Biol Invasions (2025) 27:50
https://doi.org/10.1007/s10530-024-03502-3

ORIGINAL PAPER

The known unknowns in international border interceptions of non-native insects

Rebecca M. Turner · Andrew M. Liebhold · Helen F. Nahrung · Craig B. Phillips · Takehiko Yamanaka · Eckehard G. Brockerhoff

Received: 3 September 2024 / Accepted: 26 November 2024 / Published online: 23 December 2024

Abstract Propagule pressure is one of the key drivers of establishment success of non-native species, including insects. However, border interception data, which have the potential to act as a proxy for true arrival rates (i.e., propagule pressure) of insects into a country, are seldom used to predict invasions. This can be due to the limited amount of interception data in some countries, difficulties accessing such data, and when these data are available, difficulties in addressing biases caused by variation in interception probability for different taxa due to policy changes, and operational influences of import inspections. The type of interception data required to reliably estimate arrival rates is rarely available. To improve the use of interception data as a proxy for propagule pressure, we investigated the fraction of established species which had interceptions and vice versa by taxonomic group and by biological characteristics (development type and feeding group), using several national datasets from five continents. We identified higher fractions of established species that were intercepted for plant feeding insect groups compared to fractions of non-plant feeding groups, even in countries with more general import inspection strategies. This is likely to reflect greater search effort for and recording

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10530-024-03502-3.

R. M. Turner
Scion (New Zealand Forest Research Institute), 10 Kyle Street, Riccarton, Christchurch, New Zealand
e-mail: Rebecca.Turner@Scionresearch.com

R. M. Turner · C. B. Phillips
Better Border Biosecurity (B3), Lincoln, New Zealand

A. M. Liebhold
USDA Forest Service Northern Research Station, Morgantown, WV, USA

A. M. Liebhold
Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Praha 6, Suchdol, Czech Republic

H. F. Nahrung
Forest Research Institute, University of the Sunshine Coast, Sippy Downs, QLD, Australia

C. B. Phillips
AgResearch, Lincoln, New Zealand

T. Yamanaka
Research Center for Agricultural Information Technology, NARO, Tsukuba, Japan

E. G. Brockerhoff (✉)
Swiss Federal Research Institute WSL, Birmensdorf, Switzerland
e-mail: eckehard.brockerhoff@wsl.ch

E. G. Brockerhoff
School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

of plant-feeding insect groups during inspections. To address this variation in interception probability and improve establishment predictions based on interception frequency, we developed a methodology to account for such taxonomic variation. We apply this to three hazard lists of insects, containing potential pests for an industry or potential pests for a country, to assess its effectiveness.

Keywords Biological invasions · Biosecurity · Establishment · Invasion ecology · Non-native species · Plant-feeding

Introduction

Given the extensive ecological and economic impacts that biological invasions can have, the most efficient approach to managing invasions is to prevent the introduction of non-native species in new countries or regions (Leung et al. 2002). National governments operate biosecurity programmes aimed at preventing the arrival of new and potentially damaging unwanted organisms while minimizing negative impacts on trade (Hulme 2010). Risk assessment is a crucial element of biosecurity, as it involves first identifying the likelihood of species arriving and the magnitude of their potential impacts (Montibeller et al. 2020). Certain pest taxa are predisposed to invading (Liebhold et al. 2021; 2016; Mally et al. 2022), and insect species that have already invaded one part of the world, for instance, are generally more likely to invade others (Worner and Gevrey 2006). However, predicting which insect species will invade new regions remains difficult due to their wide diversity and the millions of species that exist globally (Stork 2018).

An important resource for predicting invasions is border interception data, which can potentially be used as a proxy for arrival rates (Brockerhoff et al. 2014; McCullough et al. 2006; Turner et al. 2021). Border interception data are collected at ports of entry in many countries around the world and provide evidence for insects or other organisms associated with particular import and passenger entry pathways. Such evidence is used in biosecurity risk analyses at the pathway or individual species or taxonomic group level. This is then used to improve border biosecurity procedures (including pre-border measures, phytosanitary treatments, prohibitions, etc.) to reduce the risks associated with those pathways, or from those particular species.

However, border inspections typically do not represent a statistically based sample of all pathways (although the USDA's Agricultural Quarantine Inspection Monitoring (AQIM) program is based on statistically-based risk assessment as per US Department of Agriculture Animal and Plant Health Inspection Service [USDA APHIS] (2021)), or of all species arriving in a country; there is not an equal detection likelihood across all taxa, nor even an estimated likelihood of detection in most cases (e.g., based on slippage surveys as per Kean et al. (2008) and Ministry of Agriculture and Forestry (2003)). This complicates the use of interception frequency to represent the relative propagule pressure of different species. Most border interception data are difficult to work with, partly due to data access restrictions related to trade sensitivity, but also because they are collected under a set of priorities and operational constraints unique to each country and time period rather than with a systematic design for analysis of arrival rates (Saccaggi et al. 2016). Despite this, interception data are of value if we can understand their limitations—the known unknowns (e.g. types of species which we know will arrive but are less likely to be intercepted or identified)—and incorporate this into analysis and expressions of uncertainty.

Inferences about relative arrival rates and establishment probabilities have been attempted using interception data from several countries over the past couple of decades, including Australia (Caley et al. 2015; Nahrung and Carnegie 2021; Suhr et al. 2019), New Zealand (Bertelsmeier et al. 2018; Brockerhoff et al. 2006, 2014; Phillips et al. 2018), South Korea (Lee et al. 2016), USA (Bertelsmeier et al. 2018; Haack 2006; Jenkins et al. 2014; Liebhold et al. 2006; McCullough et al. 2006; Work et al. 2005)), Europe (Bacon et al. 2012; Eschen et al. 2015b) and South Africa (Saccaggi et al. 2021). Several of these studies generally show some of the idiosyncratic limitations of data collected in each country. Caley et al. (2015), for instance, found a positive correlation between the frequency of interceptions of species and the number of recorded insect incursions in Australia, but the overall number of intercepted species which also had incursions remained too low for their dataset to be effective for predicting future invasions. To counteract the disadvantages of sparse data and individual

country biases, some studies have used interception data from multiple countries (Brockerhoff et al. 2006; Turner et al. 2020). Focusing on particular taxa or groups rather than all insects can also help control variation in interception probability per arrival for different insect taxa or guilds. This is a key aspect when interpreting interception data because taxa vary greatly in their detectability in imports. For example, interceptions of bark and woodboring beetles in infested timber typically involve galleries which often contain some adult beetles (especially in the case of bark beetles), facilitating their identification based on morphological characteristics. By contrast, interceptions of moths and flies typically involve eggs or larvae, life stages that are difficult or impossible to identify without using molecular diagnostics.

Variation in detection probability among taxa is a key limitation of border interception data. For example, the diversity of plant pathogens (e.g., fungi, bacteria, viruses, or oomycetes) present in pathways such as soil on shoes, shipping containers and machinery (McNeill et al. 2011, 2023) is likely under-reported because the sophisticated methods needed to detect and diagnose them are usually targeted at traded plant material and pathogens of quarantine concern (Eschen et al. 2015a). Moreover, infected plants may be difficult to recognise, especially when plants are asymptomatic (Migliorini et al. 2015). Thus, the coverage of pathogens is unlikely to improve until molecular diagnostics can be applied to a greater diversity of pathways and pathogen species. In the present study, we focus on variation in coverage within insects. Among Lepidoptera, Microlepidopteran species are less likely to be intercepted internationally than Macrolepidoptera compared to members of these groups which have established historically (Mally et al. 2022). In other words, we expect fewer Microlepidoptera interceptions relative to Macrolepidoptera interceptions given equal propagule pressure. This might be because of the typically smaller size of microlepidoptera, or, alternatively (and not necessarily mutually exclusively), Microlepidoptera could be more 'successful' in establishing non-native populations given equal propagule pressure.

While interception frequency of species intercepted at least once has been found to be positively associated with the likelihood of establishment (Brockerhoff et al. 2006, 2014; Caley et al. 2015), there are still many species which have established without ever having been intercepted knowingly (e.g., Caley et al. 2015; Navia et al. 2010). Furthermore, lists of potentially high-risk species maintained by biosecurity agencies or other organisations may include many species which have neither been intercepted nor established. Therefore, it is useful if models used for predicting establishment can statistically account for the species with zero interceptions, such as the models explored in Turner et al. (2020). Examples of such lists, include the list of insects known to feed on *Pinus radiata* (Brockerhoff et al. 2023), a tree native to North America but widely planted especially in the southern hemisphere, or lists of organisms which countries target for exclusion, such as the regulated insect species from the Pest Register for New Zealand importers (https://pierpestregister.mpi.govt.nz/pest-register-importing/) and the US Dept. of Agriculture's Regulated Pest list (https://www.aphis.usda.gov/plant-imports/regulated-pest-list). Industry and government agencies use these species lists for biosecurity planning, by considering their potential impacts and likelihood of establishment.

In this study, we compiled a large dataset of interception records from multiple countries across five continents to explore how the relationship between interception frequency and establishment varies among different insect taxa. The specific goals are to: (i) describe the variation in coverage of interceptions among taxa, including testing for differences between plant-feeding and non-plant-feeding species; and (ii) assess whether accounting for taxonomic variation can improve predictions of established insect species, including those that have historically never been intercepted. Such models can provide more robust predictions of establishment probabilities, which can be applied to improve biosecurity risk assessments.

