United States Environmental Protection Agency

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS)

Donovan Blatz

Carlie A. LaLone, Ph.D.

OXFORD SOT Society of Toxicology www.toxsci.oxfordjournals.org

TOXICOLOGICAL SCIENCES, 153(2), 2016, 228–245

doi: 10.1093/toxsci/kfw119
Advance Access Publication Date: June 30, 2016
Research article

Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity

Carlie A. LaLone,*,1 Daniel L. Villeneuve,* David Lyons,† Henry W. Helgen,‡ Serina L. Robinson,§,2 Joseph A. Swintek,¶ Travis W. Saari,* and Gerald T. Ankley*

Science-based, Streamlined, Transparent, Publicly Accessible PIPELINE

Available Knowledge

Predict Susceptibility

Office of Research and Development
Center for Computational Toxicology and Exposure, Great Lakes Toxicology and Ecology Division

The views expressed in this presentation are those of the authors and do not necessarily reflect the views or policies of the U.S. EPA.

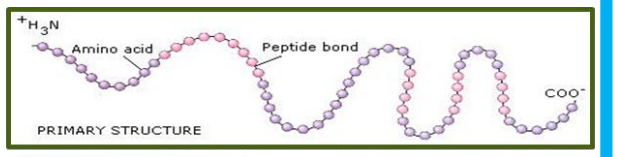
Overview

- Received Bayer bee FASTA sequences as CYP9Q-like Proteins.fasta along with TCP-metabolism data
  - Amino acids of interest- Gly111, Asn214, Ser310, Val371, Leu372
  - Template species/accession- Apis mellifera CYP9Q3: XP_006562363.1
- Perform SeqAPASS evaluations of NCBI sequences using Bayer's amino acids of interest
- Perform SeqAPASS evaluations of Bayer FASTA sequences with amino acids of interest
- Compare SeqAPASS sequences determined similar from NCBI to sequences from Bayer to see which sequences match up to be functionally similar based on sequence alone

Strategic Automated Approach for Assessing Protein Similarity

Level 1

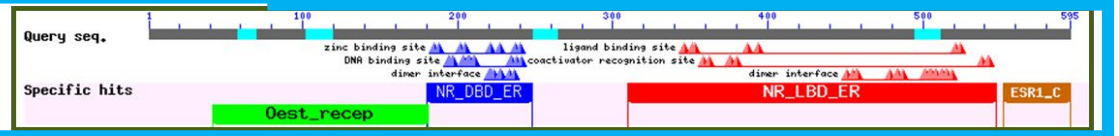
Primary Amino Acid Sequence Alignments



Ortholog Candidate Identification (RBH)

