JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Use of the Chloroplast Gene ycf1 for the Genetic Differentiation of Pine Nuts Obtained from Consumers Experiencing Dysgeusia

Sara M. Handy,[*,†] Matthew B. Parks,[‡] Jonathan R. Deeds,[†] Aaron Liston,[‡] Lowri S. de Jager,[†] Stefano Luccioli,[§] Ernest Kwegyir-Afful,[§] Ali R. Fardin-Kia,[†] Timothy H. Begley,[†] Jeanne I. Rader,[†] and Gregory W. Diachenko[†]

[†]Center for Food Safety and Applied Nutrition, Office of Regulatory Science, U.S. Food and Drug Administration, College Park, Maryland 20740, United States

[‡]Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331-2902, United States

[§]Center for Food Safety and Applied Nutrition, Office of Food Additive Safety, U.S. Food and Drug Administration, College Park, Maryland 20740, United States

Supporting Information

ABSTRACT: Pine nuts are a part of traditional cooking in many parts of the world and have seen a significant increase in availability/use in the United States over the past 10 years. The U.S. Food and Drug Administration (US FDA) field offices received 411 complaints from U.S. consumers over the past three years regarding taste disturbances following the consumption of pine nuts. Using analysis of fatty acids by gas chromatography with flame ionization detection, previous reports have implicated nuts from *Pinus armandii* (Armand Pine) as the causative species for similar taste disturbances. This method was found to provide insufficient species resolution to link FDA consumer complaint samples to a single species of pine, particularly when samples contained species mixtures of pine nuts. Here we describe a DNA based method for differentiating pine nut samples using the ycf1 chloroplast gene. Although the exact cause of pine nut associated dysgeusia is still not known, we found that 15 of 15 samples from consumer complaints contained at least some *Pinus armandii*, confirming the apparent association of this species with taste disturbances.

KEYWORDS: pine nut, dysgeusia, pine mouth, *Pinus armandii*, ycf1, DNA

INTRODUCTION

The consumption of pine nuts in the United States has increased in the last 10 years. With it, reports of taste disturbances, known as dysguesia, have also increased.[1,2] The U.S. Food and Drug Administration has received several hundred complaints of such disturbances over the past few years.[3] Twenty-nine species of pine have seeds that are recognized as suitable for human consumption according to a Food and Agricultural Organization of the United Nations (FAO) report (accessed 7/14/11: http://www.fao.org/docrep/X0453E/X0453e12.htm). Both raw and roasted pine nuts, primarily from the genus *Pinus*, are used in a variety of dishes from pesto to tarts. *Pinus armandii* or Armand Pine (in the subgenus *Strobus*) is native to China but also occurs in Taiwan and Indo China (http://www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl). Nuts from *P. armandii* have recently been implicated as the causative species for pine nut associated dysgeusia, also known as "pine mouth".[4,5] However, the actual chemical agent responsible for these taste disturbances, and whether it is natural or incurred, or the result of a particular individual hypersensitivity, have yet to be determined.[4−6]

Eight species of pine nuts are commonly exported from China including *Pinus koraiensis*, *P. sibirica*, *P. yunnanensis*, *P. griffithii* (synonym *P. wallichiana*), *P. pumila*, *P. tabulaeformis* and *P. massoniana*. They are used for a variety of commercial purposes including use in birdseed, pasta manufacturing, production of oils, etc. (accessed 7/28/11: http://www.nutfruit.org/inc-projects/chinese_pinenuts). *P. armandii* is not currently listed by the FAO as appropriate for human consumption and, according to a press release from the International Nut and Dried Fruit Council dated April 18th, 2011 (INC, 2011), is no longer being exported from China to Europe. These nuts had been previously shipped to the United States for use in pasta manufacturing (see link above). *P. armandii* was also recently reported by the Codex Committee on Pesticide Residues (para. 88 of Alinorm REP11/PR, Report of the 43rd session accessed 7/29/11: http://www.codexalimentarius.net/web/archives.jsp?lang=en) as not fit for human consumption.

Methods for DNA based species identification in food products have greatly increased in the recent past, and have been extended to include products such as dietary supplements,[7] many types of seafood,[8] spices,[9] and some types of nuts.[10,11] The development of DNA barcoding (i.e., the use of short standardized sequences in a genome to discriminate species[12]) has driven this field forward even further,[13−15] including barcoding designed to identify plants.[16−18] There are several known challenges in plant DNA based identification including the fact that plants have low substitution rates in many of the regions thus far explored as potential DNA barcodes,[16] and lack of agreement on which genes to use, either alone or in combination.[16,19]

Received: August 10, 2011
Accepted: September 20, 2011
Revised: September 19, 2011
Published: September 20, 2011

