SUPPLEMENTARY MATERIAL

African Entomology, 2024

Review of DNA barcoding alien invertebrates and biological control agents in South Africa

Tristan Pitcher, Abusisiwe Ndaba, Adriaana Jacobs, Michelle Hamer, Charlene Janion-Scheepers

Table S1. The representation of South African invertebrates on BOLD in 2014 and 2023.

Group	Number of known species in SA[b]	2014[a]			2023			Increase in number of records (%)	Increase in number of BINs (%)	Increase in number of species (%)
		Number of records	Number of BINs	Number of species	Number of records	Number of BINs	Number of species			
Insects	44088	37105	8238	513	733499	45420	5378	1876.8	451.3	948.3
Other	4195	-	-	-	65389	3540	9406	-	-	-
Arachnids	6630	1721	555	7	29919	2884	769	1638.5	419.6	10885.7
Molluscs	3154	734	214	22	2597	433	355	253.8	102.3	1513.6
Annelids	1110	357	99	3	1125	169	108	215.1	70.7	3500.0
Platyhelminthes	246	10	6	1	317	18	70	3070.0	200.0	6900.0
Cnidarians	932	15	3	1	112	15	19	646.7	400.0	1800.0
Echinoderms	410	197	6	2	842	122	104	327.4	1933.3	5100.0
Sponges	346	8	6	1	77	12	38	862.5	100.0	3700.0
Total	61111	40147	9127	550	833877	52613	16247	1977.1	476.5	2854.0

[a] da Silva & Willows-Munro (2016)

[b] Hamer (Hamer 2013)

SA = South Africa

Table S2. The number of invertebrate alien and biocontrol agent species represented on BOLD by phylum.

		Alien			Biocontrol agent		
Phylum	**Class**	**Number of species**	**Number of species in BOLD**	**Percentage of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Percentage of species in BOLD**
Arthropoda	Insecta	441	329	74.6	128	56	43.8
	Arachnida	97	50	51.5	4	0	0.0
	Malacostraca	54	42	77.8	0	0	-
	Collembola	27	25	92.6	0	0	-
	Diplopoda	9	6	66.7	0	0	-
	Maxillopoda	4	3	75.0	0	0	-
	Chilopoda	3	2	66.7	0	0	-
	Copepoda	2	2	100.0	0	0	-
	Branchiopoda	1	1	100.0	0	0	-
	Pycnogonida	1	1	100.0	0	0	-
	Thecostraca	1	0	0.0	0	0	-
Mollusca	Gastropoda	67	51	76.1	0	0	-
	Bivalvia	13	10	76.9	0	0	-
Annelida	Clitellata	55	39	70.9	0	0	-
	Polychaeta	13	7	53.8	0	0	-
	Unknown	1	0	0.0	0	0	-
Nematoda	Chromadorea	21	15	71.4	0	0	-
	Unknown	8	0	0.0	0	0	-
Platyhelminthes	Monogenea	13	2	15.4	0	0	-
	Cestoda	3	3	100.0	0	0	-
	Turbellaria	3	3	100.0	0	0	-
	Unknown	2	0	0.0	0	0	-
	Trematoda	1	0	0.0	0	0	-
Cnidaria	Hydrozoa	12	8	66.7	0	0	-
	Anthozoa	3	3	100.0	0	0	-
Chordata	Ascidiacea	9	8	88.9	0	0	-
Bryozoa	Gymnolaemata	7	4	57.1	0	0	-
Echinodermata	Asteroidea	2	2	100.0	0	0	-
	Echinoidea	1	1	100.0	0	0	-
	Ophiuroidea	1	1	100.0	0	0	-
Brachiopoda	Lingulata	2	0	0.0	0	0	-
Ctenophora	Tentaculata	1	1	100.0	0	0	-
Euglenozoa	Kinetoplastea	1	0	0.0	0	0	-
Myzozoa	Dinophyceae	1	1	100.0	0	0	-
Porifera	Demospongiae	1	1	100.0	0	0	-
Total		881	621	70.5	132	56	42.4

Table S3. The number of invertebrate alien groups per province. Bold values indicate the highest provincial abundance of a taxon.

