



Monitoring internet trade to inform species conservation actions

Valentina Vaglica¹, Maurizio Sajeve¹, H. Noel McGough¹, Dylan Hutchison²,
Claudio Russo³, Andrew D. Gordon^{3,4}, Aro Vonjy Ramarosandratana⁵,
Wolfgang Stuppy⁶, Matthew J. Smith^{7,8,*}

¹Dipartimento di Scienze e Tecnologie Biologiche, Chimiche e Farmaceutiche, Via Archirafi 18, Palermo 90123, Italy

²University of Washington, Seattle, WA 98195, USA

³Microsoft Research, Cambridge CB1 2FB, UK

⁴School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK

⁵Department of Plant Biology and Ecology, University of Antananarivo, PO Box 906, Antananarivo 101, Madagascar

⁶Royal Botanic Gardens, Kew, Wakehurst, Ardingly RH17 6TN, UK

⁷Microsoft Services, Reading RG6 1WG, UK

⁸Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AE, UK

ABSTRACT: Specimens, parts and products of threatened species are commonly traded on the internet. This could threaten the survival of some wild populations. We outline 2 methods to monitor internet sales of species to assess potential threats and inform conservation actions. Our first method combines systematic monitoring of online offers of plants for sale with expert consultation. Our second method utilises a computational model, trained to expert-classified records using probabilistic inference, to predict unknown properties of the traded taxa. We used these methods to monitor internet trade in 5 genera of succulent plant species endemic to Madagascar, some of which have recently been listed for trade regulation under the Convention on International Trade in Endangered Species (CITES). This revealed potential threats to wild populations: for instance, almost all species recorded were of high conservation concern, yet most offers for live plants were of apparently wild-collected specimens (85%). Our model predicted with 89% accuracy whether the plants were classified as propagated or wild collected by an expert, although accuracy dropped for data collected in the following summer. Our results highlight potential threats by internet trade to the survival of some CITES and non-CITES listed plant species from Madagascar. These should be addressed by further conservation actions and policy. More generally, our results reveal how standardised internet surveys can provide information on levels of trade in wild-collected threatened species that could impact on natural populations, and can provide data that can be incorporated into models to facilitate future monitoring and enforcement.

KEY WORDS: Adenia · Commiphora · Operculicarya · Uncarina · Machine learning · Infer.NET · Naive Bayes classifier

INTRODUCTION

Plants, and their parts and products, are traded extensively worldwide (RBG Kew 2016). Supplying such trade can lead to species and their populations becoming threatened with extinction. Consequently, national and international legal protection measures

have been established to prevent trade from leading to overexploitation and extinction (Flores-Palacios & Valencia-Diaz 2007, Phelps et al. 2014). Enforcing legal controls and monitoring trade is a major challenge all along the supply chain; it is typically expensive to monitor wild populations, expertise is required to identify species and their products, trade demand is

*Corresponding author: matthew.smith@microsoft.com

always changing and there is a wide range of different trade routes. New methods that facilitate such monitoring and legal enforcement of plant trade could therefore provide major benefits to the prevention of overexploitation of wild populations.

The internet has become a major mechanism through which to conduct trade in wildlife, including plants (Lavorgna 2014). The use of the internet in the illegal wildlife trade has been an area of concern, with studies implying that the internet is a common venue for conducting illegal transactions (Williamson 2004, IFAW 2005, 2008, Wu 2007, Sonricker Hansen et al. 2012, Chng 2014). In contrast, there have been few published studies of internet-traded plants. One study highlighted that at least 90% of international trade in Mexican cacti monitored over an internet auction site appeared to contravene international regulations (Sajeve et al. 2013a). A similar lack of regulation was noted by Newton & Chan (1998) for international trade in succulent plants through the postal system. An ongoing challenge remains in monitoring internet trade in species, legal and illegal, to detect whether levels might threaten the survival of wild populations.

The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) aims to regulate international trade so it does not threaten the survival of the species in the wild, including trade over the internet. CITES has lists of taxa (Appendices) for which member states (Parties) agree to regulate or monitor international trade. For example, the treaty requires export permits for international trade in wild specimens of species listed in CITES Appendix II, confirming that such trade is legal and not detrimental to the survival of the wild populations. Through the issuance of permits, CITES enables the monitoring of legal international trade in listed species. Evidence that internet trade frequently bypasses CITES regulations (e.g. Sajeve et al. 2013a) is therefore a problem for monitoring potentially legal trade, as well as preventing illegal trade.

For almost a decade, the issue of whether international trade over the internet is affecting the threats to species has been discussed at CITES meetings. It was reported at the 14th meeting (in 2007) of the Conference of the Parties (CoP) that the use of the internet for international trade was one of the factors behind a perceived increase in illegal wildlife trade (CITES 2007). A consequent workshop made recommendations on effective regulation and monitoring of international trade in CITES-listed species over the internet (CITES 2009) that were incorporated into CITES Resolution 11.3 (currently CITES 2016a) on

Compliance and Enforcement. In particular, 'regarding e-commerce of specimens of CITES-listed species' Resolution 11.3 recommends Parties evaluate and develop mechanisms to monitor internet trade.

Despite calls for methods to assist in monitoring internet trade, there have been few published examples. The International Fund for Animal Welfare (IFAW) implemented a survey to analyse internet trade in animal species (IFAW 2005, 2008). Their survey monitored the types of live animals and their products that were being traded over the internet over a 3 mo period as well as details of the traders, legal conditions and any mechanisms in place to report illegal activity. Sonricker Hansen et al. (2012) report on the application of an automated internet surveillance method to monitor illegal wildlife trade. Hinsley et al. (2016) conducted the first systematic survey of wildlife trade over social media websites, focusing on the trade in orchid species. Their analysis revealed a relatively high proportion of posts about trade in orchids relating to wild-collected plants, highlighting the importance of monitoring such trade in order to assess threats to wild populations.

