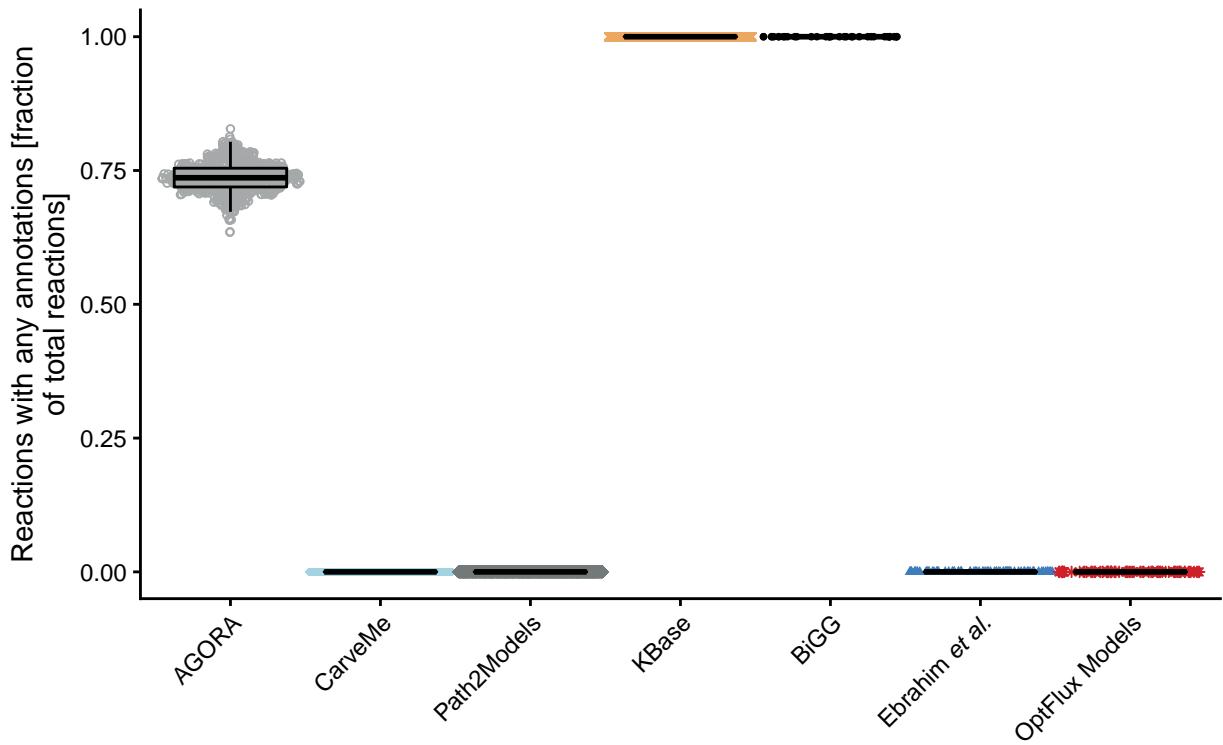


Figure S37: Presence of Reaction Annotation

3.3.3.1 Reaction Annotations Per Database

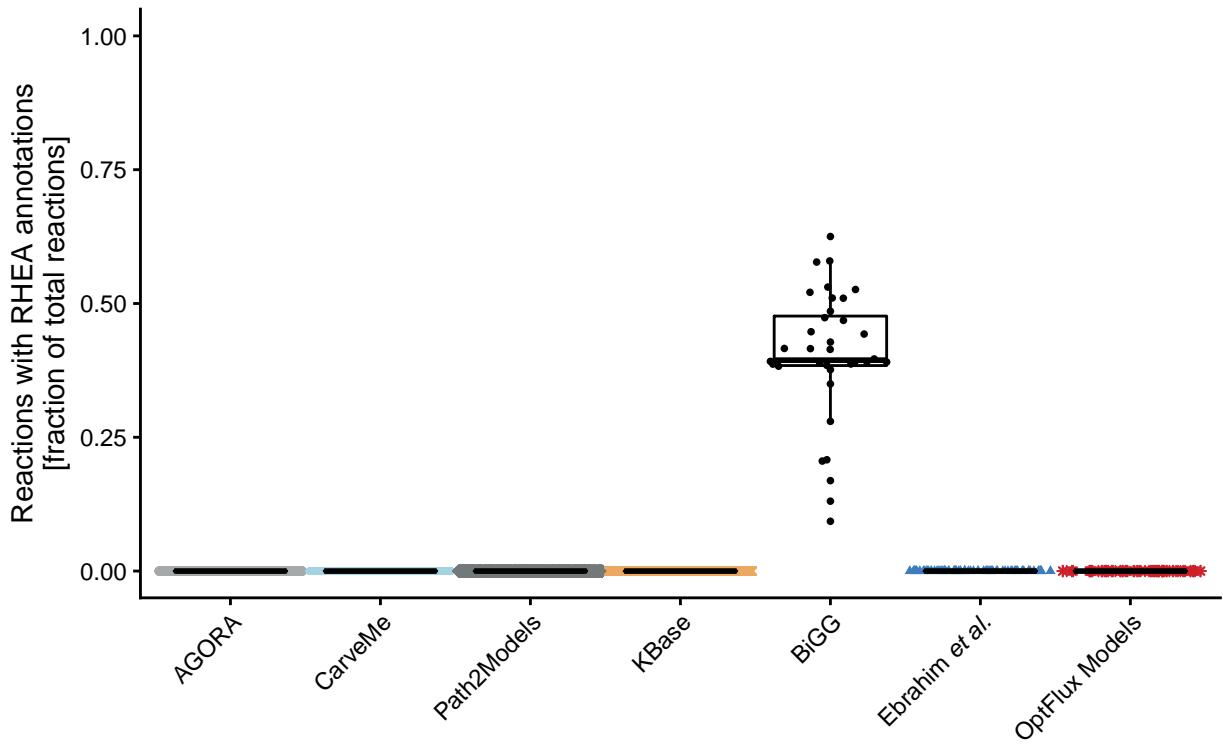


Figure S38: Reaction Rhea Annotation

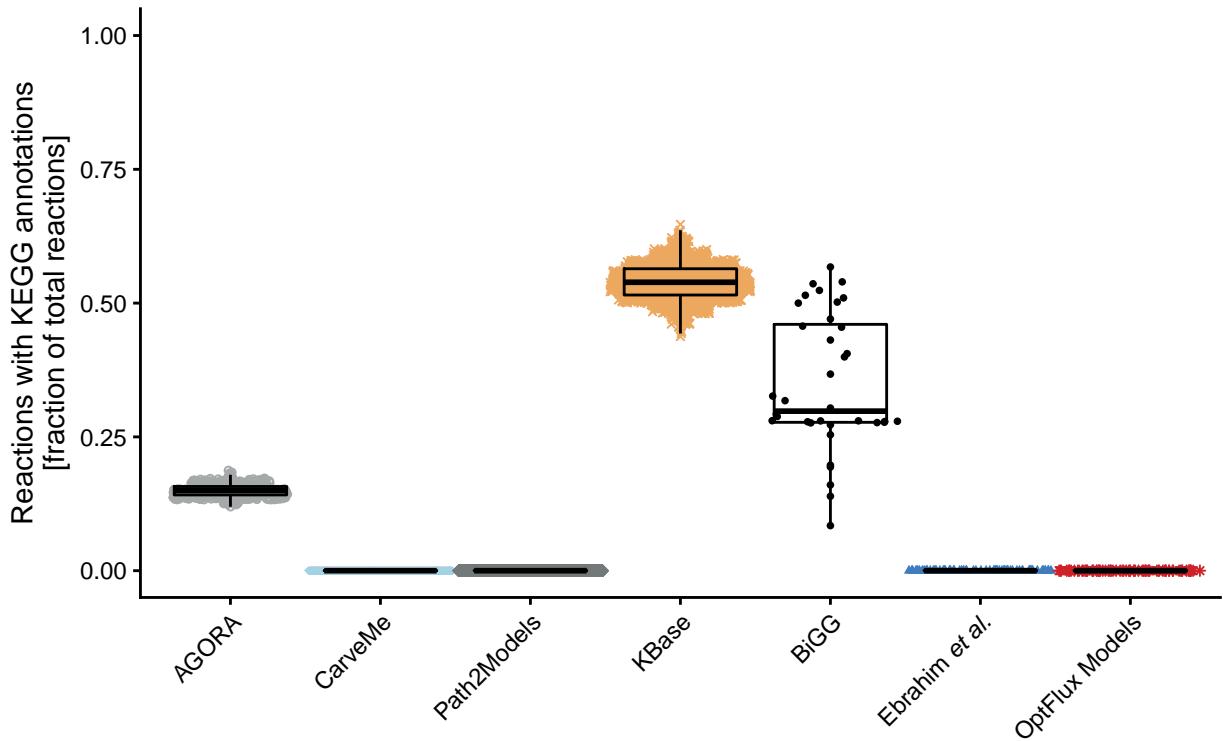


Figure S39: Reaction KEGG.reaction Annotation

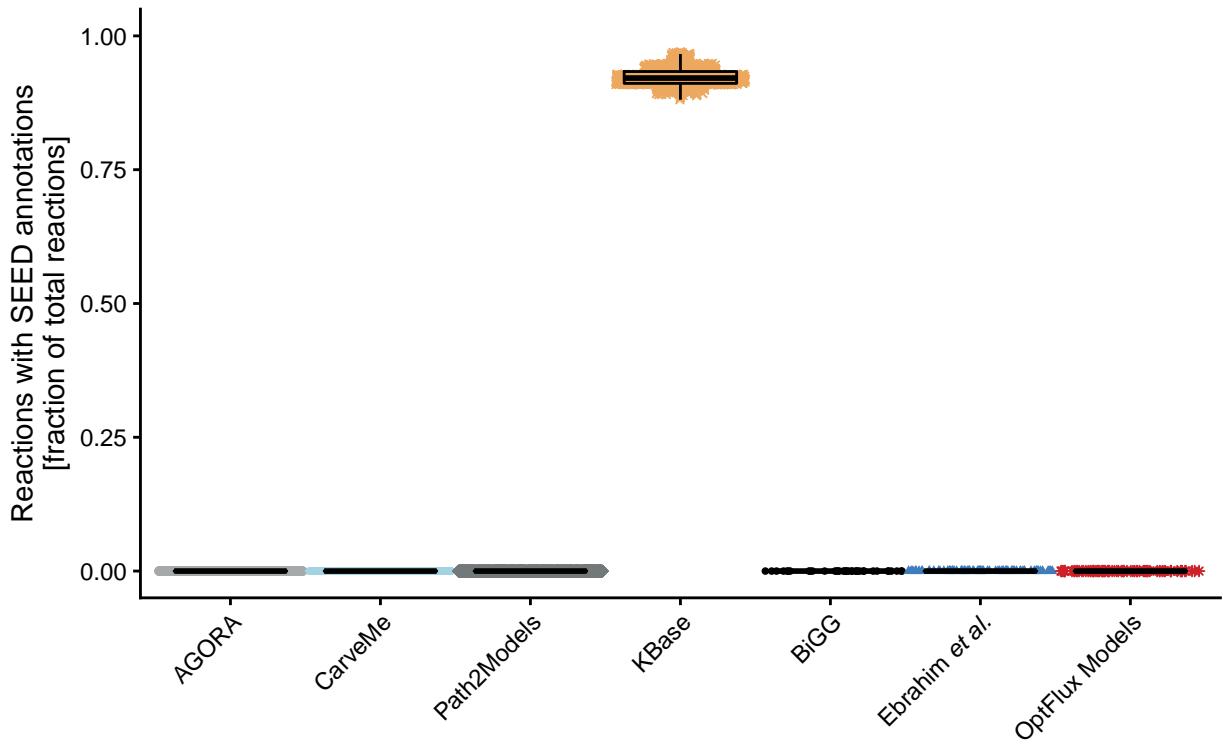


Figure S40: Reaction SEED.reaction Annotation

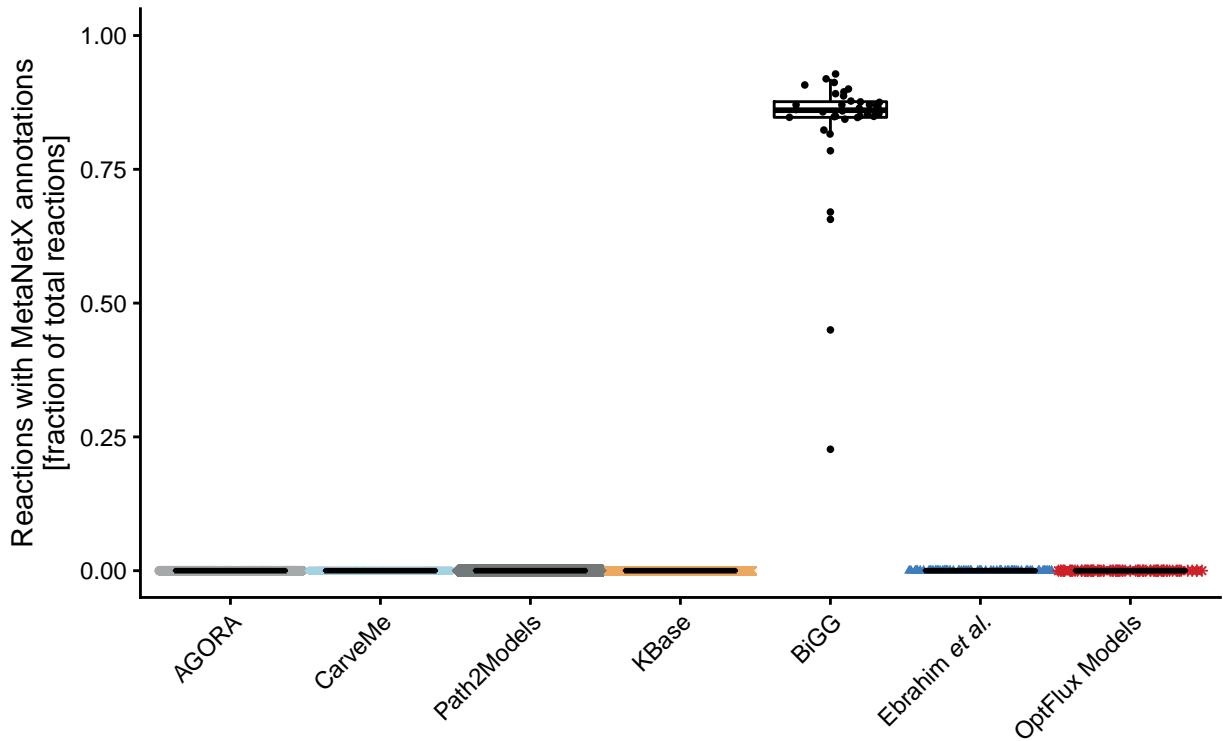


Figure S41: Reaction MetaNetX.reaction Annotation

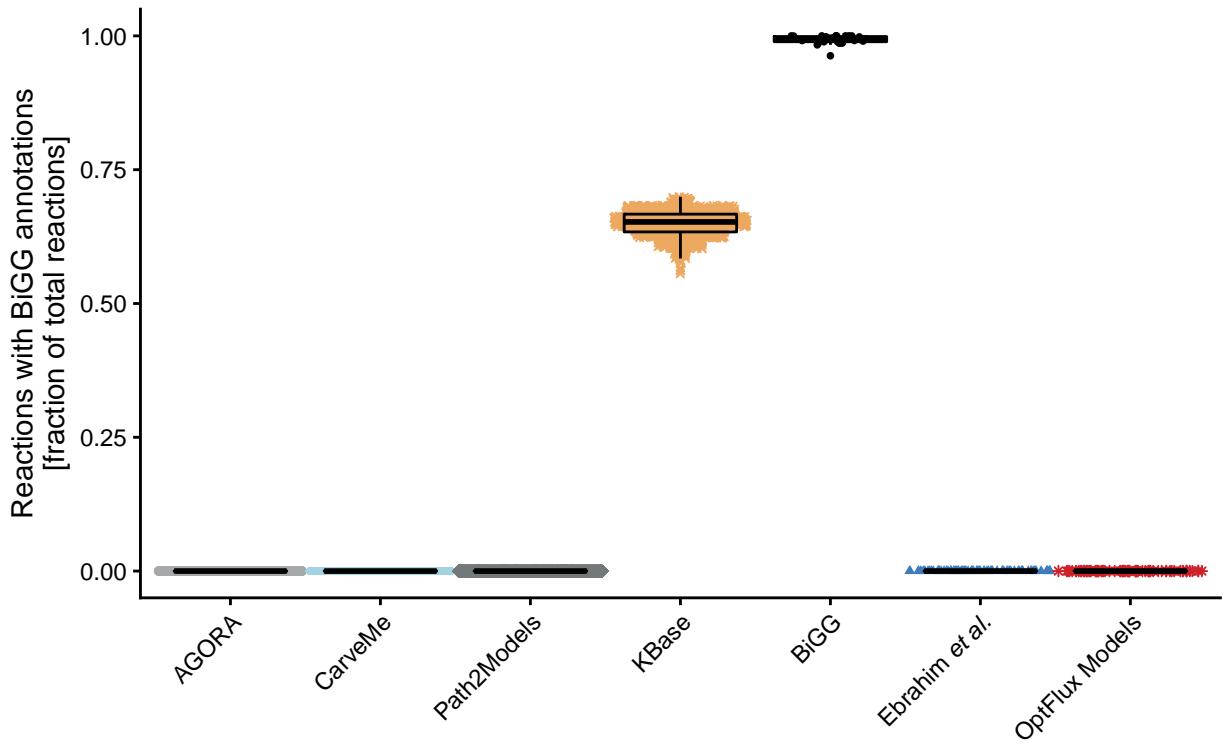


Figure S42: Reaction BiGG.reaction Annotation

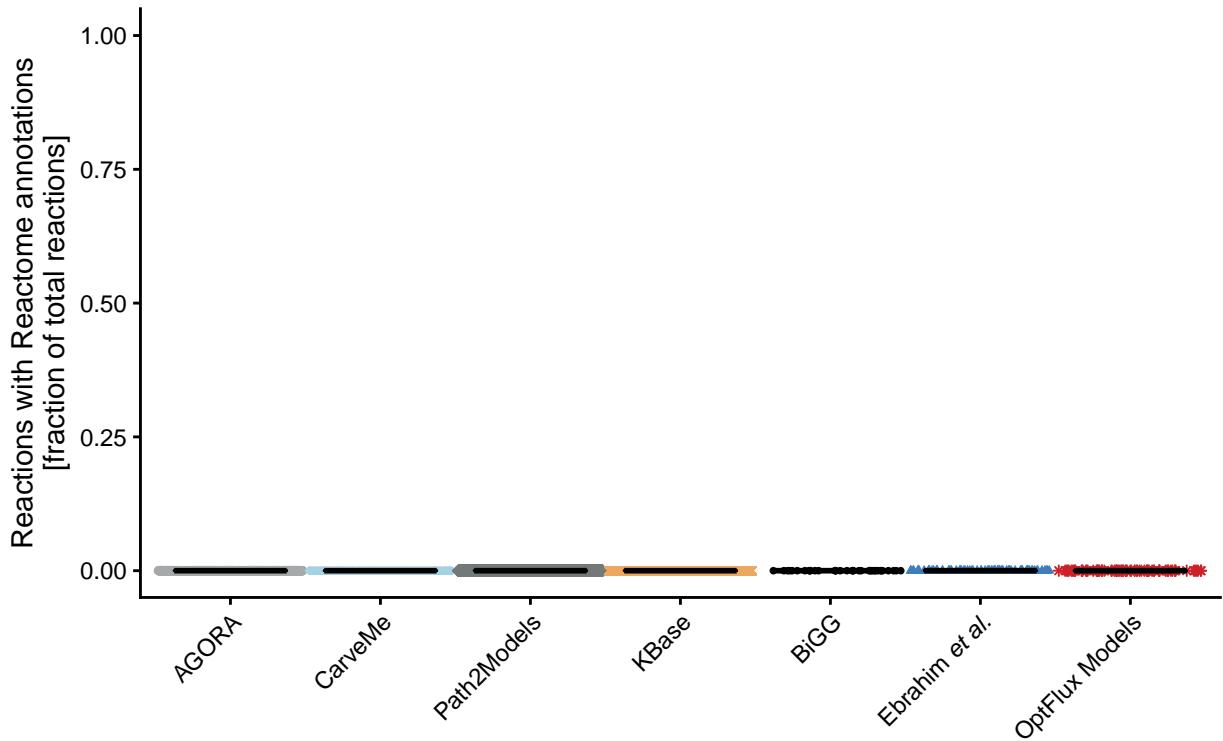


Figure S43: Reaction Reactome Annotation

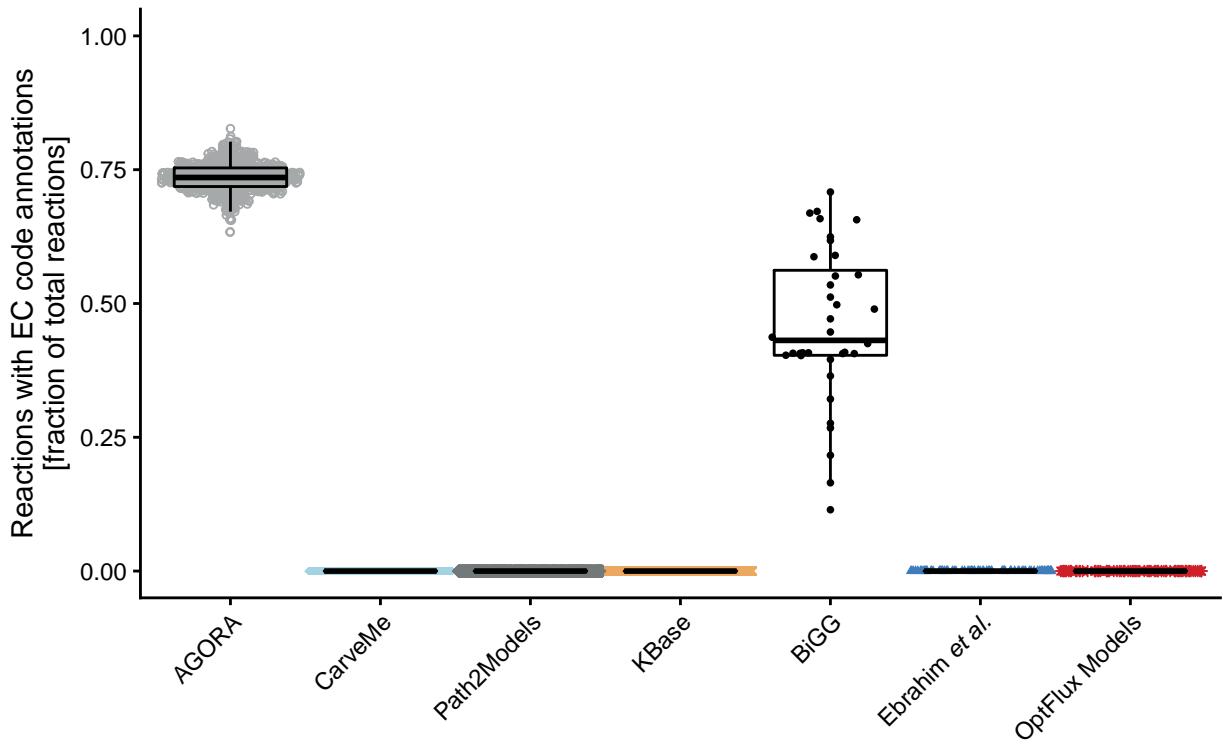


Figure S44: Reaction Enzyme Classification Annotation

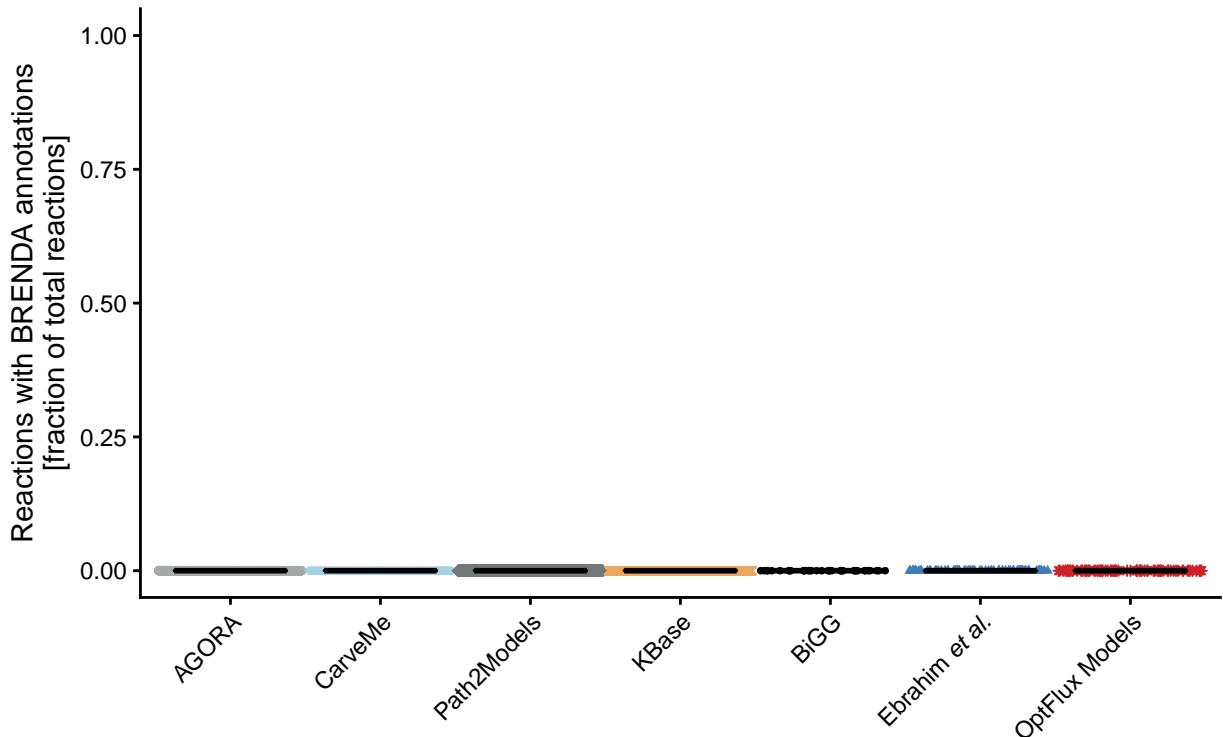


Figure S45: Reaction BRENDA Annotation

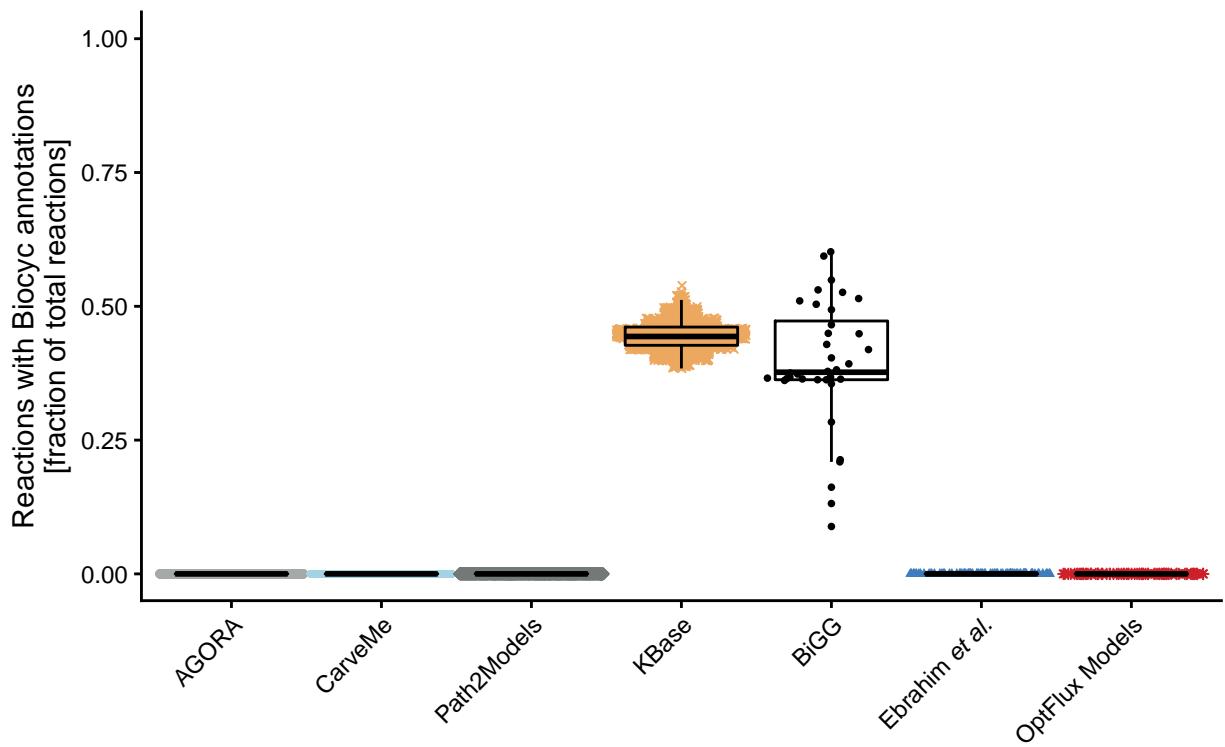


Figure S46: Reaction BioCyc Annotation

3.3.3.2 Reaction Annotation Conformity Per Database

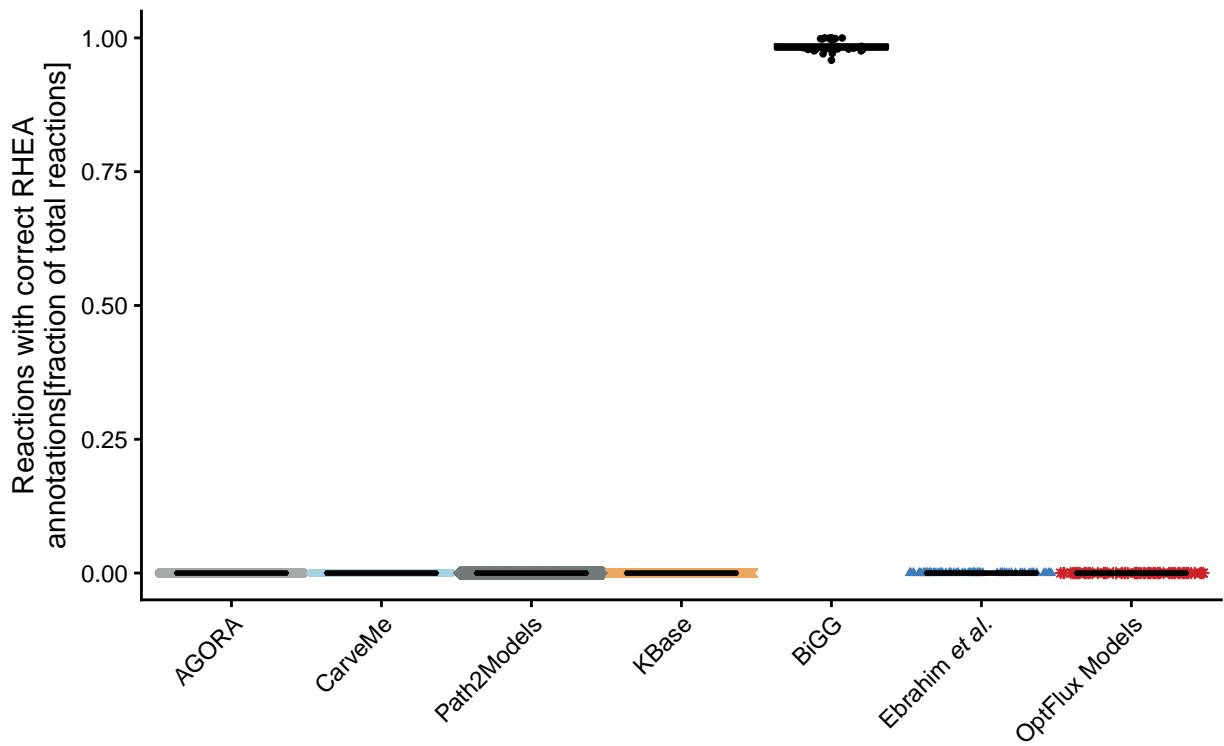


Figure S47: Correct Reaction Rhea Annotation