Materials and methods

Interception, establishment, and hazard datasets

Border interceptions

Insect border interception data from the following reporter regions: Australia, Canada, European and Mediterranean Plant Protection Organization (EPPO), Hawaii, Japan, New Zealand, Okinawa, South Africa, South Korea, the UK and the USA were compiled

from sources given in Table S1, most of which are described in more detail in Turner et al. (2021). It is important to note that usually only a small fraction of imported cargo or passenger baggage is inspected, and apart from specimens kept for diagnostics, intercepted unwanted organisms are exterminated during treatment or destruction of infested imports at ports of entry. Treatments applied to expedite biosecurity clearance can bypass the need for diagnostic results, leaving many infesting organisms remaining unidentified and unrecorded in interception databases. Consequently, interceptions are not direct measures of insect entry rates but instead represent samples from specific pathways. Each country or region implements different border inspection protocols such that there is variation in the proportion of pathways (air cargo, sea cargo, vessels, international mail, international passenger baggage) and types of commodities inspected. Moreover, criteria differ for reporting organisms and the taxonomic level used in identifications, with some jurisdictions only recording species only and others also documenting interceptions at higher taxonomic levels. We did not account for this variation because such information is often missing and may change over time, so all interceptions were included in the analyses. Given the variation in how inspections and interceptions are handled and recorded among countries, biases likely exist in the datasets; however, we combined data from multiple countries in the following analysis to minimize the effect of biases from any one country.

Establishments of non-native species

Comprehensive lists of all non-native insect species known to be established in 11 different world regions (North America, Hawaii, New Zealand, Australia, South Korea, Japan, Ogasawara, Okinawa, Great Britain, Europe, South Africa) were collected from the sources in Table S2. Miscellaneous updates were also included from various sources (Turner et al. 2024a). These establishment lists covered all insects and largely coincided with the countries for which we had interception datasets. Establishment records of species only known to occur indoors (e.g., in greenhouses) or intentionally released (e.g., biological control agents) were excluded. Records of a temporary occurrence of a species followed by an active eradication were not removed from the establishment dataset since they show potential for an intercepted species to breach a border and become established (were it not for the successful eradication). These lists also included the year of first record, when available, for each species-region combination.

Case study hazard species lists

We used three case study lists of potential hazard species to test the use of border interceptions for predicting establishment.

The first is a list of insect species feeding on *Pinus radiata* globally (Brockerhoff et al. 2023) which contains species with all combinations of intercepted, not intercepted, established internationally or not established. Thirty percent of the 649 species were present in the international interception dataset, and twenty percent were present in the international establishment dataset. Most species on this list are Coleoptera and Lepidoptera. Due to the criteria used to compile the list, they are also all plant feeders.

The second case study hazard list is a list of regulated insect species from the Official Pest Register for New Zealand importers, maintained by the NZ Ministry for Primary Industries (https://pierpestregister.mpi.govt.nz/pest-register-importing/). This list was download on the 19th of October 2023 to include all results returned using the search term "insect" in the "organism type" data field. The retrieved results were then filtered to include only "regulated" organisms, and those identified at the species level (i.e., all entries at genus level were removed and subspecies were replaced by their corresponding species). This returned a final list of 7945 species, of which 25% were present in the international interception dataset, and 18% in the international establishment dataset. Most insect species on this list belong to Hemiptera, followed by Lepidoptera and Coleoptera. Not all species are plant feeders, with some expected to cause human health, social, cultural or environmental impacts.

The third case study hazard list is from the US Regulated Plant Pest table, maintained by the US Department of Agriculture (USDA) (https://www.aphis.usda.gov/plant-imports/regulated-pest-list). This list was downloaded on the 8th of May 2024, filtered to insects, then only to those entries identified to species. This returned a final list of 2733 species, of which 78% were present in the international

interception dataset, and 18% in the international establishment dataset. Most insect species on this list belong to Hemiptera, followed by Coleoptera and Lepidoptera, and including mainly plant-feeding species.

Taxonomy and plant feeding categories

To compare species data collected from multiple sources, which may contain spelling errors or different synonyms for the same species, we applied a taxonomic cleaning process using the R taxize package (Chamberlain and Szöcs 2013). The Global Biodiversity Information Facility (GBIF) taxonomic backbone was used as the primary taxonomic standard, which accounted for approximately 90% of the names in our datasets, including fixing spelling mistakes and harmonising synonyms. The taxonomy for all datasets was updated to the GBIF taxonomic backbone GBIF (GBIF Secretariat 2021). Family structures for Coleoptera and Lepidoptera are as per Liebhold et al. (2021) and Mally et al. (2022) respectively.

In the combined international dataset, a species was counted as intercepted internationally if it was intercepted at least once regardless of the region, and a species was counted as established if it was established at least once regardless of the region (the regions in which a species was intercepted or established were not always the same).

Each family with at least 25 established species in the combined international dataset from the Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera and Thysanoptera was categorized as plant feeding or non-plant feeding based on the predominant feeding form within the family. A list of the plant feeding categorizations is provided in Turner et al. (2024b). Likewise, families were categorized as having holometabolous or hemimetabolous development.

Descriptive statistics

All analyses and data visualizations were conducted in R version 4.1.2 (Core Team 2021).

We compared the average year of first establishment record between insect orders. The analysis was restricted to orders with at least 20 dated first records of establishment, and since the variances differed significantly between orders, we used the Welch one-way test followed by the Games-Howell test (Games and Howell 1976).

For each insect order, we calculated the expected overlap of intercepted and established species to compare with the observed overlap assuming that the probabilities of a species being intercepted and of being established were independent. To do this, we first approximated the probability of a species being intercepted or established as the observed number of species intercepted or established divided by the total number of described species per order (from Zhang 2013). Then, under the assumption of independence, we multiplied those probabilities to obtain the probability of a species being both intercepted and established. Finally, this probability was multiplied by the described number of species per order to obtain the expected number of species both intercepted and established (Eq. 1).

$$E(\text{overlap}) = P(\text{interception})P(\text{establishment})(\text{number of described species in order}) \quad (1)$$

For the remaining statistics and figures in which proportions of intercepted species established were calculated, we removed the species known to be intentionally introduced in all regions where they occur from the interception dataset. Conversely, if a species was intentionally established in one region but unintentionally established in another, then it was left in the international border interception dataset.

For species from families with at least 25 species established and belonging to the top six orders (with the highest species richness in the interception dataset), we tested whether the median percentage of established species intercepted per family was higher for plant feeding compared to non-plant feeding families using Mood's median test with Fisher's exact test. A similar procedure was used for comparing between hemimetabolous

and holometabolous families. We also compared the proportions of established species intercepted between groups, using Pearson's Chi-squared test. We applied the same procedures to test the effects of both feeding modes and ontogeny on the reverse—the median percentages of intercepted species established per family, and the proportion of intercepted species established.

Establishment prediction models

Four models were compared for predicting species probabilities of establishments (i.e., presence in the international non-native species establishment dataset) based on their interception frequency in the combined international dataset. Species which were only intentionally introduced were excluded from both the interception and establishment datasets for this analysis.

Logistic regression species interception frequency only

Logistic regression models (Eq. 2) using the total species interception frequency were fitted to the training dataset of species intercepted at least once.

$$p(y = 1|x) = \frac{e^{a+bx}}{1 + e^{a+bx}} \tag{2}$$

In this model, y, is the binary response of whether or not the species has established in the international dataset, x is the interception frequency, and a and b are the two fitted parameters.

Logistic regression including family or taxon as a factor

Logistic regression models using the total species interception frequency were fitted to the dataset of species intercepted at least once including family or taxon as a factor in the model.

$$p(x) = \frac{e^{a+b_1x_1+\ldots+b_ix_i}}{1 + e^{a+b_1x_1+\ldots+b_ix_i}} \tag{3}$$

Uniform model

We used the uniform model (Eq. 3) for predicting establishment probability ($p(x)$) as a function of total international interception frequency (x) as per Turner et al. (2020) fitted for each family or taxa group using the data for species intercepted at least once:

$$p(x) = 1 - \left(\frac{p_I}{p_I + p_E(1 - p_I)}\right)^{x+1} \tag{4}$$

Turner et al. (2020) presented two possible process-based models, but for the purposes of fitting establishment models across multiple taxa we chose the uniform model as it is simpler to fit. The model assumes a uniform distribution of arrival rates. This model has two parameters, p_I (probability of an arrival event being intercepted) and p_E (the per arrival event probability of establishment, i.e., the conditional probability of establishment given a non-intercepted arrival event). However, it is difficult to fit both parameters simultaneously as they are effectively proportional. Hence, we chose to fix $p_I = 0.001$ while fitting p_E separately for each taxonomic group.

Ratio model

This model assumes that the effect of the relative interception probability versus the establishment probability per family or taxon can be approximated by normalizing the species interception frequency by the ratio of species richness per family or taxon in each dataset. I.e., the species interception frequency was multiplied by the ratio of establishment species richness to intercepted species richness (r_i) for each family or taxon to account for relative establishment probability. This is effectively the logistic regression model (Eq. 2) but with normalized species interception frequencies.

$$p(x) = \frac{e^{a+br_ix}}{1 + e^{a+br_ix}} \tag{5}$$

Each of the four models were applied in each of five scenarios (see Table 1). Two are based on intercepted species in species rich families or taxa, three are based on hazard lists.