Expert consultation plays a major role in CITES implementation and in conservation more generally (Burgman et al. 2011, CITES 2016a). In CITES enforcement, experts are consulted when taxon-specific understanding is required, such as identifying unknown species at a port of entry, identifying whether it is likely to have been collected from the wild, or advising on whether trade is likely to be detrimental to the survival of wild populations (McGough et al. 2004, Sajeve et al. 2013b). The CITES Animals and Plants Committees have compiled lists of regional experts on particular taxa who can advise when necessary (CITES 2016a). Expert consultation is time consuming and can be subject to multiple errors and biases (McBride et al. 2012). Consequently, there is interest in techniques that can replace or improve the efficacy of expert evaluation in conservation (Burgman et al. 2011, McBride et al. 2012). For example, experts could monitor trade in endangered species over the internet, such as over online auction sites, to evaluate threats to wild populations and problems in enforcement or regulation (Sajeve et al. 2013a); however, there is a considerable challenge in effectively monitoring the vast quantities of trade in species that occur over the internet.

Our goal with this study is to report the results of 2 methods aimed at improving the monitoring and enforcement of international trade in species over the internet. Our first method is a general systematic manual internet survey to enable experts to monitor

offers of species of threatened plants for sale. This method has been used for over 15 yr by some of the authors to monitor internet trade in species but has never been formally documented and there are no examples of such manual internet survey methods in the scientific literature. The method is laborious and could undoubtedly be improved, and so we describe the technique and results it generates here to support such improvements in future.

Our second method is an automated method, a classifier model, which can be trained to historical survey data to then be used to predict properties of traded specimens that would have previously been assessed by an expert. If a reliable model could be obtained, then it could speed up the process of evaluating survey data by automating some of the expert evaluations, reducing some of the dependency on experts. We therefore investigate the scenario where all details of a species offered for sale are known except plant origin, which is predicted by the model (labelled as the monitoring problem in Table 1). In the case of enforcement, we envisage 2 scenarios. First, a model could be used to predict whether an application for an import or export permit has been made for a CITES-listed species that has been falsely claimed to be artificially propagated while it was actually wild collected, or falsely claimed to be a different species in order to avoid inspection (see Table 1, enforcement 1a and 1b). Second is the case in which a species has arrived at a port of entry without any documentation but appears to be wild-collected. In that case, authorities could use the model to gain insights into what the species could be (based on historical information) and whether it should have had CITES permits (Table 1, enforcement 2).

Another aim of our study is to characterise potential international trade in a range of selected genera of succulent plant species endemic to Madagascar, to inform the conservation community and help inform further conservation actions and policy. Unsustainable international trade could threaten the survival of many species in Madagascar (Oldfield 1997, Ravao-manalina et al. 2011). As a consequence, Madagascar has proposed a number of species for inclusion in the CITES Appendices. We chose to apply our survey methods to the genera *Adenia*, *Commiphora*, *Cyphostemma*, *Operculicarya* and *Uncarina* because of the recent interest in listing species of those genera on CITES (Table S1 gives further details of why these genera were selected and Table S2 lists the CITES-

Table 1. Overview of the data collected during the survey and summary of the data assumed to be known (K; blue) and unknown (U; yellow) in the monitoring and enforcement problems specified in the 'Materials and methods'

Property	Monitoring	Enforcement		
		1a	1b	2
Data collected during survey				
Species	K	K	U	U
Price	K	U	U	K
Country	K	K	K	K
Vendor	K	U	U	K
Vendor type (nursery, auction)	K	U	U	K
Form (e.g. seed, plant)	K	K	K	K
Season	K	U	U	K
Photograph	K	U	U	K
Known properties of species				
On CITES	K	K	U	U
Genus	K	K	K	K
Year first described	K	K	U	U
Expert estimated properties of species				
Propagation status	K	K	U	U
Cultivation status	K	K	U	U
Conservation priority	K	K	U	U
Conservation concern	K	K	U	U
Wild or artificially propagated	U	U	U	K

listed species in our survey; see the Supplement at www.int-res.com/articles/suppl/n032p223_supp.pdf).

MATERIALS AND METHODS

Survey method

Our survey method consisted of recording offers of plants for sale from nursery websites and internet auction websites. These were initially found using the internet search engine Google (www.google.com) and the internet auction site eBay (www.ebay.com) as these are commonly used portals to find plants for sale. Terms used in our searches included the scientific names, the generic names alone, and spelling variants. Links to nurseries and other internet auction sites were followed from the first 100 links returned by Google searches.

Offers over eBay were recorded and links (or other references) from eBay sales sites to nurseries were investigated and recorded if target plants were offered for sale from those sites. To limit our search effort, no secondary links from nursery sites were followed. eBay searches were set so that offers for sale from any source country could be found. Only sites that were written in English, or had an accompanying English translation, were inspected.

For each site visited, we recorded the following details of plants offered for sale (Table 1): offer price (converted to US dollars; we used fixed currency exchange rates throughout the study, obtained on 27 May 2014), seller's country and web site address, whether the seller is a nursery or an individual, the plant form offered (e.g. as seeds or as mature plants), and an image of the plant offered for sale, if available. Sellers rarely noted plant origin (i.e. whether it was propagated or wild), although we did note that detail when it existed. All data were recorded in Excel spreadsheets with the exception of the associated images, which were saved separately and indexed to enable association with the data recorded in the Excel spreadsheets.

Four seasonal surveys were conducted in April, July and October 2013 and February 2014, each using the same survey protocol and taking approximately 114 h to complete, spread over 1.5 mo. We term these surveys the spring, summer, autumn and winter 2013 surveys. A further survey to collect independent test data was conducted in August 2014 using the same protocol, which we term the summer 2014 survey.

Our survey was not subject to any formal ethical approval procedure. We do not provide any specific details of the sellers here, partly to protect their confidentiality, but also because our expert evaluations detailed below are subject to a degree of error that needs to be evaluated in future.

We surveyed 5 of the approximately 50 genera with succulent species in Madagascar (out of 1698 genera; Rauh 1998)—this was the maximum number possible given the timescale of the project. We included all 121 species of these 5 genera in our survey, although only 34 species were found for sale over the internet, all endemic to Madagascar. We also looked for evidence of international trade involving the genera in our 2013 survey in the official CITES trade database (<http://trade.cites.org>). This only reports legal international trade in CITES-listed species over the period over which they were listed, which was the year 2010 at the earliest (Table S2 in the Supplement details when the CITES-listed species in our survey were added to the CITES Appendices). CITES trade data are typically up to date from 2 yr after CITES permits were issued (see https://trade.cites.org/cites_trade_guidelines/en-CITES_Trade_Database_Guide.pdf). We downloaded data covering the years 2010–2013 on 1 October 2016, assuming the data for that period were up to date. We only recorded trade identified as being for commercial

purposes (2 records were of exchanges between botanic gardens).

