Molecular Taxonomic Identification of the Oriental Fruit Fly (*Bactrocera dorsalis)* and the Melon Fly *(Bactrocera cucurbitae*): How Similar Are They at the DNA Level?

A Senior Honors Project Presented to the Faculty of the Department of Chemistry,
University of Hawai'i at Mānoa

In Partial Fulfillment of the Requirements for Bachelor of Biochemistry with Honors

By Uyen H. Freitas

May 15, 2015

Committee:

David Haymer, Mentor

Matthew Cain

Acknowledgements

Special appreciations to my project mentor Dr. David Haymer. This work was achieved at the Department of Cell and Molecular Biology (University of Hawaii at Manoa, 2500 Campus Road, Honolulu, HI 96822) under the supervision of Dr. Haymer. I would like to thank you my other thesis committee member Dr. Matthew Cain for helping me with the project. I acknowledge Dr. Le Van Vang (Can Tho University, Vietnam) for generously providing specimens of *Bacterocera cucurbitae* and *Bacterocera dorsalis.*

Abstract

Two species of fruit flies inflict massive amounts of agricultural damage throughout the Asia-Pacific region, including Hawaii. These are the oriental fruit fly, *Bactrocera dorsalis,* and the melon fly, *Bactrocera cucurbitae*. Many programs use chemical insecticides to control and eliminate the damage caused by these species, but biologically based methods are considered desirable alternatives. The first requirement of any of the biological methods is to make accurate and reliable species identifications. Unfortunately, in many cases, these species can be difficult to properly identify when traditional methods relying on morphological characters are used. The traditional methods depend almost entirely on the use of adult characters, but many specimens are captured at pre-adult stages that are difficult or impossible to identify at the species level using this approach. Fortunately, taxonomic methods based on the use of DNA characters, known as **molecular taxonomy**, have several advantages for identification of species, even for those found within closely related groupings known as species complexes. This method uses specific genetic markers found in an organism's DNA to identify it and characterize it as belonging to a particular species. One of the markers most commonly used for this purpose is the cytochrome oxidase I (COI) gene from the mitochondrial DNA. This project proposes to use molecular taxonomic methods using data from the COI gene to show the extent to which specimens of *Bactrocera dorsalis* can be discriminated from *Bactrocera cucurbitae* using DNA level markers. This will augment the use biologically based control measures, allow for more robust agricultural development, and create a stronger foundation for more sustainable farming in Hawaii as well as in other parts of the world. Another major goal of this project is collect and analyze specimens from these two species of fruit flies from Vietnam. Although these flies are known to occur throughout Southeast Asia, to date there have been no reports analyzing the genetic makeup of populations in Vietnam.

Keywords: *Bactrocera dorsalis*; *Bactrocera cucurbitae*; molecular taxonomy; haplotype; mitochondrial DNA; sequencing.

Table of Contents

Introduction

Despite advantages such as rich soil and good climate, fruit and vegetable production in Hawaii is often limited due to the presence of fruit flies that are an invasive pest species. Two of the most damaging of these pests include the melon fly, *Bacterocera cucurbitae* (*B. cucurbitae*) and the oriental fruit fly, *Bacterocera dorsalis* (*B. dorsalis)*. These flies cause significant crop losses, and they also inhibit the diversification of agriculture because of the anticipated damage they would inflict to new crops that could be grown in Hawaii.

While not much official information is available on the total dollar losses caused by these pest species, estimation caused by tephritid fruit flies affecting production, harvesting, packing, and marketing of fruits globally amounts to over US$2 billion annually. This estimation includes the Oriental fruit fly, *B. dorsalis* and the melon fly, *B. cucurbitae*. The most devastating losses in fruit commodities due to these species are observed in the Asia-Pacific region.[1]

Of these pests, the oriental fruit fly, *B. dorsalis*, is widely dispersal and is one of the most invasive, abundant and destructive pest of fruit crops. It now inhabits in more than 30 countries in the Asia-Pacific region, and is present on all major Hawaiian Islands after being accidentally introduced around 1945. These flies also cause damage in many others countries with tropical or sub-tropical climates. On the mainland US, it has invaded, but been successfully eradicated multiple times in California and in Florida. This fly attacks more than 300 kinds of fruit and vegetables including: apricot, avocado, banana, citrus, coffee, fig, guava, loquat, mango, roseapple, papaya, passion fruit, peach, pear, persimmon, pineapple, Surinam cherry and tomato. In Hawaii, the primary target is papaya.[2]

The melon fly, *B. cucubitae*, is also among the most abundant and widely distributed fruit fly species in many of these same countries and also found in Hawaii. This fly was probably introduced to Hawaii from Japan in 1985. It has been recorded to attack more than 80 different host plants, including tomato, peppers, watermelons, pumpkin, beans, eggplant, cucumber, squashes and passion fruits.[2]

Traditional methods used to control these pests involve the application of chemical pesticides. Pesticides exposure has been related to many common public health problems such as, asthma, autism and learning disabilities, birth defects and reproductive dysfunction, diabetes, Parkinson's and Alzheimer's diseases, and several types of cancer.[3] Therefore, the elimination or minimization of pesticide use is desirable. An alternative control method is biological, but in these cases, the ability to correctly identify the species involved is critical. For example, one of the most widely used biological control methods is called the Sterile Insect Release Method (SIRM). In this method, large numbers of sterile males are released into an area in the hope that they will mate with the wild females. Matings involving the sterile males will produce no offspring, and so the population simply dies out over time without the use of toxic chemicals. Clearly for this technique to work, it is essential to provide sterile males of the correct species.

Also, in nature many insects, including the oriental fruit fly, are found with groups of closed-related species known as species complexes. This added complexity makes target identification difficult when relying on morphological (physical) characters. Furthermore, taxonomy is facing difficulties because of the few experts in the field that are available. The traditional identification methods depend heavily on these experts whose knowledge can easily be lost. Fortunately, however, taxonomic methods based on the use of DNA characters, known as *molecular taxonomy*, have been shown to have a number of advantages for identification of species, and especially for discriminating between those found within species complexes.[4,5]

As previously mentioned, the oriental fruit fly is a member of a species complex, but this does not apply to the melon fly. However, not all populations of either of these pests have been adequately surveyed. Populations from some areas, such as Vietnam, have not been sampled at all using molecular taxonomic methods, and this is clearly necessary to fully document the status of these species.

What is the molecular taxonomic identification method? The method known as molecular taxonomy uses molecules found within organisms both for making taxonomic (species) identifications and for systematic (evolutionary) studies of species relationships. Although many different molecules have been proposed for this method, for a number of reasons most researchers today agree that DNA is the molecule of choice (Scientific American article). To use DNA for this purpose, a tissue sample is first taken from a collected individual and the DNA is extracted. From the extracted DNA, one or more gene regions (representing genetic markers) are amplified by PCR and the nucleotide sequence of the product is determined. The actual identification is usually accomplished by comparing the DNA sequences obtained from specimens to sequences in a reference database such as GenBank or the Barcode of Life Data Systems (BOLD).[6,7]

The molecular taxonomy method has many advantages, such as resolving questions about the taxonomic identity of specimens using material derived from any stage of the life cycle. Furthermore, these methods should be accessible to any person familiar with basic molecular biology techniques.

Objectives

The objectives of this study are to use molecular taxonomic methods to identify and characterize species of fruit flies found in Vietnam. Agriculture is an important industry in Vietnam, and in such areas it is essential to be able to identify which species of fruit flies are present in areas that have damaged fruits and vegetables. Incorrect identification can result in ineffective pest control measures, especially for biological methods for the control and elimination of the damage caused by highly invasive species including *B. dorsalis* and *B. cucurbitae.* The ability to use biological control methods in Vietnam, Hawaii and other countries in the Asia-Pacific region is essential for the robust development of agriculture, and this in turn will enable more sustainable farming to be carried out in these parts of the world.

In this research project, flies were collected from Vietnam, and DNA sequences were obtained from specific genes such as the mitochondrial cytochrome oxidase I (COI) gene. In GenBank, some COI gene sequences from *B. dorsalis* and *B. cucurbitae* were already available from various countries in the Asia-Pacific region including Hawaii, Thailand, Malaysia, Indonesia, Philippines and Taiwan, but not Vietnam. Therefore, obtaining these data for fly specimens from Vietnam was an essential part of this study in order to fully document the genetic variation within these species throughout this region. Identifications and analyses were made by comparing the DNA sequences obtained from the unknown specimens to known specimens collected from Hawaii Islands and others in the reference databases.[6,7]

Purposes

To obtain fruit fly specimens from Vietnam and characterize genetic markers from these specimens at the DNA level for molecular taxonomic identification

To use established methods for DNA sequence analysis to investigate the similarities and/or differences between various species and populations of the fruit flies *B. dorsalis* and *B.cucurbitae* within Vietnam and to compare them to other populations of these pest species from other countries in the Asia-Pacific region.

Materials and Methods

Fly collections

Flies were collected from the various locations listed by addint methyl eugenol or cue-lure, both male attractants, as well as a toxicant, to the traps, as shown in **Figure 1**. The simple traps were made with coffee cups with lids attached. Two openings on the cups were made to provide entrances for flies. Holes were made on the bottom of the coffee cups to allow rainwater pass through. A string was attached through the lid for hanging purposes. Traps were placed at different locations for each population. Three traps were placed on each location approximately 200 meters apart on various types of fruit trees such as, mango, cacao, bitter melon, etc. Samples were collected 12 hours after placement of traps. All collected specimens were preserved in 95% alcohol and stored at -20^{o}C until processed for DNA analysis.

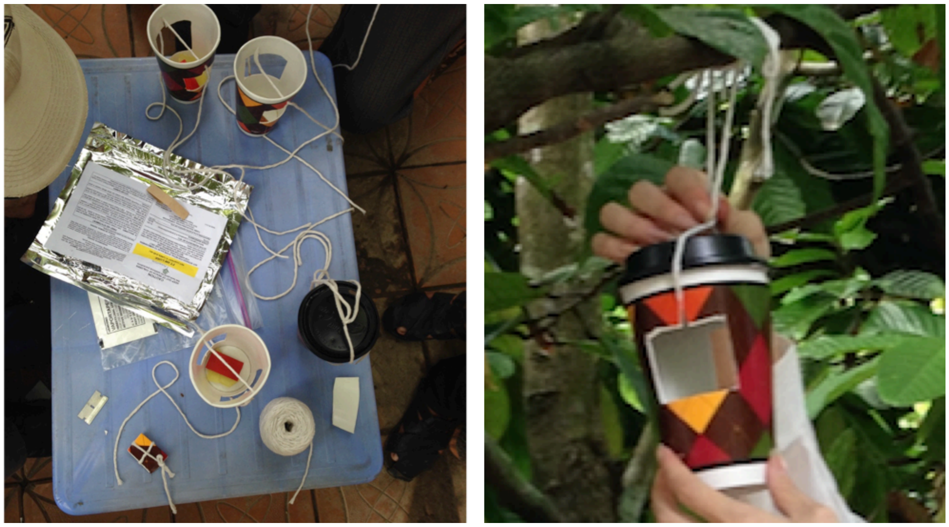


Figure 1. Trapping materials

Genomic DNA extraction

DNA purification: After removal of the ethanol, the each specimen was dried and placed in a 1.7 mL tube with labels. Solutions of PureLink^{TM} Genomic Digestion Buffer (180 μL) and Proteinase K (20 μL) were added to each tube. The specimen was then homogenized using a grinder. A new grinder was used for each specimen to avoid contamination. The ground tissue was completely immersed in the buffer mix, resulting in a yellow brown mixture due to the eye and body pigments of the specimen. The tubes were then incubated at 55^{o}C from 1 to 4 hours.

The homogenate was centrifuged at maximum speed for 3 minutes at room temperature to remove any particulate materials and the supernatant was transferred to a new, sterile microcentrifuge tube. RNase A (20 μL) was added to the lysate, mixed well and incubated at room temperature for 2 minutes. PureLink^{TM} Genomic Lysis/Binding Buffer (200 μL) was added to the tube along with absolute ethanol 100 % (200 μL), and the solution was mixed by vortexing for 5 seconds.

Binding DNA: The lysate (~640 µL) was added to a PureLink^{TM} Spin Column, and the column was centrifuged at 10,000 x g for 1 minute at room temperature. The liquid collected was discarded and the spin column was placed into a clean collection tube for washing.

Washing DNA: Wash Buffer 1 prepared with ethanol (500 µL) was added to the column. The column was then centrifuged at 10,000 x g for 1 minute at room temperature. The collection tube was discarded, and the spin column was placed into a clean PureLink^{TM} collection tube. Wash Buffer 2 prepared with ethanol (500 µL) was added to the column, centrifuged at maximum speed for 3 minutes at room temperature. The collection tube was discarded.

Eluting DNA: The spin column was placed into a sterile 1.5 mL microcentrifuge tube labeled with name of the specimen and date. PureLink^{TM} Genomic Elution Buffer (25 µL) was added to the column. The tube was then incubated at room temperature for 1 minutes, centrifuged at maximum speed for 1 minutes at room temperature. To recover more DNA, the second elution was performed using 15 µL of elution buffer. The column was centrifuged at maximum speed for 1.5 minutes at room temperature.

Collecting and storing DNA: Genomic DNA was stored at 4^{o}C for immediate and/or short term use or at -20^{o}C for long term storage. DNA was extracted from approximately 20 individuals from each population, and the appropriate markers were amplified for each using standard PCR methodology. The goal was to have results from approx. 200 individuals total (2 species x 5 populations X 20 individuals) for each marker used.