 © 2011 American Chemical Society

Table 1. Authenticated Specimen

species and authority	other information	subgenus, subsection	GenBank accession	from
P. armandii Franch.	ARMA 09: Tai-pei-shan, Shensi, China (419−84-A)	*Strobus, Strobus*	JN573391	Oregon State University
P. armandii Franch.	A140	*Strobus, Strobus*	JN573393	International Nut and Dried Fruit Foundation 288
P. armandii Franch.		*Strobus, Strobus*	JN573394	F. W. Schumacher, USA
P. armandii Franch.	sample 17 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573392	Lawyer Nursery Inc., USA
P. cembra var. sibirica Du Tour	sample 23 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573382	Lawyer Nursery Inc., USA
P. cembra L.	sample 19 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573381	Lawyer Nursery Inc., USA
P. edulis Engelm.	sample 3 in Fardin-Kia et al.[6]	*Strobus, Cembroides*	JN573378	Lawyer Nursery Inc., USA
P. gerardiana Wall. ex D. Don	sample 1 in Fardin-Kia et al.[6]	*Strobus, Gerardianae*	JN573380	Lawyer Nursery Inc., USA
P. griffithii syn. wallichiana A. B. Jacks.	sample 24 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573395	F. W. Schumacher, USA
P. kochiana Klotzsch ex K. Koch, (potentially a synonym of *P. sylvestris var. hamata* Steven)	NA74255	*Pinus, Pinus*	JN573374	US Department of Agriculture, Agriculture Research Service
P. koraiensis Siebold et Zucc.	NA71577	*Strobus, Strobus*	JN573386	US Department of Agriculture, Agriculture Research Service
P. koraiensis Siebold et Zucc.	K116	*Strobus, Strobus*	JN573387	International Nut and Dried Fruit Foundation 288
P. koraiensis Siebold et Zucc.	sample 10 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573384	F. W. Schumacher, USA
P. koraiensis Siebold et Zucc.		*Strobus, Strobus*	JN573385	Lawyer Nursery Inc., USA
P. lambertiana Douglas	sample 9 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573389	Lawyer Nursery Inc., USA
P. lambertiana Douglas		*Strobus, Strobus*	JN573388	Pinyon Penny's, USA
P. lambertiana Douglas	W6 30959	*Strobus, Strobus*	JN573390	US Department of Agriculture, Agriculture Research Service
P. massoniana Lamb.	*MASS02*, Vietnam	*Pinus, Pinus*	JN573375	Oregon State University
P. monophylla Torr. et Frém.	sample 5 in Fardin-Kia et al.[6]	*Strobus, Cembroides*	JN573379	Pinyon Penny's
P. pinea L.	sample 8 in Fardin-Kia et al.[6]	*Pinus, Pinaster*	JN573373	Sardinia/Italy
P. pumila (Pall.) Regel	sample 22 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573398	F. W. Schumacher, USA
P. pumila (Pall.) Regel		*Strobus, Strobus*	JN573397	Lawyer Nursery Inc., USA
P. sibirica Du Tour	C120	*Strobus, Strobus*	JN573383	International Nut and Dried Fruit Foundation 288
P. tabulaeformis Carriére	sample 26 in Fardin-Kia et al.[6]	*Pinus, Pinus*	JN573376	F. W. Schumacher, USA
P. taeda L.	NA76041	*Pinus, Australes*	JN573372	US Department of Agriculture, Agriculture Research Service
P. wallichiana A. B. Jacks.	sample 25 in Fardin-Kia et al.[6]	*Strobus, Strobus*	JN573396	Lawyer Nursery Inc., USA
P. yunnanensis Franch.	sample 28 in Fardin-Kia et al.[6]	*Pinus, Pinus*	JN573377	F. W. Schumacher, USA

Many species in the genus *Pinus* are closely related genetically.[20,21] The genus *Pinus* is separated by taxonomists into two subgenera, including several sections and subsections[20,22] which contain the following species known to produce edible pine nuts: subgenus *Strobus* (soft pines), *Pinus ayacahuite*, *P. albicaulis*, *P. cembra*, *P. flexilis*, *P. koraiensis*, *P. lambertiana*, *P. monticola*, *P. pumila*, *P. sibirica*, *P. gerardiana*, *P. johannis*, *P. maximartinezii*, *P. monophylla*, *P. nelsonii*, *P. pinceana*, *P. remota*, *P. cembroides*, *P. culminicola*, *P. discolor*, *P. edulis*, and *P. quadrifolia*; and subgenus *Pinus* (hard pines), which contains *P. coulteri*, *P. pinea*, *P. ponderosa*, *P. sabineana*, *P. roxburghii*, and *P. torreyana*.

Recent work done by Ran et al.[23] has shown that seven typical barcode regions do not work to differentiate species in a closely related spruce genus, *Picea* (Pinaceae). Parks et al.[24] determined that the chloroplast ycf1 gene is more variable and specifically a ca. 1−2 kbp region (region C to D) displayed high sequence divergence across the genus *Pinus*. While not strictly a "barcode" which implies a short read of DNA that is conserved among many groups, this locus potentially allows resolution of very closely related species of pine.

From July 2008 to June 2011 FDA field offices received 411 complaints from consumers who reported dysgeusia consistent with a clinical entity referred to as "pine mouth syndrome".[2] All consumers reported dysgeusia characterized by a bitter or metallic aftertaste with any food intake usually beginning hours to days after consuming pine nuts and lasting in some cases up to 2 weeks. Based on complaint reports and FDA survey questionnaires no obvious demographic, social or medical factors or conditions were found to be associated with these complaints. In order to understand the association of pine nuts with the reported dysgeusia symptoms, fifteen consumer complaint samples were collected and brought to the FDA to examine them for a variety of compounds including pesticides, lipid oxidation[3] and fatty acid profiles.[6] After initial attempts to identify complaint samples using fatty acid profiles, according to Destaillats et al.,[4] it was determined that a more specific method of species identification was required.[6] Here we describe a genetic method developed to identify pine nuts in FDA complaint samples to the species level, in most cases, using the ycf1 gene.

Table 2. Complaint Samples

sample no.	no. of seeds successfully sequenced	sample numbers in Fardin-Kia et al.[6]	package origin	genetic assessment (% composition of *P. armandii*)
1	13	36, 38, and 40	China	*P. armandii* (100%)
2	12	41	China	*P. armandii* (100%)
3	7	39	China	mixture of *P. cembra/sibirica* and *armandii* (43%)
4	6	42	China	mixture of *P. cembra/sibirica* and *armandii* (67%)
5	4	44	China	mixture of *P. cembra/sibirica*, ~*pumila* and *armandii* (25%)
6	9	30	unknown	mixture of *P. koraiensis* and *armandii* (11%)
7	5	32	China	mixture of *P. koraiensis* and *armandii* (20%)
8	4	33	China	mixture of *P. armandii* (75%) and ~*P. gerardiana*
9	6	43	China	mixture of *P. koraiensis, cembra/sibirica* and *armandii* (50%)
10	10	37	unknown	*P. armandii* (100%)
11	9	29 and 35	China	mixture of *P. koraiensis* and *armandii* (30%)
12	10	34	China	mixture of *P. koraiensis* and *armandii* (30%)
13	5	31	China	mixture of *P. koraiensis* and *armandii* (60%)
14	7	45	unknown	*P. armandii* (100%)
15	9	46	unknown	mixture of *P. cembra/sibirica* and *armandii* (20%)