Expert panel

Three experts provided additional information about the species in trade. They are all trained botanists with years of experience in identifying succulent plant species from Madagascar. Their experience includes having to identify plant species for government and CITES enforcement authorities, such as having to assess the origin of plants that appear in trade. In order to protect their anonymity we do not identify the experts here.

Prior to the surveys, the experts were independently asked 4 questions relating to each species that might potentially be found in the survey, and asked to categorize the species according to: the likelihood that a species is wild collected (cultivation status), how easy is it to propagate the species (propagation status), how threatened the wild populations are with extinction (conservation concern) and how urgent it is to undertake action to conserve the species (conservation priority; for the specific questions that were asked, see Table S3 in the Supplement). For each of these categories they could assign a low/easy, medium or high/hard rating. The experts were relatively consistent in their classifications, either being in complete agreement or choosing neighbouring categories (see 'Results'). Therefore, we assigned low/easy, medium and high/hard as the ordinal classifications 1, 3 and 5, respectively, and used 2 and 4 to denote when the experts were split between low and medium and medium and hard, respectively (Table 2).

After each seasonal survey we asked an expert whether they thought the plants in trade were likely to have been collected from the wild or artificially propagated (or whether they were unsure) using the information collected in the survey. To do this we compiled the images and data collected during each seasonal survey into a table that enabled the expert to easily view the images and associated data together and make an assessment. This document was emailed to the expert. Due to time and resource constraints we were limited to using a single expert for this evaluation and had no opportunity to formally evaluate the accuracy of our expert in their classifications. We revisit these issues in the 'Discussion'. This expert has decades of experience in aiding CITES enforcement in relation to succulent plant species from Madagascar.

Table 2. Frequency of expert classifications regarding the conservation, propagation and cultivation status of the species identified in the survey

Expert rating	Frequency of rating			
	Conservation concern	Conservation priority	Propagation	Cultivation
1 – low	0	1	0	0
2 – low or medium	0	0	5	1
3 – medium	2	3	10	1
4 – medium or high	5	10	13	6
5 – high	24	18	3	23

The sellers rarely indicated whether specimens were collected from the wild or were propagated, although this information was shared with our expert when they did.

Our expert could not classify seeds in trade as wild collected or propagated based only on information that could be obtained from the sellers. It would probably be possible to advise on the likely origin of the seeds for sale based on the reproductive biology of each genus, but such classification would be highly uncertain and, as a consequence, we chose not to attempt this task and omitted records of seeds from our analyses. Our survey results on trade in seeds are discussed in Text S1 in the Supplement because it is difficult to draw conclusive results.

Analysis and modelling

We conducted conventional (frequentist) statistical analyses of our survey results—ANOVA, chi-squared tests and Student’s *t*-tests—to identify relationships between species and trade properties and whether the species are believed to be sourced from the wild (conducted using Microsoft Excel and MATLAB’s Statistics Toolbox).

We assessed the accuracy with which a naïve Bayes classifier model could predict unknown properties of the taxa being offered for sale. This widely used classification method produces a probabilistic prediction of outcomes based on an assumption that plant sale ‘features’ (e.g. species, price, country of sale) are independent when given, regardless of whether the plant for sale is wild collected or artificially propagated, our main variable of interest. Naïve Bayes works well as a classifier even when this assumption is moderately violated (Domingos & Paz-zani 1997). We train the classifier by supplying training data and prior distributions, which we choose to be uninformative in order to remove bias by choice of prior information.

We expressed our model in Tabular (Gordon et al. 2015), a probabilistic programming language embedded in Microsoft Excel. Tabular models compile into Infer.NET (Minka et al. 2012) inference code, in which we use the Variational Message Passing inference algorithm (Winn & Bishop 2005) for training and testing. We chose Tabular for its more intuitive way of interfacing with models via Excel spreadsheets and for its more concise model descriptions (compared to C programs, for instance), both of which make the use of this method by monitoring and enforcement authorities a more realistic prospect. We also used ModelWizard (Hutchison 2015), a model-building language that generates Tabular programs, to explore many variations of classifiers before settling on the intended naïve Bayes model.

We trained the classifier model to the 2013 survey data to assess the performance of the model at correctly predicting whether a plant offered for sale was classified by an expert as wild collected or artificially propagated. So while the expert used all the survey information, including an image, to assess whether the specimen was wild collected, the trained classifier uses the same survey information, minus the image, to predict whether the specimen was wild collected. We trained and assessed the classifier using 5-fold cross-validation, a commonly used method to enable the rigorous evaluation of model performance on data that has not been used to train it. The data are randomly divided into 5 groups (folds) and the model is trained and evaluated 5 times, each time holding back one of the folds for model evaluation and using the remaining for model training. We then trained the model to the complete 2013 survey data and used the trained model to predict plant origin (wild or propagated) for the summer 2014 data. Both of these analyses are informative in terms of addressing the monitoring problem (Table 1, monitoring problem).

To address the first enforcement problem, we removed 10 records randomly from those that are CITES-listed species in the 2013 survey data and retrained the model to the remainder of the 2013 survey data to assess how well the model predicts plant origin and whether species are traded under the correct name (Table 1, enforcement problem 1). To address the second enforcement problem, we used records from the summer 2014 survey data that had

been classified by the expert as wild collected in order to assess the classifier's prediction of the species of the plant in trade (assumed to have arrived in a country without a CITES permit) and, thus, whether it should have a CITES permit (Table 1, enforcement problem 2).