Design of PCR primers

DNA sequences from cytochrome oxidase subunit I (COI) genes are used for selection of primers. For this purpose, COI DNA sequences from *B. dorsalis* and *B. cucurbitae* obtained from

GenBank are aligned using vector NTI 7.0 (Invitrogen, Carlsbad, CA), and conserved regions are chosen to screen for useful primer pairs or primers for insect species used from published papers.[8]

Similar methods can be used to identify primers to amplify target sequence from other gene sequences including other mitochondrial genes (NAD4, 16S, etc.) as well as nuclear gene sequences such the internal transcribed sequences (ITS) regions of the ribosomal genes (Douglas and Haymer 2001).

Polymerase Chain Reaction (PCR)

DNA was extracted from approximately 20 individuals from each population, and the appropriate markers were amplified for each using standard PCR methodology. The goal was to have results from approx. 200 individuals total (2 species x 5 populations X 20 individuals) for each marker used.

The reaction mixture for amplification of COI sequences contained 100 ng of genomic DNA, 1.25 μL of 10 μM primer each, 2 μL of 2.5 mM dNTPs, 0.75 μL of 50mM $MgCl_2$, 2.5 μL of 10× Taq PCR buffer minus Mg^{2+} and 0.2 μL of 1 U Taq polymerase (Invitrogen) in a final volume of 25 μL. The specific parameters for the PCR amplifications were as follows: an initial denaturing step of 94°C for 2 min, 28 cycles of 94°C for 45 s, 44°C for 75 s and 72°C for 40 s and a final extension step of 72°C for 10 min. The amplified products were stored at−20°C. Visualization of PCR products was carried out using 0.8 % agarose gel electrophoresis and SYBR® safe DNA Gel Stain under UV light.

Agarose gel electrophoresis

The electrophoresis was conducted at 90 V using 0.8 %-50 mL agarose gels made of 0.5X TBE buffer with SYBR® safe DNA Gel Stain.

PCR Purification

PCR products are purified using PureLink® PCR Purification Kit. Four volumes of the binding buffer B2 with isopropanol (usually160 uL) were added to one volume of a PCR sample (usually 40 uL) and mixed. The sample mixture was then pipetted into a PureLink® Spin Column in a Collection Tube. The columns were centrifuged for 1 minute at 10,000 x g. The flow-through was discarded. The column was reinserted into the Collection Tube and 650 μL of Wash Buffer (W1) with ethanol was added. The tubes were centrifuged at 10,000 x g for 1 minute. Again, the flow-through was discarded and the column was placed in the same collection Tube. The column was then centrifuged at maximum speed for 2.5 minutes.

The column was placed into a clean 1.7-mL Elution Tube (supplied with the kit). Elution Buffer (50 μL) was added to the center of the column. The column was incubated at room temperature for 1 minute. The column was centrifuged at maximum speed for 2 minutes. The elution tube now contained the purified PCR product. Purified PCR product was stored at 4°C for immediate use and at -20°C for long-term storage.

DNA Sequencing

PCR Purification products were sequenced using the Sanger Dideoxy chain termination DNA Sequencing Method. (Describe how the sequencing reactions were prepared (primer and products, etc). Sequencing reactions, consisting of PCR products undergoing a cycle sequencing reaction (BigDye terminator chemistry) and capillary electrophoresis in an Applied Biosystems 3730XL DNA Analyzer were done at the sequencing facility at the Center for Advanced Studies in Genomics, Proteomics and Bioinformatics (ASGPB) at Snyder Hall at UH Manoa (www.hawaii.edu/microbiology/asgpb). The sequences produced may have read lengths up to 1000 bases.

Editing of Raw DNA Sequence Files

Most reactions produce sequence files of around 800-900 DNA base pairs. There are many ‘N’ nucleotides present at the beginning of the sequence, and also at the end of the sequence. This is common and represents the limitation of the sequencing machine and/or the product in reading the beginning and end base pairs. The ‘N’ indicates that the base pair is unknown at that location.

Sequences were examined to determine their length and quality. If they had many ‘N’ reads throughout the sequence, they were considered to be of low quality sequences and were discarded. Sequences with few numbers of ‘Ns’ are acceptable and able to be corrected to yield good quality sequences. From sequence alignments, some ‘Ns’ could be manually edited based on the majority numbers of nucleotides at that position. For example, if all other sequences showed one particular nucleotide, that ‘N’ could be manually corrected to that specific nucleotide. However, if observed sequences show multiple different nucleotides at that location, the reverse sequencing reactions might be useful to confirm the identity of the nucleotide at that position. However, reversed sequences were not of adequate always quality to be used for this purpose. In such cases, the PCR amplification, purification and sequencing was repeated.

Data Analysis

Raw sequences obtained were first trimmed to remove unsequenced regions at the beginning and end of each file. After alignment, as described above some manual editing was also done to remove single ‘Ns’. After editing, the sequenced products were confirmed to have correct open reading frames using tools (ORF finder) available from the National center for Biotechnology Information website (ncbi.nlm.nih.gov). This site also hosts GenBank, the international resource for the collection of DNA sequence information. The edited sequences also are aligned using Clustal W method from the MEGALIGN program of the DNASTAR software package.

Using software packages, the DNA sequences obtained were be analyzed in different ways by comparing them to each other and to sequences derived from GenBank. Our primary package is DNASTAR. Results were also analyzed using the network based TCS and the “Popart” programs to provide alternative perspectives on the relationships of individuals within and between populations of these species.

Results

Sample collections

As described, the collection of fruit flies in Vietnam was an important part of this study. *B. dorsalis* sample specimens were collected in July 2014 from 3 different locations (**Table 1** and **Figure 2**).

Table 1. Localities of *B. dorsalis* and *B. cucurbitae* collections in Vietnam

Populations	**Locations**	**Latitude DM**	**Longitude DM**	**Collection date**
Can Tho	Can Tho University	10°01'55.5"N	105°46'06.3"E	July 2014

	My Khanh Park	9°59'20.2"N	105°42'15.2"E	July 2014
Ben Tre	Thanh Xuan 3	10°07'37.5"N	106°20'15.7"E	July 2014

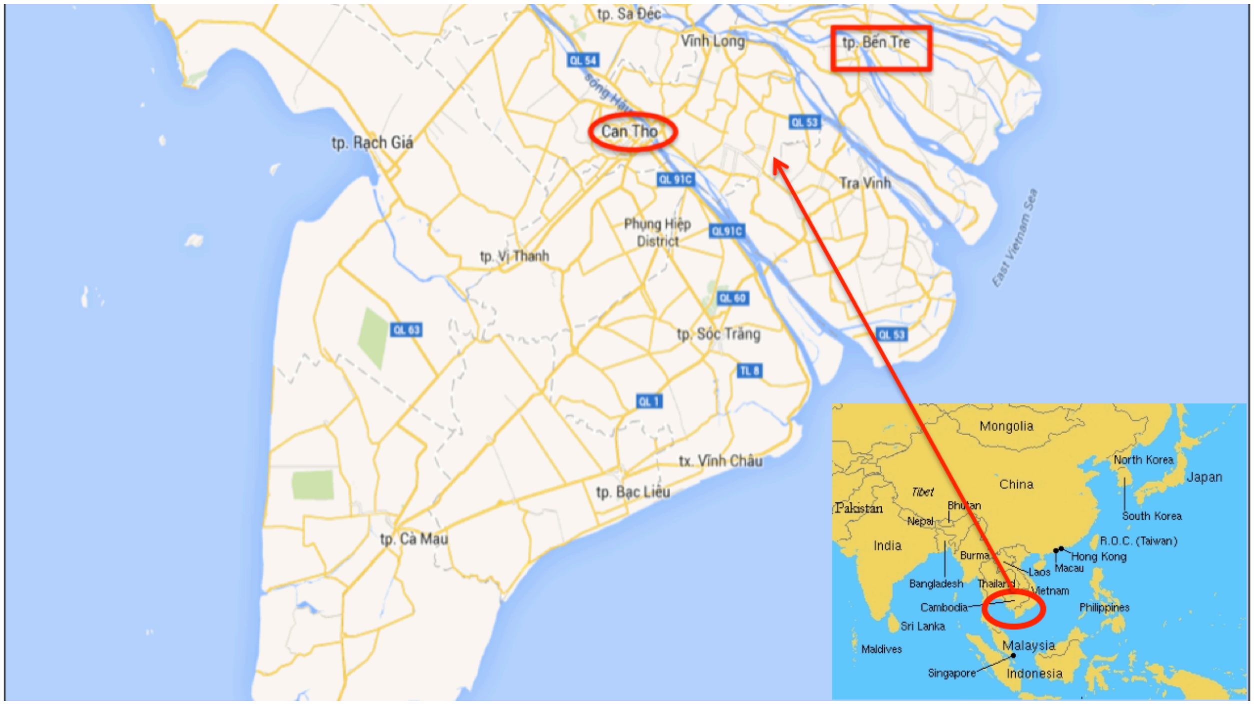


Figure 2. Geographic locations of the collections sites within the southern part of Vietnam

Several hundred specimens were collected at each of the sites and archived. **Tables 2** and **3** list the numbers of specimens analyzed by DNA sequence analysis of the two species from a total of 3 locations of *B. dorsalis* and *B. cucurbitae*. The species designation of every specimen based on morphology was initially made by a trained taxonomist, and the DNA sequences were verified by comparison with known sequences deposited in GenBank.

Table 2. Numbers of *B. dorsalis* specimens analyzed at the DNA sequence level.

Locality	DNA extractions	PCR	Number sequenced	Specimen letter designation
Can Tho University	24	24	21	BdB
My Khanh Park	20	20	20	BdA
Thank Xuan 3	20	20	20	BdC

Table 3. Numbers of *B. cucurbitae* specimens analyzed at the DNA sequence level.

Locality	DNA extractions	PCR	Number sequenced	Specimen letter designation
Can Tho University	24	24	20	BcB
My Khanh Park	20	20	20	BcA

In addition to the collections in Vietnam, *B. dorsalis* COI sequences were obtained from GenBank in order to make more extensive comparisons.[9] DNA data of *B. cucurbitae* from 3 different locations in Hawaii such as, Kauai Community College (KCC), Kapoho Papaya farm (KPF), and Kula Agriculture Park (KAP) were also obtained as shown in **Table 4**.

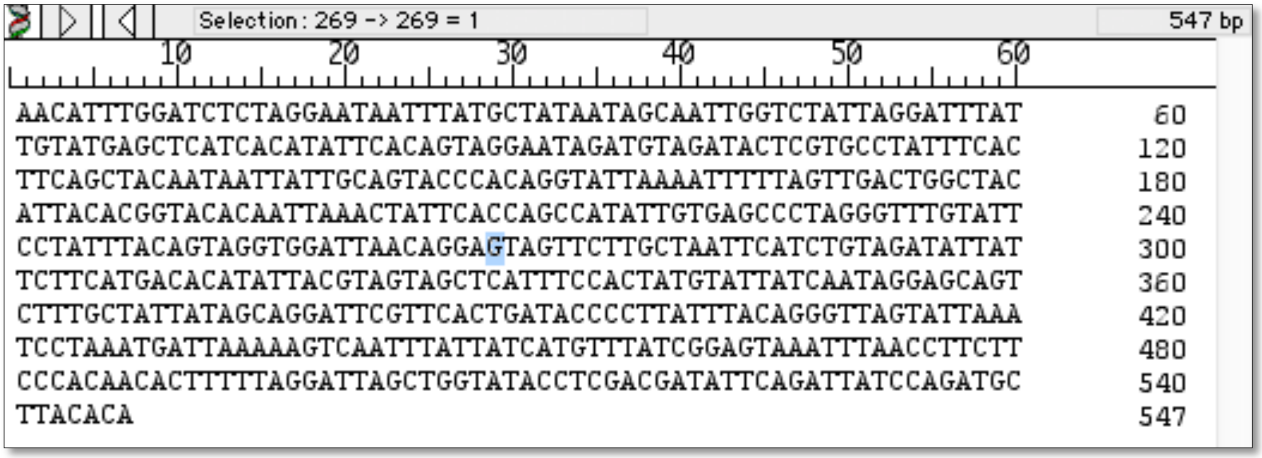
Table 4. Numbers of *B. cucurbitae* specimens analyzed at the DNA sequence level and origin of material for Hawaii population

Locality	DNA extractions	PCR	Number sequenced	Specimen designation
Kauaii Community College Orchard	10	10	6	BcC
Kapoho Papaya farm	10	10	8	BcD
Kula Agriculture Park	10	10	5	BcE
Pearl City Urban Garden Center[10]	18	16	16	BcF
Kahuku, Malaekahana Hui West[10]	18	18	17	BcG
Ewa Beach, Aloun farms[10]	18	18	12	BcH

Population level DNA sequence analysis

DNA sequences derived from the mitochondrial COI gene of *B. dorsalis* and *B. cucurbitae* have been analyzed to identify variation at the level of DNA sequences and haplotypes within and between populations. Sequences were obtained from a total of 187 individuals representing populations in Hawaii and Vietnam. After verification and editing, DNA sequences used from each individual were 547 bases in all cases.