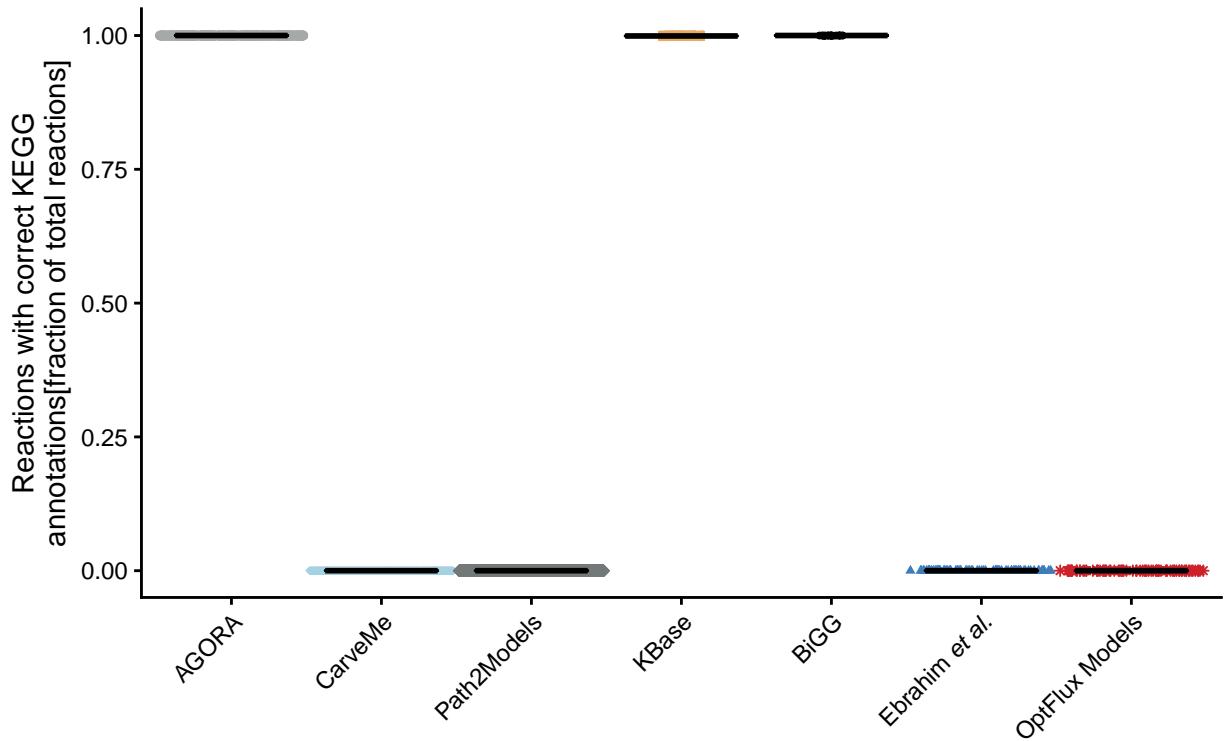


Figure S48: Correct Reaction KEGG.reaction Annotation

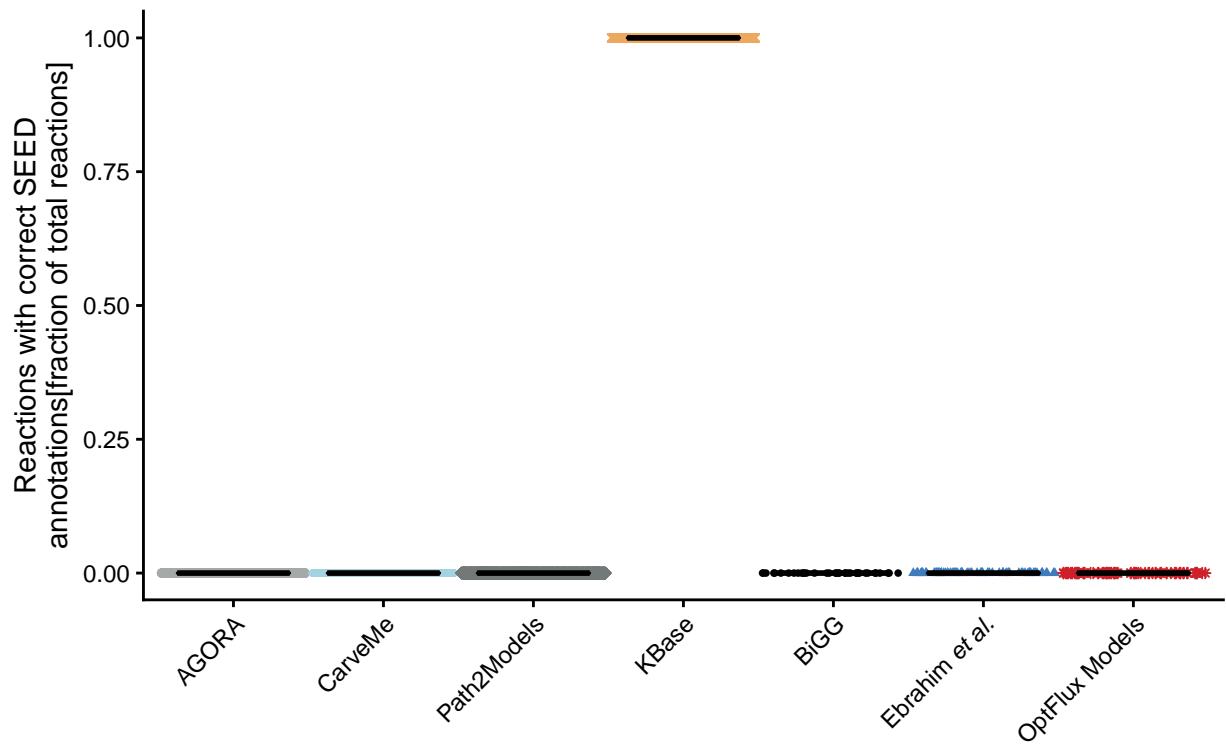


Figure S49: Correct Reaction SEED.reaction Annotation

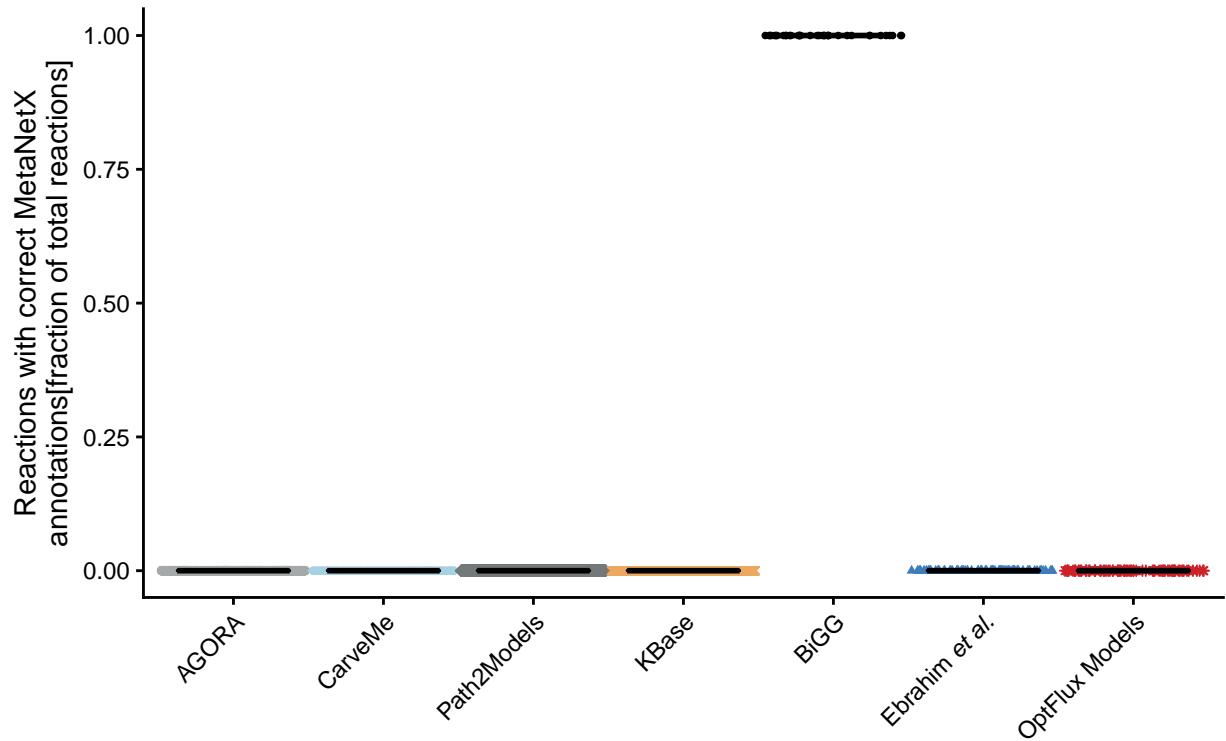


Figure S50: Correct Reaction MetaNetX.reaction Annotation

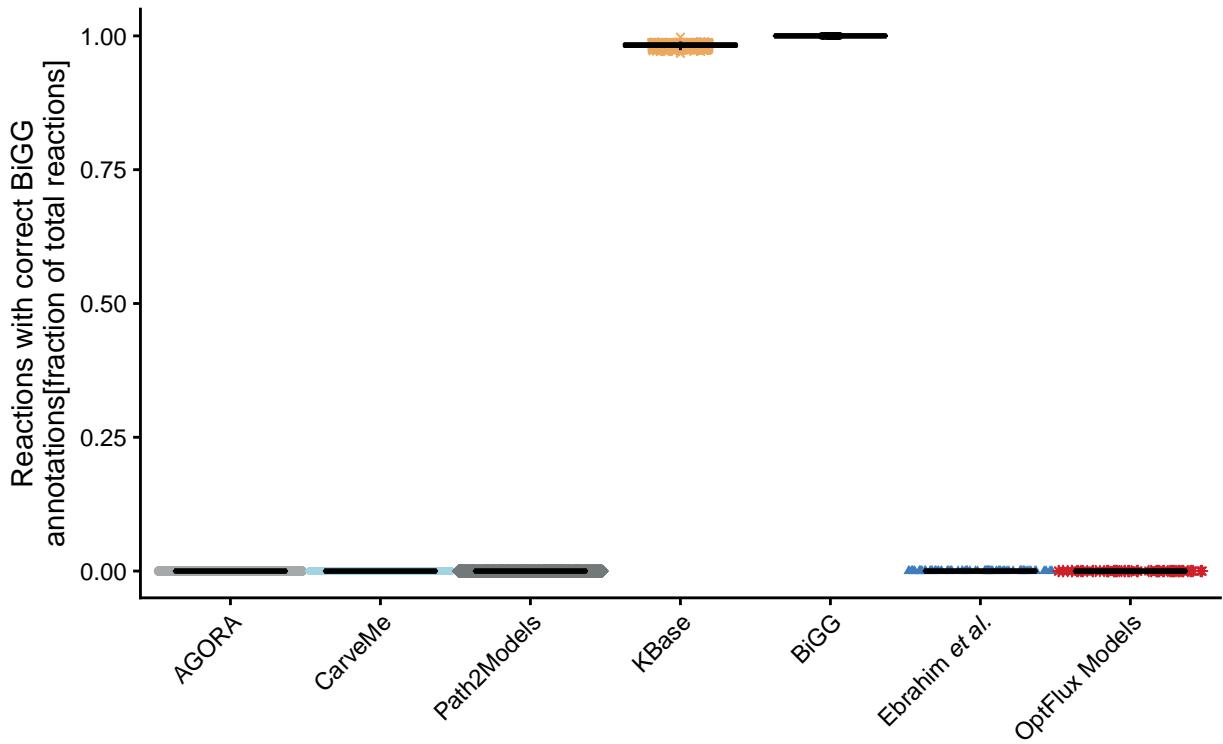


Figure S51: Correct Reaction BiGG.reaction Annotation

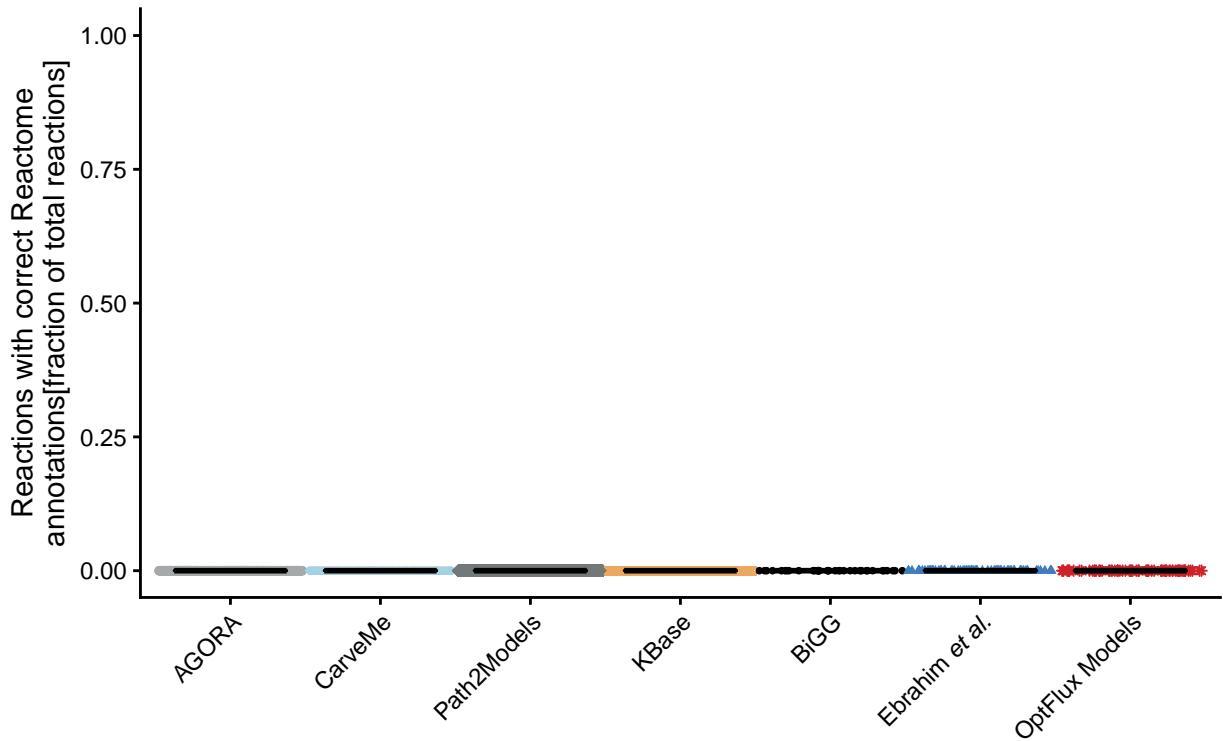


Figure S52: Correct Reaction Reactome Annotation

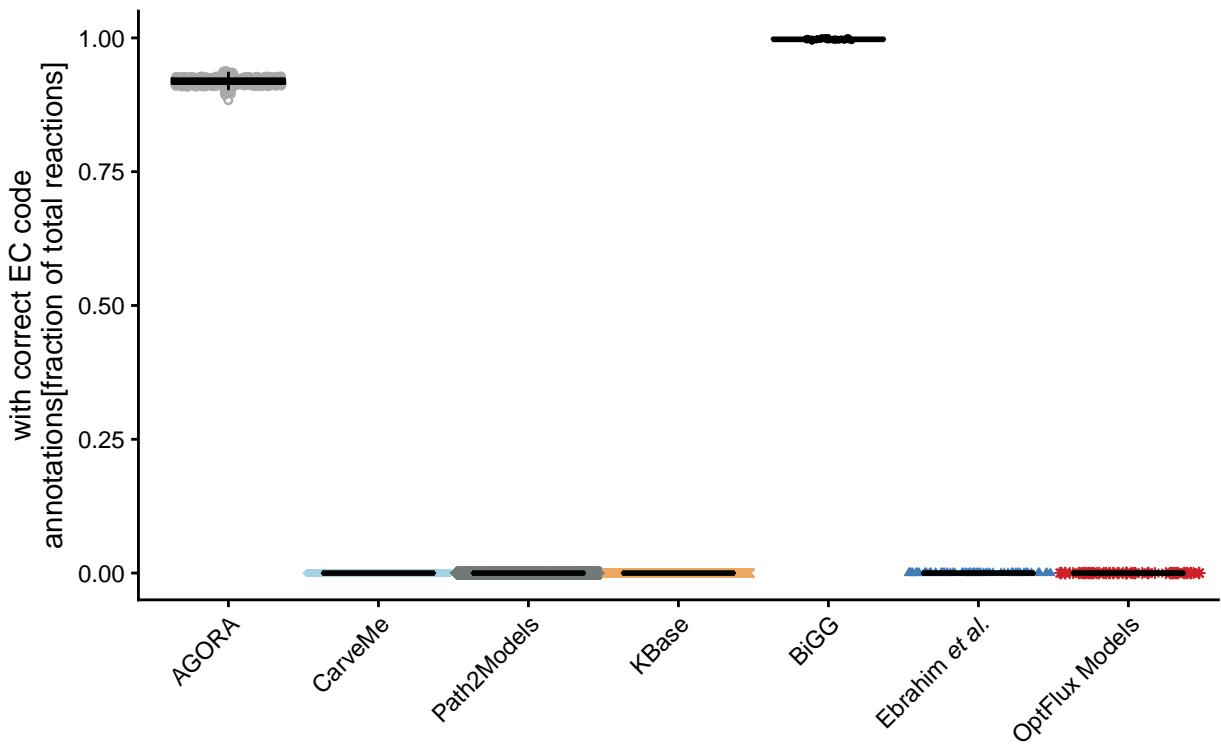


Figure S53: Correct Reaction Enzyme Classification Annotation

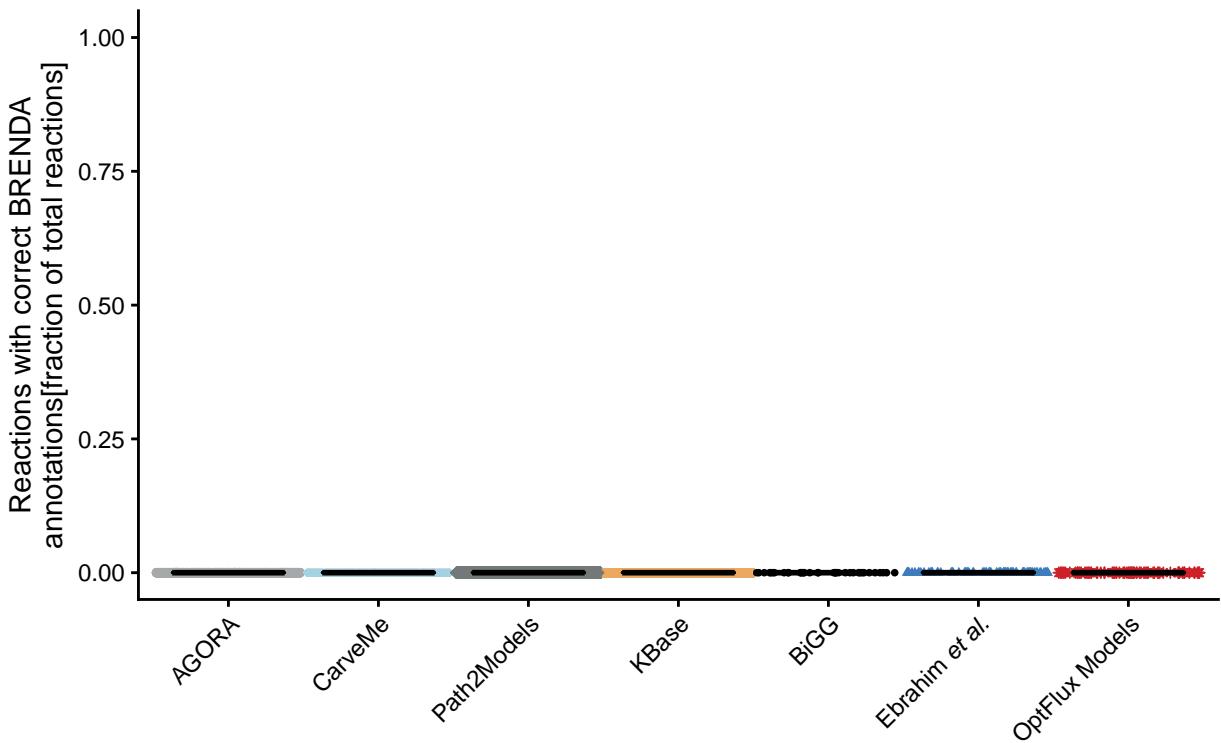


Figure S54: Correct Reaction BRENDA Annotation

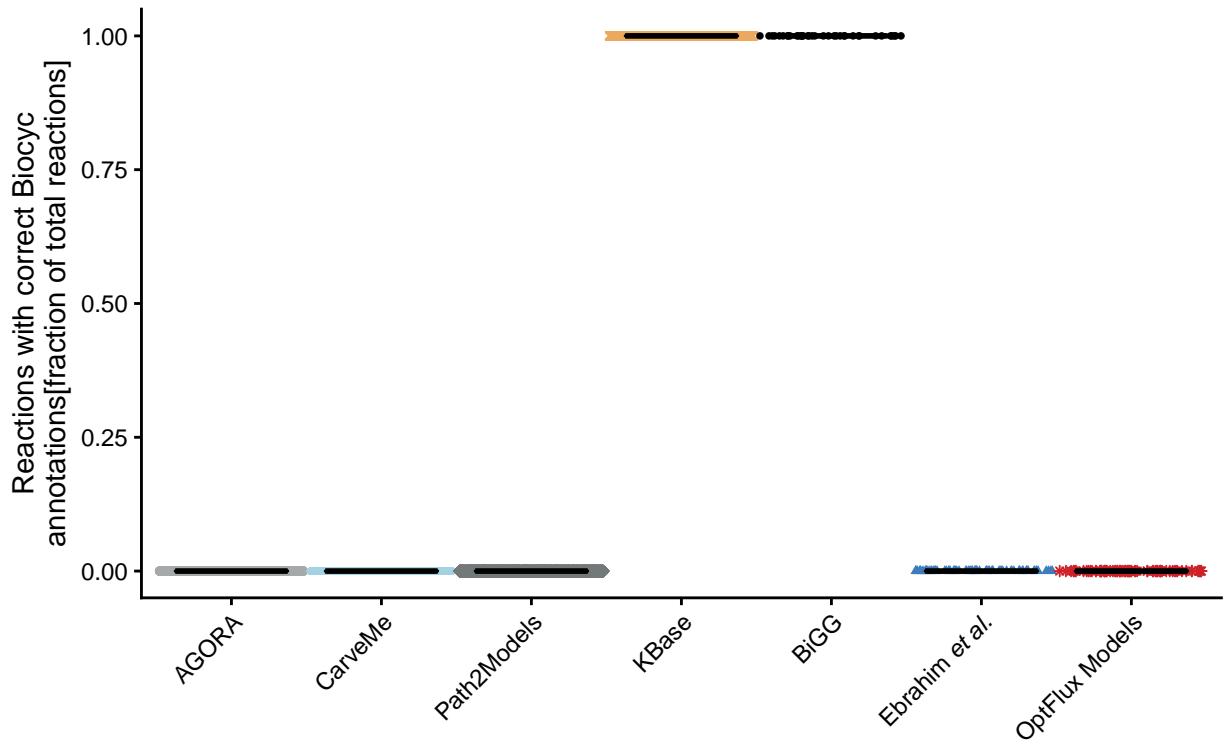


Figure S55: Correct Reaction BioCyc Annotation

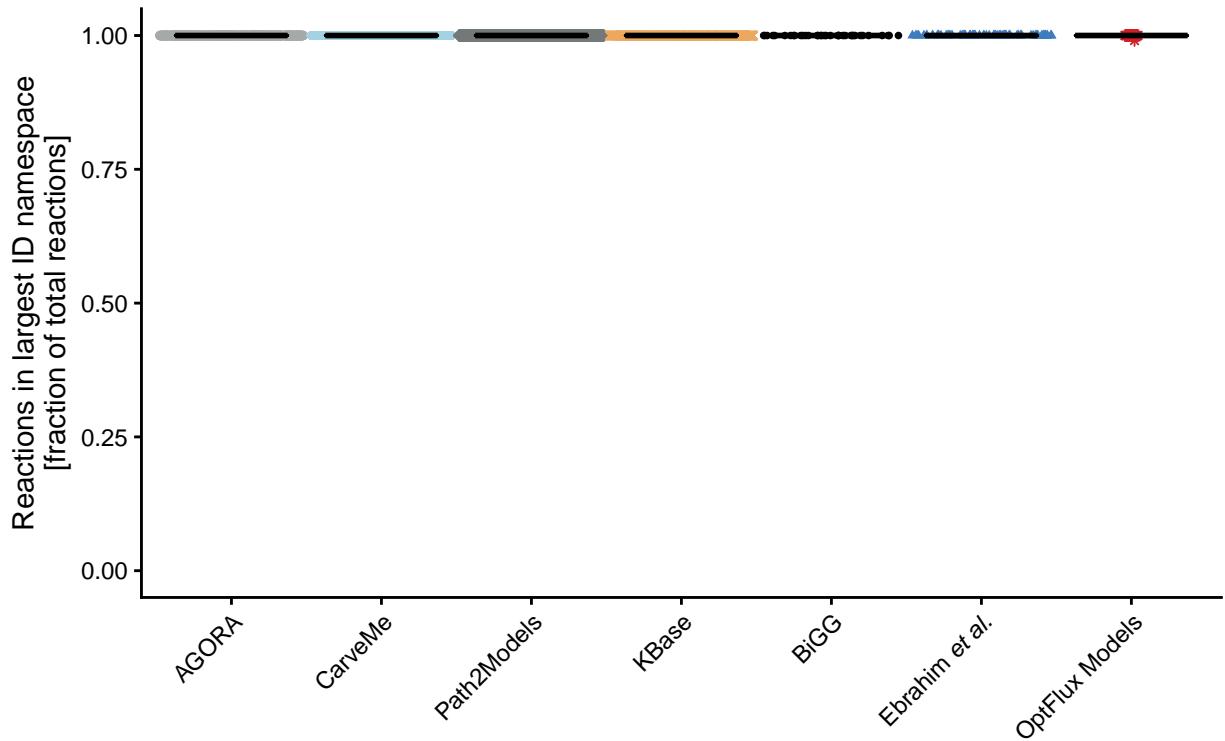


Figure S56: Uniform Reaction Identifier Namespace

3.3.4 Annotation - Genes

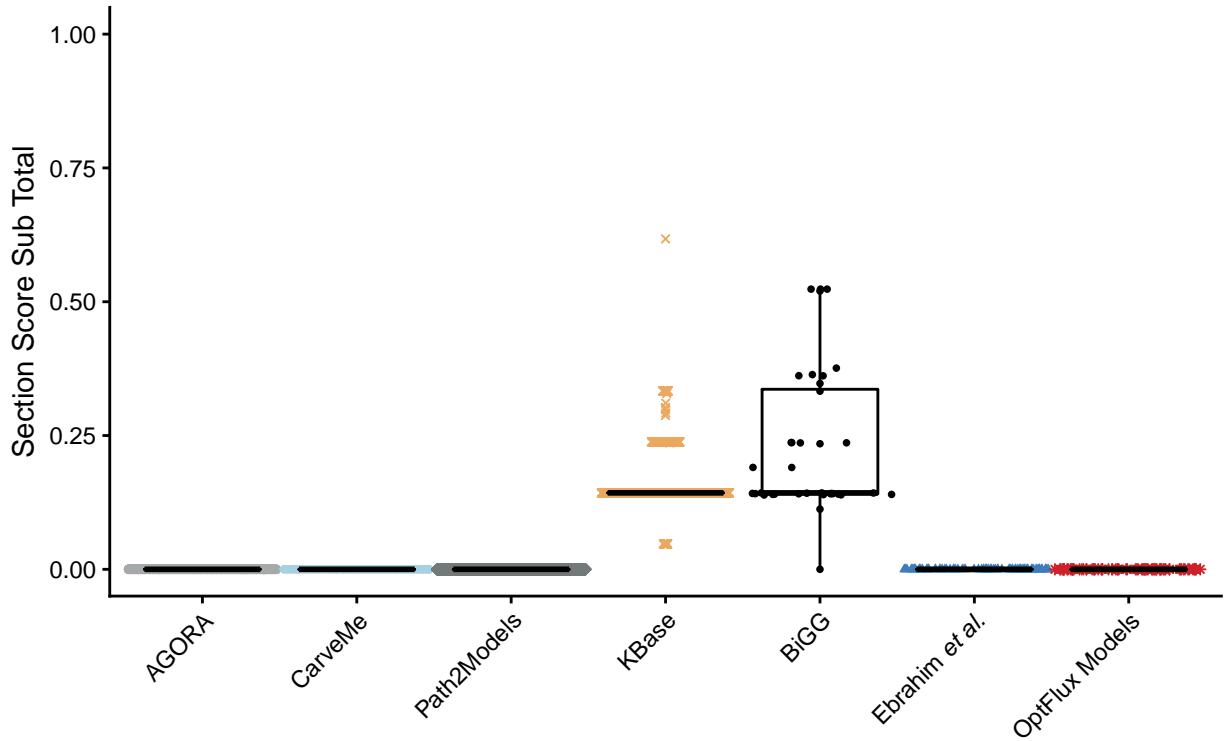


Figure S57: Annotation - Genes. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