RESULTS

Summary of survey data

Our 2013 surveys made 748 records of plants offered for sale (hereafter just 'records'), with similar numbers each survey period (spring: 208; summer: 203; autumn: 190; winter: 147). Thirty-four species were recorded, including 10 CITES-listed species. However, CITES-listed species made up 55% of the records. Our 5 chosen genera were all represented, each having >100 records and 5 species, with the exception of *Commiphora*, for which we only made 27 records (Table S1). Fifteen different countries were recorded but not Madagascar. Fifty-two different nurseries and 2 different internet auction sites were recorded. Most records were offers from nurseries in the USA or Europe (80%, 601), with the remaining 20% being from internet auction sites (see Table S4 in the Supplement). About half our records were of live plants (51%) with the remainder being seeds. Live plants were mostly represented by immature plants (60%), with mature plants (16%), cuttings (4%) and bonsai (8%) making up the minority (here, bonsai plants are those grown in small ornamental pots so as to resemble miniature versions of the larger mature plants).

Expert assessment of survey data

Nearly all the species recorded were assessed by our experts to be of high conservation concern, with only 2 assessed as of medium conservation concern (Table 2). Similarly, the majority of species were judged to be of high conservation priority; only one species was judged to be of low conservation priority and 3 of medium conservation priority (Table 2). Expert assessment of ease of propagation and likelihood in cultivation was more variable amongst the species (Table 2).

For records of live plants (not seeds), our expert was confident in identifying the origin (wild or artificially propagated) for 76% of our records about the origin of the plants, with 85% of those believed to be

of wild origin (from Madagascar) and the remainder to be artificially propagated.

Comparison with sales reported in the CITES trade database

The CITES trade database detailed international trade for commercial purposes in 10 different species in our survey between years 2010 and 2013. Thirty-six of the 39 records relate to commercial exports of live plants from Madagascar (Table 3). If these records related to plants re-sold by any sellers in our survey, then we would expect the importer and species in the CITES trade database to match the country-species combination in our survey, with the exception of trade within the European Union, in which the country of the seller could be different. The majority (60%) did not meet this criterion. Six database records of trade to the United States had the same species as live plants offered for sale from the recorded country of origin in our survey. One of these also indicated the Czech Republic as the exporting country, for which we also have a country and species match in our internet survey. Nine trade database records were to European Union countries that did not match the country-species combination of the seller but could have ended up at a seller in our survey without requiring permits.

Factors related to plant origin

Here we will focus on the records of live plants offered for sale which our expert confidently identified as being either wild collected or artificially propagated (40% of all our records).

Species-specific factors

Genus and CITES status were both related to whether the species was likely to be classified as collected from the wild ($p < 0.001$, ANOVA). CITES species were more than twice as likely to be artificially propagated (22% versus 8% of records, respectively, $p < 0.001$, χ^2 test) and different genera differed significantly in the proportion of records that were of wild-collected plants ($p < 0.001$, χ^2 test): *Adenia* was more likely to be propagated, *Uncarina* and *Commiphora* more likely to be wild collected and *Operculicarya* and *Cyphostemma* were exclusively wild

Table 3. CITES trade database records of the species featured in our survey. For number of plants, a value in parentheses is the number reported by the exporter when different from the number reported by the importer

Species	Year	Exporter	Importer	No. of plants	Artificially propagated (P) or wild collected (W)	2013 survey records from importer or exporter country?
<i>Adenia</i> spp.	2011	Philippines	Singapore	5	P	No
<i>Adenia firingalavensis</i>	2013	Madagascar	Japan	10	P	No
<i>Adenia olaboensis</i>	2013	Madagascar	Japan	10	P	No
<i>Adenia olaboensis</i>	2011	Philippines	Singapore	1	P	No
<i>Adenia olaboensis</i>	2012	Czech Republic	United States	15	P	Yes, importer and exporter
<i>Cyphostemma elephantopus</i>	2012	Madagascar	Spain	6	W	From another European country
<i>Cyphostemma elephantopus</i>	2012	Madagascar	Japan	10	W	No
<i>Cyphostemma elephantopus</i>	2012	Madagascar	United States	10	W	Yes
<i>Cyphostemma elephantopus</i>	2013	Madagascar	Germany	60	W	From another European country
<i>Cyphostemma elephantopus</i>	2013	Madagascar	Japan	30	P	No
<i>Cyphostemma elephantopus</i>	2013	Madagascar	Mexico	1	W	No
<i>Cyphostemma elephantopus</i>	2013	Madagascar	Thailand	10	P	No
<i>Cyphostemma laza</i>	2013	Madagascar	Hong Kong	10	P	No
<i>Cyphostemma laza</i>	2013	Madagascar	Japan	5	P	No
<i>Cyphostemma montagnacii</i>	2012	Madagascar	Spain	7	W	No
<i>Cyphostemma montagnacii</i>	2013	Madagascar	Germany	60	W	No
<i>Operculicarya decaryi</i>	2013	Madagascar	Hong Kong	4	P	No
<i>Operculicarya hyphaenoides</i>	2011	Madagascar	Japan	50	W	No
<i>Operculicarya hyphaenoides</i>	2011	Madagascar	United States	126 (225)	W	Yes
<i>Operculicarya hyphaenoides</i>	2012	Madagascar	Spain	6	W	No
<i>Operculicarya hyphaenoides</i>	2012	Madagascar	Japan	10	W	No
<i>Operculicarya hyphaenoides</i>	2012	Madagascar	United States	100	W	Yes
<i>Operculicarya hyphaenoides</i>	2013	Madagascar	Japan	2	W	No
<i>Operculicarya pachypus</i>	2010	Madagascar	Germany	50	W	From another European country
<i>Operculicarya pachypus</i>	2010	Madagascar	Japan	50	W	No
<i>Operculicarya pachypus</i>	2011	Madagascar	Germany	50	W	From another European country
<i>Operculicarya pachypus</i>	2011	Madagascar	Japan	30 (130)	W	No
<i>Operculicarya pachypus</i>	2011	Madagascar	United States	25 (126)	W	Yes
<i>Operculicarya pachypus</i>	2012	Madagascar	Germany	60	W	From another European country
<i>Operculicarya pachypus</i>	2012	Madagascar	Spain	10	W	From another European country
<i>Operculicarya pachypus</i>	2012	Madagascar	Japan	235 (245)	W	No
<i>Operculicarya pachypus</i>	2012	Madagascar	United States	100	W	Yes
<i>Operculicarya pachypus</i>	2013	Madagascar	Czech Republic	30	P	From another European country
<i>Operculicarya pachypus</i>	2013	Madagascar	Germany	60	W	From another European country
<i>Operculicarya pachypus</i>	2013	Madagascar	Hong Kong	88	P	No
<i>Operculicarya pachypus</i>	2013	Madagascar	Japan	130	P	No
<i>Operculicarya pachypus</i>	2013	Madagascar	Japan	22	W	No
<i>Uncarina grandidieri</i>	2013	Madagascar	Japan	3	P	No
<i>Uncarina stellulifera</i>	2013	Madagascar	Germany	50	P	From another European country