Scenarios 1 and 2 represent the idealized situation where all species of interest are intercepted

Table 1 Description of the five modelling scenarios

Scenario	Training dataset	Test dataset
1	Randomly selected 75% of the species intercepted at least once in the top families (75% of 6756 species)	The remaining 25% of the species intercepted at least once in the top families (at least 20 established species, and at least 10 intercepted species, excluding the Chironomidae as there was no overlap between intercepted and established species). (25% of 6756 species)
2	Randomly selected 75% of the species intercepted at least once in the top taxa. (75% of 8483 species)	The remaining 25% of the species intercepted at least once in the top taxa (25% of 8483 species)
3	All of the species intercepted at least once in the top taxa (8483[a] species)	The species in the *Pinus radiata* hazard list (regardless of whether they were intercepted, 649 species)
4	All of the species intercepted at least once in the top taxa (9009[a] species)	The insect species in the NZ Regulated organisms hazard list (regardless of whether they were intercepted), 7874 species
5	All of the species intercepted at least once in the top taxa (8741[a] species)	The insect species in the USDA hazard list (regardless of whether they were intercepted), 2733 species

[a]The training dataset top taxa numbers here differ for each of the hazard lists (scenarios 3, 4 and 5) as the hazard lists contain species in different sets of families. If a hazard list species is in a family which is too small by itself (i.e. did not meet the criteria of at least 20 established species and at least 10 intercepted species in the international datasets) then the family was merged up to a higher level taxon (e.g. superfamily or order) to meet these criteria

at least once. The higher-level taxonomic groupings used in scenario 2 (referred to as "taxa") were selected such that insect species on the *Pinus radiata* hazard lists were within taxa containing at least 10 intercepted species and at least 20 established species. A list of the taxa used (e.g., superfamily, order) are provided in Turner et al. (2024b). Species were randomly allotted into the training and test datasets, and subsequent training and testing of the model was repeated 10 times.

Scenarios 3, 4, and 5 represent more realistic situations where the list of species of interest, (i.e., our three case study hazard datasets), include many species with zero interceptions. These hazard datasets (the *Pinus radiata* hazard list, the NZ-regulated insects hazard list, and the US-regulated plant pest hazard list respectively) contain diverse species, and not just those from the most species-rich families. Hence, species on the list from smaller families were grouped into higher taxa (e.g., superfamily, order) such that each taxon contained at least 20 established species and at least 10 intercepted species in our interception and establishment datasets. The scenario 3 taxa were as per scenario 2 and the scenario 4 and 5 taxa were slightly different in order to encompass the broader range of species on the regulated insects hazard lists (both taxa lists are provided in Turner et al. (2024b); note that Neuroptera, Zygentoma, and Trichoptera species were excluded from the NZ regulated hazard list as data for these taxa were too few for fitting). In scenarios 3, 4, and 5, the models were fit to the species intercepted at least once in the taxa with at least 20 established species and at least 10 intercepted species (i.e., as per scenario 2). Each model was then tested on the relevant hazard dataset which included the species intercepted zero times (i.e., both the training and test datasets contain some of the species intercepted at least once).

Specifically for the ratio model, in scenarios 1 and 2, the species richness per family or taxon was calculated from samples of 75% of the full interception dataset and samples of 75% of the full establishment dataset rather than the subset of species intercepted at least once. Subsequently in scenarios 1 and 2, the ratio model was tested on the remaining 25% of the full interception dataset. In scenarios 3, 4 and 5, species richness was calculated from the full interception dataset and full establishment dataset, then applied to the full hazard dataset.

For each model test, a Receiver Operating Characteristic curve (ROC) was plotted to visualize the performance of each model for balancing sensitivity and specificity. Summary statistic Area Under the Curve ROC (AUC-ROC) was also calculated for each test. A high AUC-ROC, i.e. close to one, indicates good model performance while an AUC-ROC close to 0.5 means

that the model is no better than random at distinguishing between established and non-established species.

Results

Descriptive statistics

There were 14,136 recorded species in the combined interception and establishment datasets. About 17% of these records, 2398 species, were both established and intercepted, while 6782 species were intercepted and not established, and 4956 species were established but not intercepted. There were overall similar numbers of all taxa recorded as intercepted or established (Table 2), however, these two sets only overlapped slightly at the lower taxonomic levels (species and genera).

A third (33%) of known established species also had interceptions. For higher taxonomic level taxa, regardless of whether detections were identified to species, 59% of genera with establishments had interceptions, increasing to 91% at family level, and 95% at order level. However, if only interceptions identified to species level were included, these overlaps dropped to 47% of the recorded established genera, 72% of established families, and 82% of established orders.

Order level comparison

The establishment data included records dating back several centuries (Table S2). The six orders with the most established species were, in descending order, Coleoptera, Hemiptera, Hymenoptera, Diptera, Lepidoptera, and Psocodea. The interception datasets were collected since the 1990s (Table S1). The six orders with the most intercepted species were, in descending order, Coleoptera, Hemiptera, Lepidoptera, Hymenoptera, Diptera, and Thysanoptera (Fig. 1).

The establishment records for Thysanoptera were generally more recent than for the other insect orders

Table 2 Taxonomic richness of border interception and establishment datasets and overlap

Taxonomic level	Number intercepted	Number of taxa with interceptions identified to species level	Number established	Number intercepted and established
Species	9180	9180	7354	2398
Genus	6674	4379	3851	2259
Family	629	386	432	392
Order	25	20	22	21

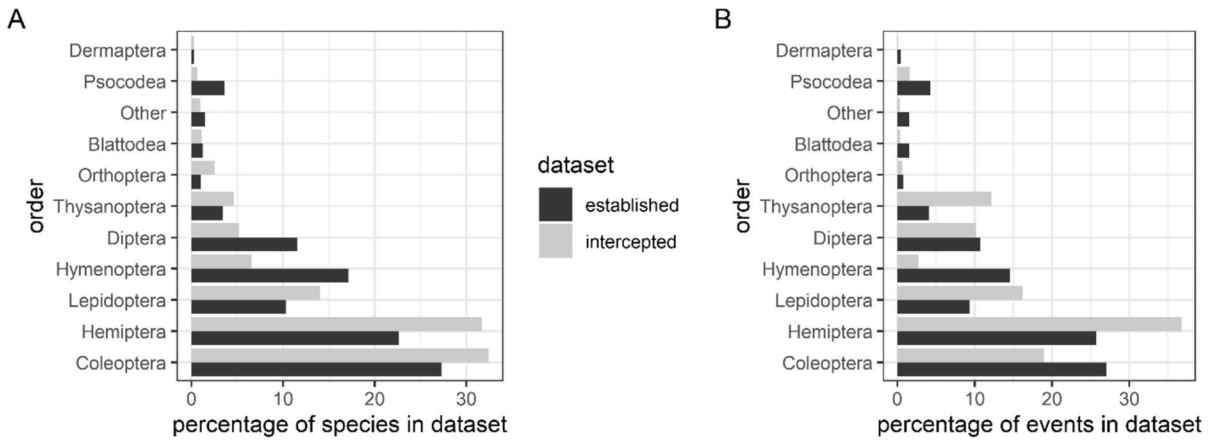


Fig. 1 Comparison of **A** species richness and **B** event frequency within orders considering both border interceptions and establishments. **B** Inclusive of interception events identified only to order, family, or genus level in addition to those identified to species level

(mean year of first record: 1963, significantly higher based on pairwise comparisons using the Games-Howell test, all $P < 0.05$). First records of Hymenoptera (mean year: 1955), Lepidoptera (1953), Hemiptera (1952) and Diptera (1951) were also more recent on average compared to first records for Coleoptera (mean year = 1938, pairwise comparisons in the Games-Howell test all $P < 0.001$). In contrast, Dermaptera (mean year = 1916) and Psocodea (1922) years of first record were significantly earlier than Coleoptera (both $P < 0.05$).

Seven of the 11 national or regional interception datasets also included records at higher taxonomic levels than species (at genus, family or order level). These datasets varied in the percentage of interceptions identified to species level (NZ: 47%, Australia: 40%, Canada 45%, mainland USA 30%, Hawaii 34%, UK: 35%, EPPO 59%, South Africa 49%). The percentage identified to species level varies among insect orders and between regions. Strikingly, with the exception of the EPPO dataset, less than 25% of Diptera interceptions were identified to species level—among the lowest percentage for any of the top six orders (Fig. S2). However, the percentage of interceptions identified to species was not uniformly low for all Diptera families in all reporter regions. Tephritidae, which contains many pests of economic importance, had a higher percentage identified to species level in New Zealand at 64%.

Among the six orders with the highest species richness in border interceptions there was a higher proportion of established species that had been intercepted in Thysanoptera, and Hemiptera, and a lower proportion in the Diptera and Hymenoptera (Fig. 2). Hymenoptera had a significantly lower proportion of established species intercepted than any other order, in addition to having the lowest total interception frequency among the top six orders (inclusive of interception events identified to species, genus, family or order). Of the six main orders, only Thysanoptera had over 50% of its established species intercepted in the international dataset (Figs. 2, S1). However, in addition to Thysanoptera, Coleoptera, Hemiptera, and Lepidoptera had over 50% of their established genera intercepted. Figure S1 shows that while the overlap between intercepted and established species was not large, the overlap was, for all orders, larger than would be expected if the probability of interception and establishment for a species were completely independent (given the number of described species per order).