Figure 3 below shows an example of one complete sequence file (547 bases). The sequence consists of the 4 types of nucleotides, cytosine (C), thymine (T), adenine (A) and guanine (G). The numbers on the top and right sides are used to locate the position of each nucleotide. For comparison purposes, all sequences were edited to begin at the same location.



```
AACATTTGGATCTCTAGGAATAATTTATGCTATAATAGCAATTGGTCTATTAGGATTTAT      60
TGTATGAGCTCATCACATATTCACAGTAGGAATAGATGTAGATACTCGTGCCTATTTCAC     120
TTCAGCTACAATAATTATTGCAGTACCCACAGGTATTAAAATTTTTAGTTGACTGGCTAC     180
ATTACACGGTACACAATTAAACTATTCACCAGCCATATTGTGAGCCCTAGGGTTTGTATT     240
CCTATTTACAGTAGGTGGATTAACAGGAGTAGTTCTTGCTAATTCATCTGTAGATATTAT     300
TCTTCATGACACATATTACGTAGTAGCTCATTTCCACTATGTATTATCAATAGGAGCAGT     360
CTTTGCTATTATAGCAGGATTCGTTCACTGATACCCCTTATTTACAGGGTTAGTATTAAA     420
TCCTAAATGATTAAAAAGTCAATTTATTATCATGTTTATCGGAGTAAATTTAACCTTCTT     480
CCCACAACACTTTTTAGGATTAGCTGGTATACCTCGACGATATTCAGATTATCCAGATGC     540
TTACACA                                                          547
```

Figure 3. An example of a complete DNA sequence file of the COI gene consisting of 547 nucleotide bases.

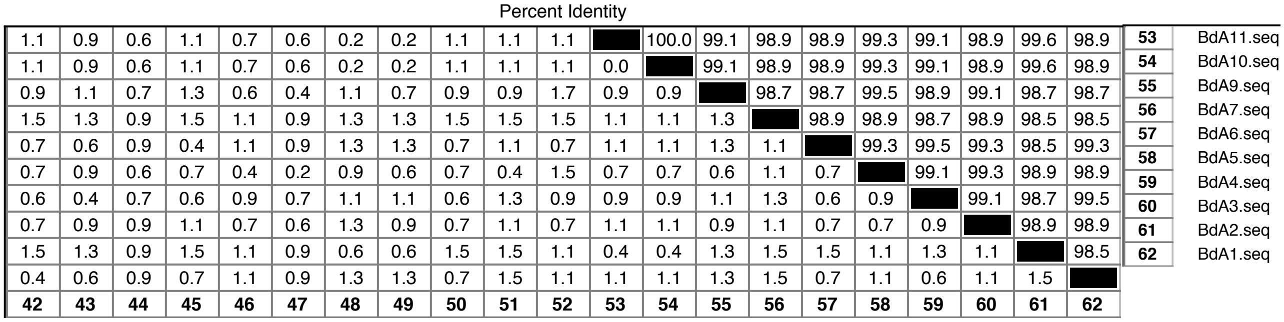
Alignment of DNA sequences from 42 individuals of *B. dorsalis* from Vietnam demonstrates how these sequences can be compared (**Figure 4)**. The sequence on the top line is complete representing the nucleotides in DNA sequence. The dots in the other sequences represent the identical bases compared to the top line sequence. Only the letters represent alternations in the sequence of nucleotides compared to the top line sequence are shown for the other sequences.

The 42 COI sequences of *B. dorsalis* generated from the Can Tho, Vietnam location were used along with 20 sequences from Ben Tre, Vietnam to obtain an alignment **(Figure 4)**. From this alignment 52 polymorphic positions were observed.

```
                   190       200       210       220       230       240       250       260       270
BdC20.seq ATTACACGGTACACAATTAAACTATTCACCAGCCATATTGTGAGCCCTAGGGTTTGTATTCCTATTTACAGTAGGAGGATTAACAGGAGT 270
BdC19.seq ...G.................................C.................................................... 270
BdC18.seq ...................................................A...................................... 270
BdC17.seq .................................T.................A...................................... 270
BdC16.seq .......................................................................................... 270
BdC15.seq .......................................A.....TT........................................... 270
BdC14.seq .......................................A.................................................. 270
BdC13.seq .......................................................................................... 270
BdC12.seq .....................................C.A.................................................. 270
BdC11.seq .......................................A.................G................................ 270
BdC10.seq .......................................................................................... 270
BdC9.seq  .......................................................................................... 270
BdC8.seq  .......................................................................................... 270
BdC7.seq  ...........................................................................G.............. 270
BdC6.seq  .......................................................................................... 270
BdC5.seq  .......................................A.................................................. 270
BdC4.seq  .......................................................................................... 270
BdC3.seq  .......................................................................................... 270
BdC2.seq  .......................................................................................... 270
BdC1.seq  ...........................................................................T.............. 270
BdB23.seq .......................................................................................... 270
BdB22.seq .......................................................................................... 270
BdB21.seq .......................................................................................... 270
BdB20.seq .......................................A.....TT........................................... 270
BdB19.seq .......................................A.................................................. 270
BdB18.seq ......T................................................................................... 270
BdB17.seq ......T................................................................................... 270
BdB16.seq ......T................................................................................... 270
BdB15.seq .......................................................................................... 270
BdB14.seq .......................................................................................... 270
BdB13.seq .......................................................................................... 270
BdB12.seq .......................................................................................... 270
BdB11.seq .......................................................................................... 270
BdB10.seq .......................................................................................... 270
BdB8.seq  .....................................C.................................................... 270
BdB7.seq  .......................................................................................... 270
BdB6.seq  .......................................A.................................................. 270
BdB4.seq  .......................................................................................... 270
BdB3.seq  .......................................................................................... 270
```

Figure 4. Alignment of DNA sequences from 62 individuals of *B. dorsalis* from Vietnam

Sequence pair distances and identity are important metrics to be used when making comparisons of variation within and between different species. **Figures 5** shows examples of values of sequence pair distances and identities for the results obtained for *B. dorsalis*. The values above the diagonal line represent percent identity and the numbers below the line was divergence values for each comparison. The numbers on the bottom and on the right side show the designation for each specimen. ‘Bd’ represents *B. dorsalis*, the following capital letter indicates the population of the specimen, and the number represents specimen number in the collection. For example, for the specimen designated ‘BdA10’ Bd represents *B. dorsalis*, ‘A’ represents My Khanh population, and the number of each specimen within this collection. The sequence pair distances and the percent identity of these *B. dorsalis* specimens ranged from 0.2 – 2.4 and 97.8 -100 % respectively.



Percent Identity

1.1	0.9	0.6	1.1	0.7	0.6	0.2	0.2	1.1	1.1	1.1	■	100.0	99.1	98.9	98.9	99.3	99.1	98.9	99.6	98.9	**53**	BdA11.seq
1.1	0.9	0.6	1.1	0.7	0.6	0.2	0.2	1.1	1.1	1.1	0.0	■	99.1	98.9	98.9	99.3	99.1	98.9	99.6	98.9	**54**	BdA10.seq
0.9	1.1	0.7	1.3	0.6	0.4	1.1	0.7	0.9	0.9	1.7	0.9	0.9	■	98.7	98.7	99.5	98.9	99.1	98.7	98.7	**55**	BdA9.seq
1.5	1.3	0.9	1.5	1.1	0.9	1.3	1.3	1.5	1.5	1.5	1.1	1.1	1.3	■	98.9	98.9	98.7	98.9	98.5	98.5	**56**	BdA7.seq
0.7	0.6	0.9	0.4	1.1	0.9	1.3	1.3	0.7	1.1	0.7	1.1	1.1	1.3	1.1	■	99.3	99.5	99.3	98.5	99.3	**57**	BdA6.seq
0.7	0.9	0.6	0.7	0.4	0.2	0.9	0.6	0.7	0.4	1.5	0.7	0.7	0.6	1.1	0.7	■	99.1	99.3	98.9	98.9	**58**	BdA5.seq
0.6	0.4	0.7	0.6	0.9	0.7	1.1	1.1	0.6	1.3	0.9	0.9	0.9	1.1	1.3	0.6	0.9	■	99.1	98.7	99.5	**59**	BdA4.seq
0.7	0.9	0.9	1.1	0.7	0.6	1.3	0.9	0.7	1.1	0.7	1.1	1.1	0.9	1.1	0.7	0.7	0.9	■	98.9	98.9	**60**	BdA3.seq
1.5	1.3	0.9	1.5	1.1	0.9	0.6	0.6	1.5	1.5	1.1	0.4	0.4	1.3	1.5	1.5	1.1	1.3	1.1	■	98.5	**61**	BdA2.seq
0.4	0.6	0.9	0.7	1.1	0.9	1.3	1.3	0.7	1.5	1.1	1.1	1.1	1.3	1.5	0.7	1.1	0.6	1.1	1.5	■	**62**	BdA1.seq
42	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**		

Figure 5. Sequence pair identities and distances for individuals of *B. dorsalis* from Vietnam

In addition to the 62 individuals analyzed from the Vietnam collection, 21 individual COI DNA sequences obtained from NCBI GenBank were used as shown in **Figure 6.** These GenBank specimens were from collections within the Hawaiian Islands reported by Barr, Ledezma, and Leblanc.[9] Sequences from GenBank were downloaded and edited down from 1500 bases to 547 bases.

```
                  190       200       210       220       230       240       250       260       270
BdD21.seq ATTACACGGTACACAATTAAACTATTCACCAGCCATATTGTGAGCCCTAGGGTTTGTATTCCTATTTACAGTAGGAGGATTAACAGGAGT 270
BdD20.seq .......................................................................................... 270
BdD19.seq .................................T........................................................ 270
BdD18.seq .................................T........................................................ 270
BdD17.seq .................................T........................................................ 270
BdD16.seq .......................................................................................... 270
BdD15.seq .......................................................................................... 270
BdD14.seq .................................T........................................................ 270
BdD13.seq .................................T........................................................ 270
BdD12.seq .......................................................................................... 270
BdD11.seq .......................................................................................... 270
BdD10.seq .......................................................................................... 270
BdD9.seq  ............................................................T............................. 270
BdD8.seq  .................................T........................................................ 270
BdD7.seq  .................................T........................................................ 270
BdD6.seq  .................................T........................................................ 270
BdD5.seq  .......................................................................................... 270
BdD4.seq  .......................................................................................... 270
BdD3.seq  .......................................................................................... 270
BdD2.seq  .................................T........................................................ 270
BdD1.seq  .......................................................................................... 270
```

Figure 6. Alignment of DNA sequences of individuals of *B. dorsalis* from Hawaii obtained from NCBI GenBank.[9]

After sequence alignment, 6 polymorphic positions were observed. These include 2 parsimony informative sites, 6 segregating sites and a value of 0.002542 for nucleotide diversity. The sequence pair distances and the percent identity of these *B. dorsalis* from the Hawaiian Island specimens ranged from 0 – 0.6 and 99.5 -100 % respectively, as shown in **Figure 7**.

Percent Identity

Divergence

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
1	■	99.8	99.6	99.6	99.8	100.0	100.0	99.6	99.6	100.0	100.0	100.0	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	1	BdD21.seq
2	0.2	■	99.5	99.5	99.6	99.8	99.8	99.5	99.5	99.8	99.8	99.8	99.6	99.5	99.3	99.5	99.8	99.8	99.6	99.5	2	BdD20.seq
3	0.4	0.6	■	100.0	99.8	99.6	99.6	100.0	100.0	99.6	99.6	99.6	99.5	100.0	99.8	100.0	99.6	99.6	99.5	100.0	3	BdD19.seq
4	0.4	0.6	0.0	■	99.8	99.6	99.6	100.0	100.0	99.6	99.6	99.6	99.5	100.0	99.8	100.0	99.6	99.6	99.5	100.0	4	BdD18.seq
5	0.2	0.4	0.2	0.2	■	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.6	99.8	99.6	99.8	99.8	99.8	99.6	99.8	5	BdD17.seq
6	0.0	0.2	0.4	0.4	0.2	■	100.0	99.6	99.6	100.0	100.0	100.0	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	6	BdD16.seq
7	0.0	0.2	0.4	0.4	0.2	0.0	■	99.6	99.6	100.0	100.0	100.0	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	7	BdD15.seq
8	0.4	0.6	0.0	0.0	0.2	0.4	0.4	■	100.0	99.6	99.6	99.6	99.5	100.0	99.8	100.0	99.6	99.6	99.5	100.0	8	BdD14.seq
9	0.4	0.6	0.0	0.0	0.2	0.4	0.4	0.0	■	99.6	99.6	99.6	99.5	100.0	99.8	100.0	99.6	99.6	99.5	100.0	9	BdD13.seq
10	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	■	100.0	100.0	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	10	BdD12.seq
11	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	0.0	■	100.0	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	11	BdD11.seq
12	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	0.0	0.0	■	99.8	99.6	99.5	99.6	100.0	100.0	99.8	99.6	12	BdD10.seq
13	0.2	0.4	0.6	0.6	0.4	0.2	0.2	0.6	0.6	0.2	0.2	0.2	■	99.5	99.3	99.5	99.8	99.8	99.6	99.5	13	BdD9.seq
14	0.4	0.6	0.0	0.0	0.2	0.4	0.4	0.0	0.0	0.4	0.4	0.4	0.6	■	99.8	100.0	99.6	99.6	99.5	100.0	14	BdD8.seq
15	0.6	0.7	0.2	0.2	0.4	0.6	0.6	0.2	0.2	0.6	0.6	0.6	0.7	0.2	■	99.8	99.5	99.5	99.3	99.8	15	BdD7.seq
16	0.4	0.6	0.0	0.0	0.2	0.4	0.4	0.0	0.0	0.4	0.4	0.4	0.6	0.0	0.2	■	99.6	99.6	99.5	100.0	16	BdD6.seq
17	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.2	0.4	0.6	0.4	■	100.0	99.8	99.6	17	BdD5.seq
18	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.2	0.4	0.6	0.4	0.0	■	99.8	99.6	18	BdD4.seq
19	0.2	0.4	0.6	0.6	0.4	0.2	0.2	0.6	0.6	0.2	0.2	0.2	0.4	0.6	0.7	0.6	0.2	0.2	■	99.5	19	BdD3.seq
20	0.4	0.6	0.0	0.0	0.2	0.4	0.4	0.0	0.0	0.4	0.4	0.4	0.6	0.0	0.2	0.0	0.4	0.4	0.6	■	20	BdD2.seq
21	0.0	0.2	0.4	0.4	0.2	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.2	0.4	0.6	0.4	0.0	0.0	0.2	0.4	21	BdD1.seq