collected. The date of botanical description of the species was also significantly related to whether a plant was sold as wild collected or artificially propagated (t -test, $p < 0.001$): the average year of description of artificially propagated plants was roughly 50 yr earlier (average 1914, 95% confidence intervals 1895 and 1962) than wild-collected plants (average 1961, 95% confidence intervals 1944 and 1996). This implies that more newly described species are more likely to appear in trade as wild-collected plants. Cultivation status, conservation status and conservation priority of the species in trade were found to be statistically unrelated to the plant source.

Trade-specific factors

As noted, plants offered for sale were more likely than not to be wild collected; however, propagated plants were 8 times more likely to be sold by nurseries than from internet auction sites (17% versus 2%, respectively, $p < 0.01$). Within sales offered by nurseries, CITES-listed plants were more likely artificially propagated than non CITES-listed plants (26 versus 8%, respectively, $p < 0.001$). The average price of plants for sale was also significantly related to their origin (t -test, $p < 0.001$), with artificially propagated plants on average selling for \$20 and wild

plants on average selling for \$37 (all prices quoted are in US dollars). This implies that more expensive plants are more likely to be wild collected. More striking were the differences in the range of prices, with the lower 2.5% confidence intervals for both sources being \$6 but the upper 97.5% being \$65 for propagated plants and \$500 for wild plants, thus the extremely expensive plants appear more likely to be wild collected. The growth stages of the plants for sale were also significantly related to plant origin, with more immature forms—seedlings, immature plants and cuttings—being more likely to be propagated (around 20% of those sales) than mature and bonsai plants (only 1% of these was thought to be propagated). This implies that larger, more mature plants are more likely to be wild collected. Neither source country nor survey season had any significant relationship with whether the live plants were artificially propagated or wild ($p > 0.1$).

Classifier model

Prediction of plant origin (monitoring problem)

A classifier model of the plant source for the 2013 survey was on average 89% correct in predicting whether the experts identified live plants for sale as wild collected or propagated (assessed by 5-fold cross-validation). This breaks down into correct predictions for ca. 8 out of 9 records (88.6%) as propagated and ca. 44 out of 49 records (89.5%) as wild.

Our expert confidently identified 20 records of live plants offered for sale in the summer 2014 survey as being either wild collected or propagated, and so these were used as a final test dataset for our model. For these data, the classifier model was less accurate than for the 2013 data—correctly predicting 6 out of the 8 plants assessed as wild collected (75%) but correctly classifying only 3 out of the 12 plants assessed as propagated (25%). The 2 false negatives (wild classified as propagated) were for species for which the expert could not classify the origin in 2013 and appear relatively inexpensive compared with wild-collected plants of the same genus offered in 2013 (\$32 and \$60 compared with an average of \$136). Three of the false positives (propagated classified as wild) were propagated plants offered for sale via internet auction sites. This appears to be because all but one record of plants for sale in the 2013 data (1 out of 44) was of wild-collected plants and so the classifier associated internet auction sites with wild-collected plants. The remaining 6 misclassifications

of propagated plants as wild plants were for species exclusively identified as wild collected in 2013.

Prediction of plant origin given uncertain identity (enforcement problem 1)

Plants assessed as wild collected in the 2013 survey were all predicted with near certainty as so by the classifier, irrespective of whether it is assumed that applicants list correct species names (Table 4; we assume that applicants always specify the correct genus). The classifier incorrectly predicted that *Adenia fringivalensis* was wild collected irrespective of whether species or genus was given, presumably because all other records from Spain in the 2013 data were of wild-collected plants. This could highlight an application that the authorities would want to investigate further because all evidence from the past has indicated that such specimens are wild collected. In contrast, *Adenia olaboensis* was predicted to be propagated irrespective of whether the species or genus was specified because all other CITES-listed *Adenia* species from the USA were propagated. The results for *Uncarina grandidieri* appear strange, with the classifier being almost certain that the specimens are wild collected when the taxon is not known, to being almost certain they are artificially propagated when the species is known. This is largely because most records of mature or seedling *Uncarina* specimens for sale from the USA and Germany were of wild-collected plants (40 records), with the only exceptions being the *U. grandidieri* specimens (4 records). A significant effect will also have come from removing these 2 records of *U. grandidieri* from the initial training data.

Prediction of identity given uncertain background information (enforcement problem 2)

The data used for the second enforcement problem were constructed by taking records identified as relating to wild-collected specimens from our 2014 survey and then removing information on species identity (see Table 1 for the data assumed to be known). Two records related to CITES-listed species (Table 5). The classifier predicted these were indeed CITES listed with $p = 0.25$. An enforcement officer could use such a probability as justification to take further steps to confirm species identity (e.g. by consulting a botanist). For records that were not of CITES-listed species, the classifier

Table 4. Inferred probabilities (Prob) that the CITES-listed species (identified by species name, exporting country and plant growth form) are of wild-collected (W) or propagated (P) stock based on observations from 2013 internet survey and expert opinion

Species	Specified by applicant for CITES permit		Inferred Prob		Expert assessment of plant source
	Country	Traded form	W; species unknown	W; species known	
<i>Adenia firingavalensis</i>	Spain	Seedling	0.9999	0.9999	P
<i>Adenia olaboensis</i>	USA	Immature	0.0004	0.0001	P
<i>Operculicarya decaryi</i>	Spain	Bonsai	0.9999	0.9999	W
<i>Operculicarya decaryi</i>	Netherlands	Immature	0.9999	0.9999	W
<i>Operculicarya decaryi</i>	USA	Mature	0.9999	0.9999	W
<i>Operculicarya pachypus</i>	France	Immature	0.9999	0.9999	W
<i>Operculicarya hyphaenoides</i>	USA	Mature	0.9999	0.9999	W
<i>Uncarina grandidieri</i>	Germany	Mature	0.9999	0.0002	P
<i>Uncarina grandidieri</i>	USA	Seedling	0.9999	0.0005	P
<i>Uncarina stellulifera</i>	France	Mature	0.9999	0.9999	W

tended to assign lower probabilities to them being CITES listed (Table 5).

