

# Categorizing endemic and introduced species using DNA sequence signatures

Jeremy Andersen, Peter Oboyski, George Roderick  
Essig Museum of Entomology  
Department of Environmental Science, Policy, and Management  
University of California, Berkeley







# The Taxonomic Impediment

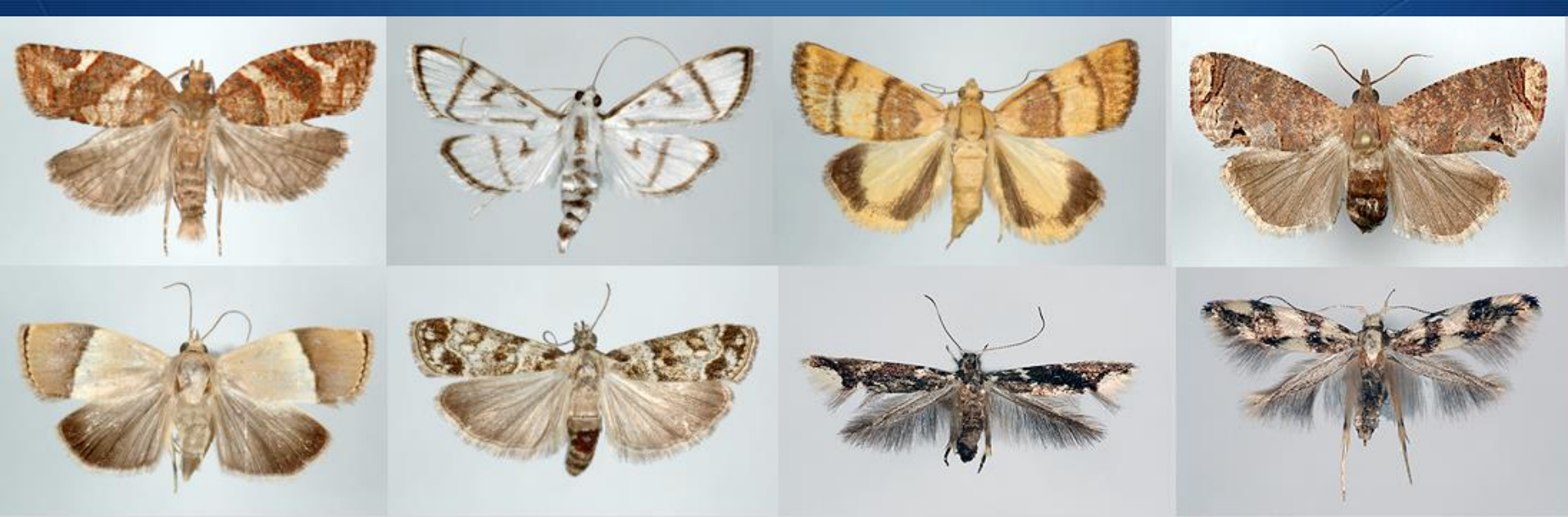






Oboyski 2015, Tahiti, French Polynesia





*Streplicrates* n.sp., part of a radiation of >8 spp.

# Taxonomic Tools

## DNA barcoding



U.S. National Library of Medicine

NCBI National Center for Biotechnology Information

Sign In to NCBI

BLAST® » blastn suite » RID-H3PP4H3T015

Home Recent Results Saved Strategies Help

### BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

[YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: Nucleotide Sequence (658 letters)

**RID** H3PP4H3T015 (Expires on 06-03 07:46 am)  
**Query ID** lc|Query\_70787  
**Description** None  
**Molecule type** nucleic acid  
**Query Length** 658

**Database Name** nr  
**Description** Nucleotide collection (nr)  
**Program** BLASTN 2.8.0+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

#### Sequences producing significant alignments:

Select: [All](#) [None](#) [Selected 0](#)

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<i>Thaumatococcus maculata</i> voucher 11ANIC-12868 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1086	1086	100%	0.0	96%	KF399439.1
<input type="checkbox"/>	<i>Thaumatococcus ayala</i> voucher 11ANIC-12873 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1053	1053	99%	0.0	96%	KF399581.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00720350 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	KY323225.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00724745 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	KY323199.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00720256 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	KY323126.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00720490 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	KY323114.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00724753 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	KY323053.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT_00666690 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1027	1027	100%	0.0	95%	GU695432.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher 11ANIC-12875 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1016	1016	100%	0.0	95%	KF403389.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher 11ANIC-12876 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1009	1009	99%	0.0	94%	KF401748.1
<input type="checkbox"/>	<i>Thaumatococcus ayala</i> voucher 11ANIC-12871 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	1003	1003	94%	0.0	96%	KF406066.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00724792 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	985	985	96%	0.0	95%	KY323135.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00720047 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	966	966	94%	0.0	95%	KY323096.1
<input type="checkbox"/>	<i>Cyclotheleba endohassa</i> voucher 11ANIC-12828 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	966	966	100%	0.0	93%	KF402830.1
<input type="checkbox"/>	<i>Cyclotheleba glava</i> voucher 11ANIC-12838 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	966	966	100%	0.0	93%	KF395867.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00721872 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	965	965	94%	0.0	95%	KY323220.1
<input type="checkbox"/>	<i>Thaumatococcus zosterophanes</i> voucher USNM ENT00724760 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	965	965	94%	0.0	95%	KY323124.1
<input type="checkbox"/>	<i>Cyclotheleba endohassa</i> voucher 11ANIC-12829 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	961	961	100%	0.0	93%	KF396530.1
<input type="checkbox"/>	<i>Cyclotheleba endohassa</i> voucher 11ANIC-12830 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	955	955	100%	0.0	93%	KF403539.1
<input type="checkbox"/>	<i>Cyclotheleba endohassa</i> voucher 11ANIC-12827 cytochrome oxidase subunit 1 (COI) gene, partial cds, mitochondrial	955	955	100%	0.0	93%	KF402417.1



## Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Lepidoptera sp. 001 PS-2011 voucher Eo00652 cytochrome oxidase subunit I (COI) gene, partial c</a>	1216	1216	100%	0.0	100%	<a href="#">JF729549.1</a>
<input type="checkbox"/>	<a href="#">Lepidoptera sp. 001 PS-2011 voucher Eo00612 cytochrome oxidase subunit I (COI) gene, partial c</a>	1199	1199	100%	0.0	99%	<a href="#">JF729515.1</a>
<input type="checkbox"/>	<a href="#">Bactra blepharopis voucher NIBGE MOT-03665 cytochrome oxidase subunit 1 (COI) gene, partial c</a>	854	854	99%	0.0	90%	<a href="#">KX862772.1</a>
<input type="checkbox"/>	<a href="#">Apotomis demissana voucher MM19905 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitc</a>	850	850	100%	0.0	90%	<a href="#">KT782601.1</a>
<input type="checkbox"/>	<a href="#">Apotomis demissana voucher MM17773 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitc</a>	850	850	100%	0.0	90%	<a href="#">KT782502.1</a>
<input type="checkbox"/>	<a href="#">Tortricidae sp. BOLD:ACM3468 voucher USNM ENT 00736892 cytochrome oxidase subunit 1 (COI</a>	850	850	100%	0.0	90%	<a href="#">KP850214.1</a>
<input type="checkbox"/>	<a href="#">Apotomis capreana voucher TLMF Lep 08079 cytochrome oxidase subunit 1 (COI) gene, partial cds</a>	850	850	100%	0.0	90%	<a href="#">KM572369.1</a>
<input type="checkbox"/>	<a href="#">Archips purpurana voucher 10BBCLP-1918 cytochrome oxidase subunit 1 (COI) gene, partial cds; r</a>	848	848	99%	0.0	90%	<a href="#">KM546243.1</a>
<input type="checkbox"/>	<a href="#">Tortricidae sp. BOLD:ACM3468 voucher USNM ENT 00739632 cytochrome oxidase subunit 1 (COI</a>	846	846	100%	0.0	90%	<a href="#">KP850914.1</a>
<input type="checkbox"/>	<a href="#">Tortrix oriarcha voucher 11ANIC-10354 cytochrome oxidase subunit 1 (COI) gene, partial cds; mitoc</a>	845	845	100%	0.0	90%	<a href="#">KF404188.1</a>
<input type="checkbox"/>	<a href="#">Tortrix hydractis voucher 11ANIC-10348 cytochrome oxidase subunit 1 (COI) gene, partial cds; mito</a>	845	845	100%	0.0	90%	<a href="#">KF400744.1</a>

Either no specimens of this species (or genus) have been sequenced, or a taxonomic expert has not yet identified sequenced specimens, or it is an undescribed species.

# Who cares?

- Do we really need to put a name on everything?
- Morpho-species (or molecular diversity) is good enough.
- Names provide ecological and evolutionary context
- Names provide links to published literature
- Ecological or behavioral information – food web relationships
- Public health information – disease vector or reservoir
- Distribution information – endemic or invasive

