Identification and characterization of *Gypsophila paniculata* color morphs in Sleeping Bear Dunes National Lakeshore, MI, USA.

Marisa L. Yang[1] , Emma Rice[2], Hailee Leimbach-Maus[2], and Charlyn G. Partridge[2]

[1]Environmental Science Policy, and Management, University of California, Berkeley, 260 Mulford Hall, Berkeley, CA 94720, email: marisayang79@gmail.com
[2]Annis Water Resources Institute, 740 W. Shoreline Dr., Muskegon, MI 49441, email: riceemm@mail.gvsu.edu; hailee.pavisich@gmail.com

Corresponding author: Charlyn Partridge, Annis Water Resources Institute, Grand Valley State University, 740 W. Shoreline Dr., Muskegon, MI; 49441, email: partridc@gvsu.edu

Abstract

Gypsophila paniculata (baby's breath) is an invasive species found throughout much of the northwest United States and western Canada. Two distinct color morphs of baby's breath were identified in areas within the coastal dunes along eastern Lake Michigan. The more common morph produces stems that are purple in color (purple morph), while the atypical morph has stems that are green-yellow (green-yellow morph). The purpose of this study was to characterize these morphs and determine if they are genetically distinct species in order to assess whether alternative management strategies should be employed to control these populations. We sequenced two chloroplast regions, RbcL and MatK, and one nuclear region, ITS2, from the purple morphs and green-yellow morphs collected from Sleeping Bear Dunes National Lakeshore, MI, USA. Sequences were aligned with reference *G. paniculata* and *G. elegans* sequences obtained from GenBank and the Barcode of Life (BOLD) databases. Phylogenetic analyses suggest that the two color morphs belong *to G. paniculata* and do not appear to be distinct genetic strains. We propose that current management practices continue to treat the two color morphs in a similar manner in terms of removal to prevent the further spread of this species.

Introduction

The Great Lakes sand dunes comprise the most extensive freshwater dune complex in the world, stretching over 1,000 km^2 in Michigan alone. Within northwest Michigan, the sand dunes ecosystem is vital both environmentally and economically. It is home to a number of threatened and endangered species, including piping plover (*Charadrius melodus*) and Pitcher's thistle (*Cirsium pitcheri).* The colonization of invasive species into this region can significantly alter the biological composition of native communities (Olson 1958, Leege and Murphy 2001, Emery et al. 2013). One invasive species of significant concern is the perennial baby's breath (*Gypsophila paniculata*). In 2015, baby's breath was listed by the Department of Natural Resources (DNR) as a "priority" invasive species for detection and control in Michigan's northern lower peninsula (DNR 2015). Since its colonization in the region it has spread along a 260 km stretch of the Michigan shoreline. Baby's breath produces a large primary taproot that can extend down to 4 meters in depth, which likely helps it outcompete native vegetation for limited resources (Darwent and Coupland 1966, Karamanski 2000). The presence of baby's breath in this system is associated with a reduction in the number of threatened Pitcher's thistle and an increase in the number of other non-native species (Emery et al. 2013). In addition, while many of the vulnerable and endangered plant species in these areas are seed limited, (e.g., Pitcher's thistle produces approximately 50-300 seeds per plant (Bevill et al. 1999)), baby's breath can produce up to 14,000 seeds per plant annually (Stevens 1957), effectively outcompeting native species in terms of overall yield. This has led to baby's breath composing approximately 50-80% of the ground cover in some areas (Karamanski 2000, Emery et al. 2013).