DISCUSSION

Internet survey methodology

Our method enables repeated assessments of offers for sale over the internet for chosen species, which, in our case study, implies potential trade in wild-collected specimens that could detrimentally affect threatened wild populations. A similar insight was made by Newton & Chan (1998) for international trade in succulent plants from South Africa through the postal system; they used a manual survey of nursery catalogues and analysis of the CITES trade database. Our survey method is laborious, taking 114 h to complete each survey period (spread over 1.5 months each), but enables us to account for the heterogeneity of ways in which sellers advertise, ranging from, for example, in-

dividual plants being sold on auction sites to lists of different sizes of the same species in electronic catalogues. Both time and effort could be saved by automating some of the process (e.g. Sonricker Hansen et al. 2012). The process of determining relevant websites is relatively easy to automate. The difficulty is accounting for the heterogeneity of ways in which species are sold and their details are specified on websites (e.g. embedded as images in pdf files, in structured tables, in free text). Machine learning techniques could greatly facilitate the process, and testing their application would be a natural avenue for future research. Given that downloading relevant content is relatively easy, the challenge will be obtaining enough training data and expressing it in an appropriate way to enable machine learning to extract the relevant content.

The laborious nature of our approach currently means that it might only be possible to sample a small subset of the overall internet trade for other groups of species and, under such circumstances, the sampling strategy becomes particularly important to minimise

Table 5. Inferred identities of species hypothetically arriving in a port of entry without documentation and identified by experts as collected from the wild. True species: true species identity; Prob(true species): predicted probability of the true species; Most likely species (probability), given the genus only: predicted most likely species (and probability) given only genus is known; On CITES: true species CITES listed; Prob(on CITES): predicted probability species is CITES listed given only genus is known

True species	Prob(true species)	Most likely species (probability), given the genus only	On CITES	Prob(on CITES)
<i>Adenia epigea</i>	0.63	<i>Adenia epigea</i> (0.63)	N	0.09
<i>Commiphora humbertii</i>	0.2	<i>Commiphora simplicifolia</i> (0.33)	N	0.003
<i>Commiphora orbicularis</i>	0.28	<i>Commiphora simplicifolia</i> (0.33)	N	0.003
<i>Cyphostemma elephantopus</i>	0.16	<i>Cyphostemma pachypus</i> (0.75)	Y	0.25
<i>Cyphostemma laza</i>	0.09	<i>Cyphostemma pachypus</i> (0.75)	Y	0.25
<i>Uncarina roeoesliana</i>	0.57	<i>Uncarina roeoesliana</i> (0.57)	N	0.06

the chances of obtaining spurious or biased results (Barber-Meyer 2010). The repeatability of our method means that temporal changes can also be detected. Temporal changes were not seen during 2013 but do appear by summer 2014, implying that a longer time course of monitoring is required to reveal temporal trends.

Our survey method misses actual recorded trade. This was addressed by Sajeve et al. (2013a) by monitoring buyer–seller dialogues to confirm that a purchased plant was received. This was not an option in our study and so we assume that the extent of offers of plants for sale over the internet reflects levels of buyer interest. Cross-checking our survey results with the CITES trade database (also employed by Sajeve et al. 2013a) indicated that if any of the sellers were trading CITES-listed species internationally between 2010 and 2013, then it was not done with official CITES permits. This could imply potentially unmonitored and unregulated international trade, or could simply be because such nursery and auction sites only trade nationally (or within the European Union). We did not attempt to quantify the number of nurseries or auction sites that could be trading internationally for the 2013 and 2014 surveys, but retrospective investigation provides evidence that some nurseries and internet auctions are open to trading internationally, e.g. by indicating how they ship internationally. Thus it seems very unlikely that all of our records relate to only national sales.

Our results also rely on expert input, which limits the scalability of our survey method and could introduce error and bias into our results. We never formally assessed the accuracy of the expert classification of plant origin as being wild collected or artificially propagated. McBride et al. (2012) assessed the evaluations of the extinction risk of Australian bird taxa from 16 different experts and revealed significant biases and errors. It therefore seems highly plausible that there are also errors and potential biases in our own expert evaluation data. At present and in our experience it is rare for CITES enforcement authorities to undertake expert consultation with any more than a few experts. Methods are needed in future to account for inevitable biases in expert evaluations. For example, a natural extension of our work here would be to obtain evaluations from multiple experts as to the origin of the plants in trade (McBride et al. 2012). This would not only immediately reveal the degree of consistency, but expert error could also be explicitly incorporated into our classifier method by inferring the extent to which the expert classification is false.

Use of classifier models

Our hypothetical applications of the classifier illustrate how such models could aid in monitoring and regulation. For example, 89% accuracy at predicting plant origin for the 2013 survey is probably high enough to be worth consulting when deciding whether to request to inspect a shipment to confirm plant origin (e.g. Table 4), although not high enough to replace expert assessment in the internet surveys. Additional insights can be obtained by investigating the reasons for a prediction being made. For example, although the relatively frequent failures of the model at predicting plant origin in 2014 would reduce confidence in any of the individual predictions of the classifier, they highlight clear contrasts between 2013 and 2014 data in the patterns of wild-collected and artificially propagated plants being offered for sale that would be useful for monitoring and enforcement authorities to know.