Figure 7. Sequence pair identities and distances of individuals of *B. dorsalis* from Hawaii obtained from NCBI GenBank.[9]

In addition to the 62 sequences from the Vietnam collection, 21 COI sequences obtained from NCBI GenBank were used as shown in **Figure 8** to give a total of 83 sequences for comparison. Using these, all together 54 polymorphic positions were observed along with 29 parsimony informative sites, 60 segregating sites and the value of 0.0117237 in nucleotide diversity.

```
                  190       200       210       220       230       240       250       260       270
BdD21.seq ATTACACGGTACACAATTAAACTATTCACCAGCCATATTGTGAGCCCTAGGGTTTGTATTCCTATTTACAGTAGGAGGATTAACAGGAGT 270
BdD20.seq .......................................................................................... 270
BdD19.seq .................................T........................................................ 270
BdD18.seq .................................T........................................................ 270
BdD17.seq .................................T........................................................ 270
BdD16.seq .......................................................................................... 270
BdD15.seq .......................................................................................... 270
BdD14.seq .................................T........................................................ 270
BdD13.seq .................................T........................................................ 270
BdD12.seq .......................................................................................... 270
BdD11.seq .......................................................................................... 270
BdD10.seq .......................................................................................... 270
BdD9.seq  ............................................................T............................. 270
BdD8.seq  .................................T........................................................ 270
BdD7.seq  .................................T........................................................ 270
BdD6.seq  .................................T........................................................ 270
BdD5.seq  .......................................................................................... 270
BdD4.seq  .......................................................................................... 270
BdD3.seq  .......................................................................................... 270
BdD2.seq  .................................T........................................................ 270
BdD1.seq  .......................................................................................... 270
BdC20.seq .......................................................................................... 270
BdC19.seq ...G.................................C.................................................... 270
BdC18.seq ...................................................A...................................... 270
BdC17.seq .................................T.................A...................................... 270
BdC16.seq .......................................................................................... 270
BdC15.seq .......................................A.....TT........................................... 270
BdC14.seq .......................................A.................................................. 270
BdC13.seq .......................................................................................... 270
BdC12.seq .....................................C.A.................................................. 270
BdC11.seq .......................................A.................G................................ 270
BdC10.seq .......................................................................................... 270
BdC9.seq  .......................................................................................... 270
BdC8.seq  .......................................................................................... 270
BdC7.seq  ...........................................................................G.............. 270
BdC6.seq  .......................................................................................... 270
BdC5.seq  .......................................A.................................................. 270
BdC4.seq  .......................................................................................... 270
BdC3.seq  .......................................................................................... 270
BdC2.seq  .......................................................................................... 270
BdC1.seq  ...........................................................................T.............. 270
BdB23.seq .......................................................................................... 270
BdB22.seq .......................................................................................... 270
BdB21.seq .......................................................................................... 270
BdB20.seq .......................................A.....TT........................................... 270
BdB19.seq .......................................A.................................................. 270
BdB18.seq ......T................................................................................... 270
BdB17.seq ......T................................................................................... 270
BdB16.seq ......T................................................................................... 270
BdB15.seq .......................................................................................... 270
BdB14.seq .......................................................................................... 270
BdB13.seq .......................................................................................... 270
BdB12.seq .......................................................................................... 270
BdB11.seq .......................................................................................... 270
```

Figure 8. Alignment of DNA sequences for individuals of *B. dorsalis* from Hawaii and Vietnam

Comparisons between the *B. dorsalis* from the Hawaiian Islands and Vietnam datasets using popart software are shown in **Table 5**. Nucleotide diversity is a measure of genetic variability, which represents patterns of molecular diversity within a sample of gene copies.[11] The *B. dorsalis* collection from Vietnam has a nucleotide diversity value that is much larger compared to that of the Hawaii Islands collection . This suggests that there is more diversity in the *B. dorsalis* populations in Vietnam. The number of segregating sites in a dataset also provides an indicator of the degree of DNA sequence variation that is present in a sample.[12] Again, *B. dorsali*s in Vietnam shows a higher number of segregating sites compared to the Hawaiian collection. Finally, parsimony-informative sites are those that contain at least two types of values (nucleotides or amino acids) where both types occur with a minimum frequency. Parsimony in general is defined as being the simplest explanation for an observation. Here again the number of parsimony-informative sites of the *B. dorsalis* in Vietnam is much greater compared to Hawaii parsimony-informative sites.

Table 5. TCS Network statistics

Statistics	***B. dorsalis* (Vietnam)**	***B. dorsalis (Hawaii)***
Nucleotide diversity	0.00976917	0.002542
Segregating sites	55	6
Parsimony-informative sites	26	2
Polymorphic positions	52	6

Figure 9 and **10** shows phylogenetic trees generated for these same individuals using the distance values as input. The phylogenetic analysis was performed following alignments using

the Clustal W method. The resulting tree **Figure 9** shows many branches of different length compared to that seen in **Figure 10**.

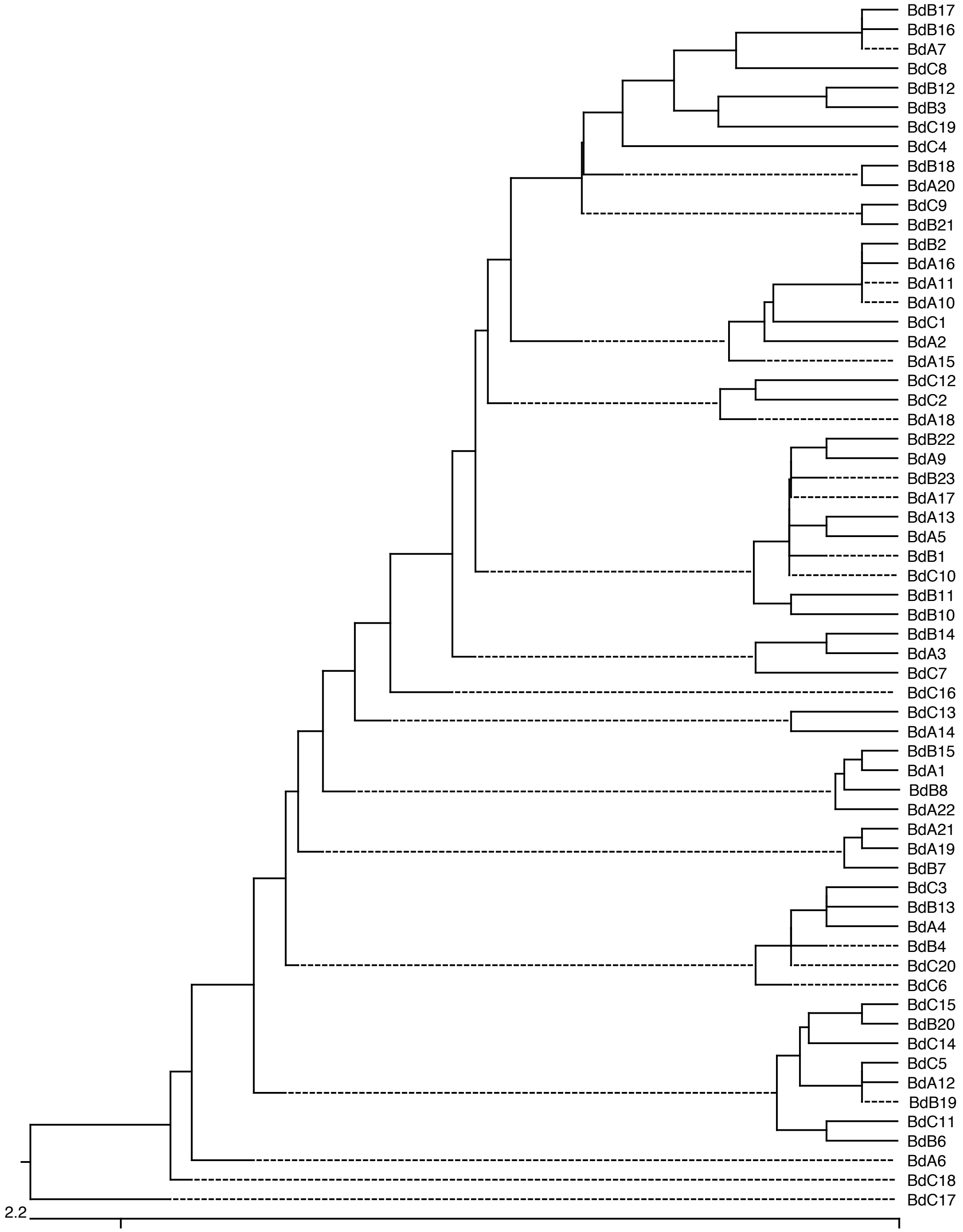


Figure 9. Phylogenetic tree of the of *B. dorsalis* collections in Vietnam

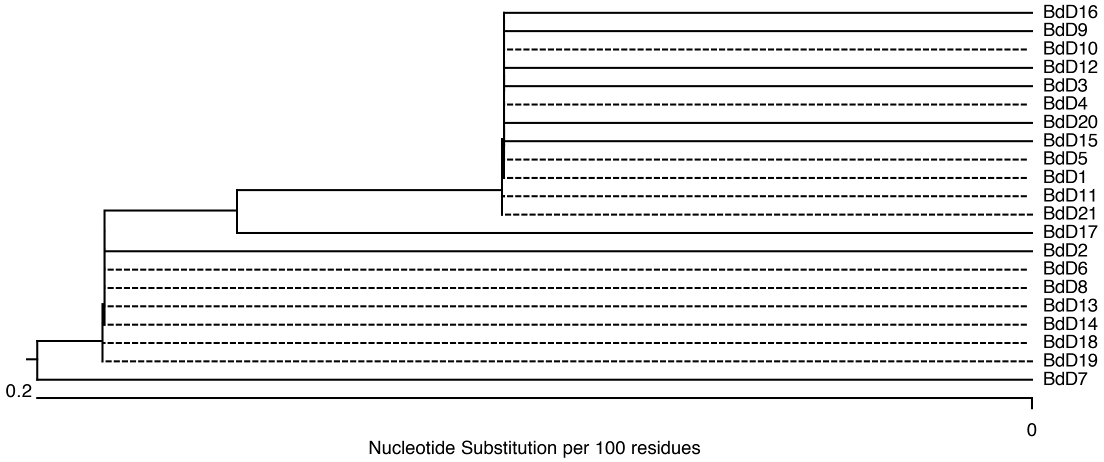


Figure 10. Phylogenetic tree of the of *B. dorsalis* collection in Hawaii

A phylogenetic tree combining the results obtained from both of the *B. dorsalis* collections in Vietnam and the Hawaii was generated to compare the two populations. No clear major groupings are observed within this tree.

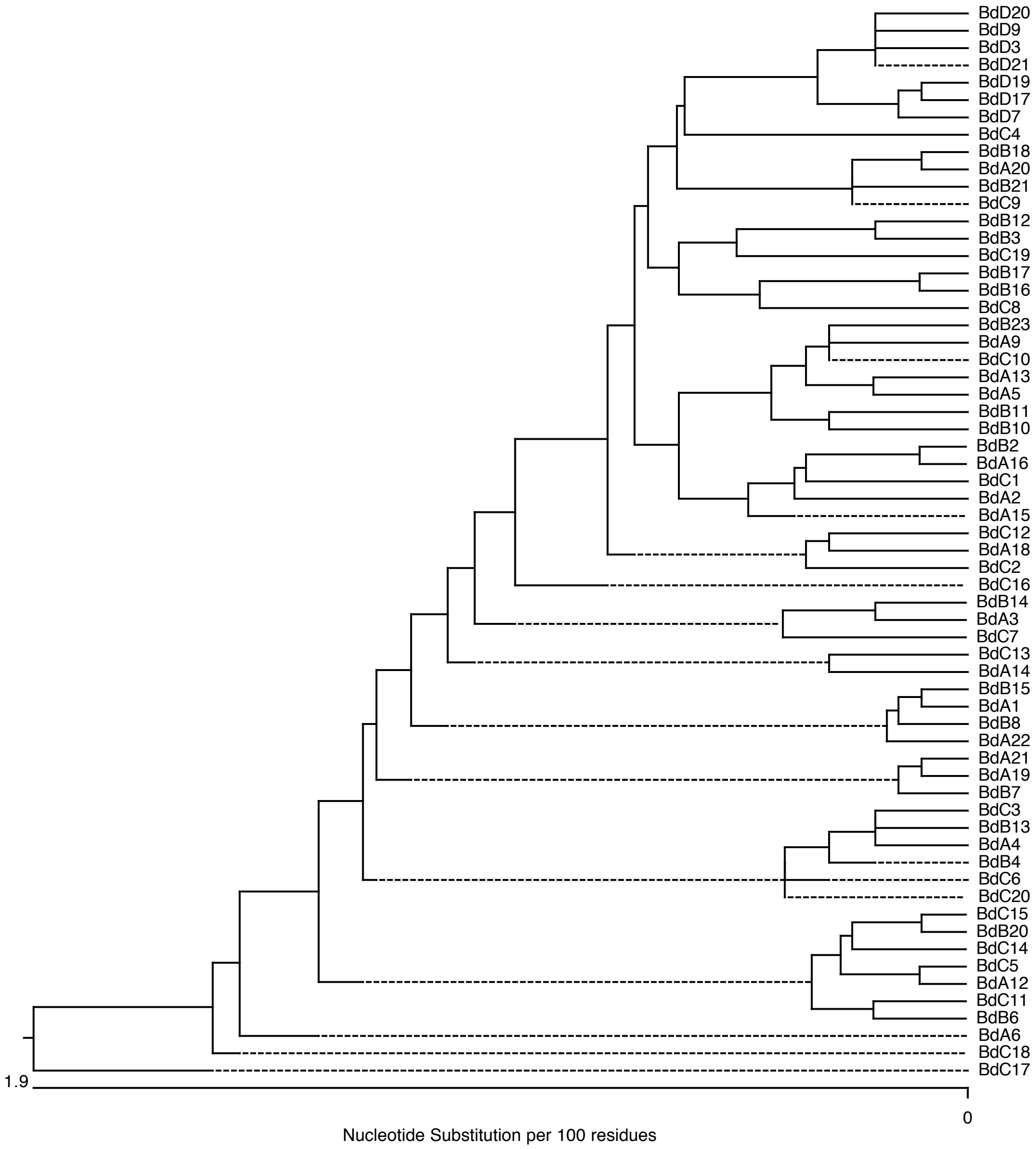


Figure 11. Phylogenetic tree of the of *B. dorsalis* collections in Vietnam and the Hawaii after elimination of duplicate sequences

Even though phylogenetic trees are very useful, they can be difficult to visualize and interpret, especially when very closely related sequences are used as input. Another method for visualizing these relationships is called TCS, which is a computer program to estimate gene genealogies as a network, as shown in **Figure 12** and **13**. TCS can also be used to visualize and estimate relationships among organisms using diagrams. In these diagrams, the short lines that cut perpendicular along the straight lines each represent one base change between the individual sequences. The TCS network analysis was performed on *B. dorsalis* data obtained from the collections from Vietnam and Hawaii.