Group	**Eastern Cape**		**Free State**		**Gauteng**		**KwaZulu-Natal**		**Limpopo**		**Mpumalanga**		**Northern Cape**		**Northwest**		**Prince Edward Islands**		**Western Cape**	
	Number of species	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**	**Number of species**	**Number of species in BOLD**
Insects	34	28	23	21	**81**	**74**	58	50	45	40	55	52	15	13	16	14	13	8	72	59
Other	21	19	9	7	6	6	19	15	3	2	8	6	15	11	5	4	5	5	**53**	**45**
Arachnids	13	7	17	7	13	11	22	11	19	11	14	8	12	9	8	4	7	2	**26**	**12**
Molluscs	19	16	4	3	9	8	11	10	4	4	3	3	8	8	3	3	1	1	**32**	**28**
Annelids	23	20	10	7	16	13	37	30	19	14	10	9	4	3	11	9	0	0	**27**	**24**
Platyhelminthes	1	1	0	0	**2**	**2**	1	1	0	0	0	0	0	0	**2**	**2**	0	0	1	1
Cnidarians	**3**	2	0	0	0	0	**3**	2	0	0	0	0	2	2	0	0	0	0	**3**	**3**
Echinoderms	0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0
Sponges	0	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0	0	0
Total	114	93	63	45	127	114	152	120	90	71	90	78	57	47	45	36	26	16	214	172