We expect that the collection of more data, over more websites and over time, would lead to a classifier model that has higher predictive accuracy because several of the species in our survey have low quantities of data associated with them (1 or 2 records), and so collecting more data would enable their relationships with other factors to be characterised with greater confidence. This is highlighted by the sensitivity of the predictions to knowing the species identify in the first enforcement problem (Table 4). This sensitivity would be removed by including more data in the training set. The contrasts between the 2013 survey data and the summer 2014 data also highlight the importance of monitoring internet sales over time to understand how relationships are changing. We also used only one method to build a classifier model even though there are multiple methods available, and future studies are needed to explore the impact of the modelling methodology and the potential benefits of using an ensemble of classifiers (Rokach 2010).

Another potential application of the classifier would be to indicate to buyers or sellers on internet trade sites that regulations may apply to particular trade items. Such details are not always apparent on internet auction sites (Magalhães & São-Pedro 2012), probably because they are irrelevant for most sales. It could be that an automatic warning system might not need a trained classifier; for example, the website could look for commonly used terms in internet trade of CITES species and only issue warnings when enough terms are used. A trained classifier would have the advantage of

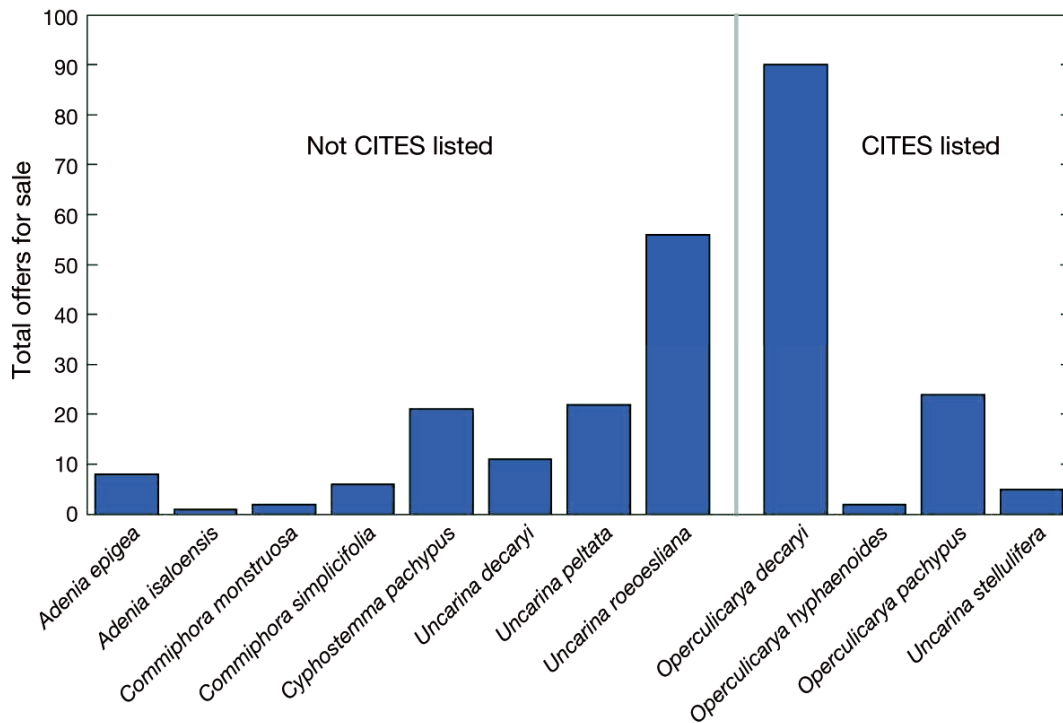


Fig. 1. Number of records made of live plants of CITES-listed and non CITES-listed species identified as wild-collected plants from the 2013 survey

being easier to retrain to account for changes in terms commonly associated with relevant trade. Ultimately, the method would predict some probability that an item or items for sale would be covered by CITES regulations and a threshold probability or other quantity would need to be identified, above which a notification should be issued. Such notifications could be issued to sellers when they post items or to buyers when they search for or proceed to purchase items.

Conservation implications from the case study

Offers of sale for wild-collected plants dominate our records for the 2013 survey, supporting the implications of the official CITES trade data that there is an ongoing trade interest. However, the lack of any official records of exports in wild-collected specimens from nurseries and auction sites outside of Madagascar implies that international trade could be unregulated and unmonitored. Similar findings were made by Sajeve et al. (2013a) and earlier by Newton & Chan (1998) relating to succulent plant species being traded through the postal system, implying that a lack of regulation in such trade is common.

The fact that most species in our survey were assessed as being of both high conservation concern and priority underlines the importance of monitoring international trade in these species. Our results also imply that international trade in wild-collected specimens of many of these species is ongoing. Eight taxa in our survey combine evidence of wild collection (Fig. 1), being of medium or high conservation priority or concern, and not being protected by CITES regulations. This implies that further action could be needed to prevent these species from being detrimentally impacted. Such action should include more extensive monitoring of the trade and of the status of wild populations. Relevant CITES Parties could list these species in CITES Appendix III to aid in the monitoring of the trade, although it has been shown that internet trade frequently bypasses CITES regulations (Sajeve et al. 2013a). Ultimately, CITES Parties will need to evaluate available information, including that presented here, and decide whether a course of action is needed. Our method to monitor trade in selected species provides a general means of gathering information on species that the CITES community (e.g. the CITES Plants Committee) could use to evaluate whether international trade is present, and the potential extent of this trade. International trade or the potential for such, is a basic requirement for a taxon to

be included in the CITES appendices and so such a method could be useful to provide such evidence.

Four CITES-listed species were offered for sale as wild-collected plants (Fig. 1). Experts judged all of them to be of high conservation priority and concern. The granting of export permits for these species by Madagascar at least implies that the authorities recognise some trade as being non-detrimental to the survival of the wild populations. While the lack of much officially recorded trade from Madagascar to the countries we recorded as selling the plants raises concerns of unregulated trade, it is possible that the wild-collected plants offered for sale outside Madagascar were acquired prior to those species' CITES listing. CITES authorities, especially Madagascar, should review our evidence to establish whether potentially detrimental trade is going unregulated. One option would be for the species to be selected for the Review of Significant Trade (CITES 2016b) process. The main aim of this process is to correct legal detrimental trade; however, it may formally highlight other problems encountered in its review of CITES Appendix II trade.