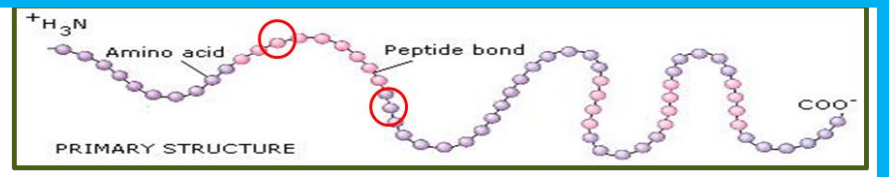
Level 2

Conserved Functional Domain Alignments



Level 3

Individual Amino Acid Residue Queries



Tertiary Protein Structure Considerations

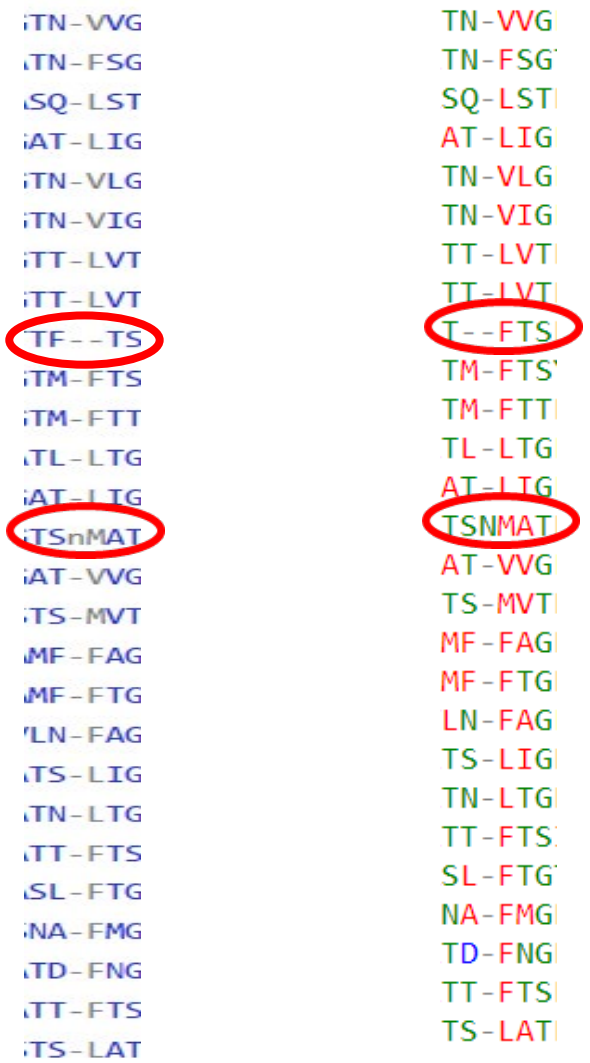
Low Level of Complexity

High level of Complexity

Developed with both researchers and risk assessors in mind

Alignment of FASTA using SeqAPASS, Cobalt and Clustal-W

NCBI Accession	Position 214	Amino Acid 214	Total Match 214	Position 215	Amino Acid 215	Total Match 215
CYP9e2_Apis_mellifera	214	N	Y	215	V	Y
CYP9DL4_Dufourea_novaeangliae	214	N	Y	215	F	N
CYP9Q1_Apis_mellifera	211	Q	Y	212	L	Y
CYP9Q2_Apis_mellifera	216	T	Y	217	L	Y
CYP9Q3_Apis_cerana	214	N	Y	215	V	Y
CYP9Q3_Apis_dorsata	214	N	Y	215	V	Y
CYP9Q3_Apis_mellifera	214	N	Y	215	V	Y
CYP9Q4_Bombus_impatiens	215	T	Y	216	L	Y
CYP9Q5_Bombus_impatiens	215	T	Y	216	L	Y
CYP9Q6_Bombus_impatiens	214	F	N	-	-	N
CYP9Q7_Eufriesa_mexicana	214	M	Y	215	F	N
CYP9Q8_Eufriesa_mexicana	214	M	Y	215	F	N
CYP9Q9_Habropoda_laboriosa	212	L	Y	213	L	Y
CYP9Q-like_1_Apis_florea	214	T	Y	215	L	Y
CYP9Q-like_1_Tetragonula_carbonaria	215	S	Y	217	M	N
CYP9Q-like_2_Apis_florea	216	T	Y	217	V	Y
CYP9Q-like_2_Tetragonula_carbonaria	214	S	Y	215	M	N
CYP9Q-like_A_vaga	211	F	N	212	F	N
CYP9Q-like_Andrena_haemorrhoa	211	F	N	212	F	N
CYP9Q-like_Colletes_cunicularius	214	N	Y	215	F	N
CYP9Q-like_E_nigrescens	214	S	Y	215	L	Y
CYP9Q-like_Lasioglossum_xanthopus	215	N	Y	216	L	Y
CYP9Q-like_Macropis_fulvipes	211	T	Y	212	F	N
CYP9Q-like_Melitta_haemorrhoidalis	214	L	Y	215	F	N
CYP9Q-like_N_Lathburiana	214	A	N	215	F	N
CYP9Q-like_N_melanderi	214	D	Y	215	F	N
CYP9Qlike_Xylocopa_violacea	214	T	Y	215	F	N
CYP9Qx(10)_Melipona_quadrifasciata	214	S	Y	215	L	Y



SeqAPASS Heat Map

This SeqAPASS Heat Map contains all the NCBI bee species found to have a similar protein based off sequence to the template sequence (Apis mellifera CYP9Q3).