Table 3. Primers for CD Region of ycf1 Gene in *Pinus*

primer	sequence (5′−3′)	use	citation
ycf1c	AAGATTTTGAAATTCGTCCTG	PCR	Parks et al.[24]
ycf1d	TACGACGTTTTGGAAGC	PCR	Parks et al.[24]
PineCDinternalF	AGAGCGGAAAAAGATAGAGGAA	sequencing	this publication
PineCDinternalR	TTCCTCTATCTTTTTCCGCTCT	sequencing	this publication
PineCDinternalMEF	GAATCAAAGTTCTAGAGGAA	sequencing	this publication
PineCDinternalMER	TTCCTCTAGAACTTTGATTC	sequencing	this publication
PineCDinternal3F	CTCTTATCATATAGTTATCTCAATTCTACAAA	sequencing	this publication
PineCDinternal3R	TTTGTTCTKAACATYTGATCTA	sequencing	this publication
PineCDinternal3EF	TAGATCAGATGTTCAGAACAAA	sequencing	this publication

MATERIALS AND METHODS

Collection of Authentic Specimens and Complaint Samples. Twenty-seven authenticated species standards representing the *Strobus* and *Pinus* subgenera of the genus *Pinus*, were collected from a variety of sources (Table 1). From each of these samples, one seed was extracted to yield DNA sequences to which the unknown complaint samples could be compared. The U.S. Food and Drug Administration and the Danish Veterinary and Food Administration collected a total of fifteen complaint samples (13 and 2, respectively) associated with consumer complaints of dysgeusia (Table 2). Samples ranged in size from ca. 50−150 seeds. At least five seeds were examined from each sample to determine their genetic identity. Additional nuts (max. 13) were analyzed from select samples depending on visual sample homogeneity. Among these, seeds that varied greatly in size and appearance were selectively subsampled to test if these represented different species (i.e., mixtures of species within an individual sample).

Cell Lysis and DNA Extraction. A sterile scalpel blade was used to cut into each pine nut (1 seed for each authenticated sample, 5−13 seeds for the complaint samples), and a small segment of megametophyte (tissue inside the seed coat not including the embryo) from the inside of the seed (~10 mg) was added to a sterile 1.5 mL microcentrifuge tube. DNA was extracted from tissue by use of a DNeasy Blood & Tissue kit. A negative control was included with each set of extractions. Reagent volumes were reduced to a quarter of the volume listed in the manual (50 μL of buffer ATL with 5.56 μL of proteinase K, followed by 55.56 μL of buffer AL and 55.6 μL of EtOH, following Handy et al.[14]). After this step, the seed fragments would not be completely dissolved, but only the liquid was transferred to the spin column, leaving the fragments behind. For the wash steps, 140 μL of AW1 and AW2 was used, followed by elution with 50 μL of buffer AE. Besides these changes, the manufacturer's protocol was followed, with the additional step of incubating the washed filters and elution buffer at 37 °C for 30 min to increase successful elution of DNA. Extracted DNA was used directly in the PCR with no dilution. Concentrations of DNA used, tested using a Nanodrop ND 1000 spectrophotometer (ThermoScientific, Wilmington, DE) ranged from 5 to 150 ng/μL, consistent with Handy et al.[14]

PCR Amplification of DNA. Trial primer sets: Four sets of primers were initially used to amplify the authenticated standards: chloroplast genes rbcL and matk,[17] ribosomal genes ITS2,[18] and finally targeting the chloroplast C−D region of the ycf1 gene.[24] Primers matKpkF4/matKpkR1 were used for matk, and rbcLaF/rbcLajf634R were used for rbcL.[16] S2F/S3R[18] were used to amplify ITS2, and ycf1c/ycf1d were used for the ycf1 gene.[24] Complaint samples were amplified by ycf1 primers only.

The PCR cocktail consisted of 6.25 μL of 10% trehalose solution, 2 μL of dd H_2O, 1.25 μL of 10× PCR buffer, 0.625 μL of 50 mM $MgCl_2$, 0.125 μL of 10 μM of both primers (each primer set above run in separate reactions), 0.062 μL of 10 mM dNTPs, 0.060 μL of Platinum Taq (5 U/μL), and 1 μL of undiluted DNA template/reaction (11.5 μL total). An Eppendorf Mastercycler ep gradient S thermocycler was used for each PCR with the following conditions. For matK: 94 °C for 2 min;

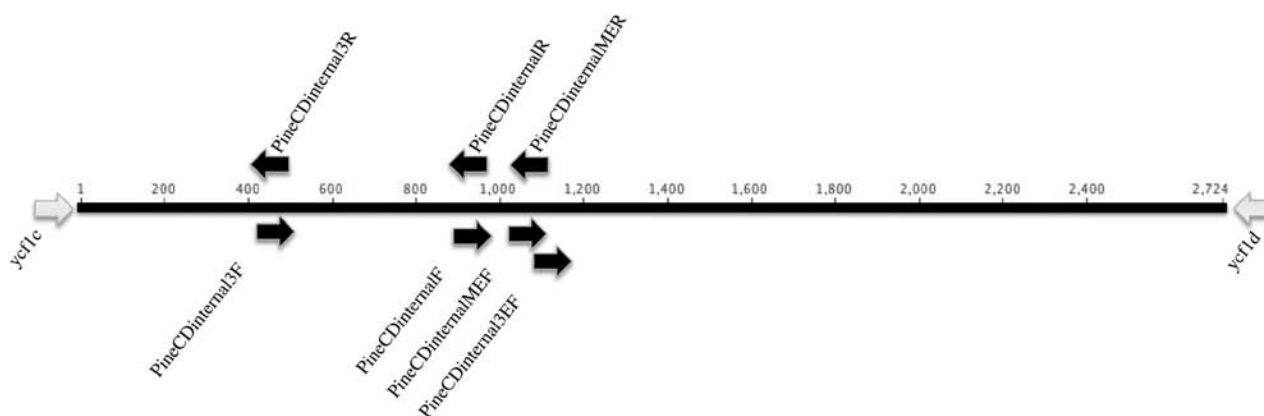

Figure 1. Map of primer locations on the ycf1 amplicon for all *Pinus* vouchers. Numbers depict positions in the full alignment (compare to Figure 3A). Gray arrows are PCR primers while black arrows are sequencing primers only. See Table 3 for primer sequences.