R code

```
install.packages("ggplot2")
install.packages("tidyr")
install.packages("tidyverse")
install.packages("dplyr")
install.packages("treemap")
install.packages("rgbif")
install.packages("bold")
install.packages("ggpubr")
install.packages("tibble")
install.packages("remotes")
remotes::install_github("mrdwab/mrdwabmisc")

library(ggplot2)
library(tidyr)
library(tidyverse)
library(dplyr)
library(treemap)
library(rgbif)
library(bold)
library(ggpubr)
library(tibble)
library(remotes)

setwd("C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding Alien Invertebrates")

inverts <- read.csv("alien_inverts_CURRENT.csv")

#### CREATE ALIEN DA SILVA DATAFRAME ####

insects <-
    inverts %>%
    filter(is.na(legallyImported), class == "Insecta") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Insects", .before = 1)
arachnids <-
    inverts %>%
    filter(is.na(legallyImported), class == "Arachnida") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Arachnids", .before = 1)
molluscs <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Mollusca") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Molluscs", .before = 1)
crustaceans <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Crustacea") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Crustaceans", .before = 1)
annelids <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Annelida") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Annelids", .before = 1)
cnidarians <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Cnidaria") %>%
```

```
    select(-c(phylum, class)) %>%
    add_column(Group = "Cnidarians", .before = 1)
echinoderms <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Echinodermata") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Echinoderms", .before = 1)
sponges <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Porifera") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Sponges", .before = 1)
platyhelminths <-
    inverts %>%
    filter(is.na(legallyImported), phylum == "Platyhelminthes") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Platyhelminths", .before = 1)
other <- inverts %>%
    filter(is.na(legallyImported),
           class != "Insecta",
           class != "Arachnida",
           phylum != "Mollusca",
           phylum != "Crustacea",
           phylum != "Annelida",
           phylum != "Cnidaria",
           phylum != "Echinodermata",
           phylum != "Porifera",
           phylum != "Platyhelminthes") %>%
    select(-c(phylum, class)) %>%
    add_column(Group = "Other", .before = 1)

invertsdaSilvaAlien <- rbind(insects,
                        arachnids,
                        molluscs,
                        crustaceans,
                        annelids,
                        cnidarians,
                        echinoderms,
                        sponges,
                        platyhelminths,
                        other)

invertsdaPlot <- invertsdaSilvaAlien %>%
    group_by(Group) %>%
    summarise(n=n()) %>%
    arrange(desc(n))

## plot da Silva
ggbarplot(data=invertsdaPlot, x="Group", y="n",
          width=0.8, fill="black",
          xlab="Group", ylab=FALSE) +
    rotate_x_text(angle=45)


#### GET TAXON NAMES W/ GBIF ####

for(i in 1:nrow(inverts2019)){
    tryCatch({
        bb <- name_backbone(name=inverts2019[i,"species"])
```

```
        if(max(dim(bb))!=-Inf){
            inverts2019[i,"phylum"] <- bb[1,"phylum"]
            inverts2019[i,"order"] <- bb[1,"order"]
            inverts2019[i,"family"] <- bb[1,"family"]
        }else{
            bbg <- name_backbone(name=inverts2019[i,"genus"])
            if(max(dim(bbg))!=-Inf){
                inverts2019[i,"phylum"] <- bbg[1,"phylum"]
                inverts2019[i,"order"] <- bbg[1,"order"]
                inverts2019[i,"family"] <- bbg[1,"family"]
            }
        }
    }, error = function(e){
        message(paste("An error occured in run #", i, ":\n"), e)
    })
}

## cross-check inverts2019 against inverts
inverts2019 <- inverts2019 %>%
    mutate(crosscheck="",.after=species)

for(i in 1:nrow(inverts2019)){
    if(inverts2019[i,"species"] %in% inverts$Species){
        inverts2019[i,"crosscheck"] <- "MATCH"
    }
}

write.csv(inverts2019, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding Alien
      Invertebrates\\alien_inverts_SANBI_CIB_2019_revised.csv")

## cross-check inverts against inverts2019
inverts <- inverts %>%
    mutate(crosscheck="",.after=Species)

for(i in 1:nrow(inverts)){
    if(inverts[i,"Species"] %in% inverts2019$species){
        inverts[i,"crosscheck"] <- "MATCH"
    }
}

write.csv(inverts, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding Alien
      Invertebrates\\alien_inverts_checked_revised.csv")


#### MINE BOLD FOR RECORD METADATA ####

for(i in 1:nrow(inverts)){
    tryCatch({
        records <- bold_specimens(taxon=inverts[i,"species"])
        if(max(dim(records))!=-Inf){