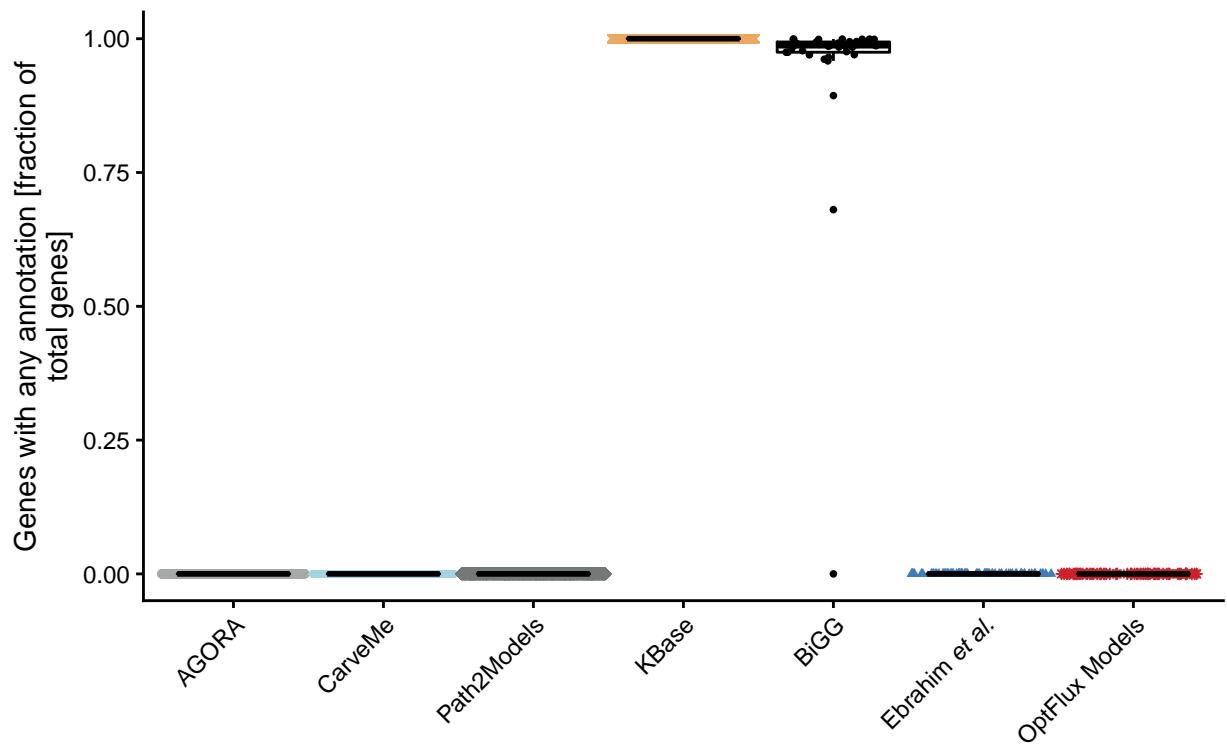


Figure S58: Presence of Gene Annotation

3.3.4.1 Gene Annotations Per Database

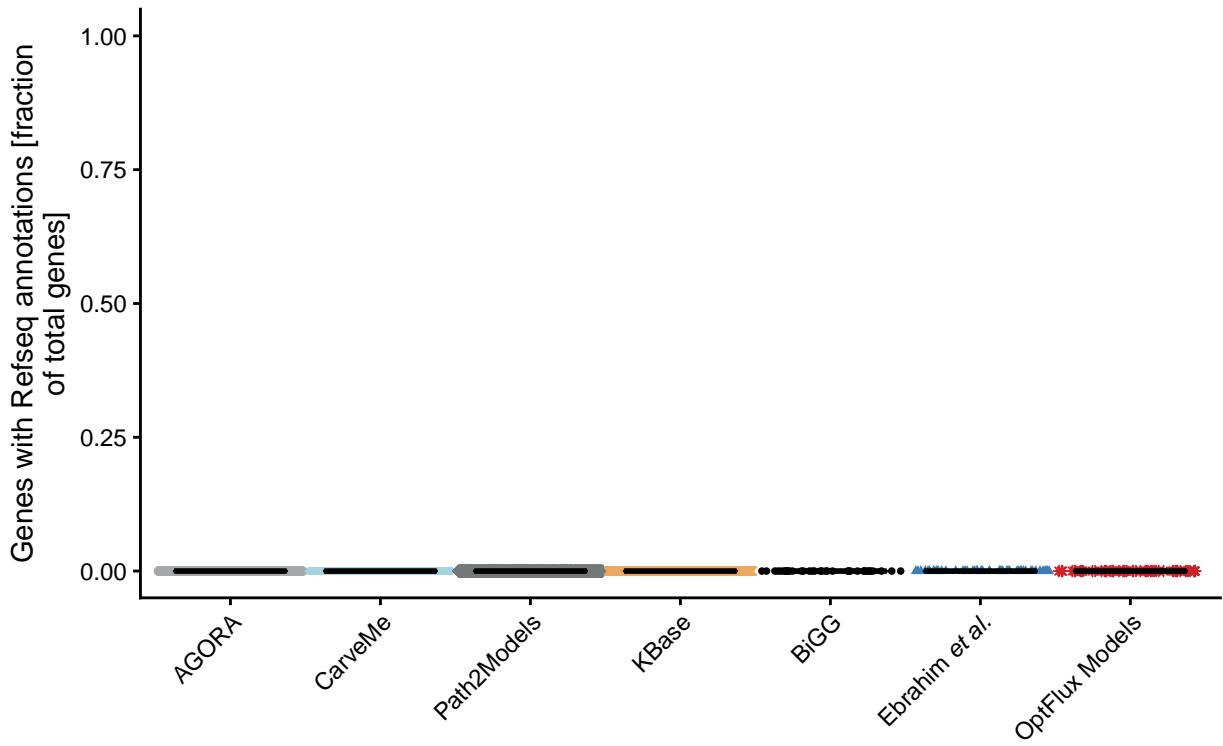


Figure S59: Gene RefSeq Annotation

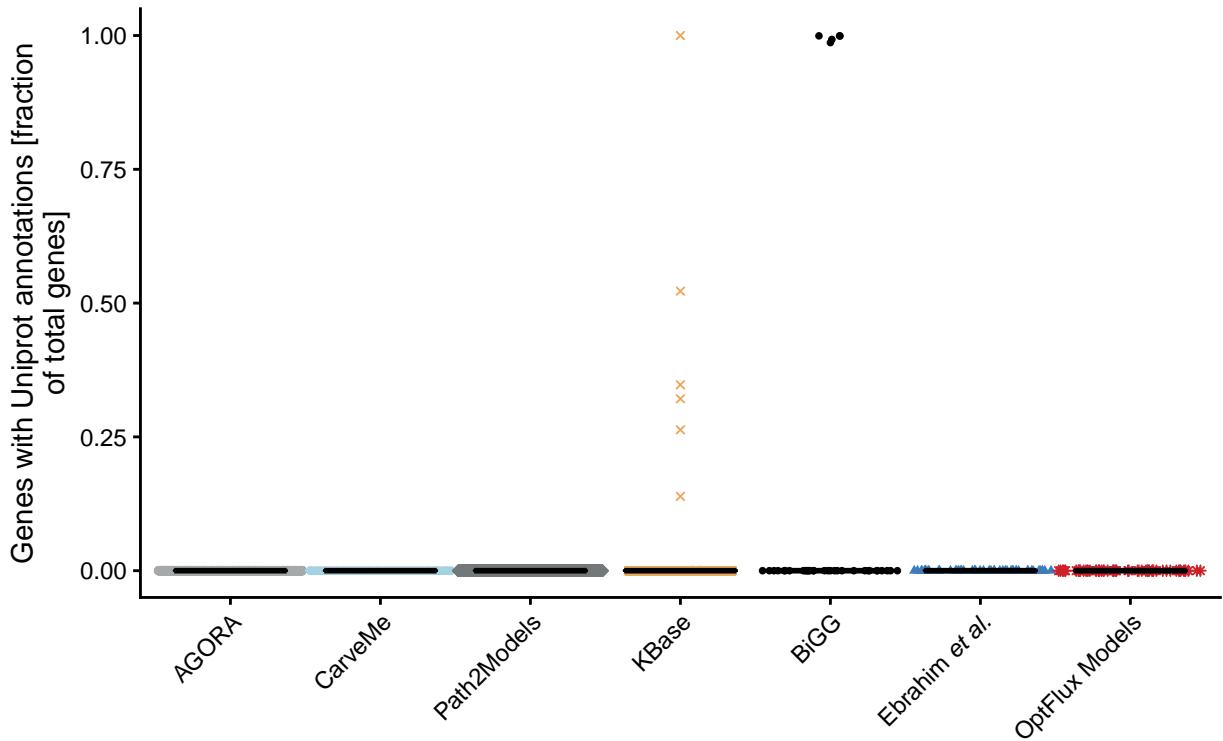


Figure S60: Gene UniProt Annotation

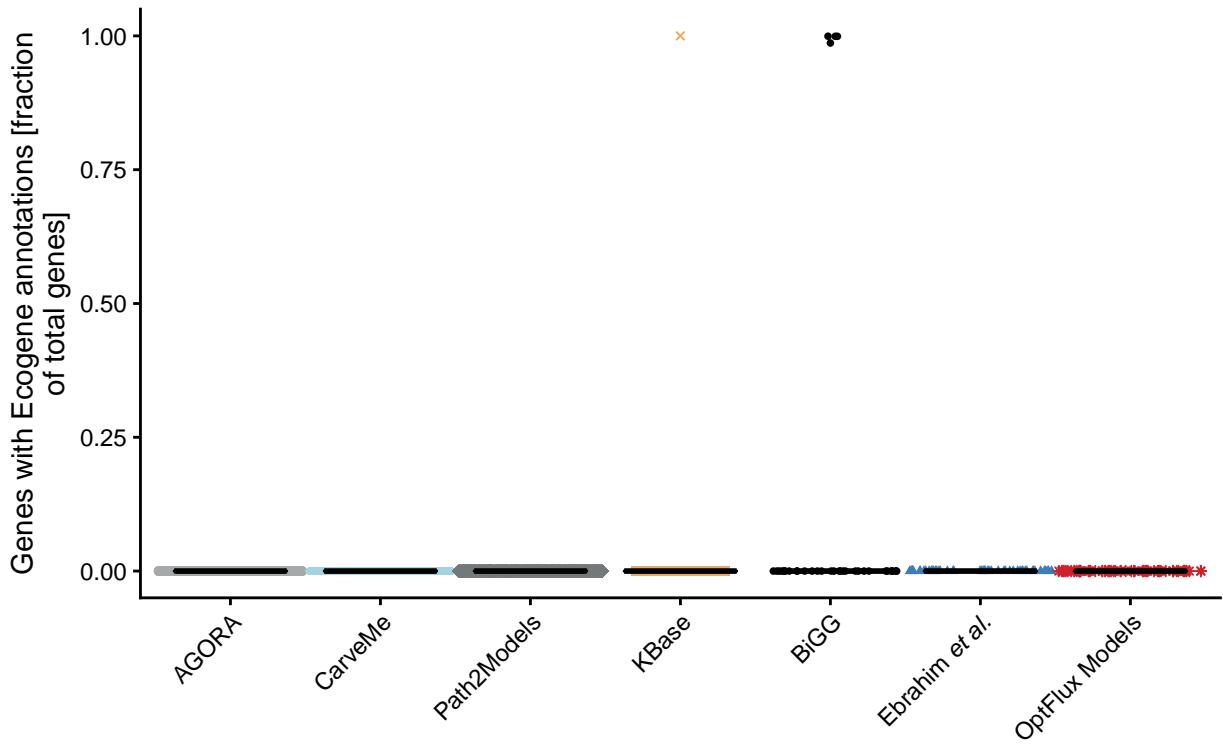


Figure S61: Gene EcoGene Annotation

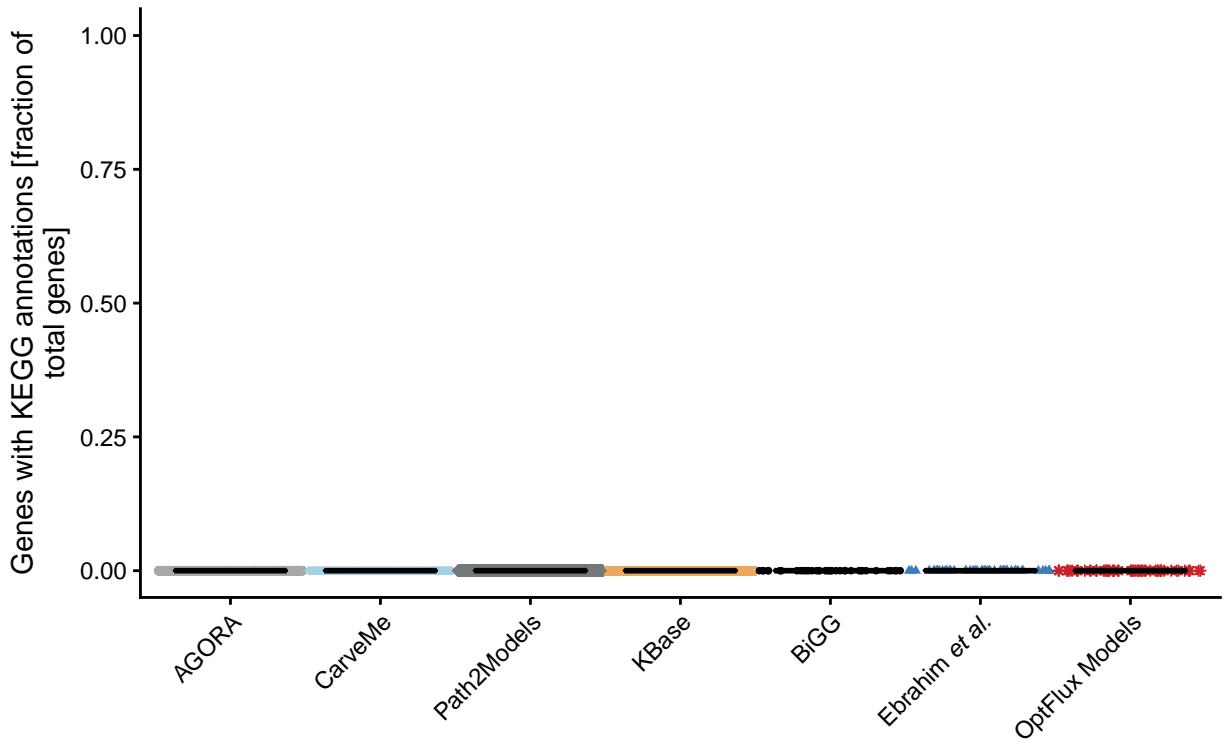


Figure S62: Gene KEGG.genes Annotation

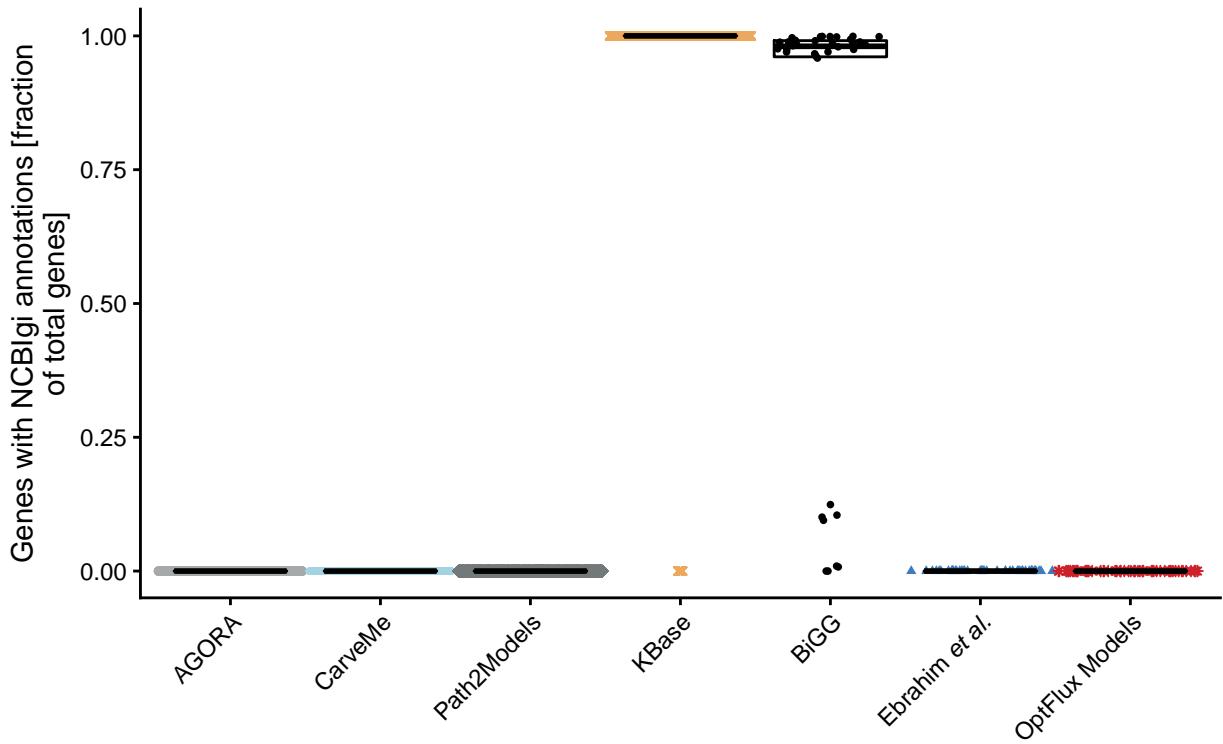


Figure S63: Gene NCBItgi Annotation

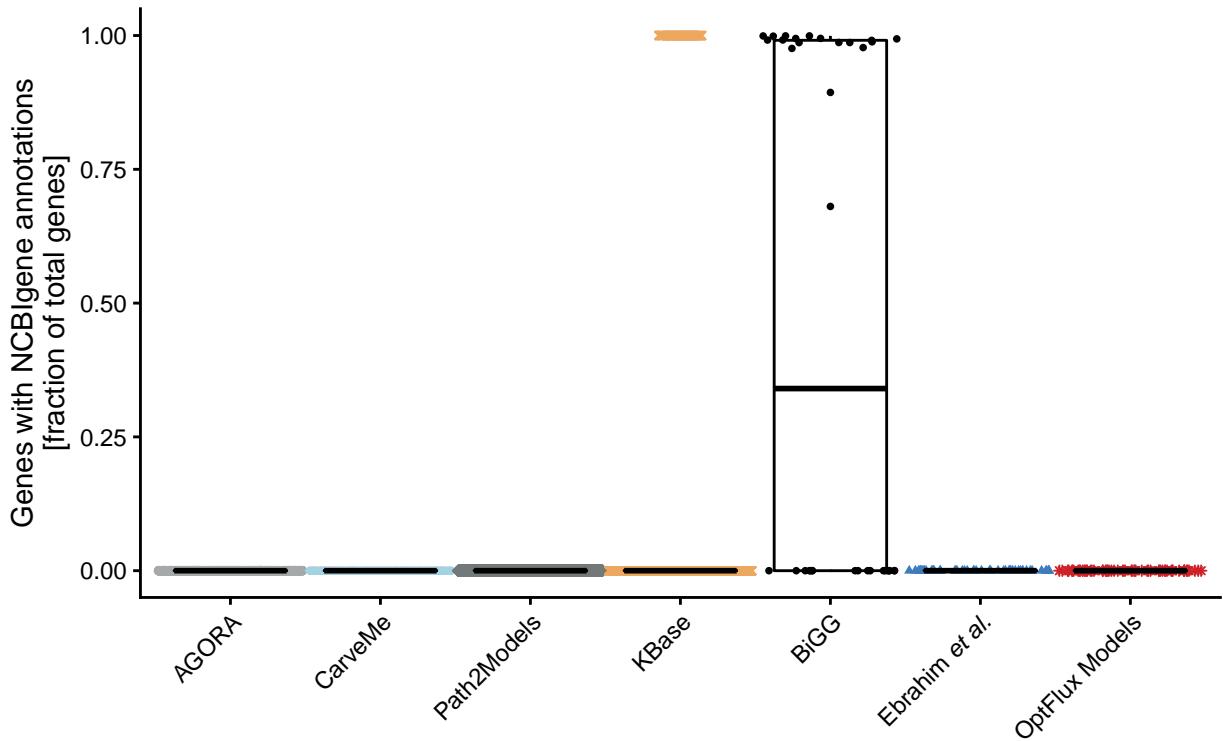


Figure S64: Gene NCBItgene Annotation

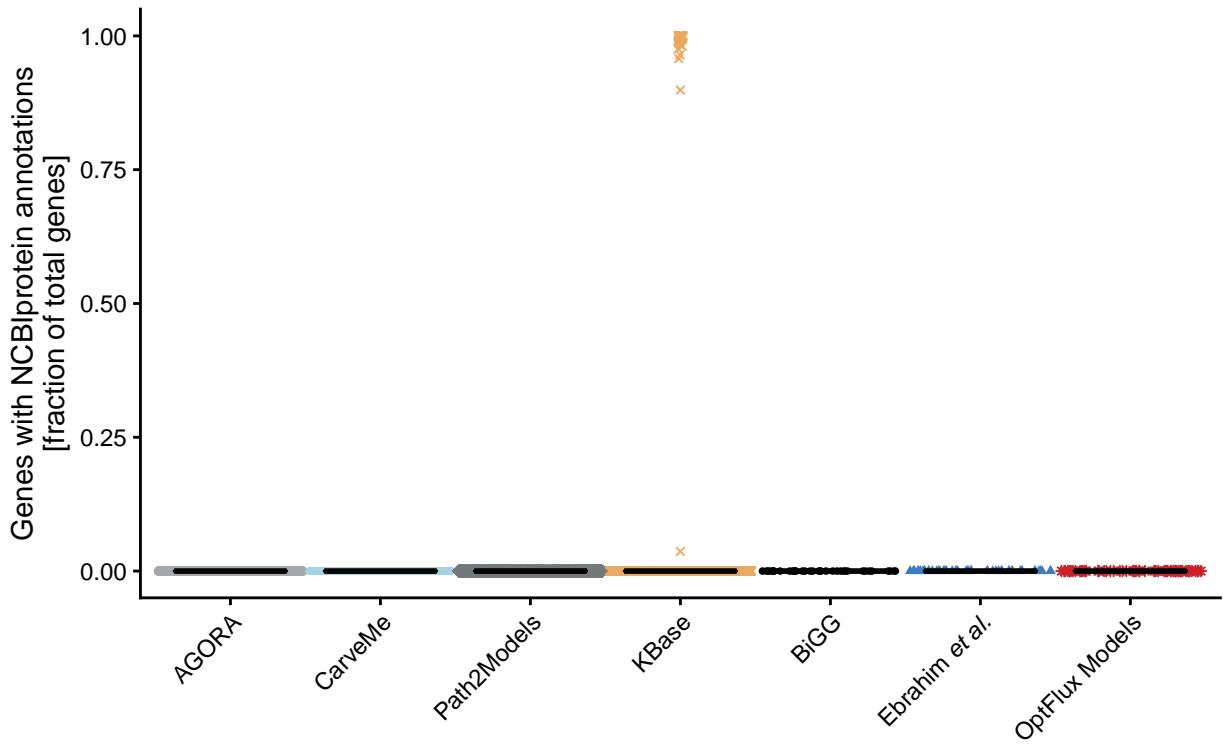


Figure S65: Gene NCBInfo Annotation

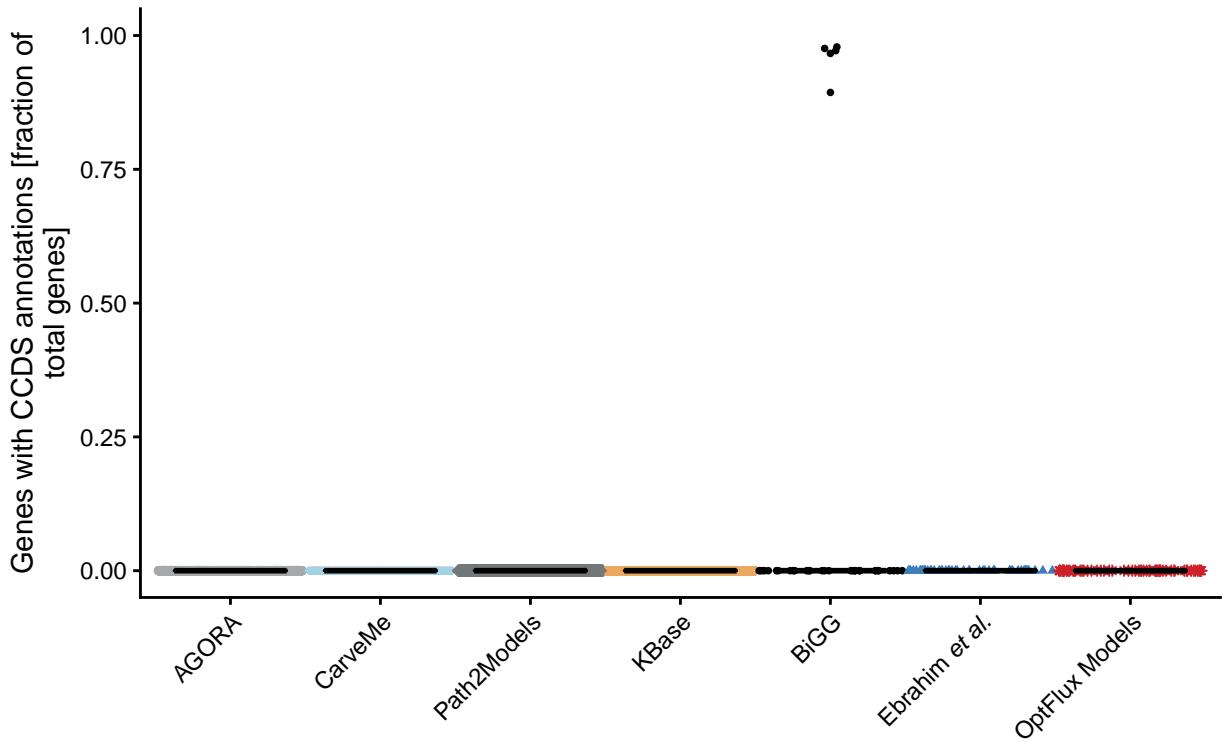


Figure S66: Gene CCDS Annotation

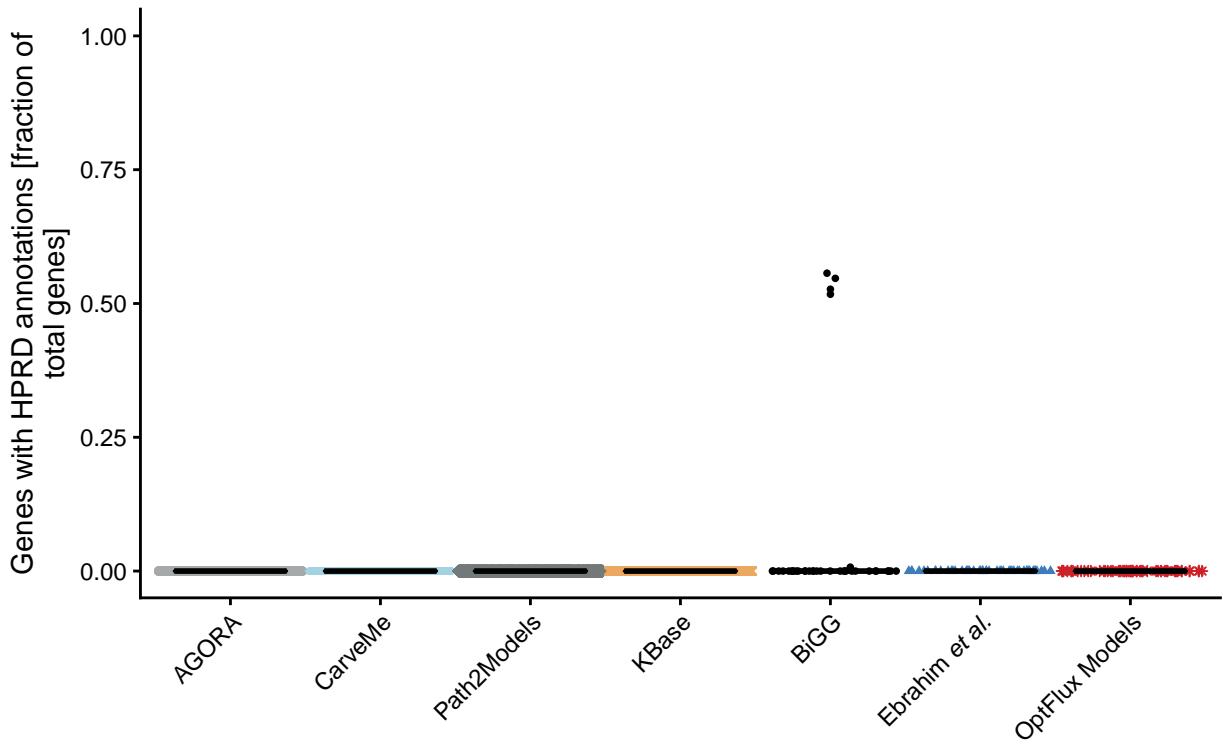


Figure S67: Gene HPRD Annotation

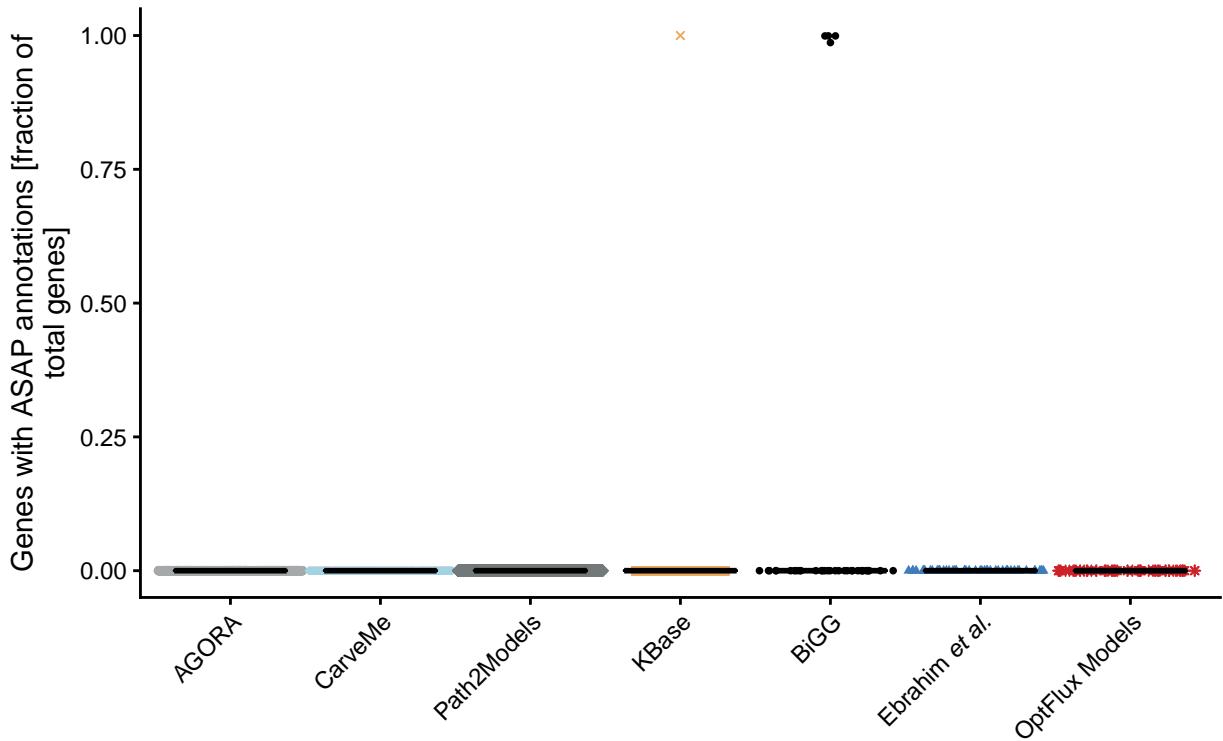


Figure S68: Gene ASAP Annotation

3.3.4.2 Gene Annotation Conformity Per Database

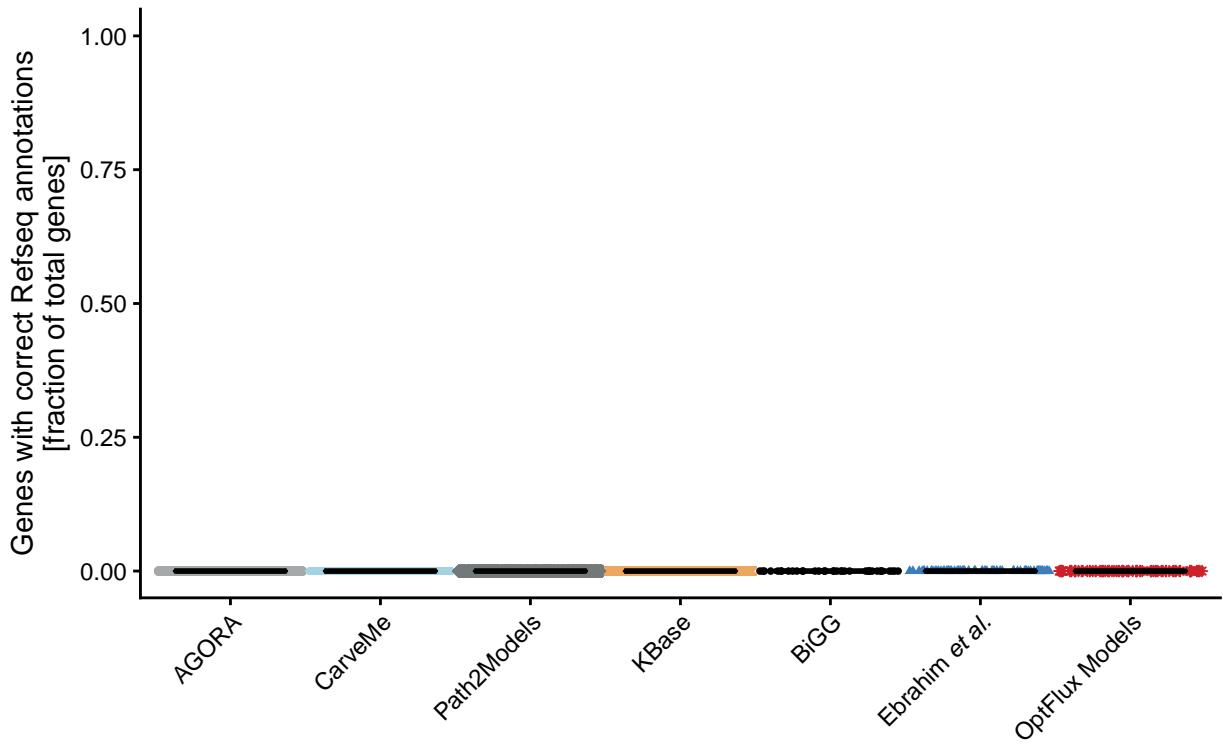


Figure S69: Correct Gene RefSeq Annotation