One concern with current management efforts is that there is some anecdotal evidence that there may be a new baby's breath variant (morph) within the Michigan dune system. In 2011 and 2012 The Nature Conservancy (TNC) removal crews reported baby's breath plants with different character traits than what is commonly observed (TNC 2014). The atypical morph has stems and leaves that are lighter in color and more yellow than the common purple morph (Figure 1a-c). The purple morph has a thick taproot (4-7 cm in diameter) just below the caudex that remains unbranched for approximately 60 – 100 cm (Darwent & Coupland 1966). Severing just below the intersection of the caudex and the taproot is where manual removal efforts target to limit regrowth. However, the atypical green-yellow morph's root system was reported by TNC to be more diffuse, making it harder to identify a primary taproot and thus, harder to sever without the potential for regrowth (TNC 2014). Currently, these green-yellow morphs are treated with herbicide application (glyphosate) when observed; however, if this morph continues to spread into areas where threatened or endangered species are present, removal methods will be a primary concern. If the green-yellow morph and the purple morph represent two discrete species, then alternative management strategies may need to be considered for these populations.

One of the first steps toward adapting current management strategies for this population is to identify whether the green-yellow morph is genetically distinct from the purple morph. While *G. paniculata* is the dominant invasive baby's breath species in northwest Michigan, there are a number of distinct *G. paniculata* commercial strains that could potentially invade these areas considering seeds are commonly sold in stores throughout the region (personal observation). In addition, other baby's breath species, including *G. elegans* are sold in this region in commercial wildflower packets. *G. elegans* is an annual species that has a smaller taproot compared to *G. paniculata,* and its coloration is similar to that of the green-yellow morph. Thus,

the goal of this work was to use traditional 'barcode' genes for plants to characterize the genetic relationship between the green-yellow morph and the common purple morph.

Methods and materials

DNA Extraction, Amplification, and Sequencing

We collected leaf tissue from 1 green-yellow morph and 16 purple morphs in 2016 and an additional 15 green-yellow morphs in 2017 from Sleeping Bear Dunes National Lakeshore (SBDNL), Empire, MI, USA (specifically: 44.884941 N, 86.062111 W and 44.875302 N, 86.056821 W). Leaf tissue was dried in silica gel until DNA extractions could take place. DNA was extracted using Qiagen DNeasy Plant Mini Kits (Qiagen, Hilden, Germany). After extraction, the DNA samples were placed through Zymo OneStep PCR inhibitor removal columns (Zymo, Irvine, CA) to remove any secondary metabolites that might inhibit PCR amplification. The DNA for each sample was then quantified using a NanoDrop 2000 (ThermoFisher, Waltham, MA).

The DNA of both green-yellow morphs and purple morphs was amplified at three proposed 'barcoding' genes for plants: large subunit of the ribulose-bisphosphate carboxylase gene (RbcL), maturase K (MatK), and internal transcribed spacer 2 (ITS2). The RbcL region was amplified using RbcL 1F and RbcL 724R primers (Chen et al. 2010), MatK was amplified using MatK 390F and MatK 1440R primers (Fior et al. 2006), and the ITS2 region was amplified using ITS2 2SF and ITS2 S3R primers (Chen et al. 2010). Individual PCR reactions for all loci consisted of 1X Taq Buffer, 2.0 mM $MgCl_2$, 0.3 µM dNTP, 0.08 mg/mL BSA, 0.4 µM forward and reverse primers, and 0.5 units of Taq polymerase in a 20 µL reaction volume. The thermal cycle protocols consisted of the following: for RbcL, an initial denaturing step of 95ºC for 2

minutes, followed by 35 cycles of 94ºC for 1 minute, 55ºC for 30 seconds, and 72ºC for 1 minute. A final elongation step was performed at 72ºC for 7 minutes. For MatK, the thermal profile consisted of 26 cycles of 94ºC for 1 minute, 48ºC for 30 seconds, and 72ºC for 1 minute, followed by a final elongation step at 72ºC for 7 minutes. For ITS2, an initial denaturing step of 95ºC for 2 minutes was applied, followed by 35 cycles of 95ºC for 30 seconds, 50ºC for 30 seconds, 72ºC for 1.5 minutes, and a final elongation step of 72ºC for 8 minutes. Successful amplification was checked by running the PCR product on a 2% agarose gel stained with ethidium bromide. PCR reactions were then cleaned using ExoSAP-IT PCR Product Cleanup Reagent (ThermoFisher, Waltham, MA). Sequencing reactions were performed for both forward and reverse primers for each of the three genes. Sequencing reactions were cleaned using a Sephadex column (GE Healthcare Life Science, Marlborough, MA) and sequenced on an ABI Genetic BioAnalyzer 3130xl (Applied Biosystems, Foster City, CA). Out of the 16 green-yellow morphs a total of 13 were successfully sequenced for RbcL, 13 were successfully sequenced for MatK, and 14 were successfully sequenced for ITS2. For the purple morphs a total of 15, 12, and 15 individuals were successfully sequenced for RbcL, MatK and ITS2, respectively.