CONCLUSIONS

Repeated internet surveys complemented by expert assessment, as described in the present study, are clearly a useful method for generating information about the potential threats posed to species populations. In our case, they imply extensive trade in wild-collected specimens of a number of species with populations of high conservation concern, and provide information for conservation and enforcement organisations about either potential contravention of CITES regulations or the need for CITES listing for certain species.

Our study has also shown that internet survey data can be used to train predictive models that are useful to inform monitoring and enforcement authorities about what data collected in the past imply about future observations. However, our results highlight that changes in patterns of trade over time and a lack of training data can limit the reliability of the classifier predictions, issues that can be at least partially addressed by collecting more survey data over a longer time, which would be facilitated by the development of more automated survey approaches. Undoubtedly, the methods conducted here should become part of a more extensive survey programme, covering more species and over longer periods of time, for species groups of conservation concern.

LITERATURE CITED

- ✦ Barber-Meyer SM (2010) Dealing with the clandestine nature of wildlife trade market surveys. *Conserv Biol* 24: 918–923
- ✦ Burgman MA, McBride M, Ashton R, Speirs-Bridge A and others (2011) Expert status and performance. *PLOS ONE* 6:e22998
- Chng SCL (2014) Escalating black spotted turtle *Geoclemys hamiltonii* trade in Asia: a study of seizures. TRAFFIC, Petaling Jaya Selangor
- CITES (2007) Internet trade in specimens of CITES listed species. CoP14 Doc 28. <https://cites.org/sites/default/files/eng/cop/14/doc/E14-28.pdf>
- CITES (2009) E-commerce of specimens of CITES listed species. SC58 Doc 22. <https://cites.org/sites/default/files/eng/com/sc/58/E58-22.pdf>
- CITES (2016a) Compliance and enforcement. Resolution Conf. 11.3 (Rev. CoP16). <https://cites.org/sites/default/files/document/E-Res-11-03-R17A.pdf>
- CITES (2016b) Significant trade in specimens of Appendix-II species. Resolution Conf. 12.8 (Rev. CoP17). <https://cites.org/sites/default/files/document/E-Res-12-08-R17.pdf>
- ✦ Domingos P, Pazzani M (1997) On the optimality of the simple Bayesian classifier under zero-one loss. *Mach Learn* 29:103–130
- Flores-Palacios A, Valencia-Diaz S (2007) Local illegal trade reveals unknown diversity and involves a high species richness of wild vascular epiphytes. *Biol Conserv* 136: 372–387
- ✦ Gordon AD, Russo CV, Szymczak M, Borgström J, Rolland N, Graepel T, Tarlow D (2015) Probabilistic programs as spreadsheet queries. *Lect Notes Comput Sci* 9032:1–25
- ✦ Hinsley A, Lee TE, Harrison JR, Roberts DL (2016) Estimating the extent and structure of trade in horticultural orchids via social media. *Conserv Biol* 30:1038–1047
- Hutchison D (2015) ModelWizard: toward interactive model construction. MS thesis, Stevens Institute of Technology, Hoboken, NJ
- IFAW (International Fund for Animal Welfare) (2005) Caught in the web: wildlife trade on the Internet. IFAW, London
- IFAW (International Fund for Animal Welfare) (2008) Killing with keystrokes: wildlife trade on the internet. IFAW, London
- ✦ Lavorgna A (2014) Wildlife trafficking in the internet age. *Crime Science* 3:1
- ✦ Magalhães ALB de, São-Pedro VA (2012) Illegal trade on non-native amphibians and reptiles in southeast Brazil: the status of e-commerce. *Phyllomedusa: J Herpetol* 11: 155–160
- ✦ McBride MF, Garnett ST, Szabo JK, Burbidge AH and others (2012) Structured elicitation of expert judgments for threatened species assessment: a case study on a continental scale using email. *Methods Ecol Evol* 3:906–920
- McGough HN, Groves M, Mustard M, Brodie C, Sajeva MJ (2004) CITES and succulents: a user's guide. Royal Botanic Gardens, Kew
- Minka T, Winn JM, Guiver JP, Knowles DA (2012) Infer.NET 2.5. Microsoft Research Cambridge. <http://research.microsoft.com/infernet>
- Newton DJ, Chan J (1998) South Africa's trade in southern African succulent plants. TRAFFIC East/Southern African report, South Africa National Office, SA

- Oldfield S (1997) Cactus and succulent plants – status survey and conservation action plan. IUCN/SSC Cactus and Succulent Specialist Group. IUCN, Gland
- Phelps J, Carrasco LR, Webb EL (2014) A framework for assessing supply-side wildlife conservation. *Conserv Biol* 28:244–257
- Rauh W (1998). Succulent and xerophytic plants of Madagascar. Strawberry Press, Mill Valley, CA
- Ravaomanalina BH, Rakotonavalona AN, Rakouth B (2011) Conservation status of some commercialized succulent species of Madagascar. *Malagasy Nature* 5:59–67
- ✦ Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33:1–39
- RBG Kew (Royal Botanic Gardens Kew) (2016) The state of the world's plants report. Royal Botanic Gardens Kew, London
- ✦ Sajeва M, Augugliaro C, Smith MJ, Oddo E (2013a) Regulating internet trade in CITES species. *Conserv Biol* 27: 429–430
- Sajeва M, McGough HN, Garrett L, Luthy J, Tse-Laurence M, Rutherford C, Sajeва G (2013b) CITES and cacti: a user's guide. Royal Botanic Gardens Kew, London
- ✦ Sonricker Hansen AL, Li A, Joly D, Mekaru S, Brownstein JS (2012) Digital surveillance: a novel approach to monitoring the illegal wildlife trade. *PLOS ONE* 7:e51156
- Williamson DF (2004) Tackling the ivories: the status of the US trade in elephant and hippo ivory. TRAFFIC North America. World Wildlife Fund, Washington, DC
- Winn JM, Bishop CM (2005) Variational message passing. *J Mach Learn Res* 6:661–694
- Wu J (2007) World without borders: wildlife trade on the Chinese-language internet. *TRAFFIC Bull* 21:75–84

Editorial responsibility: David Roberts, Canterbury, UK

*Submitted: August 23, 2016; Accepted: December 9, 2016
Proofs received from author(s): February 23, 2017*