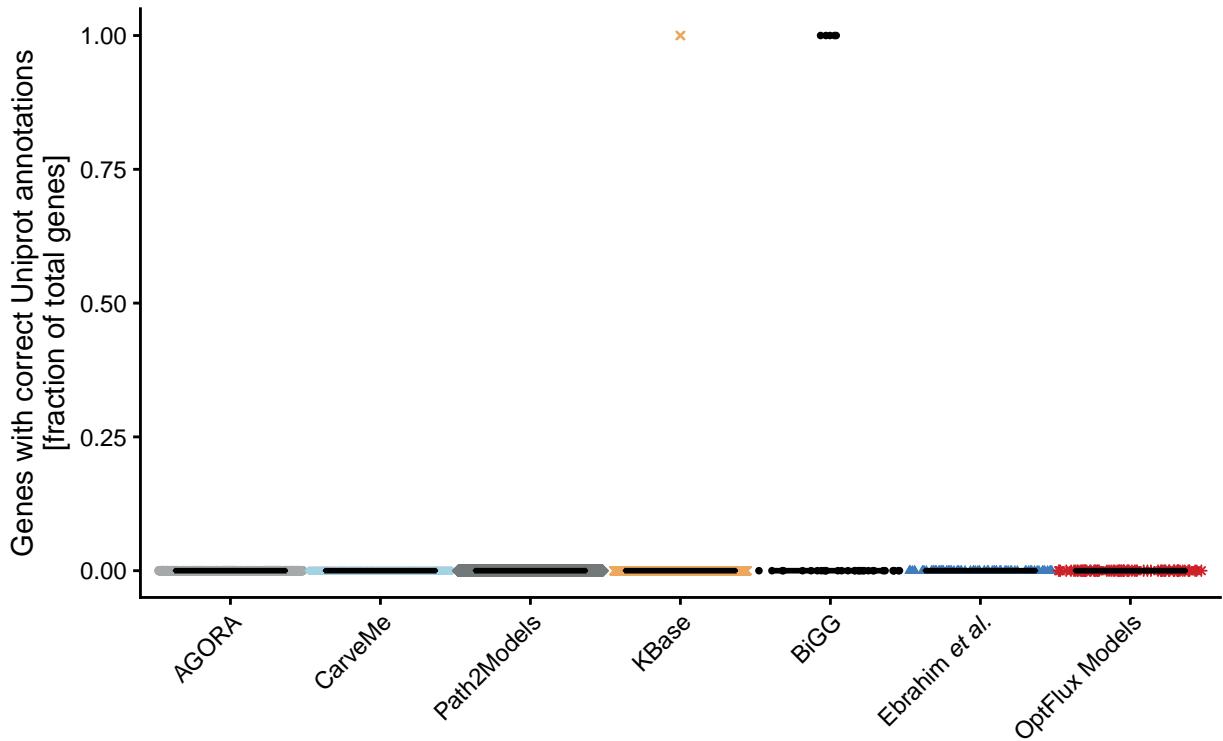


Figure S70: Correct Gene UniProt Annotation

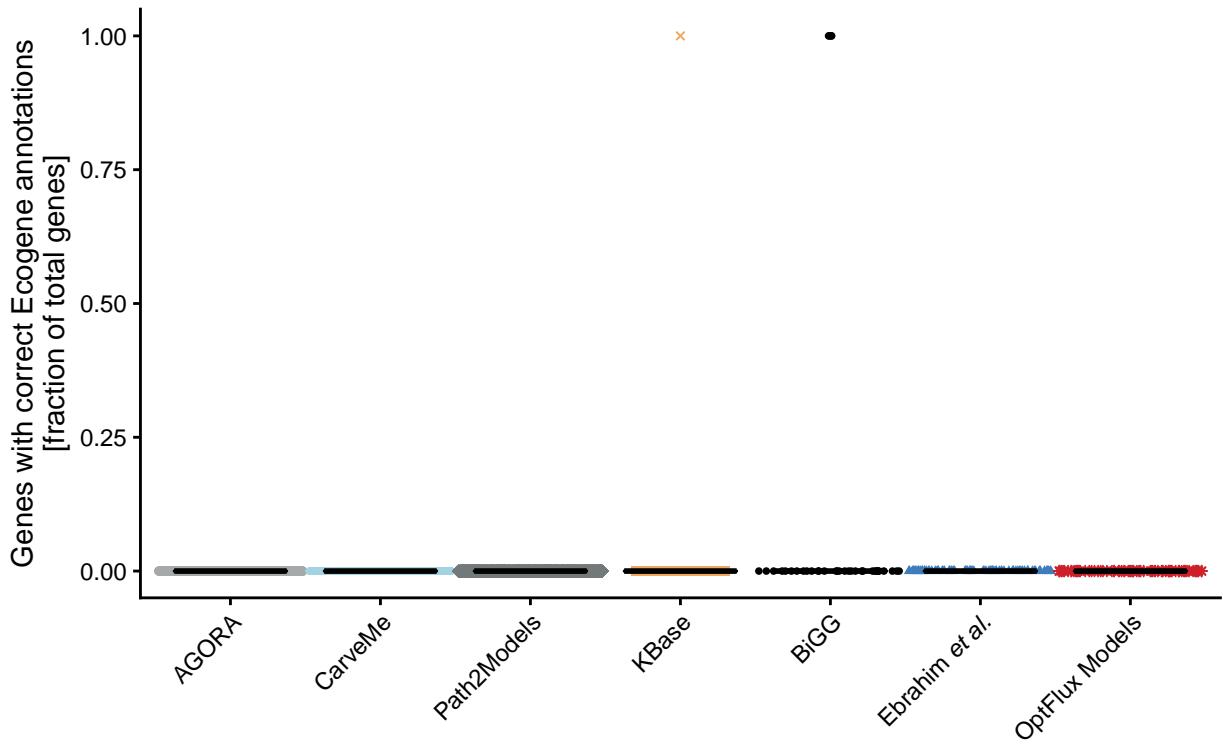


Figure S71: Correct Gene EcoGene Annotation

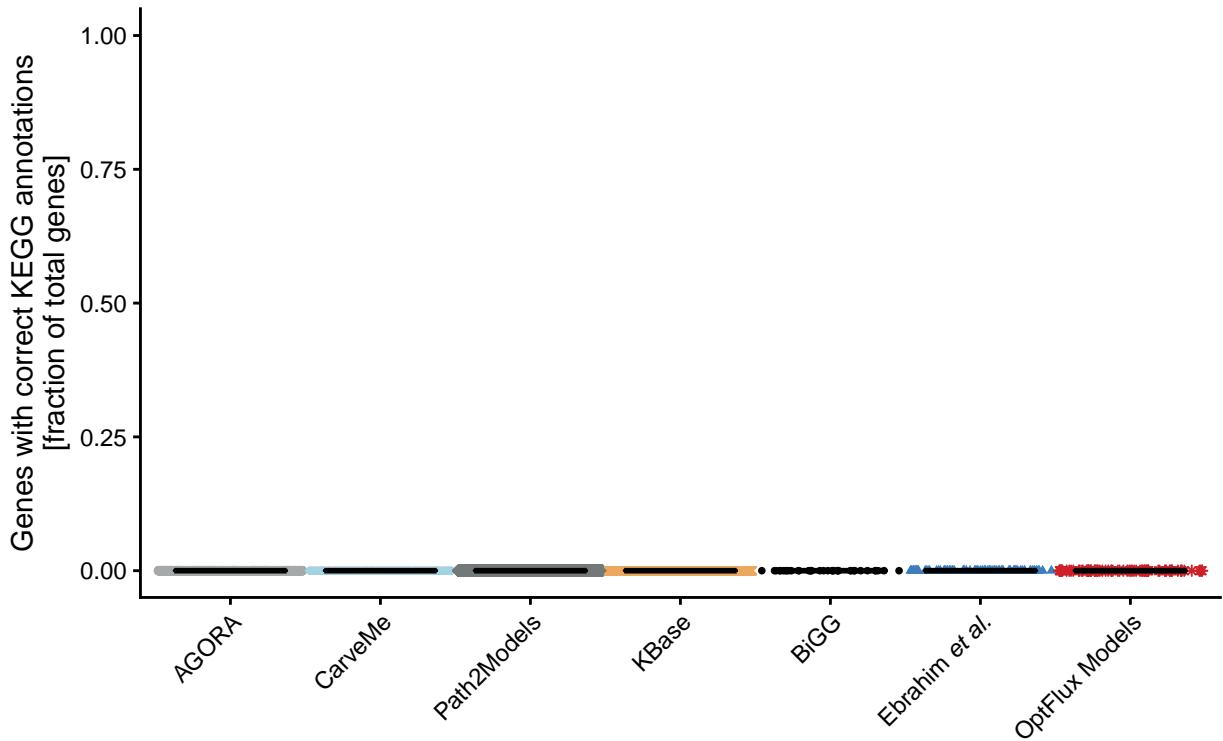


Figure S72: Correct Gene KEGG.genes Annotation

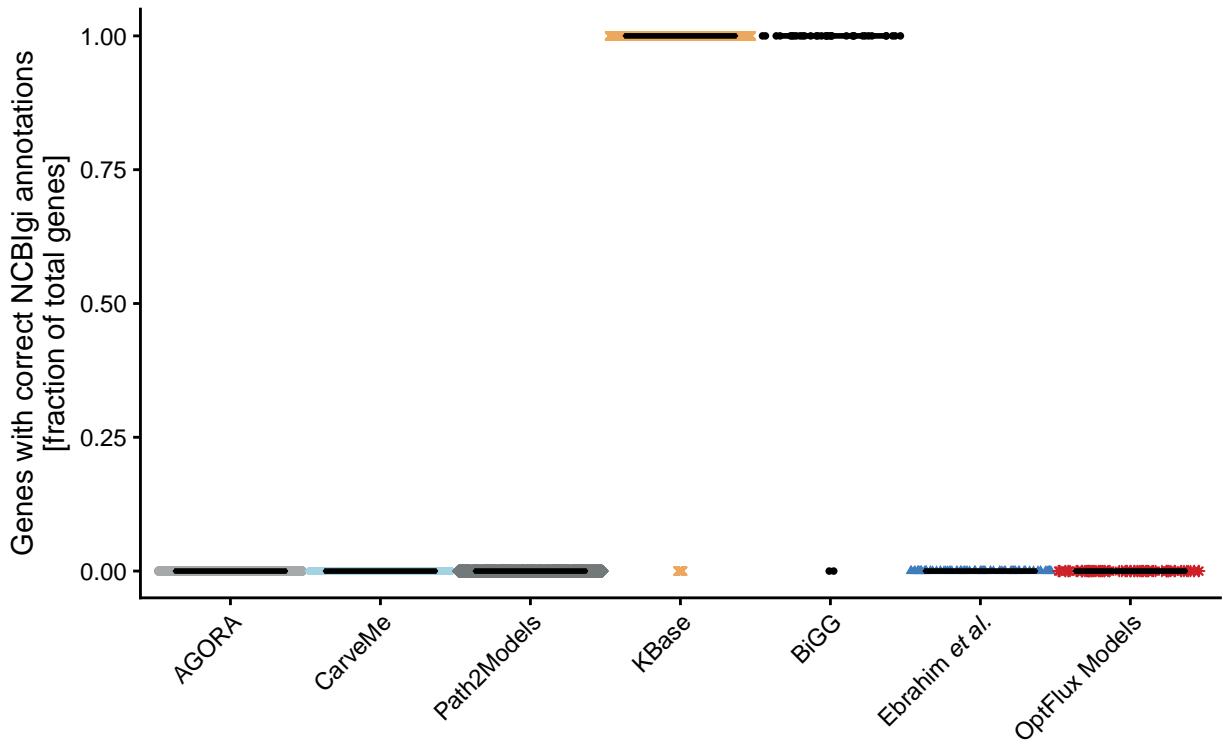


Figure S73: Correct Gene NCBIgi Annotation

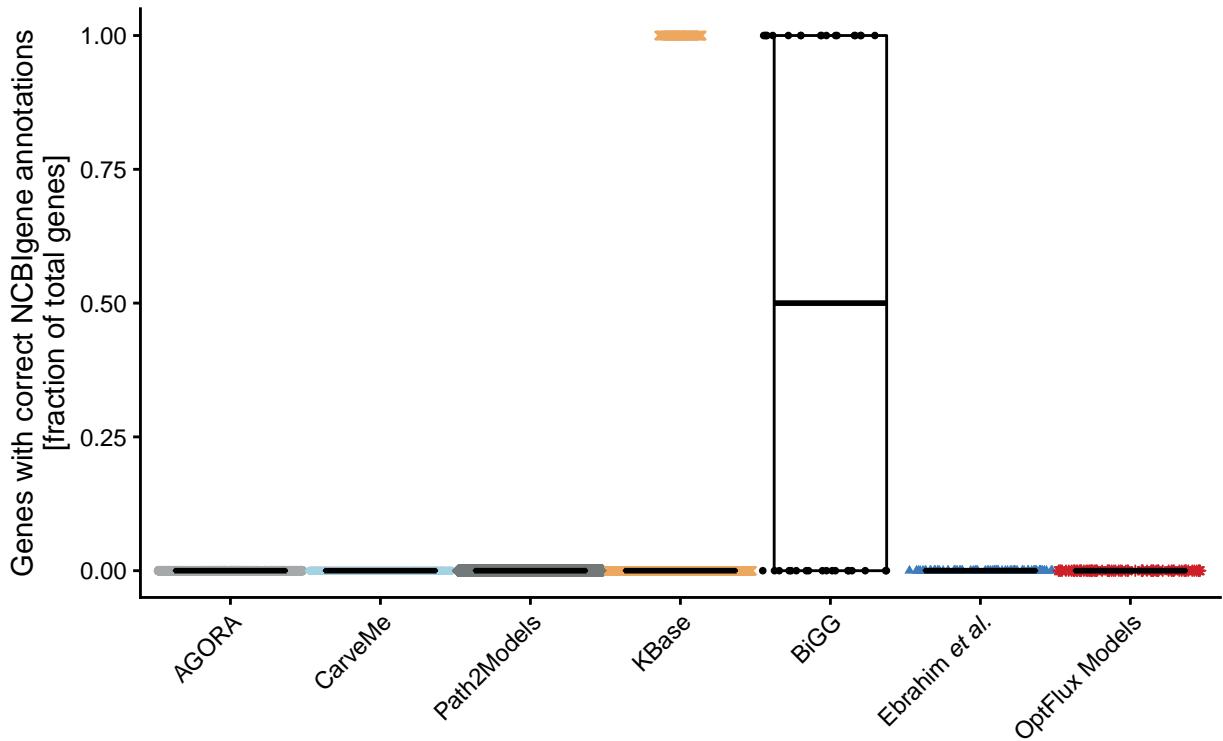


Figure S74: Correct Gene NCBIgene Annotation

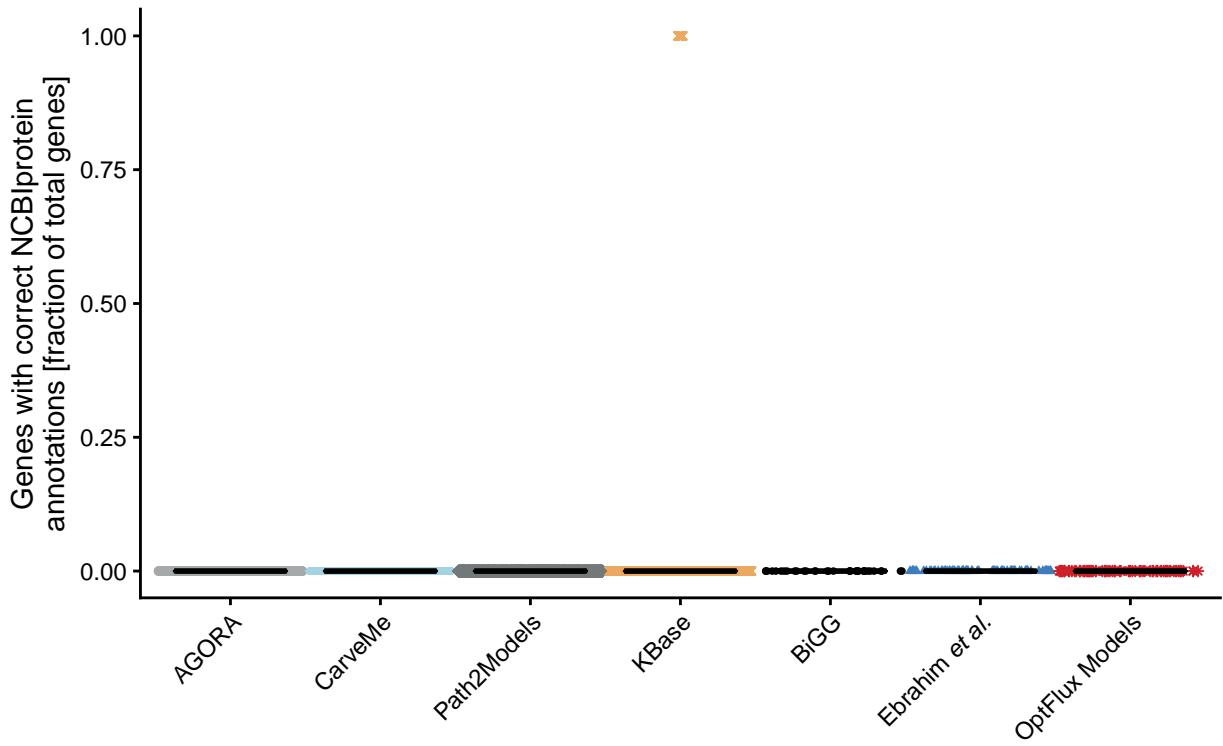


Figure S75: Correct Gene NCBIprotein Annotation

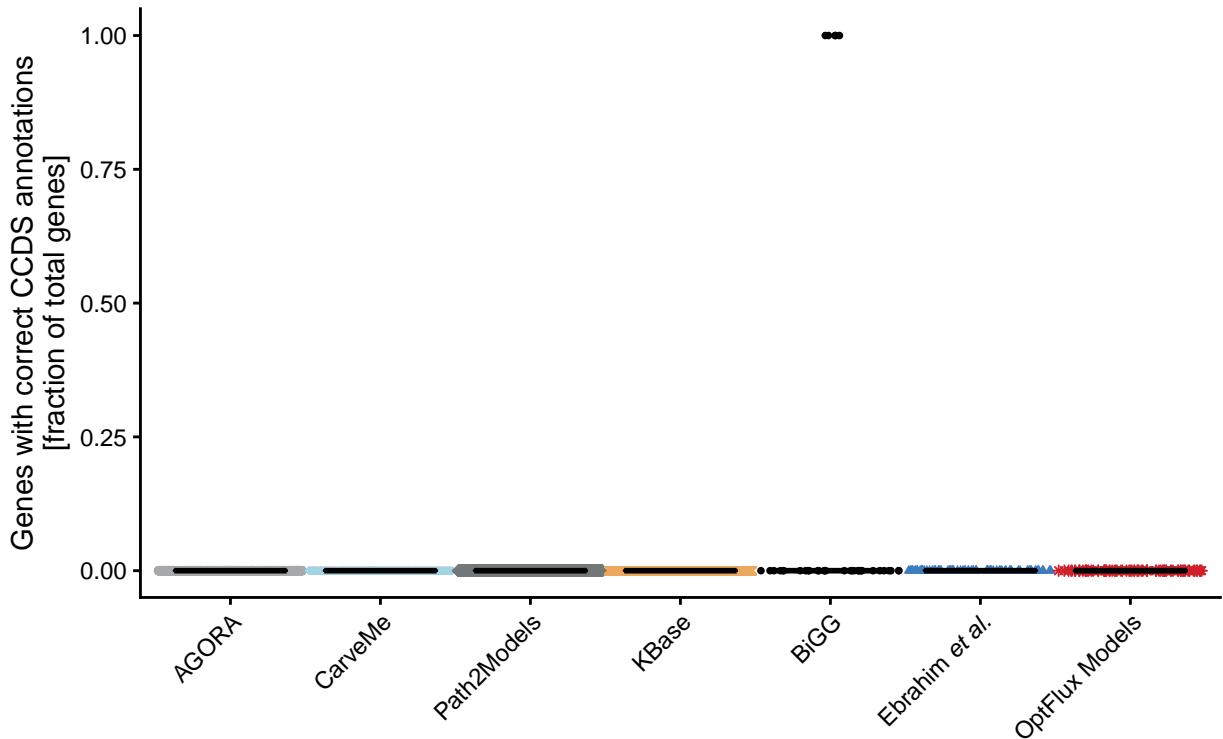


Figure S76: Correct Gene CCDS Annotation

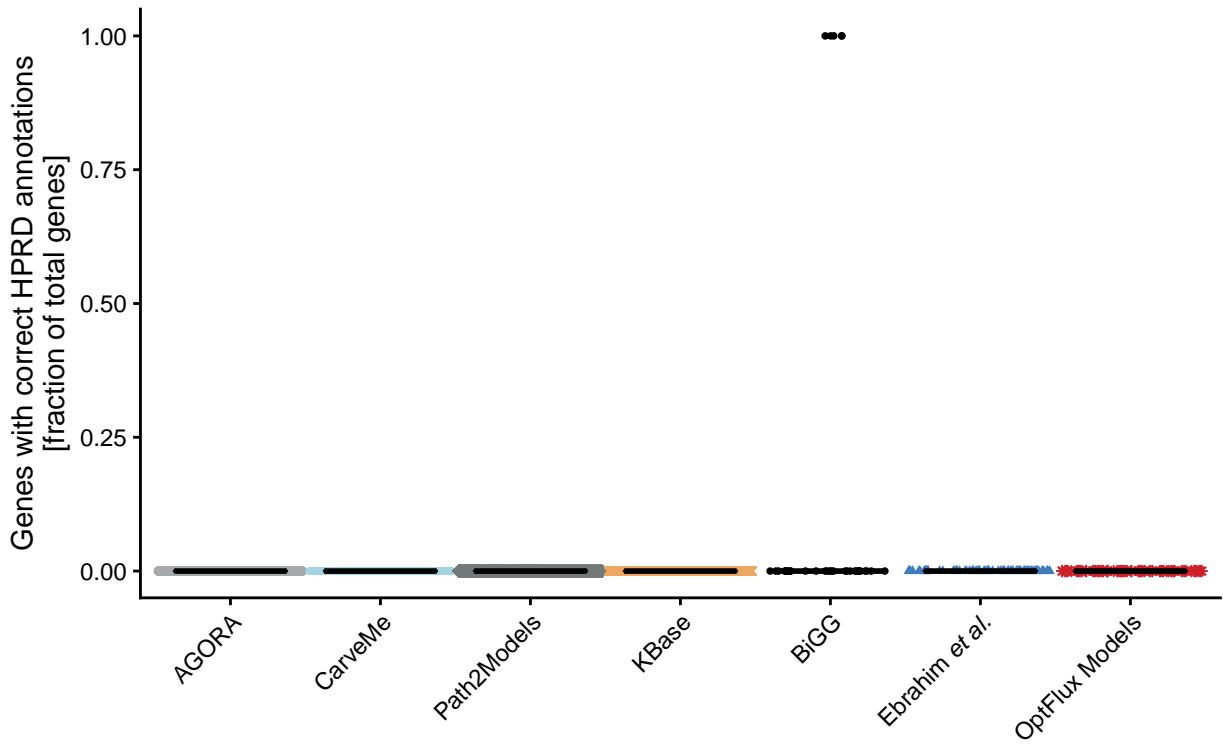


Figure S77: Correct Gene HPRD Annotation

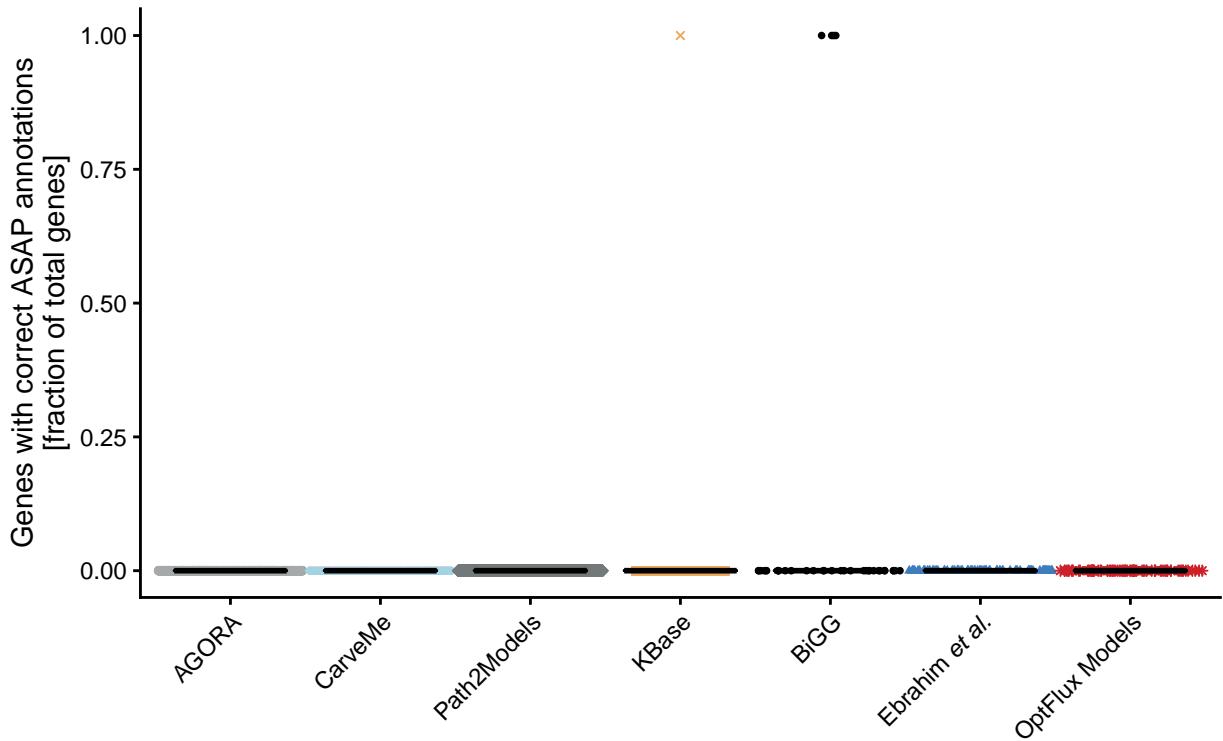


Figure S78: Correct Gene ASAP Annotation

3.3.5 Annotation - SBO Terms

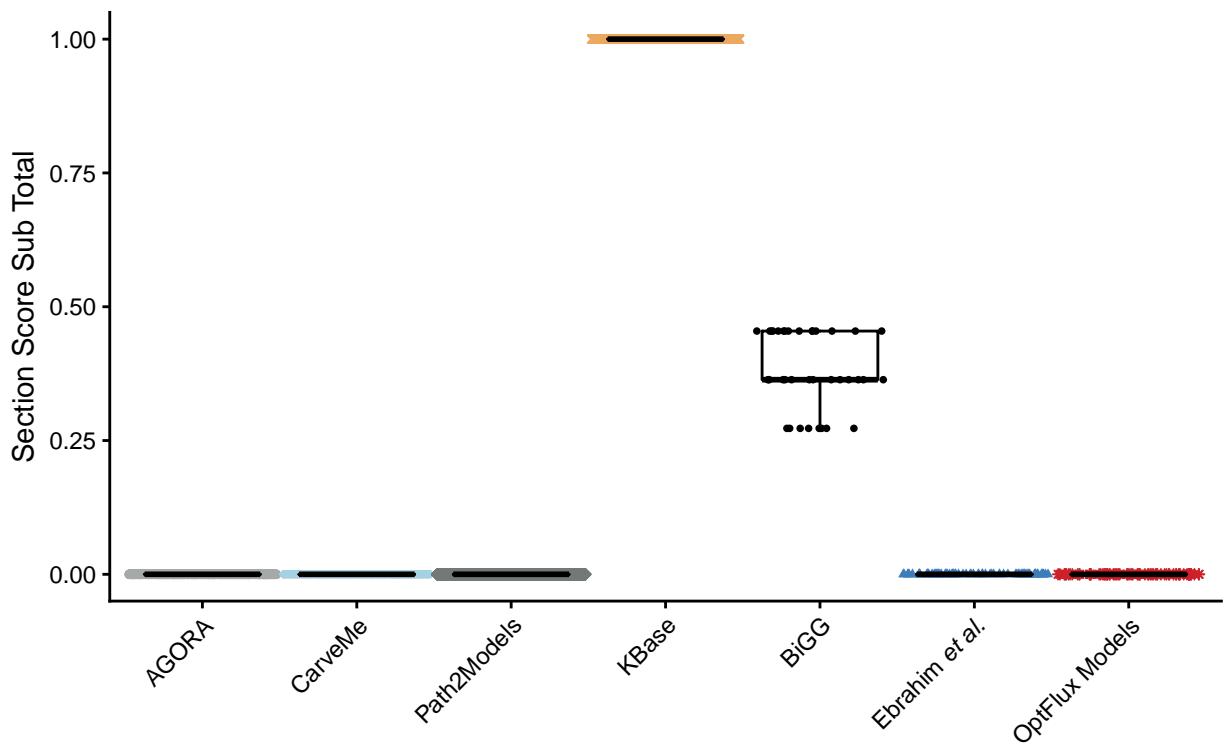


Figure S79: Annotation - SBO Terms. Depicted are the sums of all test scores in this section, applying the weights of the individual test cases as detailed in the snapshot report.

