

RESOURCE ARTICLE

Current stewardship practices in invasion biology limit the value and secondary use of genomic data

Amy L. Vaughan¹  | Elahe Parvizi²  | Paige Matheson² | Angela McGaughan²  | Manpreet K. Dhami¹ 

¹Biocontrol & Molecular Ecology, Manaaki Whenua Landcare Research, Lincoln, New Zealand

²Te Aka Mātuaatua/School of Science, University of Waikato, Hamilton, New Zealand

Correspondence

Manpreet K. Dhami and Amy L. Vaughan, Biocontrol & Molecular Ecology, Manaaki Whenua Landcare Research, Lincoln, New Zealand.

Email: dhamim@landcareresearch.co.nz and vaughana@landcareresearch.co.nz

Funding information

Genomics Aotearoa, Grant/Award Number: GA 2102

Handling Editor: Catherine E Grueber

Abstract

Invasive species threaten native biota, putting fragile ecosystems at risk and having a large-scale impact on primary industries. Growing trade networks and the popularity of personal travel make incursions a more frequent risk, one only compounded by global climate change. With increasing publication of whole-genome sequences lies an opportunity for cross-species assessment of invasive potential. However, the degree to which published sequences are accompanied by satisfactory spatiotemporal data is unclear. We assessed the metadata associated with 199 whole-genome assemblies of 89 invasive terrestrial invertebrate species and found that only 38% of these were derived from field-collected samples. Seventy-six assemblies (38%) reported an 'undescribed' sample origin and, while further examination of associated literature closed this gap to 23.6%, an absence of spatial data remained for 47 of the total assemblies. Of the 76 assemblies that were ultimately determined to be field-collected, associated metadata relevant for invasion studies was predominantly lacking: only 35% (27 assemblies) provided granular location data, and 33% ($n=25$) lacked sufficient collection date information. Our results support recent calls for standardized metadata in genome sequencing data submissions, highlighting the impact of missing metadata on current research in invasion biology (and likely other fields). Notably, large-scale consortia tended to provide the most complete metadata submissions in our analysis—such cross-institutional collaborations can foster a culture of increased adherence to improved metadata submission standards and a standard of metadata stewardship that enables reuse of genomes in invasion science.

KEYWORDS

biological invasion, invasion genomics, metadata, reference genomes, sequencing

1 | INTRODUCTION

As global trade networks expand and lifestyle-associated travel continues to rise, so does the risk associated with biological incursions,

where species establish populations outside their native range, often causing negative economic and ecological impacts (Hulme, 2021; Latombe et al., 2022; Seebens et al., 2018). The frequency and success of these incursions are only expected to increase (e.g. established alien

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

species per continent are predicted to rise by 36% from 2005 to 2050; Seebens et al., 2021) as rising temperatures driven by anthropogenic climate change act to narrow the climatic gap between tropical and temperate regions (Gutierrez et al., 2021). Of the global pests routinely transported via human-mediated or unaided pathways (as defined by Hulme et al., 2008), perhaps the most pervasive are invertebrates, with an estimated global cost exceeding US\$700 billion from 1960 to 2020 primarily due to resource loss and direct damage (Renault et al., 2022). Global health is also affected by the zoonotic vectoring of insect-borne diseases (Abubakr et al., 2022; Zhang et al., 2022), with implications for endemic biodiversity in natural habitats (Gentili et al., 2021).

As sequencing technologies have become more accessible, a surge in whole-genome sequencing (WGS) and population genomic studies of invasive species has followed (Matheson & Mc-Gaughran, 2022). Collectively, these data hold the promise of disentangling the genomic mechanisms of invasion success. In the case of invasion genomics (the study of the genome within the context of biological invasion), WGS data are demonstrating emerging trends in invasive populations that are valuable for potential mitigation and control strategies. For example, genes associated with environmental adaptation (e.g. insecticide resistance and olfaction) have been identified in a range of unrelated invasive insects, such as the navel orangeworm *Amyelois transitella* (Pyralidae; Calla et al., 2020), the brown marmorated stink bug *Halyomorpha halys* (Hemiptera; Parvizi et al., 2023) and the fruit fly *Drosophila suzukii* (Diptera; Durkin et al., 2021). Rapid evolution in translocated populations, underpinned by either abiotic (such as climatic gradients in *Drosophila melanogaster*, Behrman et al., 2018); and across beetle families in Australia, (Wardhaugh et al., 2018) or biotic factors (e.g. new predator-prey interactions; Siepielski & Beaulieu, 2017) can be detected using invasion genomic approaches (North et al., 2021; Runquist et al., 2020). A recent study by Lukicheva and Mardulyn (2021) also identified evidence of asymmetric introgression in ~2% of the genome of an invasive leaf beetle, a low enough frequency to be undetectable via reduced representation sequencing. Thus, WGS is already beginning to elucidate intrinsic mechanisms that underpin an organism's ability to invade, and previous reviews have discussed the potential of this technology in invasion biology (McCartney et al., 2019; Rius et al., 2015). When inferring invasion events, rare variant single nucleotide polymorphisms (SNPs) or those unrepresented in a reference genome may not be identified in a low coverage genome, and inference of biologically important variants may be lost. Therefore, depending on the biological question, high sequencing depth and/or whole-genome sequencing (versus reduced representation) data may be required to obtain an accurate representation of rare variant prevalence in populations (North et al., 2021).

To allow for reproducible experimentation, accurate and comprehensive metadata are required alongside published sequencing data. In the context of biological invasions, such spatiotemporal data are crucial for tracing incursion pathways (Bradhurst et al., 2021), whilst knowledge of host species may promote the study of reciprocal evolution in antagonistic species in host-parasite interactions

(Beaurepaire et al., 2019). Such metadata can also extend the use of the sequencing data beyond a project's initial intent. Global databases, such as GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), require minimum metadata standards to be met at the time of WGS data submission, including the addition of collection date, geographic location, tissue type and host/source. But this list is not exhaustive and does not set requirements for the minimum standard of data provided (e.g. the precision of spatiotemporal data is not specified). In addition, factors such as geographic coordinates, sex and developmental stage are optional contributions. While optional metadata fields may remove some barriers to data submission, a lack of metadata can severely hamper the reusability of the submitted sequencing data (Tahsin et al., 2017). In the context of biological invasions, enriched metadata have enabled expansive research outside the original collection context, providing detail on expanded populations (Schmidt et al., 2023). High-quality, complete spatiotemporal metadata records from museum samples have also allowed for a detailed analysis of divergent and parallel evolution during the invasion process (Stuart et al., 2022). Thus, without complete metadata records, the meaningful historic and biological context behind an invasion event can be lost.

To overcome the obstacles associated with metadata provision and data reuse, FAIR guiding principles of data stewardship state that data (and metadata) should be Findable, Accessible, Interoperable and Reusable (Wilkinson et al., 2016). In support of and adherence to FAIR guiding principles, recent genome sequencing projects and consortia—formed for example, with the purpose of sequencing all organisms (Lewin et al., 2022) or all life in the British Isles (Darwin Tree of Life project; Blaxter et al., 2022)—apply stringent data management policies in the recording of associated metadata. For instance, the Darwin Tree of Life project has recently proposed metadata standardization, encouraging other projects to adopt these strategies (Lawniczak et al., 2022). However, it is unclear how widely either individual researchers or consortia adhere to FAIR or other recommended data stewardship principles when capturing metadata associated with invasive species WGS experiments. To address this, we ask whether data stewardship practices in the field of invasion genomics may be limiting the value and secondary uses of genomic data.

We focus on data for terrestrial arthropods (Phylum: Arthropoda)—one of the most pervasive taxa in global incursions and among the most prevalent taxonomic groups sequenced. Collating publicly available WGS data for this group, we examine information supplied for key metadata categories with a particular focus on whether information relevant to large-scale comparative genomic analyses (i.e. potentially applicable to broadscale insights into invasive potential) is available.