            ## boldRecords
            inverts[i,"boldRecords"] <- nrow(records)
            ## SAboldRecords
            inverts[i,"SAboldRecords"] <- nrow(records %>%
                                              filter(country=="South Africa"))
            ## BINs
            inverts[i,"BINs"] <- length(unique((records %>%
                                     filter(bin_uri!=""))$bin_uri))
            ## SABINs
```

```
            inverts[i,"SABINs"] <- length(unique((records %>%
                                                    filter(country=="South Africa",
      bin_uri!=""))$bin_uri))
            ## boldImages
            inverts[i,"boldImages"] <- nrow(records %>%
                                          filter(image_ids!=""))
            ## SAboldImages
            inverts[i,"SAboldImages"] <- nrow(records %>%
                                                filter(country=="South Africa", image_ids!=""))
            ## province
            provinceBOLD <- unique(records %>%
                                       filter(country=="South Africa") %>%
                                       select(province_state))
            if(max(dim(provinceBOLD))!=-Inf){
                inverts[i, "provincialDistribution"] <-
                    paste(c(inverts[i, "provincialDistribution"]), as.character(provinceBOLD),
      collapse=", ")
            }
        }
    }, error = function(e){
            message(paste("An error occured in run #", i, ":\n"), e)
        }
    )
}

write.csv(inverts, paste0("C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\alien_inverts_BOLDED_", Sys.Date(), ".csv"))

#### PHYLA TABLE & PLOT ####

## count the number of barcoded specimens per phylum
alienPhyla1 <- inverts %>%
    filter(is.na(legallyImported)) %>%
    group_by(phylum) %>%
    summarise(no_species=n())

alienPhyla2 <- inverts %>%
    filter(is.na(legallyImported)) %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    group_by(phylum) %>%
    summarise(no_bold_records=n())

alienPhylaTable <- left_join(alienPhyla1, alienPhyla2) %>%
    arrange(desc(no_species)) %>%
    replace(is.na(.), 0)

alienPhylaTable <- rbind(alienPhylaTable, data.frame(phylum="Total",
      t(colSums(alienPhylaTable[,-1]))))

write.csv(alienPhylaTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\alienPhylaTable.csv")

## BCA phyla table
bcaPhyla1 <- inverts %>%
    filter(legallyImported=="biocontrol") %>%
    group_by(phylum) %>%
    summarise(no_species=n())

bcaPhyla2 <- inverts %>%
```

```
    filter(legallyImported=="biocontrol") %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    group_by(phylum) %>%
    summarise(no_bold_records=n())

bcaPhylaTable <- left_join(bcaPhyla1, bcaPhyla2) %>%
    arrange(desc(no_species)) %>%
    replace(is.na(.), 0)

bcaPhylaTable <- rbind(bcaPhylaTable, data.frame(phylum="Total", t(colSums(bcaPhylaTable[,-
      1]))))

write.csv(bcaPhylaTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\bcaPhylaTable.csv")

## PHYLA PLOT

## alien phyla plot
ggbarplot(data=alienPhylaTable, x="phylum", y="n",
          width=0.8, fill="black",
          xlab="Phylum", ylab=FALSE) +
    rotate_x_text(angle=45)

## alien pie plot
ggplot(alienPhylaTable, aes(x="", y=n, fill=phylum)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()


#### CLASS TABLE ####

## alien class table
alienClassTable1 <- inverts %>%
    filter(is.na(legallyImported)) %>%
    group_by(phylum) %>%
    group_by(class, add=TRUE) %>%
    summarise(no_species=n())

alienClassTable2 <- inverts %>%
    filter(is.na(legallyImported)) %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    group_by(class) %>%
    summarise(no_bold_records=n())

alienClassTable <- left_join(alienClassTable1, alienClassTable2) %>%
    arrange(desc(no_species)) %>%
    replace(is.na(.), 0)

write.csv(alienClassTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\alienClassTable.csv")

## BCA class table
bcaClassTable1 <- inverts %>%
    filter(legallyImported=="biocontrol") %>%
    group_by(phylum) %>%
    group_by(class, add=TRUE) %>%
    summarise(no_species=n())
```

```
bcaClassTable2 <- inverts %>%
    filter(legallyImported=="biocontrol") %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    group_by(class) %>%
    summarise(no_bold_records=n())

bcaClassTable <- left_join(bcaClassTable1, bcaClassTable2) %>%
    arrange(desc(no_species)) %>%
    replace(is.na(.), 0)

write.csv(bcaClassTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\bcaClassTable.csv")

#### ARTHROPOD CLASS PLOT ####

plotData <- inverts %>%
    filter(phylum=="Arthropoda", is.na(legallyImported)) %>%
    group_by(class) %>%
    summarise(n=n()) %>%
    arrange(desc(n))

## alien arthropod class plot
ggbarplot(data=plotData, x="class", y="n",
          width=0.8, fill="black",
          xlab="Class", ylab=FALSE) +
    rotate_x_text(angle=45)

## pie chart of arthropod classes
ggplot(plotData, aes(x="", y=n, fill=class)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