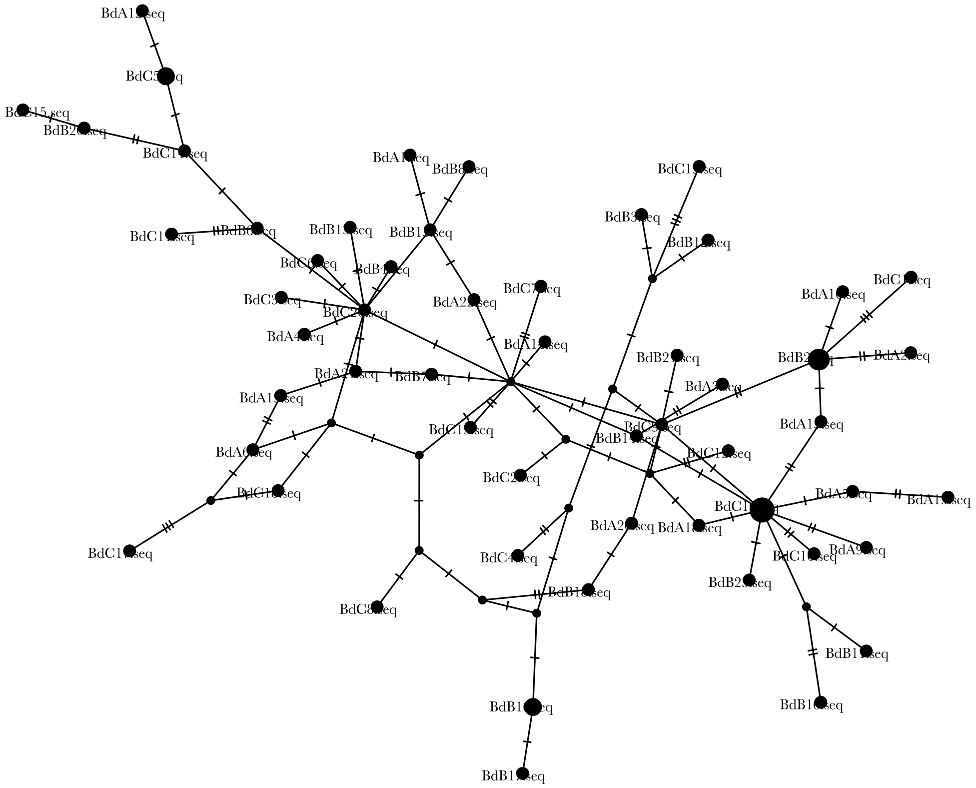


Figure 12. TCS Network of the *B. dorsalis* collections from Vietnam

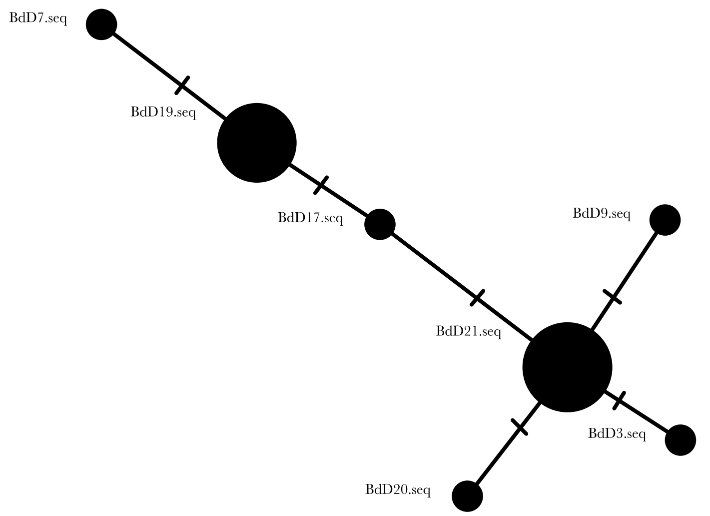


Figure 13. TCS Network of the *B. dorsalis* collection from Hawaii

A TCS Network of *B. dorsalis* collection in Vietnam and Hawaii (**Figure 14**) was also generated combining the datasets used for figures 12 and 13. This network shows extensive branching but again no clear major groupings between the two populations.

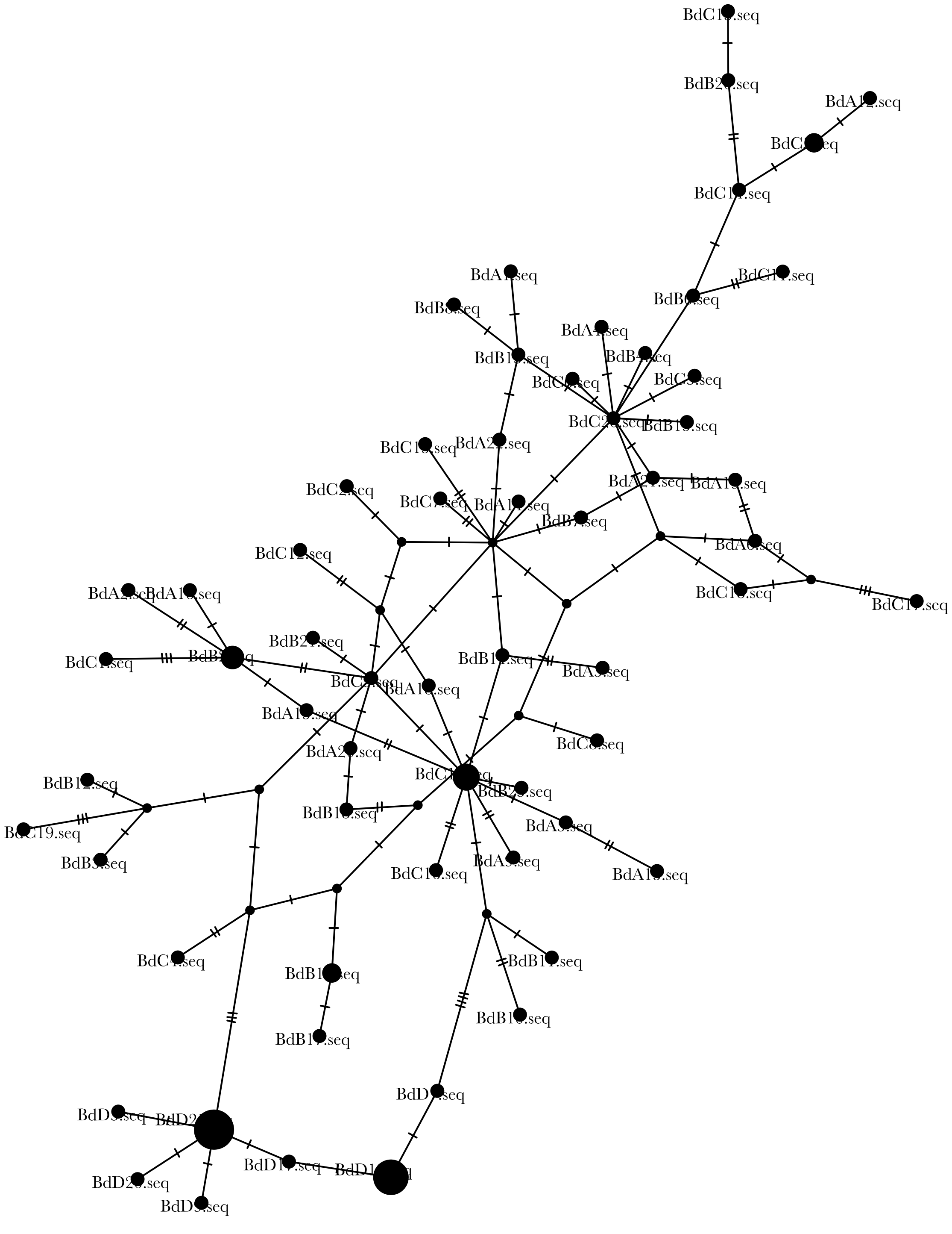


Figure 14. TCS Network of *B. dorsalis* collections in Vietnam and Hawaii

The 40 COI DNA sequences from *B. cucurbitae* individuals obtained from the collections Can Tho, Vietnam location were aligned, as shown in **Figure 14**. After alignment, 9 polymorphic positions were observed. Also included here were 2 parsimony informative sites, 10 segregating sites and a nucleotide diversity value of 0.00108752.

```
                   |         |         |         |         |         |         |         |         |
                  190       200       210       220       230       240       250       260       270
          ------------------------------------------------------------------------------------------
BcB24.seq TCTTCATGGAACACAATTAAATTATTCCCCAGCTATGCTATGAGCTTTAGGTTTTGTATTTTTATTCACAGTTGGGGGACTAACTGGAGT 270
BcB23.seq .......................................................................................... 270
BcB22.seq .......................................................................................... 270
BcB21.seq .......................................................................................... 270
BcB20.seq .......................................................................................... 270
BcB19.seq .......................................................................................... 270
BcB18.seq .......................................................................................... 270
BcB17.seq .......................................................................................... 270
BcB16.seq .......................................................................................... 270
BcB14.seq .......................................................................................... 270
BcB13.seq .......................................................................................... 270
BcB12.seq .......................................................................................... 270
BcB11.seq .......................................................................................... 270
BcB10.seq .......................................................................................... 270
BcB9.seq  .......................................................................................... 270
BcB8.seq  .......................................................................................... 270
BcB7.seq  .......................................................................................... 270
BcB6.seq  .......................................................................................... 270
BcB5.seq  .......................................................................................... 270
BcB4.seq  .......................................................................................... 270
BcA20.seq .......................................................................................... 270
BcA19.seq .......................................................................................... 270
BcA18.seq .......................................................................................... 270
BcA17.seq .......................................................................................... 270
BcA16.seq .......................................................................................... 270
BcA15.seq .......................................................................................... 270
BcA14.seq ............T............................................................................. 270
BcA13.seq .......................................................................................... 270
BcA12.seq .......................................................................................... 270
BcA11.seq .......................................................................................... 270
BcA10.seq .......................................................................................... 270
BcA9.seq  .......................................................................................... 270
BcA8.seq  .......................................................................................... 270
BcA7.seq  .......................................................................................... 270
BcA6.seq  .......................................................................................... 270
BcA5.seq  .......................................................................................... 270
BcA4.seq  .......................................................................................... 270
BcA3.seq  .......................................................................................... 270
BcA2.seq  .......................................................................................... 270
BcA1.seq  .......................................................................................... 270
```

Figure 14. Alignment of DNA sequences of individuals of *B. cucurbitae* from Vietnam

The sequence pair distances and the percent identity of these *B. cucurbitae* specimens ranged from 0.0 – 0.4 and 99.6 -100 % respectively, as shown in **Figure 15**.