2 | MATERIALS AND METHODS

A current list of invasive arthropod species with available reference genomes was compiled from the i5K Ag100Pest species list (Childers et al., 2021), the IUCN Global Invasive species database

(GISD; Poorter et al., 2005), the IUCN 100 of the World's Worst Invasive Alien Species (WAS) list (Lowe et al., 2000), and directly from the literature. The Ag100Pest Initiative is a US-specific project that aims to contribute arthropod genomes of 158 agricultural pest or potential pest species to the i5K 5000 arthropod genomes project (i5K Consortium, 2013) and Earth BioGenome Projects for cataloguing all eukaryotic diversity (Lewin et al., 2022). The WAS list comprises alien species from each taxonomic order, representing diverse taxonomic groups with large implications on human activity and species biodiversity, including 14 arthropod taxa. We searched the literature using PubMed, with search terms such as 'invasive', 'WGS' and 'arthropod', to identify any further invasive terrestrial arthropods for which WGS data were available. For each of the species on these combined lists ($n=240$), invasive status was confirmed using wider literature searches and/or checks against the Invasive Species Compendium (www.cabi.org; CABI, 2023) in June 2023, resulting in a final list of 89 species with confirmed invasive status.

As the National Centre for Biotechnology Information (NCBI) is linked directly to the i5K project and is the largest compendium of genome data, we used NCBI to locate genome and associated metadata for these species. We first collected metadata from NCBI that was associated with the reference genome of each species, identifying 'reference genome' status based on its allocation as such within the NCBI database. We refer to this dataset of 89 species as the 'reference genome' dataset hereafter. Because reference genomes are often updated, thus several versions can exist for the same species, we also collected metadata for each subsequent genome assembly that was also available for any of these species in NCBI ('assembly dataset' hereafter). Between one and nine representative genome submissions were found for each of our 89 species, resulting in metadata for the assembly dataset for 199 whole genome assembly submissions (i.e. the 'reference genome' dataset is a subset of the 'assembly' dataset; Supplementary File S1). Alternative haplotypes and subsequent assemblies that represent whole-genome resequencing population data (and not 'reference genomes') were not considered in the analysis.

2.1 | Metadata compilation for assembly and reference genome datasets

Metadata was compiled from each assembly in the assembly and reference genome datasets, including spatiotemporal data, sex, tissue type and assembly level (chromosome, scaffold, or contig). Spatiotemporal data were collected at country, region, coordinate, year, month and date levels. Tissue type defining whole, sectional or pooled samples, as well as life stage was recorded, while sex was listed as either female or male. For those genomes that did not provide spatiotemporal data to classify origin as either wild collected or laboratory/colony reared, associated literature searches with the accession/BioProject number were used to further determine this, or a classification of unknown origin was recorded. The organization and country associated with each assembly submission was assessed

to understand the distribution of invasive species research globally. A map of spatial level metadata was then generated using ArcMap 10.8.2. Frequency of submitting institutes was tallied, with 'high income' origin countries defined as those where gross national income per capita exceeds USD12,055 (www.worldeconomics.com/Regions/High-Income-Countries/). Accessibility of raw sequencing data was recorded by the presence of linked Sequence Read Archive (SRA) identifiers. Submission institute/country was also collected as described above. Finally, metadata was compared across all fields to determine overall data completeness for both the reference genome and assembly datasets.

2.2 | Field-collected subset analysis

Within invasion studies, field-collected samples are the most valuable data asset, as they represent a genomic snapshot from within the native/expanded range at a specific time point that can be related to incursions. For this reason, the subset of assemblies from the assembly dataset that were identified as field-collected was further analysed. Status as native/expanded was recorded based on the confirmed collection location, which was correlated with the known geographical home range of each species using the CABI database. The quantity of native/expanded range assemblies per species was then determined, as was the completeness of spatiotemporal data, the tissue type/sex, and submitting institute or consortia. Submission year was also compiled for this subset of field-collected assemblies to investigate temporal metadata trends.

3 | RESULTS

3.1 | Current trends in invasive species genome sequencing

As of June 2023, 55 of 158 species (35%) in the AG100Pest list had a publicly available reference genome. Thirty-four of these agricultural pests were further classified as 'invasive' using literature and the CABI database and, of these, 26 species had more than one assembly available on NCBI. Of the IUCN GISD list—when restricted to terrestrial arthropods ($n=86$)—30 species (~34.9%) had publicly available reference genomes/assemblies. Meanwhile, of the 14 invasive insects included on the WAS list, nine (64%) had a publicly available reference genome and five of these had >1 available genome assembly. This resulted in a total dataset of 89 and 199 reference genomes and assemblies, respectively.

The NCBI and i5K database searches together identified 10 orders of terrestrial arthropods for which reference genomes were available, with the largest proportion of data belonging to Hymenoptera (29%), Diptera (15%) and Coleoptera (15%), and Blattodea and Thysanoptera least represented (~0.5%, Figure 1). Genome assemblies were distributed unequally across the 89 species, with some having a low frequency and others (typically of high risk of

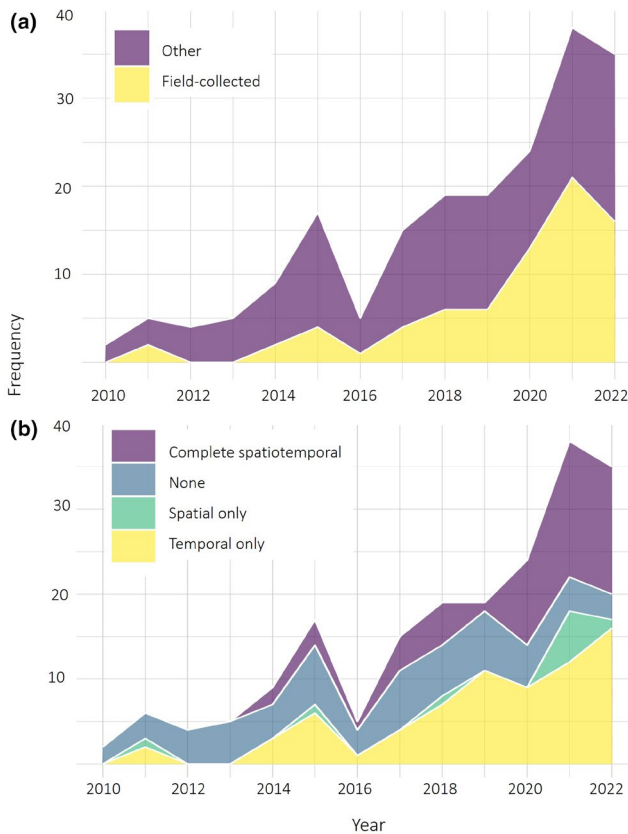


FIGURE 2 Frequency of genome assembly deposition to repositories and their relative spatiotemporal metadata entry completeness for invasive terrestrial arthropods between 2010–2022. (a) Data were separated into ‘other’—defined as the combination of lab, commercial, and unknown origins—and ‘field-collected’ based on metadata associated with 199 assemblies, using submission date for each assembly as a classifier. (b) Completeness of spatiotemporal metadata in the assembly dataset by submission year, where ‘complete’ refers to spatial data that includes collection status (field, laboratory), provision of geographic location (to at least country), and temporal (to at least year of collection) information.

level (5.6%). This contrasted with the assembly dataset, where 33% ($n=67$) and 54% ($n=108$) of assemblies had chromosome and scaffold level completeness, respectively (Figure 3a).

When identifying the origin of the reference genomes, 38% ($n=34$) were field-collected, 34% ($n=30$) were derived from laboratory/commercial colonies and 28% ($n=25$) had an unknown origin (Figure 3b). GPS coordinate level location information was provided for 25% ($n=22$) of reference genomes overall, and 14 of these were from free-living field-collected organisms (Figure 3c). The frequency of entries with no associated collection date was high (40%, $n=36$), though only one of these was found to be from a field-collected organism. However, collection date specificity was not uniform, with 18 (20%) of those that did provide a sample collection date only providing the year of collection (Figure 3c).

Approximately 32% of the total assemblies ($n=64$) were classified as field-collected, 28% ($n=56$) originated from laboratory/commercial colonies, and ~38% ($n=76$) had unknown origin (Figure 3b).

For the 76 assemblies that did not provide adequate metadata to determine sample origin, literature searches identified 12 additional assemblies derived from field-collected organisms (raising the total to $n=76$), while an additional 17 were determined to be from laboratory/commercial colonies (total increased to $n=75$). Sample origin information could not be retrieved for the remaining 47 assemblies, thus, origin information was ultimately obtainable for 76.3% of the assembly dataset. Metadata for 45 assemblies (22.6%) did not provide specific sample collection location (e.g. GPS coordinates; country), while 69 assemblies (34.6%) did not provide collection year.