G. paniculata and *G. elegans* reference sequences for RbcL, MatK, and ITS2 were downloaded either from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) or the Barcode of Life Database (BOLD) (http://www.barcodeoflife.org). Their accession numbers and sequences are provided in Supplemental Table 1. All fasta files corresponding to these data will be deposited in the Dryad database and sequences will be submitted to GenBank.

Alignment and Phylogenetic Analysis

All success sequences from our field samples, as well as sequences for *G. elegans* and *G. paniculata* obtained from GenBank or BOLD, were imported into the program

MEGA7 (version 7.0.14) (Kumar et al. 2016) and sequences for each of the three genes were aligned both individually and with all sequences combined using Muscle (Edgar 2004). The total number of base pairs aligned and analyzed for each region included: 516 base pair (bp) for RbcL, 720 bp for MatK, 264 bp for ITS2, and 1498 bp for the three regions combined. All alignment parameters were kept at their default settings. Once aligned, we used MEGA7 to identify the most appropriate substitution model (RbcL: Jukes-Cantor, MatK: Tamura 3-parameter, ITS2: Jukes-Cantor, all genes combined: Tamura 3-parameter). We then created phylogenetic trees using a maximum-likelihood (ML) approach with 500-replicated bootstrap analyzes, as well as using neighbor joining, and parsimony models. We also constructed a haplotype network based upon the combined sequences using the statistical parsimony approach (Templeton et al. 1992) in the program TCS (v 1.21) (Clement et al. 2000). For TCS, gaps were treated as missing data and we used a 95% connection limit.

Results and Discussion

Our results indicate that the green-yellow morph identified in SBDNL is not a genetically distinct species from the common purple *G. paniculata* found throughout SBDNL. The RbcL, MatK, ITS2, and combined dataset showed similar patterns with both the green-yellow morphs and the purple morphs clustering together with a *G. paniculata* reference sequence, and separately from *G. elegans* reference sequences (Figure 2a-d; Supplemental Figure 1-8). In addition, the TCS haplotype network shows that purple and green-yellow morphs have shared haplotypes. These haplotypes group within the same network as the *G. paniculata* reference, while the *G. elegans* references were grouped in an independent network (Figure 3). This suggests that both color morphs belong to *G. paniculata*.

Of the regions analyzed for the green-yellow morphs, purple morphs, and reference sequences, RbcL was the most conserved sequence with an overall mean genetic distance (*d)* = 0.001, followed by MatK (*d* = 0.002) and ITS2 (*d* = 0.018). For the ITS2 region, there were 6 purple *G. paniculata* morphs and one green-yellow morph that clustered together inside the *G. paniculata* branch (Figure 2a & 2d). Further examination of the chromatograms for these individuals show that they are likely heterozygous at position 195 of our aligned sequence and amplification bias of the 'A' SNP over the allele containing the 'G' SNP is driving this pattern.