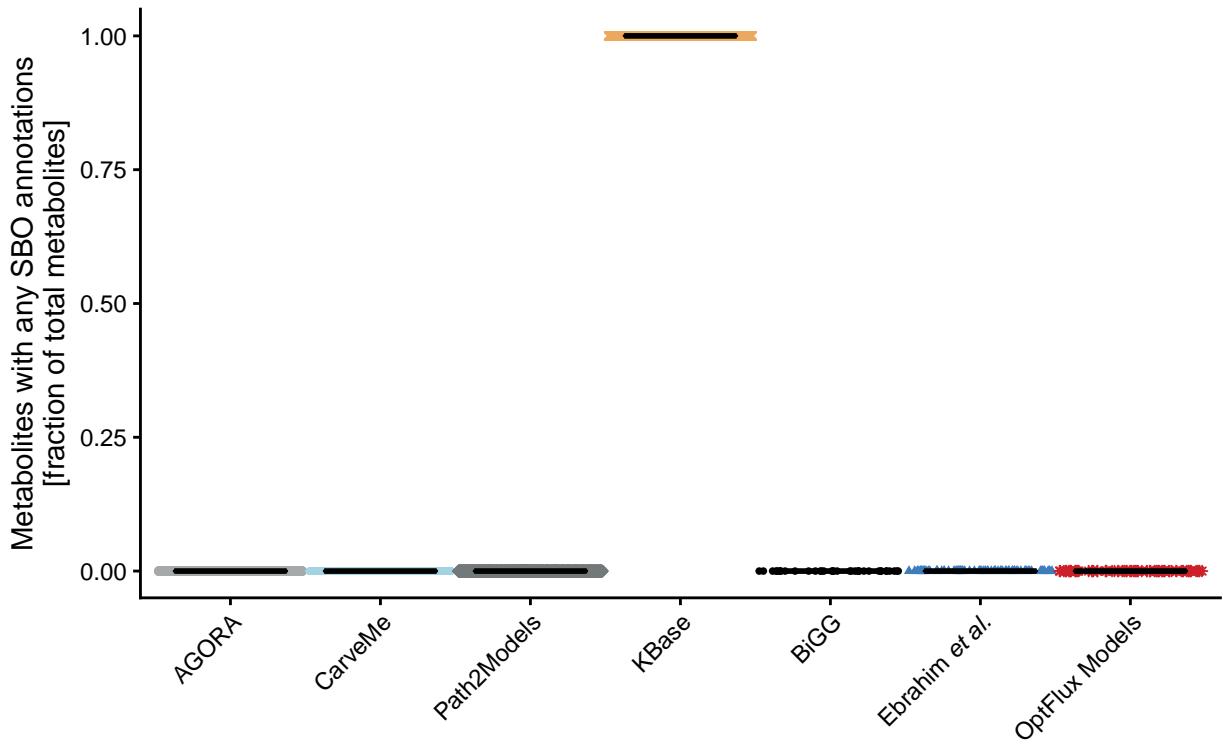


Figure S80: Metabolite General SBO Presence

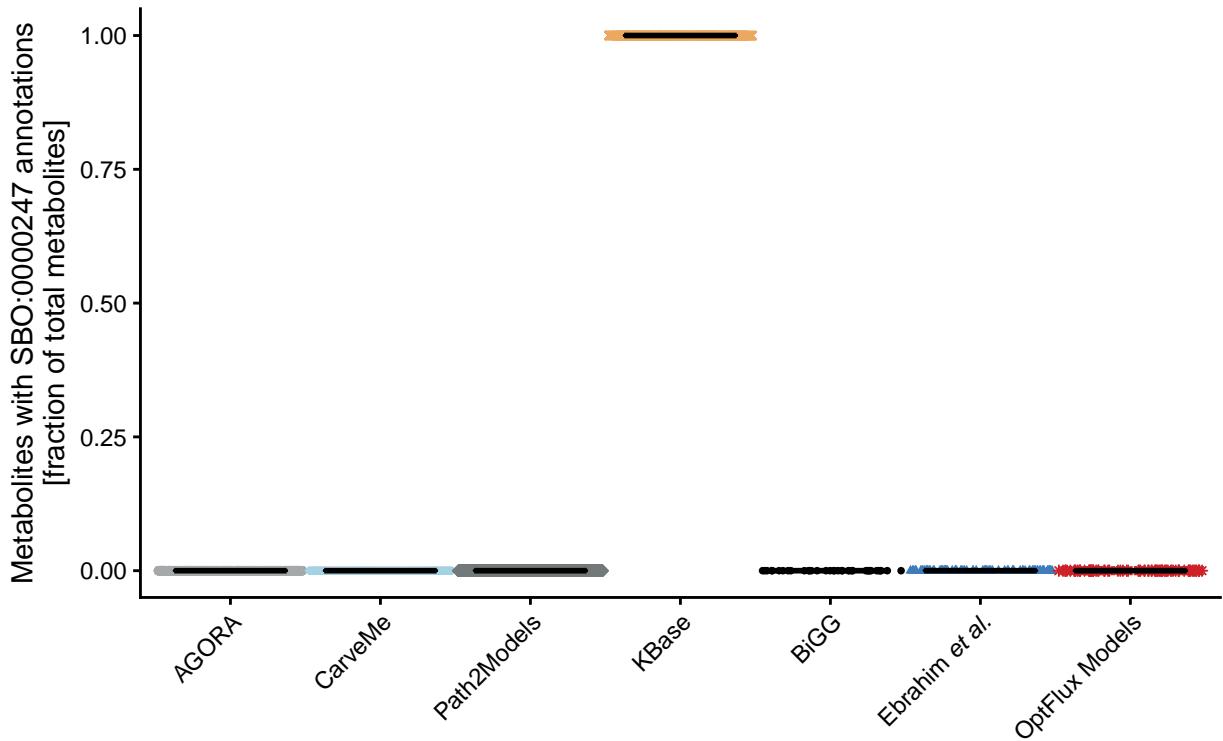


Figure S81: Metabolite SBO:0000247 Presence

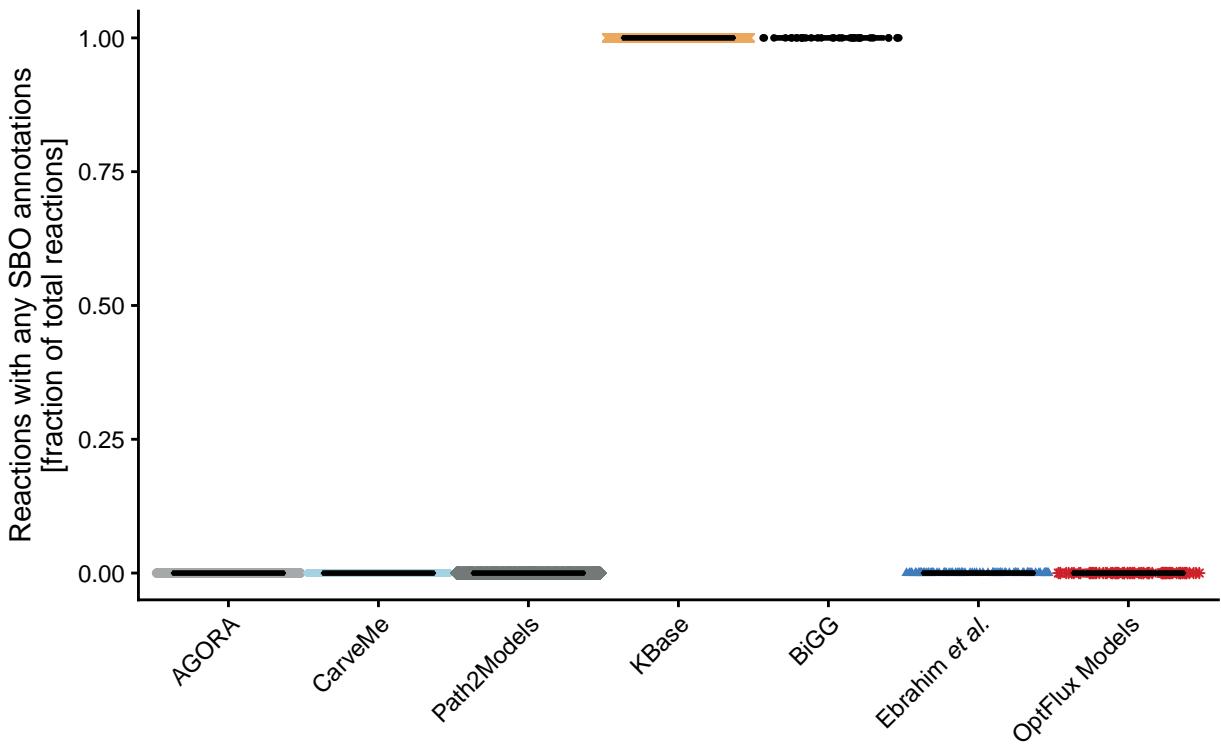


Figure S82: Reaction General SBO Presence

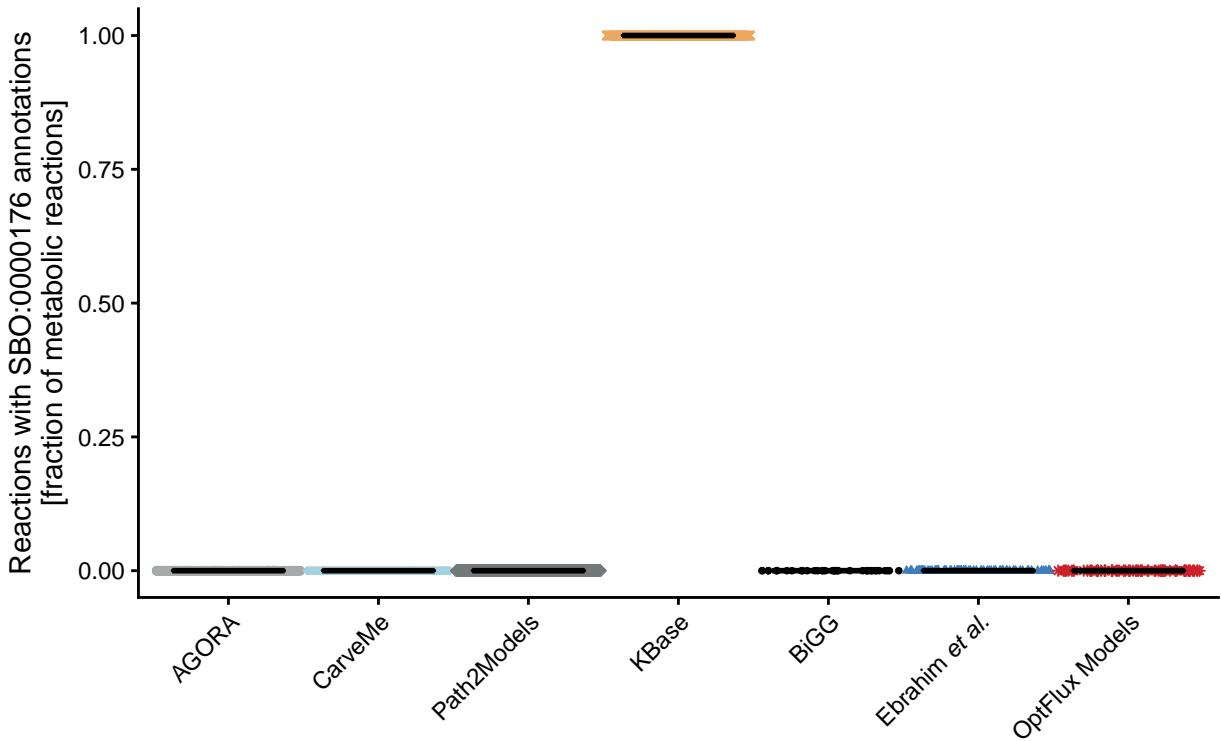


Figure S83: Metabolic Reaction SBO:0000176 Presence

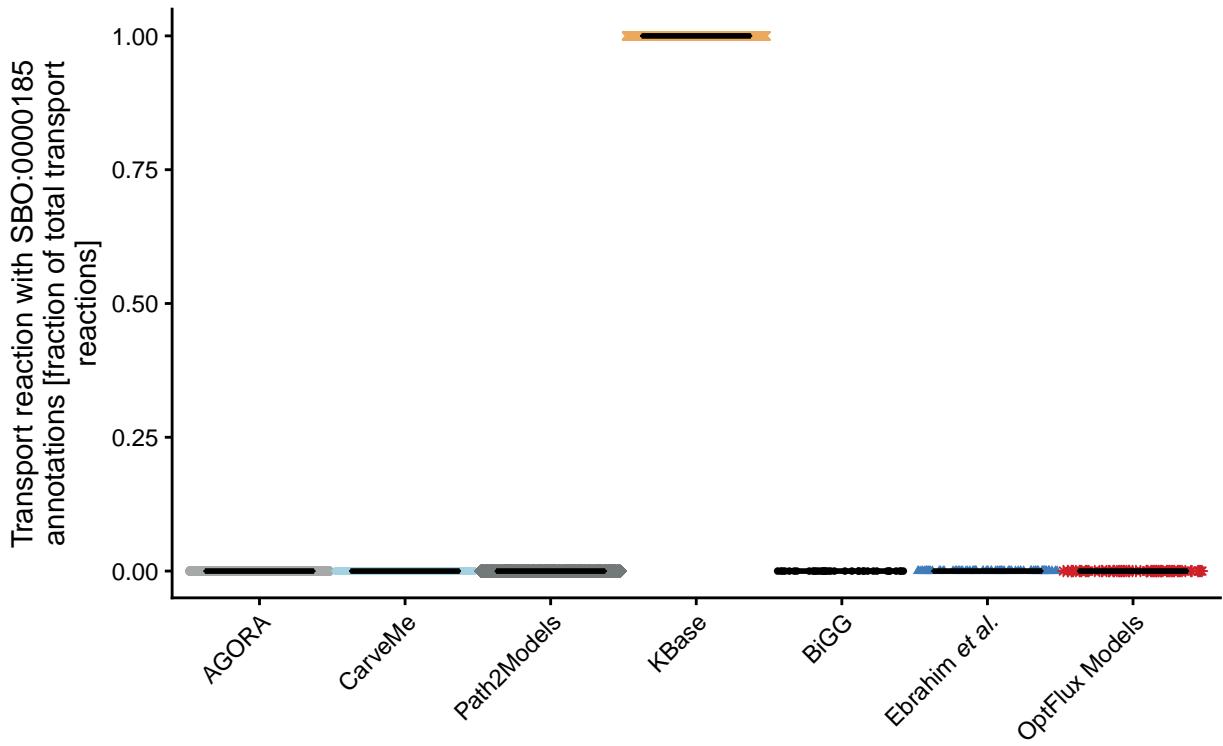


Figure S84: Transport Reaction SBO:0000185 Presence

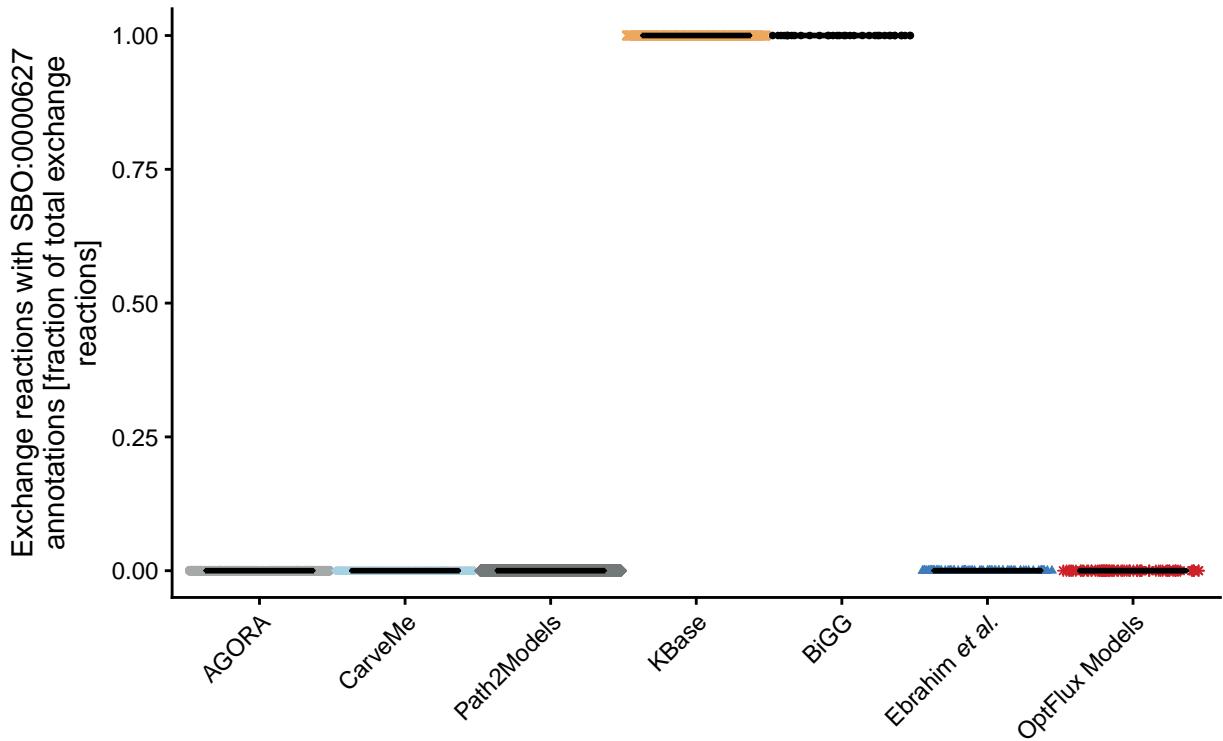


Figure S85: Exchange Reaction SBO:0000627 Presence

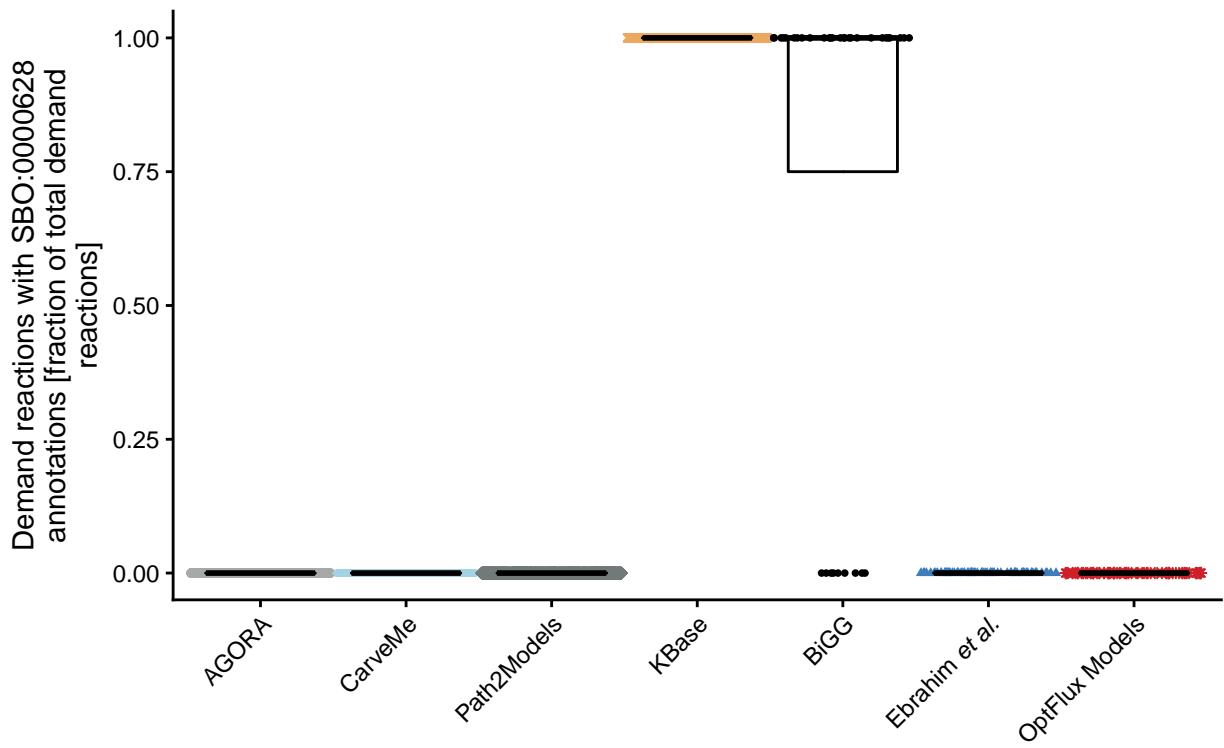


Figure S86: Demand Reaction SBO:0000628 Presence

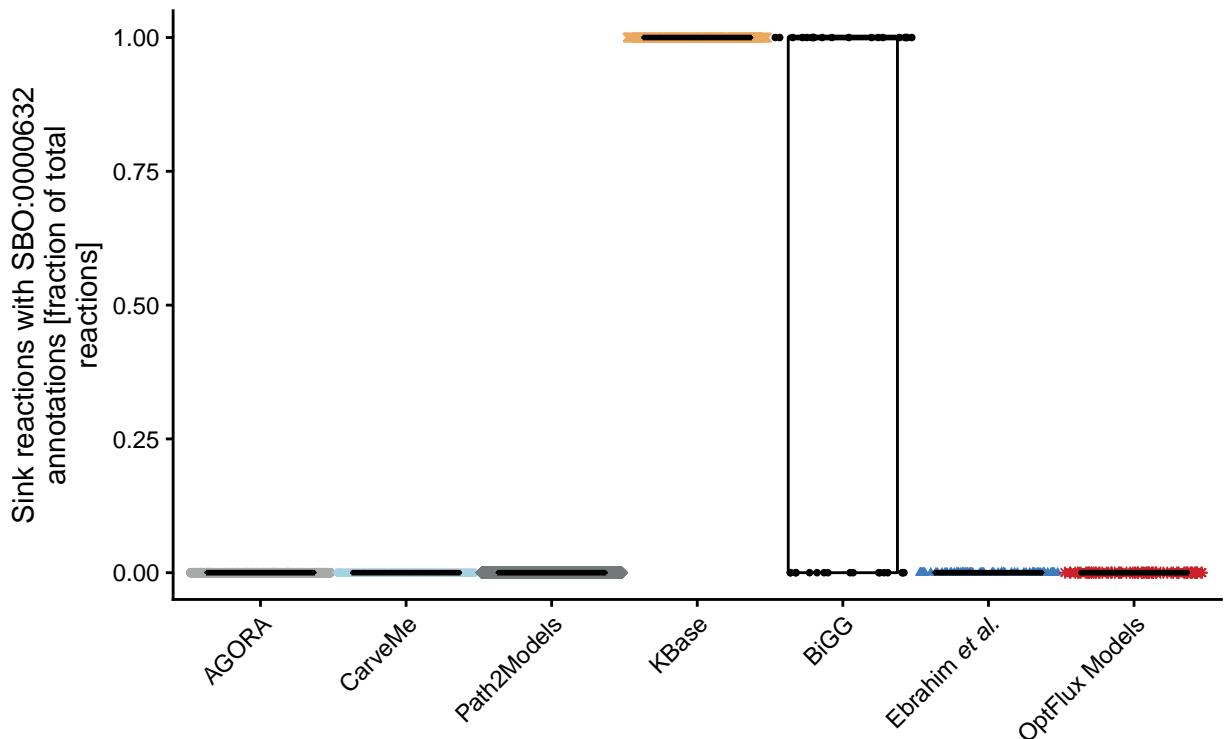


Figure S87: Sink Reaction SBO:0000632 Presence

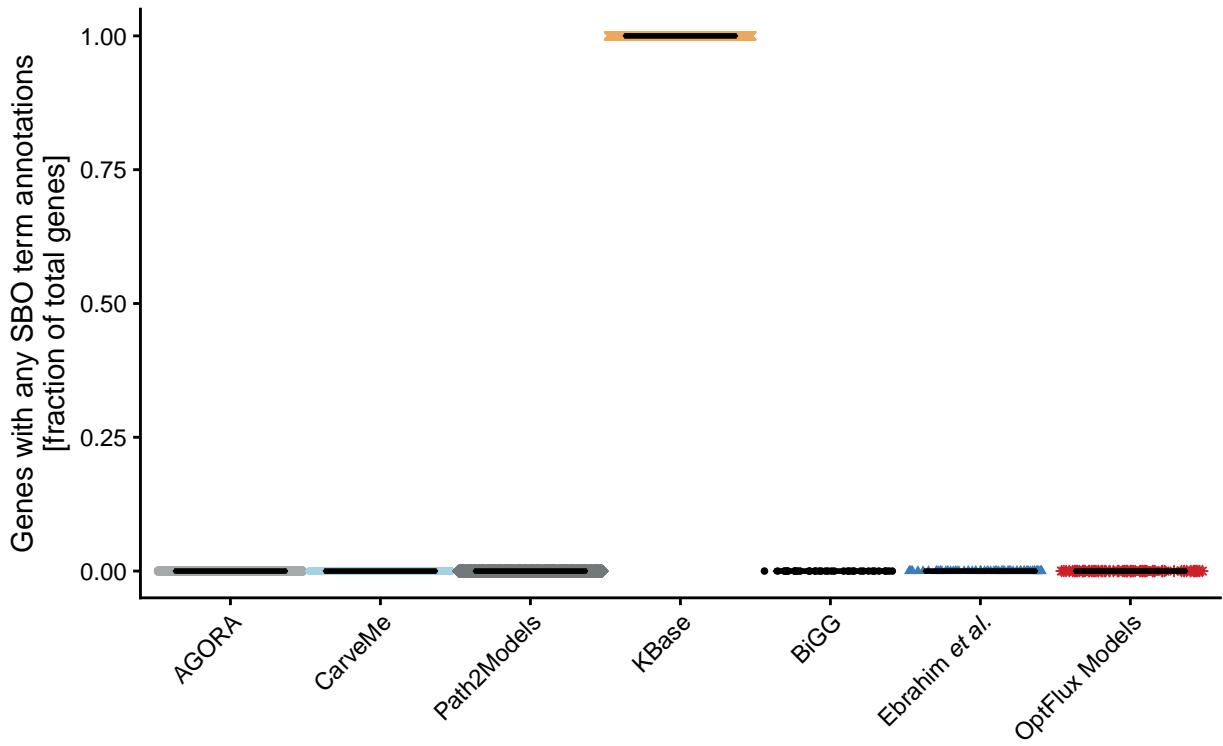


Figure S88: Gene General SBO Presence

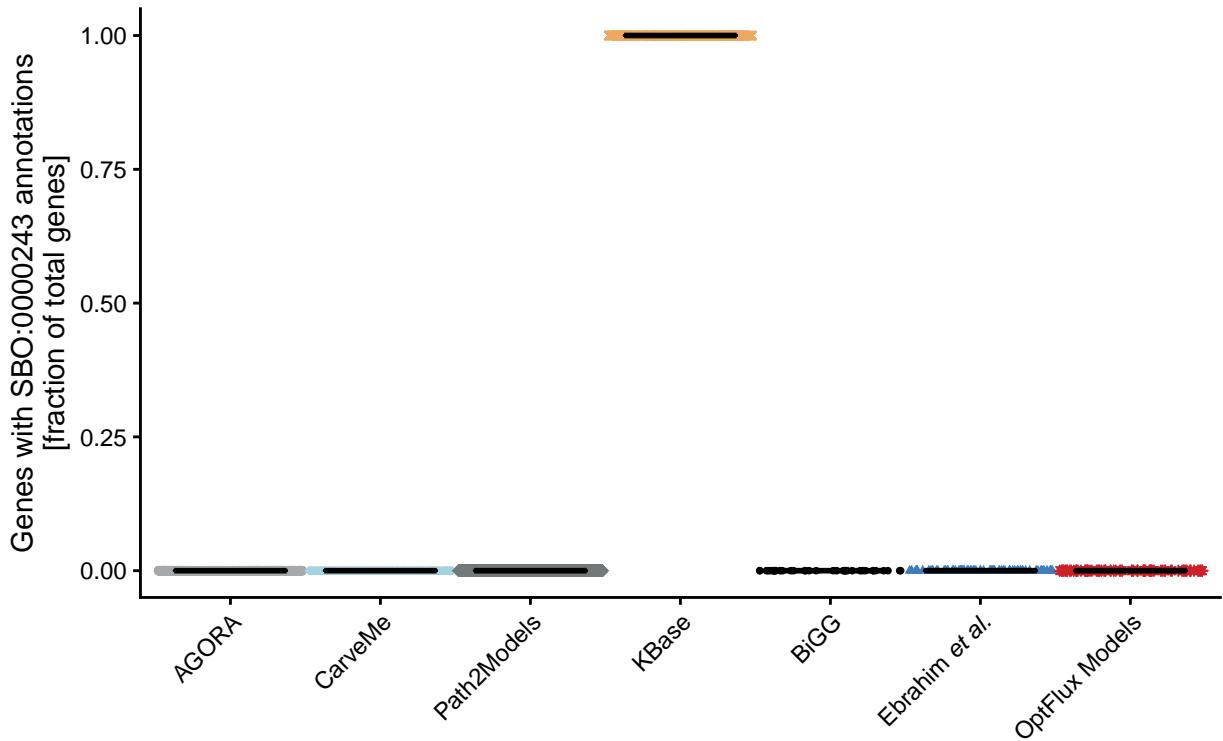


Figure S89: Gene SBO:0000243 Presence

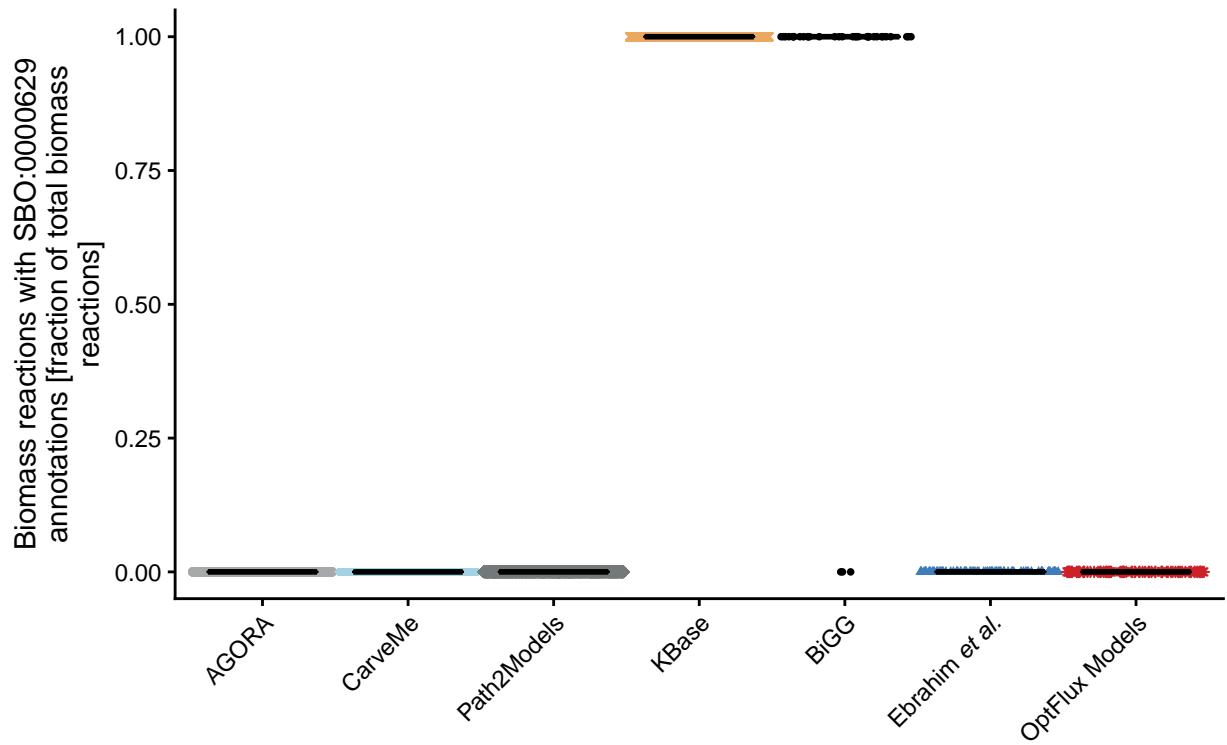


Figure S90: Biomass Reaction SBO:0000629 Presence

3.4 Specific Section

3.4.1 SBML

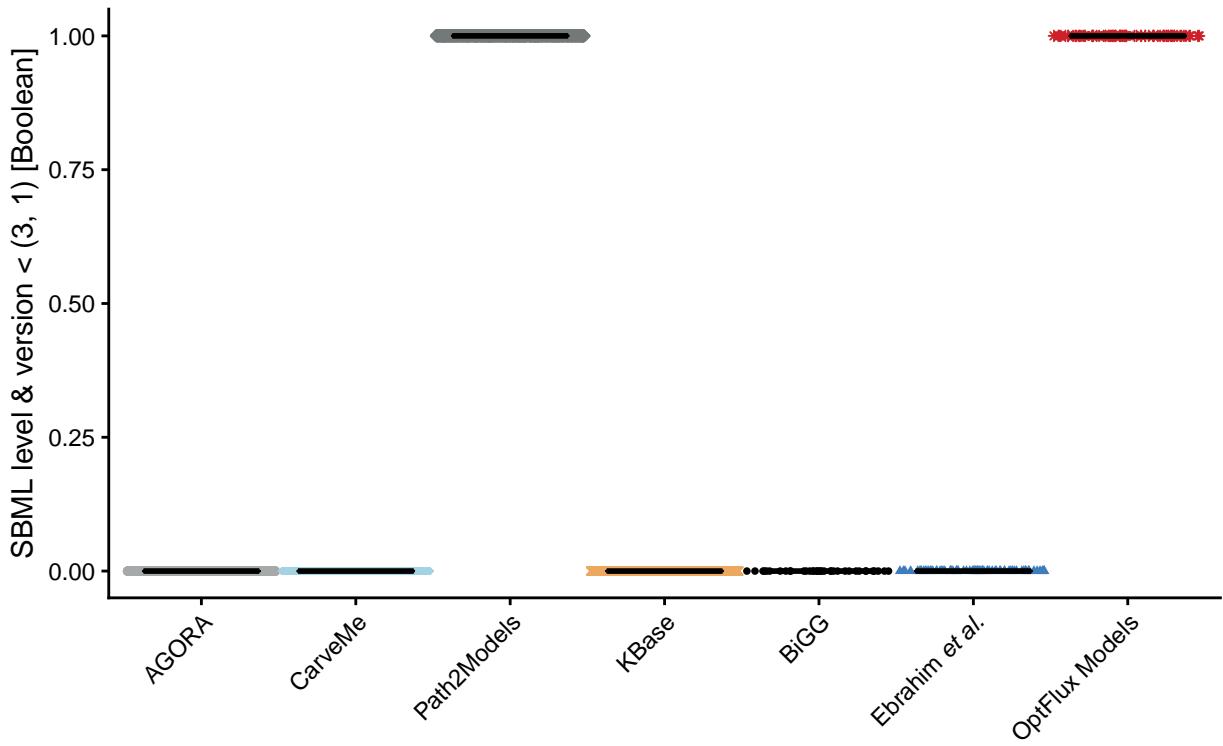


Figure S91: SBML Level and Version

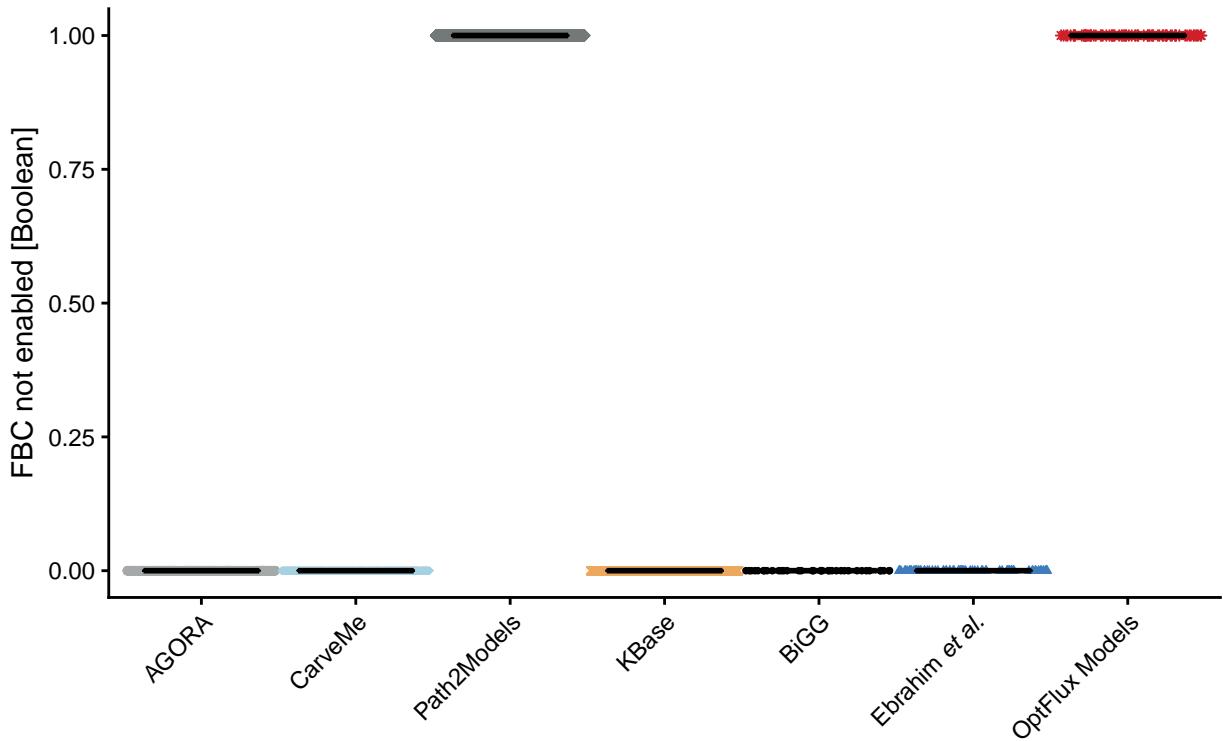


Figure S92: FBC not Enabled

3.4.2 Basic Information

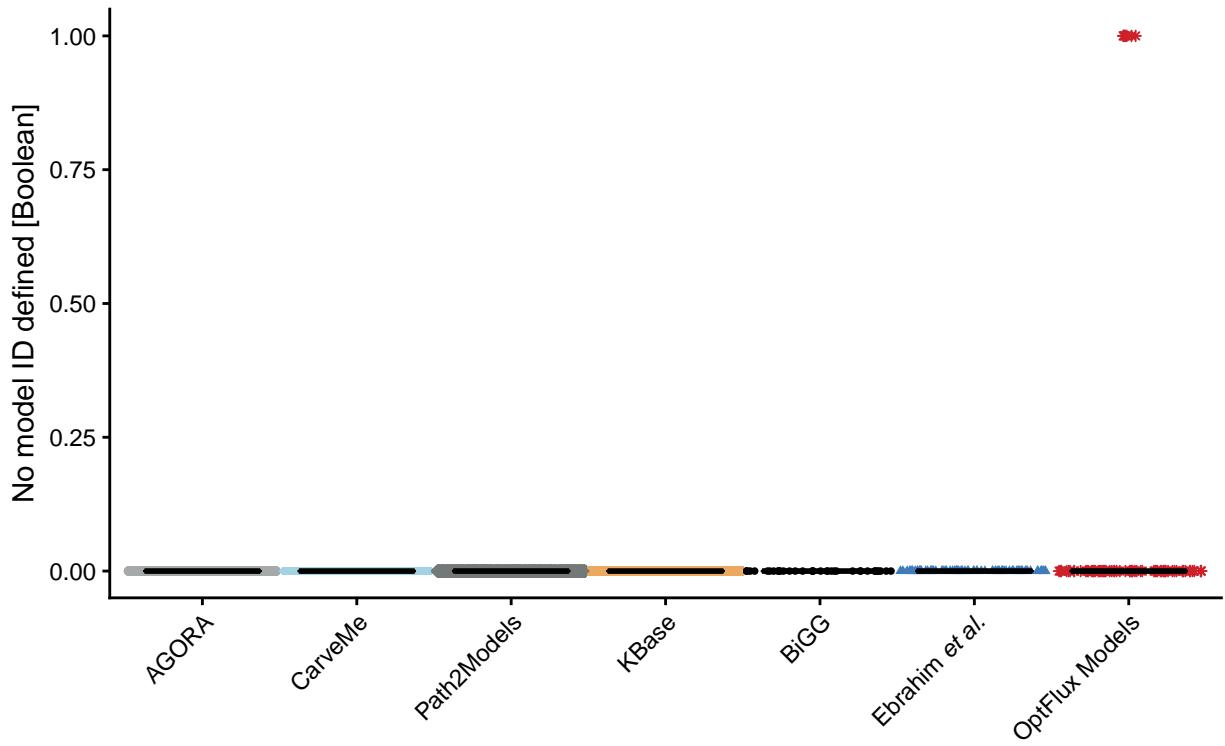


Figure S93: Model Identifier Presence

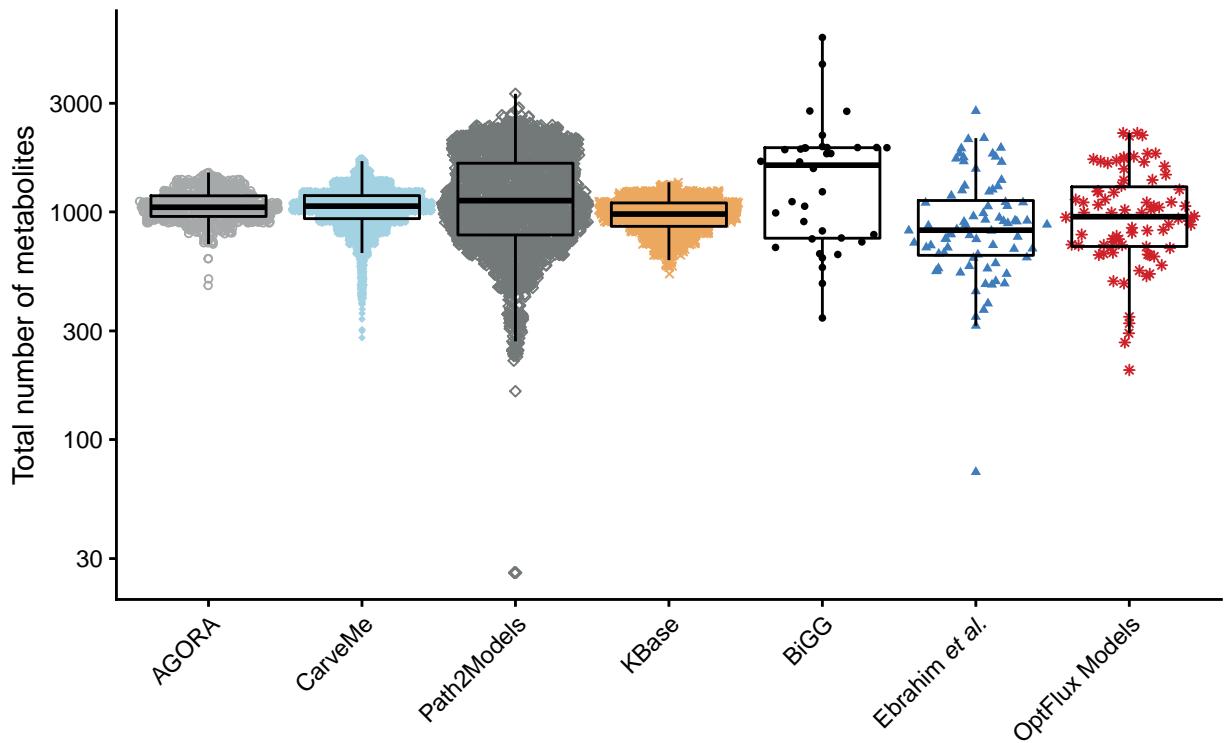