30 cycles of 94 °C for 15 s; 49 °C for 30 s; and 72 °C for 1 min, with a final extension at 72 °C for 1 min. For rbcL: 94 °C for 2 min; 30 cycles of 94 °C for 15 s; 55 °C for 30 s; and 72 °C for 1 min, with a final extension at 72 °C for 1 min. For ITS2: 94 °C for 5 min; 40 cycles of 94 °C for 30 s; 56 °C for 30 s; and 72 °C for 45 s, with a final extension at 72 °C for 10 min.[18] For ycf1: 98 °C for 30 s; 30 cycles of 98 °C for 8 s; 55 °C for 30 s; and 72 °C for 1.5 min, with a final extension at 72 °C for 5 min.[24] A selection of products were verified using precast 1.2% E-gel agarose gels according to the manufacturer's protocols with the E-Base Integrated power supply. Gels were run for 10−15 min and then visualized using a Gel Doc 2000 gel documentation system.

PCR Cleanup and Sequencing Reaction. Amplified products were purified by adding 2 μL of Exosap-IT to 5 μL of PCR product, and incubating at 37 °C for 15 min, followed by 15 min at 80 °C. For matK, rbcL, and ITS2 amplicons, the same primers used for PCR amplification were used for sequencing (see list above). For the ycf1 amplicons, at least 4 (for all subgenus *Strobus,* subsection *Strobus* samples) or up to 7 primers (for subgenus *Strobus*, subsection *Cembroides* and subgenus *Pinus*) were used in sequencing each seed (Table 3, Figure 1, more information regarding primer selection can be found below). Each reaction contained 0.25 μL of BigDye Terminator v3.1; 1.875 μL of 5× sequencing buffer; 5 μL of 10% trehalose; 1 μL of 10 μM primer (see Table 3); and 0.875 μL of molecular grade water, for a total of 9 μL, to which 1 μL of purified PCR product was added. Products were purified as illustrated in Handy et al. 2011 using an EdgeBio short well plate. At this point, samples were sequenced on an ABI 3730 instrument.

Sequence Editing. All ABI files were imported into Sequencher 4.9 (Gene Codes, Ann Arbor, MI). All sequences were individually trimmed for quality according to the following: for the 5′ end, trimming no more than 25%, trim until the first 25 bases contain fewer than three ambiguities, and trimming no more than 25%, trim until the first 25 bases contain fewer than three with confidences below KB basecaller score of 20. For the 3′ end, starting 100 bases after the 5′ trim, trim the first 25 bases containing more than three ambiguities; trim from the 3′ end until the last 25 bases contain fewer than three ambiguities; and trim from the 3′ end until the last 25 bases contain fewer than three bases with confidence below a KB basecaller score of 20. The "postfix" was set to: remove leading and trailing ambiguous bases. Next, the bidirectional sequences were assembled into contigs (with default settings: using dirty data algorithm, realigner and prefer 3 gap placement, as well as a 20-base minimum overlap and an 80% minimum match percentage). Each sample was resequenced until at least double coverage was reached and hand edited to ensure correct base calls.

Post Processing of Authenticated Specimens. Twenty-seven authenticated specimen sequences were exported from Sequencher as a FASTA file into the bioinformatics software Geneious Pro (Biomatters Ltd., Auckland, New Zealand[25]). An alignment was constructed in Geneious Pro using the "Muscle Alignment" tab,[26] with default settings. All subsequent alignments were processed this way. Sequences were analyzed using a UPGMA consensus tree (1000 replicates) with a Jukes−Cantor genetic distance model[27] option as well as a neighbor-joining consensus tree with a Jukes−Cantor genetic distance model with 1000 replicates[28] in Geneious Pro after the alignment was edited with the online software program GBLOCKS (http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=gblocks) with less stringent selection including: allow smaller final blocks, allow gap positions within the final blocks and allow less strict flanking positions.[29]

Post Processing of Complaint Samples. Sequences of all seeds for each sample were exported from Sequencher individually as FASTA files. The FASTA file was imported into Geneious Pro and merged with the authenticated sample alignment (prior to GBLOCKS trimming) into a new alignment. MUSCLE allows closely related species to group together, and therefore a determination was made about which subgenus, *Strobus* or *Pinus*, the unknown seeds aligned most closely with. In the case of seeds closely related to *P. armandii*, the subsection *Strobus* was examined by itself (Table 2). This subgenus or subsection was exported to a new alignment in Geneious Pro and refined (using the alignment tab). GBLOCKS was again used to edit each alignment in a reproducible way. Once the program had completed, the new edited alignment was imported back into Geneious Pro. Next, a UPGMA consensus tree with a Jukes−Cantor genetic distance model[28] was constructed. A species level determination was made if the seeds were imbedded in a particular group of authenticated specimens. If the seeds formed clusters outside a particular authenticated sample, but were closely related, a "∼" was used to identify it.

Primer Development. To increase sequence coverage for the ycf1 amplicons, 7 primers were developed specifically for sequencing the C−D section of the ycf1 gene (Table 3, Figure 1). Sequencing primers were designed by aligning the sequences of all 27 authenticated samples in Geneious Pro, and manually locating conserved sections. All primers were ordered from Integrated DNA Technologies (Coralville, IA). PineCDinternalF and -R were appropriate for both subgenera and usually gave bidirectional coverage when combined with the ycf1c and -d primers from Parks et al.[24] The other 5 primers listed were developed to sequence members from subgenus *Strobus*, subsection *Cembroides*, *Gerardianae* and also for subgenus *Pinus*, which have longer C−D regions.[24]

RESULTS

Selection of Appropriate Gene To Identify Pine Species. The currently accepted barcoding regions for plants, rbcL and