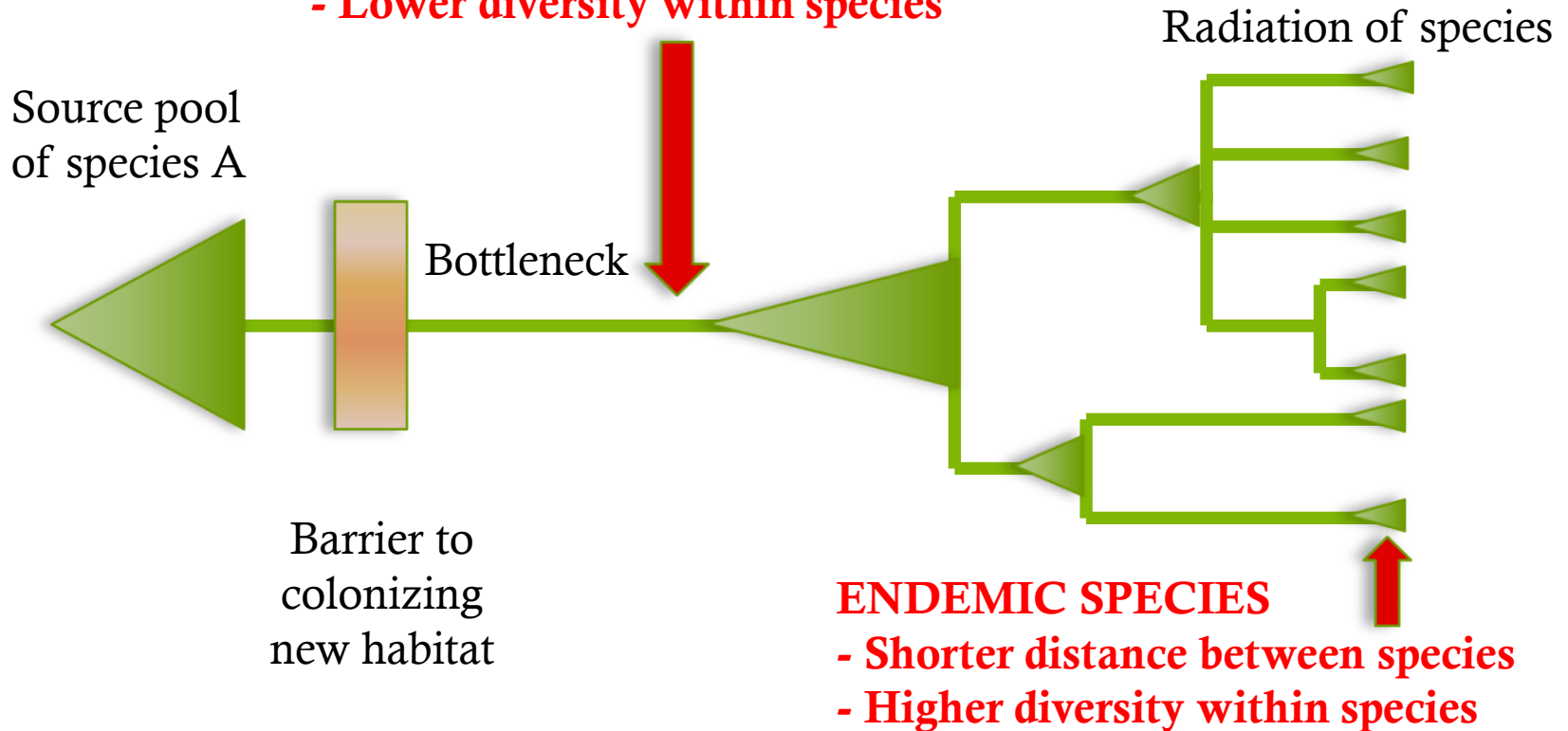
**Identifications take  
time and money  
and taxonomic expertise**



# What can we determine from DNA data?

## NEW COLONIZERS

- Greater distance between species
- Lower diversity within species



# System to test predictions



[MooreaBiocode.org](http://MooreaBiocode.org)





- ◆ Chris Meyer – Smithsonian Institution
- ◆ Rosemary Gillespie – UC Berkeley
- ◆ Sylvain Charlat – Universite de Lyon
- ◆ Curtis Ewing – UC Berkeley, Cal Fire
- ◆ Neil Davies – Gump Research Station