NCBI Accession	Scientific Name	Protein Name
XP_006562363.1	Apis mellifera	cytochrome P450 9e2
ALG65081.1	Apis cerana cerana	cytochrome P450 9Q3
XP_016922294.2	Apis cerana	LOW QUALITY PROTEIN: cytochrome P450 9e2-like
XP_006613022.1	Apis dorsata	cytochrome P450 9e2-like isoform X1
KAF3421875.1	Frieseomelitta varia	hypothetical protein E2986_11836
XP_003393376.1	Bombus terrestris	cytochrome P450 9e2
KOX69484.1	Melipona quadrifasciata	Cytochrome P450 9e2
XP_033366579.1	Bombus vosnesenskii	cytochrome P450 9e2-like
XP_017758640.1	Eufriesea mexicana	PREDICTED: cytochrome P450 9e2-like
XP_033204351.1	Bombus vancouverensis nearcticus	cytochrome P450 9e2-like
XP_033174303.1	Bombus impatiens	uncharacterized protein LOC100740972
XP_033305595.1	Bombus bifarius	cytochrome P450 9e2-like
QCY50182.1	Bombus terrestris audax	cytochrome P450 9Q6
XP_031775226.1	Apis florea	cytochrome P450 9e2-like
XP_017887159.1	Ceratina calcarata	cytochrome P450 9e2-like
XP_034180228.1	Osmia lignaria	cytochrome P450 9e2-like
QBC73079.1	Osmia bicornis	cytochrome P450 mono-oxygenase
XP_029039416.1	Osmia bicornis bicornis	uncharacterized protein LOC114874380
XP_017794730.1	Habropoda laboriosa	PREDICTED: cytochrome P450 9e2-like
XP_015439019.1	Dufourea novaeangliae	PREDICTED: cytochrome P450 9e2-like
XP_033323895.1	Megalopta genalis	cytochrome P450 9e2-like
XP_031837097.1	Nomia melanderi	cytochrome P450 9e2-like
XP_003705491.1	Megachile rotundata	PREDICTED: cytochrome P450 9e2-like

Total Match
Partial Match
Not a Match
Susceptible Yes
Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Asiatic honeybee	Y	111G	214N	310S	371V	372L
Honey bee	Y	111G	214N	310S	371V	372L
Asiatic honeybee	Y	111G	214N	310S	371I	372L
Giant honeybee	Y	111G	214N	310S	371I	372L
Stingless bees	N	111K	214T	307T	368I	369A
Stingless bees	N	346K	449S	542T	603I	604A
Buff-tailed bumblebee	N	112E	215T	308T	369V	370V
Orchid bees	N	111E	214M	307T	368I	369A
Bumble bees	N	112E	215T	308A	369V	370A
Bumble bees	N	112E	215T	308T	369I	370A
Bumble bees	N	112E	215T	308A	369V	370A
Common eastern bumble bee	N	112E	215T	308A	369V	370A
Buff-tailed bumblebee	N	111D	213T	306S	367I	368T
Little honeybee	N	113S	216T	309T	370V	371V
Carpenter bees	N	110E	213N	304T	365I	366L
Orchard mason bee	N	111K	214N	306S	367I	368A
Red mason bee	N	111K	214N	306S	367I	368S
Red mason bee	N	623K	726N	818S	879I	880S
Digger bees	N	109R	212L	305T	366I	367V
Bees	N	111N	214N	307T	368A	369A
Sweat bees	N	111E	215N	308S	369I	370A
Alkali bee	N	111D	214D	307S	368S	369P
Alfalfa leafcutting bee	N	108E	211S	303S	364V	365V

SeqAPASS Heat Map

- This Heat Map shows all the bee species from the CYP9Q-like protein.fasta file. This comparison is based off the template species- Apis Mellifera CYP9Q3
- Similar results showing Amino Acid 1 at position 111 causing a susceptibility prediction of No for most species
- Amino Acid 2: 4 species produces susceptibility prediction of No- Phenylalanine and Alanine