#### INSECT ORDER PLOT ####

plotData <- inverts %>%
    filter(class=="Insecta", is.na(legallyImported)) %>%
    group_by(order) %>%
    summarise(n=n()) %>%
    arrange(desc(n))

## alien insect order plot
ggbarplot(data=plotData, x="order", y="n",
          width=0.8, fill="black",
          xlab="Order", ylab=FALSE) +
    rotate_x_text(angle=45)

## alien insect pie plot
ggplot(plotData, aes(x="", y=n, fill=order)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## INSECT ORDER METRICS PLOT
no_of_species <- inverts %>%
    filter(class=="Insecta", is.na(legallyImported)) %>%
    group_by(order) %>%
    summarise(no.of.species=n())
```

```
records <- inverts %>%
    filter(class=="Insecta", is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    summarise(records=n())

BINs <- inverts %>%
    filter(class=="Insecta", is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(BINs), BINs!=0) %>%
    summarise(BINs=n())

images <- inverts %>%
    filter(class=="Insecta", is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(boldImages), boldImages!=0) %>%
    summarise(images=n())

daSilvaBOLD <- data.frame(group = rep(records$Group,4),
                          metric = c(rep("species",9), rep("records",9), rep("BINs",9),
      rep("images",9)),
                          frequency = c(no_of_species$no.of.species, records$records,
      BINs$BINs, images$images))
daSilvaBOLD$metric = factor(daSilvaBOLD$metric, levels=c("species", "records", "BINs",
      "images"))

ggbarplot(data=daSilvaBOLD, x="group", y="frequency", fill="metric",
          position=position_dodge(0.7),
          xlab=FALSE, ylab=FALSE,
          order = c("Insects", "Other", "Arachnids", "Molluscs", "Annelids", "Platyhelminths",
      "Cnidarians", "Echinoderms", "Sponges")) +
    theme(legend.title=element_blank()) +
    rotate_x_text(angle = 45)

#### HEMIPTERA FAMILY PLOT ####

plotData <- inverts %>%
    filter(order=="Hemiptera", is.na(legallyImported)) %>%
    group_by(family) %>%
    summarise(n=n()) %>%
    arrange(desc(n))

## alien hemipteran family plot
ggbarplot(data=plotData, x="family", y="n",
          width=0.8, fill="black",
          xlab="Family", ylab=FALSE) +
    rotate_x_text(angle=45)

## alien hemipteran family pie plot
ggplot(plotData, aes(x="", y=n, fill=family)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()




#### BOLD QUALITY ANALYSES ####
```

```
## count number of BINs
barcodeTable <- data.frame(group=c(rep("alien invertebrates",6),rep("biological control
      agents",6)),
                           metric=rep(c("records", "SA records", "BINs", "SA BINs", "images",
      "SA images"),2),
                           frequency=0)

## count alien bold records
barcodeTable[1,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(boldRecords), boldRecords!=0) %>%
    summarise(n=n())
## count alien SA bold records
barcodeTable[2,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(SAboldRecords), SAboldRecords!=0) %>%
    summarise(n=n())
## calculate international records
barcodeTable[1,3] <- barcodeTable[1,3] - barcodeTable[2,3]
## count alien BINs
barcodeTable[3,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(BINs), BINs!=0) %>%
    summarise(n=n())
## count alien SA BINs
barcodeTable[4,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(SABINs), SABINs!=0) %>%
    summarise(n=n())
## calculate international BINs
barcodeTable[3,3] <- barcodeTable[3,3] - barcodeTable[4,3]
## count alien images
barcodeTable[5,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(boldImages), boldImages!=0) %>%
    summarise(n=n())
## count alien SA images
barcodeTable[6,3] <- inverts %>%
    filter(is.na(legallyImported), !is.na(SAboldImages), SAboldImages!=0) %>%
    summarise(n=n())
## calculate international images
barcodeTable[5,3] <- barcodeTable[5,3] - barcodeTable[6,3]

## count BCA bold records
barcodeTable[7,3] <- (inverts %>%
    filter(legallyImported=="biocontrol", !is.na(boldRecords), boldRecords!=0) %>%
    summarise(n=n()))
## count BCA SA bold records
barcodeTable[8,3] <- inverts %>%
    filter(legallyImported=="biocontrol", !is.na(SAboldRecords), SAboldRecords!=0) %>%
    summarise(n=n())
## calculate international records
barcodeTable[7,3] <- barcodeTable[7,3] - barcodeTable[8,3]
## count BCA BINs
barcodeTable[9,3] <- inverts %>%
    filter(legallyImported=="biocontrol", !is.na(BINs), BINs!=0) %>%
    summarise(n=n())
## count BCA SA BINs
barcodeTable[10,3] <- inverts %>%
    filter(legallyImported=="biocontrol", !is.na(SABINs), SABINs!=0) %>%
    summarise(n=n())
## calculate international BINs
barcodeTable[9,3] <- barcodeTable[9,3] - barcodeTable[10,3]
## count BCA images
```

```
barcodeTable[11,3] <- inverts %>%
    filter(legallyImported=="biocontrol", !is.na(boldImages), boldImages!=0) %>%
    summarise(n=n())
## count BCA SA images
barcodeTable[12,3] <- inverts %>%