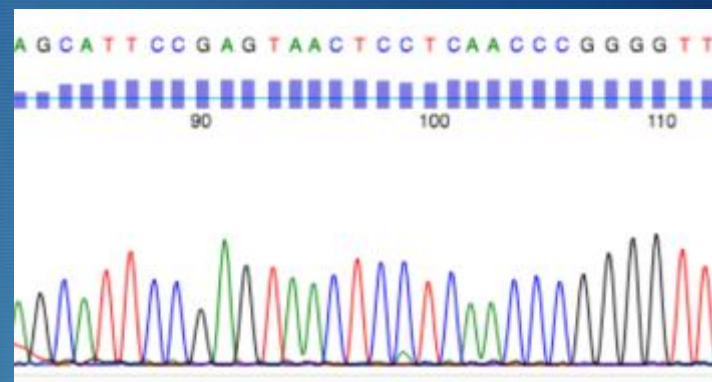
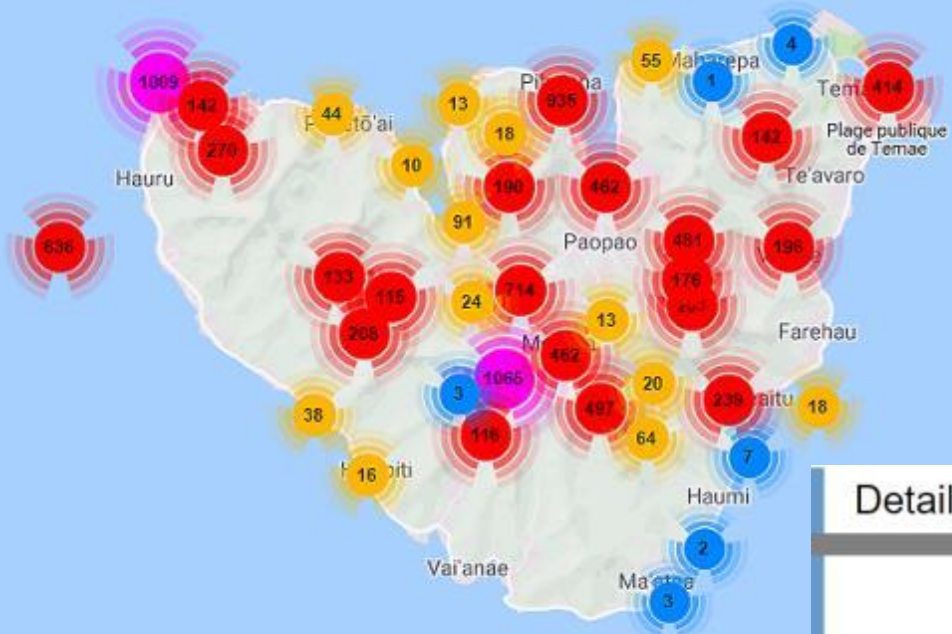
# Moorea BioCode Project



**S 17.536402, W 149.832915**







>18,000 terrestrial arthropods

### Detail of Specimen No. PTO-1048.2

[Back to: Moorea Biocode Collections](#)

**Scientific Name** *Imma*  
**Specimen No.** PTO-1048.2  
**Biocode No.** MBIO37702  
**Institution** Essig Museum of Entomology  
**Biocode Event ID** [7338](#)  
**Collector's Event ID** PTO-1048

**Common Name** moth  
**Collected by** [Peter T. Oboyski](#) (May 1, 2010)  
**Location** central ridge (Moorea Island, Society Islands, French Polynesia)

**Morpho Species Description** *Imma* sp. (dark yellow)  
**Identified by** Peter T. Oboyski  
**GUID** <http://n2t.net/ark:/21547/R2MBIO37702> (Permanent reference for this specimen)



**Phylum** Arthropoda  
**Subphylum** Hexapoda  
**Class** Insecta  
**Subclass** Neoptera  
**Infraclass** Endopterygota  
**Superorder** Amphiesmenoptera  
**Order** Lepidoptera  
**Suborder** Glossata  
**Infraorder** Neolepidoptera  
**Superfamily** Immoidea  
**Family** Immidae  
**Genus** *Imma*

**Sex/Caste**  
**Life Stage**  
**Parts**  
**Length**  
**Weight**  
**Individual Count**

PTO-1048.02

Moorea



# Native vs. Non-Native Species

- ◆ Can we predict whether a specimen belongs to a native or non-native species based on its COI signature?
- ◆ Does community composition change over elevation?
  - ◆ New species generally arrive at low elevations
  - ◆ Low elevation habitats more altered than high elevations
- ◆ Higher elevations communities dominated by endemics