Percent Identity

Divergence

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	■	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
2	0.0	■	99.8	100.0	99.8	100.0	100.0	100.0	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
3	0.2	0.2	■	99.8	99.6	99.8	99.8	99.8	99.8	99.6	99.8	99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.6
4	0.0	0.0	0.2	■	99.8	100.0	100.0	100.0	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
5	0.2	0.2	0.4	0.2	■	99.8	99.8	99.8	99.8	99.6	99.8	99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.6
6	0.0	0.0	0.2	0.0	0.2	■	100.0	100.0	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
7	0.0	0.0	0.2	0.0	0.2	0.0	■	100.0	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
8	0.0	0.0	0.2	0.0	0.2	0.0	0.0	■	100.0	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
9	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	■	99.8	100.0	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
10	0.2	0.2	0.4	0.2	0.4	0.2	0.2	0.2	0.2	■	99.8	99.6	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.6
11	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	■	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	99.8
12	0.2	0.2	0.4	0.2	0.4	0.2	0.2	0.2	0.2	0.4	0.2	■	99.8	99.8	99.8	99.8	99.8	99.8	99.8	99.6
13	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	■	100.0	100.0	100.0	100.0	100.0	100.0	99.8
14	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	■	100.0	100.0	100.0	100.0	100.0	99.8
15	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	■	100.0	100.0	100.0	100.0	99.8
16	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	■	100.0	100.0	100.0	99.8
17	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	■	100.0	100.0	99.8
18	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	■	100.0	99.8
19	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	■	99.8
20	0.2	0.2	0.4	0.2	0.4	0.2	0.2	0.2	0.2	0.4	0.2	0.4	0.2	0.2	0.2	0.2	0.2	0.2	0.2	■
21	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
22	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
23	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
24	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2

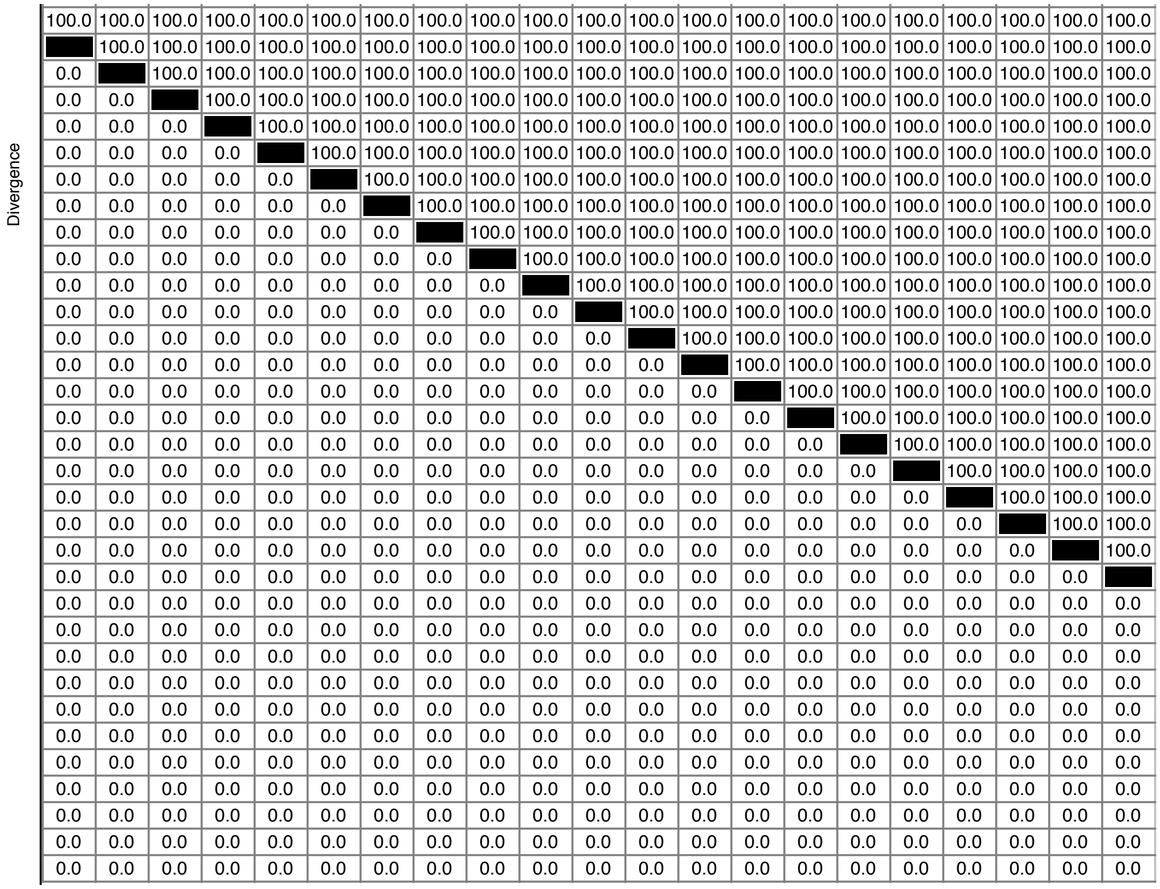
Figure 15. Sequence pair distances and identities of individuals of *B. cucurbitae* from Vietnam

B. cucurbitae individuals obtained from the Hawaiian location were aligned to generate an alignment of 64 sequences, as shown in **Figure 16**. After sequence alignment, 5 polymorphic positions were observed. These were also characterized by 0 parsimony informative sites, 5 segregating sites and a nucleotide diversity value of 0.000290183. Sequence pair distances for *B. cucurbitae* ranged from 0 – 0.4.

```
                  190       200       210       220       230       240       250       260       270
          ---------|---------|---------|---------|---------|---------|---------|---------|---------|
BcK3.seq  TCTTCATGGAACACAATTAAATTATTCCCCAGCTATGCTATGAGCTTTAGGTTTTGTATTTTTATTCACAGTTGGGGGACTAACTGGAGT 270
BcK2.seq  .......................................................................................... 270
BcK1.seq  .......................................................................................... 270
BcJ14.seq .......................................................................................... 270
BCJ13.seq .......................................................................................... 270
BcJ12.seq .......................................................................................... 270
BcJ11.seq .......................................................................................... 270
BcJ10.seq .......................................................................................... 270
BcJ9.seq  .......................................................................................... 270
BcJ8.seq  .......................................................................................... 270
BcJ7.seq  .......................................................................................... 270
BcJ6.seq  .......................................................................................... 270
BcJ5.seq  .......................................................................................... 270
BcJ4.seq  .......................................................................................... 270
BcJ3.seq  .......................................................................................... 270
BcJ2.seq  .......................................................................................... 270
BcJ1.seq  .......................................................................................... 270
BcI6.seq  .......................................................................................... 270
BcI5.seq  .......................................................................................... 270
BcI4.seq  .......................................................................................... 270
BcI3.seq  .......................................................................................... 270
BcI2.seq  .......................................................................................... 270
BcI1.seq  .......................................................................................... 270
BcH8.seq  .......................................................................................... 270
BcH7.seq  .......................................................................................... 270
BcH6.seq  .......................................................................................... 270
BcH5.seq  .......................................................................................... 270
BcH4.seq  .......................................................................................... 270
BcH3.seq  .......................................................................................... 270
BcH2.seq  .......................................................................................... 270
BcH1.seq  .......................................................................................... 270
BcG8.seq  .......................................................................................... 270
BcG7.seq  .......................................................................................... 270
BcG6.seq  .......................................................................................... 270
BcG5.seq  .......................................................................................... 270
BcG4.seq  .......................................................................................... 270
BcG3.seq  .......................................................................................... 270
```

Figure 16. Alignment of DNA sequences of individuals of *B. curcurbitae* from Hawaii

The sequence pair distances and the percent identity of the Hawaii collection *B. cucurbitae* specimens ranged from 0.0 – 0.6 and 99.8 -100 % respectively, as shown in **Figure 17**.



100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■	100.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	■
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figure 17. Sequence pair identities and distances for individuals of *B. cucurbitae* from Hawaii

The 40 COI sequences of *B. cucurbitae* generated in the Vietnam collection were merged with 64 sequences from Hawaii population to obtain a final alignment of 104 sequences, as shown in **Figure 18**. After sequence alignment, 14 polymorphic positions were observed and characterized by 2 parsimony informative sites, 15 segregating sites and a nucleotide diversity value of 0.000602076. Sequence pair distances for *B. cucurbitae* ranged from 0 - 0.4.

```
                  10        20        30        40        50        60        70        80        90
BcK3.seq  AACATTTGGTTCCTTAGGTATAATTTACGCTATAATAGCAATTGGATTACTTGGATTTATTGTATGAGCCCACCATATATTTACAGTAGG 90
BcK2.seq  .......................................................................................... 90
BcK1.seq  .......................................................................................... 90
BcJ14.seq .......................................................................................... 90
BCJ13.seq .......................................................................................... 90
BcJ12.seq .......................................................................................... 90
BcJ11.seq .......................................................................................... 90
BcJ10.seq .......................................................................................... 90
BcJ9.seq  .......................................................................................... 90
BcJ8.seq  .......................................................................................... 90
BcJ7.seq  .......................................................................................... 90
BcJ6.seq  .......................................................................................... 90
BcJ5.seq  .......................................................................................... 90
BcJ4.seq  .......................................................................................... 90
BcJ3.seq  .......................................................................................... 90
BcJ2.seq  .......................................................................................... 90
BcJ1.seq  .......................................................................................... 90
BcI6.seq  .......................................................................................... 90
BcI5.seq  .......................................................................................... 90
BcI4.seq  .......................................................................................... 90
BcI3.seq  .......................................................................................... 90
BcI2.seq  .......................................................................................... 90
BcI1.seq  .......................................................................................... 90
BcH8.seq  .......................................................................................... 90
BcH7.seq  .......................................................................................... 90
BcH6.seq  .......................................................................................... 90
BcH5.seq  .......................................................................................... 90
BcH4.seq  .......................................................................................... 90
BcH3.seq  .......................................................................................... 90
BcH2.seq  .......................................................................................... 90
BcH1.seq  .......................................................................................... 90
BcG8.seq  .......................................................................................... 90
BcG7.seq  .......................................................................................... 90
BcG6.seq  .......................................................................................... 90
BcG5.seq  .......................................................................................... 90
BcG4.seq  .......................................................................................... 90
BcG3.seq  .......................................................................................... 90
BcG2.seq  .......................................................................................... 90
BcG1.seq  .......................................................................................... 90
BcF6.seq  .......................................................................................... 90
BcF5.seq  .......................................................................................... 90
BcF4.seq  .......................................................................................... 90
BcF3.seq  .......................................................................................... 90
BcF2.seq  .......................................................................................... 90
BcF1.seq  .......................................................................................... 90
BcE10.seq .......................................................................................... 90
BcE8.seq  .......................................................................................... 90
BcE7.seq  .......................................................................................... 90
BcE6.seq  .......................................................................................... 90
BcD10.seq .......................................................................................... 90
BcD9.seq  .......................................................................................... 90
BcD8.seq  .......................................................................................... 90
BcD7.seq  .......................................................................................... 90
BcD6.seq  .......................................................................................... 90
BcD3.seq  .......................................................................................... 90
```

Figure 18. Alignment of DNA sequences of individuals of *B. cucurbitae* from Hawaii and Vietnam.

Table 6. TCS Network statistics for *B. cucurbitae*

Statistics	***B. cucurbitae* (Vietnam)**	***B. cucurbitae (Hawaii)***
Nucleotide diversity	0.00108752	0.000290183
Segregating sites	10	5
Parsimony-informative sites	2	0
Polymorphic positions	9	5

Figures 19 and **20** shows phylogenetic trees generated for these same individuals using the distance values as input. The phylogenetic analysis was performed following alignments using the Clustal W method. The resulting trees do not display many different branches for both populations. This suggests that these individuals are closely related to each other.