    filter(legallyImported=="biocontrol", !is.na(SAboldImages), SAboldImages!=0) %>%
    summarise(n=n())
## calculate international images
barcodeTable[11,3] <- barcodeTable[11,3] - barcodeTable[12,3]

barcodeTable <- barcodeTable %>%
    mutate(metric_=rep(c("records", "records", "BINs", "BINs", "images", "images"), 2),
      .before=metric) %>%
    mutate(origin=rep(c("international", "local"), 6), .after=group)

write.csv(barcodeTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\barcodeTable.csv")

## plot BOLD quality metrics
ggbarplot(data=barcodeTable, x="metric_", y="frequency", fill="origin",
          position=position_dodge(0.7), facet.by = "group",
          xlab=FALSE, ylab=FALSE) +
    scale_y_continuous(breaks=c(0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600))
    theme(legend.title=element_blank())

## DA SILVA BOLD QUALITY METRICS PLOT
no_of_species <- invertsdaSilvaAlien %>%
    group_by(Group) %>%
    summarise(no.of.species=n())

records <- invertsdaSilvaAlien %>%
    group_by(Group) %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    summarise(records=n())

BINs <- invertsdaSilvaAlien %>%
    group_by(Group) %>%
    filter(!is.na(BINs), BINs!=0) %>%
    summarise(BINs=n())

images <- invertsdaSilvaAlien %>%
    group_by(Group) %>%
    filter(!is.na(boldImages), boldImages!=0) %>%
    summarise(images=n()) %>%
    add_row(Group = "Sponges", images=0)

daSilvaBOLD <- data.frame(group = rep(records$Group,4),
                          metric = c(rep("species",9), rep("records",9), rep("BINs",9),
      rep("images",9)),
                          frequency = c(no_of_species$no.of.species, records$records,
      BINs$BINs, images$images))
daSilvaBOLD$metric = factor(daSilvaBOLD$metric, levels=c("species", "records", "BINs",
      "images"))

ggbarplot(data=daSilvaBOLD, x="group", y="frequency", fill="metric",
          position=position_dodge(0.7),
          xlab=FALSE, ylab=FALSE,
          order = c("Insects", "Other", "Arachnids", "Molluscs", "Annelids", "Platyhelminths",
      "Cnidarians", "Echinoderms", "Sponges")) +
```

```
    theme(legend.title=element_blank()) +
    rotate_x_text(angle = 45)

## INSECT ORDER BOLD QUALITYMETRICS PLOT

no_of_species_IO <- inverts %>%
    filter(class=="Insecta") %>%
    filter(is.na(legallyImported)) %>%
    group_by(order) %>%
    summarise(no.of.species=n())

records_IO <- inverts %>%
    filter(class=="Insecta") %>%
    filter(is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(boldRecords), boldRecords!=0) %>%
    summarise(records=n())

BINs_IO <- inverts %>%
    filter(class=="Insecta") %>%
    filter(is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(BINs), BINs!=0) %>%
    summarise(BINs=n())

images_IO <- inverts %>%
    filter(class=="Insecta") %>%
    filter(is.na(legallyImported)) %>%
    group_by(order) %>%
    filter(!is.na(boldImages), boldImages!=0) %>%
    summarise(images=n())

alienInsectsBOLD <- data.frame(group = c(no_of_species_IO$order, records_IO$order,
      BINs_IO$order, images_IO$order),
                          metric = c(rep("species",nrow(no_of_species_IO)),
      rep("records",nrow(records_IO)), rep("BINs",nrow(BINs_IO)),
      rep("images",nrow(images_IO))),
                          frequency = c(no_of_species_IO$no.of.species, records_IO$records,
      BINs_IO$BINs, images_IO$images))
alienInsectsBOLD$metric = factor(alienInsectsBOLD$metric, levels=c("species", "records",
      "BINs", "images"))

ggbarplot(data=alienInsectsBOLD, x="group", y="frequency", fill="metric",
          position=position_dodge(0.7),
          xlab=FALSE, ylab=FALSE,
          order = c("Hemiptera", "Coleoptera", "Diptera", "Lepidoptera", "Hymenoptera",
      "Thysanoptera", "Psocodea", "Blattodea", "Dermaptera", "Siphonaptera", "Zygentoma",
      "Phthiraptera", "Orthoptera", "Embioptera", "Phasmida")) +
    theme(legend.title=element_blank()) +
    rotate_x_text(angle = 45) +
    scale_y_continuous(breaks=seq(0,150,10))

#### PROVINCIAL PLOT ####

ECsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "EC")|str_detect(provincialDistribution, "Eastern
      Cape")) %>%
    group_by(Group) %>%
    summarise(n=n())
```

```
ECbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "EC")|str_detect(provincialDistribution, "Eastern
      Cape")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

EC <- left_join(ECsp, ECbold)

NCsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "NC")|str_detect(provincialDistribution, "Northern
      Cape")) %>%
    group_by(Group) %>%
    summarise(n=n())

NCbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "NC")|str_detect(provincialDistribution, "Northern
      Cape")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

NC <- left_join(NCsp, NCbold)

KZNsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "KZN")|str_detect(provincialDistribution, "KwaZulu-
      Natal")) %>%
    group_by(Group) %>%