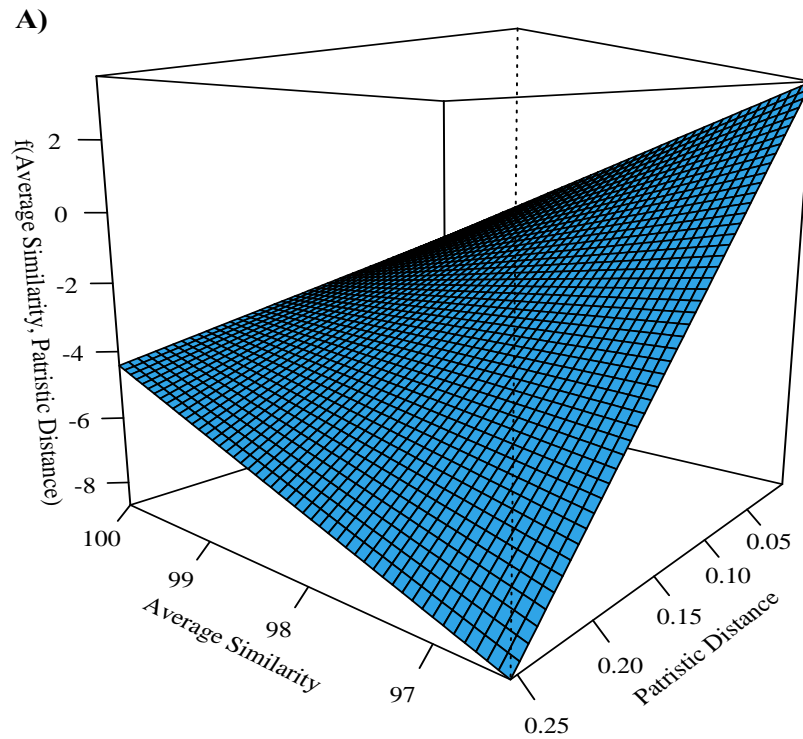
# GLMM Modelling

- ◆ Morphological identifications done by available taxonomists
- ◆ BLAST searches identified similar COI haplotypes
- ◆ Left with thousands of unidentified specimens
- ◆ Calculated within group diversity & between group distances
- ◆ Generated a Generalized Linear Mixed Model to assign non-native and endemic species using identified specimens
- ◆ Applied model to unidentified specimens to categorize as likely non-native or endemic species

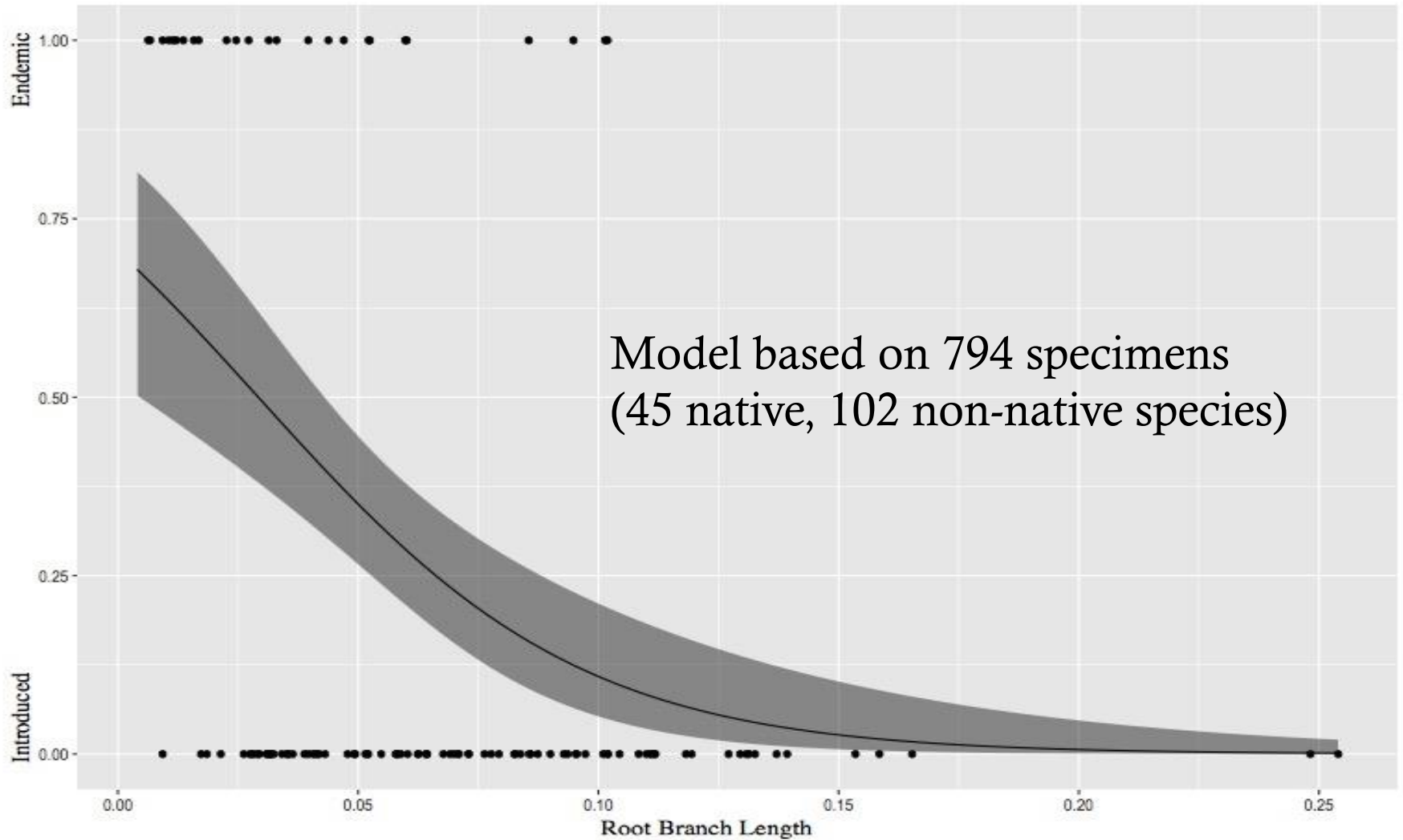


# Parameterizing

Classification  $\sim$  Average Similarity + Average Distance + (1 | Order)