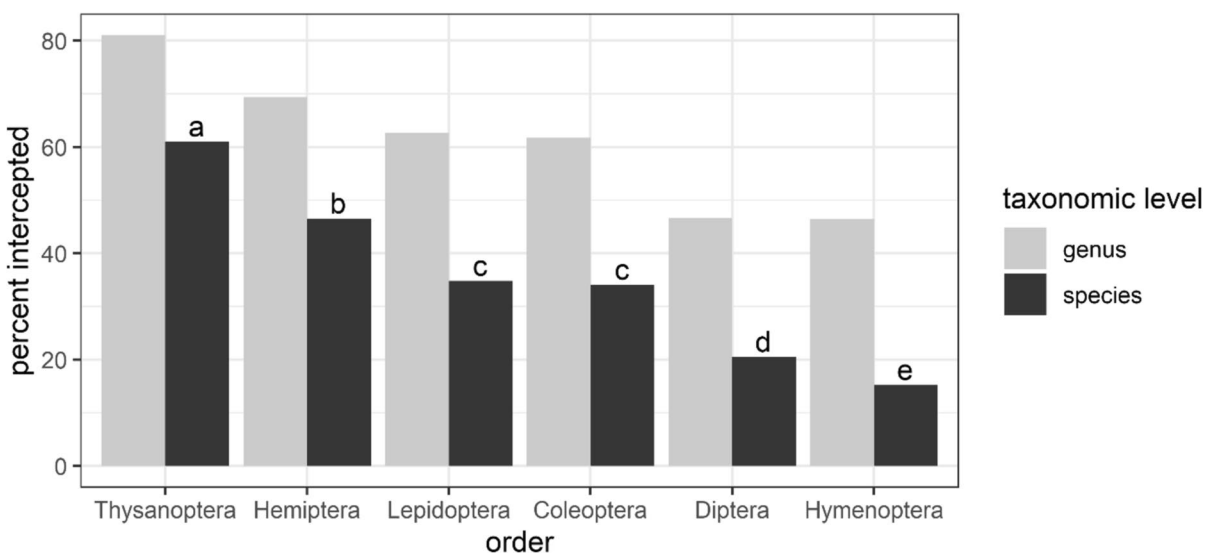
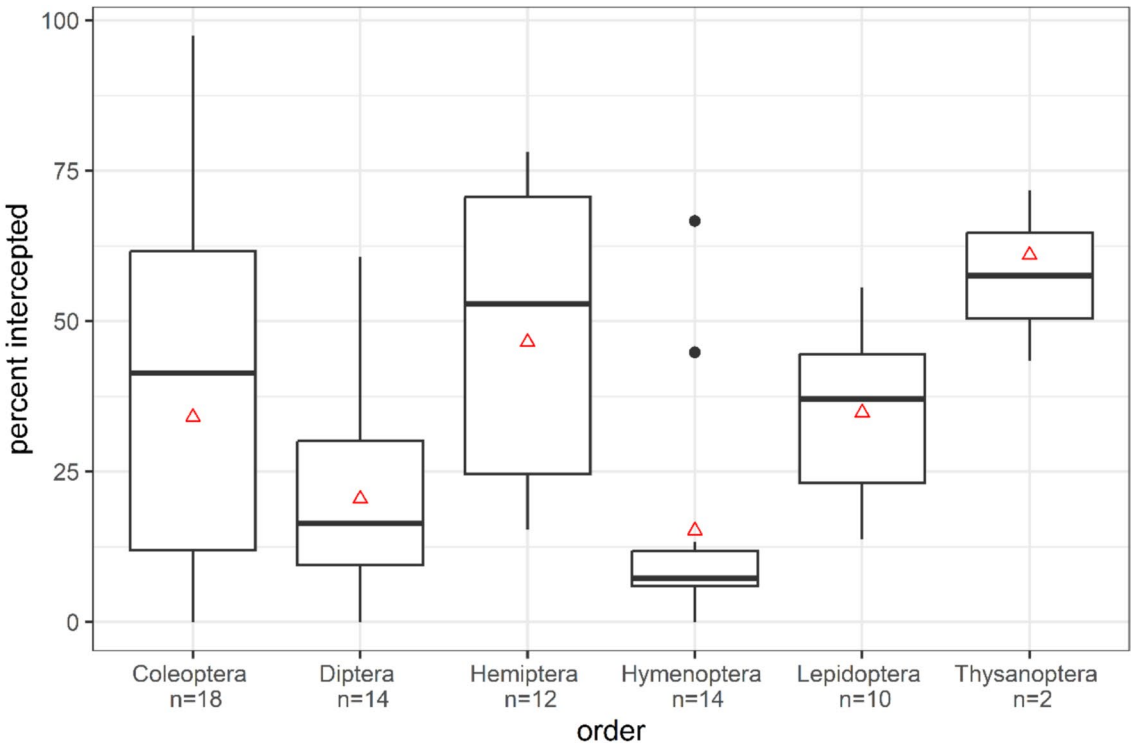
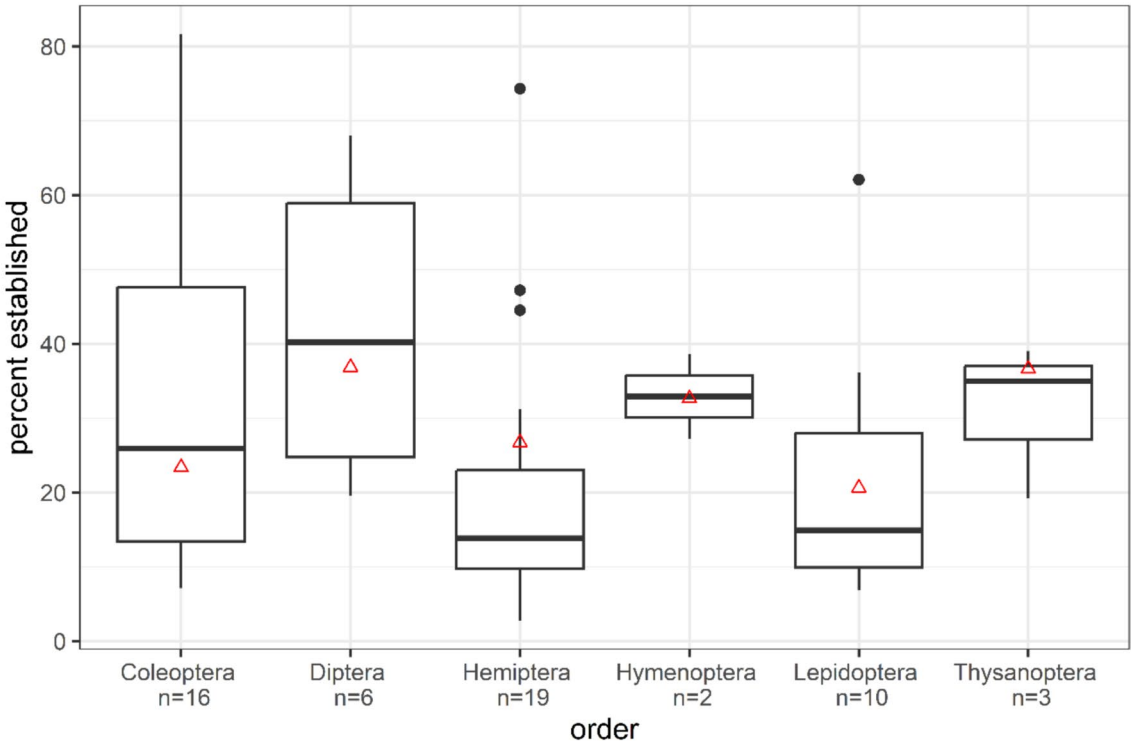


Fig. 2 Percentage of established species or genera which have been intercepted. When calculating the percentage of genera intercepted, the genera of records identified to species level were also included. Letters above species bars indicate which percentages are significantly different (pairwise comparison of proportions, with Holm (1979) method of adjustment for multiple comparisons, $P < 0.01$)


100
75
50
25
0
percent intercepted
Coleoptera
n=18
Diptera
n=14
Hemiptera
n=12
Hymenoptera
n=14
Lepidoptera
n=10
Thysanoptera
n=2
order

A


80
60
40
20
0
percent established
Coleoptera
n=16
Diptera
n=6
Hemiptera
n=19
Hymenoptera
n=2
Lepidoptera
n=10
Thysanoptera
n=3
order

B

◂**Fig. 3** **A** Boxplot showing the variation in the percentage of established species intercepted among families by order (n = families). The families included are those with at least 25 established species, from the top six orders. The two Hymenoptera outliers are Vespidae and Formicidae. Red triangles show the percentage intercepted of all established species in the order (inclusive of species in smaller families). **B** Boxplot showing the variation in the percentage of intercepted species established among families by order (n = families). The families included are those with at least 25 intercepted species, from the top six orders. The three Hemiptera outliers are Aphididae, Coccidae, Diaspididae, the Lepidoptera outlier is Tineidae. Red triangles show the percentage of all intercepted species in the order (inclusive of species in smaller families) which were established. Outliers in boxplots were defined as values above or below the upper and lower quartile boundaries, respectively, by at least 1.5 times the interquartile range

Family level comparison

Within orders, there was further variation in the coverage across families (Fig. 3). Eighteen families with at least 25 established species had at least 50% of their established species intercepted internationally. Bostrichidae had the largest percentage of established species intercepted (97%).

Established plant feeding insects were more likely to be intercepted than established non-plant feeders in all reporter regions (Figs. 4, S3). The median percentage of intercepted species established was significantly higher in plant feeding families (those with at least 25 established species in the top six orders) than non-plant feeding families (39% compared to 7% respectively, Mood's median test $P = 0.0013$). This was even apparent in reporter regions such as Australia which tended towards recording a greater diversity of intercepted species (Pearson's Chi-squared test with Yates' continuity correction: Chi-sq = 31.93, $P < 0.001$).

While the median percentage of established species intercepted was not significantly higher in hemimetabolous families compared to holometabolous families (Mood's median test, $P = 0.37$), the proportion of established hemimetabolous species intercepted was significantly higher than for holometabolous species (Pearson's Chi-squared test with Yates' continuity correction: Chi-sq = 202.84, $P < 0.001$).