Total Match
Partial Match
Susceptible Yes
Not a Match
Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Apis mellifera CYP9Q3	Y	111G	214N	310S	371V	372L
Dufourea novaeangliae CYP9DL4	N	111N	214N	307T	368A	369A
Apis mellifera CYP9Q1	Y	111G	211Q	308T	370I	371L
Apis mellifera CYP9Q2	N	113S	216T	310T	371I	372A
Apis cerana CYP9Q3	Y	111G	214N	310S	371I	372L
Apis dorsata CYP9Q3	Y	111G	214N	310S	371I	372L
Apis mellifera CYP9Q3	Y	111G	214N	310S	371V	372L
Bombus impatiens CYP9Q4	N	112E	215T	308A	369V	370A
Bombus impatiens CYP9Q5	N	112E	215T	308T	369V	370A
Bombus impatiens CYP9Q6	N	111D	214F	306S	367I	368T
Eufriesa mexicana CYP9Q7	N	111D	214M	307S	368I	369A
Eufriesa mexicana CYP9Q8	N	111E	214M	307T	368I	369A
Habropoda laboriosa CYP9Q9	N	109R	212L	305T	366I	367V
Apis florea CYP9Q-like-1	N	111N	214T	306T	367I	368A
Tetragonila carbonaria CYP9Q-like-1	N	111K	215S	309S	370I	371A
Apis florea CYP9Q-like-2	N	113S	216T	309T	370V	371V
Tetragonila carbonaria CYP9Q-like-2	N	111K	214S	307S	368I	369A
Andrena vaga CYP9Q-like	N	111A	211F	304T	365I	366G
Andrena haemorrhoa CYP9Q-like	N	111A	211F	304T	365I	366G
Colletes cunicularius CYP9Q-like	N	111E	214N	307T	368I	369A
E nigrescens CYP9Q-like	N	111E	214S	307T	368I	369V
Lasioglossum xanthopus CYP9Q-like	N	111D	215N	308S	369V	370A
Macropis fulvipes CYP9Q-like	N	108K	211T	301S	362V	363A
Melitta haemorrhoidalis CYP9Q-like	N	111E	214L	306T	367I	368A
N lathburiana CYP9Q-like	N	111E	214A	307S	368I	369A
N melanderi CYP9Q-like	N	111D	214D	307S	368S	369P
Xylocopa violacea CYP9Q-like	N	111D	214T	307S	368I	369V
Melipona quadrifasciata CYP9Q-like	N	111K	214S	307T	368I	369A

SeqAPASS Heat Map

Reverse approach performed with Megachile rotundata as template species with previously aligned amino acids.

NCBI Accession	Scientific Name	Protein Name
XP_003705491.1	Megachile rotundata	PREDICTED: cytochrome P450 9e2-like
XP_029039348.1	Osmia bicornis bicornis	cytochrome P450 9e2-like
QBC73083.1	Osmia bicornis	cytochrome P450 mono-oxygenase
XP_034180498.1	Osmia lignaria	LOW QUALITY PROTEIN: uncharacterized protein LOC117604504
XP_017794737.1	Habropoda laboriosa	PREDICTED: cytochrome P450 9e2-like
XP_015439018.1	Dufourea novaeangliae	PREDICTED: cytochrome P450 9e2
XP_033366576.1	Bombus vosnesenskii	cytochrome P450 9e2-like
XP_033305593.1	Bombus bifarius	cytochrome P450 9e2-like
XP_033204350.1	Bombus vancouverensis nearcticus	cytochrome P450 9e2-like
XP_012250003.1	Bombus impatiens	cytochrome P450 9e2-like
XP_017758638.1	Eufriesea mexicana	PREDICTED: uncharacterized protein LOC108549653
XP_003393379.1	Bombus terrestris	cytochrome P450 9e2
XP_031848282.1	Nomia melanderi	cytochrome P450 9e2-like
XP_033323887.1	Megalopta genalis	uncharacterized protein LOC117219102
KAF3426864.1	Frieseomelitta varia	hypothetical protein E2986_12664
KOX69483.1	Melipona quadrifasciata	Cytochrome P450 9e2
XP_026672903.1	Ceratina calcarata	uncharacterized protein LOC108629176
XP_028525692.1	Apis cerana	cytochrome P450 9e2
PBC28686.1	Apis cerana cerana	Cytochrome P450
XP_031365041.1	Apis dorsata	uncharacterized protein LOC102673719
XP_031775207.1	Apis florea	uncharacterized protein LOC100867182
XP_026300962.1	Apis mellifera	uncharacterized protein LOC725621