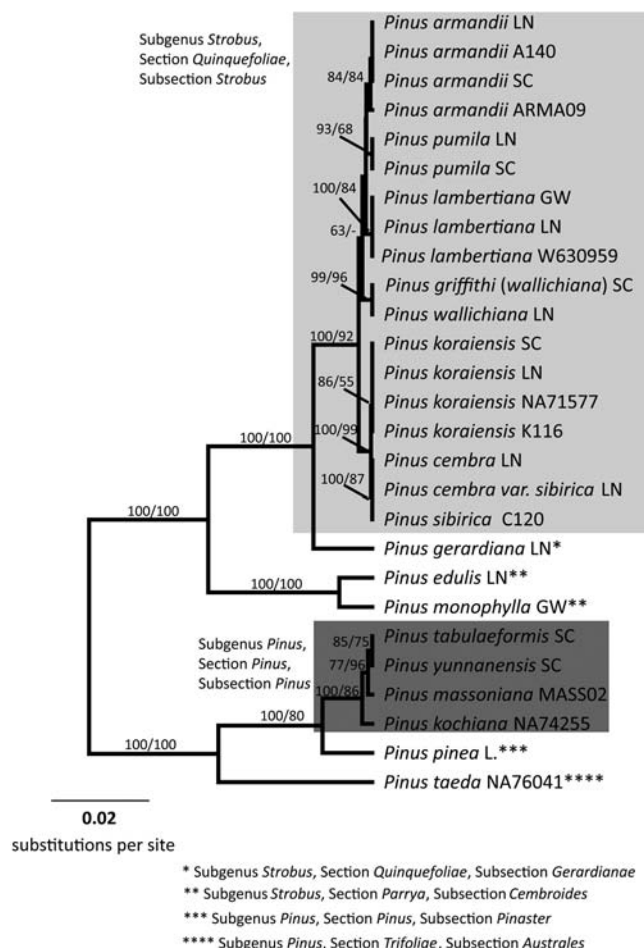

Figure 2. A UPGMA consensus tree using a Jukes−Cantor genetic distance model with 27 authenticated pine species. Values located on the branches are UPGMA bootstraps/neighbor-joining bootstraps each based on 1000 replications, all generated in the program Geneious Pro. Subgenus, section and subsection based on Gernandt et al. 2005[20] are noted.

matk[17] as well as the ribosomal locus ITS2[18] were used to attempt to differentiate among 27 authenticated pine nut species (data not shown). Initially, it was determined that the 2 subgenera of pine could be distinguished. However, among the *Strobus* subgenus and specifically the *Strobus* subsection, which is the group into which *P. armandii* falls, the short portions of the genes used were not variable enough to resolve the relationships among many of the species, even when the genes were combined in the analysis (data not shown). For example, in a combined alignment of rbcL and matK barcode regions (combined length 1318 base pairs) for subsection *Strobus*, none of the alignment positions were variable (data not shown). On the other hand, in an alignment of the C−D section of the ycf1 gene[24] for subsection *Strobus* (length was 1376 base pairs after trimming with GBLOCKS), 8.2% of alignment positions were variable, resulting in bootstrap support values ranging from 63 to 100% (UPGMA) and 55−100% (neighbor-joining) (Figure 2).

While the clustering method used here was not meant to be phylogenetically informative (i.e., suggest a specific evolutionary history), the tree generated was congruent with previously published phylogenetic analyses of *Pinus*.[30] Nonetheless, some closely related pine species could not be distinguished and were therefore grouped together in later analyses. These groups were: *P. cembra*/*sibirica* and *P. tabulaeformis*/*yunnanensis* (Figure 2).

Properties of Sequences. Twenty-seven sequences from the C−D region of the ycf1 gene of authenticated samples were deposited in GenBank (JN573372−JN573398, Table 1). They varied in length from 1390 to 1852 bases. 116 complaint sample sequences representative of multiple individual nuts from 15 complaint samples (Table 2) were deposited in GenBank (JN573399−JN573398). They varied in length from 1343 to 1742 bases.

Properties of Alignments. Figure 3A illustrates an alignment from Geneious Pro of all full length sequences of the C−D section of the ycf1 gene with a total length of 2724 bases, while Figure 3B shows the alignment trimmed to usable regions by the online program GBLOCKS. This reduces the usable alignment to 1355 bases. If just the *Strobus* subsection of the *Strobus* subgenus is extracted, the usable alignment increases to 1472 (Figure 3C).

Results of Genetic Analysis on Complaint Samples. Based on our analysis of 15 complaint samples, all 15 (100%) contained at least some *P. armandii*. We estimate that 4/15 (27%) contained only *P. armandii*, while 11/15 (73%) contained mixtures of two or more species. Of the mixed samples, 5/11 were *P. armandii* mixed with *P. koraiensis*, 3/11 (27%) were *P. armandii* mixed with *P. cembra*/*sibirica*, 1/11 (9%) was *P. armandii* mixed with ∼*P. gerardiana*, and 2/11 (18%) contained mixtures of 3 or more types of seeds (Table 2). Further, *P. armandii* represented 11−100% of the seeds we sampled, and all samples contained at least one seed of this species (Table 2). Figure 4 is an example of the trees used to identify the species that were found in a particular sample.

■ DISCUSSION

While Destaillats et al.[4] and Köbler et al.[5] have both proposed a method for identifying pine nuts using gas−liquid chromatography of fatty acids and nuclear magnetic resonance spectroscopy and chemometrics, respectively, their methods do not work definitively with mixtures,[6] which comprised at least 11/15 of the FDA consumer complaint samples tested here. To date, no DNA based method for differentiating pine nuts in commercial products has been proposed. The method described in this manuscript would allow regulators to use a method other than morphology and size to distinguish seeds, and to verify the presence or absence of *P. armandii* in a sample, which, as of this date, is assumed to be the "problem species" for dysgeusia.