Analysis of further metadata fields found a high degree of missing data, with absent tissue type information for 27 (30%) of reference genomes and 65 (32%) of assemblies. Similarly, developmental stage information was missing for 38 (42%) reference genomes, and 65 (32.6%) assemblies, and sex was not provided for 33 (37%) and 99 (49.7%) of reference genome and assembly data, respectively. Finally, we found that 64% of assemblies had associated raw data that was submitted with an accessible SRA identifier.

3.3 | Spatiotemporal origins of field-collected samples remain elusive

Of the 76/199 total assemblies identified as field-collected (including 12 assemblies which we added via literature review; see above), 39 (50.6%) were also classified as ‘reference genomes’ for their species, and 50 (56%) of the 89 total species included in this study were represented (Supplementary File S1).

Field-collected assemblies had a higher proportion of data relating to tissue type compared with the complete assembly dataset, with 46 (60.5%) and 47 (61.8%) including information on the developmental stage and sex, respectively. The representation of complete chromosome level sequences was slightly higher in field-collected assemblies versus the whole assembly dataset (34.2% vs. 33.7%) but lower than reference genomes (34.2% vs. 42.7%). Overall, ~70% of field-collected assemblies were of scaffold level or higher.

A slightly higher proportion of invasive species assemblies that came from field-collected samples were collected from their native ($n=40$, 51.9%) rather than expanded range ($n=33$, 42.9%) (Figure 4). Without the addition of assemblies from the literature review, these numbers reduced further for expanded range assemblies ($n=26$, 40%) though increased for native range assemblies ($n=35$, 53.8%). For four assemblies with field-collected status, native/expanded status could not be determined, as specific location was unknown, or the species was listed as feral.

Specific location data for most of these assemblies were restricted to region or city level, with location data provided only at the country level for eight (10%) of assemblies. Only 35.5% provided GPS coordinates, including six (26%) from the Wellcome Sanger Institute. Of the 76 field-collected assemblies, 58% ($n=44$) were collected from high-income group countries (Figure 5).

Seven species had assemblies from field-collected specimens that were collected from both the native and expanded ranges

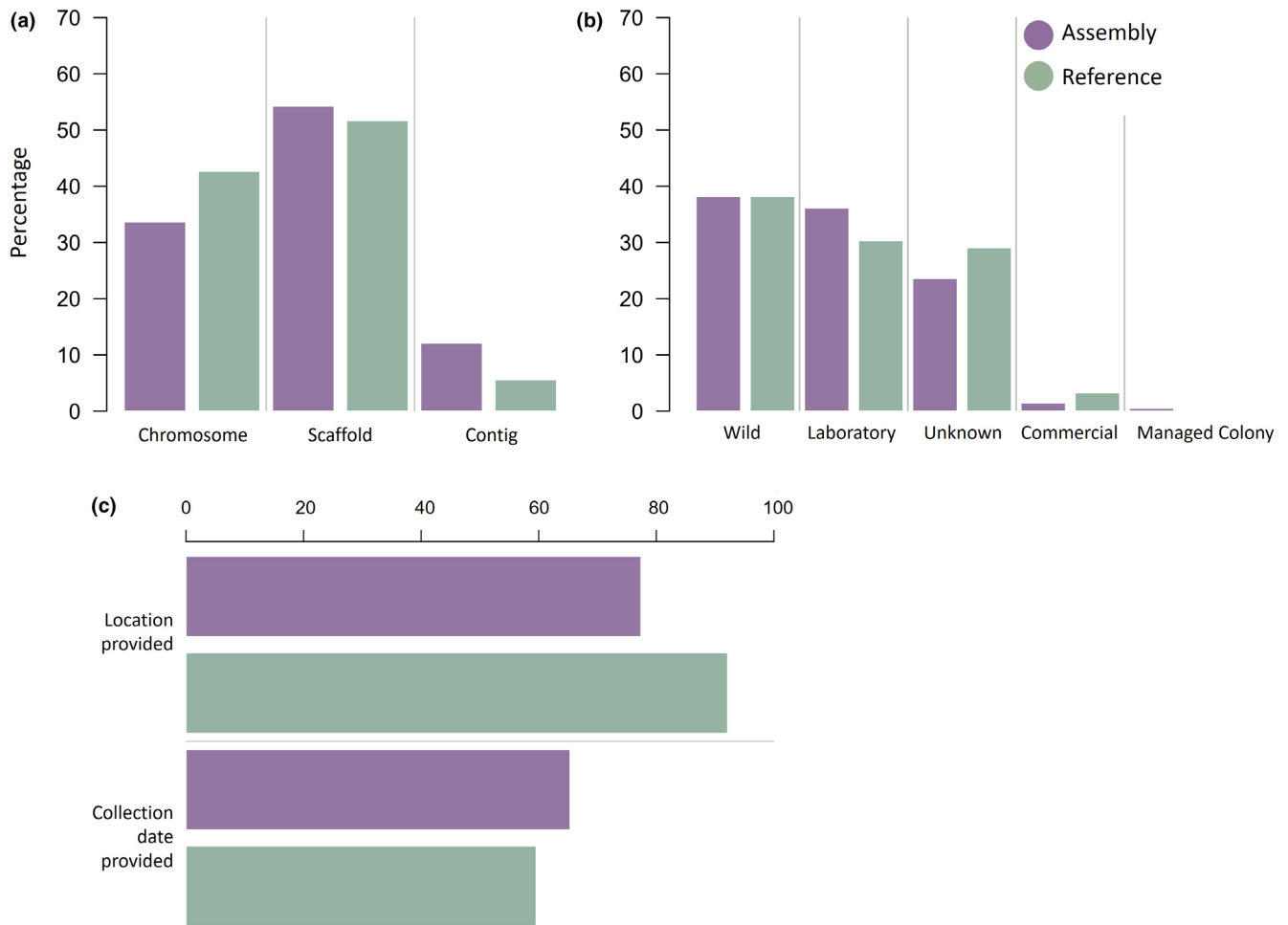


FIGURE 3 Metadata trends for reference genome ($n=89$) and assembly ($n=199$) datasets. (a) Assembly completeness (chromosome, scaffold, or contig level); (b) Sample origin (wild, laboratory, commercial, and managed colony); and (c) Completeness of spatiotemporal metadata (location and/or collection date provided).

(Table 1). These included six insects and one arachnid species (glassy-winged sharpshooter; *Homalodisca vitripennis*: Hemiptera), common wasp (*Vespula vulgaris*: Hymenoptera), small white butterfly (*Pieris rapae*: Lepidoptera), harlequin ladybird (*Harmonia axyridis*: Coleoptera), asian citrus psyllid (*Diaphorina citri*: Hemiptera), european wasp (*Vespula germanica*: Hymenoptera) and asian longhorned tick (*Haemaphysalis longicornis*: Ixodida; added via literature review), representing just 7.8% of the complete list of 89 species in our analysis, and 7.5% of the genome assemblies.

4 | DISCUSSION

With the rise in accessibility of next-generation sequencing technologies in the last decade, research into the genomics of invasive species is burgeoning. Yet, as the number of reference genomes and genome assemblies has increased, we found an increasing tendency for incomplete metadata associated with submissions, limiting the usability of these genomic resources for the study of mechanisms that underpin invasion success.

Only 38.2% of reference genomes, and 38.1% of assemblies in our respective datasets contained enough metadata to be comprehensively classified as field-collected. Yet, undertaking large-scale, comprehensive comparative genomic analysis to identify signals of invasiveness requires genome assemblies of species collected from their native and expanded ranges (Dematteis et al., 2020; North et al., 2021; Turner et al., 2021) to identify rare variants that may not be present within a single reference genome (e.g. following post-invasion adaptation). For example, Carlson et al. (2022) sequenced three high-quality chromosome scale assemblies of *Diaphorina citri* (from two expanded and one native range individuals) and found that the native range strain was more similar to one of the expanded range strains in a manner congruent with the known invasion history. Such studies cement the value of using high-quality chromosome level genomes from both invaded and native ranges in invasion studies. The absence of environmental context for most invasive insects effectively excludes these genomes/assemblies from future analysis (without additional sequencing) where the aim is to understand adaptation in the invaded range. For example, studies of *D. sukuzii*