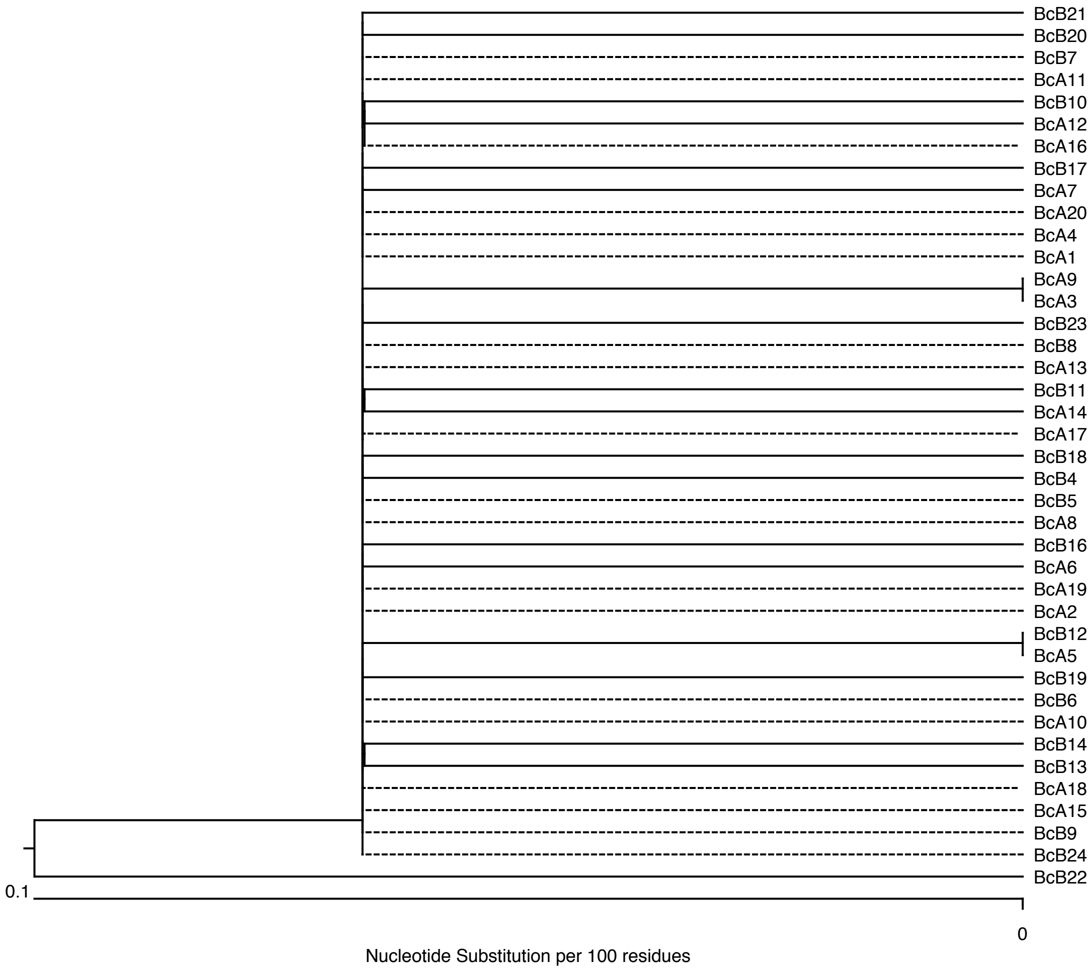


Figure 19. Phylogenetic tree of the *B. cucurbitae* collections in Vietnam

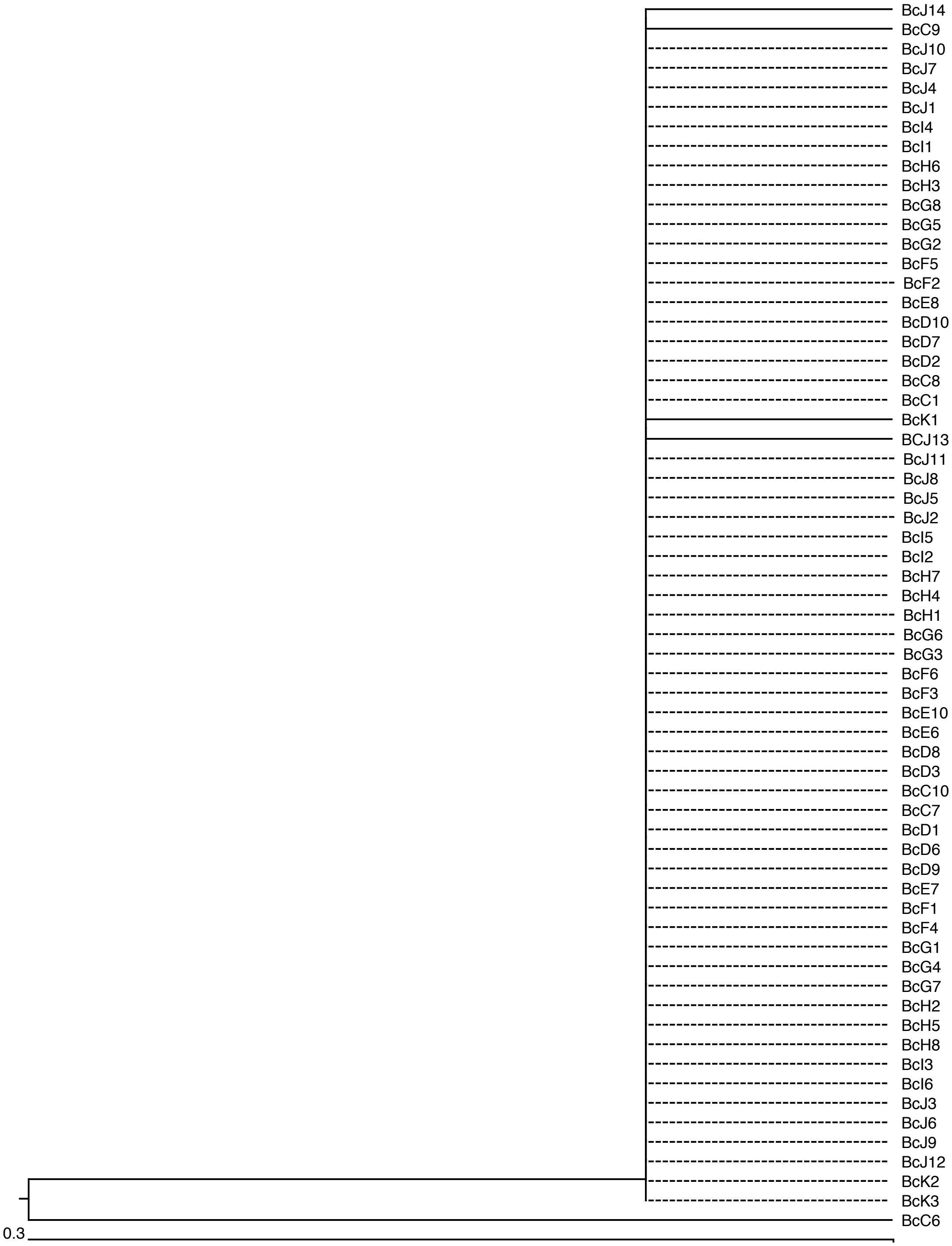


Figure 20. Phylogenetic tree of the *B. cucurbitae* collections in Hawaii

Combing the two populations with the elimination of identical individuals, phylogenetic tree in **Figure 21** was generated. The phylogenic tree displayed the close relationship between individuals from Hawaii. Both collections do not displace multiple branching.

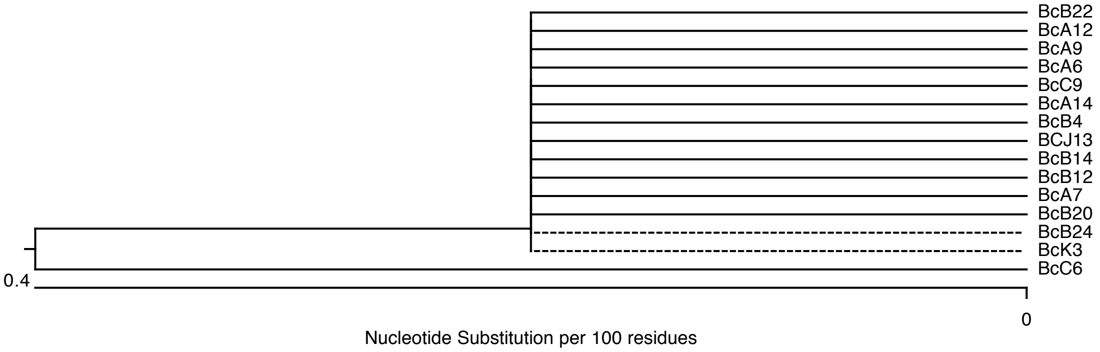


Figure 21. Phylogenetic tree of the *B. cucurbitae* collections in Vietnam and Hawaii after elimination of identical sequences

The TCS network analysis was performed on *B. cucurbitae* data obtained from the collections from Vietnam and Hawaii, as shown in **Figures 22** and **23**. The TCS Network of *B. cucurbitae* collection in Vietnam was more spread out compared to the TCS Network of *B. cucurbitae* collection from Hawaii. There were a large number of identical individuals in the Hawaii collection.

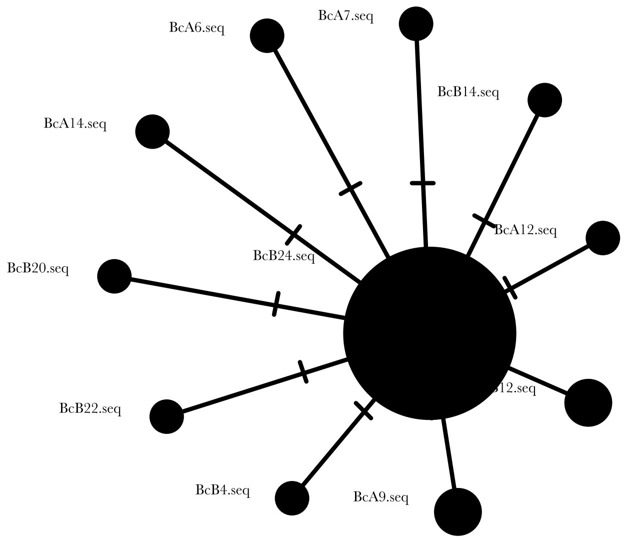


Figure 22. TCS Network of *B. cucurbitae* collection in Vietnam

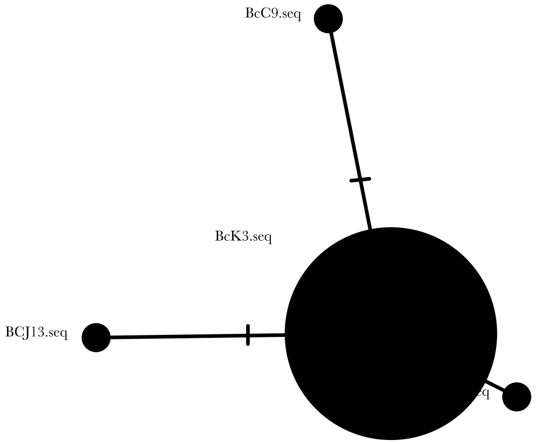


Figure 23. TCS Network of *B. cucurbitae* collection in Hawaii

The TCS Network of the *B. cucurbitae* collection in Vietnam and Hawaii (**Figure 24**) was generated as the combining of the two TCS Network in **Figures 22** and **23**. This network displayed a more compact structure with high numbers of identical individuals as the large black circle indicate. The 12 non-identical sequences differ only by 1 base pair.

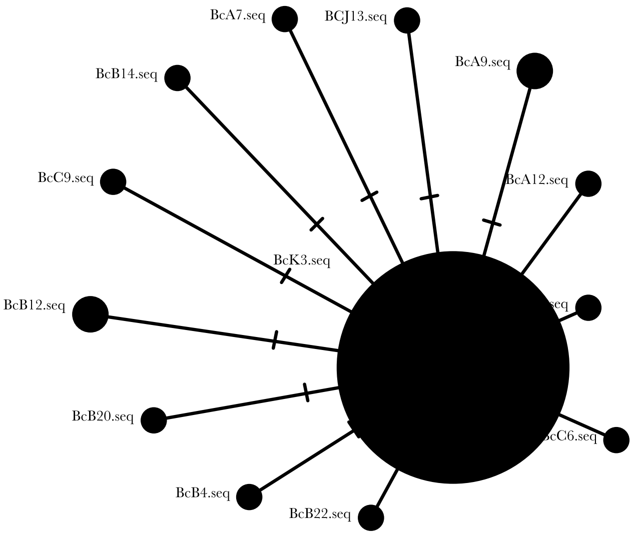


Figure 24. TCS Network of *B. cucurbitae* collections in Vietnam and Hawaii

The 83 COI sequences of *B. dorsalis* were merged with 104 sequences of *B. cucurbitae* to obtain a final alignment of 187 sequences, as shown in **Figure 25**.

```
BdB4.seq  .........T......................................C......................................... 360
BdB3.seq  .........T......................................C......................................... 360
BdB2.seq  .........T......................................C......................................... 360
BdB1.seq  .........T......................................C......................................... 360
BdA22.seq .........T......................................C......................................... 360
BdA21.seq .........T......................................C......................................... 360
BdA20.seq .........T......................................C......................................... 360
BdA19.seq .........T.............................T........C......................................... 360
BdA18.seq .........T......................................C......................................... 360
BdA17.seq .........T......................................C......................................... 360
BdA16.seq .........T......................................C......................................... 360
BdA15.seq .........T......................................C......................................... 360
BdA14.seq .........T......................................C......................................... 360
BdA13.seq .........T.............................T........C.................T....................... 360
BdA12.seq .........T......................................C......................................... 360
BdA11.seq .........T......................................C......................................... 360
BdA10.seq .........T......................................C......................................... 360
BdA9.seq  .........T......................................C......................................... 360
BdA7.seq  .........T......................................C......................................... 360
BdA6.seq  .........T.............................T........C......................................... 360
BdA5.seq  .........T.............................T........C......................................... 360
BdA4.seq  .........T......................................C......................................... 360
BdA3.seq  .........T......................................C......................................... 360
BdA2.seq  .........T......................................C......................................... 360
BdA1.seq  .........T......................................C......................................... 360
BcK3.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcK2.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcK1.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ14.seq ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BCJ13.seq ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ12.seq ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ11.seq ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ10.seq ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ9.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ8.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ7.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ6.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ5.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ4.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ3.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ2.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcJ1.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI6.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI5.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI4.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI3.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI2.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcI1.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH8.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH7.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH6.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH5.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH4.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH3.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH2.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcH1.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
BcG8.seq  ...AT.A.....C..T.....T....................T..C..C.................T.....T................. 360
```

Figure 25. Alignment of DNA sequences from individuals representing both *B. dorsalis* and *B. cucurbitae* from Hawaii and Vietnam

Phylogenetic trees were also generated to show the evolutionary relationships based on DNA sequences among individuals of the two species, as shown in **Figure 26**. The phylogenetic tree of *B. dorsalis* was extensively branched. However, the tree for *B. cucurbitae* is simpler without many branches. This result is also consistent with the alignments and the sequence pair distance results. In **Figure 26**, there was obvious grouping in the alignment branching pattern.