Several "barcoding" loci were examined for use in species identification of pine nuts including matk, rbcL and ribosomal ITS genes (data not shown). They will not be discussed further here because none had comparable resolution to that found with the gene ycf1, most likely because of the short sequence length examined. While ycf1 is capable of discriminating among very closely related species,[24,31] which is essential for questions like ours, it does not meet the strict definition of a barcode locus due to the unavailability of "universal" (i.e., plant specific) primers.[32] Considering the variability of this locus, it may be necessary to design unique primers for most plant families. The problem of primer design has also been found with the plastid matK locus,[32] and it seems that different primers will be required for angiosperms and gymnosperms, for example.[33] It is likely that a two-tiered strategy, with identification to genus or family with rbcL, followed by more precise identification with matK or ycf1, when required, may be needed for many plant groups.

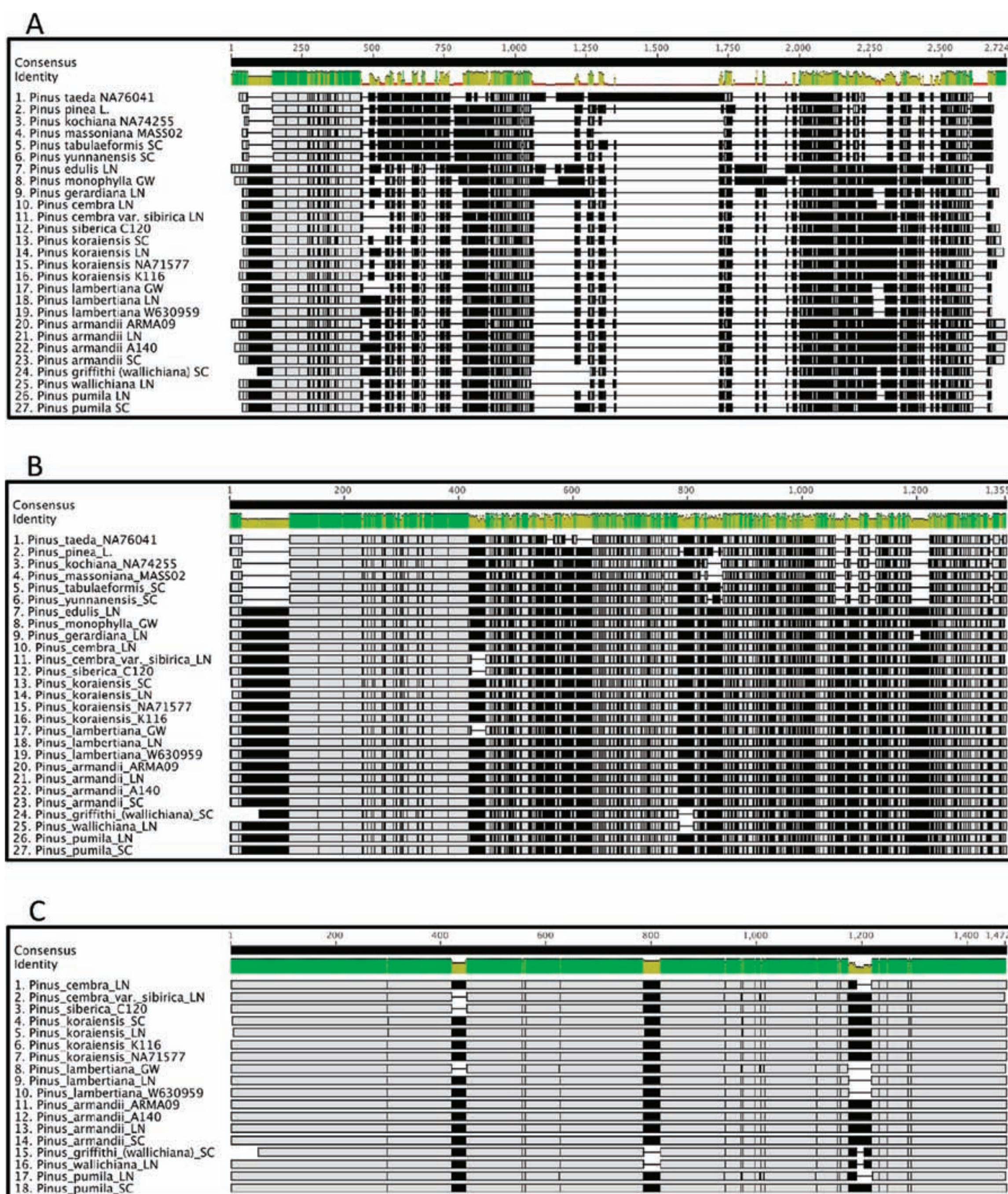

Figure 3. Geneious Pro alignments representing the 27 authenticated samples full length with a total length of 2724 bases (A), after trimming with GBLOCKS to a length of 1355 (B) and consisting of the *Strobus* subsection after extracting it from the original full length alignment and editing with GBLOCKS total length 1472 (C). The darker the area, the more variability at that particular location, while light gray regions signify conserved areas of the alignment.

Between the two subgenera and even between some of the subsections of pine, there are many regions of inserts and repeats. The original alignment in Figure 3A clearly shows these gap regions that are not appropriate to use in any sort of clustering analysis since they cannot be compared between species. These regions were removed by the program GBLOCKS (Figure 3B), however information is also lost in this trimming. The longest alignment possible as determined by GBLOCKS could be retained by extracting closely related members (giving researchers the best chance at determining species identity in a repeatable manner).

Although some authenticated species of pine could not be differentiated, using this region of the ycf1 loci, in these cases (i.e., *P. yunnanensis*/*P. tabulaeformis*, and *P. cembra*/*sibirica*, Figure 2) the seeds are all considered edible. *P. armandii* was easily resolved using the C−D section of the ycf1 gene from the other closely related *Strobus* species, including *P. lambertiana*, *P. cembra*, *P. sibirica*, and *P. koraiensis* (Figure 2).