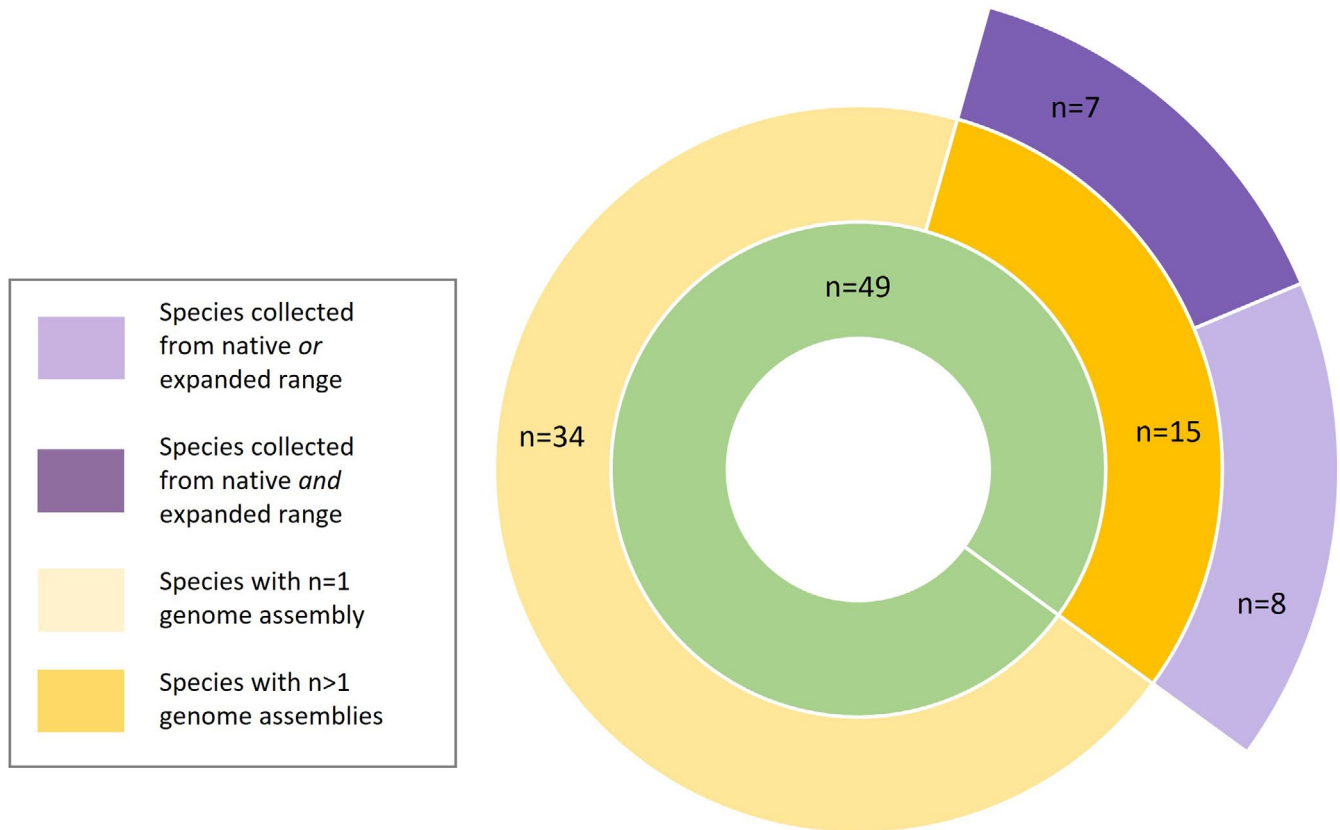


FIGURE 4 Metadata trends for field-collected species assemblies. The proportion (from a total of $n=49$; green inner circle) of species with field-collected data, including those with single or multiple assemblies (shades of yellow), and those that included information on native or expanded range (shades of purple). Includes two assemblies of *Aphis gossypii*, where the origins of the species were unknown but were specified as field-collected.

demonstrate that the genome content of laboratory colonies does not necessarily mirror that of free-living individuals, with samples derived from laboratory populations in Japan and Hawai'i segregating in population structure analyses (Lewald et al., 2021). Similarly, a study by Lainhart et al. (2015) on laboratory isolates of the mosquito *Anopheles darlingi* after 21 generations showed the effects of genetic bottlenecks within populations, including decreased heterozygosity compared with the founder population. Studies like these suggest that the genetic drift commonly observed in long-term laboratory populations would make them unsuitable for large-scale comparisons between invasive and non-invasive species, as well as potentially for intraspecific comparisons between native and expanded ranges. Nevertheless, our data indicate that invasive species genomic data are commonly derived from laboratory and commercial lines (30/89 reference genomes; 75/199 assemblies).

In our dataset, 28% of reference genomes and 23.6% of genome assemblies had an unknown or unlisted origin. This compares with a large-scale analysis by Toczydlowski et al. (2021), which showed that spatiotemporal metadata was present in 51% of 'wild' datasets of eukaryote species (and just 39% where location data was specified as GPS coordinates). Our findings likely also translate to other species, though the number of field-collected samples with

geotagged locations is likely to be much higher in some studies (e.g. those focussed on population genomics). Original study objectives are therefore likely to be a limiting factor in the reuse of data, and generation of 'intention free' data that includes all available metadata would have far-reaching benefits for invasion biology and the wider genomics research community.

Overall, 40% of the 89 species, and 35% of the 199 assemblies also lacked a collection date. An important aspect to invasion biology are comparisons of historic and contemporary data to identify key genomic changes over time. For example, findings of parallel selection across historic and contemporary samples of *Sturnus vulgaris* in the UK and Australia suggest occasional convergent responses in both the native and expanded range, demonstrating the value of studying historic data (Stuart et al., 2022, 2023). Historical collections (e.g. museums, herbaria) are commonly used to address such questions, where samples have not been previously collected and retained for this purpose but the provision of relevant temporal data would allow the future reuse of contemporary samples in time series studies. A crucial area of focus lies in reconstructing invasion histories and examining genomic and genetic connections between native and expanded range populations. Reconstruction of invasion histories facilitates applied strategies for prevention and mediation of biological invasions, though the completion of

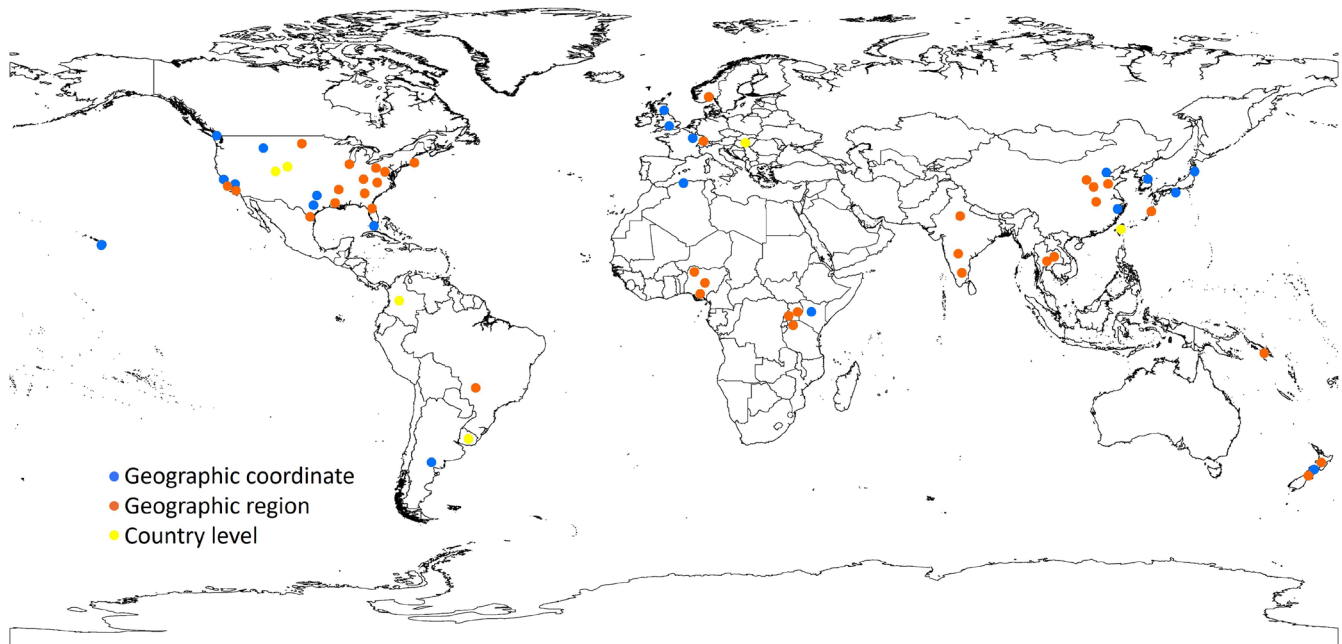


FIGURE 5 Level of geographic metadata provided for field-collected sample assemblies. Sampling sites for field-collected sample assemblies were reported at three different resolutions, as indicated by the provide key.

comprehensive spatiotemporal metadata is crucial for facilitating such transgenerational analysis (Estoup & Guillemaud, 2010). Completion of collection date or location metadata can allow researchers to trace gene flow among populations, potentially identifying the source of initial bridgehead populations (Vallejo-Marín et al., 2021), and/or infer whether single or sequential incursion events occurred to establish the invasive population(s) (Blumenfeld et al., 2021; Puckett et al., 2020). Completion of comprehensive metadata therefore allows for a better understanding of invasion dynamics.