Figure S94: Number of Metabolites

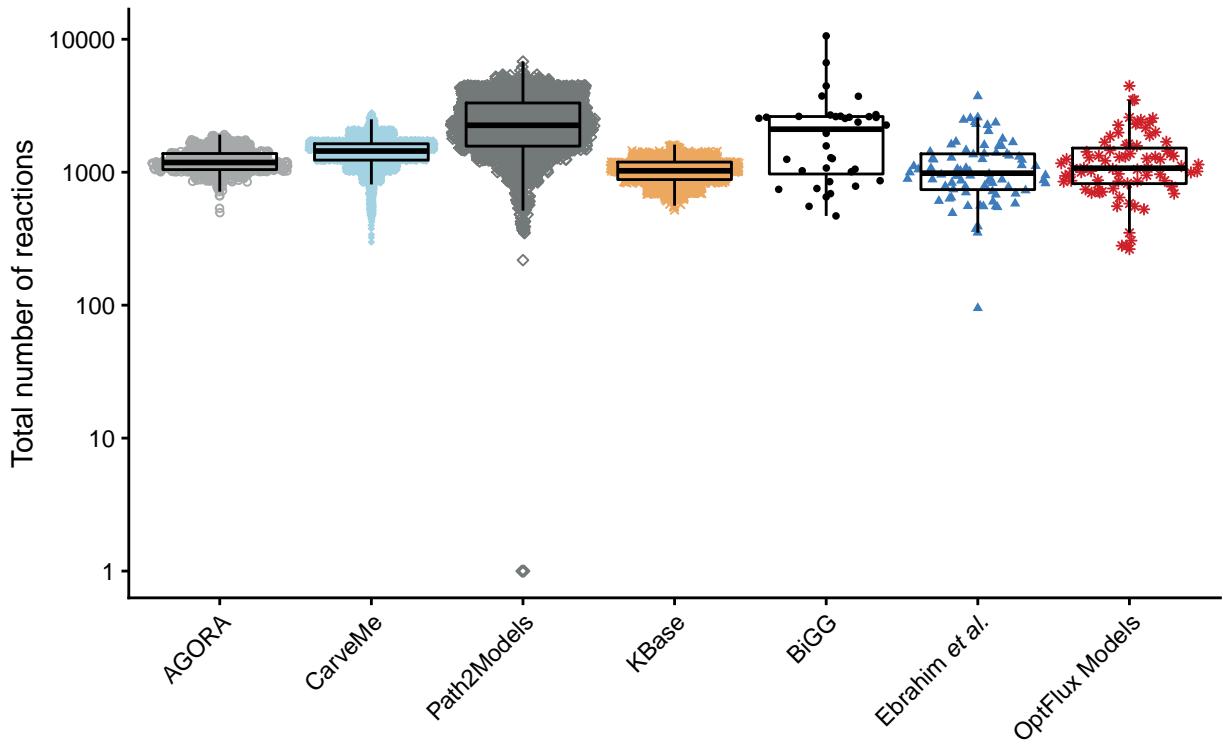


Figure S95: Number of Reactions

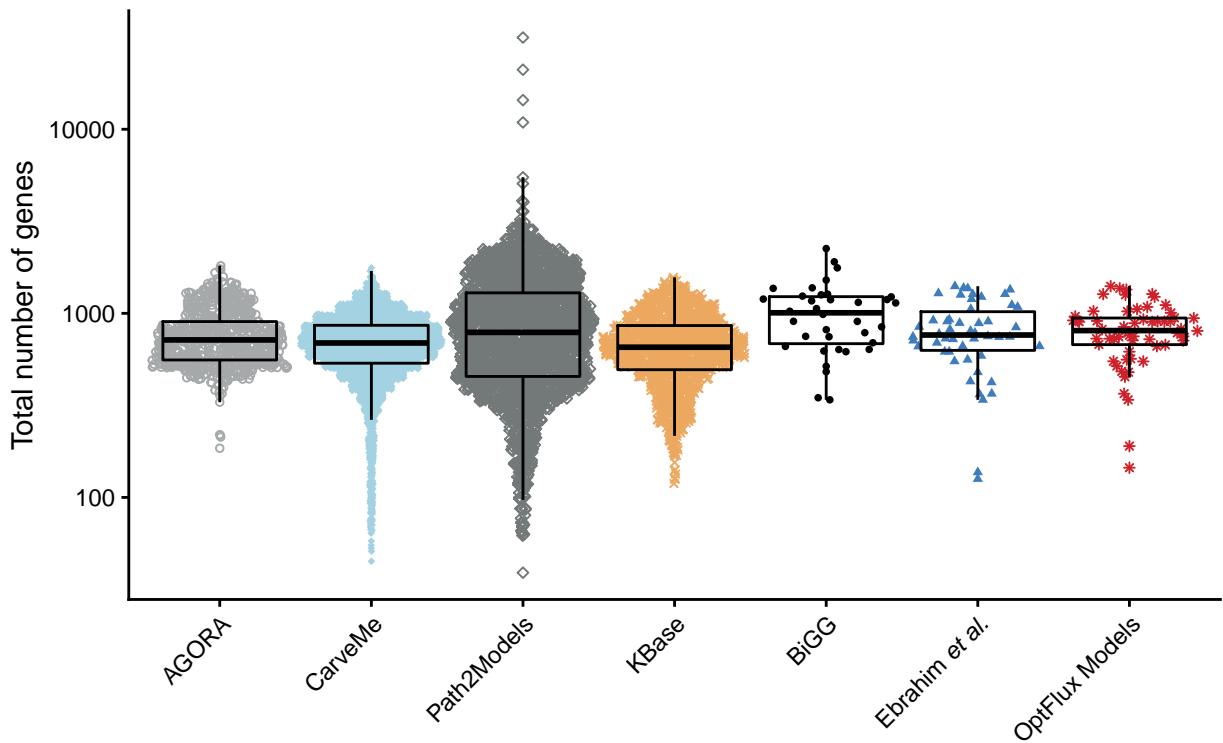


Figure S96: Number of Genes

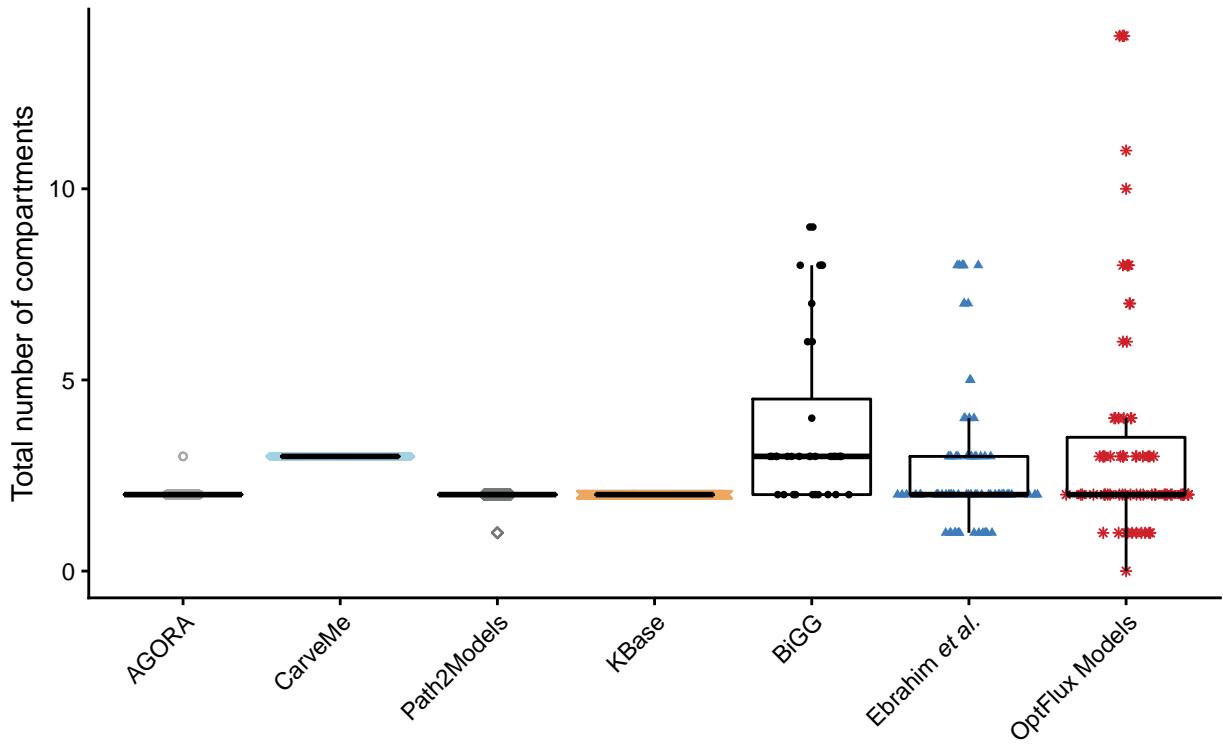


Figure S97: Number of Compartments

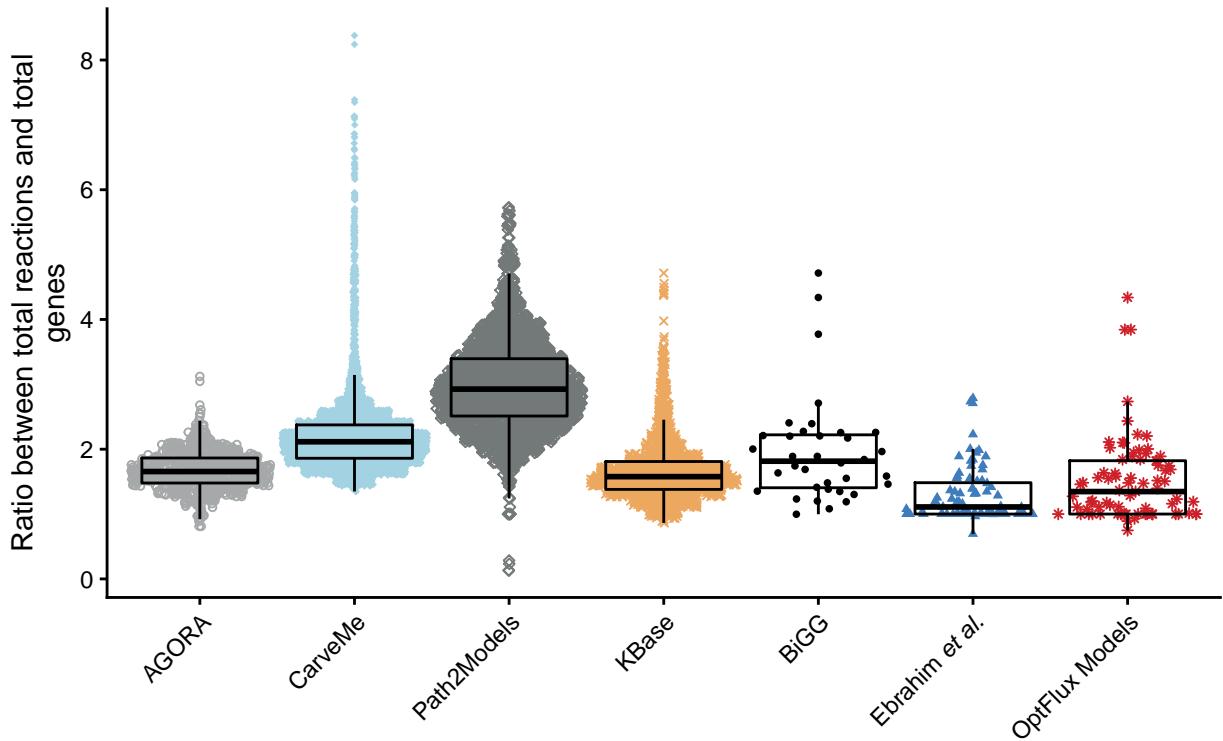


Figure S98: Metabolic Coverage

3.4.3 Metabolite Information

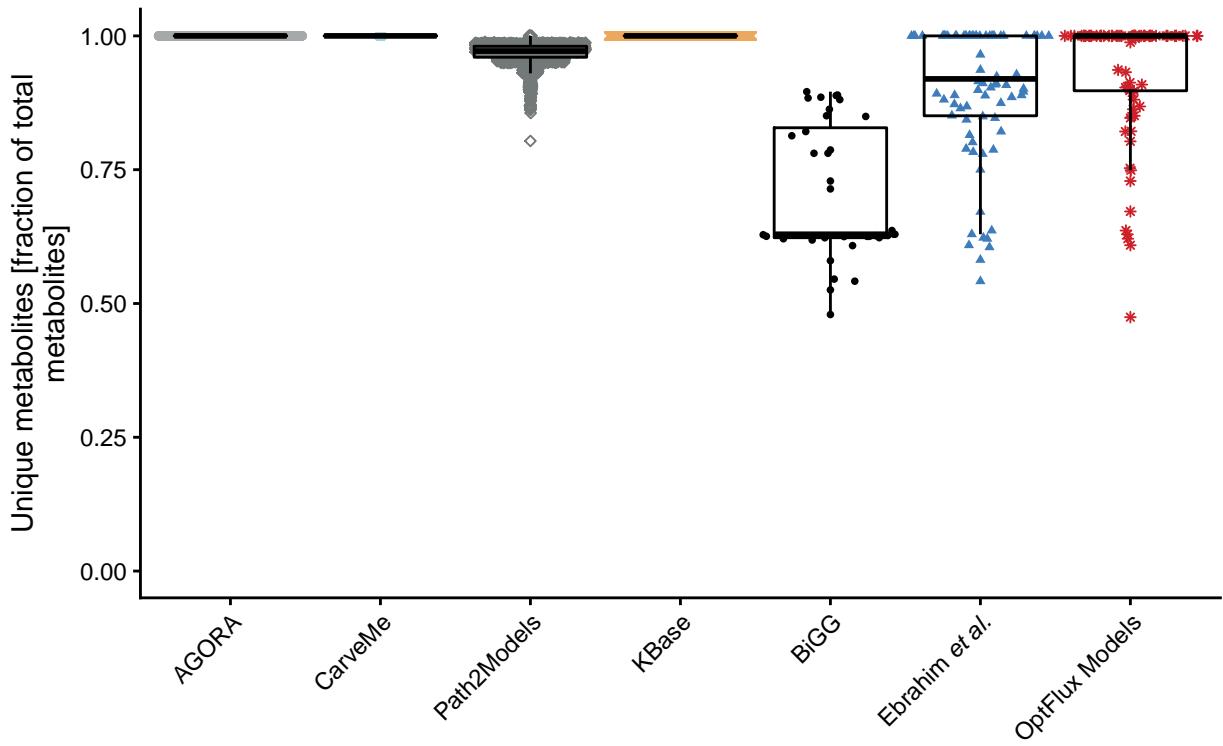


Figure S99: Unique Metabolites

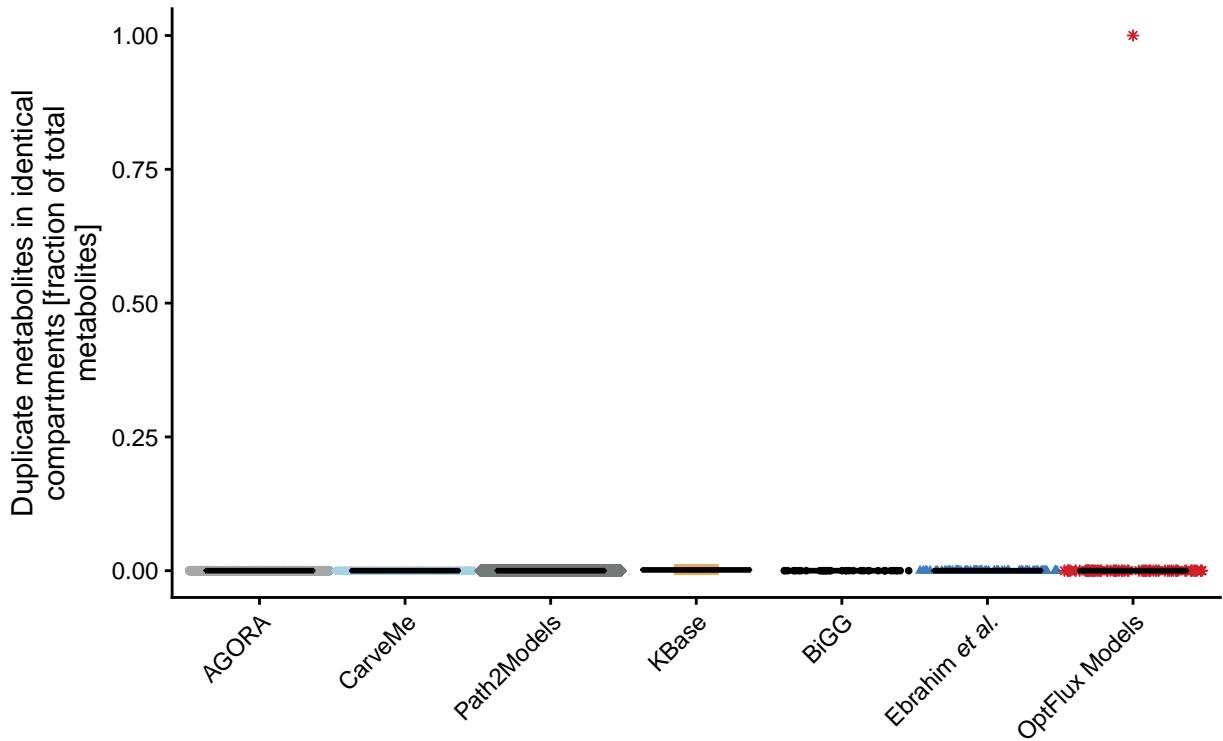


Figure S100: Duplicate Metabolites in Identical Compartments

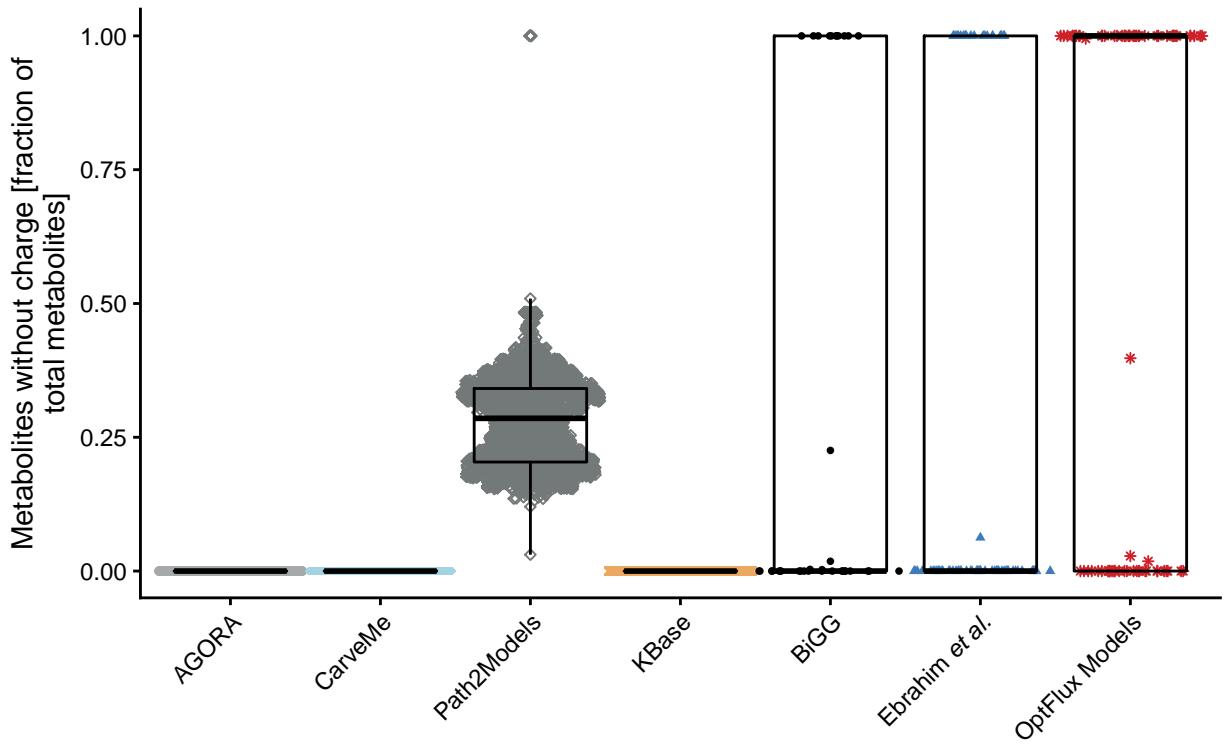


Figure S101: Metabolites Without Charge

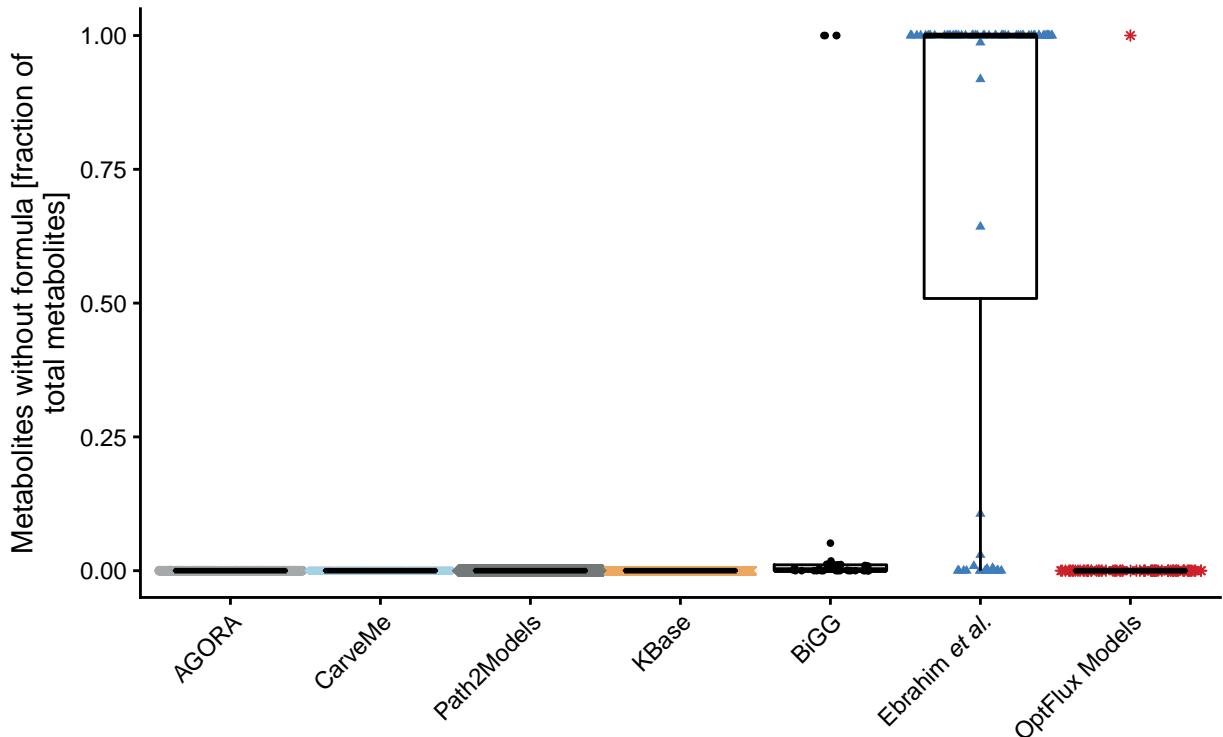


Figure S102: Metabolites Without Formula

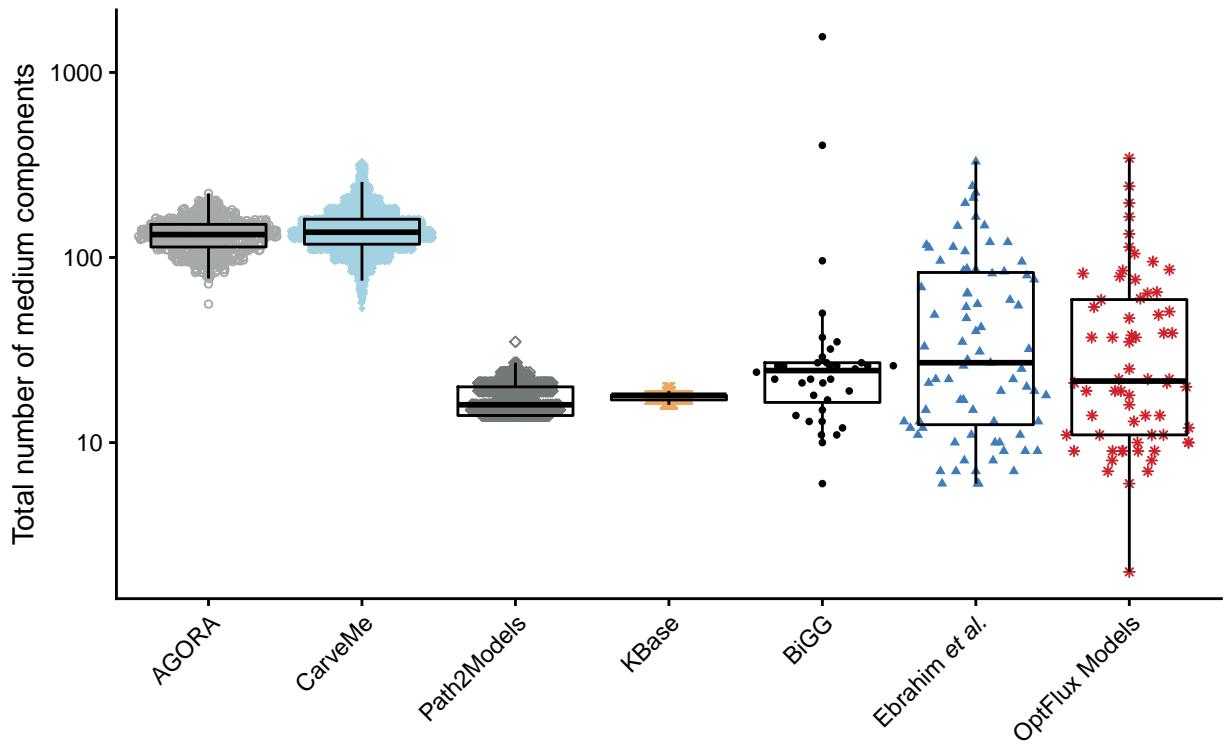


Figure S103: Number of Medium Components

3.4.4 Reaction Information

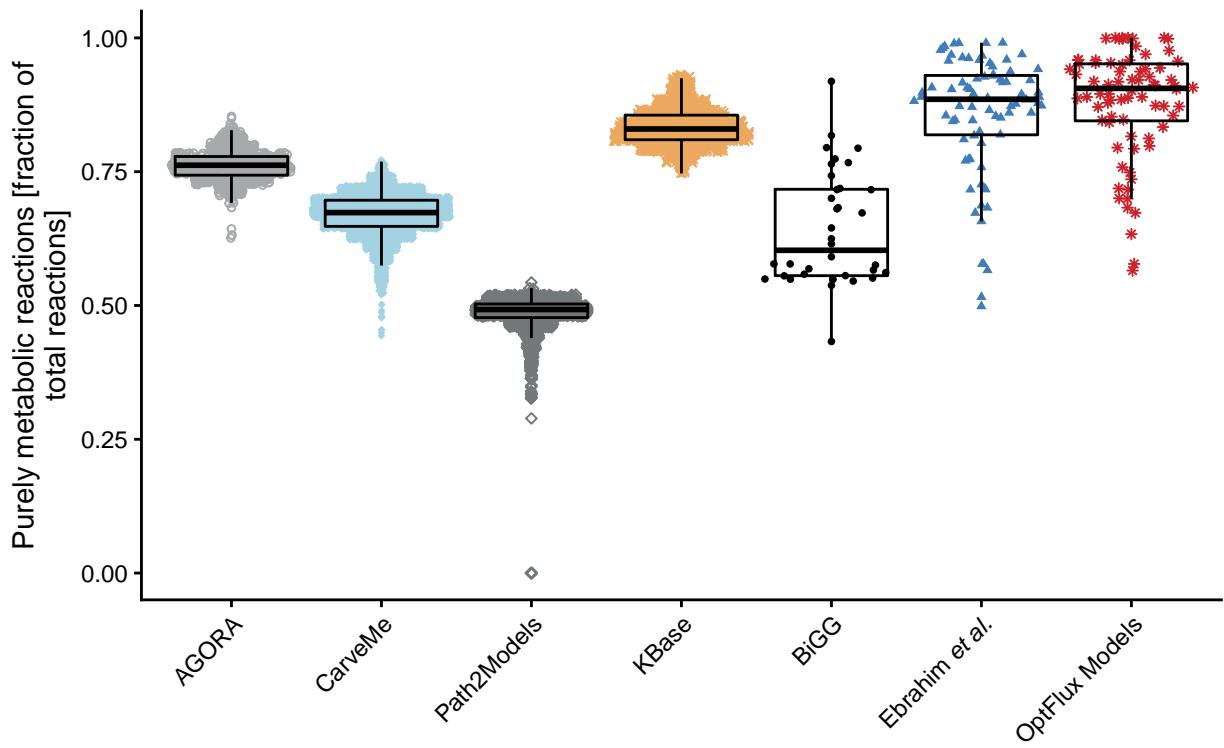


Figure S104: Purely Metabolic Reactions

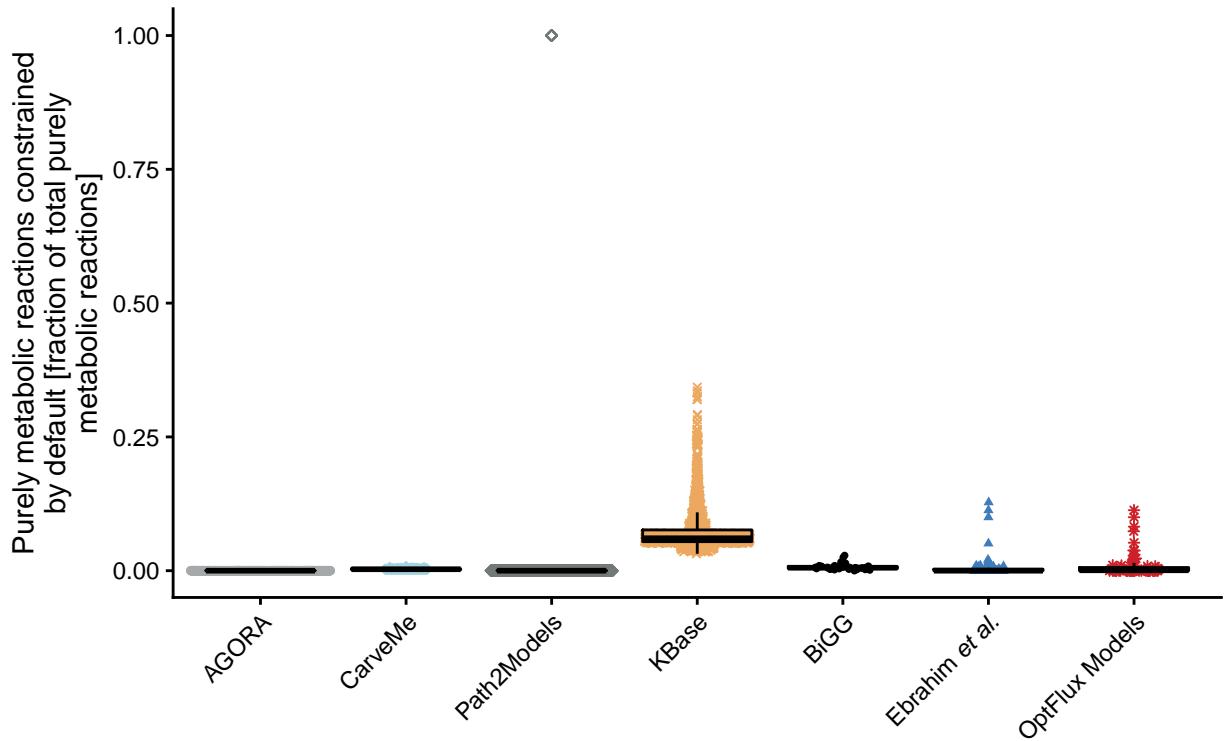


Figure S105: Purely Metabolic Reactions with Constraints

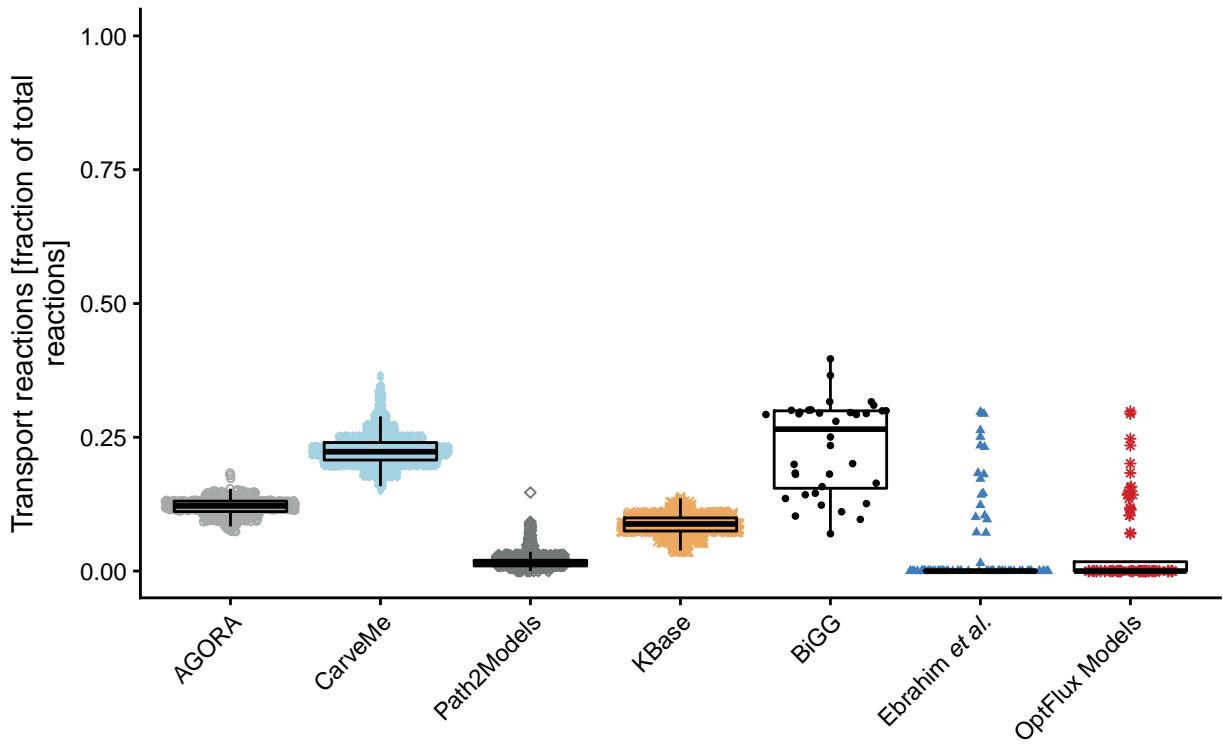


Figure S106: Transport Reactions

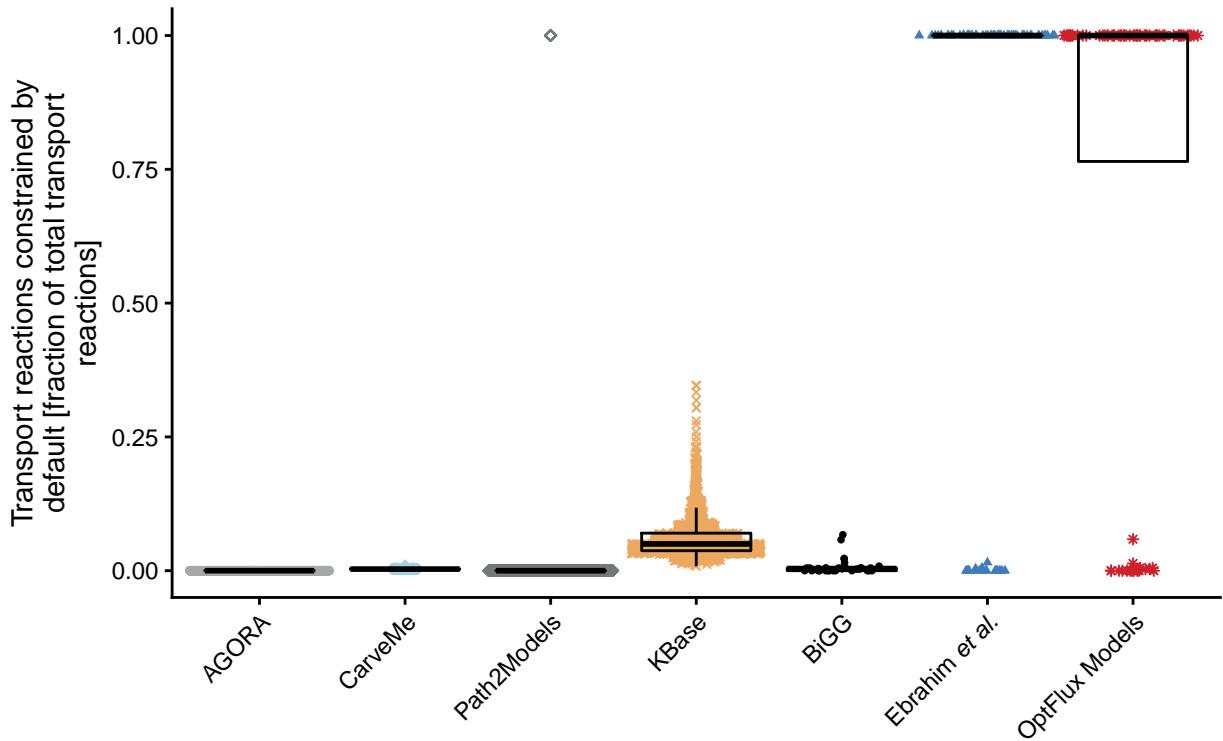


Figure S107: Transport Reactions with Constraints

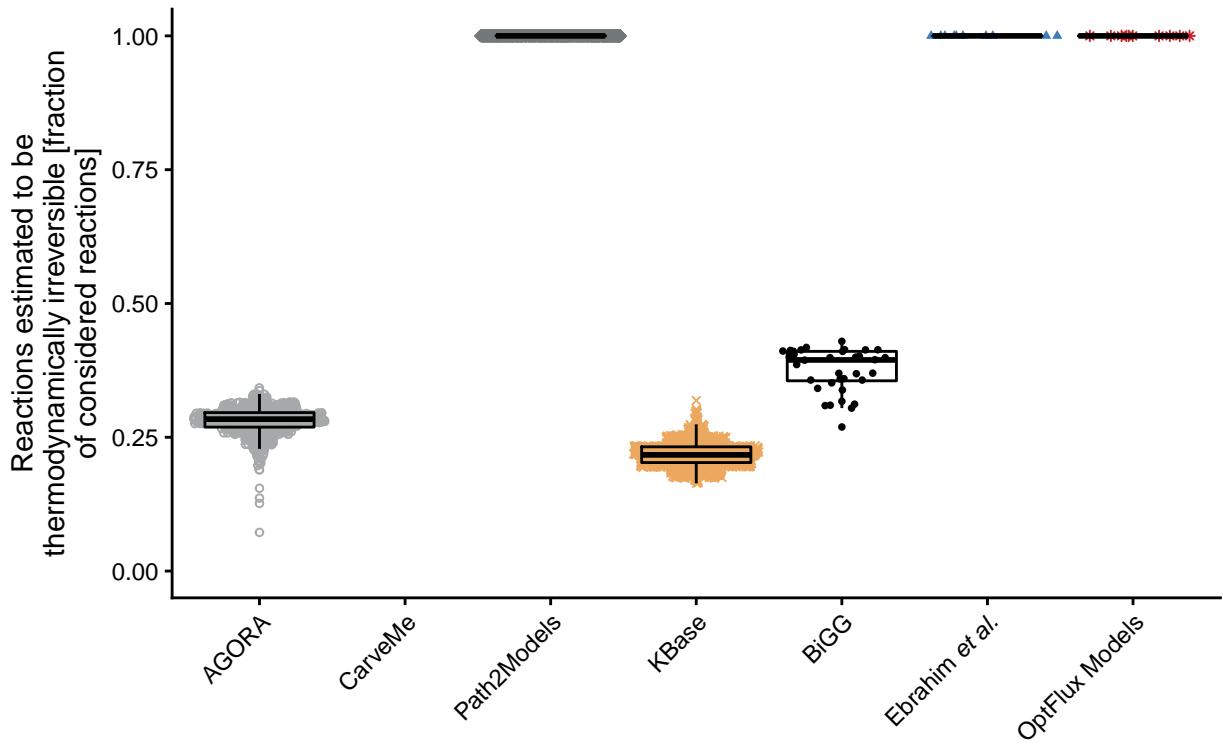


Figure S108: Thermodynamic Reversibility of Purely Metabolic Reactions

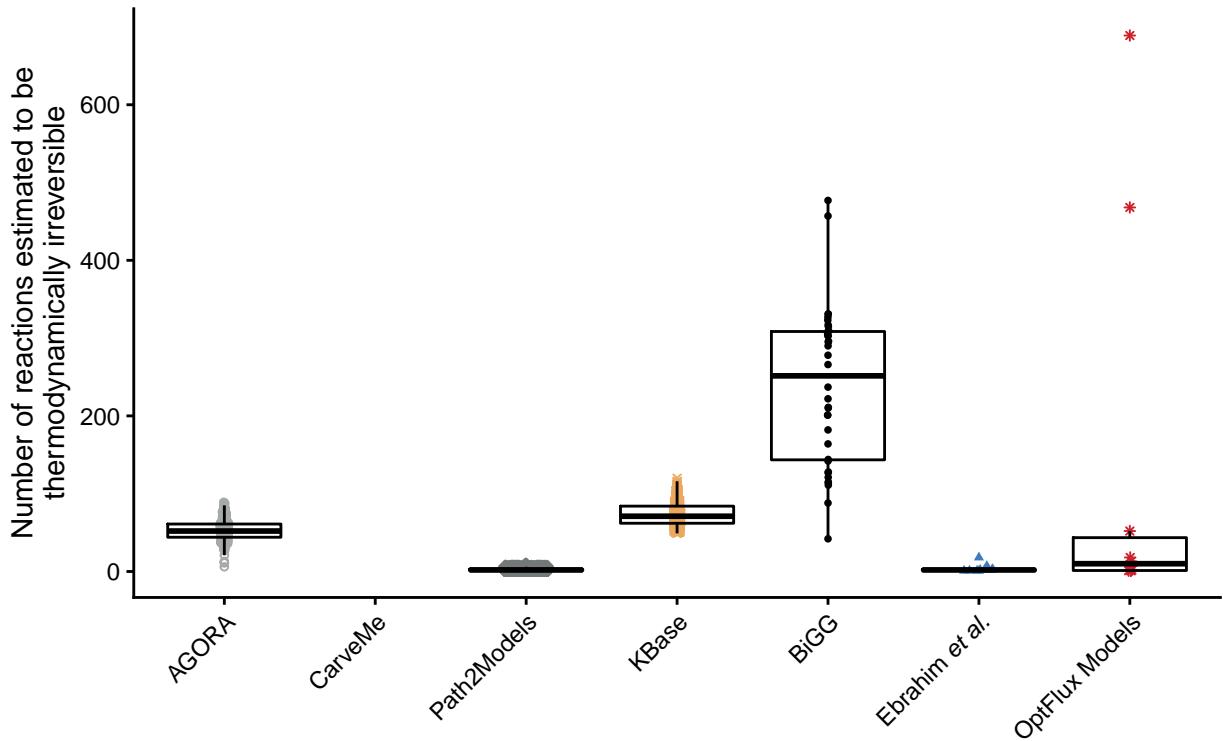