The phylogeny constructed from the RbcL sequence showed one *G. paniculata* reference sequence (SDH831-14) grouping together with *G. elegans*. All reference sequences for RbcL were obtained from the BOLD database, and it is possible that this specific individual was misidentified. Further examination of this specimen from the herbarium image posted to BOLD indicates a small taproot compared to the upper foliage (http://www.boldsystems.org/pics/SDH/CCDB-24938-B09_h235508%2B1424896806.jpg), which is one characteristic of the annual *G. elegans*. While there is also sequence data for both the ITS2 and MatK regions for this specimen in BOLD, these were not used for this dataset due to higher levels of sequence fragmentation compared to our samples. However, both the ITS2 and MatK sequences from this *G. paniculata* reference sequence aligned most closely to *G. elegans* in BLAST.

While our data suggest that both the purple and green-yellow morphs are *G. paniculata*, whether the green-yellow morph is a different commercial genetic strain is still unclear. RbcL, MatK, and ITS2 are common 'barcode' genes used to delineate plant species, but their ability to distinguish genetically distinct strains of the same species can be limited. RbcL and MatK are highly conserved and would likely not distinguish between different cultivars; however, ITS2 is

more variable and has been used to distinguish between different varieties of domestic and imported teas (*Camellia sinensis*) (Lee et al. 2017). In our dataset, the purple and green-yellow morphs from SBDNL grouped separately from the *G. paniculata* reference sequence, but within the same clade. The *G. paniculata* reference sequence for ITS2 was obtained from GenBank and is part of an unpublished study, so we do not have information regarding where this sample was collected. However, given the distinct grouping between the *G. paniculata* reference sequence and our samples, ITS2 may be a promising region to assess the ability to distinguish among different commercial strains or regional varieties of *G. paniculata*. If this is the case, then our data would suggest that the purple and yellow-green morphs not only belong to the same species (*G. paniculata)*, but are likely derived from the same or similar initial source populations.

The mechanism driving the color difference between the purple and green-morphs is currently unknown. Within SBDNL, the purple morph is the most common form, with green-yellow individuals found interspersed in a couple of locations throughout the dunes. The largest observed group of green-yellow morphs consists of a few hundred plants clumped within approximately an acre-sized area and interspersed throughout large groups of purple morphs. Occasionally, small pockets of green-yellow individuals have also been found together in other areas of SBDNL and Zetterberg Preserve (approximately 38 km from SBDNL). Based upon the dispersal patterns of the two morphs throughout the dunes, the color difference observed does not appear to be solely environmentally driven, and likely has a genetic component. Potential candidate genes could include those involved in the anthocyanin pathway, which influences red – purple coloration in a number of plants (Asen et al. 1972, de Pascual-Teresa et al. 2002, Abdel-Aal et al. 2006, Knievel et al. 2009). Further work will begin to elucidate the specific mechanism

195 influencing these color difference in invasive *G. paniculata* populations, as well as to explore
196 whether this color difference drives functional differences between the morphs.

197 In terms of management strategies, the purple and green-yellow morph are not distinct
198 species and are likely derived from the same genetic strain, so there does not appear to be a need
199 to manage these two morphs differently. One concern with the green-yellow morph initially
200 noted by TNC removal crews was that the taproot tended to be more diffuse than the purple
201 morph potentially making manual removal of these plants less effective. We are currently
202 examining root growth between the green-yellow and purple morphs under controlled
203 conditions; however, our lab's personal observations in the field have not found any indication
204 that large differences in root structure occur between mature plants of the two color morphs.
205 Therefore, current management approaches for these populations should be maintained to control
206 the further spread of *G. paniculata* throughout the Michigan coastal dune system.

207 **Acknowledgements**

208 This project was funded through an EPA-Great Lakes Restoration Initiative Grant (C.G.P) and
209 through an NSF-REU grant (REU-QUEST) based at Annis Water Resources Institute. We would
210 like to thank Shaun Howard from The Nature Conservancy for his help in identifying the green-
211 yellow morph and Benjamin Giffin for his help with sequencing analysis. We would also like to
212 thank Kurt Thompson, Doug Haywick, and Brayden Partridge for assisting with the figure
213 construction.