Contrary to trends in spatiotemporal data, developmental stage and sex were well-represented in the reference genome and assembly datasets. Ninety-five percent of assembly data included the development stage, though only 57% of the reference genomes followed suit. Meanwhile 50% and 62% of the assembly and reference genome datasets stated the sex of the organism, respectively. Demonstrating the importance of these data fields, functional transcriptome analysis of adult female *Drosophila* spp., including the highly invasive *D. sukukii*, allowed a qualitative overview of gene expression across ovipositor shapes, the mechanism by which female *D. sukukii* can expand its host range (Crava et al., 2020). Thus, incorporating physiologically meaningful information into metadata fields could improve understanding of how genomic traits underpin physiological flexibility in invasive species. While these additional fields of metadata may have a less direct relevance for some applications (e.g. comparative genomics), they are important for the purposes of future reuse of data, and to adhere to the FAIR principle of interoperability (i.e. to set baseline standards of metadata collection regardless of experimental intent).

Seven species in our analysis had both native and expanded range representation, with temporal (93.4%), developmental stage (93.4%) and sex (80%) metadata provided (Boyes & Crowley, 2021; Carlson et al., 2022; Crowley, 2022; Ettinger et al., 2021; Guerrero et al., 2019; Harrop et al., 2020; Jia et al., 2020; Li et al., 2022; Lohse et al., 2021; Shen et al., 2016). Interestingly, none of these studies were conducted within a principle invasion context, though Ettinger et al. (2021) acknowledged that the sequencing of the invasive representative was a critical resource for management strategies in the glassy-winged sharpshooter, and this sentiment was reiterated by Harrop et al. (2020) for the common wasp. Of the studies conducted on these eight species (15 genome assemblies), 10 were within the last 5 years, emphasizing the recently improved accessibility of high-throughput sequencing and the promising trend of improving metadata quality we identified for the insects in our dataset. In particular, the number of assemblies that came from samples that were classified as field-collected increased from 29 to 65 between 2016 and 2021. A recent meta-analysis confirmed this trend, showing a 13.5% yearly decrease in the probability of retrieving missing metadata across multiple sequencing projects—indicating that new metadata is more likely to be complete (Crandall et al., 2023). Indeed, ten Hoopen et al. (2016) demonstrated that post-hoc manual sample curation was laborious and difficult to implement on a large scale. FAIR-adhering initiatives, such as the Genomic Observatories Metadatabase (GEOME; Deck et al., 2017), are a practical solution for authors to make metadata of previously published genomic and genetic datasets available retrospectively.

Perhaps unsurprisingly, the most consistent and thorough metadata submissions in our analysis came from the Wellcome Sanger

TABLE 1 Seven species for which both native and expanded range assemblies were available.

Order	Species	Common name	Range	GenBank accession	Reference	Useful insights from metadata and potential future uses
Hemiptera	<i>Homalodisca</i>	Glassy-winged sharpshooter	Native: Texas, United States Expanded: California, United States	GCA_021130785.2	(Li et al., 2022)	Confirmation of reference genome from invasive populations, future studies on population structure and development of nonbiological controls
	<i>vitripennis</i>			GCA_019364655.1	(Ettinger et al., 2021)	
Hymenoptera	<i>Vespula vulgaris</i>	Common wasp	Native: Wytham Great Wood, United Kingdom Expanded: Pelorus, New Zealand	GCA_905475345.1 GCA_0144466185.1	(Crowley, 2022) (Harrop et al., 2020)	Provided genomic targets for applied management. Comparisons between native/expanded populations would need to eliminate targets that can harm populations within the native range.
Lepidoptera	<i>Pieris rapae</i>	Small white butterfly	Native: West Linton, Scotland, United Kingdom Expanded: Texas, United States	GCA_905147795.1 GCA_001856805.1	(Lohse et al., 2021) (Shen et al., 2016)	Further insights of the species ecology by sequencing from the native range after Shen et al. (2016) sequenced individuals from an invasive population.
Ixodida	<i>Haemaphysalis longicornis</i>	Asian longhorned tick	Native: Shandong, China Expanded: New Zealand	GCA_013339765.2 GCA_008122185.1	(Jia et al., 2020) (Guerrero et al., 2019)	Study of pathogenesis related genes from a population that solely utilizes parthenogenetic reproduction.
Coleoptera	<i>Harmonia axyridis</i>	Harlequin ladybird	Native: Japan Expanded: Wytham Great Wood, United Kingdom	GCA_003402655.1 GCA_914767665.1	— (Boyes & Crowley, 2021)	
Hemiptera	<i>Diaphorina citri</i>	Asian citrus psyllid	Native: Taiwan Expanded: Uruguay Expanded: Los Angeles and California, United States	GCA_024506325.2 GCA_024506275.2 GCA_024506315.2	(Carlson et al., 2022) (Carlson et al., 2022) (Carlson et al., 2022)	Creation of three genomes from the same BioProject with the goal to compare between native and expanded ranges, demonstrating congruent results to known invasion histories. Temporal data provided context of population to known invasion history. Temporal data provided context of population to known invasion history.
Hymenoptera	<i>Vespula germanica</i>	European wasp	Native: Wytham Rough Common, United Kingdom Expanded: Lincoln, New Zealand	GCA_905340365.1 GCA_0144466195.1	(Crowley, 2022) (Harrop et al., 2020)	As with the common wasp, any method of targeted management would need to eliminate genomic targets that can harm populations within the native range.

Institute (Blaxter et al., 2022), where the code of practice for sampling specifically aims for the maximum reuse of each genome assembly, following FAIR principles (Wilkinson et al., 2016). These project submissions included geographic coordinates (and altitude in m), specific collection dates, and detailed sample notes on developmental stage and tissue type as per the Darwin Tree of Life's sample manifest (www.github.com/darwintreeoflife/metadata/). In invasion genomics, these metadata fields are especially useful in allowing users to determine the range type and invasion status of the population the individual was sampled from, as demonstrated in recent population genomic studies on invasive species (e.g. Parvizi et al., 2023; Schmidt et al., 2021).

As research questions develop and technology continues to evolve, the ability to use genomic tools to answer fundamental questions in invasion biology, such as identifying the genes responsible for an organism's mechanistic ability to invade, will increase (Gill et al., 2021). Our study supports calls for standardized metadata in genome assembly data submissions (Díaz et al., 2020; Toczydlowski et al., 2021) to facilitate this exciting future work. In particular, we have demonstrated that current metadata stewardship falls short of what is required for large-scale comparative studies in invasion biology. We suggest four recommendations that would improve metadata curation:

1. Mandatory minimum metadata requirements for spatiotemporal data, ideally to include decimal latitude/longitude, environmental medium/habitat, and date to mm/yyyy level. Current standards utilized by institutes such as Wellcome Sanger have stringent, comprehensive metadata requirements and would be appropriate for adoption by the wider community.
2. Standardized field codes for metadata input, with date and decimal latitude/longitude to be in uniform field descriptions, in addition to drop-down selections for wild collected, laboratory or colony strain origin states. Optional strain number, source or voucher accession number could also be valuable.
3. Greater inter-institute collaboration in defining baseline standards for metadata curation, with demonstrable and uniform adherence to FAIR principles. Establishment of working groups or active discussions at conferences and workshops to define these baselines would be a good starting point for establishing widespread metadata standards.
4. Authors should retrospectively enrich metadata curation associated with publicly available genome datasets where possible, using open access infrastructure such as GEOME, or any other FAIR-compliant open repositories.

While these improvements would have far-reaching benefits, in invasion research they would specifically allow for determination of a samples' status regarding field versus laboratory collection, pre-versus post-invasion, and native versus expanded range, facilitating broader comparative research and exciting new insights into the role of genome architecture in biological invasion.