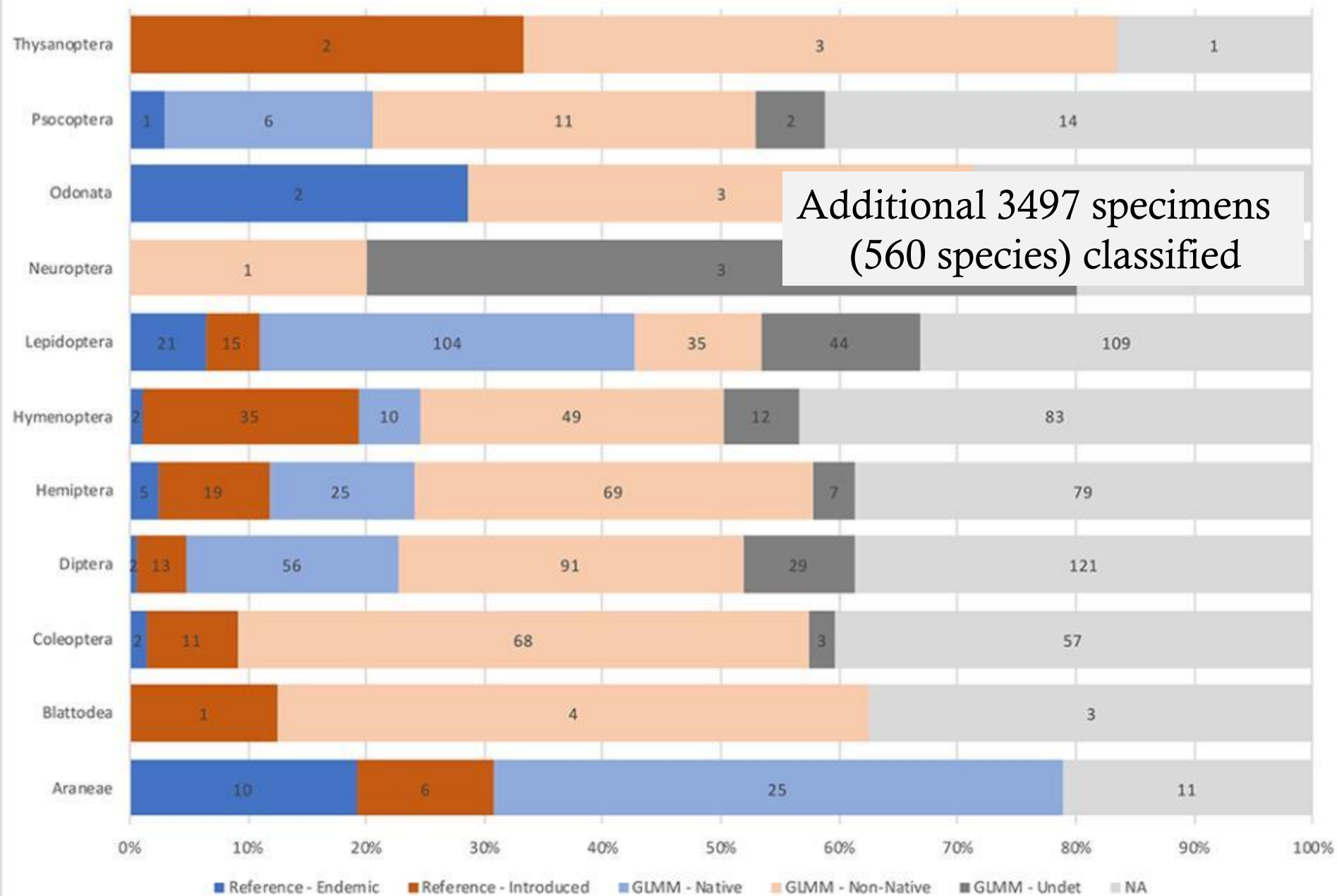


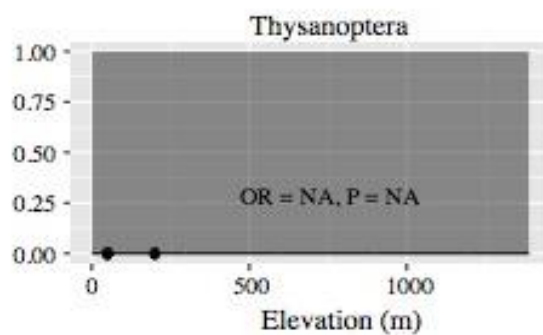
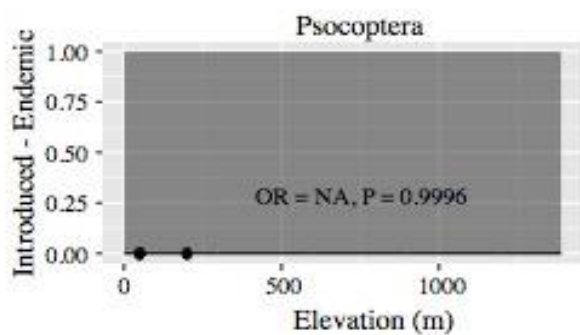
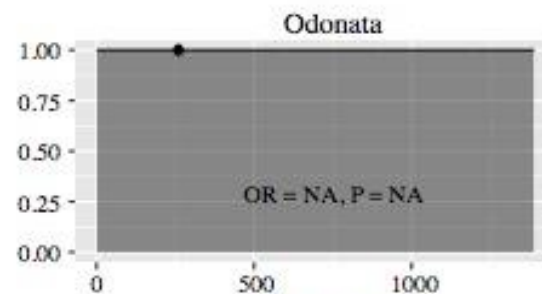
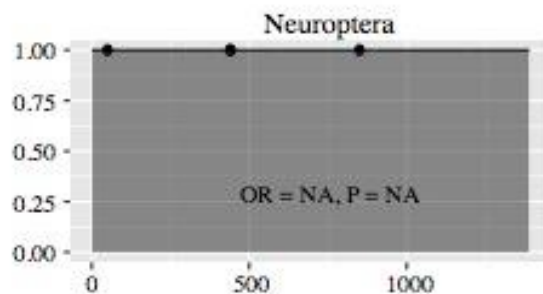
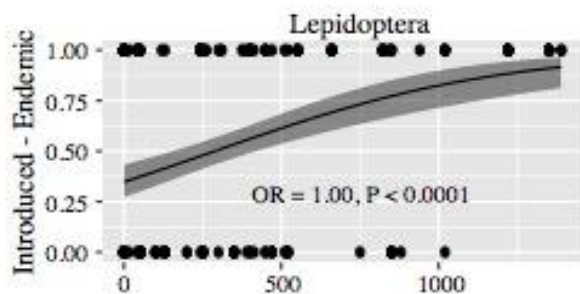