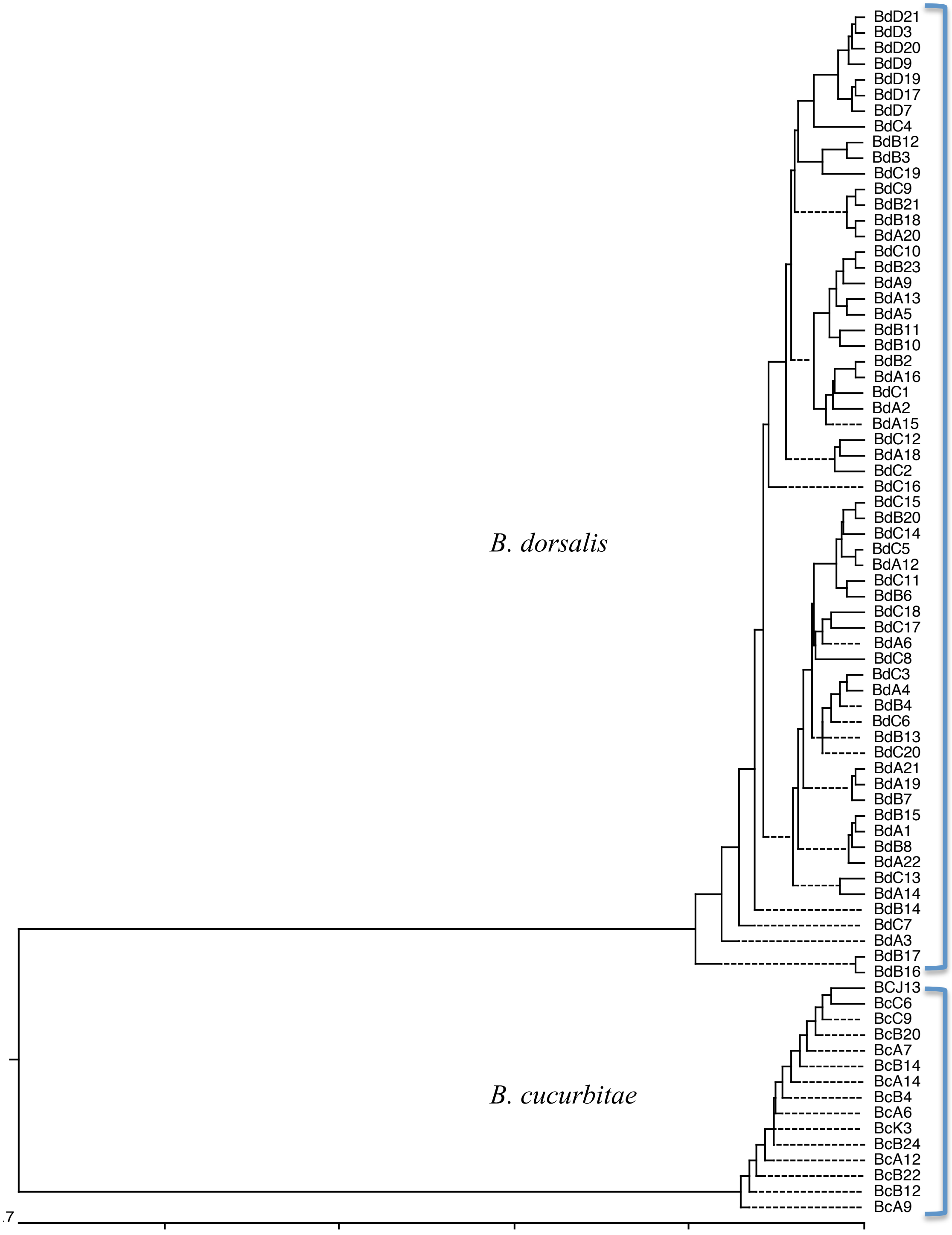


Figure 26. Phylogenetic tree of individuals of *B. dorsalis* and *B. cucurbitae* from Hawaii and Vietnam

The TCS network analysis of *B. dorsalis* and *B. curcurbitae* collections in Vietnam and Hawaii was performed and as can be seen, many changes were observed separating *B. curcurbitae* from *B. dorsalis* (**Figure 27**). The dash lines that cut perpendicular to the straight lines represents single changes between sequence types. The large black filled circle represents identical sequences. The large numbers of dashed lines separating the *B. dorsalis* network from that of *B. curcurbitae* network indicates that these two species are well separated at the DNA level.

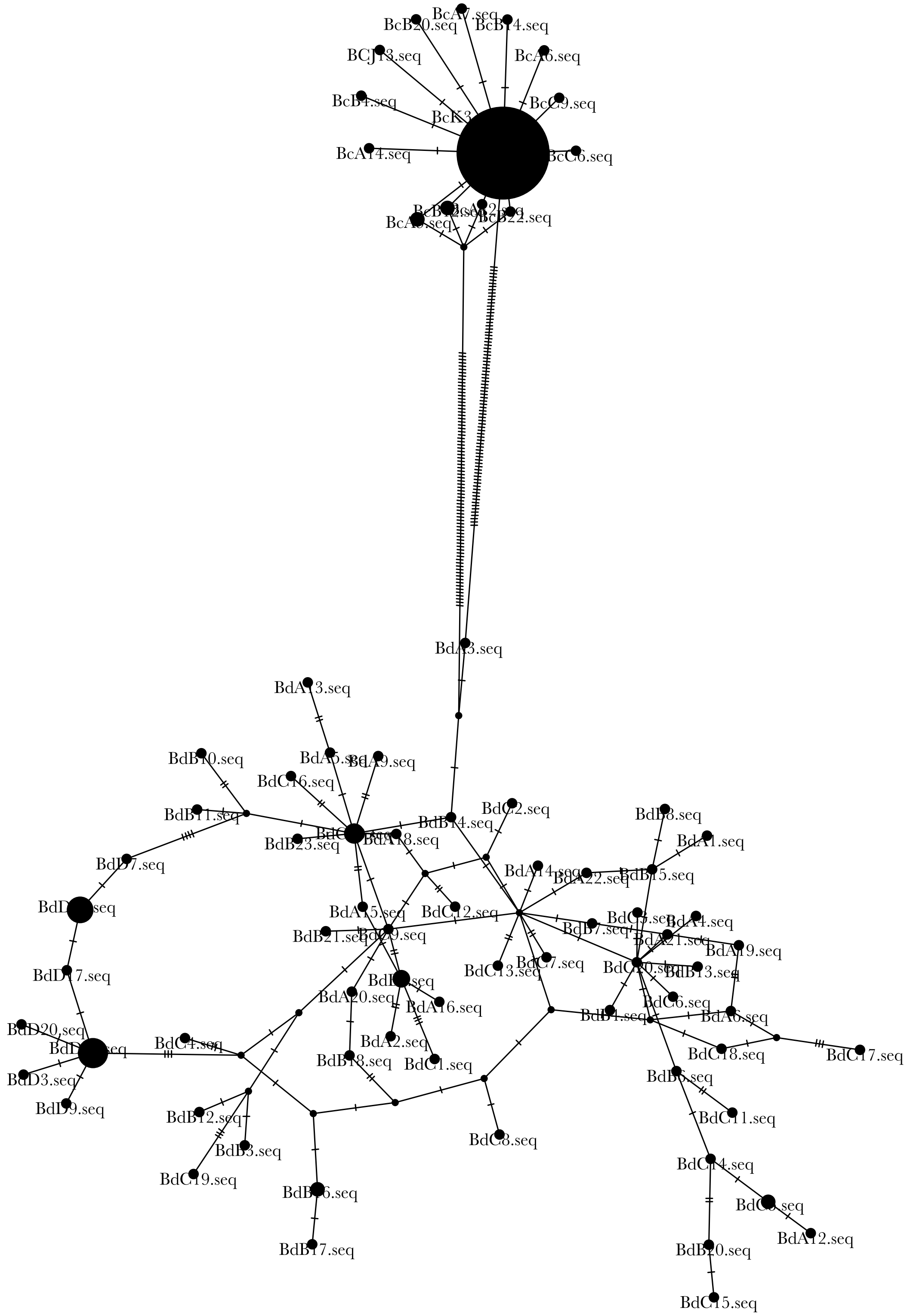


Figure 27. TCS Network of *B. dorsalis* and *B. curcurbitae* collections in Vietnam and Hawaii

Diversity measures between the *B. dorsalis* and *B. cucurbitae* datasets using popart software are shown in **Table 7**.

Table 7. TCS Network statistics

Statistics	***Bacterocera dorsalis***	***Bacterocera cucurbitae***
Nucleotide diversity	0.0117237	0.000602076
Segregating sites	60	15
Parsimony-informative sites	29	2

In *B. dorsalis*, the value of nucleotide diversity is much larger compared to that of *B. cucurbitae*. This shows that there is more diversity in *B. dorsalis* than in *B. cucurbitae.* The number of segregating sites provides an indicator of the degree of DNA sequence variation that is present in a sample.[12] Again, *B. dorsali*s shows higher segregating site compared to *B. cucurbitae*. The parsimony-informative sites of the *B. dorsalis* is much greater compared to *B. cucurbitae* parsimony-informative sites.

Conclusion and Discussion

The molecular taxonomic identification method is important for implementing alternatives to the application of chemical pesticides. This requires the ability to make correct species identification, and the results have shown that *B. dorsalis* and *B. cucurbitae* are easy to correctly identify and distinguish.

The geographic populations of *B. dorsalis* and *B. cucurbitae* from Vietnam, as well as samples from the Hawaii have been studied using a 547 base pair fragment of COI gene sequence. The observed genetic diversity of *B. cucurbitae* was exceedingly low, considering the geographic scale of the sampling that is relatively widespread and diverse in terms of biotypes. It can be concluded that *B. cucurbitae* exists Vietnam and Hawaii as a single phyletic lineage with no sign of cryptic species or historical separation.

For *B. dorsalis,* the collection from Vietnam shows much more variation compared to Hawaii. In contrast, the *B. cucurbitae* shows little variation in both Vietnam and Hawaii. At the DNA level, and *B. dorsalis* is significantly different compared to *B. cucurbitae*. Within *B. dorsalis*, at the population level, individuals showed fairly high levels of nucleotide diversity with relatively little sharing of haplotypes among populations. The result suggests that the *B. dorsalis* complex represents a rapidly evolving species complex with sign of historical separation and great evolutionary radiation.

The percentage identity of *B. dorsalis* is relatively lower compared to *B. cucurbitae* since there are many polymorphisms observed in *B. dorsalis.* The divergence value of *B. dorsalis* is much higher when compared to *B. cucurbitae* divergence values. The phylogenetic analysis was performed following alignments using the Clustal W method. The phylogenetic tree of *B. dorsalis* was highly branched, while the phylogenetic tree of *B. cucurbitae* showed very few branches. The TCS network also highlighted these same differences. The *B. cucurbitae* network was relatively compact with high numbers of identical individuals, while the TCS network of *B. dorsalis* was spread out. *B. cucurbitae* also showed much smaller values of nucleotide diversity, segregating sites and parsimony-informative sites. The TCS network also displayed great differences between two species by many dash lines cutting between the two networks. As a result, one can conclude that *B. dorsalis* is very different compared to *B. cucurbitae*.

B. dorsalis sequences obtained from Vietnam shows greater diversity compared to the sequences collected on the Hawaii obtained from NCBI GenBank. This may be explained by gene flow and migration between populations. In fact, Hawaii is isolated; therefore, migration is limited compared to Vietnam.

One of the great values of a molecular taxonomic approach is that traditional taxonomy is facing a crisis because fewer specialists are available. Therefore, the use of molecular taxonomy helps taxonomy overcome this problem. With highly developed computational technology, DNA sequences have become the major source of new information for advancing our understanding of evolution and genetic relationships. Nowadays, the cost of labor is rising rapidly while the cost of computational technology keeps falling. The molecular taxonomy method provides us with the tools that make a universal DNA-based taxonomy system that anyone can access.

The National Center for Biotechnology Information contained DNA data; however, it is not directly suitable for taxonomic purpose because there is no established taxonomic standards that the submitters have to follow. Therefore, there is no guarantee that the correct species names were assigned by the submitter of the sequence. Molecular taxonomy could help us out of the need for a new database. Furthermore, the collected sequences are not only useful for identification, but they also constitute an invaluable resource for phylogenetic analysis. Phylogenetic trees will show differences, molecular evolution, diversifications, biodiversity and etc.

Beside DNA sequences from the mitochondria COI genes, mitochondrial genes (NAD4, 16S, etc.), as well as nuclear gene sequences such the internal transcribed sequences (ITS) regions of the ribosomal genes can provide useful markers for the analysis of genetic relationships among closely related species. Knowing the molecular levels of these closely related species allows researchers to apply biological control methods more accurately.

References

[1] Shelly TE, Epsky N, Jang EB, Reyes-Flores J, Vargas RI (Eds) (2014) Trapping and the Detection, Control, and Regulation of Tephritid Fruit Flies, Springer, Dordrecht, 638 pp.

[2] Mau (April 2007). Bactrocera dorsalis (Hendel). Crop Knowledge Master. http://www.extento.hawaii.edu/Kbase/Crop/Type/bactro_d.htm (24 June 2008).

[3] Kagan Owens, Jay Feldman, John Kepner. (2010). *Wide Range of Diseases*. Retrieved from Pesticide-Induced Diseases Database: http://www.beyondpesticides.org/health/pid-database.pdf

[4] Scheffer, S., M. L. Lewis, and R. C. Joshi. 2006. DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines. Ann. Entomol. Soc. Am. 99: 204-210.

[5] Rugman-Jones, P. F., R. T. Wharton, Van Noort and R., Stouthamer. 2009. Molecular differentiation of the Psyttalia concolor (Szepligeti) species complex (Hymenoptera: Braconidae) associated with olive fly, Bactrocera oleae (Rossi) (Diptera: tephritidae), in Africa. Biol. Control 49: 17-26.

[6] Clare, E. L., B. K. Lim, M.D. Engstrom, J.L. Eger, and P.D.N. hebert. 2007. DNA barcoding of Neotropical bats: species identification and discovery within Guyana. Mol. Ecol. Res. 7: 184-190.

[7] Waugh, J. 2007. DNA Barcoding in animal species: progress, potential ans pitfalls. Bioessays 29: 188-197.

[8] Tusnády GE, Simon I, Váradi A, Arányi T. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. Nucleic Acids Research. 2005;33(1):e9. doi:10.1093/nar/gni012.

[9] Barr, N. B., Ledezma, L. A. and Leblanc, L. Genetic diversity of Bactrocera dorsalis (Diptera: Tephritidae) on the Hawaiian Islands: Implications for an introduction pathway into California. http://www.ncbi.nlm.nih.gov/nuccore/KF801366.1

[10] Liat's data (a student who worked in Dr. Haymer research lab).

[11] Nei, M.; Li, W. *Proc. Natl. Acad. Sci. USA*, **1979**, 76 (10), 5269-5273.

[12] Fu Y. X. *Theorical Population Biology journal*, **1955**, 48(@), 172-197.