    summarise(n=n())

KZNbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "KZN")|str_detect(provincialDistribution, "KwaZulu-
      Natal")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

KZN <- left_join(KZNsp, KZNbold)

GPsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "GP")|str_detect(provincialDistribution, "Gauteng")) %>%
    group_by(Group) %>%
    summarise(n=n())

GPbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "GP")|str_detect(provincialDistribution, "Gauteng")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

GP <- left_join(GPsp, GPbold)

FSsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "FS")|str_detect(provincialDistribution, "Free State"))
      %>%
    group_by(Group) %>%
    summarise(n=n())
```

```
FSbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "FS")|str_detect(provincialDistribution, "Free State"))
      %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

FS <- left_join(FSsp, FSbold)

MPsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "MP")|str_detect(provincialDistribution, "Mpumalanga"))
      %>%
    group_by(Group) %>%
    summarise(n=n())

MPbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "MP")|str_detect(provincialDistribution, "Mpumalanga"))
      %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

MP <- left_join(MPsp, MPbold)

WCsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "WC")|str_detect(provincialDistribution, "Western
      Cape")) %>%
    group_by(Group) %>%
    summarise(n=n())

WCbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "WC")|str_detect(provincialDistribution, "Western
      Cape")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

WC <- left_join(WCsp, WCbold)

NWsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "NW")|str_detect(provincialDistribution, "North West"))
      %>%
    group_by(Group) %>%
    summarise(n=n())

NWbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "NW")|str_detect(provincialDistribution, "North West"))
      %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

NW <- left_join(NWsp, NWbold)

LPsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "LP")|str_detect(provincialDistribution, "Limpopo")) %>%
    group_by(Group) %>%
    summarise(n=n())
```

```
LPbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "LP")|str_detect(provincialDistribution, "Limpopo")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

LP <- left_join(LPsp, LPbold)

PEIsp <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "PEI")|str_detect(provincialDistribution, "Prince Edward
      Islands")) %>%
    group_by(Group) %>%
    summarise(n=n())

PEIbold <- invertsdaSilvaAlien %>%
    filter(str_detect(RangeBroadAdmin, "PEI")|str_detect(provincialDistribution, "Prince Edward
      Islands")) %>%
    filter(!is.na(boldRecords), boldRecords != 0) %>%
    group_by(Group) %>%
    summarise(records=n())

PEI <- left_join(PEIsp, PEIbold)

combined <- rbind(
    transform(EC, Dataset="EC"),
    transform(NC, Dataset="NC"),
    transform(KZN, Dataset="KZN"),
    transform(GP, Dataset="GP"),
    transform(FS, Dataset="FS"),
    transform(MP, Dataset="MP"),
    transform(WC, Dataset="WC"),
    transform(NW, Dataset="NW"),
    transform(LP, Dataset="LP"),
    transform(PEI, Dataset="PEI")
)

write.csv(combined, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding Alien
      Invertebrates\\provinceGroupTable.csv")

## produce province table
provinceTable <- combined %>%
    reframe(Province = Dataset, Group = Group, Species = n, Records = records) %>%
    group_by(Province) %>%
    arrange(desc(Species), .by_group = TRUE) %>%
    group_by(Group, .add = TRUE)

write.csv(provinceTable, "C:\\Users\\15tri\\Documents\\Admin\\Masters\\R Scripts\\Barcoding
      Alien Invertebrates\\provinceTable.csv")

## plot ALL
ggplot(combined, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white", position="fill") +
    coord_polar("y", start=0) +
    theme_void() +
    facet_wrap(~ Dataset, ncol = 3)

## plot NC
ggplot(NC, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
```

```
    coord_polar("y", start=0) +
    theme_void()

## plot KZN
ggplot(KZN, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot GP
ggplot(GP, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot FS
ggplot(FS, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot MP
ggplot(MP, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot WC
ggplot(WC, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot NW
ggplot(NW, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot LP
ggplot(LP, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## plot PEI
ggplot(PEI, aes(x="", y=n, fill=Group)) +
    geom_bar(stat="identity", width=1, color="white") +
    coord_polar("y", start=0) +
    theme_void()

## count number of species with unknown provincial distributions
noProvince <- invertsdaSilvaAlien %>%
    filter(is.na(RangeBroadAdmin), is.na(provincialDistribution))
```