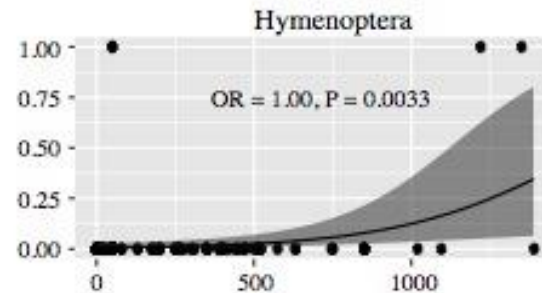
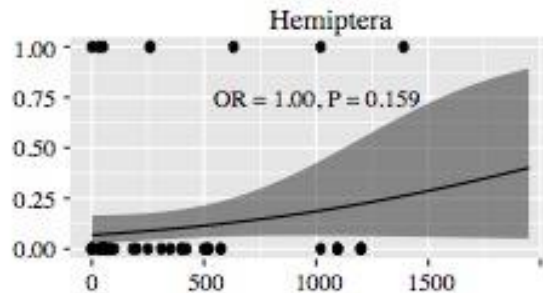
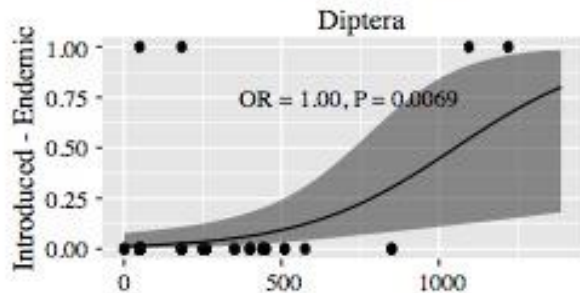
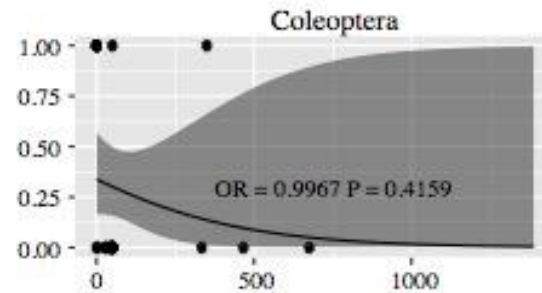
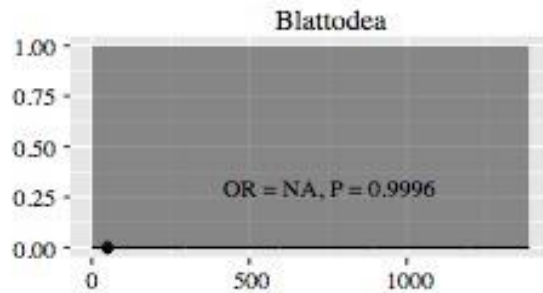
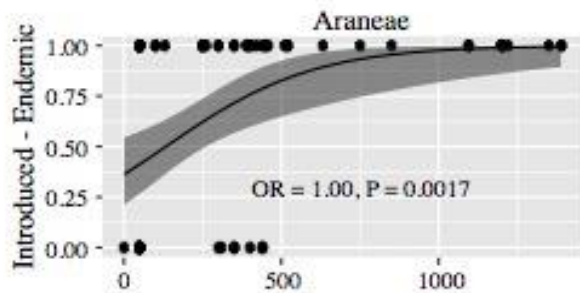
# Results