Figure S109: Thermodynamic Reversibility of Purely Metabolic Reactions

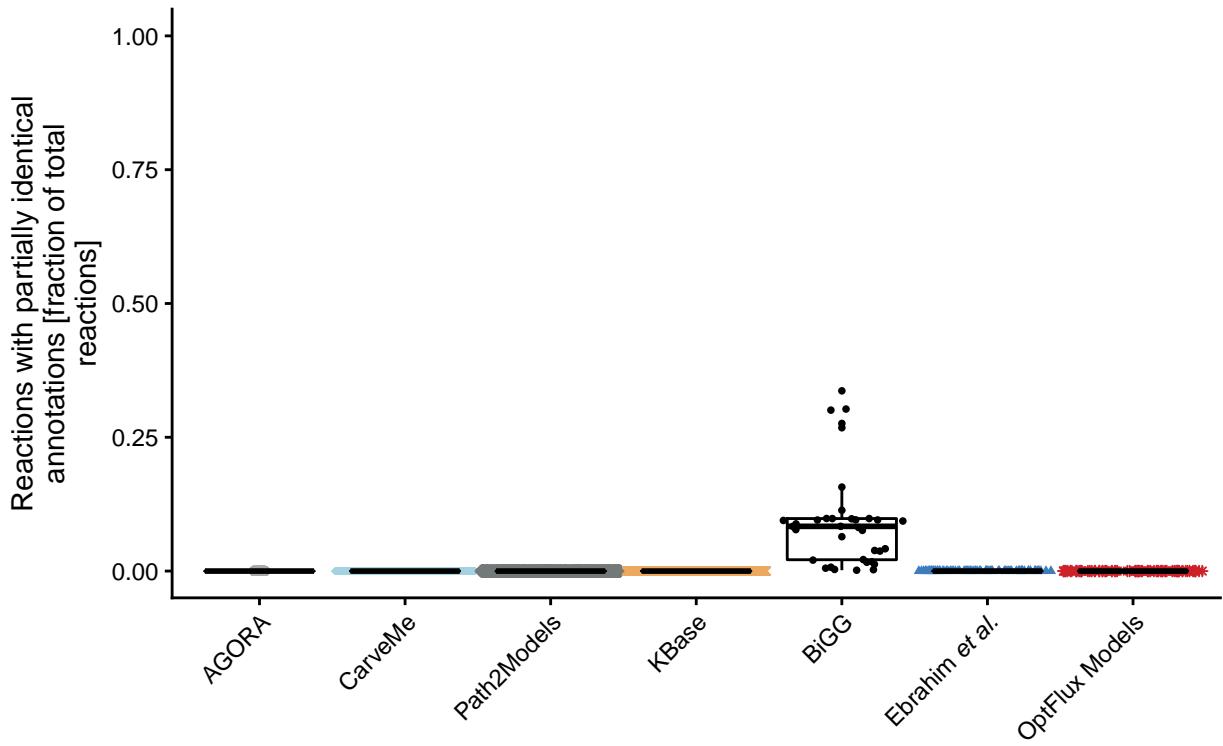


Figure S110: Reactions with Partially Identical Annotations

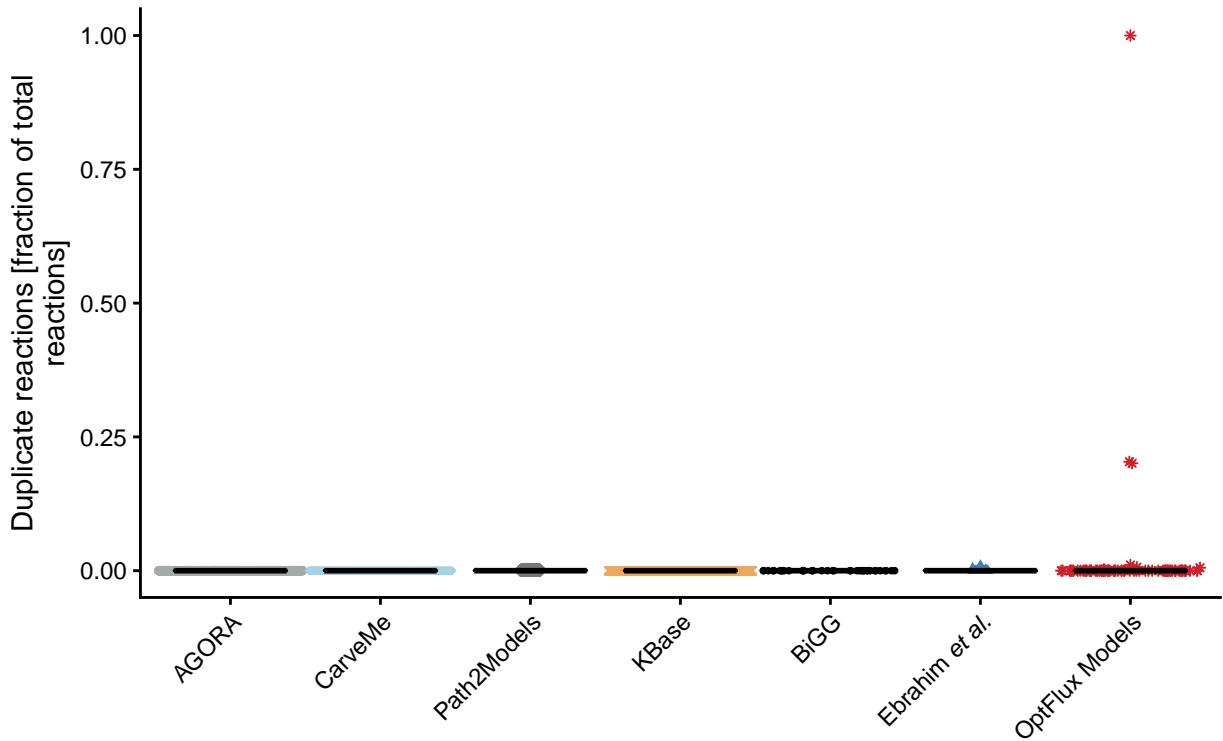


Figure S111: Duplicate Reactions

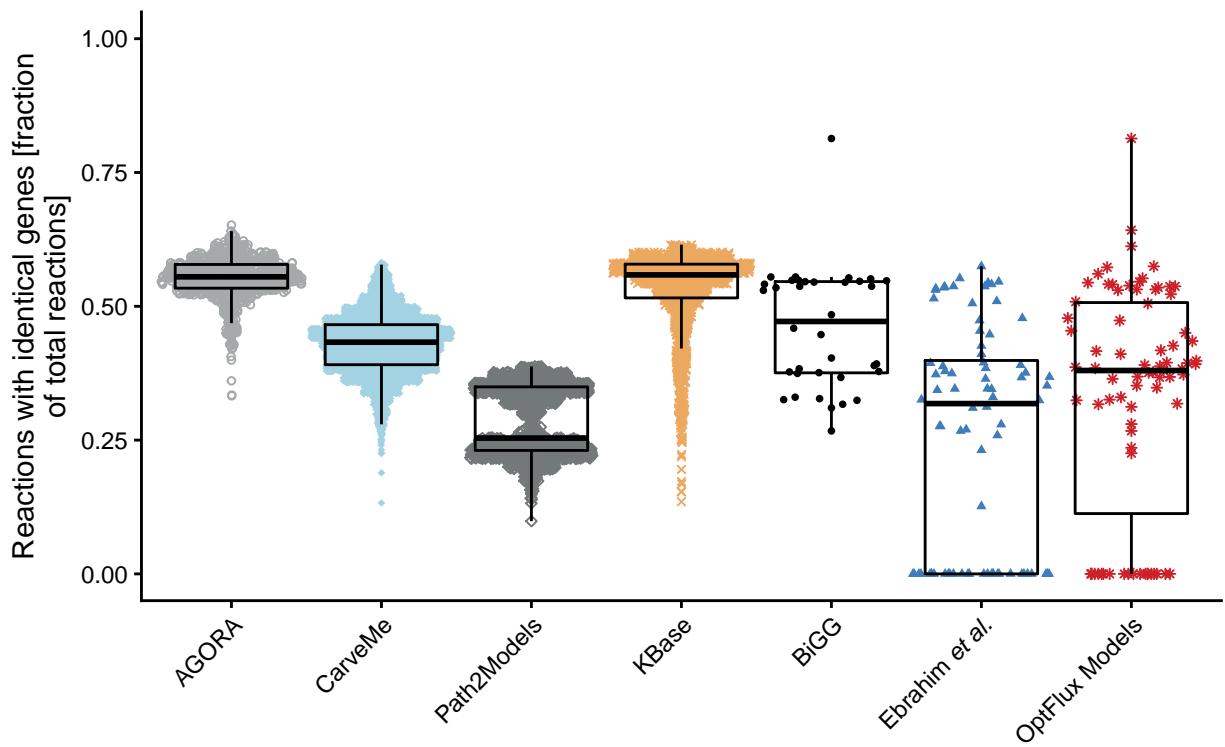


Figure S112: Reactions with Identical Genes

3.4.5 Gene-Protein-Reaction (GPR) Association

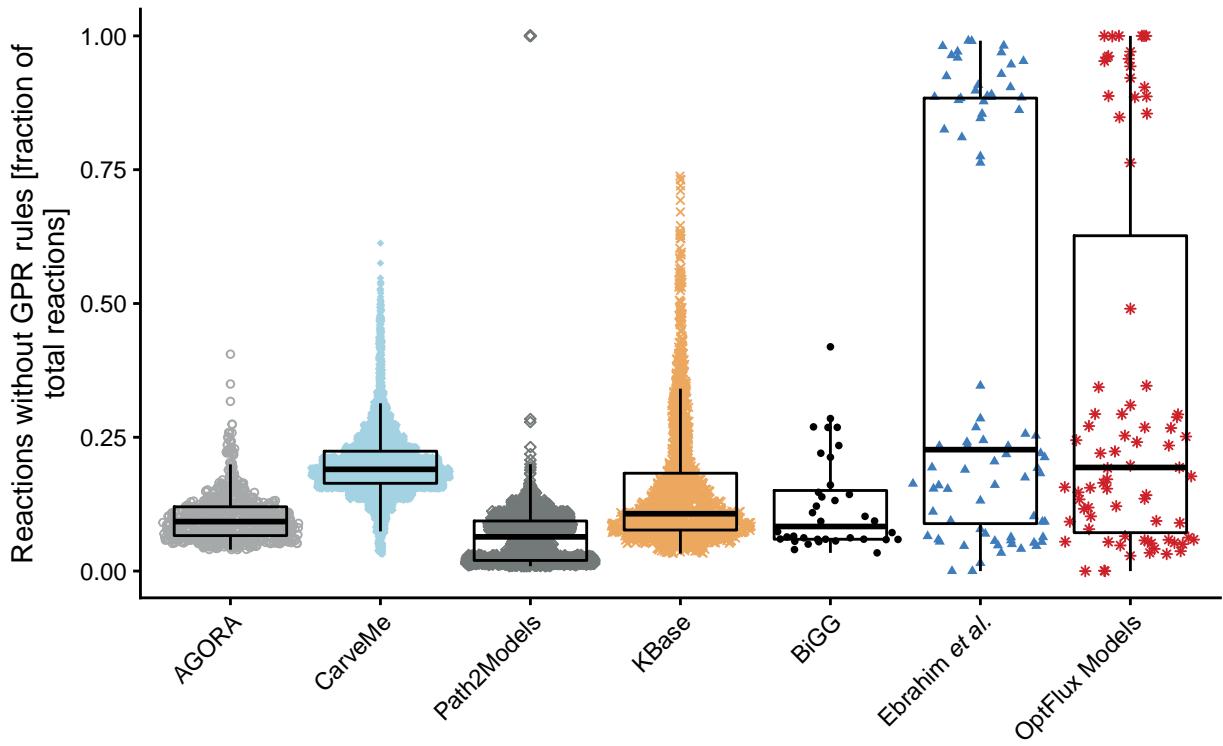


Figure S113: Reactions without GPR

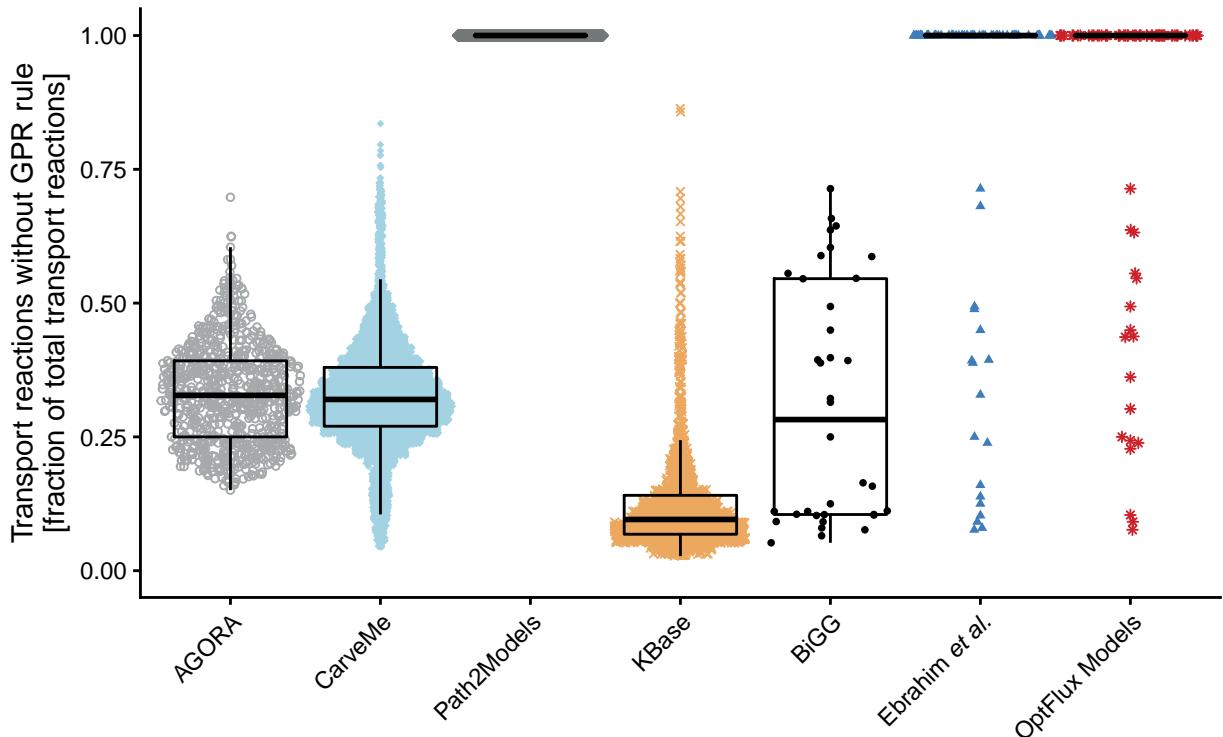


Figure S114: Fraction of Transport Reactions without GPR

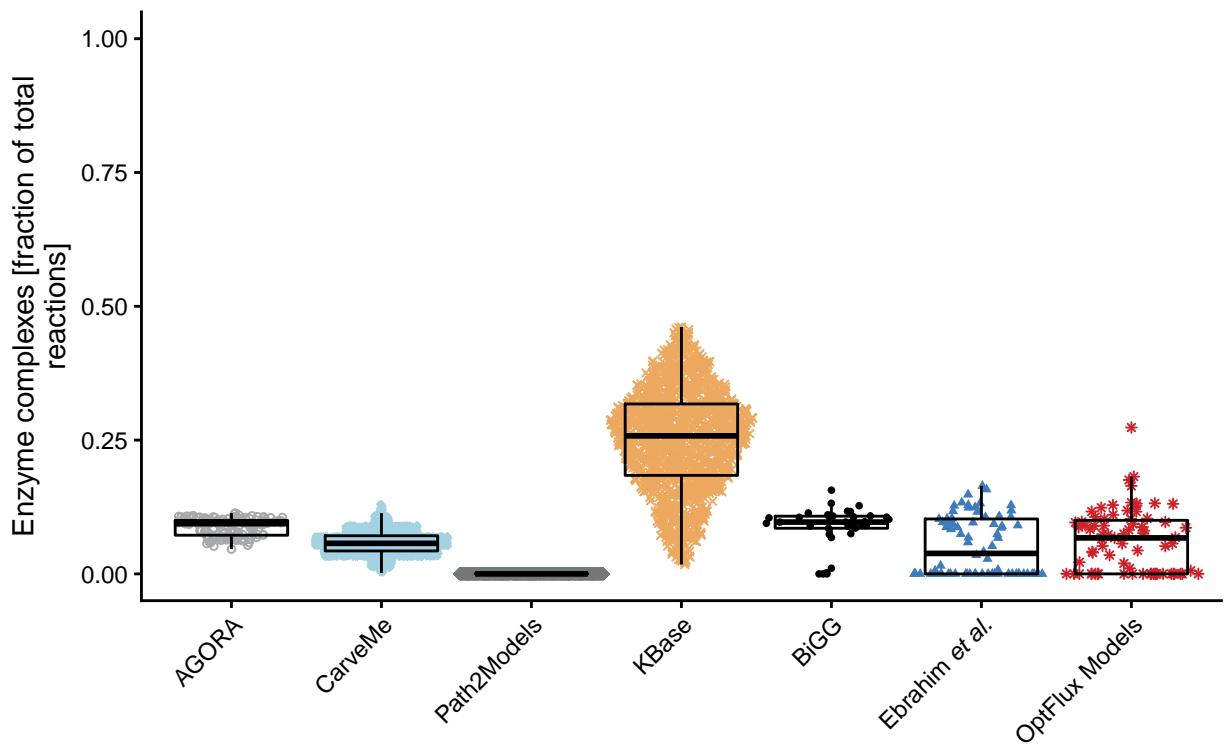


Figure S115: Enzyme Complexes

3.4.6 Biomass

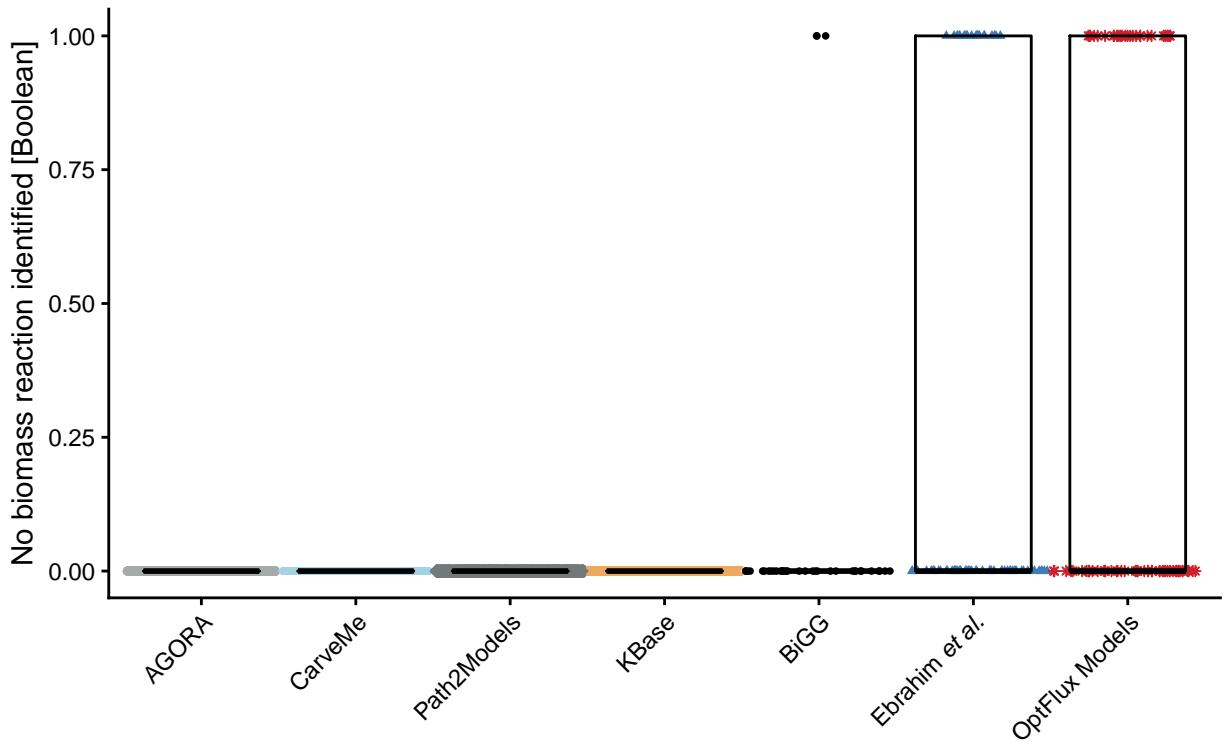


Figure S116: Biomass Reactions Identified

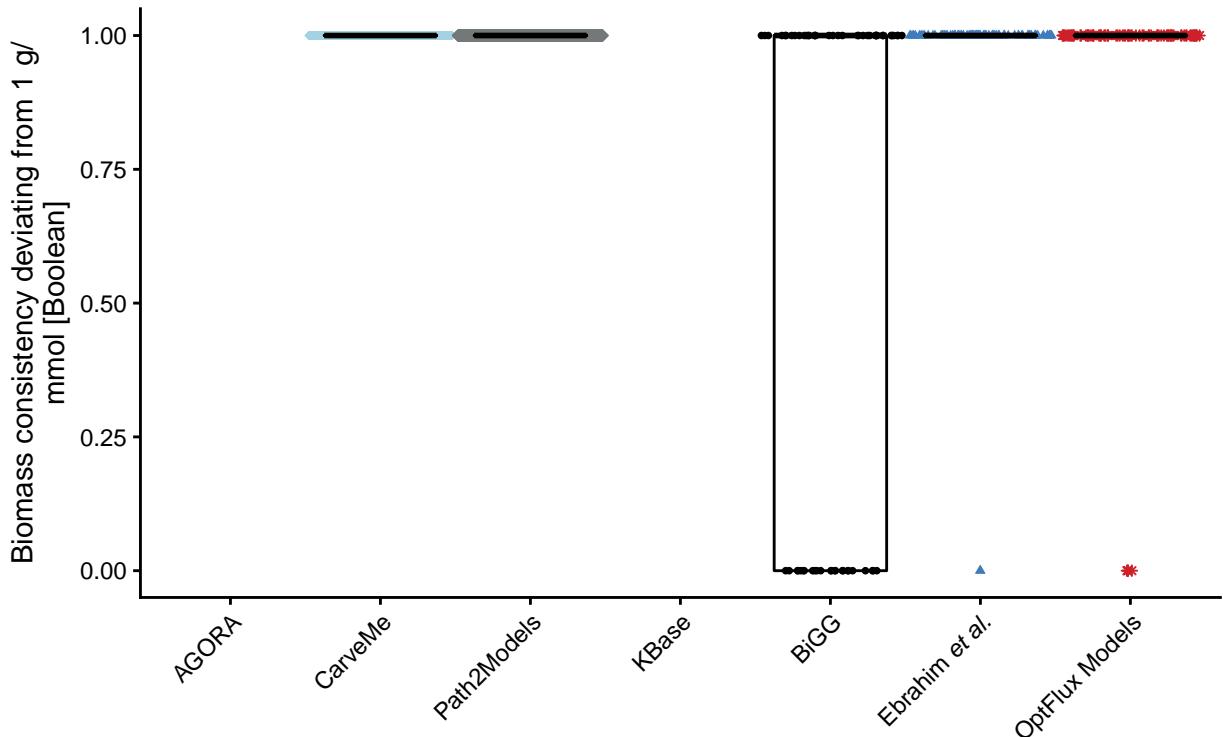


Figure S117: Biomass Consistency

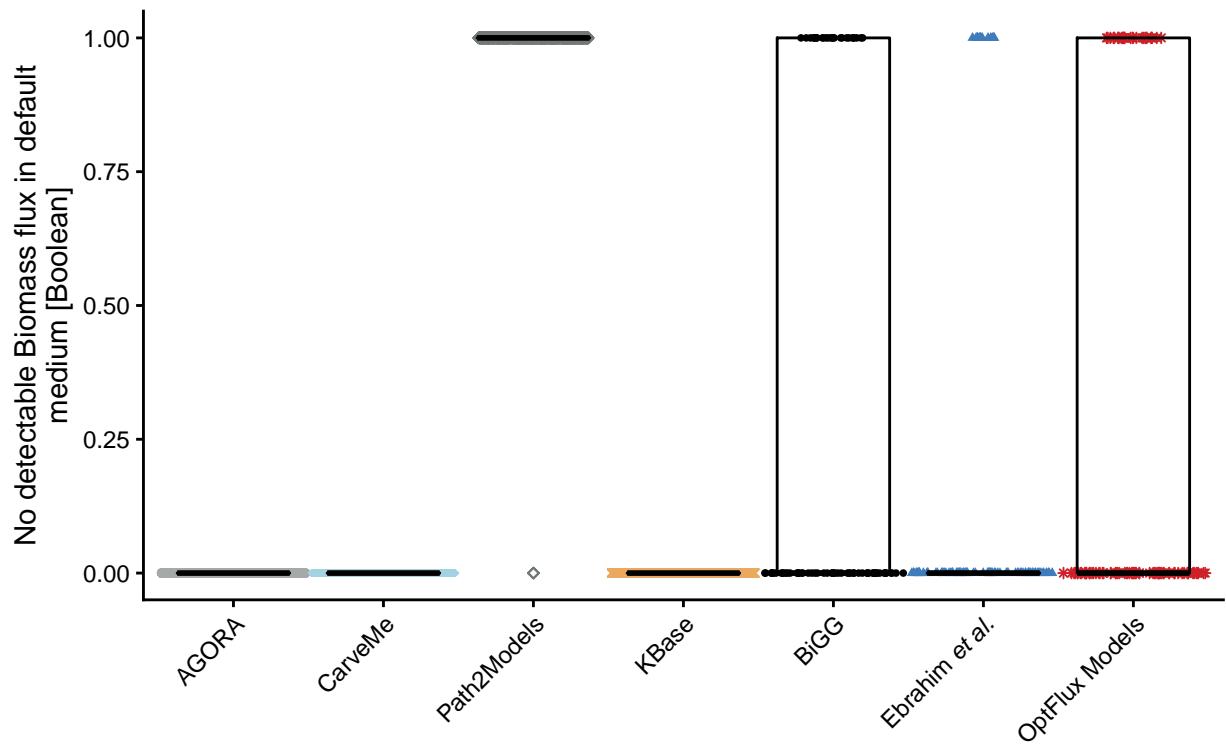


Figure S118: Biomass Production in Default Medium

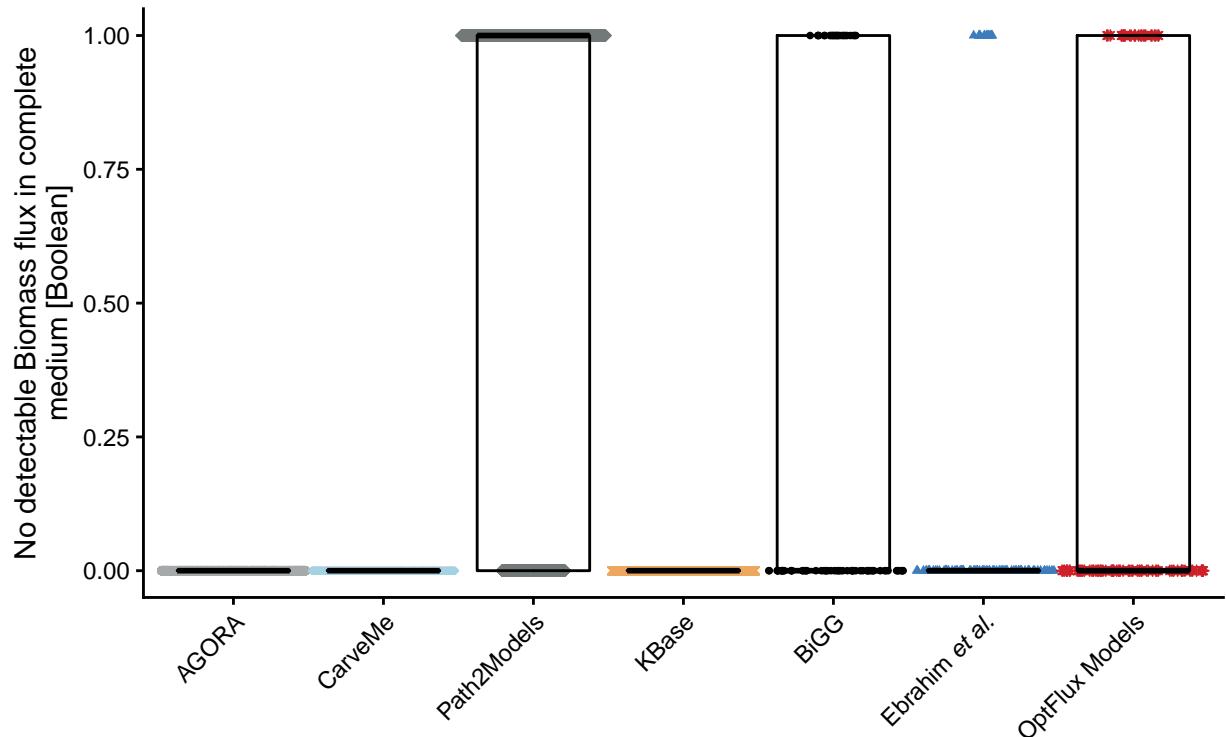


Figure S119: Biomass Production in Complete Medium

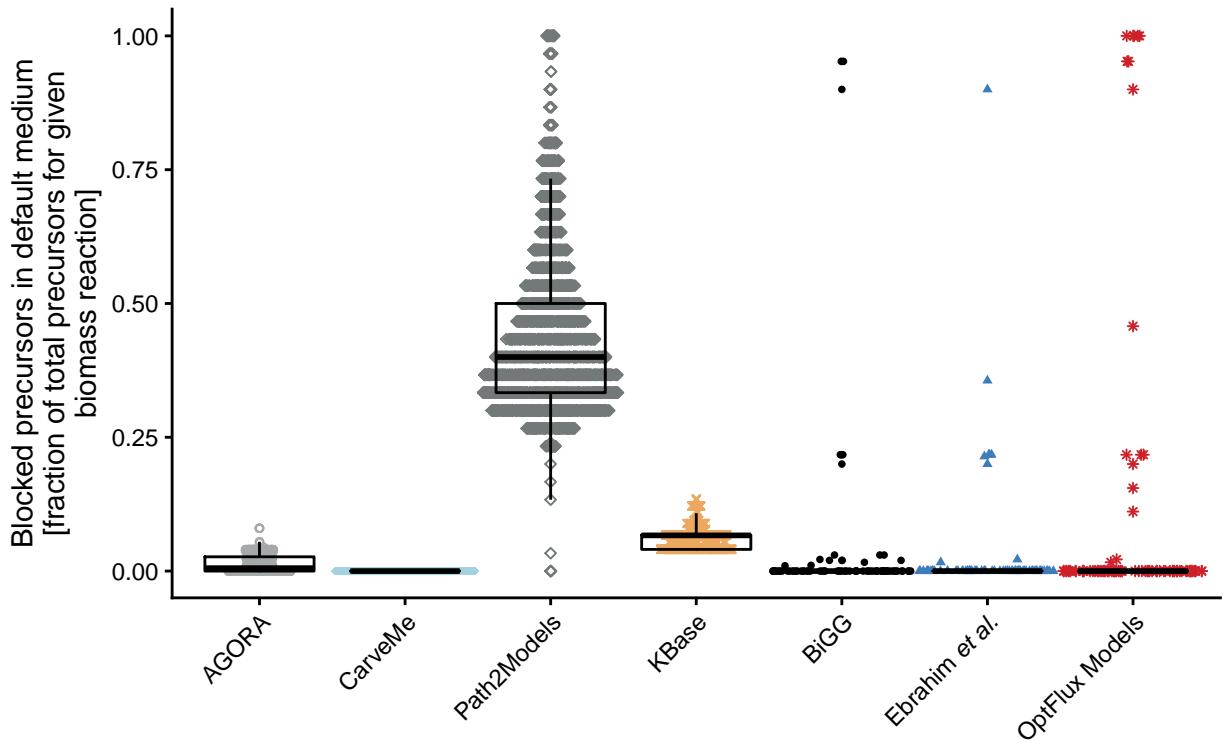


Figure S120: Blocked Biomass Precursors in Default Medium

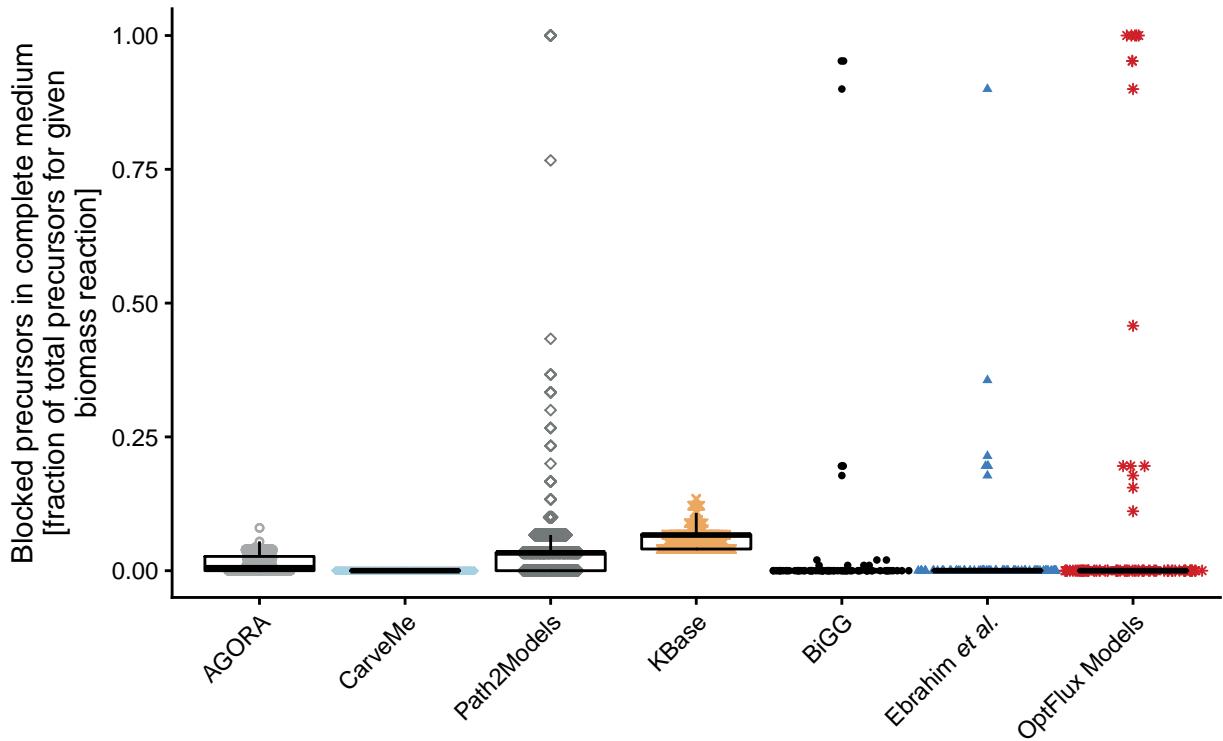


Figure S121: Blocked Biomass Precursors in Complete Medium

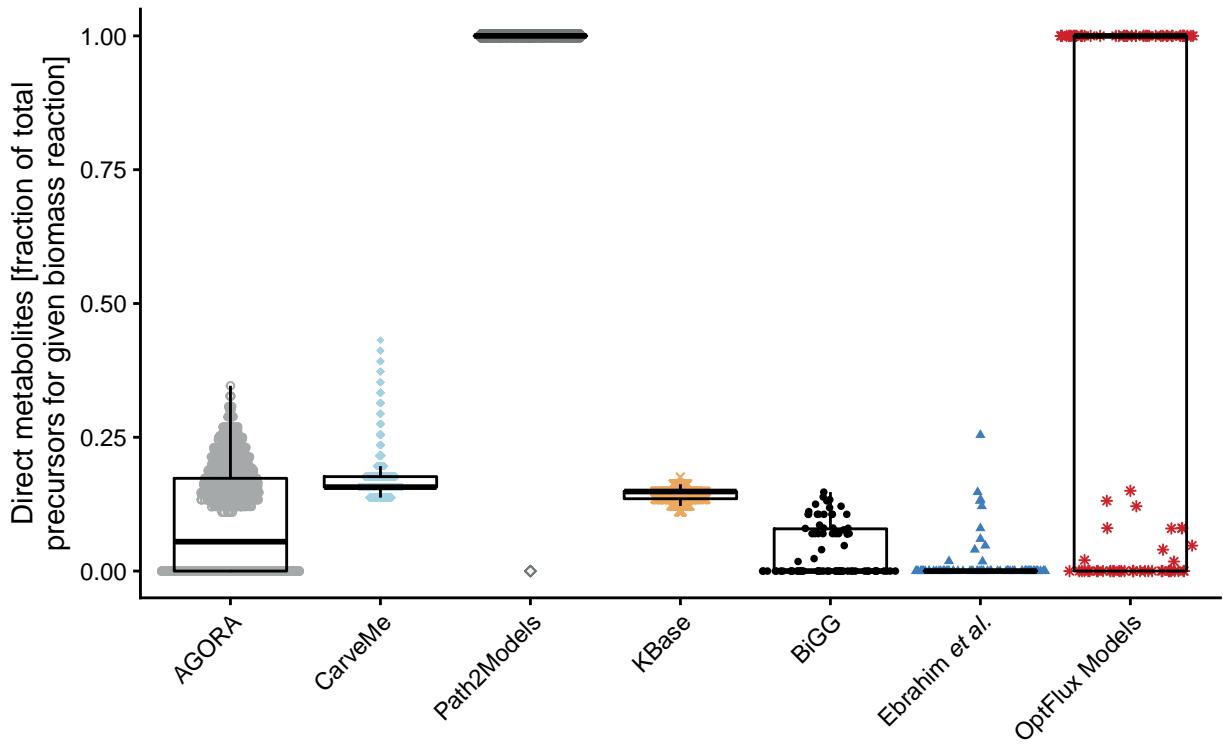


Figure S122: Ratio of Direct Metabolites in Biomass Reaction