214
215 **References**
216
217 Abdel-Aal, E.S.M., Young, J.C., and Rabalski, I. 2006. Anthocyanin composition in black, blue,
218 pink, purple, and red cereal grains. J. Agric. Food Chem. **54**(13): 4696-4704.
219

Asen, S., Stewart, R.N., and Norris, K.H. 1972. Co-pigmentation of anthocyanins in plant tissues and its effect on color. Phytochemistry **11**(3): 1139-1144.

Bevill, R.L., Louda, S.M., and Stanforth, L M. 1999. Protection from Natural Enemies in managing rare plant species. Conserv. Biol. **13**(6): 1323-1331.

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, X., Jai, X., Lin, Y., and Leon, C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PloS One **5**(1): e8613. doi: 10.1371/journal.pone.0008613L

Clement, M., Posada, D., and Crandall, K. A. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657-1659.

de Pascual-Teresa, S., Santos-Buelga, C., and Rivas-Gonzalo, J.C. 2002. LC–MS analysis of anthocyanins from purple corn cob. J. Sci. Food Agric. **82**(9): 1003-1006.

Darwent, A.L., and Coupland, R.T. 1966. Life history of *Gypsophila paniculata*. Weeds **14**(4): 313-318.

Department of Natural Resources (DNR). 2015. Michigan Invasive Species Grant Program Handbook. http://www.michigan.gov/documents/dnr/2015-MISGP-hndbk_491809_7.pdf?source=govdelivery

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**(5): 1792-1797.

Emery, S.M., Doran, P.J., Legge, J.T., Kleitch, M., and Howard, S. 2013. Aboveground and belowground impacts following removal of the invasive species baby's breath (*Gypsophila paniculata*) on Lake Michigan sand dunes. Restoration Ecol. **21**(4): 506-514.

Fior, S., Karis, P. O., Casazza, G., Minuto, L., and Sala, F. 2006. Molecular phylogeny of the Caryophyllaceae (Caryophyllales) inferred from chloroplast matK and nuclear rDNA ITS sequences. Am. J. Bot. **93**(3): 399-411.

Jukes, T.H. and Cantor, C.R. 1969. Evolution of protein molecules. *In* Mammalian Protein Metabolism. Edited by H. N. Munro. Academic Press, New York. pp. 21-132

Karamanski, T.J. 2000. Nationalized Lakeshore: The Creation and Administration of Sleeping Bear Dunes National Lakeshore. Online Book. National Park Service, Department of the Interior. http://www.cr.nps.gov/history/online_books/slbe/

Kumar, S., Stecher, G., & Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. **33**(7): 1870-1874.

Lee, S. C., Wang, C. H., Yen, C. E., and Chang, C. 2017. DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. J. Food Drug Anal. **25**(2): 260-274. doi: 10.1016/j.jfda.2016.06.008

Leege, L. M., Murphy, P. G. 2001. Ecological effects of the non-native *Pinus nigra* on sand dune communities. Can. J. Bot. **79**(4): 429-437.

Stevens, O. A. 1957. Weights of seeds and numbers per plant. Weeds **5**: 46-55.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Mol. Biol. Evol. **9**(4): 678-687.

Templeton, A. R., Crandall, K. A., Sing, C. F. 1992. A cladistics analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics. **132**: 619-633.

The Nature Conservancy (TNC). 2014. Lake Michigan Coastal Dunes Restoration Project. 2013 Field Season Report. The Nature Conservancy and Partners, Lansing, Michigan