The median percentages of intercepted species established per family was not significantly different between feeding types nor between development types (Mood's median test, $P = 0.47$ and $P = 0.17$, respectively). However, the proportion of intercepted species established was higher for non-plant feeding species compared to plant feeding species (Pearson's Chi-squared test with Yates' continuity correction: Chi-sq = 9.86, $P = 0.0008$), and higher for hemimetabolous species compared to holometabolous species (Pearson's Chi-squared test with Yates' continuity correction: Chi-sq = 22.16, $P < 0.001$).

Establishment prediction

If the probabilities of interception are reasonably similar across different species, then border interception frequencies could be used as a proxy for propagule pressure and thereby predict which species are most likely to establish. However, as seen in the previous results, different insect families had substantially different average probabilities of detection at the border *relative* to their establishment likelihood. In the following establishment prediction models, variation among families or other taxa was taken into account, though we could not distinguish whether the variation was due to differences in detection or diagnosis rates at the border, or differences in the probability that an arriving species establishes. A summary of all model results is shown in Table 3.

In the scenarios in which the models were applied to species intercepted at least once (scenario 1 and scenario 2), only the uniform model with family/taxa term and the ratio model improved both the sensitivity and specificity for historical establishments to above 0.7 (Fig. 5) on our test datasets.

In contrast to the application to datasets in which all species were intercepted at least once (scenarios 1 and 2, Fig. 5), the models behaved quite differently when applied to insect lists which also contain species with no interceptions (scenarios 3, 4 and 5, Fig. 6). When applied to our three case study lists, the logistic regression models with a taxa term (fitted to datasets with species intercepted at least once) performed worse than using just total interception frequency as the sole predictor (see also Fig. S4 for an example model fit to several taxa in scenario 4). Neither the uniform nor ratio models offered a large improvement if aiming for balanced sensitivity and specificity. However, if the model's application requires a greater degree of sensitivity (e.g., if the cost of failing to predict a species establishment outweighed the cost of needlessly preparing for additional species which would not establish), then the uniform model was best able to improve the

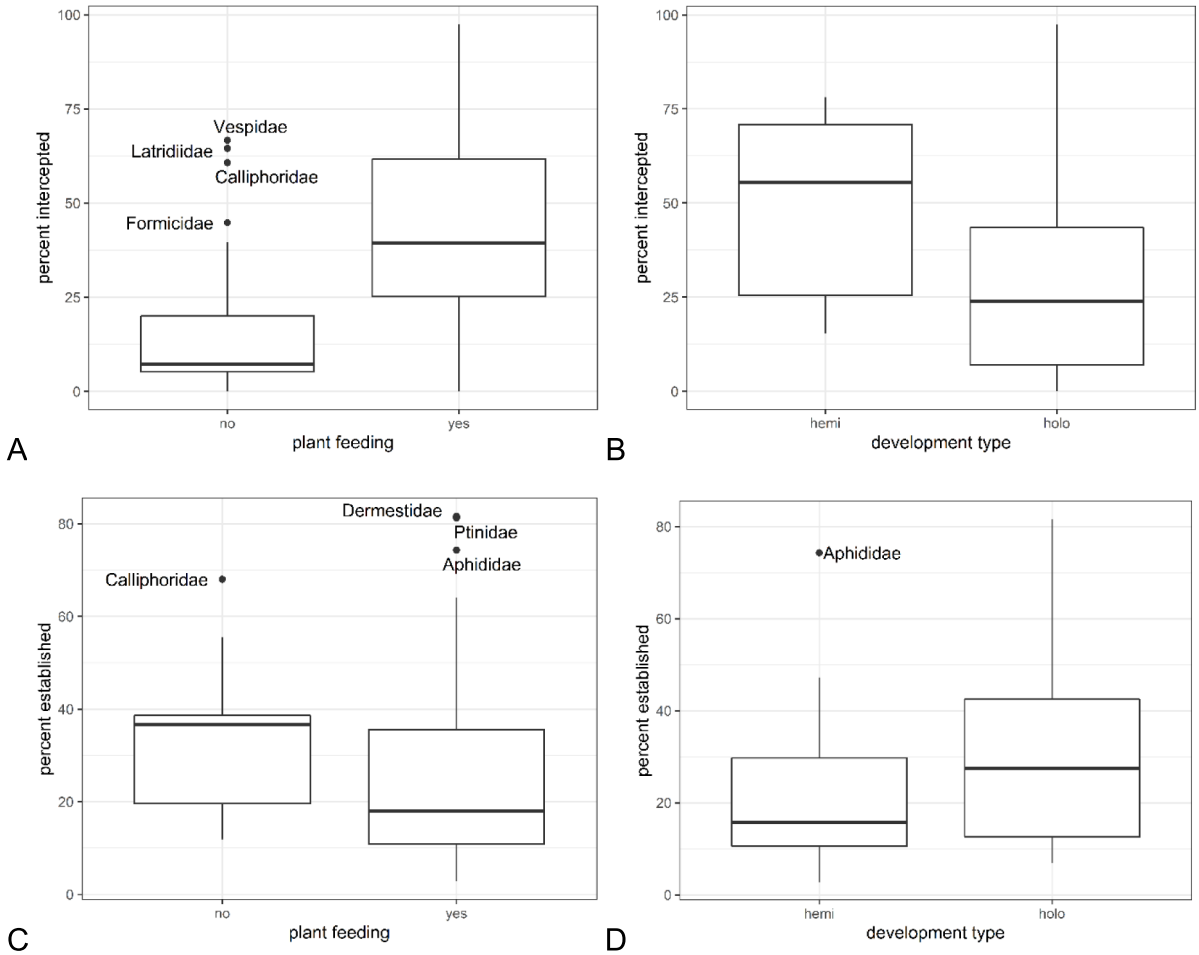


Fig. 4 Boxplots showing the variation in the percentage of established species intercepted among families by **A** feeding group or **B** type of development (hemimetabolous or holometabolous). The families included in **A** and **B** are those with at least 25 established species, from the top six orders. **C** and **D** show the variation in percentage of intercepted species established among families by feeding group or type of development respectively. The families included in **C** and **D** are those with at least 25 intercepted species, from the top six orders. Outliers in boxplots were defined as values above or below the upper and lower quartile boundaries, respectively, by at least 1.5 times the interquartile range

sensitivity while retaining a specificity of more than >0.6. That said, the improvement in sensitivity when the uniform model was applied to the NZ regulated pest list (scenario 4) was much weaker, and if the use case required a greater emphasis on specificity (e.g., aiming for specificity above 0.8), then the uniform model performed worse than the default total interception frequency model.

Table 3 Evaluation of establishment models based on species interception frequency and family or taxon. AUC-ROC is the area under the receiver operator curve

Scenario	Models	AUC-ROC (sd)
1	Logistic regression based on species interception frequency	0.73 (0.0095)
	Logistic regression with **family** term	0.73 (0.0056)
	Uniform model by **family**	0.77 (0.0126)
	Interceptions normalized by establishment/interception species richness ratio by **family**	0.78 (0.0096)
2	Logistic regression based on species interception frequency	0.72 (0.012)
	Logistic regression with **taxa** term	0.73 (0.009)
	Uniform model by **taxa**	0.77 (0.010)
	Interceptions normalized by establishment/interception species richness ratio by **taxa**	0.78 (0.009)
3	Logistic regression based on species interception frequency	0.81
	Logistic regression with **taxa** term	0.76
	Uniform model by **taxa**	0.83
	Interceptions normalized by establishment/interception species richness ratio by **taxa**	0.81
4	Logistic regression based on species interception frequency	0.77
	Logistic regression with **taxa** term	0.67
	Uniform model by **taxa**	0.77
	Interceptions normalized by establishment/interception species richness ratio by **taxa**	0.77
5	Logistic regression based on species interception frequency	0.7
	Logistic regression with **taxa** term	0.69
	Uniform model by **taxa**	0.74
	Interceptions normalized by establishment/interception species richness ratio by **taxa**	0.71

"top" families or taxa are defined as those with at least 20 established species and at least 10 intercepted species in the interception and establishment datasets. Taxa for Scenario 2–5 are listed in Turner et al. (2024b)