The goal of this particular study was to present a more definitive method for differentiating pine nuts in response to reports that a specific pine nut species, *P. armandii*, is responsible

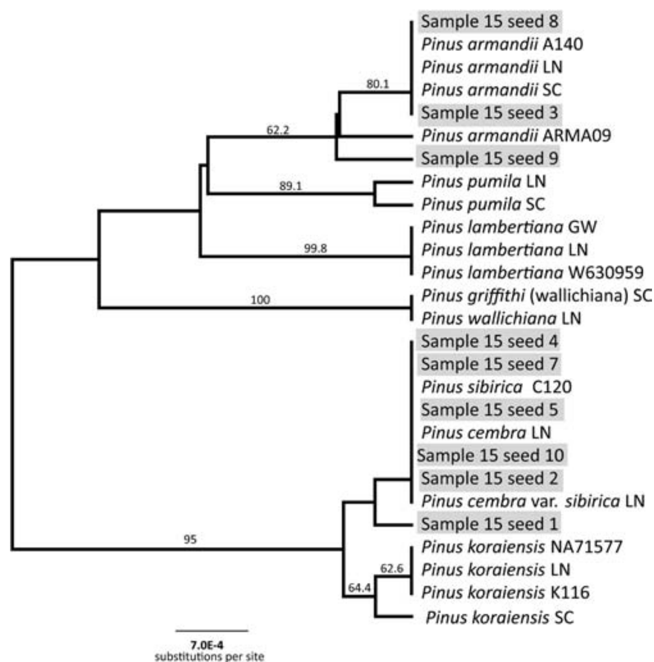

Figure 4. A UPGMA consensus tree using a Jukes−Cantor genetic distance model with an example complaint sample (sample 15). Values located on the branches are UPGMA bootstraps based on 1000 replications, all generated in the program Geneious Pro. This particular sample was a mixture of *Pinus armandii* and *Pinus cembra/sibirica*.

for the syndrome known as "pine mouth".[2] Without additional knowledge of the exact causes of dysguesia associated with the consumption of pine nuts, it is impossible to conclusively attribute this syndrome to one species. Moreover, studies prior to this one were forced to draw conclusions in the absence of information identifying whether samples were composed of one or more than one species. The method described here makes it possible to identify the exact species present, in most cases, and thereby provide a more reliable basis on which to develop additional regulatory strategies to address this issue. It also provides a stronger foundation on which to formulate hypotheses about the exact cause of "pine mouth".

All three reports on consumer complaints of pine nut associated dysgeusia (ref 5 and this publication) have identified at least some *P. armandii* in all complaint samples. The Chinese government and the International Tree and Nut Association have supported this assertion but stress in one report that *P. armandii* is exported to the United States only for use for pasta manufacturing (http://www.nutfruit.org/inc-projects/chinese_pinenuts) and will no longer be exported to the EU (Press release, April 18, 2011). All complaint samples analyzed in this study were collected as a ready to use product (i.e., bulk or loose pine nuts appropriate to be consumed without further preparation). Codex has also recently classified *P. armandii* as unfit for human consumption (REP 11/PR from the 43rd session of the Codex Committee on Pesticide Residues). Since we found that a majority of complaint samples were in fact mixtures, and the methods provided to date cannot reliably distinguish pine nut mixtures composed of the subsection *Strobus* group, the methods described here provide a reliable way to identify *P. armandii*. Based on the weight of the evidence (15/15 complaint samples contained at least some *P. armandii*), we are currently attempting to develop a method to identify *P. armandii* that relies solely on PCR (i.e., does not require subsequent sequencing) so that a larger number of individual seeds or combined mixtures can be tested at any one time. Ideally, such a method would provide a more rapid and inexpensive method for both researchers and regulators who need to indentify *P. armandii* in their samples. However, based on the high degree of sequence conservation seen in our alignments, it may not be possible to effectively develop this test. Nonetheless, the work presented herein at least provides an accurate, reasonably cost-effective method for screening pine nut samples for the presence of *P. armandii*.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. The full-length alignment generated by Geneious Pro, prior to trimming with GBLOCKS. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Sara.Handy@fda.hhs.gov. Phone: 240-402-3063.

■ ACKNOWLEDGMENT

The authors thank Dr. Charles Delwiche from the University of Maryland and the Norton-Brown Herbarium (MARY) for facilitating collaborations with the Oregon State University and also thank the University of Maryland Genomics core (run by Amy Beaven and Dr. Delwiche). Thanks to Danish Veterinary and Food Administration for providing us with consumer complaint samples and to the U.S. Department of Agriculture for providing authenticated pine specimen, and finally to Dr. Matthew Birck (US Customs and Border Protection) for helpful discussions. The authors also appreciate input from three anonymous reviewers who greatly improved the manuscript.

■ ABBREVIATIONS USED

PCR, polymerase chain reaction; DNA, deoxyribonucleic acid; UPGMA, unweighted pair group method with arithmetic mean; MUSCLE, multiple sequence comparison by log-expectation; FAO, Food and Agriculture Organization

■ REFERENCES

(1) Mostin, M. Taste disturbances after pine nut ingestion. *Eur. J. Emerg. Med.* **2001**, *8* (1), 76.

(2) Munk, M. D. Pine Mouth" syndrome: cacogeusia following ingestion of pine nuts (genus: *Pinus*). An emerging problem? *J. Med. Toxicol.* **2010**, *6* (2), 158–159.

(3) Kwegyir-Afful, E.; DeJager, L.; Begley, T.; Luccioli, S. Pine mouth syndrome- an emerging food hypersensitivity? *Clin. Transl. Allergy* **2011**, *1* (Suppl. 1), 25.

(4) Destaillats, F; Cruz-Hernandez, C; Giuffrida, F; Dionisi, F; Mostin, M; Verstegen, G. Identification of the Botanical Origin of Commercial Pine Nuts Responsible for Dysgeusia by Gas-Liquid Chromatography Analysis of Fatty Acid profile. *J. Toxicol.* **2011**, 1–7.

(5) Köbler, H.; Monakhova, Y. B.; Kuballa, T.; Tschiersch, C.; Vancutsem, J.; Thielert, G.; Mohring, A.; Lachenmeier, D. W. Nuclear magnetic resonance spectroscopy and chemometrics to identify pine nuts that cause taste disturbance. *J. Agric. Food Chem.* **2011**, *59* (13), 6877–6881.