AUTHOR CONTRIBUTIONS

All authors participated in study conceptualisation. ALV analysed the data and wrote the manuscript draft. All authors edited the manuscript and approved the final version.

ACKNOWLEDGEMENTS

This work was funded and supported by Genomics Aotearoa, a New Zealand Ministry of Business, Innovation and Employment funded research platform (Grant GA 2102 awarded to MKD and AM). Open access publishing facilitated by Landcare Research New Zealand, as part of the Wiley - Landcare Research New Zealand agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

No new genomic data were generated during the course of this study. All accession numbers and related SRA and BioProject numbers for each reference genome and assembly can be found in Supplementary File S1.

BENEFIT-SHARING STATEMENT

This study promotes international scientific partnership in the development of metadata standards for facilitation of broader future studies. Our research addresses limitations in the recording of whole genome sequencing metadata for invasive species and encourages the broader scientific community to initiate data standards conversations for the benefit of future studies.

ORCID

Amy L. Vaughan  <https://orcid.org/0000-0003-0309-8851>

Elahe Parvizi  <https://orcid.org/0000-0002-1695-8817>

Angela McGaughan  <https://orcid.org/0000-0002-3429-8699>

Manpreet K. Dhani  <https://orcid.org/0000-0002-8956-0674>

REFERENCES

- Abubakr, M., Sami, H., Mahdi, I., Althahir, O., Abdelbagi, H., Mohamed, N. S., & Ahmed, A. (2022). The phylodynamic and spread of the invasive Asian malaria vectors, *Anopheles stephensi*, in Sudan. *Biology (Basel)*, 11(3), Article 409. <https://doi.org/10.3390/biology11030409>
- Beaurepaire, A. L., Moro, A., Mondet, F., Le Conte, Y., Neumann, P., & Locke, B. (2019). Population genetics of ectoparasitic mites suggest arms race with honeybee hosts. *Scientific Reports*, 9(1), 11355. <https://doi.org/10.1038/s41598-019-47801-5>
- Behrman, E. L., Howick, V. M., Kapun, M., Staubach, F., Bergland, A. O., Petrov, D. A., Lazzaro, B. P., & Schmidt, P. S. (2018). Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. *Proceedings Biological Sciences*, 285(1870), Article 20172599. <https://doi.org/10.1098/rspb.2017.2599>
- Blaxter, M., Mieszkowska, N., Di Palma, F., Holland, P., Durbin, R., Richards, T., Berriman, M., Kersey, P., Hollingsworth, P., Wilson, W., Twyford, A., Gaya, E., Lawniczak, M., Lewis, O., Broad, G., Howe, K., Hart, M., Flicek, P., & Barnes, I. (2022). Sequence locally, think globally: The Darwin tree of life project. *Proceedings of the*

- National Academy of Sciences, 119(4), e2115642118. <https://doi.org/10.1073/pnas.2115642118>
- Blumenfeld, A. J., Eyer, P.-A., Husseneder, C., Mo, J., Johnson, L. N. L., Wang, C., Kenneth Grace, J., Chouvinc, T., Wang, S., & Vargo, E. L. (2021). Bridgehead effect and multiple introductions shape the global invasion history of a termite. *Communications Biology*, 4(1), 196. <https://doi.org/10.1038/s42003-021-01725-x>
- Boyes, D., & Crowley, L. (2021). The genome sequence of the harlequin ladybird, *Harmonia axyridis* (Pallas, 1773). *Wellcome Open Research*, 6(300). <https://doi.org/10.12688/wellcomeopenres.17349.1>
- Bradhurst, R., Spring, D., Stanaway, M., Milner, J., & Kompas, T. (2021). A generalised and scalable framework for modelling incursions, surveillance and control of plant and environmental pests. *Environmental Modelling & Software*, 139, 105004. <https://doi.org/10.1016/j.envsoft.2021.105004>
- CABI. (2023). *CABI Compendium*. <https://www.cabidigitallibrary.org/>
- Calla, B., Demkovich, M., Siegel, J. P., Viana, J. P. G., Walden, K. K. O., Robertson, H. M., & Berenbaum, M. R. (2020). Selective sweeps in a nutshell: The genomic footprint of rapid insecticide resistance evolution in the almond agroecosystem. *Genome Biology and Evolution*, 13(1), Article evaa234. <https://doi.org/10.1093/gbe/evaa234>
- Carlson, C. R., Ter Horst, A. M., Johnston, J. S., Henry, E., Falk, B. W., & Kuo, Y. W. (2022). High-quality, chromosome-scale genome assemblies: Comparisons of three *Diaphorina citri* (Asian citrus psyllid) geographic populations. *DNA Research*, 29(4), Article dsac027. <https://doi.org/10.1093/dnares/dsac027>
- Childers, A. K., Geib, S. M., Sim, S. B., Poelchau, M. F., Coates, B. S., Simmonds, T. J., Scully, E. D., Smith, T. P. L., Childers, C. P., Corpuz, R. L., Hackett, K., & Scheffler, B. (2021). The USDA-ARS Ag100pest initiative: High-quality Genome assemblies for agricultural pest arthropod research. *Insects*, 12(7), 626. <https://doi.org/10.3390/insects12070626>
- Crandall, E. D., Toczydlowski, R. H., Liggins, L., Holmes, A. E., Ghoojaei, M., Gaither, M. R., Wham, B. E., Pritt, A. L., Noble, C., Anderson, T. J., Barton, R. L., Berg, J. T., Beskid, S. G., Delgado, A., Farrell, E., Himmelsbach, N., Queeno, S. R., Trinh, T., Weyand, C., ... Toonen, R. J. (2023). The importance of timely metadata curation to the global surveillance of genetic diversity. *Conservation Biology*, 37, e14061. <https://doi.org/10.1111/cobi.14061>
- Crava, C. M., Zanini, D., Amati, S., Sollai, G., Crnjar, R., Paoli, M., Rossi-Stacconi, M. V., Rota-Stabelli, O., Tait, G., Haase, A., Romani, R., & Anfora, G. (2020). Structural and transcriptional evidence of mechanotransduction in the *Drosophila suzukii* ovipositor. *Journal of Insect Physiology*, 125, 104088. <https://doi.org/10.1016/j.jinphys.2020.104088>
- Crowley, L. M. (2022). The genome sequence of the German wasp, *Vespula germanica* (Fabricius, 1793). *Wellcome Open Research*, 7, 60. <https://doi.org/10.12688/wellcomeopenres.17703.1>
- Deck, J., Gaither, M. R., Ewing, R., Bird, C. E., Davies, N., Meyer, C., Riginos, C., Toonen, R. J., & Crandall, E. D. (2017). The Genomic Observatories Metadatabase (GeOME): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biology*, 15(8), e2002925. <https://doi.org/10.1371/journal.pbio.2002925>
- Dematteis, B., Ferrucci, M. S., Ortega-Baes, P., & Coulleri, J. P. (2020). Genome size variation between the native and invasive ranges of *Senecio madagascariensis* (Asteraceae). *Systematic Botany*, 45(1), 212–218. <https://doi.org/10.1600/036364420X15801369352487>
- Díaz, S., Zafra-Calvo, N., Purvis, A., Verburg, P. H., Obura, D., Leadley, P., Chaplin-Kramer, R., De Meester, L., Dulloo, E., Martín-López, B., Shaw, M. R., Visconti, P., Broadgate, W., Bruford, M. W., Burgess, N. D., Cavender-Bares, J., DeClerck, F., Fernández-Palacios, J. M., Garibaldi, L. A., ... Zanne, A. E. (2020). Set ambitious goals for biodiversity and sustainability. *Science*, 370(6515), 411–413. <https://doi.org/10.1126/science.abe1530>
- Durkin, S. M., Chakraborty, M., Abrieux, A., Lewald, K. M., Gadau, A., Svetec, N., Peng, J., Kopyto, M., Langer, C. B., Chiu, J. C., Emerson, J. J., & Zhao, L. (2021). Behavioral and genomic sensory adaptations underlying the pest activity of *Drosophila suzukii*. *Molecular Biology and Evolution*, 38(6), 2532–2546. <https://doi.org/10.1093/molbev/msab048>
- Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, 19(19), 4113–4130. <https://doi.org/10.1111/j.1365-294X.2010.04773.x>
- Ettinger, C. L., Byrne, F. J., Collin, M. A., Carter-House, D., Walling, L. L., Atkinson, P. W., Redak, R. A., & Stajich, J. E. (2021). Improved draft reference genome for the glassy-winged sharpshooter (*Homalodisca vitripennis*), a vector for Pierce's disease. *G3 (Bethesda)*, 11(10), Article jkab255. <https://doi.org/10.1093/g3journal/jkab255>
- Gentili, R., Schaffner, U., Martinoli, A., & Citterio, S. (2021). Invasive alien species and biodiversity: Impacts and management. *Biodiversity*, 22(1–2), 1–3. <https://doi.org/10.1080/14888386.2021.1929484>
- Gill, N. S., Mahood, A. L., Meier, C. L., Muthukrishnan, R., Nagy, R. C., Stricker, E., Duffy, K. A., Petri, L., & Morissette, J. T. (2021). Six central questions about biological invasions to which NEON data science is poised to contribute. *Ecosphere*, 12(9), e03728. <https://doi.org/10.1002/ecs2.3728>
- Guerrero, F. D., Bendele, K. G., Ghaffari, N., Guhlin, J., Gedye, K. R., Lawrence, K. E., Dearden, P. K., Harrop, T. W. R., Heath, A. C. G., Lun, Y., Metz, R. P., Teel, P., Perez de Leon, A., Biggs, P. J., Pomroy, W. E., Johnson, C. D., Blood, P. D., Bellgard, S. E., & Tompkins, D. M. (2019). The Pacific biosciences de novo assembled genome dataset from a parthenogenetic New Zealand wild population of the longhorned tick, *Haemaphysalis longicornis* Neumann, 1901. *Data in Brief*, 27, 104602. <https://doi.org/10.1016/j.dib.2019.104602>
- Gutiérrez, A. P., Ponti, L., Neteler, M., Suckling, D. M., & Cure, J. R. (2021). Invasive potential of tropical fruit flies in temperate regions under climate change. *Communications Biology*, 4(1), 1141. <https://doi.org/10.1038/s42003-021-02599-9>
- Harrop, T. W. R., Guhlin, J., McLaughlin, G. M., Permina, E., Stockwell, P., Gilligan, J., Le Lec, M. F., Gruber, M. A. M., Quinn, O., Lovegrove, M., Duncan, E. J., Remnant, E. J., Van Eeckhoven, J., Graham, B., Knapp, R. A., Langford, K. W., Kronenberg, Z., Press, M. O., Eacker, S. M., ... Dearden, P. K. (2020). High-quality assemblies for three invasive social wasps from the *Vespula*. *Genus G3 (Bethesda)*, 10(10), 3479–3488. <https://doi.org/10.1534/g3.120.401579>
- Hulme, P. E. (2021). Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth*, 4(5), 666–679. <https://doi.org/10.1016/j.oneear.2021.04.015>
- Hulme, P. E., Bacher, S., Kenis, M., Klotz, S., Kühn, I., Minchin, D., Nentwig, W., Olenin, S., Panov, V., Pergl, J., Pyšek, P., Roques, A., Sol, D., Solarz, W., & Vilà, M. (2008). Grasping at the routes of biological invasions: A framework for integrating pathways into policy. *Journal of Applied Ecology*, 45(2), 403–414. <https://doi.org/10.1111/j.1365-2664.2007.01442.x>
- i5K Consortium. (2013). The i5K initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, 104(5), 595–600. <https://doi.org/10.1093/jhered/est050>
- Jia, N., Wang, J., Shi, W., Du, L., Sun, Y., Zhan, W., Jiang, J.-F., Wang, Q., Zhang, B., Ji, P., Bell-Sakyi, L., Cui, X.-M., Yuan, T.-T., Jiang, B.-G., Yang, W.-F., Lam, T. T.-Y., Chang, Q.-C., Ding, S.-J., Wang, X.-J., ... Cao, W.-C. (2020). Large-scale comparative analyses of tick genomes elucidate their genetic diversity and vector capacities. *Cell*, 182(5), 1328–1340.e1313. <https://doi.org/10.1016/j.cell.2020.07.023>
- Lainhart, W., Bickersmith, S. A., Moreno, M., Rios, C. T., Vinetz, J. M., & Conn, J. E. (2015). Changes in genetic diversity from field to laboratory during colonization of *Anopheles darlingi* root (Diptera: Culicidae). *American Journal of Tropical Medicine and Hygiene*, 93(5), 998–1001. <https://doi.org/10.4269/ajtmh.15-0336>