Figure Legends

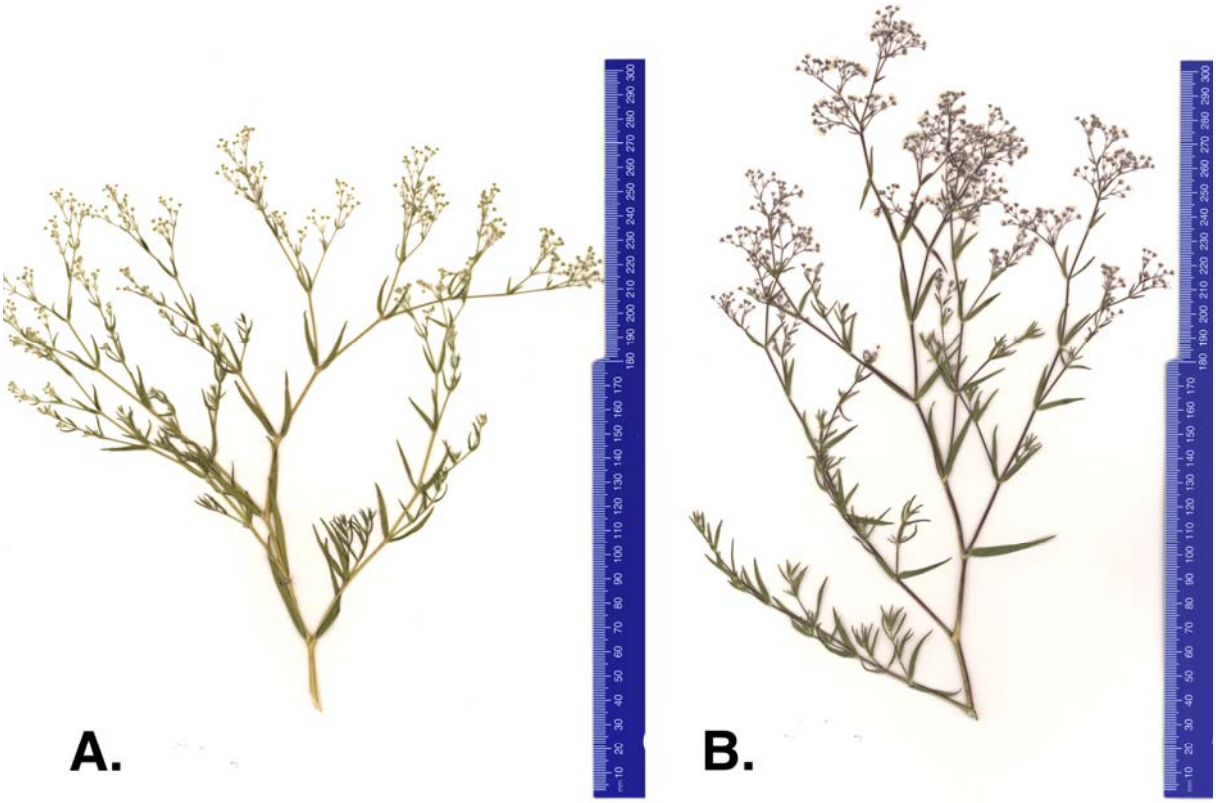
Figure 1: (A) Green-yellow morph, (B) common purple morph and (C) Stem of the green-yellow and purple baby's breath morph found in Sleeping Bear Dunes National Lakeshore.

Figure 2: Phylogenetic construction of the purple and green-yellow baby's breath color morphs in relationship to *G. paniculata* and *G. elegans* reference sequences. All evolutionary histories were inferred using maximum likelihood methods. (A) Phylogeny based on RbcL, MatK, and ITS2 combined, (B) phylogeny based on the RbcL region, (C) phylogeny based on the MatK region, (D) phylogeny based on the ITS2 region. For RbcL and ITS2 we used a Jukes Cantor model of molecular evolution (Jukes and Cantor, 1969). For MatK and the combined data set we used a Tamura 3-parameter model of molecular evolution (Tamura 1992).

Figure 3: A TCS haplotype network based on RbcL, MatK and ITS2 combined for the purple and green-yellow baby's breath color morphs, *G. paniculata* reference, and *G. elegans* references. The haplotypes with the highest outgroup probabilities are displayed as squares and the size of the square or oval corresponds to their haplotype frequency. The empty dots represent missing intermediate haplotypes. GY = Green-Yellow.