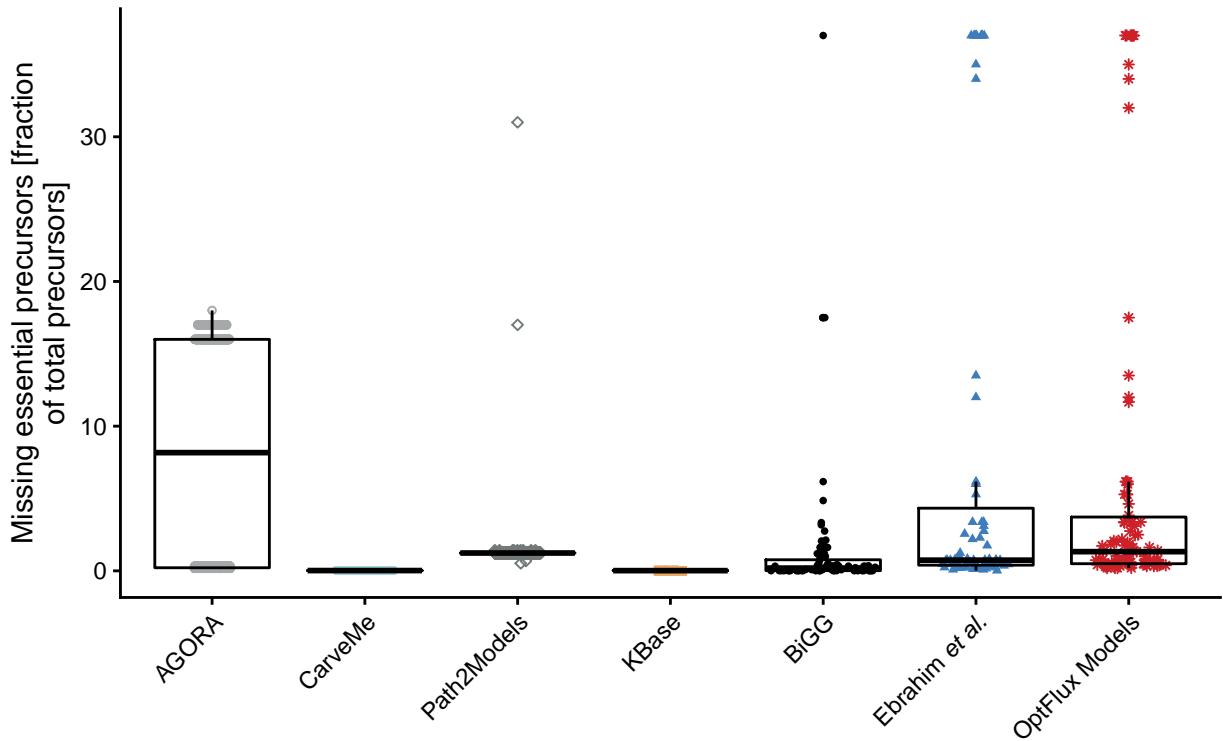


Figure S123: Number of Missing Essential Biomass Precursors

3.4.7 Energy Metabolism

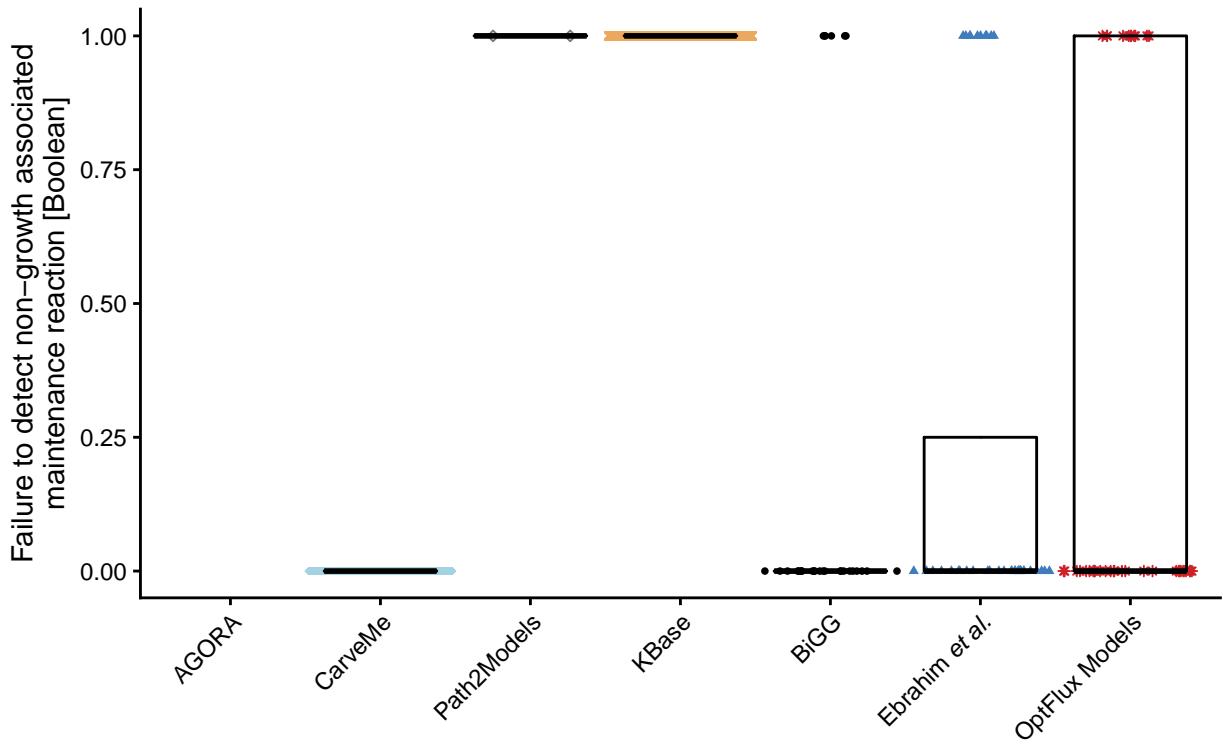


Figure S124: Non-Growth Associated Maintenance Reaction

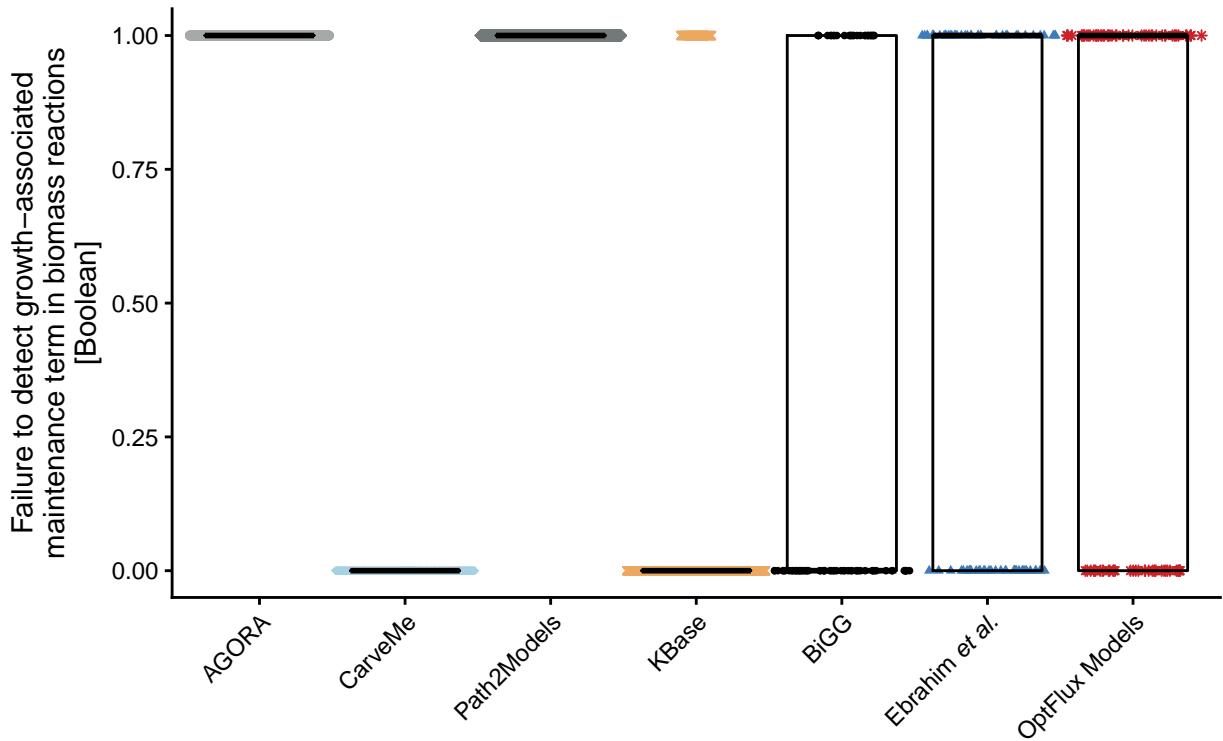


Figure S125: Growth-associated Maintenance in Biomass Reaction

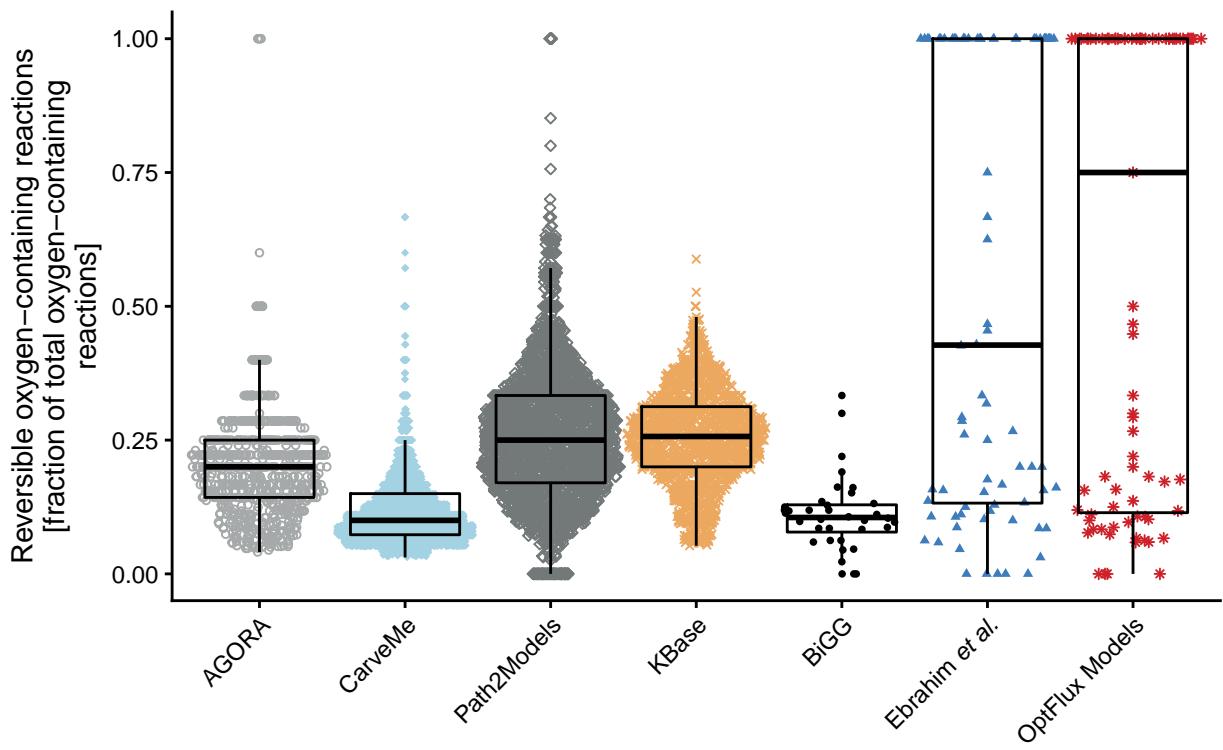


Figure S126: Number of Reversible Oxygen-Containing Reactions

3.4.8 Network Topology

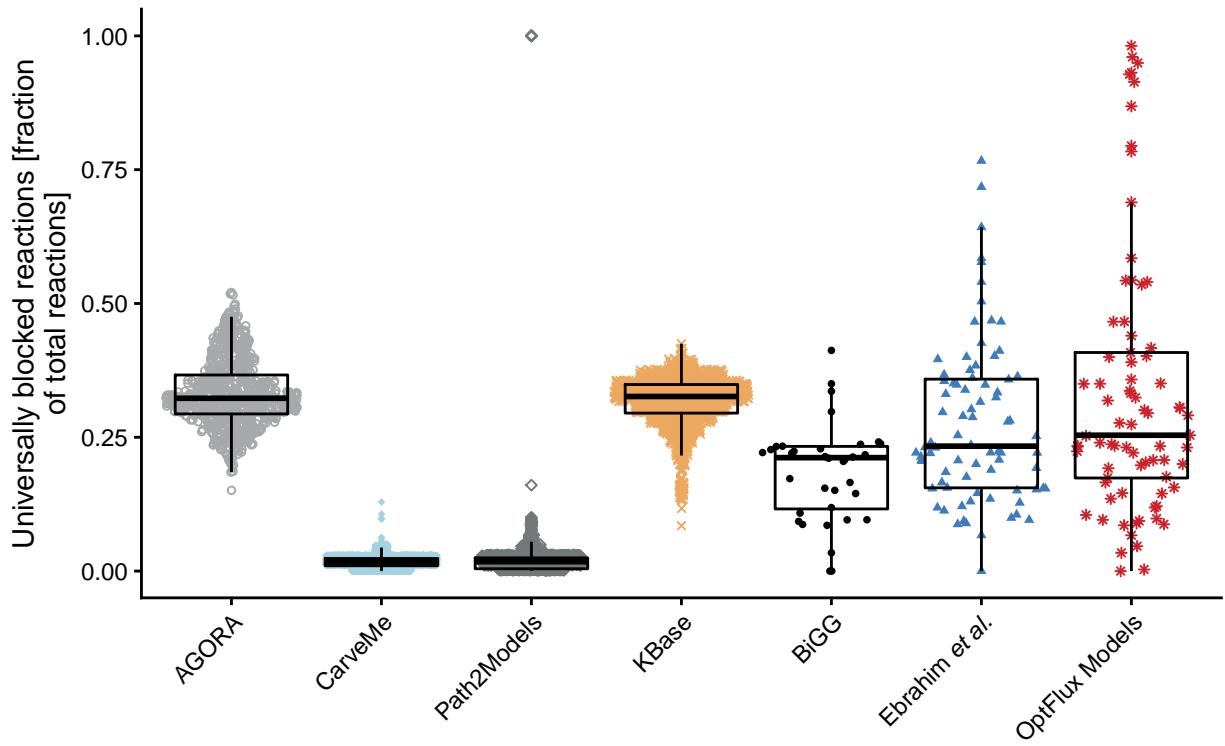


Figure S127: Universally Blocked Reactions

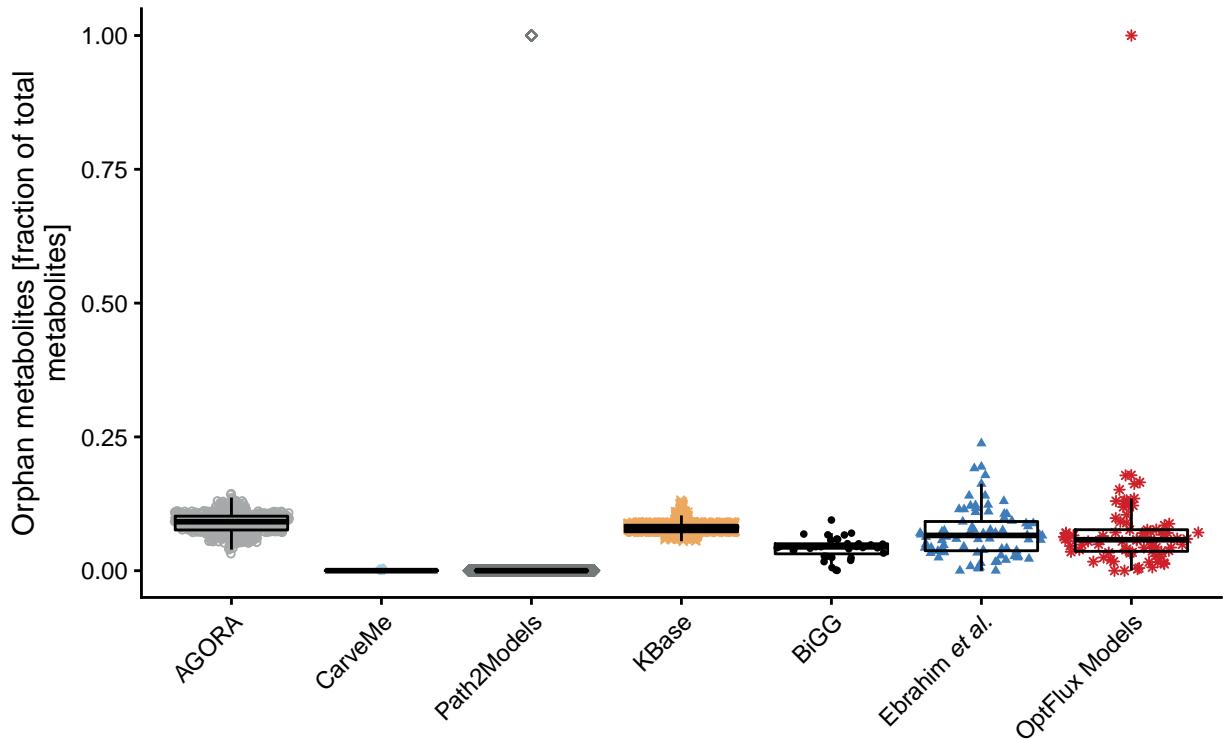


Figure S128: Orphan Metabolites

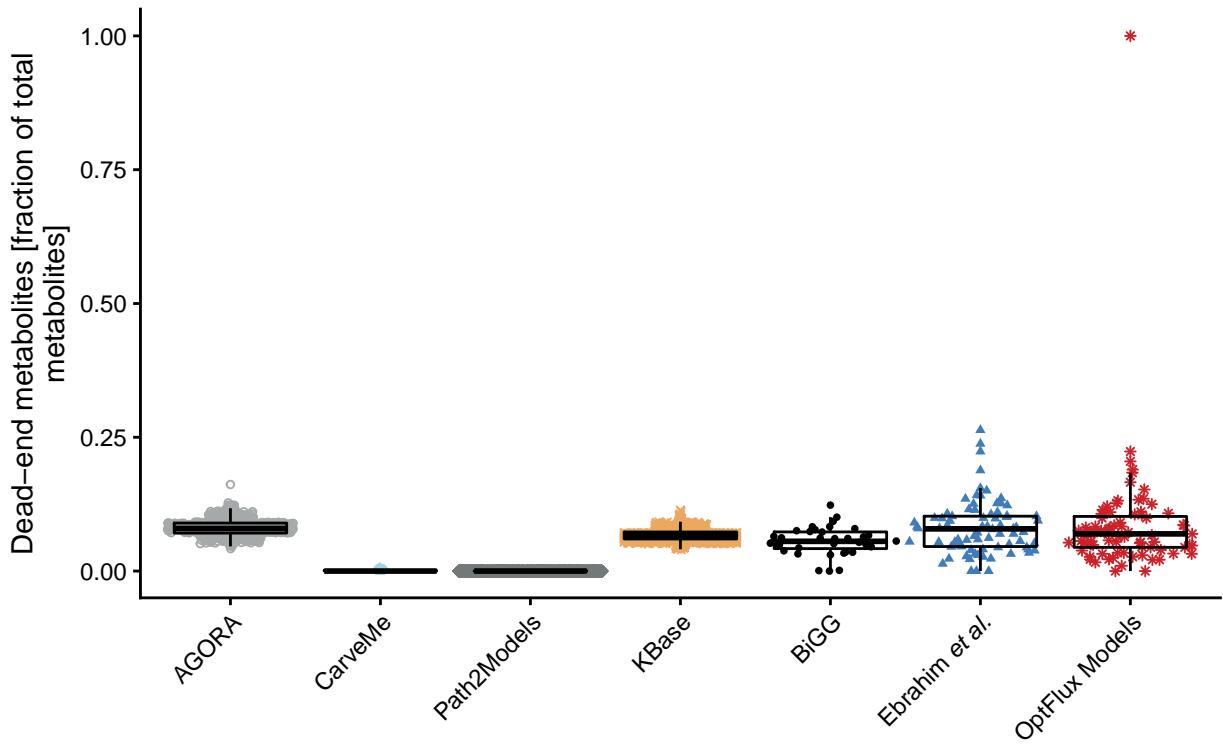


Figure S129: Dead-end Metabolites

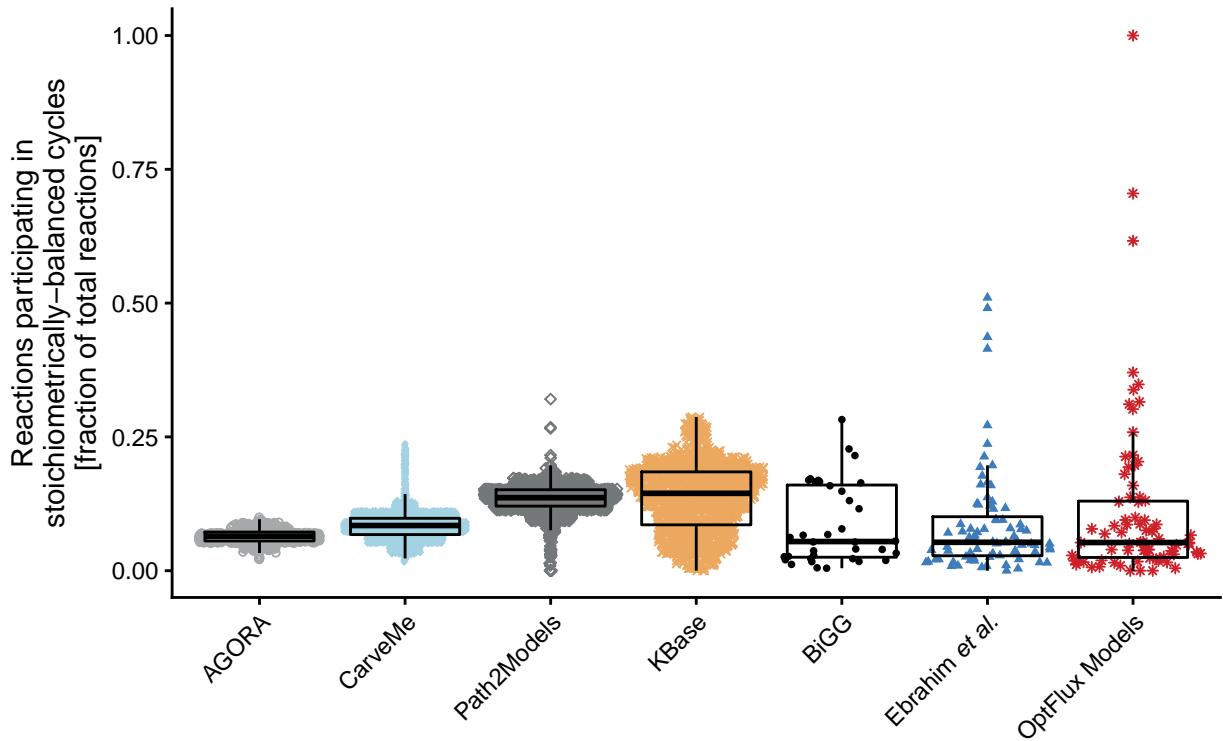


Figure S130: Stoichiometrically Balanced Cycles

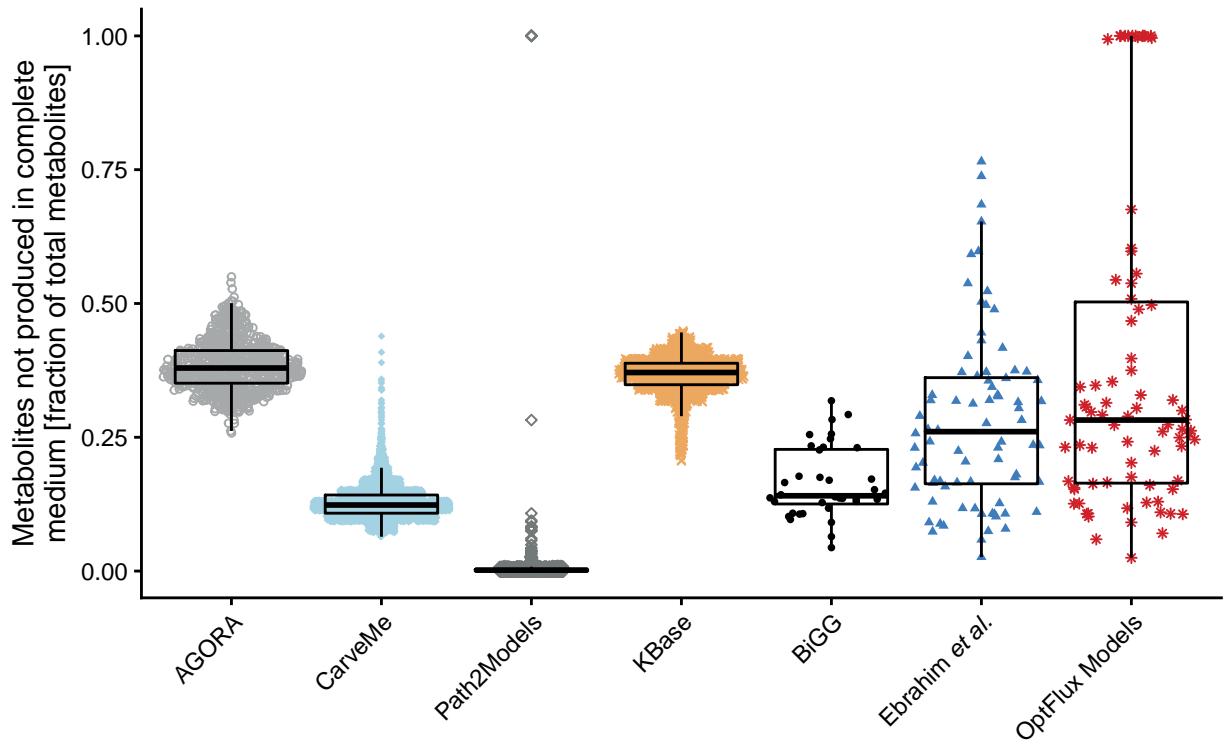


Figure S131: Metabolite Production in Complete Medium

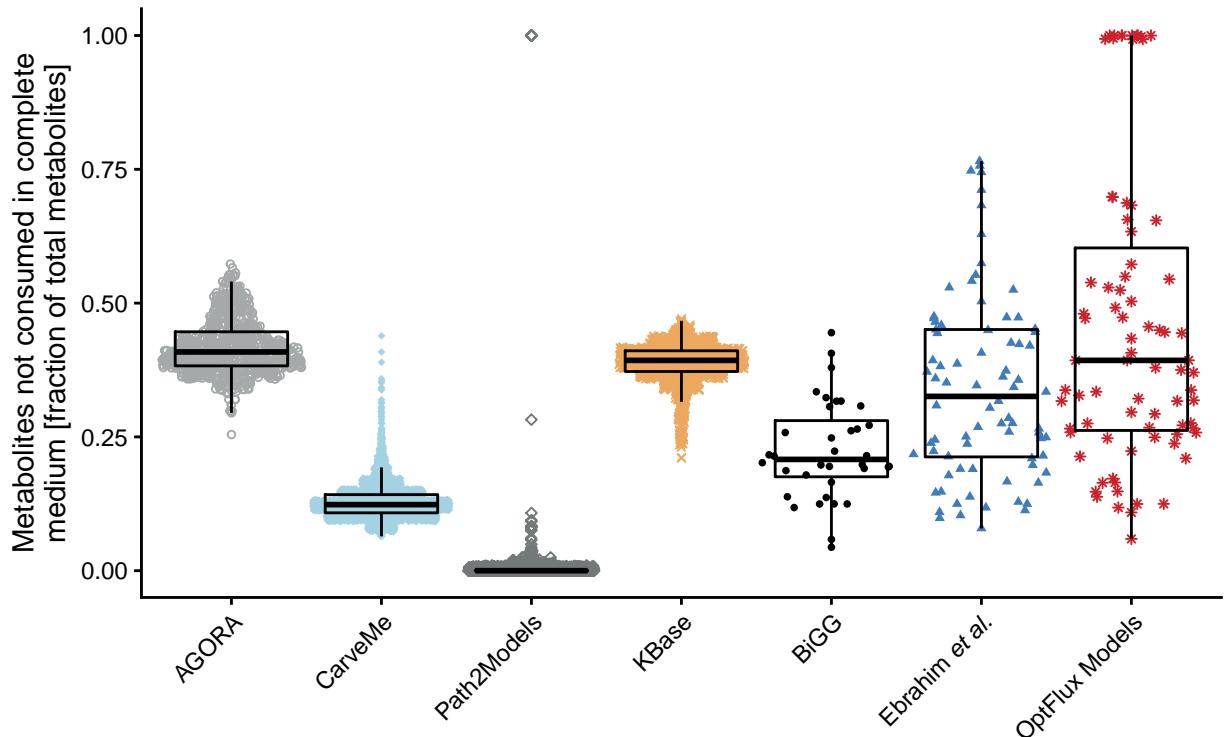


Figure S132: Metabolite Consumption in Complete Medium

3.4.9 Matrix Conditioning

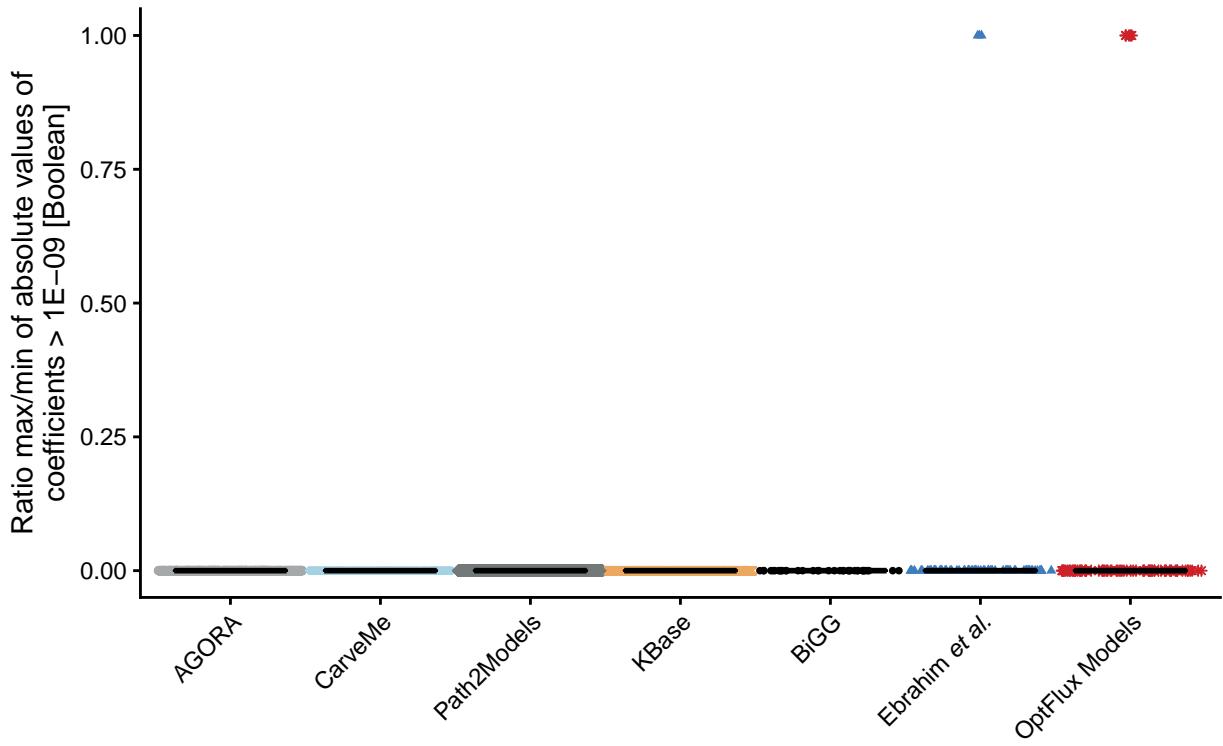


Figure S133: Ratio Min/Max Non-Zero Coefficients

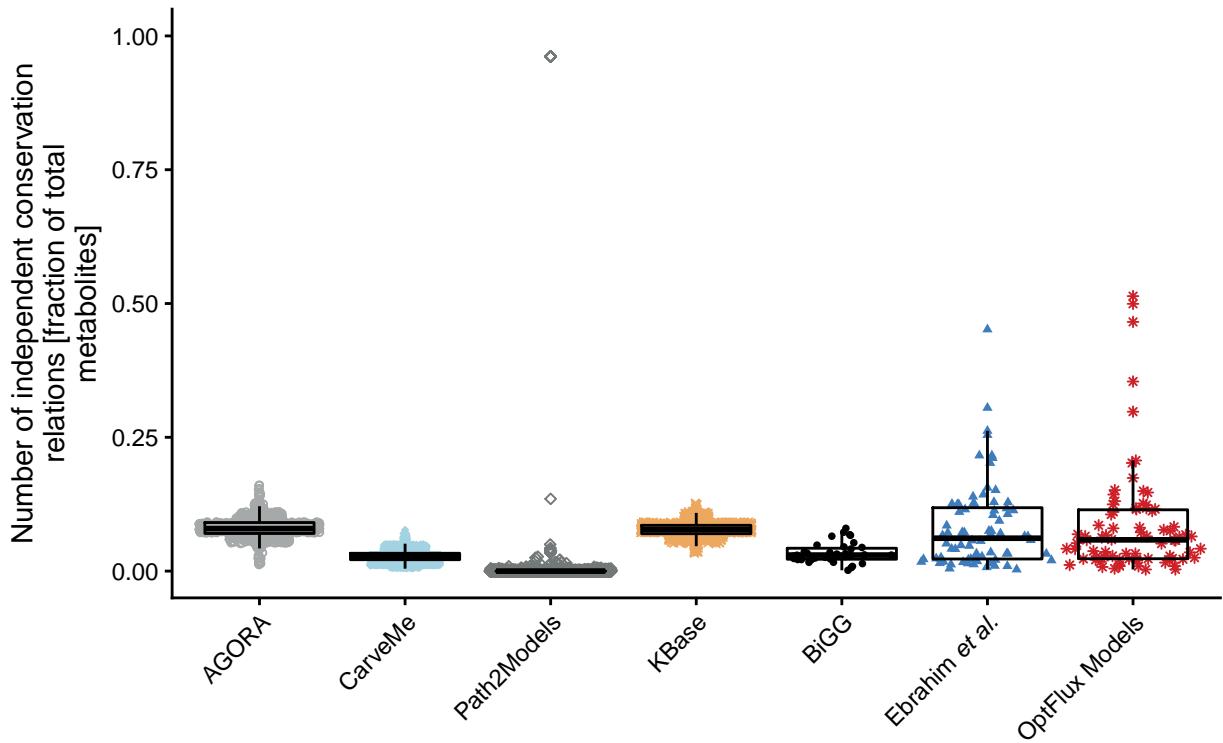


Figure S134: Independent Conservation Relations

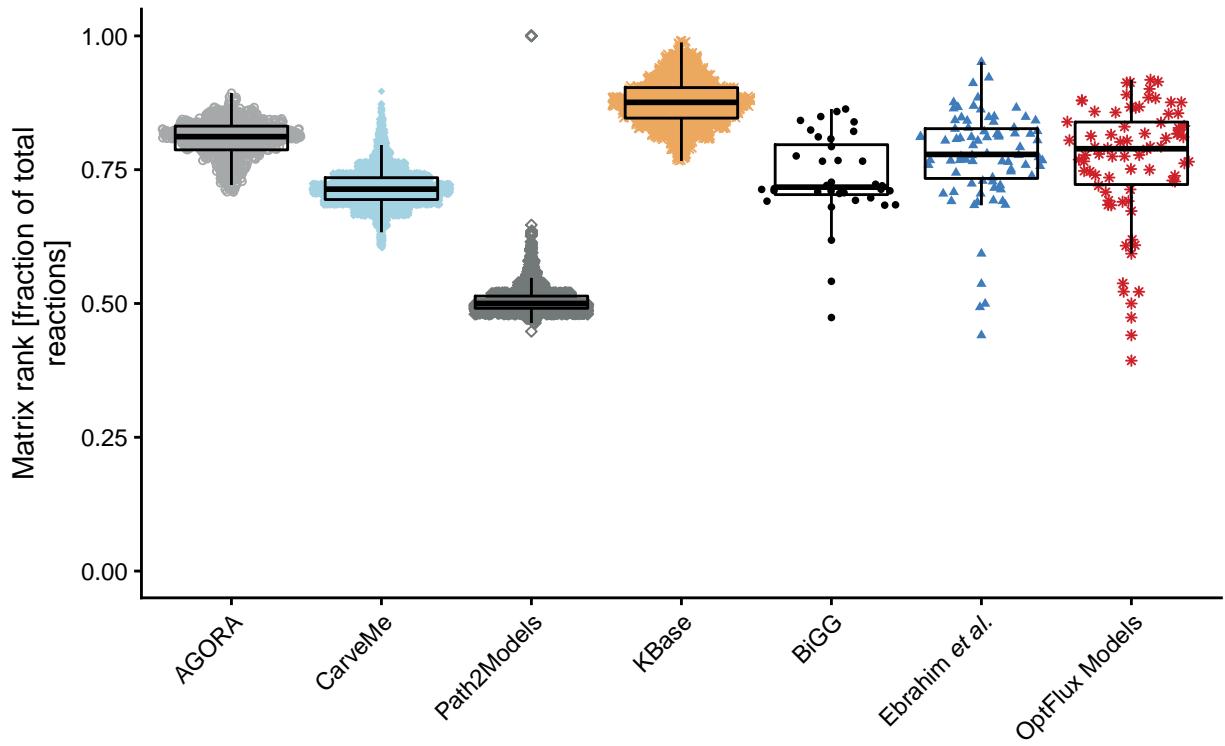


Figure S135: Rank

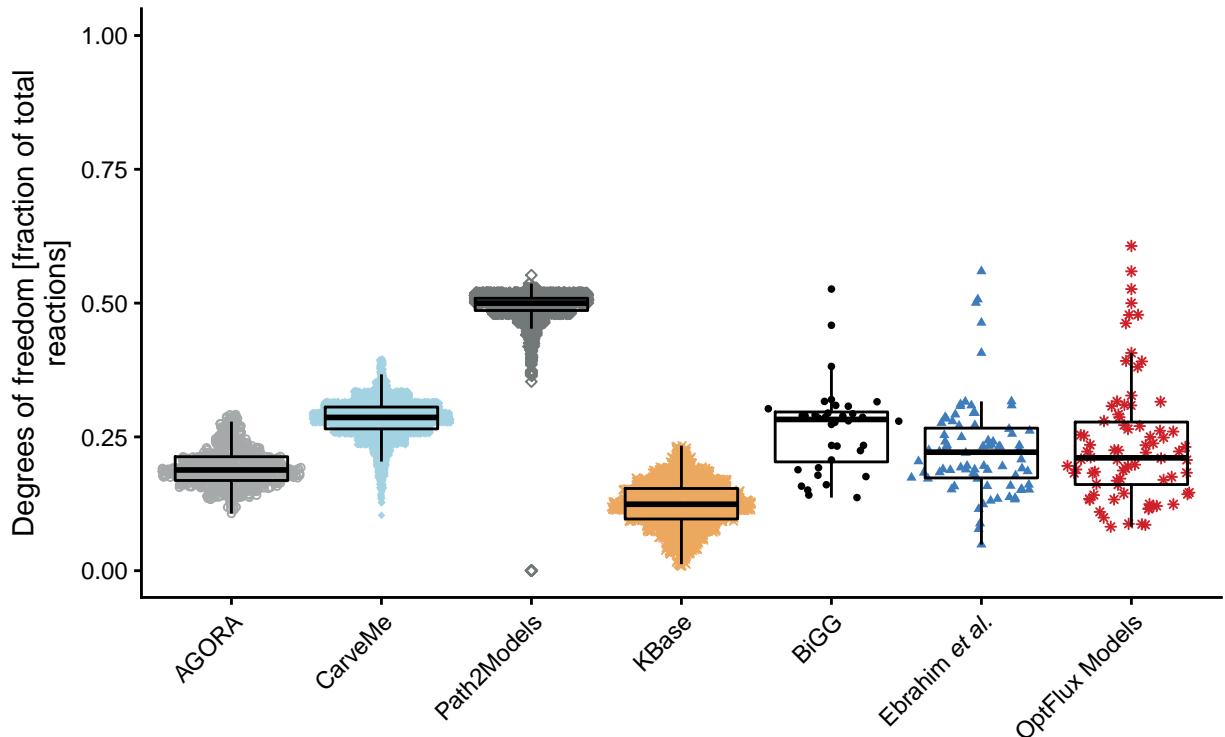


Figure S136: Degrees of Freedom