- Latombe, G., Seebens, H., Lenzner, B., Courchamp, F., Dullinger, S., Golivets, M., Kühn, I., Leung, B., Roura-Pascual, N., Cebrian, E., Dawson, W., Diagne, C., Jeschke, J. M., Pérez-Granados, C., Moser, D., Turbelin, A., Viscconti, P., & Essl, F. (2022). Capacity of countries to reduce biological invasions. *Sustainability Science*, 18, 771–789. <https://doi.org/10.1007/s11625-022-01166-3>
- Lawniczak, M., Davey, R., Rajan, J., Pereira-da-Conceicao, L., Kiliyas, E., Hollingsworth, P., Barnes, I., Allen, H., Blaxter, M., Burgin, J., Broad, G., Crowley, L., Gaya, E., Holroyd, N., Lewis, O., McTaggart, S., Mieszkowska, N., Minotto, A., Shaw, F., & Sivess, L. (2022). Specimen and sample metadata standards for biodiversity genomics: A proposal from the Darwin tree of life project. *Wellcome Open Research*, 7, 187. <https://doi.org/10.12688/wellcomeopenres.17605.1>
- Lewald, K. M., Abrieux, A., Wilson, D. A., Lee, Y., Conner, W. R., Andreazza, F., Beers, E. H., Burrack, H. J., Daane, K. M., Diepenbrock, L., Drummond, F. A., Fanning, P. D., Gaffney, M. T., Hesler, S. P., Ioriatti, C., Isaacs, R., Little, B. A., Loeb, G. M., Miller, B., ... Chiu, J. C. (2021). Population genomics of *Drosophila suzukii* reveal longitudinal population structure and signals of migrations in and out of the continental United States. *G3 (Bethesda)*, 11(12), Article jkab343. <https://doi.org/10.1093/g3journal/jkab343>
- Lewin, H. A., Richards, S., Lieberman Aiden, E., Allende, M. L., Archibald, J. M., Bálint, M., Barker, K. B., Baumgartner, B., Belov, K., Bertorelle, G., Blaxter, M. L., Cai, J., Caperello, N. D., Carlson, K., Castilla-Rubio, J. C., Chaw, S.-M., Chen, L., Childers, A. K., Coddington, J. A., ... Zhang, G. (2022). The earth BioGenome project 2020: Starting the clock. *Proceedings of the National Academy of Sciences*, 119(4), e2115635118. <https://doi.org/10.1073/pnas.2115635118>
- Li, Z., Li, Y., Xue, A. Z., Dang, V., Holmes, V. R., Johnston, J. S., Barrick, J. E., & Moran, N. A. (2022). The genomic basis of evolutionary novelties in a leafhopper. *Molecular Biology and Evolution*, 39(9), Article msac184. <https://doi.org/10.1093/molbev/msac184>
- Lohse, K., Ebdon, S., & Vila, R. (2021). The genome sequence of the small white, *Pieris rapae* (Linnaeus, 1758). *Wellcome Open Research*, 6(273). <https://doi.org/10.12688/wellcomeopenres.17288.1>
- Lowe, S., Browne, M., Boudjelas, S., & De Poorter, M. (2000). *100 of the world's worst invasive alien species: A selection from the global invasive species database* (Vol. 12). Invasive Species Specialist Group Auckland.
- Lukicheva, S., & Mardulyn, P. (2021). Whole-genome sequencing reveals asymmetric introgression between two sister species of cold-resistant leaf beetles. *Molecular Ecology*, 30(16), 4077–4089. <https://doi.org/10.1111/mec.16011>
- Matheson, P., & McGaughan, A. (2022). Genomic data is missing for many highly invasive species, restricting our preparedness for escalating incursion rates. *Scientific Reports*, 12(1), 13987. <https://doi.org/10.1038/s41598-022-17937-y>
- McCartney, M. A., Mallez, S., & Gohl, D. M. (2019). Genome projects in invasion biology. *Conservation Genetics*, 20(6), 1201–1222. <https://doi.org/10.1007/s10592-019-01224-x>
- North, H. L., McGaughan, A., & Jiggins, C. D. (2021). Insights into invasive species from whole-genome resequencing. *Molecular Ecology*, 30(23), 6289–6308. <https://doi.org/10.1111/mec.15999>
- Parvizi, E., Dhami, M. K., Yan, J., & McGaughan, A. (2023). Population genomic insights into invasion success in a polyphagous agricultural pest, *Halyomorpha halys*. *Molecular Ecology*, 32(1), 138–151. <https://doi.org/10.1111/mec.16740>
- Poorter, M. D., Browne, M., Alford, D. V., & Backhaus, G. F. (2005). The global invasive species database (GISD) and international information exchange: Using global expertise to help in the fight against invasive alien species. In *Plant prot. plant health Europe: Introduction and spread of invasive species held Humboldt* (Vol. 2005, pp. 49–54). Univ.
- Puckett, E. E., Magnussen, E., Khlyap, L. A., Strand, T. M., Lundkvist, Å., & Munshi-South, J. (2020). Genomic analyses reveal three independent introductions of the invasive brown rat (*Rattus norvegicus*) to The Faroe Islands. *Heredity*, 124(1), 15–27. <https://doi.org/10.1038/s41437-019-0255-6>
- Renault, D., Angulo, E., Cuthbert, R. N., Haubrock, P. J., Capinha, C., Bang, A., Kramer, A. M., & Courchamp, F. (2022). The magnitude, diversity, and distribution of the economic costs of invasive terrestrial invertebrates worldwide. *Science of the Total Environment*, 835, 155391. <https://doi.org/10.1016/j.scitotenv.2022.155391>
- Rius, M., Bourne, S., Hornsby, H. G., & Chapman, M. A. (2015). Applications of next-generation sequencing to the study of biological invasions. *Current Zoology*, 61(3), 488–504. <https://doi.org/10.1093/czoolo/61.3.488>
- Runquist, R. D. B., Gorton, A. J., Yoder, J. B., Deacon, N. J., Grossman, J. J., Kothari, S., Lyons, M. P., Sheth, S. N., Tiffin, P., & Moeller, D. A. (2020). Context dependence of local adaptation to abiotic and biotic environments: A quantitative and qualitative synthesis. *The American Naturalist*, 195(3), 412–431. <https://doi.org/10.1086/707322>
- Schmidt, T. L., Endersby-Harshman, N. M., Kurucz, N., Pettit, W., Krause, V. L., Ehlers, G., Muzari, M. O., Currie, B. J., & Hoffmann, A. A. (2023). Genomic databanks provide robust assessment of invasive mosquito movement pathways and cryptic establishment. *Biological Invasions*. <https://doi.org/10.1007/s10530-023-03117-0>
- Schmidt, T. L., Swan, T., Chung, J., Karl, S., Demok, S., Yang, Q., Field, M. A., Muzari, M. O., Ehlers, G., Brugh, M., Bellwood, R., Horne, P., Burkot, T. R., Ritchie, S., & Hoffmann, A. A. (2021). Spatial population genomics of a recent mosquito invasion. *Molecular Ecology*, 30(5), 1174–1189. <https://doi.org/10.1111/mec.15792>
- Seebens, H., Bacher, S., Blackburn, T. M., Capinha, C., Dawson, W., Dullinger, S., Genovesi, P., Hulme, P. E., van Kleunen, M., Kühn, I., Jeschke, J. M., Lenzner, B., Liebhold, A. M., Pattison, Z., Pergl, J., Pyšek, P., Winter, M., & Essl, F. (2021). Projecting the continental accumulation of alien species through to 2050. *Global Change Biology*, 27(5), 970–982. <https://doi.org/10.1111/gcb.15333>
- Seebens, H., Blackburn, T. M., Dyer, E. E., Genovesi, P., Hulme, P. E., Jeschke, J. M., Pagad, S., Pyšek, P., van Kleunen, M., Winter, M., Ansong, M., Arianoutsou, M., Bacher, S., Blasius, B., Brockerhoff, E. G., Brundu, G., Capinha, C., Causton, C. E., Celesti-Grapow, L., ... Essl, F. (2018). Global rise in emerging alien species results from increased accessibility of new source pools. *Proceedings of the National Academy of Sciences*, 115(10), E2264–E2273. <https://doi.org/10.1073/pnas.1719429115>
- Shen, J., Cong, Q., Kinch, L., Borek, D., Otwinowski, Z., & Grishin, N. (2016). Complete genome of *Pieris rapae*, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. *F1000Research*, 5(2631). <https://doi.org/10.12688/f1000research.9765.1>
- Siepielski, A. M., & Beaulieu, J. M. (2017). Adaptive evolution to novel predators facilitates the evolution of damselfly species range shifts. *Evolution*, 71(4), 974–984. <https://www.jstor.org/stable/48577377>
- Stuart, K. C., Hofmeister, N. R., Zichello, J. M., & Rollins, L. A. (2023). Global invasion history and native decline of the common starling: Insights through genetics. *Biological Invasions*, 25, 1291–1316. <https://doi.org/10.1007/s10530-022-02982-5>
- Stuart, K. C., Sherwin, W. B., Austin, J. J., Bateson, M., Eens, M., Brandley, M. C., & Rollins, L. A. (2022). Historical museum samples enable the examination of divergent and parallel evolution during invasion. *Molecular Ecology*, 31(6), 1836–1852. <https://doi.org/10.1111/mec.16353>
- Tahsin, T., Weissenbacher, D., Jones-Shargani, D., Magee, D., Vaiente, M., Gonzalez, G., & Scotch, M. (2017). Named entity linking of geospatial and host metadata in GenBank for advancing biomedical research. *Database: The Journal of Biological Databases and Curation*, 2017, Article bax093. <https://doi.org/10.1093/database/bax093>