309
310
311

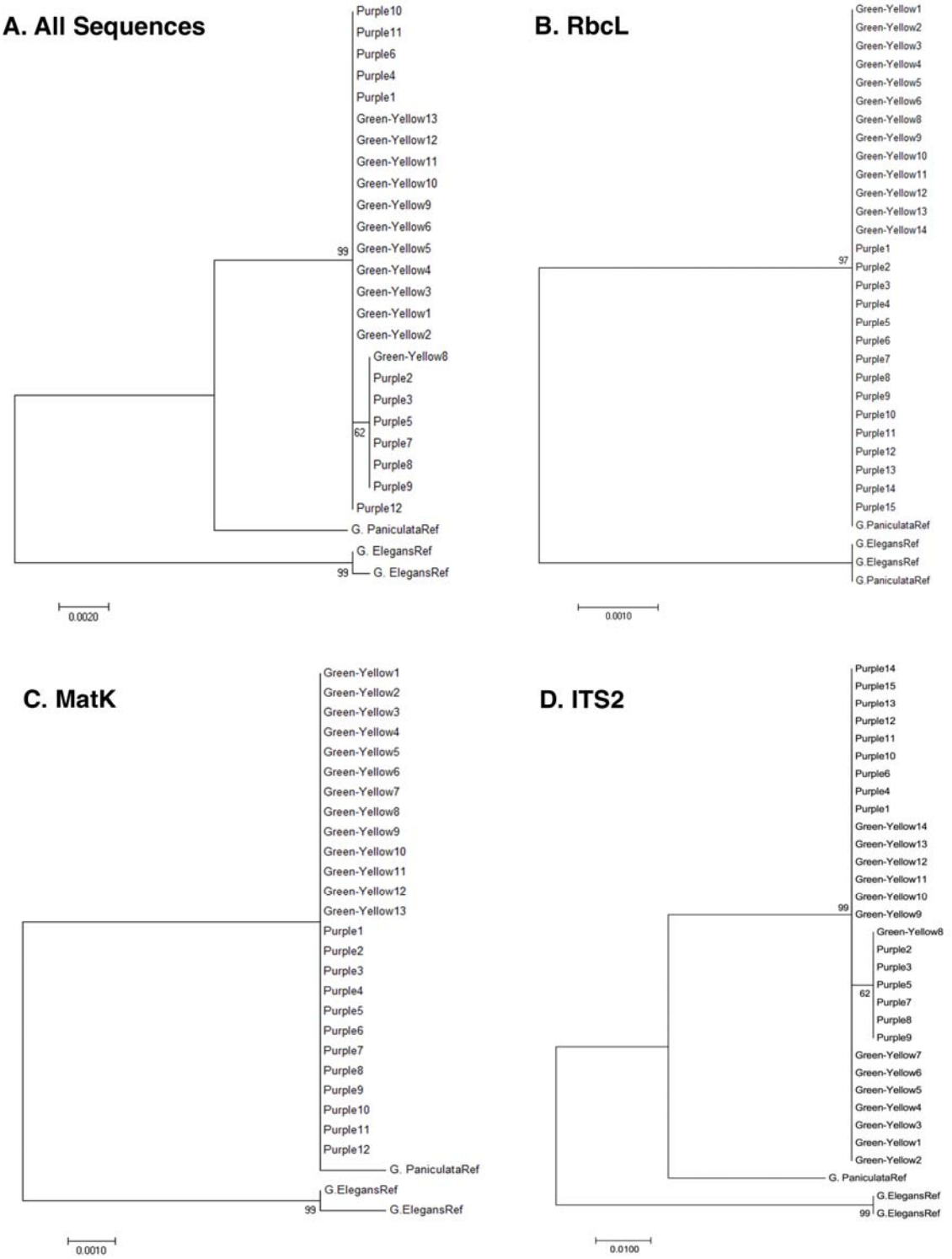
Figure 1



312

313
314