Total Match
Partial Match
Susceptible Yes
Not a Match
Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Alfalfa leafcutting bee	Y	108E	211S	303S	364V	365V
Red mason bee	Y	108E	211T	303S	364V	365I
Red mason bee	Y	108E	211T	303S	364V	365I
Orchard mason bee	N	685G	788T	880S	941V	942I
Digger bees	Y	111K	214N	306S	367V	368I
Bees	N	112A	215S	307S	368V	369I
Bumble bees	N	113D	216H	308S	369T	370I
Bumble bees	N	113D	216H	308S	369T	370I
Bumble bees	N	113D	216H	308S	369T	370I
Common eastern bumble bee	N	113D	216H	308S	369T	370I
Orchid bees	Y	626D	729N	821S	882A	883I
Buff-tailed bumblebee	N	113D	216H	308S	369T	370I
Alkali bee	N	112S	215H	307S	368V	369I
Sweat bees	N	1199S	1302N	1394S	1455A	1456I
Stingless bees	Y	876D	979S	1071S	1132S	1133A
Stingless bees	Y	303D	406N	499S	560A	561I
Carpenter bees	Y	637D	740N	830T	891S	892I
Asiatic honeybee	N	112A	215T	308S	369V	370V
Asiatic honeybee	N	112A	215T	308S	369V	370V
Giant honeybee	N	621A	724T	817S	878V	879V
Little honeybee	N	619A	722T	815S	876V	877V
Honey bee	N	642A	745T	838S	899V	900V

Species with 100% similarity between SeqAPASS and Bayer species

NCBI Accession	Scientific Name	Protein Name
XP_006562363.1	Apis mellifera	cytochrome P450 9e2
XP_016922294.2	Apis cerana	LOW QUALITY PROTEIN: cytochrome P450 9e2-like
XP_017758640.1	Eufriesea mexicana	PREDICTED: cytochrome P450 9e2-like
XP_031775226.1	Apis florea	cytochrome P450 9e2-like
XP_017794730.1	Habropoda laboriosa	PREDICTED: cytochrome P450 9e2-like
XP_031837097.1	Nomia melanderi	cytochrome P450 9e2-like

Total Match
Partial Match
Not a Match
Susceptible Yes
Susceptible No

Common Name	Similar Susceptibility	Amino Acid 1	Amino Acid 2	Amino Acid 3	Amino Acid 4	Amino Acid 5
Honey bee	Y	111G	214N	310S	371V	372L
Asiatic honeybee	Y	111G	214N	310S	371I	372L
Orchid bees	N	110K	214M	307T	368I	369A
Little honeybee	N	112K	216T	309T	370V	371V
Digger bees	N	108K	212L	305T	366I	367V
Alkali bee	N	110K	214D	307S	368S	369P

This evaluation would be performed by using the species and protein of interest while look for species that are included within the FASTA file that was sent over. Although these species are present, SeqAPASS looks alone at similarity of sequence to the template species. It cannot look at similarity of function between isoforms and species.

Other information and data

- Sequence alignment and non-targeted comparison for bee species of interest
  - Apis mellifera (CYP9Q3) as template species with Bayer protein sequences compared
  - Apis mellifera (CYP9Q3) as template species with NCBI protein sequences compared
  - Megachile rotundata (9e2-like) as template species with NCBI protein sequences compared
- Apis mellifera as template species aligned to three Megachile rotundata 9e2 sequences