(6) Fardin-Kia, A. R.; Handy, S. M.; Rader, J. I. Use of the highly polar SLB-IL111 GC column to determine the fatty acid composition of

different species of pine nuts and comparison with "Pine Mouth" associated pine nuts. Unpublished.

(7) Saker, M. L.; Welker, M.; Vasconcelos, V. M. Multiplex PCR for the detection of toxigenic cyanobacteria in dietary supplements produced for human consumption. *Appl. Microbiol. Biotechnol.* **2007**, *73*, 1136–1142.

(8) Rasmussen-Hellberg, R. S.; Morrissey, M. T. Advances in DNA-based techniques for the detection of seafood species substitution on the commercial market. *J. Lab. Autom.* **2011**, *16* (4), 308–321.

(9) Focke, F.; Haase, I.; Fischer, M. DNA-based identification of spices: DNA isolation, whole genome amplification, and polymerase chain reaction. *J. Agric. Food Chem.* **2011**, *59* (2), 513–520.

(10) Brzezinski, J. L. Detection of sesame seed DNA in foods using real-time PCR. *J. Food Prot.* **2007**, *70* (4), 1033–6.

(11) Brezná, B.; Dudásová, H.; Kuchta, T. A novel real-time polymerase chain reaction method for the detection of Brazil nuts in food. *J. AOAC Int.* **2010**, *93* (1), 197–201.

(12) Hebert, P. D. N.; Cywinska, A.; Ball, S. L.; deWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. London, Ser. B.* **2003**, *270*, 313–321.

(13) Colombo, F; Chessa, S; Cattaneo, P; Bernardi, C. Polymerase chain reaction products (PCR) on "DNA barcode zone" resolved by temporal temperature gradient electophoresis: A tool for species identification of mixed meat specimens - A technical note on preliminary results. *Food Control* **2010**, *22* (8), 1471–1472.

(14) Handy, S. M.; Deeds, J.; Ivanova, N.; Hebert, P.; Hanner, R.; Ormos, A; Weigt, L.; Moore, M.; Yancy, H. A single laboratory validated method for the generation of DNA barcodes for the identification of fish for regulatory compliance. *J. AOAC Int.* **2011**, *94*, 1.

(15) Rasmussen-Hellberg, R.; Naaum, A.; Handy, S. M.; Hanner, R.; Yancy, H.; Deeds, J.; Morrissey, M. Interlaboratory evaluation of a real-time multiplex polymerase chain reaction method for identification of salmon and trout species in commercial products. *J. Agric. Food Chem.* **2011**, *59*, 876–884.

(16) Fazekas, A. J.; Burgess, K. S.; Kesanakurti, P. R.; Graham, S. W.; Newmaster, S. G.; Husband, B. C.; Percy, D. M.; Hajibabaei, M.; Barrett, S. C. H. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* **2008**, *3* (7), e2802.

(17) Hollingsworth, P. M.; Forrest, L. L.; Spouge, J. L.; et al. CBOL Plant Working Grp. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106* (31), 12794–12797.

(18) Chen, S.; Yao, H.; Han, J.; Liu, C.; Song, J.; Shi, L.; Zhu, Y.; Ma, X.; Gao, T.; Pang, X.; Luo, K.; Li, Y.; Li, X.; Jia, X.; Lin, Y.; Leon, C. Validation of the ITS2 Region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* **2010**, *5* (1).

(19) Pennisi, E. Wanted: A barcode for plants. *Science* **2007**, *318*, 190–191.

(20) Gernandt, D. S.; Lopez, G.; Garcia, S. O.; Liston, A. Phylogeny and classification of *Pinus*. *Taxon* **2005**, *54*, 29–42.

(21) Whittall, J. B.; Syring, J.; Parks, M.; Buenrostro, J.; Dick, C.; Liston, A.; Cronn, R. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Mol. Ecol.* **2010**, *19* (Suppl. 1), 100–114.

(22) Farjon, A. *Pines: Drawings and descriptions of the genus Pinus*, 2nd ed.; Koninklijke Brill NV: Leiden, The Netherlands, 2005; pp 232−235.

(23) Ran, J. H.; Wang, P. P.; Zhao, H. J.; Wang, X. Q. A test of seven candidate barcode regions from the plastome in *Picea* (Pinaceae). *J. Integr. Plant Biol.* **2010**, *52* (12), 1109–1126.

(24) Parks, M.; Liston, A.; Cronn, R. Newly developed primers for complete YCF1 amplification in *Pinus* (Pinaceae) chloroplasts with possible family-wide utility. *Am. J. Bot.* **2011**, *98* (7), e185–e188.

(25) Drummond, A. J.; Ashton, B.; Buxton, S.; Cheung, M.; Cooper, A.; Heled, J.; Kearse, M.; Moir, R.; Stones-Havas, S.; Sturrock, S.; Thierer, T,; Wilson, A. Geneious v5.1, available from http://www.geneious.com, 2010.

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

(27) Jukes, T. H.; Cantor, C. R. Evolution of protein molecules. In *Mammalian protein metabolism*; Munro, H. N., Ed.; Academic Press: New York, 1969; pp 21−123.

(28) Saitou, N.; Nei, M. The neighbor-joining method: a new method for reconstruction phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

(29) Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552.

(30) Parks, M.; Cronn, R.; Liston, A. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* **2009**, *7* (1), 84.

(31) Neubig, K. M.; Whitten, W. M.; Carlsward, B. S.; Blanco, M. A.; Endara, L.; Williams, N. H.; Moore, M. Phylogenetic utility of ycf1 in orchids: a plastid gene more variable than matK. *Plant Syst. Evol.* **2009**, *277* (1−2), 75–84.

(32) Hollingsworth, P. M.; Graham, S. W.; Little, D. P. Choosing and Using a Plant DNA Barcode. *PLoS ONE.* **2011**, *6* (5), e19254.

(33) Li, D.-Z.; Liu, J.-Q.; Chen, Z.-D.; Wang, H.; Ge, X.-J.; Zhou, S.-L.; Gao, L.-M.; Fu, C.-X.; Chen, S.-L. Plant DNA barcoding in China. *J. Syst. Evol.* **2011**, *49* (3), 169–175.