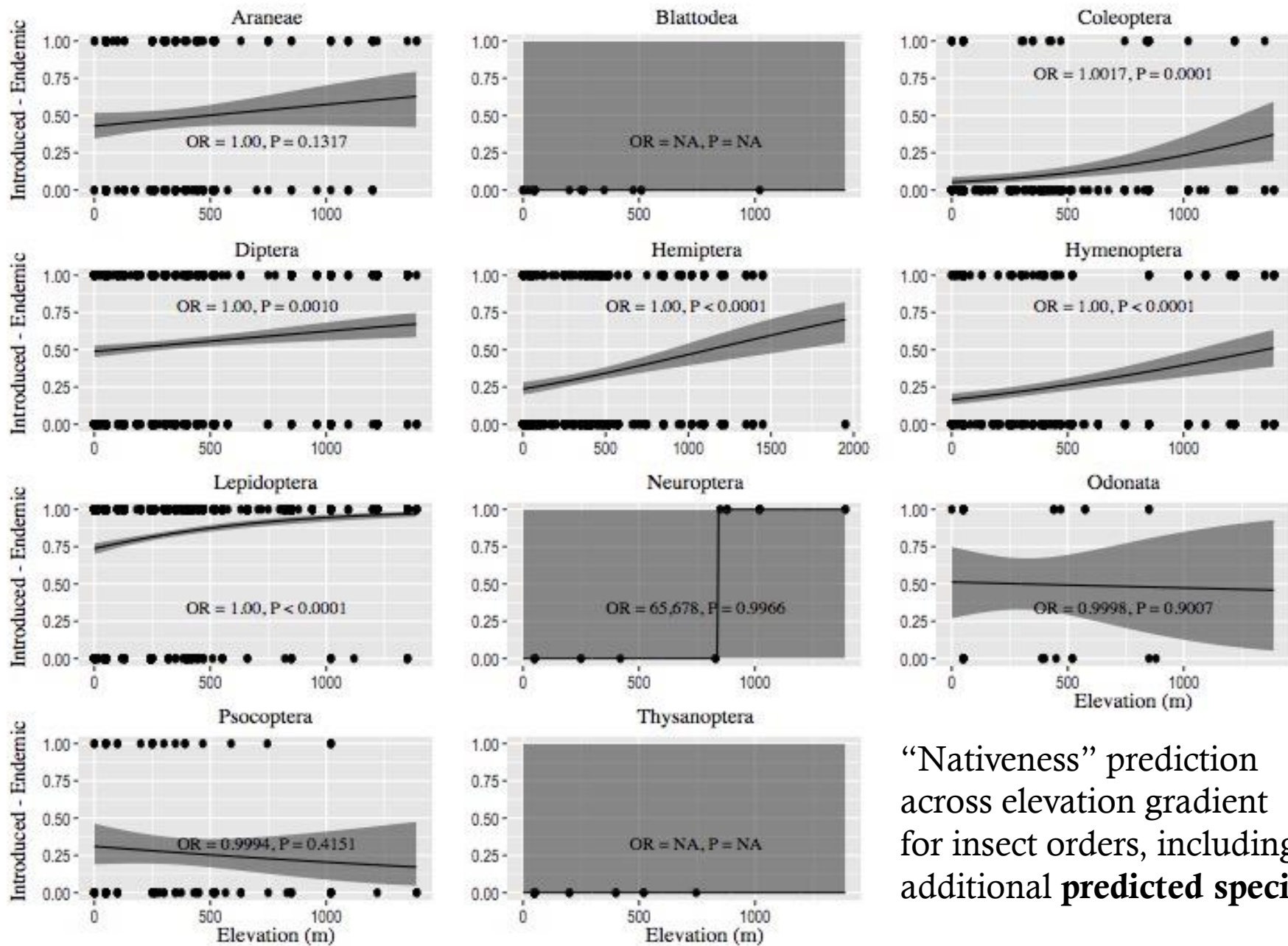
## BioCode Classifications Based on Reference Library and GLMM Classifications





“Nativeness” prediction  
across elevation gradient  
for insect orders, based on  
**species-level identifications**





“Nativeness” prediction across elevation gradient for insect orders, including additional predicted species.

# Limitations

- ◆ Requires a large data set with multiple individuals of each species to generate molecular diversity metrics.
- ◆ Most applicable to “islands” or other isolated regions with more distinct DNA signatures for native and non-native.
- ◆ Distance and similarity metrics fall along a continuum, which probably accounts for many “unassignable” species

# Conclusions

- ◆ Given a large enough, diversified data set representative of the geographic region of interest, molecular diversity metrics can provide a first approximation of native versus non-native species for unidentified specimens
- ◆ Useful for prioritizing identifications for invasive species, public health concerns, and conservation