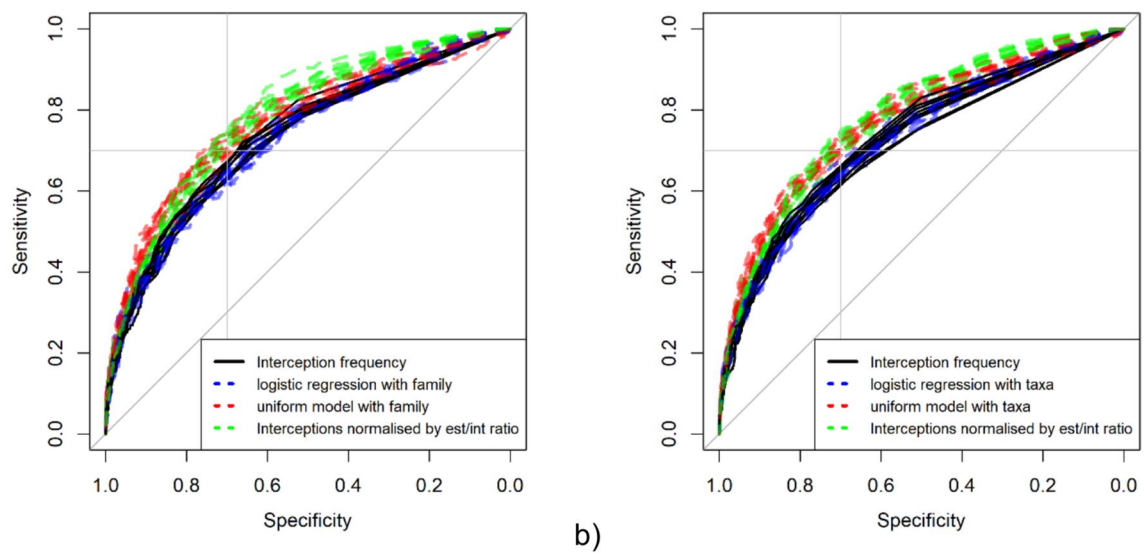


Fig. 5 Receiver operator curves for **a** scenario 1 (insects intercepted at least once in families with at least 10 intercepted species and at least 20 established species—6756 species total) and **b** scenario 2 (insects intercepted at least once in taxa with at least 10 intercepted species and at least 20 established species—8483 species total). Note that the ROC curve for the model based on total interception frequency only is equivalent to the logistic regression model without a family or taxa term, and to the uniform model if fitted without breaking the dataset into family or taxa, and to the ratio model if all ratios were the same. The grey horizontal and vertical lines mark 0.7 sensitivity and specificity respectively

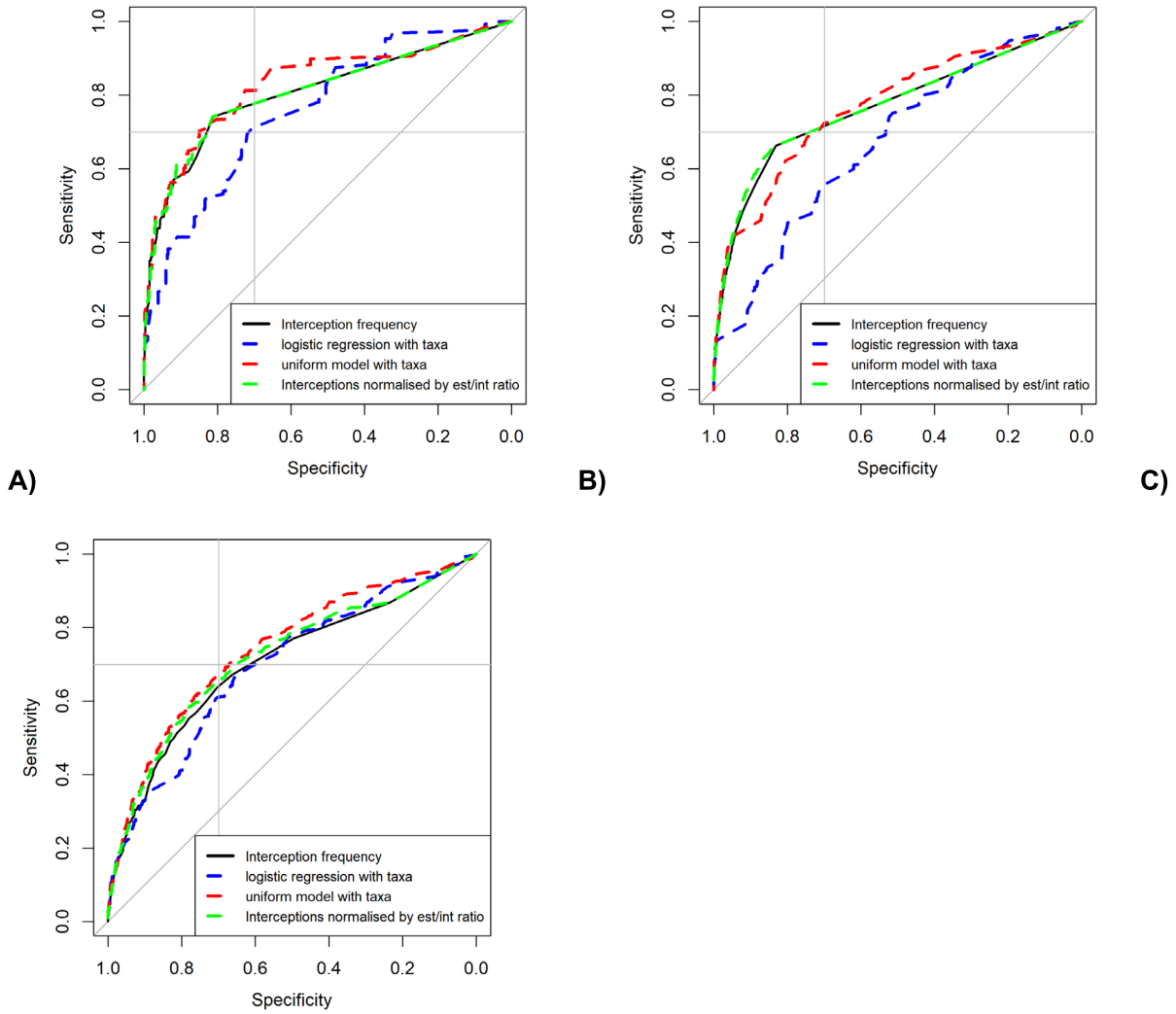


Fig. 6 Receiver operator curves for **a** scenario 3 (insects on the *Pinus radiata* hazard list—649 species total), **b** scenario 4 (insects on the list of New Zealand regulated pests—7874 species total) and **c** scenario 5 (USDA hazard list). Note that the ROC curve for the model based on total interception frequency only is equivalent to the logistic regression model without a taxa term, and to the uniform model if fitted without breaking the dataset into taxa, and to the ratio model if all ratios were the same. The grey horizontal and vertical lines mark 0.7 sensitivity and specificity respectively

Discussion

While there were similar numbers of insect species in the establishment and interception databases (Table 2), the overlap was small (approximately 30%, Figs. 2, S1). However, if the intercepted and established species were randomly sampled from the same pool of potential species, e.g., all described species, an even smaller overlap would be expected. We instead observed a much greater overlap which we suggest is attributed to the underlying distribution of arrival rates and therefore the potential for interceptions to predict invasions.

We also observed that the overlap between intercepted and established species differed between insect groups (Figs. S1, 1, 2, 4, 5). For example, Hemiptera and Thysanoptera had larger proportions of their established species intercepted (Fig. 1) and were the same two orders with the highest border interception sample coverage observed for species at ports of entry in Turner et al. (2021). We argue that some of this variation is non-random. Border interception

systems are complex and driven by multiple policy and operational constraints. But some influences are consistent (or mostly consistent) among countries, leading to predictable effects in interception data (e.g. plant feeder vs non-plant feeders, Fig. S3). There are clear differences in the distribution of species among taxa comparing international border interception data with establishment data (Fig. 1). These differences may reflect variation in detection (and identification) probability, or variation in establishment probabilities among families, though we cannot fully distinguish these two factors without independent data.

There are some groups of insects with low interception probabilities relative to their historical establishment probability—the known unknowns. We know that insects in these groups are likely to establish without being intercepted. Ideally, a large proportion of established species will have also been intercepted, although there may be historical invasions which were facilitated by pathways not monitored by border inspections, or because the species were in fact intercepted but not identified to species. Low coverage may be due to multiple factors including low detection probability, difficulty of species identification, arrival via non-anthropogenic means such as wind dispersal, or historic arrivals due to pathways which are no longer are common (e.g., naval provisions) due to modern transport technology, and pre-biosecurity measures. The general influences identified in this analysis are (i) lower interception probabilities for non-plant feeding insects, (ii) lower interception probabilities for Diptera, and (iii) lower interception probabilities for Hymenoptera. The following sections will discuss the drivers of each trend and the exceptions to the general trend which illustrate strategies for uncovering blind spots (e.g. taxonomic groups or pathways lacking in border interception data).

Low interception probabilities for non-plant feeding insects might be expected given that border biosecurity is often primarily driven by phytosanitary measures. This reflects the historical focus of biosecurity measures on excluding insect pests of plants. However, there is increasing recognition of ecological impacts of insect groups such as predators, pollinators and detritivores (McGeoch et al. 2016; Brockerhoff et al. 2010). There are exceptions to the general trends in some reporter regions, which shows that, with deliberate interventions at the border, some of the blind spots can be removed. Some exceptions to the low coverage for certain groups are the focus by Australia and New Zealand on Formicidae due to an aspiration to exclude non-native ants. The example of substantial interceptions of Formicidae in some countries (Bertelsmeier et al. 2018) demonstrates that it is possible to improve detection probability for these species.

Diptera often arrive at the border as immature stages which are difficult or impossible to quickly identify to the species level (as for many Holometabolous insects, Fig. 4, and note the low proportion of Diptera interceptions identified to species level in many of the reporter regions, Fig. S2). One exception to the general trend of low Diptera interception probabilities is the richness of Tephritidae interceptions by New Zealand. Internationally, Tephritidae are a recognized high impact group of insects (Trombik et al. 2023). However, New Zealand protocols include a priority on sending suspected Tephritidae larvae to a government diagnostics laboratory for identification, and a greater time allowance (compared to other biosecurity samples) to complete molecular diagnostics. Diptera are not the only insect group which tend to arrive as immatures (eggs, larvae or pupae); however, we could not find a consistent source of data (among reporter regions) on the most common life stage of arriving insect taxa. Some border interception datasets do include information on life stage, but it is generally incomplete. Differences in the rate of establishments and interceptions are partially dependent on the extent to which phytosanitary measures are implemented and enforced in different countries. However, it is difficult to synthesize a set of indicators that reflect the extent of biosecurity measures for individual countries because of the large number of phytosanitary measures and operational processes which vary among countries but also within countries, among commodity types and pathways, date of implementation, degree of enforcement, and because many pests can enter via multiple pathways.