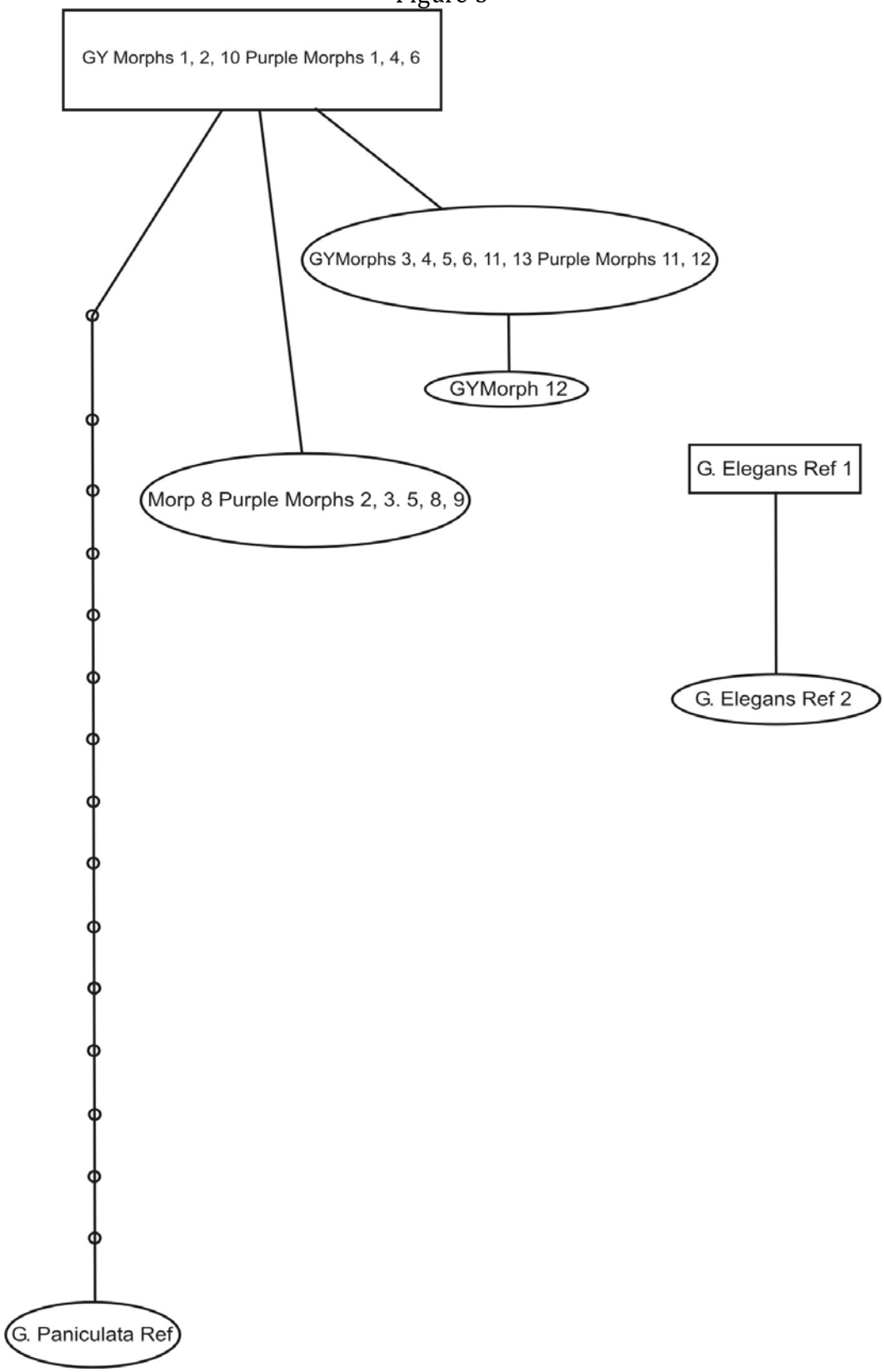
Figure 2



315
316
317

318
319

Figure 3



320