- ten Hoopen, P., Amid, C., Luigi Buttigieg, P., Pafilis, E., Bravakos, P., Cerdeño-Tárraga, A. M., Gibson, R., Kahlke, T., Legaki, A., Narayana Murthy, K., Papastefanou, G., Pereira, E., Rossello, M., Luisa Toribio, A., & Cochrane, G. (2016). Value, but high costs in post-deposition data curation. *Database*, 2016, Article bav126. <https://doi.org/10.1093/database/bav126>
- Toczydlowski, R. H., Liggins, L., Gaither, M. R., Anderson, T. J., Barton, R. L., Berg, J. T., Beskid, S. G., Davis, B., Delgado, A., Farrell, E., Ghoojaei, M., Himmelsbach, N., Holmes, A. E., Queeno, S. R., Trinh, T., Weyand, C. A., Bradburd, G. S., Riginos, C., Toonen, R. J., & Crandall, E. D. (2021). Poor data stewardship will hinder global genetic diversity surveillance. *Proceedings of the National Academy of Sciences*, 118(34), e2107934118. <https://doi.org/10.1073/pnas.2107934118>
- Turner, K. G., Ostevik, K. L., Grassa, C. J., & Rieseberg, L. H. (2021). Genomic analyses of phenotypic differences between native and invasive populations of diffuse knapweed (*Centaurea diffusa*). *Frontiers in Ecology and Evolution*, 8. <https://doi.org/10.3389/fevo.2020.577635>
- Vallejo-Marín, M., Friedman, J., Twyford, A. D., Lepais, O., Ickert-Bond, S. M., Streisfeld, M. A., Yant, L., van Kleunen, M., Rotter, M. C., & Puzey, J. R. (2021). Population genomic and historical analysis suggests a global invasion by bridgehead processes in *Mimulus guttatus*. *Communications Biology*, 4(1), 327. <https://doi.org/10.1038/s42003-021-01795-x>
- Wardhaugh, C. W., Stone, M. J., & Stork, N. E. (2018). Seasonal variation in a diverse beetle assemblage along two elevational gradients in the Australian wet tropics. *Scientific Reports*, 8(1), 8559. <https://doi.org/10.1038/s41598-018-26216-8>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zhang, L., Rohr, J., Cui, R., Xin, Y., Han, L., Yang, X., Gu, S., Du, Y., Liang, J., Wang, X., Wu, Z., Hao, Q., & Liu, X. (2022). Biological invasions facilitate zoonotic disease emergences. *Nature Communications*, 13(1), 1762. <https://doi.org/10.1038/s41467-022-29378-2>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Vaughan, A. L., Parvizi, E., Matheson, P., McGaughran, A., & Dhami, M. K. (2025). Current stewardship practices in invasion biology limit the value and secondary use of genomic data. *Molecular Ecology Resources*, 25, e13858. <https://doi.org/10.1111/1755-0998.13858>