Hymenoptera are generally underrepresented in border interception data. We suggest that this is, in part, due to many Hymenoptera being predatory or parasitic rather than plant feeders. Endoparasitic Hymenoptera arriving inside host insects would be particularly difficult to detect. Fenn-Moltu et al. (2024) found that parasitic Hymenoptera are reported in interception data, albeit at a low rate. Many

hymenopterous parasitoids tend to be small adding to the detection difficulty. As mentioned above, a key exception for Hymenoptera is the Formicidae which include many species that are widely considered to be important invasive pests, which is likely to have led to their being considered a priority, resulting in enhanced targeting for inspections and an improved detection probability.

Psocodea, while containing more historically established species than Thysanoptera, are poorly recorded in interception data. However, this is, in general, not likely to be an issue as most establishment records in this order occurred on average nearly a century ago. Many of these establishments will have been associated with early human migration including lice of humans and domestic animals (e.g., Philopteridae and Menoponidae). Families within the order with more recent years of first record (e.g., Liposcelididae—booklice) are still intercepted (primarily by NZ and Australia).

There are other insect groups that are very successful at establishing after arrival. For example, a large fraction of species in the Hemiptera subfamily Sternorrhyncha exhibit parthenogenic reproduction which facilitates their establishment at low densities (Vershinina and Kuznetsova 2016). In these groups, the ratio of intercepted species to established species can also be low. This may be explained by the parthenogenetic reproduction of many Hemiptera, such as aphids, which may make them highly successful invaders as even the arrival of very few individuals could lead to successful invasions (Liebhold et al. 2024).

When border interception frequencies are used to predict probabilities of species establishment, the preferred approach depends on the context and availability of data. In this study, we considered the context of predictions for hazard lists which typically contain several taxa groups. Our results indicate that both the sensitivity and specificity of predictions can be improved by taking information about the family/taxon into account (Fig. 5) when applied to lists of species which have been intercepted at least once.

When using such models to prioritize species from existing hazard lists, it can be anticipated that many insect species on these lists have never been intercepted. In this case, models which include consideration of the probability of establishment for species with zero interceptions have an advantage. The uniform model is one of these and as demonstrated in Fig. 6, has the particular advantage of retaining reasonable specificity (>0.7) while improving sensitivity. However, the magnitude of this advantage depends on the composition of the hazard list. As seen in Fig. 6b, for the NZ hazard list, the uniform model actually performed worse than the default total interception frequency model if wanting to retain specificities above 0.7. This was largely due to the large number of species from the infraorder Aphidomorpha (e.g., aphids) on the NZ hazard list. The uniform model fit to the Aphidomorpha gave a relatively high predicted probability of establishment for the species with zero interceptions (0.34) whereas only 11% of the Aphidomorpha on the NZ hazard list actually have established. This meant that all Aphidomorpha would be predicted to have established given a low enough threshold, in contrast to the other insect taxa on the list. This overprediction of establishment for Aphidomorpha species can be partially attributed to the assumption of the model that arrival rates are uniformly distributed rather than following a right skewed distribution which might be more realistic. Fitting a model to Aphidomorpha interceptions is particularly challenging given that, while Aphidomorpha are plant feeders and we would therefore expect them to be relatively well detected, many are also parthenogenic (Vershinina and Kuznetsova 2016) which means that their probability of establishment is relatively high per arrival (Rubio-Meléndez et al. 2019). Furthermore, establishment predictions for species that have not yet been established may be realised in future. It has been shown that new species become successful invaders all the time, and hundreds of arthropod species, probably mainly insects, that have not yet been recorded as invaders, are predicted to invade in coming decades (Seebens et al. 2021).

There is considerable variation in coverage of border interception data in relation to historical establishments, a large part of which is likely due to variation in detection probability. Some of this variation is due to country-specific operations at ports of entry (as well as during diagnostics) as illustrated by the variation between reporter regions (we leave further investigation of reporter region variation to future research). However, some trends are consistent across all reporter regions such as the tendency

for higher interceptions (relative to establishments) of plant feeding insects relative to non-plant feeders reflecting the historical emphasis on phytosanitary inspections at the border. Models, such as those presented here provide a means for predicting the risk of species establishment even for insect species that are not detected. Improved establishment predictions ultimately should lead to improved risk assessments and general increases in biosecurity programmes meeting their goals of excluding damaging non-native species.

Acknowledgements AML received support from grant EVA4.0, No. CZ.02.1.01/0.0/0.0/16_019/0000803 financed by Czech Operational Programme "Science, Research, and Education". RMT was supported by USDA, Forest Service International Programs (grant # 21-IG-11132762-241) and with funding from the Ministry of Business, Innovation and Employment, New Zealand, through the Better Border Biosecurity (B3) research collaboration (www.b3nz.org) as project B21.3. The authors would like to thank all of the biosecurity staff involved in collecting border interception data which has made such analysis possible. In particular thanks to the Australian Department of Agriculture, Fisheries and Forestry who facilitated the data agreement for the updated Australian border interception data.

Funding Open Access funding provided by Lib4RI – Library for the Research Institutes within the ETH Domain: Eawag, Empa, PSI & WSL. AML received support from grant EVA4.0, No. CZ.02.1.01/0.0/0.0/16_019/0000803 financed by Czech Operational Programme "Science, Research, and Education". RMT was supported by USDA, Forest Service International Programs (grant # 21-IG-11132762–241) and with funding from the Ministry of Business, Innovation and Employment, New Zealand, through Strategic Science Investment Funding and the Better Border Biosecurity (B3) research collaboration (www.b3nz.org) as project B21.3.

Data availability Data sources for the international establishments and border interceptions are provided in Supplementary material Tables S1 and S2. Datasets used in the establishment prediction modelling are provided in Turner et al. (2024b).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Bacon SJ, Bacher S, Aebi A (2012) Gaps in border controls are related to quarantine alien insect invasions in Europe. PLoS ONE 7(10):e47689

Bertelsmeier C, Ollier S, Liebhold AM, Brockerhoff EG, Ward D, Keller L (2018) Recurrent bridgehead effects accelerate global alien ant spread. Proc Natl Acad Sci 115(21):5486–5491

Brockerhoff EG, Bain J, Kimberley M, Knížek M (2006) Interception frequency of exotic bark and ambrosia beetles (*Coleoptera*: *Scolytinae*) and relationship with establishment in New Zealand and worldwide. Can J Res 36(2):289–298

Brockerhoff EG, Barratt BIP, Beggs JR, Fagan LL, Kay MKN, Phillips CB, Vink CJ (2010) Impacts of exotic invertebrates on New Zealand's indigenous species and ecosystems. N Z J Ecol 34(1):158–174

Brockerhoff EG, Kimberley M, Liebhold AM, Haack RA, Cavey JF (2014) Predicting how altering propagule pressure changes establishment rates of biological invaders across species pools. Ecology 95(3):594–601

Brockerhoff EG, Gresham BA, Meurisse N, Nahrung HF, Perret-Gentil A, Pugh AR, Sopow SL, Turner RM (2023) Pining away and at home: global utilisation of *Pinus radiata* by native and non-native insects. NeoBiota 84:137–167

Caley P, Ingram R, De Barro P (2015) Entry of exotic insects into Australia: does border interception count match incursion risk? Biol Invasions 17(4):1087–1094. https://doi.org/10.1007/s10530-014-0777-z

Chamberlain SA, Szöcs E (2013) Taxize: taxonomic search and retrieval in R. F1000Research. https://doi.org/10.12688/f1000research.2-191.v2

Core Team R (2021) R: A Language and environment for statistical computing. . R Foundation for Statistical Computing. https://www.R-project.org/. Accessed 27/06/2024

Eschen R, Britton K, Brockerhoff E, Burgess T, Dalley V, Epanchin-Niell RS, Gupta K, Hardy G, Huang Y, Kenis M, Kimani E, Li HM, Olsen S, Ormrod R, Otieno W, Sadof C, Tadeu E, Theyse M (2015a) International variation in phytosanitary legislation and regulations governing importation of plants for planting. Environ Sci Policy 51:228–237. https://doi.org/10.1016/j.envsci.2015.04.021

Eschen R, Roques A, Santini A (2015b) Taxonomic dissimilarity in patterns of interception and establishment of alien arthropods, nematodes and pathogens affecting woody plants in Europe. Divers Distrib 21(1):36–45. https://doi.org/10.1111/ddi.12267

Fenn-Moltu G, Liebhold AM, Weber DC, Bertelsmeier C (2024) Pathways for accidental biocontrol: The human-mediated dispersal of insect predators and parasitoids. Ecol Appl. https://doi.org/10.1002/eap.3047

Games PA, Howell JF (1976) Pairwise multiple comparison procedures with unequal N's and/or variances: a Monte Carlo study. J Edu Stat 1(2):113–125. https://doi.org/10.2307/1164979

GBIF Secretariat (2021) GBIF Backbone Taxonomy. Checklist dataset https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c. Accessed 02/09/2022

Haack RA (2006) Exotic bark- and wood-boring *Coleoptera* in the United States: recent establishments and interceptions. Can J for Res 36(2):269–288. https://doi.org/10.1139/x05-249

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6(2):65–70

Hulme PE (2010) Biosecurity: the changing face of invasion biology. In: Richardson DM (ed) Fifty years of invasion ecology: the legacy of Charles Elton. Wiley, pp 301–314

Jenkins DA, Mizell RF III, Van Bloem S, Whitmire S, Wiscovitch L, Zaleski C, Goenaga R (2014) An analysis of arthropod interceptions by APHIS-PPQ and customs and border protection in puerto rico. Am Entomol 60(1):44–57. https://doi.org/10.1093/ae/60.1.44

Kean J, Phillips C, McNeill M (2008) Surveillance for early detection: lottery or investment. In: KJ Froud, AI Popay, Zydenbos S (eds) Surveillance for Biosecurity: pre-border to pest management, pp 11–17

Lee W, Lee Y, Kim S, Lee J-H, Lee H, Lee S, Hong K-J (2016) Current status of exotic insect pests in Korea: comparing border interception and incursion during 1996–2014. J Asia-Pac Entomol 19(4):1095–1101. https://doi.org/10.1016/j.aspen.2016.09.003

Leung B, Lodge DM, Finnoff D, Shogren JF, Lewis MA, Lamberti G (2002) An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proc R Soc Lond B 269(1508):2407–2413. https://doi.org/10.1098/rspb.2002.2179

Liebhold AM, Work TT, McCullough DG, Cavey JF (2006) Airline baggage as a pathway for alien insect species invading the United States. Am Entomol 52(1):48–54. https://doi.org/10.1093/ae/52.1.48

Liebhold AM, Yamanaka T, Roques A, Augustin S, Chown SL, Brockerhoff EG, Pyšek P (2016) Global compositional variation among native and non-native regional insect assemblages emphasizes the importance of pathways. Biol Invasions 18(4):893–905. https://doi.org/10.1007/s10530-016-1079-4

Liebhold AM, Turner RM, Blake RE, Bertelsmeier C, Brockerhoff EG, Nahrung HF, Pureswaran DS, Roques A, Seebens H, Yamanaka T (2021) Invasion disharmony in the global biogeography of native and non-native beetle species. Divers Distrib 27(11):2050–2062. https://doi.org/10.1111/ddi.13381

Liebhold AM, Turner RM, Bartlett CR, Bertelsmeier C, Blake RE, Brockerhoff EG, Causton CE, Matsunaga JN, McKamey SH, Nahrung HF, Owen CL, Pureswaran DS, Roques A, Schneider SA, Sanborn AF, Yamanaka T (2024) Why so many hemiptera invasions? Divers Distrib 30(12):e13911. https://doi.org/10.1111/ddi.13911

Mally R, Turner RM, Blake RE, Fenn-Moltu G, Bertelsmeier C, Brockerhoff EG, Hoare RJB, Nahrung HF, Roques A, Pureswaran DS, Yamanaka T, Liebhold AM (2022) Moths and butterflies on alien shores: global biogeography of non-native Lepidoptera. J Biogeogr 49(8):1455–1468. https://doi.org/10.1111/jbi.14393

McCullough DG, Work TT, Cavey JF, Liebhold AM, Marshall D (2006) Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. Biol Invasions 8(4):611–630. https://doi.org/10.1007/s10530-005-1798-4

McGeoch MA, Genovesi P, Bellingham PJ, Costello MJ, McGrannachan C, Sheppard A (2016) Prioritizing species, pathways, and sites to achieve conservation targets for biological invasion. Biol Invasions 18(2):299–314. https://doi.org/10.1007/s10530-015-1013-1

McNeill M, Phillips C, Young S, Shah F, Aalders L, Bell N, Gerard E, Littlejohn R (2011) Transportation of nonindigenous species via soil on international aircraft passengers' footwear. Biol Invasions 13:2799–2815

McNeill M, Phillips C, Richards N, Alders L, Van Koten C, James T, Young S, Bell N, Laugrad A (2023) Defining the biosecurity risk posed by soil found on sea freight. Neobiota 88:103–133. https://doi.org/10.3897/neobiota.88.98440

Migliorini D, Ghelardini L, Tondini E, Luchi N, Santini A (2015) The potential of symptomless potted plants for carrying invasive soilborne plant pathogens. Divers Distrib 21(10):1218–1229

Ministry of Agriculture and Forestry (2003) MAF Discussion Paper No. 35. Prepared for MAF Biosecurity Authority by the Border Management Group. Ministry of Agriculture and Forestry, Wellington, New Zealand

Montibeller G, Franco LA, Carreras A (2020) A risk analysis framework for prioritizing and managing biosecurity threats. Risk Anal 40(11):2462–2477. https://doi.org/10.1111/risa.13542

Nahrung HF, Carnegie AJ (2021) Border interceptions of forest insects established in Australia: intercepted invaders travel early and often. NeoBiota 64:69–86

Navia D, Ochoa R, Welbourn C, Ferragut F (2010) Adventive eriophyoid mites: a global review of their impact, pathways, prevention and challenges. Exp Appl Acarol 51(1):225–255. https://doi.org/10.1007/s10493-009-9327-2

Phillips CB, Kean JM, Vink CJ, Berry JA (2018) Utility of the CLIMEX 'match climates regional' algorithm for pest risk analysis: an evaluation with non-native ants in New Zealand. Biol Invasions 20(3):777–791. https://doi.org/10.1007/s10530-017-1574-2

Rubio-Meléndez ME, Barrios-SanMartin J, Pina-Castro FE, Figueroa CC, Ramirez CC (2019) Asexual reproduction of a few genotypes favored the invasion of the cereal aphid Rhopalosiphum padi in Chile. PeerJ 7:e7366. https://doi.org/10.7717/peerj.7366

Saccaggi DL, Karsten M, Robertson MP, Kumschick S, Somers MJ, Wilson JRU, Terblanche JS (2016) Methods and approaches for the management of arthropod border incursions. Biol Invasions 18(4):1057–1075. https://doi.org/10.1007/s10530-016-1085-6

Saccaggi DL, Arendse M, Wilson JRU, Terblanche JS (2021) Contaminant organisms recorded on plant product imports to South Africa 1994–2019. Scientific Data 8(1):83. https://doi.org/10.1038/s41597-021-00869-z

Seebens H, Bacher S, Blackburn TM, Capinha C, Dawson W, Dullinger S, Genovesi P, Hulme PE, van Kleunen M, Kühn I, Jeschke JM, Lenzner B, Liebhold AM, Pattison Z, Pergl J, Pyšek P, Winter M, Essl F (2021) Projecting the continental accumulation of alien species through to 2050. Glob Change Biol 27(5):970–982. https://doi.org/10.1111/gcb.15333

Stork NE (2018) How many species of insects and other terrestrial arthropods are there on earth? Annu Rev Entomol 63(1):31–45. https://doi.org/10.1146/annurev-ento-020117-043348

Suhr EL, O'Dowd DJ, Suarez AV, Cassey P, Wittmann TA, Ross JV, Cope RC (2019) Ant interceptions reveal roles of transport and commodity in identifying biosecurity risk pathways into Australia. NeoBiota 53:1–24

Trombik J, Ward SF, Norrbom AL, Liebhold AM (2023) Global drivers of historical true fruit fly (*Diptera*: *Tephritidae*) invasions. J Pest Sci 96(1):345–357. https://doi.org/10.1007/s10340-022-01498-0

Turner RM, Plank MJ, Brockerhoff EG, Pawson S, Liebhold A, James A (2020) Considering unseen arrivals in predictions of establishment risk based on border biosecurity interceptions. Ecol Appl 30(8):e02194. https://doi.org/10.1002/eap.2194

Turner RM, Brockerhoff EG, Bertelsmeier C, Blake RE, Caton B, James A, MacLeod A, Nahrung HF, Pawson SM, Plank MJ, Pureswaran DS, Seebens H, Yamanaka T, Liebhold AM (2021) Worldwide border interceptions provide a window into human-mediated global insect movement. Ecol Appl 31(7):e02412. https://doi.org/10.1002/eap.2412

RMTurnerALiebholdHNahrungCPhillipsTYamanakaEG-Brockerhoff2024The known unknowns in international border interceptions of non-native insects datasetsZenodo10.5281/zenodo.14085684Turner RM, Liebhold A, Nahrung H, Phillips C, Yamanaka T, Brockerhoff EG (2024b) The known unknowns in international border interceptions of non-native insects datasets. Zenodo. https://doi.org/10.5281/zenodo.14085684

Turner R, Blake R, Mally R, Liebhold AM (2024a) International Non-native Insect Establishment Data (0.2.1) . Zenodo. https://doi.org/10.5281/zenodo.11519262

US Department of Agriculture Animal and Plant Health Inspection Service [USDA APHIS] (2021) Agricultural quarantine inspection monitoring (AQIM) handbook. 3rd Edn edn. US Department of Agriculture, Animal and Plant Health Inspection Service, Riverdale, MD

Vershinina AO, Kuznetsova VG (2016) Parthenogenesis in *Hexapoda*: *Entognatha* and non-holometabolous insects. J Zool Syst Evol Res 54(4):257–268. https://doi.org/10.1111/jzs.12141

Work TT, McCullough DG, Cavey JF, Komsa R (2005) Arrival rate of nonindigenous insect species into the United States through foreign trade. Biol Invasions 7(2):323–332. https://doi.org/10.1007/s10530-004-1663-x

Worner SP, Gevrey M (2006) Modelling global insect pest species assemblages to determine risk of invasion. J Appl Ecol 43(5):858–867. https://doi.org/10.1111/j.1365-2664.2006.01202.x

Zhang ZQ (2013) Phylum Arthropoda. In: Zhang ZQ (eds) Animal biodiversity An outline of higher-level classification and survey of taxonomic richness Zootaxa, pp 17–26